Simple Estimators For Hard Problems: Endogeneity and Dependence in Binary Choice Related Models

Arthur Lewbel
Boston College

SIMPLE ESTIMATORS:

1. Few or no smoothers.
2. Steps resembling common estimators.
3. Few or no numerical searches.

Why simple instead of general?

1. Smoothers affect finite sample performance.
2. Simpler variance estimation, and bootstrapping feasible.
3. Fewer coding errors, ease of use.
4. Avoid numerical failures.
5. Provides good starting values for hard.

TWO STEP ESTIMATORS

Assume first stage $\widehat{\gamma}$ from

$E[g_1(Z, \gamma)] = 0,$

second stage estimates $\widehat{\beta}$ from

$E[g_2(Z, \beta, \gamma) \mid \gamma = \widehat{\gamma}] = 0.$

Inefficient, wrong standard errors.

Example "Heckit." First stage, get $\widehat{\gamma}$ from max selection probit likelihood $L(Z, \gamma)$:

$$E\left[\partial L(Z, \gamma)/\partial\gamma\right] = 0$$

Second stage, get $\widehat{\beta}$ regressing Y on X and on inverse mills ratio $m(X'\widehat{\gamma})$.

$$Y = X'\beta + m(X'\widehat{\gamma})\lambda + e$$

Simple solution: Efficient with correct standard errors is joint GMM:

Proposition 1: Can apply efficient GMM to

$$E\begin{pmatrix} g_1(Z,\gamma) \\ g_2(Z,\beta,\gamma) \end{pmatrix} = 0, \text{ write as: } E\left[g(Z,\beta,\gamma)\right] = 0$$

$$\widehat{\gamma}, \widehat{\beta} = \arg\min_{\gamma,\beta} \sum_{i=1}^{n} g(Z_i,\beta,\gamma)'\Omega_n \sum_{i=1}^{n} g(Z_i,\beta,\gamma).$$

Heckit example:

$$E\begin{bmatrix} \partial L(Z,\gamma)/\partial\gamma \\ \left[Y - X'\beta - m(X'\gamma)\lambda\right]\begin{pmatrix} X \\ m(X'\gamma) \end{pmatrix} \end{bmatrix} = 0$$

If score $\partial L(Z,\gamma)/\partial\gamma$ hard to calculate use computer numerical or analytic derivative.

Start with two step estimates, obtain asymptotic efficiency with one iteration of the efficient GMM estimator.

CONVENIENT REPRESENTATION OF BINOMIAL RESPONSE MODELS

$$D = I(\widetilde{X}'\widetilde{\beta} + \widetilde{\varepsilon} \geq 0), \qquad var(\widetilde{\varepsilon}) = 1$$

$$D = I(X'\beta + V + \varepsilon \geq 0), \qquad var(\varepsilon) = \sigma_{\varepsilon}^{2}$$

Scaled Probit log likelihood is $\sum_{i=1}^{n} L(D_i, X_i, V_i, \beta, \sigma_{\varepsilon})$ where $L =$

$$D \ln \Phi\left(\frac{X'\beta + V}{\sigma_{\varepsilon}}\right) + (1-D) \ln\left[1 - \Phi\left(\frac{X'\beta + V}{\sigma_{\varepsilon}}\right)\right]$$

Choice probabilities are given by $\Pr(D = 1) = 1 - F_{\varepsilon}[-(X'\beta + V)]$ instead of the equivalent $1 - F_{\widetilde{\varepsilon}}[-(\widetilde{X}'\widetilde{\beta})]$.

For normal errors $F_{\widetilde{\varepsilon}}(\cdot) = \Phi(\cdot)$ and $F_{\varepsilon}(\cdot) = \Phi(\cdot/\sigma_{\varepsilon})$.

Alternative scaled representation preferred for many semiparametric estimators because:

β often easier to estimate, or faster converging than $\widetilde{\beta}$.

For some semiparametric models, β is identified while $\widetilde{\beta}$ is not.

β can provide a natural scaling. If D is purchase decision, V negative logged price, then $X'\beta$ is log reservation price (willingness to pay).

$$\begin{aligned} D &= I(X'\beta + V + \varepsilon \geq 0) \\ &= I(-V \leq X'\beta + \varepsilon) \end{aligned}$$

Similarly, V can be a bid in survey design.

The estimators in this paper will all use this scaled representation.

