On the Identification of the Censored Regression Model with a Stochastic and Unobserved Threshold

Geert Ridder*
Department of Econometrics
Vrije Universiteit
and Tinbergen Institute Amsterdam

Kees van Montfort
Department of Econometrics
Vrije Universiteit

March 9, 1998

Abstract

We show that a sufficient condition for the identification of all parameters of the censored regression model with a stochastic and unobserved threshold is that the errors are jointly normally distributed. Exclusion restrictions are not needed.

JEL Classification: C34
Keywords and phrases: *censored regression, unobserved threshold, identification*

*Mailing address: Vakgroep Econometrie,Faculteit der Economische Wetenschappen en Econometrie, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Fax: 31-20-4446020, Email: gridder@econ.vu.nl

1 Introduction

In this note we show that all parameters of the censored regression model with a stochastic and unobserved threshold are identified if the errors of the regression and threshold equations are jointly normally distributed. No exclusion restrictions in either equation are needed.

2 A censored regression model with a stochastic and unobserved threshold

We consider the censored regression model with normally distributed errors

$$\begin{aligned} y_1^* &= \beta_1' x + u_1 & (1) \\ y_2^* &= \beta_2' x + u_2 & (2) \\ y_1 &= y_1^* I(y_1^* \geq y_2^*) & (3) \\ y_2 &= I(y_1^* \geq y_2^*) & (4) \end{aligned}$$

with

$$u_1, u_2 \sim N(0, \Sigma) \qquad (5)$$

The model applies to all members of an infinitely large population. We omit subscripts that identify individual members. The parameters of the model are the vectors of regression coefficients β_1, β_2 and the components of the variance-covariance matrix Σ of the errors. To avoid trivial instances of the model, we restrict the variances in Σ to be strictly positive and the correlation to be strictly greater than -1 and strictly less than 1. The latent variables y_1^*, y_2^* are the dependent variables in a SUR model with normally distributed errors. Equations (3) and (4) relate the observed y_1, y_2 to the latent variables. By (3) y_1^* is observed if it exceeds the latent threshold y_2^*. If it is below this threshold, we do not observe y_1^*, and the missing value is labeled by the number 0. The missing value label is in the range of the observed values of y_1^*, but the probability of such an observation, *i.e.* of the event $y_1^* = 0, y_1^* > y_2^*$, is 0. We also observe the indicator y_2 of the event that the latent threshold is exceeded.

This model was introduced by Nelson (1977), who also gives some examples. Unobserved thresholds occur in optimization problems with possibly binding restrictions ,*e.g.* the demand for consumer durables and labor supply, or dynamic choice problems, in which the optimal strategy is characterized by a reservation value, *e.g.* optimal stopping problems as in job search models. In these applications both the (distribution of) the partially observed dependent variable and that of the unobserved threshold are of interest. For that reason, it is important to know under what conditions their joint distribution can be recovered from the distribution of the observations.

Nelson (1977) and Maddala (1983) (section 8.4) discuss the identification of the parameters of this model. Nelson reparametrizes the model and inspects the resulting likelihood function. He uses an analogy between the new parametrization and the original one, and the parameters of the structural and reduced form in a system of simultaneous equations, to derive sufficient conditions for identification. Maddala derives the same sufficient conditions by considering the estimation problem as a two-stage problem. In the first stage a probit model is estimated for the dummy-dependent y_2. The probit estimates are used in the second stage to obtain the conditional expectation of y_1 given x. Both authors conclude that

1. The parameters of equation (1), β_1, σ_1^2 are identified, if the regression equation (2) has at least one explanatory variable (that may also appear in regression of the partially observed variable).

2. The other parameters, $\beta_2, \sigma_2^2, \sigma_{12}$ are identified, if, in addition, either $\sigma_{12} = 0$ or, if there is at least one explanatory variable x_k with $\beta_{2k} = 0, \beta_{1k} \neq 0$.

