REVERSE REGRESSIONS FOR LATENT VARIABLE MODELS

by

David Levine

University of California, Los Angeles

UCLA Working Paper #319

January 1984

REVERSE REGRESSIONS FOR LATENT VARIABLE MODELS*

by

David Levine**

University of California, Los Angeles

ABSTRACT

Under joint normality of all regressors, the errors-in-variables bounds for linear regression may be extended to probit and related models of censorship and truncation.

---

*I am grateful to Tim Erickson and Ed Leamer for stimulating my interest in this problem, and for helpful discussions.

**UCLA Department of Economics, Los Angeles, CA 90024

Klepper and Leamer (1984) show how to find bounds and other diagnostics in the normal errors-in-variables model. In this paper I show how to extend their results to a normal model with a latent dependent variable, such as probit or normal censorship or truncation. Klepper and Leamer's results are based on the idea of reverse regressions, that is, regressing each of the explanatory variables measured with error on the remaining explanatory variables and the endogenous variable. When the endogenous variable is a latent variable so that it is observed only via a proxy the direct analogue of a reverse regression is not useful. The problem is that the covariance between the latent variable and explanatory variables cannot be estimated by the sample moments, since the latent variable is unobservable. However, the covariance between the latent and explanatory variables can be estimated from knowledge of the regression coefficients of the latent on the explanatory variables, and these coefficients can be estimated by maximum likelihood.

Adopting the notation and assumptions of Leamer and Klepper y_t is normal with mean $\beta_o + \beta'\chi_t$ and variance σ^2 where χ_t is a (kxl) vector of unobservables. The unobservables χ_t are measured by the vector x_t which, conditional on χ_t is normal with mean χ_t and diagonal covariance D. The unobservables χ_t are normal with mean $\bar{\chi}$ and covariance Σ. The unconditional variance of y_t is s_y^2, the covariance between y_t and both x_t and χ_t is the k vector r and the covariance of x_t is N. Thus the covariance of the vector (y_t, x_t') is the matrix

$$V(y_t, x_t') = \begin{vmatrix} s_y^2 & r' \\ r & N \end{vmatrix}$$

from which the reverse regression and other diagnostics can be computed.

Klepper and Leamer assume y_t is directly observable so that likelihood estimates of s_y^2 and r are the sample moments. Suppose instead that y_t is not observed, but instead only $z(y_t)$ is observable. For example in probit

$$z(y_t) = \begin{matrix} 1 & y_t > 0 \\ 0 & y_t \leq 0 \end{matrix}$$

and censorship and truncation models can be similarly represented. Obviously s_y^2 and r can not be consistently estimated by sample moments using z_t in place of y_t.

Let us now formally state the Klepper and Leamer problem in the latent variable context. The problem is to maximize the joint likelihood function derived from the joint density $f(z_t, x_t | \beta, r, N, s_y^2, D)$ with respect to β, r, N, s_y^2 and D. Since the model is not identified the solution will be a set rather than a point.

Joint normality, however, implies that conditional on x_t the latent variable y_t is normally distributed with mean $b_o + b'x_t$ and variance σ_b^2. Joint normality also implies the identities

$$\begin{aligned} r &= Nb \\ \beta &= (N-D)^{-1} r \\ s_y^2 &= \sigma_b^2 + b'Nb \end{aligned} \tag{1}$$

Thus the invariance property of maximum likelihood imlies that we can maximize the joint likelihood then compute β from (1). Since given b, σ_b^2 and N the joint normal distribution of y_t and x_t is independent of D $f(z_t, x_t | b, \sigma_b^2, N, D) = f(z_t, x_t | b, \sigma_b^2, N)$ and the maximum likelihood estimates of b, σ_b^2 and N are independent of D. On the other hand D is restricted to be positive semi-definite and

$$V(y_t, \chi'_t) = \begin{vmatrix} s_y^2 & r' \\ r & N-D \end{vmatrix}$$

must also be positive semi-definite. These two restrictions characterize the feasible D's and thus by (1) the feasible β's.

Klepper and Leamer show how to find the feasible set of β and other diagnostics making use of these restrictions. Their method is based on analyzing the inverse of the estimated matrix $V(y_t, x_t')$ which using (1) may be written as

$$(2) \qquad V(y_t, x_t') = \begin{vmatrix} \sigma_b^2 + b'Nb & b'N \\ Nb & N \end{vmatrix}$$

where σ_b^2, b and N are maximum likelihood estimates.

Analyzing the likelihood function for b, σ_b^2 and N we see that $f(z_t, x_t | b, \sigma_b^2, N) = f(z_t | x_t, b, \sigma_b^2)\, f(x_t | N)$ since given x_t, b and σ_b^2 the distribution of y_t doesn't depend on N, and given N the distribution of x_t doesn't depend on b or σ_b^2. Thus each factor may be maximized separately. Since x_t is observed and joint normal the maximum likelihood estimate of N is just the sample moment; maximizing the likelihood from $f(z_t | x_t, b, \sigma_b^2)$ simply involves doing probit, censorship or truncation as appropriate, since $f(y_t | x_t, b, \sigma_b^2)$ is normal with mean $b_o + b'x_t$ and variance σ_b^2 by the joint normality hypothesis. Note that in probit $\sigma_b^2 = 1$ is usually imposed as a identifying restriction.

The upshot is that b and σ_b^2 should be estimated ignoring measurement error and using x_t in place of χ_t. The matrix N is estimated by the empirical moment matrix for x_t. These estimates are then combined in (2) to which the Klepper and Leamer methods apply.

What happens in the non-normal case? In the OLS case the consistency of the sample moments implies that the Klepper and Leamer methods give consistent bounds on the feasible parameter values. The latent variable case has two difficulties. The first is that the family of distributions for the latent variable depends on whether we condition on x_t or χ_t. For example if logit is appropriate when χ_t is conditioned on, it will generally not provide the correct family of distributions when x_t is conditioned on. In the case of discrete choice this problem is probably insurmountable. In the case of censorship or truncation the robust methods of Powell (1981) would be consistent for a location parameter linear in either x_t or χ_t regardless of how the latent variable is distributed (provided it is symmetric in the case of censorship). A more serious problem is that (1) requires the location parameter for the latent variable to be linear in x_t. Unless χ_t and x_t are joint normal, linearity of the conditional expectation in χ_t does not guarantee linearity in x_t. An open question is whether there are plausible non-normal joint distributions of χ_t and x_t for which a location parameter for the latent variable is linear in both sets of variables. If so, then (1) combined with Powell's methods yield consistent bounds; otherwise the Klepper and Leamer method breaks down.

Reference

Klepper, S. and E. Leamer, Consistent Sets of Estimates for Regressions with Errors in All Variables, Econometrica, 52, #1 (January 1984), 163-83.

Powell, James L., Least Absolute Deviations Estimation for Censored and Truncated Regression Models, IMSSS Technical Report #356, December 1981.