A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors

Vella, Frank

Working Paper No. 199
August 1989

A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors

Frank Vella

Rochester Center for Economic Research
Working Paper No. 199

A SIMPLE ESTIMATOR FOR SIMULTANEOUS MODELS WITH CENSORED ENDOGENOUS REGRESSORS

Frank Vella

Department of Economics
University of Rochester
Rochester, NY 14627

Working Paper No. 199

Revised August 1989

I owe a great debt of gratitude to Tom Cooley and Adrian Pagan for their encouragement and support. I am also very grateful to Adrian Pagan for his extremely helpful advice and comments on earlier versions. I also benefitted from discussions with Tony Hall and Siu Fai Leung and from seminars at the Australian National University, LaTrobe University, the University of Melbourne and the University of Rochester. I alone am responsible for any remaining errors.

ABSTRACT

This paper presents a simple consistent estimator for models with censored endogenous explanatory variables and generalized selectivity bias. The procedure relies upon the use of generalized residuals to account for the inconsistency caused by the endogeneity and is similar to the method proposed by Hausman (1978) for specification testing. As the approach is capable of dealing with selectivity bias as well as the endogeneity of censored regressors it unifies two areas of the econometrics literature. Two new tests of endogeneity are also discussed. The first is simply a by-product of the estimation procedure while the second is an application of the conditional moment testing framework. To illustrate the utility of the estimation procedure two empirical examples are presented. The first examines the trade off between fringe benefits and wages while the second is based upon trichotomous selection bias among working women. In both examples the estimator and the tests perform well.

1. Introduction.

The endogeneity of censored explanatory variables and the presence of sample selection bias are two frequently encountered problems in estimating econometric models from unit record data. The first is due to the nature of questionnaire based data and is most commonly observed in the form of explanatory binary dummy variables. The second results from individuals systematically choosing different courses of action. While both problems are generated by the same underlying mechanism they have been typically analyzed in separate frameworks. Simultaneous systems with limited dependent variables were initially examined by Amemiya (1978,1979), Heckman (1978), Lee (1978,1979), Nelson & Olson (1978), and later Newey (1987), who largely adopted an instrumental variable approach while recently some additional issues have been addressed by Smith & Blundell (1986), Rivers & Vuong (1988) and Blundell & Smith (1989) in a conditional maximum likelihood framework. Estimation from non-random samples was pioneered by Heckman (1974,1976,1979) and Lee (1978) who at first employed a full information likelihood framework but subsequently employed a two step estimator.

In this paper we provide an unifying framework for analyzing both problems. We do so by considering a simultaneous system comprising of a structural equation, of primary interest, where one or more of the explanatory variables are censored and endogenous while the remaining equations in the system comprise the reduced form representations of these censored variables. A method of producing consistent estimates is discussed for where, first, we wish to estimate the parameters over the entire data set and, second, where the observations in the primary equation can be sorted into sub-samples according to the values of these censored explanatory variables. An example of the first case 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.

A feature of this model is that estimation is complicated by the presence of endogeneity. For example, in the case of union status, it would be valuable to establish that while union status is not exogenously determined it is weakly exogenous, in the terminology of Engle, Hendry & Richard (1983), to wages. This is not only useful for hypothesis testing but also implies that the use of dummy variables will lead to consistent and fully efficient 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 explanatory variables appearing in the structural equation are censored or the observations for the structural equation can be sorted according to the value of the endogenous explanatory variables. The approach provides consistent estimates of both the reduced form and the structural equation parameters and provides a test of endogeneity. The second task is to introduce an easily implementable alternative test of endogeneity. This test is based upon the methods described in Newey (1985) and Tauchen (1985) and 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. Second, the procedure produces a test of weak exogeneity. Finally, unlike the existing estimators in the literature, the method discussed here can be easily extended to include various forms of censoring and selection bias. In this way it not only unifies a growing literature but also provides a link between two areas of research.

The following section discusses the general model and also presents 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 two empirical examples are presented. First we examine a model with an endogenous explanatory variable censored below at zero and the example explored is the effect of fringe benefits on wages where a non-trivial proportion of individuals in the sample receive 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 five. It should be noted that while the models to be covered are quite extensive the analysis that follows is reliant upon the assumption of normality. Accordingly the models that follow are restricted to be members of that family.

