MONASH

7/90

MONASH UNIVERSITY

AUSTRALIA

A SIMPLE ESTIMATOR FOR SIMULTANEOUS MODELS WITH CENSORED ENDOGENOUS REGRESSORS

Francis Vella

Working Paper No. 7/90

June 1990

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 86746 954 4

A SIMPLE ESTIMATOR FOR SIMULTANEOUS MODELS WITH CENSORED ENDOGENOUS REGRESSORS

Francis Vella

Working Paper No. 7/90

June 1990

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS AND POLITICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors

Francis Vella

Department of Economics
Rice University
Houston, TX, 77251
and
Department of Econometrics
Monash University
Clayton, Victoria 3168

July 1989
This Version June 1990

This paper presents a simple two step estimator for models with censored endogenous regressors and sample selection bias. The approach unifies the literature on censored endogenous regressors and sample selection bias and provides some extensions. The procedure relies upon the use of generalized residuals to adjust for the inconsistency caused by the endogeneity of the censored regressors. The methodology is very simple and easily implementable with existing computer programs. The paper also provides two tests of weak exogeneity. Two empirical examples, based on issues from the labor economics literature, indicate that the estimation procedure and the tests perform well.

JEL Classification Nos: 211,821

Keywords: Endogenous censored regressors, sample selection bias, generalized residuals, conditional moment restrictions, compensating differentials.

I am very grateful to Tom Cooley and Adrian Pagan for their encouragement. I also wish to thank Adrian Pagan, Whitney Newey, Trevor Breusch, Andrew Chesher, Tony Hall, Roger Klein, Siu Fai Leung, Jim Powell, Chris Skeels and Chris Alouze for their comments on earlier versions. I alone am responsible for any remaining errors.

1. Introduction.

Endogenous censored regressors and sample selection bias are frequently encountered problems in estimating econometric models from unit record data. However, while both are generated by the same underlying mechanism they have been typically analyzed in separate frameworks. Simultaneous systems with limited dependent variables were initially treated by Amemiya (1978,1979), Heckman (1978), Lee (1978,1979), Nelson & Olson (1978), and later Newey (1987), in an instrumental variable framework while Smith & Blundell (1986), Rivers & Vuong (1988) and Blundell & Smith (1989) have recently employed a conditional maximum likelihood approach. Estimation from non-random samples was pioneered by Heckman (1974,1976,1979) and Lee (1978) through a full information likelihood framework and subsequently employing a two step estimator.

In this paper we discuss an approach which encompasses these two aspects of microeconometrics. In doing so we provide an unifying framework for the various available estimators and present some extensions. The system we consider comprises a structural equation, of primary interest, where one or more of the endogenous regressors are censored while the remaining equations represent the reduced forms of the censored variables. We discuss estimation procedures for several models including; A) the estimation of the parameters over the entire data set and B) where the observations in the primary equation can be systematically sorted into sub samples. An example of the first is evaluating the impact on wages of union status, job choice and fringe benefits while an example of the second is the effect of personal characteristics on wages, conditional on a particular job choice.

Estimation in this family of models is complicated by the presence of endogeneity. For example, in the union status case, it is valuable to establish whether union status is weakly exogenous, in the Engle, Hendry & Richard (1983) sense, to wages. This is not only useful for hypothesis testing but also implies that the use of a dummy variable for union status will often lead to consistent estimates. Accordingly an easily implementable test of such a proposition would be useful.

The objectives of this paper are the following. First we develop a simple two step estimator for models where the endogenous regressors in the structural equation are censored or the observations for the structural equation can be sorted according to the value of the endogenous regressors. The approach provides consistent estimates of the reduced form and structural equation parameters and produces a test of endogeneity. Second, we introduce an easily implementable alternative test of endogeneity based upon the methodology of Newey (1985) and Tauchen (1985) which focuses on the conditional moments implied by the model. This approach to diagnostic tests in the limited dependent variable framework is discussed at length in Pagan & Vella (1989).

The procedure that follows has several features to commend it over existing estimators. First, the estimator is easily implementable and requires little additional computation above estimating a sequence of equations. This is a major attraction as it is often difficult to estimate these models by maximum likelihood and frequently impossible with existing computer programs. Second, it produces a test of weak exogeneity. Finally, unlike the existing estimators in the literature, the method can be easily extended to other forms of censoring and selection bias. Thus it not only unifies a growing literature but also provides a link between estimators. The following section discusses the general model and the estimation procedure. We derive the sample selection procedures of Heckman (1979), Barnow, Cain & Goldberger (1981) and Garen (1984) as special cases of the general model and introduce some new extensions. Section three discusses the endogeneity tests from this framework and presents the conditional moment test. In section four we discuss the extensions to account for departures from normality. Section five presents two empirical examples. First we examine a model with an endogenous censored regressor and the example presented is the effect of fringe benefits on wages where a non-trivial proportion of individuals in the sample report receiving no fringes. The second example is that of polychotomous selectivity where adjustment is made for the amount of time spent in the labor market. A wage equation is estimated correcting for the bias introduced by individuals revealing varying degrees of labor market commitment. Concluding comments are presented in section six.

2. The General Model.

Consider the following M equation model comprising of one structural equation and M-1 reduced form equations

$$w_i=\alpha' X_i+\sum_{j=1}^{M-1}\beta_j y^*_{ji}+\sum_{j=1}^{M-1}\psi_j y_{ji}+\sum_{j=1}^{M-1}\delta_{jk} d_{jki}+u_i \qquad i=1..n:\ k=1..K \quad (1)$$

$$y^*_{ji}=\gamma_j' Z_i+v_{ji} \qquad i=1..n;\ j=1.M-1 \quad (2)$$

where w_i is the dependent variable in the equation of primary interest; y^*_j are unobserved endogenous variables; y_j are censored variables and d_{jk} are indicator functions determined by the values of y^*_j; Z is a vector of exogenous variables, of which X is a subset, observed for the n individuals in the sample; $\alpha, \beta, \psi, \delta$ and γ are parameters to be estimated; and the u's and v's represent zero mean error terms.

Assumption 1: The latent variables are censored by the functions h_j such that the variables y_{ji} are observed.

$$y_{ji} = h_j(y^*_{ji}) \tag{3}$$

Assumption 2: The latent variables can be assigned into various groups by the following indicator functions

$$d_{jki} = 1 \text{ iff } I_A(y^*_{ji}) \text{ and } d_{jki} = 0 \text{ otherwise} \tag{4}$$

where I_A denotes the occurrence of the event A. Note that it is possible for each latent y to generate a sequence of binary dummy variables.

Assumption 3. The triplet (Z_i, u_i, v_{ji}) are independently and identically distributed.

Assumption 4: u_i and the v_{ji}'s are, conditional on Z_i, jointly normal with zero means and covariance matrix

$$\begin{bmatrix} \sigma^2_u & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix}$$

where Σ_{vv} is a diagonal matrix.

Assumption 5: The parameters of the model are identified up to some normalization. This may require the imposition of exclusion restrictions on the number of variables from each censored regressor which may appear in the structural equation. For example, the logical consistency requirement of Heckman (1978) must be satisfied. (Note, where the censoring produces an indicator function we cannot separately identify δ and ψ).

This model clearly has nested within it several familiar models. Consider some of the more interesting noting that we limit our discussion to where y^* is a scalar although the extensions to the vector case is obvious.

