Estimating and Interpreting Models With Endogenous Treatment Effects

Francis VELLA
Department of Economics, Rutgers University, New Brunswick, NJ 08901
(*fvella@rci.rutgers.edu*)

Marno VERBEEK
Center of Economic Studies, K. U. Leuven, Belgium, and CentER, Tilburg University, The Netherlands
(*marno.verbeek@econ.kuleuven.ac.be*)

This article examines the relationship between two alternative approaches, instrumental variables and control function procedures, for estimating the impact of endogenous treatment effects. Although it is well known that the two approaches generate comparable estimates, the relationship between the estimators and their accompanying endogeneity tests appears not to be well understood. We show that the two procedures are closely related. We also examine the implications of the two procedures for the underlying economic sorting behavior.

KEY WORDS: Control function estimators; Endogenous treatment effects; Instrumental variables; Roy model.

The estimation of endogenous treatment effects is a common feature of empirical work in economics. When the treatment can be characterized by a dichotomous indicator function, its effect is typically estimated via instrumental variables (IV) or variants of the control function (CF) approach motivated by Heckman (1978, 1979). Although it is well known that the two approaches generate similar estimates, the relationship between the estimators and their accompanying endogeneity tests appears not to be well understood. This article explores this relationship and the associated economic implications of these estimators.

Section 1 presents a simultaneous model featuring a response variable, which we shall refer to as a wage, and an endogenous treatment. Section 2 discusses two alternative procedures for obtaining the parameters of this model, IV and the CF estimator, and the conditions they respectively require for consistency. Section 3 compares the estimators in more detail and features a small Monte Carlo experiment. Section 4 further explores the economics of sorting implicit in these procedures and considers the different economic restrictions they impose. Section 5 concludes.

1. A SIMULTANEOUS MODEL OF WAGES AND TREATMENTS

Let W_{ti} and W_{ni} denote individual i's wage with and without a given treatment, respectively, noting that for each individual we only observe the wage in one state or the other. Furthermore, let the conditional expectations of these variables, given a vector of observable characteristics X_i, be given by $X_i\gamma_t$ and $X_i\gamma_n$, where γ_t and γ_n are unknown parameters. We can then write

$$W_{ti} = X_i\gamma_t + \varepsilon_{ti} \tag{1}$$

and

$$W_{ni} = X_i\gamma_n + \varepsilon_{ni}, \tag{2}$$

where ε_{ti} and ε_{ni} are zero mean error terms, assumed to be independent of X_i. Let $T_i = 1$ if individual i received a treatment and $T_i = 0$ otherwise so that the observed wage is given by $W_i = T_iW_{ti} + (1 - T_i)W_{ni}$.

In many economic examples, assignment to the subsample receiving the treatment is not random and potentially endogenous to the outcome variable in the primary equation. The objective is to estimate the effect of undertaking the treatment while accounting for its endogeneity. To complete the model, assume that nonwage-related considerations are possibly relevant to the decision to undertake the treatment. Let C_i denote the cost of acquiring the treatment, where

$$C_i = S_i\delta + \varepsilon_{ci}, \tag{3}$$

where S_i is a vector of exogenous variables, independent of $(\varepsilon_{ti}, \varepsilon_{ni}, \varepsilon_{ci})$, δ is an unknown parameter vector, and ε_{ci} is a zero mean error term. The variables in S_i may overlap with those in X_i. Assuming that individuals wish to maximize wages, the decision to undergo the treatment can be written as

$$T_i = I(W_{ti} - W_{ni} - C_i > 0) = I(Z_i\pi + \varepsilon_i > 0), \tag{4}$$

where $I(\cdot)$ is an indicator function, $\varepsilon_i = \varepsilon_{ti} - \varepsilon_{ni} - \varepsilon_{ci}$, Z_i is a vector containing all elements found in X_i and S_i, and π is a vector of reduced-form parameters. The errors $(\varepsilon_{ti}, \varepsilon_{ni}, \varepsilon_{ci})'$ are assumed to be independent of the exogenous variables, with variances σ_j^2 and covariances σ_{jk}, for $j, k = t, n, c$, noting that it is not uncommon to assume joint normality. The covariances for the pairs $(\varepsilon_{ti}, \varepsilon_i)$ and $(\varepsilon_{ni}, \varepsilon_i)$ are denoted $\sigma_{t\varepsilon}$ and $\sigma_{n\varepsilon}$, respectively. The specification in (4) is based on the assumption that, when mak-