2. The General Model.

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

$$m_i=\alpha' X_i+\sum_{j=1}^{M-1}\beta_j y^*_{ji}+v_i \qquad i=1..n;\ j=1..M-1 \quad (1)$$

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

where m is the dependent variable in the equation of primary interest; y^*_j are unobserved endogenous variables; X and Z are vectors of exogenous variables observed for the n individuals in the sample; α,γ and β are parameters to be estimated; and the 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.

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

Assumption 2. The triplet (X_i,v_i,v_{ji}) are independently and identically distributed.

Assumption 3: v_i and the v_{ji}'s are, conditional on X_i, jointly normal with zero means and covariance matrix

$$\begin{matrix}\sigma^2 & \sigma_{12}\\ \sigma_{21} & \Sigma\end{matrix}$$

Assumption 4: The parameters of the model are identified.

To illustrate the issues consider two simple examples. Assume M is equal to two, allowing us to suppress the j subscript, and suppose that the latent variable y_i^* reflects i's productivity as a union member. The variable m_i denotes the observed wage received by i. The individual will be a union member if the level of y_i^* is beyond some threshold producing an observed value of y_i which is a binary dummy variable. As the decision to join the union is likely to be dependent on the offered wage rates, producing a non zero covariance between v_i and v_{ji}, least squares on the the structural equation after substituting y_i for y_i^* will produce inconsistent estimates.

Now consider where y_i^* represents some index reflecting productivity in different occupations and the censoring function produces an indicator function denoting the chosen occupation. Assume m_i is only observed for a subset of the sample corresponding to specific values of y_i. Least squares over the various subsets of y_i will not produce consistent estimates while regressing m on the explanatory variables and the observed values of y_i will also lead to inconsistent estimates. In fact such an approach is often not sensible as the censoring may impose a quantitative structure on y_i. For example in the case of occupational choice data different outcomes may be assigned different numerical values (i.e. $y_{ji}=1$ if individual i is a doctor; $y_{ji}=2$ if individual i is a lawyer etc). Note that even if the whole sample is observed the problem of selection bias remains as the choice of occupations may be a function of the wages offered in the various outcomes.

To contain consistent estimates in this model take expectations conditional[1] on the observed values of y_{ji}.

$$E(m_i|y_{ji})=E(\alpha' X_i + \sum_{j=1}^{M-1} \beta_j y^*_{ji}|y_{ji})+E(v_i|y_{ji}) \quad i=1..n;\ j=1..M-1 \quad (4)$$

$$E(y^*_{ji}|y_{ji})=E(\gamma_j' X_i|y_{ji})+E(v_{ji}|y_{ji}) \quad i=1..n;\ j=1..M-1 \quad (5)$$

The expected values of the error terms are now conditional on the value 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} and note that 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.

$$\begin{aligned} E(E(v_i|v_{ji})|y_{ji}) &= \Sigma^{-1}\sigma_{12}E(v_{ji}|y_{ji})+\eta_i \quad (6) \\ &= \Sigma^{-1}\sigma_{12}\upsilon_{ji}+\eta_i \\ &= \lambda'\upsilon_{ji}+\eta_i \end{aligned}$$

where η_i has an expectation of zero and is independent of υ_{ji} by construction and λ is a $j \times 1$ vector with λ_j as the j^{th} element. Now rewrite the structural equation as

$$E(m_i|y_{ji})=E(\alpha' X_i + \sum_{j=1}^{M-1} \beta_j y^*_{ji}+\lambda'\upsilon_{ji}|y_{ji})+E(\eta_i|y_{ji}) \quad (7)$$

---

[1] The term conditional expectations is used for the remainder of the paper to refer to the expectation taken with respect to y_{ji} in the information set.

and consistent estimation of α, β and λ is possible by least squares after substituting y^*_{ji} with y_{ji}. Thus where estimates of α and β are desired 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 residuals and insert them into the structural equation as additional regressors[2] and estimate α, β and λ by least squares. (Note that 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).