In this note we argue that these conditions are not necessary. In particular, we show that no exclusion restrictions are needed to identify all parameters of the model. Our result is a warning against the use of informal arguments or analogies in the discussion of the identification of models with limited dependent variables.

3 Identification without additional parameter restrictions

Without loss of generality we omit the explanatory variables x. Let $f(y_1, y_2; \mu, \Sigma)$ be the density function of the observed variables y_1, y_2 with μ the vector of means of the latent variables. We say that two sets of parameters μ, Σ and $\tilde{\mu}, \tilde{\Sigma}$ are *observationally equivalent* if for all y_1, y_2 with $-\infty < y_1 < \infty$, $y_2 = 0, 1$,

$$f(y_1, y_2; \mu, \Sigma) = f(y_1, y_2; \tilde{\mu}, \tilde{\Sigma}) \tag{6}$$

We prove the following theorem

Theorem 1 *Let the latent variables y_1^*, y_2^* have a bivariate normal distribution with mean μ and variance-covariance matrix Σ. Let the observed variables y_1, y_2 be defined as in equations (3) and (4). Then (μ, Σ) and $(\tilde{\mu}, \tilde{\Sigma})$ are observationally equivalent if and only if $(\mu, \Sigma) = (\tilde{\mu}, \tilde{\Sigma})$.*

Proof The joint density of (y_1, y_2) is

$$\begin{aligned} f(y_1, y_2; \mu, \Sigma) = &\left\{ \Phi\left(\frac{y_1 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1)}{\sigma_2\sqrt{1-\rho^2}} \right) \frac{1}{\sigma_1} \phi\left(\frac{y_1 - \mu_1}{\sigma_1} \right) \right\}^{y_2} \cdot \\ &\cdot \left\{ 1 - \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \right) \right\}^{1-y_2} \end{aligned} \tag{7}$$

with $-\infty < y_1 < \infty$, $y_2 = 0, 1$.

Let μ, Σ and $\tilde{\mu}, \tilde{\Sigma}$ be observationally equivalent. Consider the joint density for $y_2 = 0$. Observational equivalence implies the following restriction on the parameters

$$\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} = \frac{\tilde{\mu}_1 - \tilde{\mu}_2}{\sqrt{\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2 - 2\tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2}} \tag{8}$$

For $y_2 = 1$ observational equivalence implies

$$\begin{aligned} &\Phi\left(\frac{y_1 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1)}{\sigma_2\sqrt{1-\rho^2}} \right) \frac{1}{\sigma_1} \phi\left(\frac{y_1 - \mu_1}{\sigma_1} \right) = \\ &\quad \Phi\left(\frac{y_1 - \tilde{\mu}_2 - \tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}(y_1 - \tilde{\mu}_1)}{\tilde{\sigma}_2\sqrt{1-\tilde{\rho}^2}} \right) \frac{1}{\tilde{\sigma}_1} \phi\left(\frac{y_1 - \tilde{\mu}_1}{\tilde{\sigma}_1} \right) \end{aligned} \tag{9}$$

for all $-\infty < y_1 < \infty$.

We express equation (9) as a ratio by division by the right-hand side that is strictly positive if the variance of y_1^* is strictly positive, as is assumed throughout. Observational equivalence implies that this ratio is identically equal to 1. Note that the ratio is equal to an individual contribution to the likelihood ratio for these two sets of parameters. Our identification proof considers the behavior of this likelihood ratio contribution over the support of the observed variables. In particular, we look for values of the observed variables that favor one set of parameters over the other, with observational equivalence implying that such values can not be found. Without loss of generality we assume that $\sigma_1 > \tilde{\sigma}_1$. Then for the second factor in the ratio

$$\lim_{|y_1| \to \infty} \frac{\phi\left(\frac{y_1-\mu_1}{\sigma_1}\right)}{\phi\left(\frac{y_1-\tilde{\mu}_1}{\tilde{\sigma}_1}\right)} = \infty \tag{10}$$