Case 1: $\beta \neq 0$; $\psi = 0$; $\delta = 0$.

Here we are interested in the effect of the latent variable on the dependent variable. There is no emphasis on the observed version of the latent variable nor is there any structural shift across groups corresponding to different values of y^*_{ji}.

Case 2: $\beta \neq 0$; $\psi = 0$; $\delta \neq 0$.

This is case 1 with structural shift. We allow the observations in the various sub samples to have different intercepts.

Case 3: $\beta = 0$; $\psi \neq 0$; $\delta = 0$.

This case is interested in the effect of the censored variable upon the dependent variable. There is no structural shift.

Case 4: $\beta = 0$; $\psi \neq 0$; $\delta \neq 0$.

Case 3 with structural shift.

Case 5: $\beta \neq 0$; $\psi = 0$ **for sub-sample corresponding to** $d_{ji} = 1$.

This case focuses on the effect of the latent variable upon the dependent variable **conditional** on the observation having a value of y^*_{ji} in some range.

Case 6: β=0; $\psi\neq$0 for sub-sample corresponding to d_{ji}=1.

This is similar to case 5 except we focus on the effect of the censored variable upon the dependent variable.

Case 7: β=0; ψ=0 for sub-sample corresponding to d_{ji}=1.

In this model we are interested in the effect of the X's upon the dependent variable conditional on the latent variable falling in a particular range.

While these seven cases are not exhaustive they are quite extensive and include a number of interesting models. To motivate their relevance consider each in light of the empirical examples we explore below. Suppose we wish to estimate the effect of fringe benefits upon wages noting that for many observations there are no reported values on the actual level of fringes. That is, a level of fringe benefits that falls below some threshold will be reported as a zero. Also note that it is likely that fringes and wages are simultaneously determined and estimating the wage equation without accounting for this will lead to inconsistent estimates. Case 1 represents the trade off between the **actual (latent)** fringes received and the wage recalling that the actual fringe level is not always observed. Case 2 allows a structural difference between those who receive fringes and those who do not. The third case indicates that the appropriate trade off is between the **observed** fringe benefits received and the wage while Case 4 augments this with an intercept dummy to capture structural differences. Case 5 focuses on the trade off between the actual fringes and the wage **conditional** on the individual receiving a specified level of fringes while case 6 examines the trade off between the observed fringes and wages conditional on a level of fringe being received. Case 7 focuses on the sub samples of the entire data set where attention is upon the

α's conditional on y^*_{ji} being in some range.

The model also nests instances of polychotomous discrete choice such as occupational choice. In that instance y^*_{ji} reflects the level of utility from various occupations and d_{jki} represents a vector of binary dummy variables denoting whether an individual is a member of a particular occupation. This is an important case of these models as the censoring often produces a variable which cannot be employed or enable the computation of the expectation of the latent variable. For example, often in occupational choice data the censoring produces different numerical values for the different occupations. The use of such a variable would impose an unrealistic structure upon the latent variable.

To obtain consistent estimates in this model first take expectations conditional on the vector of observed values of $y_{\sim ji}$, noting that the subscripted tilde denotes a vector.

$$E(w_i|y_{\sim ji})=E(\alpha' X_i|y_{\sim ji})+E(\sum_{j=1}^{M-1}\beta_j y^*_{ji}|y_{\sim ji})+E(\sum_{j=1}^{M-1}\psi_j y_{ji}|y_{\sim ji})+ E(\sum_{j=1}^{M-1}\delta_{jk} d_{jki}|y_{\sim ji})$$

$$+ E(u_i|y_{\sim ji}) \qquad i=1..n \qquad (5)$$

$$E(y^*_{ji}|y_{\sim ji})=E(\gamma_j' Z_i|y_{\sim ji})+E(v_{ji}|y_{\sim ji}) \qquad i=1..n;\ j=1..M-1 \quad (6)$$

The conditional error terms are dependent on the values of y_{ji} and can be described as generalized errors in the sense of Cox & Snell (1968). Denote these generalized errors as υ_i and υ_{ji} noting their values are dependent upon the form of the censoring functions h_j. Employing our assumption of joint normality and the law of iterated expectations rewrite υ_i in the following manner:

$$E(E(u_i|\underset{\sim}{v}_{ji})|\underset{\sim}{y}_{ji})=\Sigma_{uv}\Sigma_{vv}^{-1}E(\underset{\sim}{v}_{ji}|\underset{\sim}{y}_{ji}) \tag{7}$$
$$=\Sigma_{uv}\Sigma_{vv}^{-1}\underset{\sim}{\upsilon}_{ji}$$
$$=\lambda'\underset{\sim}{\upsilon}_{ji}$$

where λ is a (M-1) x 1 vector with λ_j as the j^{th} element. Now rewrite the estimable form of the structural equation as

$$w_i=\alpha' X_i+ E(\sum_{j=1}^{M-1}\beta_j y^*_{ji}|\underset{\sim}{y}_{ji})+\sum_{j=1}^{M-1}\psi_j y_{ji}+\sum_{j=1}^{M-1}\delta_{jk}d_{jki}+\lambda'_j\underset{\sim}{\upsilon}_i+\eta_i \tag{8}$$

as we can directly replace the expectation of the censored variables and the indicator functions with their observed values. Further note that η_i is a zero mean error term which is uncorrelated with the regressors by construction. We can now estimate the parameters by least squares after obtaining estimates of the remaining expectations. Thus the procedure is as follows. First estimate the M-1 reduced form equations to obtain consistent estimates of γ_j by maximum likelihood using the observed values of y_{ji} in place of y^*_{ji}. The form of the likelihood functions will be determined by the nature of the censoring functions h_j. Employing these estimates of γ_j we compute estimates of the generalized errors and the expected values of the latent variables and insert them into the structural equation as additional regressors[1] and estimate the remaining parameters by least squares[2].

In general the distribution of η_i will not be normal, or in fact, even known, and the conditional maximum likelihood approach of Smith & Blundell (1986), Rivers & Vuong (1988) and Blundell & Smith (1989) will not be applicable. That approach is appropriate where y^*_i is uncensored, producing generalized residuals that coincide with OLS residuals and values of η_i which are also normally distributed[3]. The intractability of the distribution of η_i is a substantial constraint as it restricts the dependent variable in the structural equation to be uncensored to enable estimation[4].

Implementation of this procedure requires estimates of the generalized errors and the latent variable. The generalized residuals are obtained through the results of Gourieroux, Monfort, Renault & Trognon (1987) which show that the best prediction of the error is the score with respect to the intercept, for each observation, evaluated at the maximum likelihood estimates[5]. An outline of their results, and the derivation of the generalized residuals for the models discussed in this paper, are contained in the Appendix. To obtain the expected value of the latent variable we evaluate the first moment of the truncated normal distribution where the form of truncation depends upon the censoring function h_j.