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. The conditional maximum likelihood approach is appropriate where y^*_i is uncensored, producing generalized residuals that coincide with OLS residuals, which then produces values of η_i which are also normally distributed[3]. The intractability of the distribution

---

[2] Note that in many instances we will require more than just the slope parameter estimates to obtain the generalized residuals.

[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

of η_i is in fact a substantial constraint as it restricts the dependent variable in the structural equation to be uncensored to enable estimation[4].

The implementation of this estimation procedure requires estimates of the generalized errors. These are obtained through the results of Gourieroux, Monfort, Renault & Trognon (1987) which show that the best prediction of the residual 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 Appendix A.

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

---

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 will also have some 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.

$$m_i=\beta' X_i+\delta y_i^*+u_{1i} \qquad i=1..n \qquad (8)$$

$$y_i^*=\alpha' Z_i+u_{2i} \qquad i=1..n \qquad (9)$$

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 (9) is probit. The generalized residuals, see Appendix A, are given by

$$\hat{\upsilon}_{2i}=E(v_{2i}|y_i)=(y_i-\hat{\Phi}_i)\hat{\phi}_i(1-\hat{\Phi}_i)^{-1}\hat{\Phi}_i \quad i=1..n \qquad (10)$$

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 $(\hat{\alpha/\sigma_2})$. Invoking our assumption of joint normality rewrite u_{1i} conditional on the observed value of υ_{2i}. That is

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

where λ is equal to σ_{12}/σ_2^2 . Now rewrite equation (8) in terms of its conditional expectation

$$E(m_i|y_i)=E(\beta' X_i+\delta y_i+\lambda\hat{\upsilon}_{2i}|y_i)+E(\eta_i|y_i) \qquad i=1..n \qquad (11)$$

Equation (11) can be estimated by least squares to produce consistent estimates of β, δ and λ. Those familiar with the selectivity bias literature will identify this estimator, where X does not contain an intercept and only values of m corresponding to specific values of y are observed, as Heckman's two step estimator while equation (11) is that proposed by Barnow et.al (1981). While this example is not the interesting contribution of this paper it is valuable to see the selectivity bias estimator derived in this context. It also indicates that the method can be extended to where y_i is a vector.

Employing this approach also produces the continuous selectivity bias estimator of Garen (1984). In Garen's model the dependent variable in the selection equation, (9), is able to take a continuum of values over a given range and is uncensored. To produce Garen's estimator we estimate (9) by OLS, as it corresponds to MLE, compute the generalized residuals, given by the least squares residuals (see Appendix A), and include them as an additional regressor in equation (8)[6,7].

---

[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 four 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

Although the models discussed above have appeared elsewhere in the literature they are easily derived in this present framework. More importantly this methodology can be extended to models with less conventional types of censoring which often appear in models in labor economics. For example consider the model outlined in equations (8) and (9) but where the censoring now takes the form

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

Given this type of censoring, combined with our normality assumption, an appropriate means of estimating α and σ_2 from equation (9) is Tobit. The generalized residuals now take the form

$$\upsilon_{2i}=E(v_{2i}|y_i)= -\tilde{\sigma}_2(1-I_i)\tilde{\phi}_i(1-\tilde{\Phi}_i)^{-1}+I_i\tilde{v}_{2i} \quad (11)$$

where $\tilde{\alpha}$ and $\tilde{\sigma}_2$ are the Tobit maximum likelihood estimates of α and σ_2; $\tilde{\phi}$ and $\tilde{\Phi}$ are evaluated at these estimates; $\tilde{v}_{2i}=y_i-\tilde{\alpha}'Z_i$; 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 $y_i>0$.

These examples illustrate estimation of both β and δ. However

---

procedure.

consider where the values of β are of primary interest and individuals can be sorted by their observed value of y_i. For example, consider the following model

$$m_{ji}=\beta_j'X_{ji}+u_{1i} \qquad i=1..n;\ j=1..k \qquad (12)$$

$$y^*_{ji}=\alpha'Z_i+u_{2i} \qquad i=1..n;\ j=1..k \qquad (13)$$

and the selection rule is

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

The value of m_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. However we can identify the category type for each individual i by observing the value of the k indicator functions determined by the selection rule. Take expectations conditional on the value of these indicator functions and rewrite (12) as