BINOMIAL RESPONSE WITH ENDOGENOUS REGRESSORS

$Y =$ vector of endogenous or mismeasured regressors
$W =$ vector of exogenous covariates. Let

$$\begin{aligned} X &= (Y', X_2')' \\ W &= (Z_1', X_2', V)' \end{aligned}$$

The model is

$$\begin{aligned} D &= I(X'\beta + V + \varepsilon \geq 0) \\ &= I(Y'\beta_1 + X_2'\beta_2 + V + \varepsilon \geq 0) \end{aligned}$$

Latent error ε uncorrelated with exogenous W, correlated with endogenous Y, and has unknown heteroskedasticity.
Data are n observations of D_i, Y_i, W_i.

Why not ML? Can be numerically hard, Must completely specify joint distribution of $\varepsilon, Y \mid W$, sensitive to hard to estimate parameters like $cov(\varepsilon, Y \mid W)$.

Common inconsistent coefficient estimators: probit or logit, ignoring the endogeneity of Y, or two stage least squares linear probability model, regressing of D on X, V using instruments W.

Goal: numerically simple estimators, consistent under relatively general conditions.

TYPES OF ENDOGENEITY

With $X = (Y', X_2')$, will consider

$$\begin{aligned} D &= I(X'\beta + V + \varepsilon \geq 0) \\ Y &= g(W, U) \end{aligned}$$

for some function g and vector of unobservables U.
Endogeneity arises from dependence between U and ε, due to measurement error in Y or simultaneity in determining Y and D.
Higher moments of ε can depend on W, U.

CONTROL FUNCTION ESTIMATORS

$$D = I(X'\beta + V + \varepsilon \geq 0)$$
$$Y = h(W) + U$$

with restrictions on condition distribution ε given U, W, and entails explicitly Y model and U. Requires continuous Y.

INSTRUMENTAL VARIABLES WITH A VERY EXOGENOUS V

$$D = I(X'\beta + V + \varepsilon \geq 0)$$
$$E(ZY) \neq 0, \quad E(Z\varepsilon) = 0$$

With strong restrictions on V, does not estimate Y model or U, permits continuous discrete, truncated, or otherwise limited Y, and heteroskedastic ε.

CONTROL FUNCTION ESTIMATORS

$$
\begin{aligned}
D &= I(X'\beta + V + \varepsilon \geq 0) \\
Y &= W'b + U \qquad E(WU) = 0 \\
\varepsilon &= U'\lambda + \eta, \quad \eta \perp U, W, \qquad \eta \sim N(0, \sigma_\eta^2)
\end{aligned}
$$

Substituting out ε in the D equation yields

$$D = I(X'\beta + V + U'\gamma + \eta \geq 0)$$

where $\gamma = \lambda\beta_1$. Is a probit with regressors X, V, and U. Result is simple estimator:

ESTIMATOR A

1. For each observation i, construct data $\widehat{U}_i = Y_i - W_i'\widehat{b}$ and as the fitted values of an ordinary least squares regression of Y on W (or seemingly unrelated regression if Y is a vector).

2. Let $\widehat{\beta}$ be the estimated coefficients of X in an ordinary scaled probit regression of D on X, V and $\widehat{U}$.

Scaled probit normalizes the coefficient of V to be one. This conveniently keeps β unchanged whether error is η or ε. Up to scaling, Estimator A is Rivers and Voung (1988), Blundell and Smith (1986), Nelson and Olsen (1978), like Heckman (1978).

Define $\widetilde{D} = X'\beta + V + U'\gamma$ and

$$R = D\frac{\phi\left(\widetilde{D}/\sigma_{\eta}\right)}{\Phi\left(\widetilde{D}/\sigma_{\eta}\right)} + (1-D)\frac{-\phi\left(\widetilde{D}/\sigma_{\eta}\right)}{1-\Phi\left(\widetilde{D}/\sigma_{\eta}\right)}$$

ESTIMATOR B:

Use ordinary GMM to estimate the parameters $b, \beta, \gamma, \sigma_{\eta}$ based on the moments

$$\begin{aligned} E\left[W(Y - W'b)\right] &= 0 \\ E\left[R(D, X, V, Y - W'b, \beta, \gamma, \sigma_{\eta})X\right] &= 0 \\ E\left[R(D, X, V, Y - W'b, \beta, \gamma, \sigma_{\eta})\left(Y - W'b\right)\right] &= 0 \\ E\left[R(D, X, V, Y - W'b, \beta, \gamma, \sigma_{\eta})V\right] &= 0 \end{aligned}$$

Amemiya's GLS, Newey (1987).