Journal of Business & Economic Statistics
October 1999, Vol. 17, No. 4

ing the decision to undergo the treatment, individuals know their individual gains from it. Relaxing this and assuming, for example, that the individual's expected gain from treatment is $X_i(\gamma_t - \gamma_n)$ would not change the reduced form in (4), although ε_i would not include ε_{ti} and ε_{ni}. This has no implications for the statistical properties of the estimators we consider because these will depend on $\sigma_{t\varepsilon}$ and $\sigma_{n\varepsilon}$, but it does have implications for the economic interpretation of the results.

When $C_i = 0$, the model collapses to that considered by Roy (1951), and individuals choose to undergo the treatment purely on the basis of a comparison of the potential wages. Generally, costs are allowed to be relevant, and Z_i includes variables in addition to X_i that capture differences in costs across individuals. Because observable characteristics that are likely to influence the cost of acquiring the treatment appear in S_i, it is likely that unobservable characteristics also influence the cost of acquiring the treatment independently of their influence on the wages in the two sectors. Accordingly, ε_{ci} has a positive variance. This is potentially an important point and one to which we shall return later. Assuming that the treatment effect operates through the intercept only (e.g., see Heckman and Robb 1985; Robinson 1989a,b) the observed wage can be written as

$$W_i = \alpha_n + (\alpha_t - \alpha_n)T_i + X_i\gamma + \eta_i, \qquad (5)$$

where $\eta_i = T_i\varepsilon_{ti} + (1 - T_i)\varepsilon_{ni}$. The treatment effect for individual i is $W_{ti} - W_{ni}$, and $\alpha = \alpha_t - \alpha_n$ is the average treatment effect of an individual "randomly" assigned to the treatment. Note that the terminology is slightly misleading because there are no individuals who are randomly assigned. The term is meant, however, to capture the idea that the unobservables capturing the treatment decision that are correlated with the wage have been controlled for. Heckman (1990) referred to this as the experimental treatment average. The average treatment effect of those undergoing the treatment is given by

$$E\{W_{ti} - W_{ni}|T_i = 1, Z_i\} = \alpha + E\{\eta_i|T_i = 1, Z_i\}, \qquad (6)$$

which comprises the sum of the experimental average and the return to the unobservables as they are priced in the treated sector. This is a special case of what Imbens and Angrist (1994) defined as "local average treatment effects." Imbens and Angrist focused on the estimation of the treatment effect but allowed it to vary for location in the population. We focus here purely on the estimation of the experimental treatment average, α.

2. ESTIMATION OF THE EXPERIMENTAL TREATMENT AVERAGE

Consider estimation of the treatment effect from Equation (5). Under the usual regularity conditions and assuming that

$$E\{\eta_i|X_i, T_i\} = 0, \qquad (7)$$

α can be estimated consistently by ordinary least squares (OLS). Under normality, a necessary and sufficient condition for (7) to hold is $\sigma_{t\varepsilon} = \sigma_{n\varepsilon} = 0$, which implies that the unobservable components of wages are irrelevant to the treatment decision. That is, T_i is exogenous to wages. In cases in which (7) does not hold, it is necessary to account for the endogeneity of the treatment.

