Using Copulas in Statistical Models of Switching Regimes

M. D. Smith
Econometrics and Business Statistics, University of Sydney, Sydney NSW 2006, Australia
(Murray.Smith@econ.usyd.edu.au)

Abstract: By a theorem due to Sklar, a multivariate distribution can be represented in terms of its underlying margins by binding them together using a copula function. By exploiting this representation, the "copula approach" to statistical modelling proceeds by specifying distributions for each margin and a copula function. In this paper, a number of families of copula functions are given, with attention focusing on those that fall within the Archimedean class. Members of this class of copulas are rich in various distributional attributes that are desired when modelling. The paper then proceeds by applying the copula approach to construct statistical models for the Roy model of switching regimes. When models are constructed using copulas from the Archimedean class, the resulting expressions for the log-likelihood and score facilitate maximum likelihood estimation. The literature on sample selection models is almost exclusively based on multivariate normal specifications. The copula approach permits modelling based on multivariate non-normality.

Keywords: *Roy model of switching regimes; Sample selection; Copula; Sklar's theorem; Copula representation; Copula approach; Families of copulas; Archimedean.*

1. INTRODUCTION

This article sets out to demonstrate the application of the "copula approach" to model specification in the context of the extended or utility-based Roy model of switching regimes (e.g. see Vijverberg (1993)). Over the last thirty to forty years, a large volume of literature on sample selection models, including the Roy model, has been built up in economics and econometrics; see, for example, Vella (1998) for a recent survey. However, the vast majority of analyses have depended on the statistical assumption of multivariate normality. Although ubiquitous throughout all facets of econometric modelling, the adequacy of inference based on the assumption of multivariate normality has often been questioned, and often found to be wanting in the context of sample selection models. Unfortunately, relaxing multivariate normality by replacing it with an alternative multivariate distribution has received relatively little attention. In the main, this was because of the additional computational burdens that were expected to arise. Instead, the literature developed by focusing on semi-parametric and non-parametric versions of these models, where modelling improvements might be brought about by the use of flexible functions of parameters and the covariates of the random variables. The aim of this article is to return to the issue of replacing multivariate normality with an alternative multivariate distribution (or, more precisely, a class of multivariate distributions). The adverse computational consequences are, if anything, mitigated under the proposed method of model specification: the so-called copula approach.

The copula approach is a modelling strategy whereby a joint distribution is induced by specifying marginal distributions, and a function that binds them together: the copula. The copula parameterises the dependence structure of the random variables, thereby capturing all of the joint behaviour. This then frees the location and scale structures to be parameterised through the margins, one at a time. Most importantly, the copula approach permits specifications other than multivariate normality, although it does retain that distribution as a special case. Examples of its use in economics and econometrics includes Bouyé *et al*

(2000), Patton (2001) and Dardanoni and Lambert (2001). Perhaps the most accessible contribution to date is a series of five studies reported in Joe (1997, Ch.11) that estimate copula models for various multivariate and longitudinal data sets. The specification method suggested by Lee (1983) for modelling self-selection provides an example of the copula approach; Smith (2003) provides extensive details.

As all multivariate distributions have a copula representation (per Sklar's Theorem), it might seem that the copula approach is nothing more than the reworking of an old theme. Might the advantage derived by the copula approach simply be that econometricians are better practiced at modelling univariate distributions than they are multivariate ones? The ideal, of course, is to choose the right statistical model *a priori*, and hence the right copula. However, when working with empirical data it is rare to have such insight. The specification problem is further compounded in most sample selection models due to latency of the underlying utilitarian variables, and the presence of covariates. When faced with such difficulties, it is advantageous to have at hand a range of potential candidate models from which a preferred fit can emerge. Under a copula approach, families of models can be constructed according to classes of copula functions: of particular interest here is the class of Archimedean copulas. Archimedean copulas can display a range of distributional behaviour such as joint asymmetry, excess joint skewness and joint kurtosis. When applied in the specification of selectivity models, relatively simple formulae for likelihood and score functions result, thereby facilitating estimation by maximum likelihood (ML hereafter).