$$E(m_{ji}|y_{ji})=E(\beta_j'X_{ji}|y_{ji})+E(u_{1i}|y_{ji})$$

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

$$E(u_{2i}|y_{ji})=D_{ji}\pi_{ji}\Pi_{ji}^{-1}(1-\Pi_{ji})^{-1}(y_{ji}-\Pi_{ji}) \qquad (15)$$

where D_{ji} is an indicator function taking the value 1 if individual i is in category j and zero otherwise; Π_{ji} is the estimated probability that individual i is in the j^{th} category while π_{ji} is the estimated value of the density at that point. As shown in Appendix A equation (15) 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 β_j by estimating k separate regressions over the sub-samples corresponding to $y_{ji}=1$ 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

---

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.

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 models discussed above is that the estimation procedure is complicated by the presence of endogeneity. For example, where exogeneity can be established the use of y_{ji} in place of y_{ji}^* will often produce estimates which are not only consistent but also efficient. 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. For example, is the choice to become a union member or seek government employment independent of the factors that determine wages?

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 hetereoskedasicity arising in the first step. This is done for the Tobit and ordered probit models in Appendix B.

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. An alternative approach is to perform the test under the null hypothesis that the correlation is equal to zero. By estimating under the null 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)[9]. 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 value of σ_{1j} will be equal to zero. Thus a relevant test of such a proposition would be to examine the sample estimate of σ_{1j}. This can be easily shown to be equal to $\hat{\tau}=n^{-1}\Sigma\hat{\upsilon}_{1i}\hat{\upsilon}_{ji}$ (see Pagan & Vella (1989)) where the $\hat{\upsilon}$'s represent

---

[9] 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.

the estimated values of the generalized residuals computed under the null of weak exogeneity. The difficulty now lies in deriving the distribution of $\hat{\tau}$. 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 $\hat{\tau}$ is equal to zero[10].

While this conditional moment approach requires some additional computation it is often more readily applicable than the alternative method. That is, where exogeneity is not rejected it is not necessary to compute the second step covariance matrix as the usual computer output is appropriate. Given that the covariance matrix may be cumbersome to compute in some instances there may exist large incentives to perform these conditional moment tests first.

4. 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.

---

[10] 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.

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

$$\text{hourly wage} = \alpha + \Sigma\alpha_j\text{*personal characteristics} + \Sigma\alpha_l\text{*region dummies} + \Sigma\alpha_i\text{*industry dummies} + \alpha_f\text{*hourly fringe benefits} \quad (16)$$

It is likely that the level of fringe benefits is determined simultaneously with wages so the problem of endogeneity is obvious. Furthermore, many individuals receive no fringe benefits and as 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 (16) 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[11]. 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

$$\text{hourly fringe benefits} = \beta + \Sigma\beta_j * \text{personal characteristics} + \Sigma\beta_l * \text{region dummies} + \Sigma\beta_f * \text{industry dummies} \quad (17)$$

and assuming the error for this equation is normally distributed we can estimate the β's by Tobit[12]. Following the estimation of equation (17) 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 (16) 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.

---

11 I am grateful to Paul Yakaboski for making this data set available.

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

To examine the fringe\wage trade off we first estimate equation (16) without entering the generalized residuals as an 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[13]. 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 (16) was re-estimated and the results are reported in column 3 of Table 2[14]. The coefficient on fringe benefits continues to be statistically significant at conventional levels of confidence but now displays the expected negative sign. The coefficient also appears to be of reasonable magnitude. 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 the relationship.

Now focus upon the estimation of wage equations with trichotomous

---

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

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

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[15]. Accordingly we will

---

15 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

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.

$$\text{log hourly wage} = \alpha + \Sigma\alpha_j * \text{personal characteristics} \qquad (18)$$

The variables employed are described in Table 3 and equation (18) 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 (18), 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

---

(1989) produces evidence based on a larger data set that the only statistically significant step is at 35 hours per week.

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 dichotomy[16] is not weakly exogenous to wages[17].

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 (15) and insert them as an additional regressor in equation (18). 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

---

16 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.

17 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.

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[18]. 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

---

18 Vella (1989) interprets the coefficient on FT as the value of non wage labor income received by full-time workers.

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.