Before examining some less conventional forms of censoring first consider the most common case and how the above procedure produces the two step estimator of Heckman (1979) and the selectivity bias estimator of Barnow et.al (1981). The model has the following two equation representation and can be treated as either an example of cases 3 or 7 above noting that y_i is now a scalar as we have set M equal to 2.

$$w_i=\alpha' X_i+\psi y_i+u_{1i} \qquad i=1..n \tag{9}$$

$$y_i^*=\gamma' Z_i+u_{2i} \qquad i=1..n \tag{10}$$

where the u_i's are normally distributed error terms with zero means, variances σ_1^2 ,σ_2^2 and covariance σ_{12}. The censoring takes the form

$y_i=1$ if $y_i^*>0$

$y_i=0$ otherwise

An appropriate estimation procedure for estimating the parameters from equation (10) is probit. The generalized residuals, see Appendix, are given by

$$\hat{\upsilon}_{2i}=E(v_{2i}|y_i)=(y_i-\hat{\Phi}_i)\hat{\phi}_i(1-\hat{\Phi}_i)^{-1}\hat{\Phi}_i^{-1} \quad i=1..n \tag{11}$$

where $\hat{\Phi}$ and $\hat{\phi}$ are the cumulative distribution function and probability density function of the standard normal distribution evaluated at the probit estimates of $(\widehat{\gamma/\sigma_2})$. From joint normality rewrite u_{1i} conditional on the observed value of $\hat{\upsilon}_{2i}$. That is

$$E(u_{1i}|y_i)=\lambda\hat{\upsilon}_{2i} \quad i=1..n$$

where λ is equal to σ_{12}/σ_2^2. Rewrite equation (9) in terms of its conditional expectation

$$E(w_i|y_i)=\alpha' X_i+\psi y_i+\lambda E(\hat{\upsilon}_{2i}|y_i) \quad i=1..n \tag{12}$$

Least squares on the regression form of (12) now produces consistent estimates of α, ψ and λ. Those familiar with the selectivity bias literature will identify this estimator, where X does not contain an intercept and only values of w corresponding to specific values of y are observed, (corresponding to case 7) as Heckman's two step estimator while equation (12) is that proposed by Barnow et.al (1981). It is valuable to see the selectivity bias estimator derived in this manner and note that the inverse mills ratio is the generalized residual for the probit model. It also indicates that the method can be extended to where y_i is a vector.

This approach also produces the continuous selectivity bias estimator of Garen (1984). In that model the dependent variable in the selection equation, (10), is able to take a continuum of values over a given range and is uncensored, corresponding to our case 3. To produce Garen's estimator we estimate (10) by OLS, as it corresponds to MLE, compute the generalized residuals, given by the least squares residuals (see Appendix), and include them as an additional regressor in equation (9)[6,7].

Although these models have appeared elsewhere we feel they are more easily derived in this present framework. Furthermore this methodology can be extended to models with less conventional types of censoring. Consider the model in equations (9) and (10) but with the following censoring

$y_i=y_i^*$ if $y_i^*>0$ and
$y_i=0$ otherwise.

An appropriate means of estimating γ and σ_2 from equation (10) is Tobit and the generalized residuals, see Appendix, take the form

$$\tilde{\upsilon}_{2i}=E(v_{2i}|y_i)= -\tilde{\sigma}_2(1-I_i)\tilde{\phi}_i(1-\tilde{\Phi}_i)^{-1}+I_i\tilde{v}_{2i} \qquad (13)$$

where $\tilde{\gamma}$ and $\tilde{\sigma}_2$ are the Tobit maximum likelihood estimates of γ and σ_2; $\tilde{\phi}$ and $\tilde{\Phi}$ are evaluated at these estimates; $\tilde{v}_{2i}=y_i-\tilde{\gamma}'Z_i$; and I_i is an indicator function taking the value one if y_i is uncensored and zero otherwise. Now substitute the structural equation error term with its conditional expectation plus a zero mean error and estimate the parameters consistently using ordinary least squares. Consistent estimates of these parameters could also be obtained from a regression over the sub sample corresponding to $y_i>0$.

This model also has an alternative interpretation depending on our views regarding the relevant variable that should appear as a regressor. It may be more appropriate that the model should be specified as

$$w_i=\alpha' X_i+\beta y_i^*+u_{1i} \qquad i=1..n$$

with the latent variable appearing rather than the observed variable. The expectation of the structural equation should be written

$$E(w_i|y_i)=\alpha' X+\beta E(y_i^*|y_i)+\lambda E(u_{2i}|y_i)$$

and the regression should include the expectation of the latent variable as a regressor rather than the censored variable. Given this form of censoring the expectation of y_i^* can be written as

$$E(y_i^*|y_i)=I_i\tilde{\gamma}' Z_i+(1-I_i)\{\tilde{\gamma}' Z_i-\tilde{\sigma}_2\tilde{\phi}_i(1-\tilde{\Phi}_i)^{-1}\} \qquad i=1..n \qquad (14)$$

These examples illustrate estimation of α,β and ψ. Consider where the values of α are of primary interest and individuals can be sorted by their observed value of y_i. This corresponds to our case 7. For example, consider the following model

$$w_{ji}=\alpha_j'X_{ji}+u_{1i} \qquad i=1..n;\ j=1..k \tag{15}$$

$$y^*_{ji}=\gamma'Z_i+u_{2i} \qquad i=1..n;\ j=1..k \tag{16}$$

and the selection rule is

$$d_{ji}=1 \quad \text{if } y^*_{ji}>y^*_{pi} \quad \text{for all } j\neq p$$
$$d_{ji}=0 \text{ otherwise}$$

The value of w_j is only observed for the category of j chosen by individual i where the k different categories may or may not have some natural ordering. We can identify the category type for each individual by the k indicator functions. Taking expectations conditional on these indicator functions gives

$$E(w_{ji}|y_{ji})=\alpha_j'X_{ji}+E(u_{1i}|y_{ji}) \tag{17}$$

We now need the generalized residuals. If the k categories have no underlying order we estimate γ by multinomial probit while if some natural ordering does exist we employ ordered probit. In both instances the generalized residuals take the following form

$$E(u_{2i}|d_{ji})=d_{ji}\hat{\pi}_{ji}\hat{\Pi}_{ji}^{-1}(1-\hat{\Pi}_{ji})^{-1}(d_{ji}-\hat{\Pi}_{ji}) \tag{18}$$

where d_{ji} is an indicator function taking the value 1 if individual i is in category j and zero otherwise; $\hat{\Pi}_{ji}$ is the estimated probability that individual i is in the j^{th} category while $\hat{\pi}_{ji}$ is the estimated value of the density at that point. As shown in the appendix equation (18) represents the scores of the respective likelihood functions with respect to the intercept. Note however that the probabilities Π_{ji} will, in general, differ depending on whether some ordering of categories is imposed upon the model. We can now obtain consistent estimates of α_j by estimating k separate regressions over the sub-samples corresponding to $d_{ji}=1$ and including the generalized residuals as a regressor. Similarly, if we are interested in shift differences across groups we can estimate one regression and include, along with the generalized residuals, k-1 dummy variable reflecting group type[8].

These two new models illustrate the wide applicability of the proposed approach and how the techniques employed for the more conventional types of models can be easily applied to the less conventional cases. While we explore only two new types of censoring it is apparent that the model can be easily adapted to various other forms of censored variables such as the various Tobit models discussed in Amemiya (1984). The model can also be extended to handle different types of selectivity bias as well as multiple selection rules. Further, the model can also be applied to structures where there are multiple endogenous explanatory variables generated by different censoring functions.