2. COPULA THEORY

2.1. Sklar's Theorem

With a view to the main result that is embodied in Sklar's theorem, the copula for an n-dimensional multivariate distribution function F with given one-dimensional marginal distribution functions $F_1, ..., F_n$, is the function that binds together the margins in such manner as to form precisely the joint distribution function. The action performed by the copula implies that it serves to represent the dependence characteristics that associate each of the underlying random variables, irrespective of the form the margins take. To date, most uses of copula theory have concentrated on the study of the association between random variables and, to a slightly lesser extent, the establishment of limiting (Fréchet) bounds on distributions. For these details see Dall'Aglio (1991), Schweizer (1991) and Nelsen (1999).

The main result of interest here is a theorem due to Sklar (given below for the bivariate case). Sklar's theorem shows that there exists a copula function which acts to represent the joint cumulative distribution function (cdf hereafter) of random variables in terms of its underlying one-dimensional margins. Let the margins $F_1(x_1)$ and $F_2(x_2)$ denote, respectively, the cdf of the random variables X_1 and X_2; that is, $F_i(x_i) = \Pr(X_i \leq x_i)$, where $x_i \in \overline{\mathbb{R}}$ ($i = 1, 2$; $\overline{\mathbb{R}}$ denotes the extended real line $\mathbb{R} \cup \{-\infty, +\infty\}$), and let $F(x_1, x_2) = \Pr(X_1 \leq x_1, X_2 \leq x_2)$ denote the joint cdf. Then, for some two-place function C, the joint cdf has the representation (e.g. Nelsen, 1999, Theorem 2.3.3)

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \qquad (1)$$

where C is termed the copula function. The copula representation is a re-formulation of the joint cdf such that it separates the margins F_1 and F_2 from their interaction. So while the copula function takes as arguments the margins F_1 and F_2 in the representation (1), the function itself is independent of those margins. The copula serves to capture the dependence characteristics that exist between the random variables X_1 and X_2. Nelsen (1999, Section 2.3) provides a proof of (1) that follows the method given in Schweizer and Sklar (1983, Ch.6) where the multivariate version of the theorem is proved.

If F_1 and F_2 are continuous functions, then (1) is unique for any $(x_1, x_2) \in \overline{\mathbb{R}}^2$. On the other hand, if either or both X_1 and X_2 are discrete random variables that take values on some lattice of points Ω, then (1) is unique provided $(x_1, x_2) \in \Omega$, but not elsewhere; this does not cause any great harm, for regions outside of the supporting lattice are rarely of interest. Implicit in (1) is $C(u, v) = 0$ if either or both u and v are zero, and $C(1, v) = v$ and $C(u, 1) = u$, where the pair $(u, v) \in \mathbb{I}^2$ ($\mathbb{I}$ denotes the closed interval $[0, 1]$ of the real line).

2.2. Examples of Copulas

Three bivariate copulas of some importance are, respectively, the Product copula $\Pi = uv$, the Fréchet lower bound for (bivariate) copulas $W = \max(u + v - 1, 0)$ and the Fréchet upper bound for copulas $M = \min(u, v)$, where

$(u,v) \in \mathbb{I}^2$. Π corresponds to stochastic independence; that is, if two random variables are independent, then Π is the copula of their joint distribution. The closed interval $[W, M]$ has the property of containing all bivariate copulas; namely, for all copulas C on $\mathbb{I}^2$, $W \leq C \leq M$. These bounds - the Fréchet bounds for copulas - were obtained by Hoeffding.

For the purposes of statistical modelling it is desirable to parameterise the copula function so that data can be used to shed light on the extent of association between the random variables of interest. Let θ denote the association parameter of the bivariate distribution (possibly vector valued) and write the parameterised copula as per

$$C_\theta(u,v)$$

where $(u,v) \in \mathbb{I}^2$. This notation denotes a family of copulas, where the members are indexed according to values assigned to θ. Provided that the margins F_1 and F_2 do not depend on θ, the representation (1) holds for all members of a given family; this assumption is imposed hereafter. There are numerous examples of families of bivariate copulas given in Joe (1997) and Nelsen (1999). For example, the family of Bivariate Normal copulas is given by

$$C_\theta(u,v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta) \qquad (2)$$

where $-1 \leq \theta \leq 1$, here, $\Phi(\cdot)$ denotes the cdf of a standard normal variate, and $\Phi_2(\cdot,\cdot;\theta)$ the cdf of a bivariate standard normal variate with Pearson's product moment correlation coefficient θ. Note that setting $u = \Phi(x_1)$ and $v = \Phi(x_2)$ in (2) recovers the bivariate standard normal cdf. A second example is the Farlie-Gumbel-Morgenstern family of copulas (FGM hereafter):

$$C_\theta(u,v) = uv + \theta uv(1-u)(1-v) \qquad (3)$$

where $-1 \leq \theta \leq 1$. The FGM family can be useful in analytic work due to its mathematical simplicity.