5. 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
CONSTANT	-.775*	.538*	-.057*
	(.213)	(.113)	(.193)
AGE	.010*	.003*	.012*
	(.002)	(.001)	(.002)
MAR	.129*	.033	.138*
	(.057)	(.031)	(.035)
RACE	.088	.069	.145*
	(.081)	(.047)	(.045)
MALE	.287*	.246*	.485*
	(.062)	(.034)	(.059)
EDUC	.072*	.030*	.091*
	(.010)	(.006)	(.015)
OFFICE	----	.064	.007
		(.034)	(.037)
HFRINGE	----	.437*	-.457*
		(.025)	(.183)
FGRES	----	----	.862*
			(.174)
σ	.564	----	----
Log-likelihood	-529.14	----	----
$\bar{R}^2$	----	.599	.614
Observations	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) * 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 Variables		
	WORK	FT	TIME
Constant	-.930* (.289)	1.361* (.393)	-.447 (.232)
AGE1	-.581* (.146)	-.590* (.191)	-.614* (.131)
AGE2	-.224 (.119)	-.240 (.147)	-.222* (.103)
AGE3	-.233* (.110)	-.048 (.137)	-.170 (.095)
CIT	.282* (.082)	.062 (.106)	.239* (.073)
MAR	-.541* (.161)	.003 (.269)	-.500* (.139)
HLT	-.510* (.115)	-.158 (.995)	-.469* (.099)
ENG	.245* (.118)	.056 (.160)	.184 (.099)
EDUC	.148* (.021)	-.012 (.026)	.110* (.017)
CHLD	-1.786* (.115)	-1.043* (.188)	-1.769* (.106)
SPINC	.002* (.0004)	-.0003 (.0007)	.0015* (.00003)
TINC	.00003 (.00005)	.0001 (.00008)	.00006 (.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	----	-.500*	----	-.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 A: 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.

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

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

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

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

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

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

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 (10) in the text.

Proof: GMRT p14.

Result 6. The generalized residuals for the tobit model are given by equation (11) in the text.

Proof: GMRT p 17.

Result 7. The generalized residuals for the ordered probit model are given by equation (15) in the text.

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

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

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

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

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

Differentiating with respect to the intercept and employing Result 3 gives

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

which is equivalent to equation (15) in the text.

Appendix B: Asymptotic Covariance Matrix of the Two Step Estimator

In deriving the covariance matrix of the estimator discussed in section 2 we follow the approach of Newey (1984) and Pagan (1986). Without loss of generality we consider the two equation form of the model where the extension to multiple equations is obvious.

(B1) $m_i=\beta' X_i+\delta' y_i+\lambda'\hat{\upsilon}_i+\eta_i$

(B2) $y_i=\alpha' Z_i+u_i$

where $\hat{\upsilon}_i$ is some specified function of $\hat{\alpha}'Z_i$ and the errors are uncorrelated across individuals and equations. The system can be estimated in a sequential manner in that we can first estimate (B2) to get $\hat{\alpha}$ which we then employ to generate $\hat{\upsilon}_i$. Denote the vector of parameters from (B1) as $\theta=[\beta,\delta,\lambda]$; the variables $W=[X,y,\hat{\upsilon}]$; and make the following assumptions.

b1. The estimate of α, $\hat{\alpha}$, is obtained by solving the j moment conditions[19] $h(\hat{\alpha})=0$.

b2. $\hat{\alpha}$ converges to a well defined limit α_0.

b3. The estimate of θ, $\hat{\theta}$, is obtained by solving the l moment conditions $g(\hat{\alpha},\hat{\theta})=0$.

b4. The conditions for the mean value theorem for random functions are satisfied.

b5. $n^{1/2}(\hat{\alpha}-\alpha_0)\sim N(0,V_\alpha)$.

---

[19] Note that the α's may include more than just the slope parameters.

b6. $n^{1/2}(g(\alpha_0,\theta_0)) \sim N(0,V_\theta)$.