3. Tests of Endogeneity.

A feature of the above models is that estimation is complicated by the endogeneity. Furthermore in examining the economic behavior of agents it is often of interest to establish whether particular explanatory variables can be treated as exogenous to the variables of primary interest. One test of this proposition in the above framework is to examine whether λ_j is equal to zero as this is the parameter which captures the correlation between the structural equations error and the

j^{th} reduced form equation's error. As it is possible to obtain a consistent estimate of λ_j we need to now derive an estimate of it's variance. As this class of model is a member of the sequential method of moments models examined by Newey (1984) we can estimate the covariance matrix in the manner outlined there, and in Pagan (1986), adjusting for the heteroskedasticity arising in the first step[9]. If we only wish to evaluate the statistic under the null that the λ_j's are equal to zero we can use the standard errors from the original output as under the null hypothesis the uncertainty introduced from having to estimate the generalized residuals disappears. However, by computing the standard errors under the alternative we are able to perform tests of endogeneity on the individual coefficients with the assurance that both the estimates of the coefficients and the standard errors are consistent.

This test of endogeneity is evaluated while accounting for the correlation that exists between equations. This is precisely the approach adopted in the conditional maximum likelihood literature although in Rivers & Vuong (1989) and Smith & Blundell (1986) they estimate the standard errors under the null. An alternative approach is to perform and evaluate the test under the null hypothesis that the correlation is equal to zero. By doing so we are able to estimate each of the equations by maximum likelihood as the distribution of the error terms is known. We then develop a test in the conditional moment framework of Newey (1985) and Tauchen (1985) and discussed in relation to limited dependent variable models by Pagan & Vella (1989)[10]. As the methodology of these tests is discussed at length in the above mentioned papers it is inappropriate to do so here. However for the sake of motivating the test a brief review, in the context of the current example, will be given.

In the case of weak exogeneity the population values of the elements of Σ_{uv} will be equal to zero. Thus a relevant test of such a proposition would be to examine the sample estimate of each of the elements of Σ_{uv}. This can be easily shown to be equal to $\hat{\tau}_{1j}=n^{-1}\Sigma\hat{\upsilon}_{1i}\hat{\upsilon}_{ji}$ $j=1..M-1$ (see Pagan & Vella (1989)) where the $\hat{\upsilon}$'s represent the estimated values of the generalized residuals computed under the null of weak exogeneity. The difficulty now lies in deriving the distribution of $\hat{\tau}_{1j}$. This is done by employing the results of Newey (1985), Tauchen (1985) and the methods outlined in Pagan & Vella. These papers show that it is possible to test the restriction that τ is equal to zero by regressing $\hat{\tau}_i$ against the scores for the model and an intercept. The t-test on the intercept being different from zero represents a test of whether τ is equal to zero[11]. A joint test on the τ's being jointly equal to zero would require joint estimation of these auxillary regressions and a joint tests on the intercepts. Note while the first test, based on the t statistics in the structural regression, is robust against Σ_{vv} being non-diagonal the conditional moment test is not.

4. Extensions for Non-normality.

Thus far we have relied heavily upon our assumption of normality in two places. First, it determines the form of our likelihood function in the estimation of the γ's. It also features in the manner we represent our structural error as a linear combination of the reduced form errors. As the generalized residual results are applicable to the exponential family we are not restricted to normality for this reason. If we wish to extend our results to the case of non-normality we need to make the assumption directly that

$$E(u_i | v_{ji}) = f(v_{ji}) \tag{19}$$

where f denotes a function mapping the reduced form residuals into the structural equation error. In the case of joint normality explored above the f function is a linear mapping and this allows us to substitute the structural error with a linear combination of the reduced form errors. Now consider where the errors are not jointly normal. One approach is to borrow results from the semi-parametric literature, see for example Gallant & Nychka (1987), and the diagnostic testing literature (see Lee (1984) and Pagan & Vella (1989)). In those papers it is argued that departures from normality can be approximated by multiplying the normal distribution by some suitably designed polynomial. This can be employed here by capturing the departures from normality by including higher order terms in the generalized residuals as shown in (20).

$$E(u_i | v_{ji}) = \sum_{q=1}^{Q} \sum_{j=1}^{M-1} \lambda_j^q \upsilon_{ji}^q \tag{20}$$

Accordingly, we adjust the procedure as follows. We make our distributional assumptions regarding v and estimate γ by MLE. Providing the assumed distribution is in the exponential distribution the generalized results are still applicable. We then assume that the structural error term can be approximated by some linear combination of a higher order polynomial in the generalized residuals.

This now extends the available estimators in the first step to those beyond normality while also relaxing the constraint that the structural error is normally distributed. The exogeneity tests still apply although the conditional moment tests require distributional assumptions about the structural equations error.

5. Applications.

To illustrate the methodology discussed above we present two examples. The first examines the trade off between wages and fringe benefits. The second considers the possibility of trichotomous selectivity bias where wage equations are estimated accounting for the varying degrees of individuals commitment to market work.

A feature of the compensating differential literature in labor economics, of which fringe benefits is a special case, is the inability to find the expected relationships in the data. For example, consider the following equation where the objective is to establish the trade off between wages and fringe benefits

$$\text{hourly wage} = \alpha + \sum\alpha_j\text{*personal characteristics} + \sum\alpha_l\text{*region dummies} + \sum\alpha_i\text{*industry dummies} + \alpha_f\text{*hourly fringe benefits} \qquad (21)$$

It is likely that the level of fringe benefits is determined simultaneously with wages so the problem of endogeneity is obvious. Furthermore, many individuals report receiving no fringe benefits and as reported fringe benefits are strictly positive the level of fringes is censored at zero.

Prior to estimation consider the expected sign of α_f. Most theoretical models in labor economics, for example those presented in the compensating differential literature, unambiguously predict that a negative relationship exists between wages and fringes although the size of the trade off is not clear. The intuition behind this result is the following. Individuals facing an overall level of financial compensation can choose to receive it either directly in pay or in the form of fringe benefits. This may represent some desire to avoid higher tax rates or simply may reflect the preferences of the worker. However as the total value of compensation is fixed the worker must trade off fringes for pay thus producing a negative relationship between the two. Empirical attempts to establish such a relationship have failed miserably. For example the work of Smith & Ehrenberg (1983), Leibowitz (1983), Kuehneman (1986) and Yakaboski (1988) all present theoretical models predicting a negative relationship between wages and fringes but produce empirical results indicating a positive relationship. We argue that these models fail due their inability to adequately account for the simultaneity.

To estimate equation (21) we employ data constructed by matching the 1977 Quality of Employment Survey with the 1977 Employer Expenditures for Employee Compensation Survey. This produces a data set which has information on individuals' earnings, receipt of fringe benefits, personal characteristics and work place characteristics[12]. The variables employed are described in Table 1.

The first step of the estimation procedure is to estimate the reduced form equation of the fringe benefit receipts. This takes the form

$$\text{hourly fringe benefits} = \beta + \sum \beta_j * \text{personal characteristics} + \sum \beta_l * \text{region dummies} + \sum \beta_f * \text{industry dummies} \qquad (22)$$

and assuming the error for this equation is normally distributed we can estimate the β's by Tobit[13]. Following the estimation of equation (22) we employ the estimates of β and σ, reported in column (1) of Table 2, to compute the generalized residuals. We insert the generalized residuals into equation (21) as an additional regressor and obtain consistent estimates of the α's by OLS. Further, the t-test on the coefficient of the generalized residuals is a test of weak exogeneity.