Next consider the first factor in the ratio that we rewrite as

$$\frac{\Phi\left(\frac{\left(1-\rho\frac{\sigma_2}{\sigma_1}\right)y_1-\mu_2+\rho\frac{\sigma_2}{\sigma_1}\mu_1}{\sigma_2\sqrt{1-\rho^2}}\right)}{\Phi\left(\frac{\left(1-\tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}\right)y_1-\tilde{\mu}_2+\tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}\tilde{\mu}_1}{\tilde{\sigma}_2\sqrt{1-\tilde{\rho}^2}}\right)} \tag{11}$$

We distinguish four cases

I $\left(1-\rho\frac{\sigma_2}{\sigma_1}\right) \geq 0, \left(1-\tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}\right) \geq 0$

II $\left(1-\rho\frac{\sigma_2}{\sigma_1}\right) \geq 0, \left(1-\tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}\right) < 0$ or $\left(1-\rho\frac{\sigma_2}{\sigma_1}\right) = 0, \left(1-\tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}\right) > 0$

III $\left(1-\rho\frac{\sigma_2}{\sigma_1}\right) < 0, \left(1-\tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}\right) \geq 0$ or $\left(1-\rho\frac{\sigma_2}{\sigma_1}\right) > 0, \left(1-\tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}\right) = 0$

IV $\left(1-\rho\frac{\sigma_1}{\sigma_2}\right) < 0, \left(1-\tilde{\rho}\frac{\tilde{\sigma}_2}{\tilde{\sigma}_2}\right) < 0$

Only the instance that both weak inequalities in case I are equalities, belongs to case I. If one of these weak inequalities is an equality and the other strict, we are in case II or III, as indicated. The limit for $y_1 \to \infty$ or $y_1 \to -\infty$ of the ratio obtained by dividing the left-hand side of equation (9) by the right-hand side is equal to the product of the limit in equation (10) and the limit of equation (11). The product of the limits is ∞ if either $y_1 \to \infty$ (cases

I and II) or $y_1 \to -\infty$ (cases III and IV). Because this limit should be equal to 1, we have a contradiction, and we conclude that

$$\sigma_1 = \tilde{\sigma}_1 \tag{12}$$

We substitute this equality in equation (9). The ratio of the right- and left-hand sides of that equation simplifies to

$$\frac{\Phi\left(\frac{\left(1-\rho\frac{\sigma_2}{\sigma_1}\right)y_1-\mu_2+\rho\frac{\sigma_2}{\sigma_1}\mu_1}{\sigma_2\sqrt{1-\rho^2}}\right)}{\Phi\left(\frac{\left(1-\tilde{\rho}\frac{\tilde{\sigma}_2}{\sigma_1}\right)y_1-\tilde{\mu}_2+\tilde{\rho}\frac{\tilde{\sigma}_2}{\sigma_1}\mu_1}{\tilde{\sigma}_2\sqrt{1-\tilde{\rho}^2}}\right)} \cdot e^{\frac{\mu_1-\tilde{\mu}_1}{\sigma_1^2}y_1} \cdot e^{\frac{\tilde{\mu}_1^2-\mu_1^2}{2\sigma_1^2}} \tag{13}$$

Again,without loss of generality we may assume that

$$\mu_1 > \tilde{\mu}_1 \tag{14}$$

We consider the cases I-IV separately, where the instances with equalities are subsumed under the same cases as before. In the cases I and II, we let $y_1 \to \infty$. The ratio of c.d.f.'s converges to 1 and ∞, respectively, and the exponential function increases to ∞, so that the expression diverges to ∞. This contradicts the assumed observational equivalence. In the cases III and IV, we let $y_1 \to -\infty$. In case IV the ratio of c.d.f.'s converges to 1 and the exponential function to 0. Again, this it at odds with observational equivalence. Case III is more involved. In the instance that the weak inequality in case III is an equality the denominator is a positive constant and the limit is 0. Again, this contradicts observational equivalence. Next, we consider the case that the inequality is strict. Because the c.d.f. in the numerator converges to 1, we can restrict attention to the exponential function in the numerator and the c.d.f. in the denominator that both converge to 0. Using l' Hospital's rule we find that the limit is ∞, again a contradiction.