The first estimator for α that we consider is the IV estimator in which Z_i is employed as an instrument vector for X_i, T_i. Estimation requires at least one regressor in Z_i that is not contained in X_i, and consistency requires that

$$E\{\eta_i|Z_i\} = E\{\varepsilon_{ti}|T_i = 1, Z_i\}P\{T_i = 1|Z_i\} + E\{\varepsilon_{ni}|T_i = 0, Z_i\}P\{T_i = 0|Z_i\} = 0. \qquad (8)$$

Thus, under normality, a necessary condition for (8) is $\sigma_{t\varepsilon} = \sigma_{n\varepsilon}$, implying that IV estimation of the experimental average under normality requires equality of the covariances. The standard IV estimator thus only allows for restrictive forms of endogenous selection into treatment in that it imposes restrictions on the values of the covariances between the error terms. Heckman (1997), not assuming any model for the treatment decision, also showed that IV estimation imposes strong restrictions on how persons make the decision to undergo the treatment. Only if the response to treatment does not vary among persons (implying that $\varepsilon_{ti} = \varepsilon_{ni}$), or the idiosyncratic gain $(\varepsilon_{ti} - \varepsilon_{ni})$ is unobserved to the individual when making the decision, the condition in (8) is trivially satisfied and an instrument that enters (4) with nonzero coefficient would identify the experimental treatment average.

It is possible to adopt alternative forms of the IV procedure. For example, one could estimate the term $P\ \{T_i = 1|Z_i\}$ parametrically, which, in turn, could be employed as an instrument for T_i. It is also possible to estimate the $P\ \{T_i = 1|Z_i\}$ nonparametrically and employ this as an instrument. Finally, one could estimate $P\ \{T_i = 1|Z_i\}$ either parametrically or nonparametrically and insert it in place of T_i in Equation (5) and estimate the model by OLS. Although these different approaches are variants on the IV estimator, they all require (8) to hold. Irrespective of what assumptions are made or not made, in estimation a necessary condition for these estimators' consistency, if the underlying distribution is normal, is $\sigma_{t\varepsilon} = \sigma_{n\varepsilon}$. In the instances in which $P\ \{T_i = 1|Z_i\}$ is estimated parametrically, however, there is no need to impose the exclusion restriction that Z_i contains something not found in X_i provided that the mapping from Z_i to $P\ \{T_i = 1|Z_i\}$ is nonlinear.

An alternative estimator that produces consistent estimates of the treatment effect is based on the CF method of Heckman (1978, 1979). This approach is based on the conditional expectation of W_i given T_i and Z_i,

$$E\{W_i|Z_i, T_i\} = \alpha_n + \alpha T_i + X_i\gamma + E\{\eta_i|Z_i, T_i\}. \qquad (9)$$

The key feature is related to the inclusion of the latter term, which can be written as

$$E\{\eta_i|Z_i, T_i\} = T_iE\{\varepsilon_{ti}|Z_i, T_i = 1\} + (1 - T_i)E\{\varepsilon_{ni}|Z_i, T_i = 0\}.$$

Under the joint normality assumption, the two conditional expectations on the right side can be written as

$$E\{\varepsilon_{ji}|Z_i, T_i\} = \sigma_{j\varepsilon}\lambda_i(Z_i\pi), \qquad j = t, n, \qquad (10)$$

where

$$\lambda_i(Z_i\pi) = E\{\varepsilon_i|Z_i, T_i\} = (1 - T_i)\frac{-\phi(Z_i\pi)}{\Phi(-Z_i\pi)} + T_i\frac{\phi(-Z_i\pi)}{1 - \Phi(-Z_i\pi)} \qquad (11)$$

is the generalized residual (Gourieroux, Monfort, Renault, and Trognon 1987) of the probit model describing the treatment decision and where $\phi(\cdot)$ and $\Phi(\cdot)$ represent the probability density and cumulative density functions of the standard normal distribution, respectively. As usual, we have normalized σ_ε^2 to unity.