To examine the fringe\wage trade off we first estimate equation (21) without entering the generalized residuals as a regressor. These results are reported in column 2 of Table 2 and an inspection of this table reveals that the coefficient on the fringe benefits variable is highly significant and positive[14]. As noted above this violates the expectations generated by conventional models in labor economics. Following the estimation of the reduced from and the calculation of the generalized residuals equation (21) was re-estimated and the results are reported in column 3 of Table 2[15]. The coefficient on fringe benefits continues to be statistically significant at conventional levels of confidence but now displays the expected negative sign. The coefficient on the correction factor, FGRES, is also significant indicating, as expected, that the level of fringe benefits is endogenous to the wage determining process. Furthermore the positive coefficient on this variable indicates that the unobserved factors that result in a high level of fringe benefits are also producing a higher level of wages. This is an important result as it is clearly this relationship that is dominating previous attempts to estimate this fringe/wage trade off.

As noted in section 2 it is not obvious which trade off captures the relationship between fringe benefits and wages. That is, is the appropriate relationship that between the observed level of fringes and wages or that between the expectation of the level of fringes and the wage? To investigate this we re-estimate (21) but replace the censored values of the fringes with their expectation computed from equation

(14). These results are reported in column 4 of table 2. Again they reflect that after accounting for the endogeneity of fringes the effect of fringes upon wages is negative. To compare the meaning of these coefficients we calculate the implied trade off from columns 3 and 4 for a white, married male individual who is 37 years of age, with 12 years of education, is an office worker living in the western region. For column 3 we assign them a hourly fringe benefit rate of 91 cents while for column 4 they are given an expected value of $1.08. The expected wage rates from this two columns for a worker with these characteristics is $6.62 and $6.48 respectively. Now we evaluate the impact upon each wage rate of increasing the hourly rate of fringe benefits by $1.00. The effect upon the column 3 wage rate is to reduce it by approximately $2.40 to $4.18. This would appear to be an unreasonably large decrease although it may be acceptable for those facing very high marginal tax rates. Now consider the column 4 wage. Increasing the level of hourly fringe benefits by $1 decreases the hourly wage rate by about $1 to $5.51. This appears to be a more acceptable trade off and may indicate that the appropriate regressor should be the expectation of the latent variable and not the censored value.

Now focus upon the estimation of wage equations with trichotomous selection bias. The proposition that estimating wage equations over a sample of working women will lead to biased parameter results is perhaps the most empirically supported argument in labor economics. This, of course, results from the systematic self selection of individuals into the work on not work category. However it is not clear that this dichotomous characterization of market work behavior is satisfactory as there exist varying degrees of involvement in the labor force by females. For example, the fixed costs labor supply model of Cogan

(1981) predicts that the cost of market work involvement encountered by each individual will affect the minimum number of hours they are willing to work. Accordingly it is possible that some "selection bias" is contained in the sub sample of working women. To investigate this possibility we examine data on females aged between 15 and 26 years living in the two most populous states in Australia, namely Victoria and New South Wales. The data refer strictly to women who have left school and are taken from the 1985 wave of the Australian Longitudinal Survey.

To explore the degree of labor force participation we first examine the distribution of working hours. This revealed that the majority of women either worked zero hours or worked over 35 hours per week. The remainder of the sample, comprising about 10 per cent of the data, were uniformly distributed over the interval 1-35 hours. This suggests that the usual dichotomy of work/not work is inadequate as there appears to be at least three types of labor force commitment[16]. Accordingly we will refer to those who work zero hours as non workers; those who work one to thirty five hours as part-time workers; and those who work above thirty five hours as full-time workers.

To investigate this possibility of trichotomous selection we first estimated the following simple wage regression over the sample of workers.

$$\text{log hourly wage} = \alpha + \sum \alpha_j {*}\text{personal characteristics} \qquad (23)$$

The variables employed are described in Table 3 and equation (23) was initially estimated adjusting for the possibility of selection bias resulting from the work/not work decision. The adjustment took the form of the Heckman two step correction after estimating the reduced form equation explaining the work decision. The results from the reduced form probit are reported in column (1) of Table 4 and the results from estimating equation (23), reported in the first column of Table 5, provide the expected finding that selection bias is present.

To investigate the further possibility of selection bias amongst the working women we adopted the following two strategies. First we include, in the adjusted equation, a dummy variable, denoted FT, indicating whether an individual worked full-time or part-time. These results are shown in column (1) of Table 5 and the statistically significant coefficient on this variable indicates some difference between the two groups although it does not indicate whether the decision to participate full-time or part-time is weakly exogenous to wages. The second approach was to ignore the existence of selection bias in the work/not work decision and to perform the selection correction over the sub-sample of workers adjusting for the degree of participation. These results are reported in column (3) and provides some evidence that the part-time/full-time decision[17] is not weakly exogenous to wages[18].

This evidence suggests the methods described in section 2 are appropriate. As the dependent variable in the censoring equation has a natural ordering (not work=0; part-time work=1; full-time work=2) we can employ the ordered probit method of McKelvey & Zavoina (1975) to estimate the reduced form parameters. These are reported in column (3) of Table 4. Employing these estimates we calculate the generalized residuals according to equation (18) and insert them as an additional regressor in equation (23). We then re-estimated this wage equation over the sub-sample of workers accounting for the endogeneity of FT. We also estimated the wage equations over the sub-samples of full-time and part-time workers while accounting and testing for this possibility. The results confirm our suspicion that the degree of market work involvement is endogenous to the level of wages and adopting the two step estimator of Heckman over the whole sample of workers will lead to biased estimates. The major effect of the bias appears to be reflected in the parameters on the education variable and the shift variable FT[19]. It appears that adopting the simpler approach encourages misleading inferences regarding the differential part-time workers receive.

It should be noted that the coefficient on the correction factor for the part-time wage equation is not significant at conventional levels of statistical significance. This appears to be primarily attributable to the very small coefficient on this variable compared to the corresponding value for the full-time group. A closer examination of this part-time equation however reveals that all of the parameters are very different to those of the full-time group. It would appear on the basis of this that the two markets operate in quite different manners in determining wages. This provides even stronger evidence for employing the approach outlined above and suggests a greater need to examine the operation of these markets.

To further investigate the possibility of endogeneity in the above models the conditional moment tests discussed in section 3 were performed. The results for both models are reported in Table 6.

For both models the conditional moments were evaluated under the null hypothesis of weak exogeneity. The conditional moment test was first performed on the fringe benefit data set. The resulting value of the t-statistic on the intercept is 4.635 reinforcing our prior of the endogeneity and supporting the finding of the alternative test. The tests were also performed for the working female data set and similar results were obtained although the t-statistic was surprisingly small for the working women sub-sample.

6. Conclusions.

The objective of this paper is to provide a simple consistent estimator for simultaneous models with censored endogenous explanatory variables. The method developed employs the use of generalized residuals as a means of adjusting for the inconsistency caused by the endogeneity. The approach is applicable to various forms of censoring and is also capable of handling unconventional forms of selection bias. In this sense the paper provides an unifying approach to two areas of the econometric literature which have been considered separately.

Two simple tests of endogeneity are also provided. The first can be derived directly from the estimation procedure and requires no additional computation. The second is derived in the conditional moment framework and relies on directly testing particular sample moment values implies by the model under the null hypothesis of correct specification. While this test requires some additional computation, in that scores from the model need to be evaluated, it may often be the case that it is easier to evaluate the correlation prior to performing the adjustment procedure.