We conclude that

$$\mu_1 = \tilde{\mu}_1 \tag{15}$$

Upon substitution of (12) and (15) in (8) and (9), we obtain the following system of three equations

$$\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} = \frac{\mu_1 - \tilde{\mu}_2}{\sqrt{\sigma_1^2 + \tilde{\sigma}_2^2 - 2\tilde{\rho}\sigma_1\tilde{\sigma}_2}} \tag{16}$$

$$\frac{1 - \rho\frac{\sigma_2}{\sigma_1}}{\sigma_2\sqrt{1-\rho^2}} = \frac{1 - \tilde{\rho}\frac{\tilde{\sigma}_2}{\sigma_1}}{\tilde{\sigma}_2\sqrt{1-\tilde{\rho}^2}} \tag{17}$$

$$\frac{\rho\frac{\sigma_2}{\sigma_1}\mu_1 - \mu_2}{\sigma_2\sqrt{1-\rho^2}} = \frac{\tilde{\rho}\frac{\tilde{\sigma}_2}{\sigma_1}\mu_1 - \tilde{\mu}_2}{\tilde{\sigma}_2\sqrt{1-\tilde{\rho}^2}} \tag{18}$$

Consider $\tilde{\mu}_2, \tilde{\sigma}_2, \tilde{\rho}$ as the unknowns of this system. We must show that the unique solution is $\tilde{\mu}_2 = \mu_2, \tilde{\sigma}_2 = \sigma_2, \tilde{\rho} = \rho$. By solving equation (17) we can express $\tilde{\sigma}_2$ as a function of $\tilde{\rho}$, and by solving equation (16) we can express $\tilde{\mu}_2$ as a function of $\tilde{\rho}, \tilde{\sigma}_2$. Substitution of these expressions in equation (18) gives an equation in $\tilde{\rho}$. After some calculations we obtain

$$\frac{\tilde{\rho}}{\sqrt{1-\tilde{\rho}^2}} = \frac{a + \sigma_1^2 b^2 - bc - \mu_1 b^2}{\frac{c}{\sigma_1} + \frac{\mu_1}{\sigma_1}b} \tag{19}$$

with a, b, c the left-hand sides of equations (16), (17), (18). Because (19) has a unique solution, the system has a unique solution $\tilde{\mu}_2 = \mu_2, \tilde{\sigma}_2 = \sigma_2, \tilde{\rho} = \rho$. This completes the proof.

4 Conclusion

Explanatory variables x do not play any role in the proof. By repeating the argument for every subpopulation defined by a vector of explanatory variables x, we see that $\mu_1(x), \mu_2(x)$, and $\sigma_1(x), \sigma_2(x), \rho(x)$ for that matter, are identified for all x.

The identification of the pair μ, Σ poses a number of questions. First, do distinct pairs have joint densities of observables that differ with respect to a suitable metric, *e.g.* the Kullback-Leibler metric associated with Maximum Likelihood estimation? Second, are the parameters identified in the sense of Rothenberg (1971), *i.e.* is the information matrix of constant rank in a neighborhood of the population parameters and of full rank? Third, given the nature of the identification proof that is based on the behavior of the likelihood ratio for extreme observations and relies on the tail behavior of the normal density, can we distinguish between distinct pairs in practice? The answer to these questions is beyond the scope of the current note.

5 References

MADDALA, G.S. (1983), *Limited dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge.

NELSON, F.D. (1977), Censored regression models with unobserved, stochastic censoring thresholds. *Journal of Econometrics*, **6**, p. 309-327.

ROTHENBERG, T.J. (1971), Identification in parametric models, *Econometrica*,**39**, p. 577-591.