The ability of a given family of copulas to represent differing degrees of association can be examined in terms of the extent to which it covers the interval between the lower and upper Fréchet bounds for copulas $[W, M]$. This is generally determined at the extremes of the parameter space for θ. For example, for the Bivariate Normal family (2), $C_{-1}(u,v) = W$ and $C_1(u,v) = M$, so that this family has full coverage. Furthermore, the family of Bivariate Normal copulas is said to be comprehensive, where this nomenclature means that a given family includes W, M and Π amongst its members, or as limiting cases. Comprehensive families of copulas therefore parameterise the full range of association and, by (1), this property holds irrespective of the form of the margins. However, there are typically many other features of the data that are of interest, and these may not necessarily be well-modelled if attention is restricted to using comprehensive families of copulas.

There are many copula families that are not comprehensive, one example is the FGM family (3): it includes Π, but not W and M. For such families it is desirable to assess coverage in terms of measures of association. The most familiar measure is Pearson's product moment correlation coefficient, but due to its lack of invariance with respect to the margins, the properties of this measure are dominated by others such as Kendall's τ (Joe (1997, Section 2.1.9)). τ is a concordance measure that is bounded between $[-1, 1]$: equal to -1 at W, 1 at M and 0 for Π. Importantly, it is invariant to strictly increasing transformations of the variables, implying that it depends only on the copula of the joint distribution, and not the margins. For independent pairs (X_{1i}, X_{2i}), $i = 1, 2$, that are copies of (X_1, X_2), τ is defined as

$$\tau = \Pr((X_{11} - X_{12})(X_{21} - X_{22}) > 0) - \Pr((X_{11} - X_{12})(X_{21} - X_{22}) < 0)$$

Should (X_1, X_2) be a pair of continuous random variables, with the copula of the joint distribution given by C, then τ may be simplified:

$$\begin{aligned} \tau &= 4 \int\int_{\mathbf{I}^2} C(u,v)\,dC(u,v) - 1 \\ &= 4E[C(U,V)] - 1 \end{aligned}$$

Here, U and V denote standard uniform random variables with joint cdf C. For the FGM family of copulas it can be shown that $\tau = 2\theta/9$; clearly $-2/9 \leq \tau \leq 2/9$ for this family.

2.3. The Archimedean Class

Of particular importance in this article is the class of Archimedean copulas. This class encompasses many families of copulas, a number of which can be of use in statistical modelling. The mathematical properties of the Archimedean class are captured by an additive generator function $\varphi : \mathbb{I} \rightarrow [0, \infty]$, which is a continuous, convex decreasing function ($\varphi'(t) < 0$ and $\varphi''(t) > 0$, for $0 < t < 1$), with terminal

$\varphi(1) = 0$. φ may also be indexed by the association parameter θ, thus an entire family of copulas can be Archimedean. Any function φ that satisfies these conditions can be used to generate a valid bivariate cdf. The advantage in mathematics of working with Archimedean copulas is the achievement of reduction in dimensionality: while the copula of an n-variate distribution is an n-place function, the generator φ only ever takes a single argument. In econometrics, this property of Archimedean copulas has the potential to be of use in models of limited dependent variables, especially those requiring some probabilistic enumeration on high-dimensional subspaces, for evaluation then becomes essentially a univariate task.

In the bivariate case, the means by which φ generates the copula is according to:

$$\varphi\left(C(u,v)\right) = \varphi(u) + \varphi(v) \tag{4}$$

Note that the generator is unique up to a scaling constant. Particular examples are $\varphi(t) = -\log t$ and $\varphi(t) = \left(t^{-\theta} - 1\right)/\theta$, which are, respectively, the generators of the Product copula Π and the Clayton family of copulas:

$$C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta} \tag{5}$$

where $\theta \geq 0$. Note that neither the Bivariate Normal family nor the FGM families are members of the Archimedean class. Nelsen (1999, Table 4.1) lists numerous single-parameter families of Archimedean copulas.