Control function generalizes. Blundell and Powell (2003):

$$
\begin{aligned}
D &= I(X'\beta + V + \varepsilon \geq 0) \\
Y &= h(W) + U, \quad E(U \mid W) = 0 \\
\varepsilon &\mid X, U \sim \varepsilon \mid U
\end{aligned}
$$

Estimation not simple, requires high dimensional nonparametric first step and semi-parametric multiple index estimation.

Control function mostly not applicable if endogenous Y are discrete, censored, truncated, or otherwise limited.

Requires estimating a model for each element of Y

Unlike IV methods if control function assumptions hold for a set of instruments Z, then they will not hold in general using some smaller subset of instruments.
Can't omit questionable instruments.

INSTRUMENTAL VARIABLES WITH A VERY EXOGENOUS V

Advantages of linear 2SLS:
Numerical simplicity,
No explicit modeling of endogenous regressors Y,
Same if Y are continuous, discrete, limited, truncated, etc.,.
No explicit modeling of higher moments of latent error ε,
Only needs instruments correlated with Y, uncorrelated with latent error ε.

Here provide an estimator that preserves these features of 2SLS, without imposing a linear probability model. Extends ideas in Lewbel (2000) and Lewbel, Linton, and McFadden (2003) to be very simple to implement.

WHAT IS A VERY EXOGENOUS V?

Let $S = (X, Z)$ be all regressors and instruments (including Y) except V.

V is very exogenous if:

1. $h(X) + V + \varepsilon$ for some h is a latent variable in a model.

2. $V = g(\nu, S)$ for some g,
   $\nu \perp S, \varepsilon$,
   ν is continuously distributed,
   $\partial g(\nu, S)/\partial \nu > 0$.

3. $supp[h(X) + \varepsilon] \subseteq supp(-V \mid S)$.

Example:

$$D = I(X'\beta + V + \varepsilon \geq 0)$$
$$V = S'b + \nu, \quad \nu \perp S, \varepsilon, \quad \nu \sim N(0, \sigma^2)$$

Sufficient is an independent normal additive error in a model for V.

1. $h(X) + V + \varepsilon$ for some h is a latent variable in a model.

2. $V = g(\nu, S)$, $\nu \perp S, \varepsilon$, implies $\varepsilon \mid V, S \sim \varepsilon \mid S$, (ordinary exogeneity).

3. $supp(X'\beta + \varepsilon) \subseteq supp(-V \mid S)$. Large support for identification (Manski 1975, 1985, Horowitz 1992, Han 1987, Cavanagh and Sherman 1998). Can replace large support with ε tail restrictions (Magnac and Maurin 2003).

Examples:
ε unobserved ability and V age.
ε preferences and V supply side.
V experimental design.

Empirical Applications: Anton, Sainz, and Rodriguez-Poo (2001) Maurin (2002), Cogneau and Maurin (2002), Lewbel (2003), Lewbel, Linton, and McFadden (2003).

BINOMIAL RESPONSE WITH VERY EXOGENOUS V

Let $Z = (Z_1', X_2')'$ = all exogenous covariates except for V.

Let $S = (Y', Z')'$ = all the available covariates except for V.

$X = (Y', X_2')'$ = all regressors in the D equation except for V.

The entire model is

$$\begin{aligned} D &= I(X'\beta + V + \varepsilon \geq 0), \quad E(Z\varepsilon) = 0 \\ V &= S'b + \nu, \quad \nu \perp S, \varepsilon, \quad \nu \sim N(0, \sigma_\nu^2) \end{aligned}$$
$$rank[E(ZX')] = rank[E(XX')]$$

Instead of modeling and restricting Y, here model we model and restrict V. Other than V and D models, nothing is known or estimated regarding the DGP of X and hence of Y, except $E(Z\varepsilon) = 0$, $E(ZX') \neq 0$.