The CF estimator is implemented in the following way. First, the probit model for the treatment decision is estimated by maximum likelihood to obtain an estimate of $\pi, \hat{\pi}$, after which an estimated value of (11), interacted with T_i, is included in the wage regression to be estimated by least squares. Alternatively, we can constrain $\sigma_{t\varepsilon}$ to equal $\sigma_{j\varepsilon}$ and add $\lambda_i(\cdot)$ as a single regressor. We refer to this latter approach as the *restricted* CF estimator. Because the correction term $\lambda_i(\cdot)$ is derived under the assumption of normality, the CF approach generally does not provide consistent estimates in the absence of normality. Note, however, that it is possible to account for some departures from normality through the use of variants of this approach.

3. THE RELATIONSHIP BETWEEN INSTRUMENTAL VARIABLES AND CONTROL FUNCTION ESTIMATORS

Although it is well known that the IV and CF approaches generally produce similar estimates of the experimental treatment average (e.g., see Robinson 1989a) the relationship between the two is often not well understood. First, consider the conventional IV estimator. From the two-stage least squares result we know that the IV estimator can be obtained via OLS on

$$W_i = \alpha_n + \alpha\hat{T}_i + X_i\gamma + \xi_i, \qquad (12)$$

where $\hat{T}_i$ is the linear predictor of T_i, given Z_i, and ξ_i is a zero mean error. Hausman (1978), noted that the algebraically equivalent estimator for α_n, α, and γ is obtained by estimating

$$W_i = \alpha_n + \alpha T_i + X_i\gamma + \hat{v}_i\theta + \zeta_i \qquad (13)$$

by OLS, where $\hat{v}_i = T_i - \hat{T}_i$ and ζ_i is equal to a zero mean error term uncorrelated with all the exogenous variables in the model. On the other hand, imposing $\sigma_{t\varepsilon} = \sigma_{n\varepsilon}$, the restricted CF estimator is based on OLS in

$$W_i = \alpha_n + \alpha T_i + X_i\gamma + \hat{\lambda}_i(\cdot)\sigma_{t\varepsilon} + \nu_i, \qquad (14)$$

where the only difference from IV is the inclusion of the generalized probit residual instead of the least squares residual based on the same explanatory variables. Although (13) is typically not employed to implement IV, partially due to the requirement to adjust the standard errors for the two-step nature of the procedure, the use of residuals to account for endogeneity is commonly encountered in models with censored endogenous regressors. For example, Heckman (1978, 1979) adopted this approach to account for sample-selection bias and endogeneity bias in models in which the treatment is captured through an indicator function. Vella (1993) employed the same approach for a range of models involving selection bias or censored endogenous regressors. Smith and Blundell (1986) and Rivers and Vuong (1988) adopted the same idea in accounting for endogeneity in models in which the dependent variable is censored and the endogenous regressor is continuous.

Equation (14) helps understand the relationship between the two estimators, IV and restricted CF, noting that (for known λ) it has the interpretation of a conditional expectation. If we would estimate (14) by IV with instruments Z_i, the $\hat{\lambda}_i$ would be eliminated and the standard IV estimator for α_n, α, and γ would result. The reason for this is that $\sum_{i=1}^N \hat{\lambda}_i(\cdot)Z_i = 0$ from the first-order conditions of the probit model determining whether an individual undergoes treatment. Because estimating by IV will never be more efficient than estimating by OLS, the OLS estimator in (14), which reproduces the restricted CF estimator, is at least as efficient as IV. The rationale for this is that the latter does not exploit the functional form of λ_i and thus does not exploit normality.

It is obvious from (13) and (14) that the IV and restricted CF approaches would produce identical results if $\hat{v}_i$ were equal, or proportional, to $\hat{\lambda}_i$. When the first-stage probit is replaced by a linear probability model (see Olsen 1980), however, the two approaches are identical. In addition, for many applications the empirical correlation coefficient between $\hat{v}_i$ and $\hat{\lambda}_i$ will be very high, thereby generating similar estimates for both procedures. Moreover, nonnormality will only increase the differences between the IV and CF approaches to the extent that it affects the empirical correlation between the generalized residual and the OLS residual.