If the terminal $\varphi(0) = \infty$, the generator is termed strict, and the inverse function φ^{-1} exists. The generators of Π and (5) are strict. In this instance, from (4), the copula is recovered by:

$$C(u,v) = \varphi^{-1}\left(\varphi(u) + \varphi(v)\right)$$

Nelsen (1999, Ch.4) gives extensive details about Archimedean copulas (strict and non-strict); see also Genest and MacKay (1986), Genest and Rivet (1993), Jouini and Clemen (1996) and Mari and Kotz (2001, Section 4.6).

2.4. The Copula Approach

For the purposes of statistical modelling, it is the converse of the copula representation of the joint cdf given by Sklar's theorem that is relevant. In other words, given models for the margins and a copula function that binds them together, this then has the effect of constructing a statistical model for the random variables of interest, as a joint cdf is specified. Consider, for example, a bivariate setting in which X_1 and X_2 denote the variables of interest. Required is a statistical model for the true, but unknown joint distribution of X_1 and X_2; naturally, this distribution may depend on parameters and covariates. Under a copula approach, models for the margins $F_1(x_1)$ and $F_2(x_2)$ are proposed, as well as a selection of a copula family C_θ. Then, by (1), these selections have the effect of specifying the joint cdf of X_1 and X_2. Intuitively, the copula approach determines each component of the overall model, then engineers them together using a copula function.

An added boon for modelling that results by adopting a copula approach concerns the freedom to specify each margin; for example, identicality in distribution of the margins need not be imposed. Indeed, because the copula representation is unique on the domain of support of the random variables in question, multivariate models can be constructed using a copula approach whose margins can be either continuous or discrete, or mixtures of both.

3. THE ROY MODEL

3.1. Model and Likelihood

Sample stratification, or sample selection, is commonplace amongst microeconometric data, whereby underlying individual choices can themselves influence the observations collected on the random variables of interest. Models of increasing complexity have been constructed to account for stratification in its various guises, should it be present, and a number of these are discussed in texts such as Amemiya (1985) and Maddala (1983). In this section, attention focuses on the Roy model of switching regimes, based on a binary indicator S that governs which regime is observed.

The Roy model of switching regimes arises when observations on the trio of random variables (S, Y_2, Y_3) are generated according to the following observation rules

$$S = 1\{Y_1^* > 0\} \quad Y_2 = S\,Y_2^* \quad Y_3 = (1-S)\,Y_3^*$$

where $1\{A\}$ denotes the indicator function, taking value 1 if event A holds, and 0 otherwise. Here, (Y_1^*, Y_2^*, Y_3^*) denotes latent utilitarian variables with margins that have cdf and pdf denoted respectively by $F_i(y_i^*)$ and $f_i(y_i^*)$, for $y_i^* \in \overline{I\!R}$ $(i = 1, 2, 3)$. It is assumed that these margins depend on covariates and parameters, however, their specification is not of concern at this stage. Basically, Y_2^* is observed whenever

$Y_1^* > 0$, otherwise it is Y_3^* that is observed; the switching mechanism S is binary. Note that dummy values of 0 are assigned to Y_2 and Y_3 as required, according to the outcomes of the switch S. Here, $F_i(y_i^*)$ is assumed continuous throughout the support of Y_i^*, for $i = 2, 3$. Vijverberg (1993) cites a number of empirical applications of this model.

Let (s_j, y_{2j}, y_{3j}) denote the jth observation on (S, Y_2, Y_3), $j = 1, ..., n$. For a random sample of size n, the likelihood is given by

$$\prod_0 \frac{\partial}{\partial y_3} F_{13}(0, y_3) \prod_1 \left(f_2 - \frac{\partial}{\partial y_2} F_{12}(0, y_2) \right)$$

where $\prod_0$ indicates the product over those observations for which $s_j = 0$, and $\prod_1$ the product over those observations for which $s_j = 1$, and the bivariate margin $F_{1i}(y_1^*, y_i^*) = \Pr(Y_1^* \leq y_1^*, Y_i^* \leq y_i^*)$ with pdf f_{1i}, $i = 2, 3$. The likelihood is expressed in terms of the underlying margins, where the presence of the differentials are due to the continuity assumptions on Y_2^* and Y_3^*. For convenience, the observation index j is suppressed throughout.