To find the asymptotic distribution of $n^{1/2}(\hat{\theta}-\theta_0)$ take a Taylor series expansion of $g(\alpha_0,\theta)$ around $g(\alpha_0,\theta_0)$. This gives

$$0=g(\hat{\alpha},\hat{\theta})=g(\alpha_0,\theta_0)+H_{\theta\theta}(\tilde{\alpha},\tilde{\theta})(\hat{\theta}-\theta_0)+H_{\theta\alpha}(\tilde{\alpha},\tilde{\theta})(\hat{\alpha}-\alpha_0)+o_p(1)$$

where $(\tilde{\alpha},\tilde{\theta})$ lies between (α_0,θ_0) and $(\hat{\alpha},\hat{\theta})$ and H denotes the Hessian matrix. Now through minor manipulation and rearrangement we get

$$n^{1/2}(\hat{\theta}-\theta_0)=I_{\theta\theta}^{-1}\ n^{-1/2}g(\alpha_0,\theta_0)+I_{\theta\theta}^{-1}I_{\theta\alpha}n^{1/2}(\hat{\alpha}-\alpha_0)$$

where I denotes $-n^{-1}H$. Assuming that the cross products are equal to zero gives the following covariance matrix.

$$\text{(B4)}\quad I_{\theta\theta}^{-1}V_\theta I_{\theta\theta}^{-1}+I_{\theta\theta}^{-1}I_{\theta\alpha}V_\alpha I_{\theta\alpha}' I_{\theta\theta}^{-1}$$

To compute (B4) for the models discussed in section 2 we need to specify the moment conditions which determine the estimates and the form of the function which maps $\hat{\alpha}'Z_i$ into $\hat{\upsilon}_i$. However first note that the first part of (B4) can be consistently estimated using White's (1980) procedure. For the tobit model the moment conditions that determine $\hat{\alpha}$ are given by

$$\text{(B5)}\quad h_\alpha= \sum_0 \phi((\hat{\alpha}'Z_i/\sigma))Z_i/(1-\Phi(\hat{\alpha}'Z_i/\sigma) +\sigma^{-2}\sum_+ (y_i-\hat{\alpha}'Z_i))Z_i$$

$$\text{(B6)}\quad h_\sigma 2= 1/2\sigma^3 \sum_0 (\hat{\alpha}'Z_i\phi(\hat{\alpha}'Z_i/\sigma)/(1-\Phi(\hat{\alpha}'Z_i/\sigma)) -p/2\sigma^2 +1/2\sigma^4\sum_+(y_i-\alpha'Z_i)^2$$

where $\Phi(.)$ and $\phi(.)$ are the cdf and pdf of the normal distribution; the subscripts 0 and + denote that the summation is over the zero and positive observations of y respectively; and p is equal to the number of non-zero observations. The moment conditions specified by g are simply the least squares normal equations. Thus the terms required for (B4) are all immediately available except for the component capturing the uncertainty generated by having to estimate υ_i and this is contained in $I_{\theta\alpha}$. As the constructed variable is computed as the first order condition with respect to the intercept this is shown in (B5) where the α is the intercept's coefficient. To obtain $I_{\theta\alpha}$ we first differentiate the (B5) which corresponds to the intercept with respect to the variables which enter the generalized residuals. This produces the tobit second derivatives for the intercept taken with respect to the other parameters. These are given by

$$\text{(B7)}\quad dh_{\alpha}/d\alpha = -\sum_{0}\hat{\phi}_i(1-\hat{\Phi}_i)^{-2}[\hat{\phi}_i-\hat{\sigma}^{-2}(1-\hat{\Phi}_i)\hat{\alpha}'Z_i]Z_i'Z_i - \hat{\sigma}^{-2}\sum_{+} Z_i'Z_i$$

$$\text{(B8)}\quad dh_{\alpha}/d\sigma = -1/2\hat{\sigma}^{2}\sum_{0}\hat{\phi}_i(1-\hat{\Phi}_i)^{-2}[\hat{\sigma}^{-2}(1-\hat{\Phi}_i)(\hat{\alpha}'Z_i)^{2}-(1-\hat{\Phi}_i)-\hat{\alpha}'Z_i\hat{\phi}_i]Z_i$$