To illustrate this latter point, we simulated the following model in which the response variable is wages and the endogenous treatment is a dummy variable capturing union membership. To ensure that our experimental design is realistic we simulated the endogenous variables from the model outlined here using exogenous variables for 2,121 working male youth drawn from the 1985 wave of the Australian Longitudinal Survey. The model had the following form:

$$\begin{aligned}\text{union} &= I(-1.00 + .06\text{ educ} + .61\text{ govt} \\ &\quad - .01\text{ hours} - .18\text{ health} + .12\text{ ex} - .004\text{ ex}^2 \\ &\quad + .313\text{ appa} + \varepsilon_1 > 0)\end{aligned}$$

$$\begin{aligned}\ln(\text{wage}) &= .69 + .09\text{ educ} - .028\text{ govt} \\ &\quad - .008\text{ hours} - .006\text{ health} + .10\text{ ex} \\ &\quad - .005\text{ ex}^2 + .25\text{ union} + \varepsilon_2,\end{aligned}$$

where ε_1 and ε_2 are error terms and $I(\cdot)$ is an indicator function. The remaining variables are as follows: educ—years of education; govt—a dummy denoting employment in the government sector; hours—weekly hours worked; health—dummy denoting health limits on type of work; ex—age minus education minus 6; appa—dummy indicating that the individual underwent an apprenticeship. The parameter values used to generate the data are the two-step restricted CF estimates, although for the sake of the simulations we increased the value of the union dummy coefficient to .25 and chose the value of the constant to ensure that between 30% and 50% of the observations are assigned to the union group. We examined nine different cases corresponding to different assumptions regarding the errors. The two features we control are the distribution of the errors in the union equation and the degree of the correlation between the error terms. First, we assume that (a) $\varepsilon_{1i} \sim \chi^2_1$, (b) $\varepsilon_{1i} \sim$ U[0, 1], and (c) $\varepsilon_{1i} \sim$ N(0, 1). In each instance we subtract the mean of the errors to ensure that each has zero mean and then divide by the standard deviation. We then generate the wage equation error as $\varepsilon_2 = \theta\varepsilon_1 + .3$ N(0, 1), where we set θ equal to (a) .01, (b) .075, and (c) .1. Note that these parameter variables generated an error for the wage equation that had a similar variance to the empirical wage residual distribution. We stress that the simulations are very limited, and we only employ them for illustrative purposes. The results from these simulations are shown in Table 1. The reported estimates are those for the union coefficients.

The first notable feature of Table 1 is the performance of the CF estimator. For the Case 1 simulations, the CF and IV estimators both appear relatively unbiased and are clearly superior to the OLS estimates. For the Case 2 simulations, with the uniform errors in the reduced form, the CF estimator is performing, in terms of bias, as well as the IV estimator. Note that this is in the presence of a reasonable degree of bias as is highlighted by the OLS estimates. Note, however, that this is not a general result and mainly caused by the imposed linear relationship between ε_1 and ε_2. In the Case 3 simulations the CF and IV estimators both perform well, as is expected. A feature of the results for Cases 2 and 3 is the similarity of the mean values of the IV and CF estimates.

The primary conclusion from Table 1 is that the IV and CF estimates are similar, and the similarity does not require normality. In Table 1, under the heading of ρ_1, we report the correlation between the OLS and the generalized residuals from the reduced-form treatment-decision equation. The figures in this column show that the correlation between these two adjustment factors is remarkably high even in the absence of normality. Under the heading ρ_2, we report the correlation between the IV and CF estimates for each simulation. It is clear from these final two columns that the correlation between the OLS residuals and the inverse mills ratios drives the similarity, in terms of comovement, of the IV and CF estimates. Note also that in additional simulations we increased the variance of the disturbance in the union equation. Although we do not report the results, the values of ρ_1 and ρ_2 remained remarkably high.