The empirical examples presented provide encouraging results for both the estimation and testing procedure. In both cases the results were consistent with prior reasoning and the resulting parameters were of acceptable magnitude.

Finally it should be noted that while the analysis has been discussed purely in a cross section framework the procedure is appropriate for many time series orientated empirical questions. These include, for example, analyses of income policies and other various goverment policies which are often measured by indicator functions.

TABLE 1: Variables used in Fringe Benefit Analysis

Variable Name	Definition	Mean
MALE	Is individual male: yes=1 no=0.	.64
MAR	Is individual married:yes=1 no=0	.69
AGE	Individuals age (years)	37.65
RACE	Is individual white: yes=1 no=0	.91
EDUC	Individual's years of education	12.51
OFFICE	Does individual work in office: yes=1 no=0	.48
LPAY	Log of hourly wage rate ($)	1.68
HFRINGE	Hourly level of fringe benefits ($)	.91

TABLE 2: Reduced Form Tobit & Structural OLS

	Dependent Variable			
	HFRINGE	LPAY	LPAY	LPAY
CONSTANT	-.775*	.538*	-.057*	.162
	(.213)	(.113)	(.193)	(.130)
AGE	.010*	.003*	.012*	.011*
	(.002)	(.001)	(.002)	(.002)
MAR	.129*	.033	.138*	.100*
	(.057)	(.031)	(.035)	(.027)
RACE	.088	.069	.145*	.117*
	(.081)	(.047)	(.045)	(.041)
MALE	.287*	.246*	.485*	.408*
	(.062)	(.034)	(.059)	(.037)
EDUC	.072*	.030*	.091*	.071*
	(.010)	(.006)	(.015)	(.008)
OFFICE	----	.064	.070	.070
		(.034)	(.037)	(.038)
HFRINGE	----	.437*	-.457*	----
		(.025)	(.183)	
EHRINGE	----	----	----	-.163*
				(.062)
FGRES	----	----	.862*	.403*
			(.174)	(.031)
σ	.564	----	----	----
Log-likelihood	-529.14	----	----	----
$\bar{R}^2$	----	.599	.614	.614
Observations	616	616	616	616

NOTES: i) Standard errors are reported in parentheses.
ii) All three models include dummy variables controlling for region and industry type.
iii) FGRES denotes the generalized residuals computed from the reduced form results
iv) EHFRINGE denotes the expectation of the hourly level of fringe benefits computed from the reduced form results.
v) * denotes significance at 5% level.

TABLE 3: Variables used in Female Wage Equation Analysis

Variable Name	Definition	Mean
AGE1	Individual aged 15-17 years	.11
AGE2	Individual aged 18-20 years	.32
AGE3	Individual aged 21-23 years	.35
AGE4	Individual aged 24-26 years	.22
CIT	Individual lives in city	.75
MAR	Individual has legal or defacto spouse	.29
HLT	Individual has work limiting disability	.09
EDUC	Individual's years of schooling	11.48
ENG	Individual speaks english well	.89
CHLD	Individual has child/children	.19
SPINC	Spouse's weekly income ($): (for MAR=1)	300.3
TINC	Total family weekly income ($)	219.8
WORK	Individual engaged in market work	.69
FT	Individual works > 34 hours per week	.57
PT	Individual works 0-34 hours per week	.12
LPAY	Log of hourly wage rate ($): (for WORK=1)	1.89
UNION	Individual in union: (for WORK=1)	.44
GOVT	Individual employed by government: (for WORK=1)	.30
ASSIS	Individual employed under employment scheme: (for WORK=1)	.07
TIME	WORK + FT	

TABLE 4: Reduced Form Probit and Ordered Probit Equations

	Dependent Variable		
	WORK	FT	TIME
Constant	-.930*	1.361*	-.447
	(.289)	(.393)	(.232)
AGE1	-.581*	-.590*	-.614*
	(.146)	(.191)	(.131)
AGE2	-.224	-.240	-.222*
	(.119)	(.147)	(.103)
AGE3	-.233*	-.048	-.170
	(.110)	(.137)	(.095)
CIT	.282*	.062	.239*
	(.082)	(.106)	(.073)
MAR	-.541*	.003	-.500*
	(.161)	(.269)	(.139)
HLT	-.510*	-.158	-.469*
	(.115)	(.995)	(.099)
ENG	.245*	.056	.184
	(.118)	(.160)	(.099)
EDUC	.148*	-.012	.110*
	(.021)	(.026)	(.017)
CHLD	-1.786*	-1.043*	-1.769*
	(.115)	(.188)	(.106)
SPINC	.002*	-.0003	.0015*
	(.0004)	(.0007)	(.00003)
TINC	.00003	.0001	.00006
	(.00005)	(.00008)	(.00004)
MU(1)	----	----	.428*
			(.028)
Log-Likelihood	-759.15	-515.21	-1286.60
Observations	1715	1715	1715

TABLE 5: Female Wage Regressions: Dependent Variable LPAY

	All Workers (1)	All Workers (2)	Full Time (3)	Full Time (4)	Part Time (5)	Part Time (6)
Constant	2.054*	1.990*	1.897*	1.984*	1.904*	1.919*
	(.077)	(.065)	(.064)	(.064)	(.231)	(.241)
AGE1	-.687*	-.685*	-.676*	-.688*	-.653*	-.653*
	(.031)	(.032)	(.034)	(.030)	(.074)	(.093)
AGE2	-.348*	-.345*	-.358*	-.362*	-.274*	-.272*
	(.022)	(.022)	(.022)	(.023)	(.074)	(.074)
AGE3	-.068*	-.069*	-.080*	-.077*	-.029	-.026
	(.021)	(.022)	(.022)	(.022)	(.074)	(.074)
EDUC	.018*	.019*	.020*	.016*	.030	.031
	(.005)	(.005)	(.004)	(.004)	(.016)	(.019)
ENG	-.045	-.043*	-.052*	-.052*	.005	-.005
	(.025)	(.025)	(.026)	(.022)	(.080)	(.083)
GOVT	.102*	.101*	.099*	.098*	.163*	.164*
	(.017)	(.018)	(.017)	(.017)	(.072)	(.074)
UNION	.023	.023	.015	.016	.066	.067
	(.016)	(.016)	(.016)	(.016)	(.051)	(.053)
ASSIS	-.062*	-.062*	-.062*	-.061*	-.069	-.068
	(.029)	(.029)	(.028)	(.025)	(.109)	(.112)
FT	-.136*	-.077*	----	----	----	----
	(.021)	(.030)				
LAMBDA	-.068*	----	-.149*	----	-.012	----
	(.030)		(.071)		(.100)	
WGRES	----	-.050*	----	-.067*	----	-.001
		(.022)		(.022)		(.046)
$\bar{R}^2$	.487	.488	.524	.525	.369	.369
Obs	1187	1187	984	984	203	203

NOTES: i) LAMBDA denotes the appropriate Heckman correction
ii) WGRES denotes the generalized residuals

TABLE 6: Conditional Moment Tests for Endogeneity

MODEL	T-STAT for Intercept
Fringe benefits	4.635
Working women (whole sample)	1.421
Working women (full-time)	3.158
Working women (part-time)	.255

Appendix : Generalized Residuals

To illustrate the derivation of the generalized residuals we restate the relevant results of Gourieroux, Monfort, Renault & Trognon (hereafter GMRT) providing the page references for their proofs. The family of models we consider are nested in the exponential family and the log likelihood for the latent variable has the following representation.

$$\text{(A1)}\quad L^{*}(y_i^{*} \mid X_i,\beta)=\sum\{Q(X_i,\beta)T(y_i^{*})+A(X_i,\beta)+B(y_i^{*},X_i)\}$$

where Q,T,A and B are given numerical functions. Following GMRT we give the latent model the following second order representation

$$\text{(A2)}\quad T(y_i^{*})=m(X_i,\beta)+v_i(\beta)$$

where $E[v_i(\beta)]=0$ and $m(X_i,\beta)$ is the mean of $T(y_i^{*})$.