From the general form of the likelihood it is clear that any association parameters that may exist between Y_2^* and Y_3^* cannot be identified as L is not a function of these parameters (L does not depend on the bivariate margin F_{23} nor on the trivariate F). This implies that it is superfluous to specify F, the trivariate distribution of (Y_1^*, Y_2^*, Y_3^*). However, under the further assumption that Y_1^* is continuous, it is possible to give bounds on the association between Y_2^* and Y_3^* in terms of Kendall's concordance measure τ_{23} that can be tighter than $-1 \leq \tau_{23} \leq 1$, namely:

$$-1 + |\tau_{12} + \tau_{13}| \leq \tau_{23} \leq 1 - |\tau_{12} - \tau_{13}|$$

where τ_{1i} is Kendall's measure between Y_1^* and Y_i^* ($i = 2, 3$; see Joe (1997, Theorem 3.12)). This contrasts against the bounds given by Viverberg (1993, equation (3)) that are expressed in terms of Pearson's product moment correlation coefficients.

Under the copula approach, modelling proceeds by specifying margins F_i, and the copulas that represent the bivariate margins F_{12} and F_{13}. Supposing that the copula of F_{12} is Archimedean with generator φ, and, likewise, that the copula of F_{13} is Archimedean with generator η, then the derivatives appearing in the general form of the likelihood simplify to

$$\begin{aligned} \frac{\partial}{\partial y_i} F_{1i}(0, y_i) &= \frac{\partial}{\partial v} C_{\cdot}(F_1, v)\bigg|_{v \to F_i} \times \frac{\partial F_i}{\partial y_i} \\ &= \begin{cases} \frac{\varphi'(F_2)}{\varphi'(C_\theta^{12})} f_2 & \text{when } i = 2 \\ \frac{\eta'(F_3)}{\eta'(C_\lambda^{13})} f_3 & \text{when } i = 3 \end{cases} \end{aligned}$$

where $\varphi'(t) = \frac{\partial}{\partial t}\varphi(t)$, $\eta'(t) = \frac{\partial}{\partial t}\eta(t)$, and θ and λ collect the association parameters between, respectively, Y_1^* and Y_2^*, and Y_1^* and Y_3^*. The notation $F_1 = F_1(0)$, $F_2 = F_2(y_2)$, $F_3 = F_3(y_3)$, $f_2 = f_2(y_2)$, $f_3 = f_3(y_3)$, $C_\theta^{12} = \varphi^{-1}(\varphi(F_1) + \varphi(F_2))$ and $C_\lambda^{13} = \eta^{-1}(\eta(F_1) + \eta(F_3))$. Substitution into the likelihood yields

$$\prod_0 \frac{\eta'(F_3)}{\eta'(C_\lambda^{13})} f_3 \prod_1 \left(1 - \frac{\varphi'(F_2)}{\varphi'(C_\theta^{12})} \right) f_2 \qquad (6)$$

As the functional forms of $\varphi'(t)$ and $\eta'(t)$ are generally quite easy to derive, the likelihood is relatively easy to code. For example, if both φ and η are Clayton (5), then the likelihood is

$$\prod_0 \left(\frac{C_\lambda^{13}}{F_3} \right)^{\lambda+1} f_3 \prod_1 \left(1 - \left(\frac{C_\theta^{12}}{F_2} \right)^{\theta+1} \right) f_2$$

Of course, there is no need to restrict φ and η to generate the same family of copulas. It is also straightforward to derive the score function, and then to evaluate it and the log-likelihood for purposes of ML estimation using a Quasi-Newton algorithm.

3.2. Remarks

Under the copula approach, parametric models for the margins can be constructed using generalised linear methods. This flexibility is a distinct advantage of the copula approach, as the margins need not be restricted to the same family of distributions. However, other approaches are also possible; for example, using semi- and non-parametric methods to specify the margins. For the copula functions, this article advocates selecting families of copulas from the Archimedean class.