Table 1. Monte Carlo Simulation Results

	OLS		IV		CF			
Case	Est.	S.E.	Est.	S.E.	Est.	S.E.	ρ_1	ρ_2
1a	.317	.014	.250	.041	.247	.041	.998	.985
b	.352	.014	.250	.043	.244	.042	.998	.987
c	.385	.015	.252	.044	.245	.044	.998	.988
2a	.334	.014	.252	.077	.252	.077	.999	.995
b	.376	.014	.250	.083	.251	.083	.999	.996
c	.417	.013	.257	.085	.258	.084	.999	.996
3a	.327	.013	.251	.061	.251	.060	.999	.992
b	.365	.013	.250	.062	.250	.061	.999	.992
c	.405	.013	.253	.066	.253	.065	.999	.993

NOTE: The results are average estimates and Monte Carlo standard errors over 1,000 replications, ρ_1 is the correlation coefficient between the generalized residual and OLS residual, and ρ_2 is the correlation between the IV and CF estimates.

This relationship highlights the similarity of the IV and CF methods. Often it is implied in empirical work, however, that the methods differ in some fundamental way. One such common example is the manner in which Z_i is chosen. It is clear from the preceding that the choice of Z_i is equally important in both methodologies. IV is often preferred over CF, however, on the basis that it is easier to specify the instruments and there is less concern about economic justification for inclusion in Z_i. Conversely, variables included in the selection equation on economic grounds, should be valid as instruments as well [see Hausman and Wise (1979) for an example in which this problem is ignored].

Thus, the choice between CF and IV is mainly a trade-off between efficiency and robustness against nonnormality. The CF approach, however, does have some advantages. First, CF estimates an additional number of parameters, including $\sigma_{t\varepsilon}$ and provides information about the sorting process. This is a major advantage because the economics of treatment effects is captured in these covariances. For this reason, the CF approach is straightforwardly extended to the case in which $\sigma_{t\varepsilon} \neq \sigma_{n\varepsilon}$. Second, the inherent nonlinearity in the CF approach enables the testing of the validity of the instruments even when there is only one exclusion restriction.

The similarity of the IV and (restricted) CF estimators has implications for their associated exogeneity tests. A test in the IV framework can be based on the estimated coefficient of the residual in (13), which is equivalent to a Hausman test based on the difference between the OLS and IV estimators for α (see Hausman 1978). Equivalently, the test can be based on the empirical correlation between $\hat{\eta}_i$, the OLS residual from (5), and $\hat{v}_i$. In the CF framework, the exogeneity test is also simply a t test on the coefficient for the inverse mill's ratio (λ_i) in (14). This test is thus based on the empirical correlation between $\hat{\eta}_i$ and $\hat{\lambda}_i$. Because $\hat{\lambda}_i$ and $\hat{v}_i$ are highly correlated, both tests are expected to perform similarly. In fact, the primary difference in the tests will be due to the different small-sample properties rather than any difference in what they are testing.

4. THE ECONOMICS OF TREATMENT EFFECTS

To discuss the economics of sorting into the two sectors (the treatment and no-treatment sector), we return to the more general model with unrestricted $\sigma_{t\varepsilon}$ and $\sigma_{n\varepsilon}$. To ease the discussion, we shall refer to ε_{ti} and ε_{ni} as sector-specific skills (e.g., see Heckman and Honoré 1990) although we retain the assumptions that they are independent of the X_i's. When $\sigma_{tn} < 0$, the sector-specific skills are negatively correlated and we have a *comparative advantage* structure. That is, on average those who perform well, relative to others, with the treatment perform relatively worse without the treatment. Alternatively, $\sigma_{tn} > 0$ characterizes a *hierarchical* structure in which, on average, those individuals who perform well, relative to others, with the treatment would also perform relatively well without it. Note that σ_{tn} cannot be directly estimated (cf. Vijverberg 1993).