$$-\hat{\sigma}^{-4}\sum_{+}(y_i-\hat{\alpha}'Z_i)Z_i$$

where $\hat{\phi}_i$ and $\hat{\Phi}_i$ are the normal pdf and cdf evaluated at $\hat{\alpha}'Z_i$. Premultiplying (B7) and (B8) by W produces $I_{\theta\alpha}$ and this takes the following form. Denoting the moment conditions for the k+1=j parameters in the following manner $h_{\alpha1}, h_{\alpha2}, \ldots h_{\alpha k}$, h_{σ} such that $h_{\alpha1}$ is the derivative with respect to the intercept, and the value of this derivative for the i^{th} individual as $h_{\alpha1i}$. $I_{\theta\alpha}$ thus takes the form of the product of the two following matrices

$$\begin{matrix} W_{11} \cdot W_{12} \cdot \cdots W_{1n} \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ W_{M1} \cdot W_{M2} \quad W_{Mn} \\ M \times n \end{matrix} \qquad \begin{matrix} dh_{\alpha 11}/d\alpha_1, & dh_{\alpha 11}/d\alpha_2, & . & dh_{\alpha 11}/d\sigma \\ . & . & . & . \\ . & . & . & . \\ dh_{\alpha 1n}/d\alpha_1, & dh_{\alpha 1n}/d\alpha_2, & . & dh_{\alpha 1n}/d\sigma \\ & n \times j & & \end{matrix}$$

In deriving the covariance matrix for the model where the first step is estimated by ordered probit we follow the same steps although the h moment conditions now take a different form. To illustrate this we first require some additional notation. Define $D_{i,j}=1$ if the i^{th} individual is in the j^{th} category and this occurs when $\mu_{i,j}<y_{i,j}<\mu_{i,j}$, and zero otherwise. Now define $\delta_{i,j}=1$ if the i^{th} individual is in the j^{th} category and zero otherwise. Now set

$$y_{i,j}=\mu_j-\alpha' Z_i; \ \phi_{i,j}=\phi(y_{i,j}) \text{ and } \Phi_{i,j}=\Phi(y_{i,j})$$

The moment conditions for the α's are given by

$$\text{(B9)} \quad h_\alpha=\sum_i \sum_j D_{i,j}[(\hat{\phi}_{i,j-1}-\hat{\phi}_{i,j})(\hat{\Phi}_{i,j}-\hat{\Phi}_{i,j-1})^{-1}]Z_i$$

and the moment conditions for the μ's are given by

$$\text{(B10)} \quad h_\mu=\sum_i \sum_j D_{i,j}[(\hat{\phi}_{i,j}\delta_{i,j}-\hat{\phi}_{i,j-1}\delta_{i,j-1})(\Phi_{i,j}-\Phi_{i,j-1})^{-1}]$$

The corresponding terms to (B7) and (B8) are given by

$$(B11)\quad dh_{\alpha}/d\alpha = \sum_i \sum_j D_{i,j}[\{\Phi_{i,j}-\Phi_{i,j-1})(\phi_{i,j-1}y_{i,j-1}-\phi_{i,j}y_{i,j})$$

$$-(\phi_{i,j-1}-\phi_{i,j})^2\}\{\Phi_{i,j}-\Phi_{i,j-1}\}^{-2}]Z_iZ_i$$

$$(B12)\quad dh_{\alpha}/d\mu = \sum_i \sum_j D_{i,j}[\{(\hat{\Phi}_{i,j}-\hat{\Phi}_{i,j-1})(\hat{\phi}_{i,j}y_{i,j}\delta_{i,j}-\hat{\phi}_{i,j-1}y_{i,j-1}\delta_{i,j-1})$$

$$-(\hat{\phi}_{i,j-1}-\hat{\phi}_{i,j})(\hat{\phi}_{i,j}\delta_{i,j}-\hat{\phi}_{i,j-1}\delta_{i,j-1})\}\{\hat{\Phi}_{i,j}-\hat{\Phi}_{i,j-1}\}^{-2}]Z_i$$

It should be noted that in this second example the sample sizes will vary depending on the category of j being considered. However the information matrix for α, which reflects the uncertainty introduced by having to estimate α, is always based on the entire sample.

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

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

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 forthcoming

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

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 a Simultaneous Model of Wages and Hours" unpublished paper

White,H. (1980) "A Hetereoskedasicity Consistent Covariance Matrix Estimator and a Direct Test for Hetereoskedasicity" <u>Econometrica</u> 48:817-38

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