Given the relatively simple functional form for the likelihood function under a pair of Archimedean copulas (6), ML estimation can be employed to jointly estimate all parameters. As general analytical expressions for the score function can be derived, it is relatively easy to implement well-known Quasi-Newton optimisation algorithms such as DFP and BFGS. Accordingly, the use of Archimedean copulas in model specification satisfies the need identified by Vella (1998, p.132) to maintain ease of implementation as the model assumption departs from multivariate normality, while remaining in the framework of ML estimation. Unfortunately, obtaining the analytic form of the Hessian of the log-likelihood is a tedious exercise, so if implementation of the Newton-Raphson algorithm is desired, then, when deriving the Hessian matrix, it is perhaps better to use numerical methods that can approximate derivatives. These considerations also impact on estimation of the asymptotic variance-covariance matrix of the ML estimator. The method advocated here is to use as the estimate the final iterate of the approximation to the inverse Hessian that is generated at each step of the BFGS algorithm. Other variance-covariance matrix estimators include the OPG estimator, although this is known to be prone to inflate standard errors in small samples. Estimation using the inverse Information matrix does not seem practicable here due to the difficulties induced by non-linearity in the variables of the model.

It seems quite plausible that the usual suite of asymptotic properties of the ML estimator will hold. However, it remains an open question for research to prove those regularity conditions under which the ML estimator is consistent, asymptotically normal and efficient under Archimedean copulas (or more generally for any pair of copulas).

Given empirical data, model selection across differing specifications is an *a posteriori* consideration, for it is rare that the true data generating mechanism is known *a priori*. Setting aside for now the specification of the margins, differing families of copulas are, in general, parametrically non-nested, even if the families compared are Archimedean. Consequently, following the suggestion of Joe (1997, Section 10.3), information measures such as AIC and BIC applied to each fitted model can be used as the selection criterion amongst competing models.

4. REFERENCES

Amemiya, T. *Advanced Econometrics*. Cambridge MA: Harvard, 1985.

Bouyé, E, Durrleman, V, Nikeghbali, A, Riboulet, G, and Roncalli, T. Copulas for finance: a reading guide and some applications. All About Value at Risk Working Papers, 2000.

Dall'Aglio, G. Frechet classes: the beginnings. Chapter 1 in *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, by Dall'Aglio, G, Kotz, S, and Salinetti, G (eds.). Dordrecht: Kluwer, 13-50, 1991.

Dardanoni, V, and Lambert, P. Horizontal inequity comparisons. *Social Choice and Welfare* 18, 799-816, 2001.

Genest, C, and MacKay, J. The joy of copulas: bivariate distributions with uniform marginals. *American Statistician* 40, 280-283, 1986.

Genest, C, and Rivet, L-P. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* 88, 1034-1043, 1993.

Lee, L-F. Generalized econometric models with selectivity. *Econometrica* 51, 507-512, 1983.

Joe, H. *Multivariate Models and Dependence Concepts*. London: Chapman and Hall, 1997.

Jouini, M N, and Clemen, R T. Copula models for aggregating expert opinions. *Operations Research* 44, 444-457, 1996.

Maddala, G S. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press, 1983.

Mari, D D, and Kotz, S. *Correlation and Dependence*, London: Imperial College Press, 2001.

Nelsen, R B. *An Introduction to Copulas*, New York: Springer-Verlag, 1999.

Patton, A J. Modelling time-varying exchange rate dependence using the conditional copula. Department of Economics Discussion Paper 2001-09, UCSD, 2001.

Schweizer, B. Thirty years of copulas. Chapter 2 in *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, by Dall'Aglio, G, Kotz, S, and Salinetti, G (eds.), Dordrecht: Kluwer, 13-50, 1991.

Schweizer, B, and Sklar, A. *Probabilistic Metric Spaces*. New York: North-Holland, 1983.

Smith, M D. Modelling sample selection using Archimedean copulas. *Econometrics Journal*, 6, 99-123, 2003.

Vella, F. Estimating models with sample selection bias: a survey, *The Journal of Human Resources* 33, 127-143, 1998.

Vijverberg, W P M. Measuring the unidentified parameter of the extended Roy model of selectivity. *Journal of Econometrics* 57, 69-90, 1993.