Denote the log-likelihood for the latent variable as $L^{*}(y_i^{*} \mid X_i,\beta)$ and the observed log likelihood as $L(y_i \mid X_i,\beta)$.

Result 1. $E[(dL^{*}(y_i^{*} \mid X_i,\beta)/d\beta) \mid y_i)=E[dL(y_i \mid X_i,\beta)/d\beta]$

Proof: GMRT p31

Result 2. $dL^{*}(y_i^{*} \mid X_i,\beta)/d\beta=\{dQ(X_i,\beta)/d\beta\}v_i(\beta)$

Proof: GMRT p9

Definition: The generalized error $\upsilon_i(\beta)=E[v_i(\beta) \mid y_i]$

Result 3. $dL(y_i \mid X_i,\beta)/d\beta=\{dQ(X_i,\beta)/d\beta)\}\upsilon_i(\beta)$

Proof: GMRT p12

Result 1 states that the expected value of score of the latent model is equal to the score of the observed model. Results 2 and 3 show that the scores for each model can be expressed as the product of the explanatory variables and the generalized residuals. Thus by obtaining the scores of the observed model with respect to the intercept we have derived the generalized residuals.

Result 4. The generalized residuals for the model where $y_i^*=y_i$ are given by the OLS residuals.

Proof: For the model $y_i=\beta' X_i+v_i$ the log likelihood has the following representation

$L^*=L=\sum[(\beta' X_i/\sigma^2)y_i-y_i^2/2\sigma^2-\ln\sqrt{2\pi\sigma^2}-(\beta' X_i)^2/2\sigma^2)]$

where $Q=\beta' X_i/\sigma^2$; $T(y_i)=y_i$; $A=(\beta' X_i)^2$; $B=-y_i^2/2\sigma^2-\ln\sqrt{2\pi\sigma^2}$;

Employing Results 1,2 and 3 gives

$dL^*/d\beta=dL/d\beta=\sigma^{-2}X_i v_i=\sigma^{-2}X_i \upsilon_i$

Thus $v_i=\upsilon_i$ where v_i are the OLS residuals.

Result 5. The generalized residuals for the probit model are given by equation (11) in the text.

Proof: GMRT p14.

Result 6. The generalized residuals for the tobit model are given by equation (13) in the text.

Proof: GMRT p 17.

Result 7. The generalized residuals for the ordered probit model are given by equation (18) in the text.

Proof: First introduce some additional notation. Following McKelvey & Zavoina (1975) define k ordinal outcomes. Now define the variables $d_{ji}=1$ iff individual i is in the j^{th} category and this is satisfied when

$$\mu_{j-1,i,}<y_i<\mu_{ji}$$

where $y_{ji}=\mu_j-\beta' X_i$ and $\Phi_{ji}=\Phi(y_{ji})$ and $\phi_{ji}=\phi(y_{ji})$ and Φ and ϕ denote the cumulative distribution function (cdf) and the probability density function (pdf) of the standard normal distribution.

Now $Pr[d_{ji}=1]=\Phi_{ji}-\Phi_{j-1,i}$ and the likelihood function for the ordered probit model can be written as

$$L=\sum_j \sum_i d_{ji} \log(\Phi_{ji}-\Phi_{j-1,i})$$

Differentiating with respect to the intercept and employing Result 3 gives

$$\sum_j \sum_i d_{ji}[(\phi_{j-1,i}-\phi_{ji})(\Phi_{ji}-\Phi_{j-1,i})]^{-1}$$

Note that the denominator of the above expression is the probability of the i^{th} observation being in the j^{th} category which we can denote as Π_{ji}, while the numerator is the density function evaluated at that probability. Denote this value as π_{ji}. Further as we evaluate this term for each observation we do not sum over the entire sample. Thus we can write the above derivative as

$$\sum_j d_{ji}\pi_{ji}\Pi_{ji}^{-1}$$

setting $y_{ji}=1$ when i is in the j^{th} category allows us to rewrite this as

$$\sum_j d_{ji}\pi_{ji}(1-\Pi_{ji})^{-1}\Pi_{ji}^{-1}(y_{ji}-\Pi_{ji})$$

which is equivalent to equation (18) in the text.

FOOTNOTES:

1 In many instances we will require more than just the slope parameter estimates to obtain the generalized residuals.

2 This approach is similar to that proposed by Hausman (1978) who argued that inconsistency due to the endogeneity of regressors can be adjusted by the inclusion of the residuals in place of the predicted values of the endogenous variable. This is the basis for the Hausman test of endogeneity.

3 While the result that the generalized residuals are the OLS residuals where the true value of y^*_{ji} is observed is trivially implied by the definition provided by Cox & Snell (1968) it is shown in the appendix for completeness.

4 In some instances it may be possible to employ estimators available in the non parametric and semi parametric literature although these procedures typically require restrictions upon the behavior of the error terms.

5 The results of Gourieroux et.al (1987) apply to models contained in the exponential family. Thus, given our assumption of normality, their results are relevant for the models discussed in this paper.

6 In Garen's empirical example the censoring variable he considers is years of education. As Garen notes his approach is not strictly applicable as years of education cannot be treated as a continuous variable. A more appropriate procedure, as also noted by Garen, is to treat y as an ordinal variable and estimate the censoring equation by ordered probit. This is the methodology pursued in section five of this paper.

7 Note that while this is not precisely Garen's estimator it captures the essence of his method and can be easily adjusted to replicate his procedure.

[8] This approach is somewhat similar to that proposed by Terza (1987) for models with ordinal qualitative explanatory variables although he does not consider the case where the qualitative variables are determined endogenously. In a subsequent paper, Terza (1989), he addresses this issue and the resulting estimator is similar to that outlined here. The major difference in the respective approaches is that in this paper we derive conditional expectations of the reduced form error while in Terza's work he focuses upon the expectation of the ordinal variable itself. It should be noted however that Terza's results are specific to the framework he examines and, unlike here, does not result as the special case of a more general model.

[9] The derivation of the covariance matrices for the models discussed in this paper are presented in Vella (1989).

[10] The application of these tests requires the data to satisfy certain conditions. These are all satisfied by the framework of the models discussed here. The application of the tests also requires that the models are estimated by maximum likelihood methods which, given the nature of the problem, is the method most likely to be employed.

[11] It is not necessary to perform these tests in the regression based framework as the test statistic is directly computable. As discussed in Pagan & Vella (1989) there may be certain advantages and disadvantages in employing this approach.

[12] I am grateful to Paul Yakaboski for making this data set available.

[13] The number of censored observations in the sample is forty one. This constitutes approximately seven percent of the sample.

[14] The semi-log specification reported in this table was chosen over the linear form on the basis of simple equation diagnostics.

[15] This model is identified by the non-linearity of the function that maps $\alpha' Z_i$ into generalized residuals. Alternative specifications which were identified through conventional exclusion restrictions produced similar results to those reported here.