We characterize the selection into the treatment as *positive* if, on average, those who have the most to benefit from the treatment undergo the treatment. That is, $E\{\varepsilon_{ti}|Z_i, T_i = 1\} > E\{\varepsilon_{ti}|Z_i\} = 0$. Conversely, selection into the no-treatment sector is positive if $E\{\varepsilon_{ni}|Z_i, T_i = 0\} > 0$. Under the assumptions of Section 1, it holds that

$$E\{\varepsilon_{ti}|Z_i, T_i = 1\} = (\sigma_t^2 - \sigma_{tn} - \sigma_{tc})E\{\varepsilon_i|\varepsilon_i > -Z_i\pi\} \quad (15)$$

and

$$E\{\varepsilon_{ni}|Z_i, T_i = 0\} = (\sigma_{tn} - \sigma_n^2 + \sigma_{nc})E\{\varepsilon_i|\varepsilon_i \leq -Z_i\pi\}, \quad (16)$$

where the conditional expectation on the right side is described by (11). Consequently, the economics of sorting into the two sectors is captured by the single index $Z_i\pi$ and the (co)variances of $(\varepsilon_{ti}, \varepsilon_{ni}, \varepsilon_{ci})'$.

Let us first consider the case in which the unobserved cost component is irrelevant or uncorrelated with sector skills ($\sigma_{tc} = \sigma_{nc}$). In this case $\sigma_{t\varepsilon} = \sigma_t^2 - \sigma_{tn}$ and $\sigma_{n\varepsilon} = \sigma_{tn} - \sigma_n^2$. In a comparative advantage structure, ($\sigma_{tn} < 0$), this implies that positive selection into both treatment and nontreatment requires a positive $\sigma_{t\varepsilon}$ and a negative $\sigma_{n\varepsilon}$. Because the IV and restricted CF estimators impose equality of these two covariances, however, they a priori exclude a comparative advantage structure. In the hierarchical structure, with $\sigma_{tn} > 0$, the signs of $\sigma_{t\varepsilon}$ and $\sigma_{n\varepsilon}$ are less clear. One can once again have a positive $\sigma_{t\varepsilon}$ and a negative $\sigma_{n\varepsilon}$, but also—depending on which sector has the higher variance—two negative or two positive covariances. Only a degenerate hierarchical structure in which a unit correlation between sector-specific skills is imposed can be consistent with equality of the two covariances as imposed by the IV and restricted CF estimators. This suggests that these estimators thus impose severe restrictions on the sorting process. Not only do they require a hierarchical structure, implying that the better treatment workers would also make the better nontreatment workers, but also that ε_{ti} and ε_{ni} are perfectly correlated. This implies that sector-specific skills are perfectly correlated.

When the unobserved cost is correlated with sector-specific skills, thereby implying that σ_{tc} and/or σ_{nc} are nonzero, the equality of $\sigma_{t\varepsilon}$ and $\sigma_{n\varepsilon}$, as imposed by the estimators, does not necessarily imply a degenerate hierarchical structure. It does, however, require that $\sigma_t^2 + \sigma_n^2 - 2\sigma_{tn} + \sigma_{nc} - \sigma_{tc} = 0$, which can in principle be consistent with any type of structure provided that $\sigma_{tc} - \sigma_{nc}$ is sufficiently large and positive. This requires that costs are relatively important in the sorting process. For a comparative advantage structure, a necessary, but not sufficient, condition is that σ_c^2 is larger than both σ_t^2 and σ_n^2.

The restrictions imposed by the IV approach are thus most easily satisfied in a hierarchical structure and require strong parametric restrictions in a comparative advantage structure. The reason for this is that IV imposes different signs on the conditional expectations $E\{\varepsilon_{ti}|Z_i, T_i = 1\}$ and $E\{\varepsilon_{ni}|Z_i, T_i = 0\}$. Note that it would be possible to relax this somewhat by assuming that $E\{\eta_i|Z_i\}$ in (8) equals a constant rather than 0 (see Robinson 1989a), but under the current distributional assumptions this does not provide any consolation. Consequently, it is not possible that, after selection, the average skills in both sectors are above the population average.