[16] This choice of categories is rather arbitrary and further investigation is required to establish the robustness of the results to variations in the separation points for the categories. However Vella (1990) produces evidence based on a larger data set that the major step is at 35 hours per week.

[17] This required the estimation of a reduced form probit explaining the decision to work full-time or part-time to enable the calculation of the relevant correction factor. This was performed and the results are reported in column (2) of Table 4.

[18] Strictly speaking this approach is not appropriate as it ignores the presence of the already established selection bias in the work decision. It does provide some indication however if further bias exists.

[19] Vella (1990) interprets the coefficient on FT as the value of non wage labor income received by full-time workers.

REFERENCES:

Amemiya,T. (1978) "The Estimation of a Simultaneous Equation Generalized Probit Model." Econometrica 46:1193-205

----- (1979) "The Estimation of a Simultaneous Tobit Model." International Economic Review 20(1): 169-81

----- (1984) "Tobit Models: A Survey." Journal of Econometrics 24:3-61

Barnow,B.,G.Cain & A.Goldberger (1981) "Issues in the Analysis of Selectivity Bias." in W.Stromsdorfer & G.Farkas (eds), Evaluation Studies Review Annual Vol.5 pp 43-59. Beverly Hills: Sage

Blundell,R. & R.Smith (1989) "Estimation in a Class of Simultaneous Limited Dependent Variable Models" Review of Economic Studies 56(1) 185

Cogan,J. (1981) "Fixed Costs and Labor Supply" Econometrica 49: 945-64

Cox,D. & E.Snell (1968) "A General Definition of Residuals" Journal of the Royal Statistical Society: Series B

Engle,R.,D.Hendry & J-F.Richard (1983) "Exogeneity" Econometrica 51:277-304

Gallant,A. & D.Nychka (1987) "Semi-nonparametric maximum likelihood estimation" Econometrica 55: 363-90

Garen,J. (1984) "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable" Econometrica 52:1199-1218

Gourieroux,C.,A.Monfort.,E.Renault & A.Trognon (1987) "Generalized Residuals" Journal of Econometrics 34:5-32

Hausman,J. (1978) "Specification Tests in Econometrics" Econometrica 46:1251-71

Heckman,J. (1974) "Shadow Prices, Market Wages, and Labor Supply" Econometrica 42:679-94

----- (1976) "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variable Models and a Simple Estimator for Such Models" Annals of Economic and Social Measurement 5:475-92

----- (1978) "Dummy Endogenous Variables in a Simultaneous Equation System." Econometrica 46:931-59

----- (1979) "Sample Selection Bias as a Specification Error" Econometrica 47:153-61

Kuehneman,G. (1986) "The Role of Fringe Benefits in Compensation: A Theoretical and Empirical Investigation" unpublished Doctoral Dissertation, University of Chicago

Lee,L. (1978) "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables" International Economic Review 19:415-33

----- (1979) "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables. Econometrica 47:977-96

----- (1984) "Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity" Econometrica 52: 843-63

Leibowitz,A. (1983) "Fringe Benefits in Employee Compensation" in The Measurement of Labor Costs NBER Studies in Income and Wealth Vol. 48 edited by J.Triplett

McKelvey,R. & W.Zavoina (1975) "A Statistical Model for the Analysis of Ordinal Level Dependent Variables" Journal of Mathematical Sociology 4: 103-20

Nelson,F. & L.Olson (1978) "Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables" International Economic Review 19:695-710

Newey,W. (1984) "A Method of Moments Interpretation of Sequential Estimators" Economic Letters 14:201-206

----- (1985) "Maximum Likelihood Specification Testing and Conditional Moment Tests" Econometrica 53:1047-70

----- (1987) "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables" Journal of Econometrics 36:231-50

Pagan,A. (1986) "Two Stage and Related Estimators and their Applications" Review of Economic Studies 53:517-538

-----& F.Vella (1989) "Diagnostic Tests for Models Based on Individual Data: A Survey" Journal of Applied Econometrics 4: S29-S60

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

Smith,J & R.Ehrenberg (1983) "Estimating Wage-Fringe Trade Offs: Some Data Problems" in The Measurement of Labor Costs NBER Studies in Income and Wealth Vol. 48 edited by J.Triplett

Smith,R. & R.Blundell (1986) "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply" Econometrica 54:679-85

Tauchen,G. (1985) "Diagnostic Testing and Evaluation of Maximum Likelihood Models" Journal of Econometrics 30:415-43

Terza,J. (1987) "Estimating Linear Models with Ordinal Qualitative Regressors" Journal of Econometrics 34:275-91

----- (1989) "Estimating Linear Models Involving Endogenous Ordinal Qualitative Variables" unpublished paper

Vella,F. (1989) "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors" Rochester Center of Economic Research working paper No 199

---- (1990) "Non-Wage Benefits in a Simultaneous Models of Wages and Hours: Labor Supply Functions of Young Females" Monash University, Department of Econometrics working paper No 6

White,H. (1980) "A Hetereoskedasicity Consistent Covariance Matrix Estimator and a Direct Test for Hetereoskedasticity" Econometrica48:817-38

Yakaboski,P. (1988) "Productive Fringe Benefits: Theory and Evidence" unpublished paper

MONASH UNIVERSITY

DEPARTMENT OF ECONOMETRICS

WORKING PAPERS

4/89 Muhammad I. Bhatti, "Null Distribution of the Small Sample Mean Correlation Coefficient: An Application to Medical Research."

5/89 Grant H. Hillier, "On the Normalisation of Structural Equations: Properties of Direction Estimators".

6/89 Brett A. Inder, "A New Test for Autocorrelation in the Disturbances of the Dynamic Linear Regression Model".

7/89 Russel J. Cooper, Dilip B. Madan and Keith R. McLaren, "A 'Gormanesque' Approach to the Solution of Intertemporal Consumption Models".

8/89 Keith R. McLaren, "A Variant on the Arguments for the Invariance of Estimators in a Singular System of Equations".

9/89 T.R.L. Fry, "Univariate and Multivariate Burr Distributions: A Survey".

10/89 Merran Evans, "Robustness and Size of Tests of Autocorrelation and Heteroscedasticity to Non-normality".

11/89 R.D. Snyder. "A Computerized System for Forecasting Spare Parts Sales: A Case Study".

12/89 Maxwell L. King and Ping X. Wu. "Small-Disturbance Asymptotics and the Durbin-Watson and Related Tests in the Dynamic Regression Model".

13/89 W.T.M. Dunsmuir and R.D. Snyder. "ABC Analysis in Inventory Control - The Issue of Stability".

1990

1/90 P. Burridge. "The Functional Central Limit Theorem: An Introductory Exposition with Application to Testing for Unit Roots in Economic Time Series".

2/90 Maxwell L. King and Ping X. Wu. "Locally Optimal One-Sided Tests for Multiparameter Hypotheses."

3/90 Grant H. Hillier. "On Multiple Diagnostic Procedures for the Linear Model."

4/90 Jean-Marie Dufour and Maxwell L. King. "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors."

5/90 Keith R. McLaren. "A Reappraisal of the Neoclassical Approach to Modelling Business Investment."

6/90 Francis Vella. "Non-Wage Benefits in a Simultaneous Model of Wages and Hours: Labor Supply Functions of Young Females."