In the unrestricted CF approach, the equality of $\sigma_{t\varepsilon}$ and $\sigma_{n\varepsilon}$ is not imposed. Moreover, both parameters can be estimated consistently. It is, however, not necessarily the case that one can identify the type of skill structure (hierarchical or comparative advantage), as indicated by σ_{tn}. Recall that $\sigma_{t\varepsilon} = \sigma_t^2 - \sigma_{tn} - \sigma_{tc}$ and $\sigma_{n\varepsilon} = \sigma_{tn} - \sigma_n^2 - \sigma_{nc}$. Only when unobserved cost components are irrelevant to the model or are uncorrelated with the respective skills in the two sectors is it possible to infer the sign of σ_{tn} from σ_t^2, σ_n^2, and the two covariances. In general, however, the signs of $\sigma_{t\varepsilon}$ and $\sigma_{n\varepsilon}$ say nothing about the sorting in the model because any pair of signs is possible and consistent with any type of sorting.

Another point to be made is that the IV approach requires one to specify the elements of Z_i such that it includes at least one variable not found in X_i. This clearly excludes IV estimation of the purest form of the Roy model in which sorting takes place on the basis of relative wages only. In addition, it seems natural, though not necessary, to allow for unobserved cost components whenever Z_i includes observed cost components as instruments. Consequently, it cannot be the case that individuals are sorting purely on potential wages. The CF approach, either restricted or unrestricted, does not impose this and is able to identify through the nonlinearity implied by joint normality. If one does not want to impose any distributional assumptions and individuals sort purely on the basis of wages, the average treatment effect is no longer identified in the absence of additional restrictions.

5. CONCLUDING REMARKS

This article examines several issues in the estimation and interpretation of models with endogenous treatment effects. We argue that the two conventional methods of estimating these models, CF and IV, are closely related and that many of the issues raised with respect to the specification of the CF procedure are equally relevant to IV estimation. We also show that the procedures have important implications for the nature of the sorting patterns.

ACKNOWLEDGMENTS

Helpful comments by two anonymous referees and an associate editor are gratefully acknowledged. Financial support from the Research Programme of the Fund for

Scientific Research-Flanders (Belgium), (F.W.O) grant G.0198.97, is gratefully acknowledged.

[Received September 1996. Revised October 1998.]

REFERENCES

Gourieroux, C., Monfort, A., Renault, E., and Trognon, A. (1987), "Generalized Residuals," *Journal of Econometrics*, 52, 5–32.

Hausman, J. A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

Hausman, J. A., and Wise, D. (1979), "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica*, 47, 455–473.

Heckman, J. J. (1978), "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica*, 46, 931–960.

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

——— (1990), "Varieties of Selection Bias," *American Economic Review*, Papers and Proceedings, 80, 313–318.

——— (1997), "Instrumental Variables. A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441–462.

Heckman, J. J., and Honoré, B. (1990), "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1159.

Heckman, J. J., and Robb, R. (1985), "Alternative Methods for Evaluating the Impact of Interventions: An Overview," *Journal of Econometrics*, 30, 239–267.

Imbens, G. W., and Angrist, J. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Olsen, R. J. (1980), "A Least Squares Correction for Selectivity Bias," *Econometrica*, 48, 1815–1820.

Rivers, D., and Vuong, Q. (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39, 347–366.

Robinson, C. (1989a), "The Joint Determination of Union Status and Union Wages Effects: Some Tests of Alternative Models," *Journal of Political Economy*, 97, 639–667.

——— (1989b), "Union Endogeneity and Self-Selection," *Journal of Labor Economics*, 7, 106–112.

Roy, A. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 33, 135–146.

Smith, R., and Blundell, R. (1986), "An Exogeneity Test for a Simultaneous Equation Tobit Model With an Application to Labor Supply," *Econometrica*, 54, 679–685.

Vella, F. (1993), "A Simple Estimator for Simultaneous Equations Models With Censored Endogenous Regressors," *International Economic Review*, 34, 441–457.

Vijverberg, W. (1993), "Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity," *Journal of Econometrics*, 57, 69–89.