VERY EXOGENOUS
ESTIMATOR DERIVATION

$$D^* = X'\beta + \varepsilon, \quad \nu = V - S'b, \quad T = \frac{D - I(V \geq 0)}{f(\nu)}$$

so $D = I(D^* + V \geq 0)$. Then, by the definition of conditional expectation,

$$E(T \mid S, \varepsilon) =$$

$$\int_{-\infty}^{\infty} \left(\frac{I(D^* + V \geq 0) - I(V \geq 0)}{f(V - S'b)} \right) f_V(V \mid S, \varepsilon) dV$$

$$= \int_{-\infty}^{\infty} \left[I(D^* + V \geq 0) - I(V \geq 0) \right] d\nu$$

$=$ probability a uniform V lies between $-D^*$ and 0

$$= \int_{-D^*}^{0} 1 d\nu = D^*$$

Therefore $T = X'\beta + \varepsilon + e$ and

$$E(ZT) = E[Z(X'\beta + \varepsilon)] = E(ZX')\beta$$

$$
\begin{aligned}
D &= I(X'\beta + V + \varepsilon \geq 0), \quad E(Z\varepsilon) = 0 \\
V &= S'b + \nu, \quad \nu \perp S, \varepsilon, \quad \nu \sim N(0, \sigma_\nu^2)
\end{aligned}
$$

$$
rank[E(ZX')] = rank[E(XX')]
$$

ESTIMATOR C

1. For each i, construct $\widehat{\nu}_i = V_i - S_i'\widehat{b}$ as residuals of an OLS regression of V on S.
2. For each i, let $f(\widehat{\nu}_i,\widehat{\sigma}_\nu^2)$ be the pdf of $\widehat{\nu}_i$,

$$
f(\widehat{\nu}_i, \widehat{\sigma}_\nu^2) = \frac{1}{\sqrt{2\pi\widehat{\sigma}_\nu^2}} \exp\left(\frac{-\widehat{\nu}_i^2}{2\widehat{\sigma}_\nu^2}\right)
$$

where $\widehat{\sigma}_\nu^2$ is the sample mean of $\widehat{\nu}^2$.
3. For each observation i, construct

$$
\widehat{T}_i = \frac{D_i - I(V_i \geq 0)}{f(\widehat{\nu}_i, \widehat{\sigma}_\nu^2)}
$$

4. Let $\widehat{\beta}$ be estimated coefficients of an ordinary linear 2SLS regression of $\widehat{T}$ on X, using instruments Z.

Estimator C is a numerically trivial, no steps more complicated than linear regressions.

This estimator assumes $V = S'b + \nu$, $\nu \sim N(0, \sigma_\nu^2)$ while Lewbel (2000) uses a nonparametric conditional density estimator for V. The latter is not strictly more general than the Estimator C, since $V = S'b + \nu$ allows V to depend on Y.

It provides root n CAN estimates of the coefficients β, though for $var(T)$ finite need $X'\beta + \varepsilon$ bounded or very thin tailed, or based on Magnac and Maurin (2003), a symmetry or moment restriction on the tails of ε. As a result, estimator is sensitive to the choice of V, more accurate when var(ν) is large relative to var($X'\beta + \varepsilon$).

May also be sensitive to outliers, since density in T means large values of ν_i generate extremely large values of T_i. May want to use robust moment estimators, for example, in the 2SLS step, discarding all observations having very small $f(\widehat{\nu}_i, \widehat{\sigma}_\nu^2)$.

More efficient estimates with correct standard errors can be obtained by combining the above steps.

ESTIMATOR D

Use GMM to estimate the parameters β, b, σ_ν^2 based on the moment conditions

$$E[S(V - S'b)] = 0$$
$$E[\sigma_\nu^2 - (V - S'b)^2] = 0$$
$$E\left[Z\left([D - I(V \geq 0)]\,\sigma_\nu e^{\left(\frac{(V-S'b)^2}{2\sigma_\nu^2}\right)} - X'\beta\right)\right] = 0$$

MORE THAN ONE VERY EXOGENOUS REGRESSOR

If lucky enough to have more than one very exogenous regressor in the model, then efficiently use the information in both by taking all the moments of estimator D based on each such regressor, and doing a large GMM on the entire set, normalizing the coefficient of one of them to be one.

MORE GENERAL V MODEL

$$V = S'b + \exp(S'c)\nu, \quad \nu \perp S, \varepsilon, \quad \nu \sim N(0,1)$$

add moment

$$E[S\left(\exp(2S'c) - \exp(S'c)(V - S'b)^2\right)] = 0$$

Can easily test simple vs general V models.

UNKNOWN ν DISTRIBUTION

Drop the assumption that ν is normal. Could use a one dimensional kernel estimator of $f(\nu_i)$, but simpler is to sort ν_i lowest to highest. Let ν_i^- and ν_i^+ be sorted before and after ν_i values. Then

$$f(\nu_i) = \frac{\partial F(\nu_i)}{\partial \nu} \approx \frac{F(\nu_i^+) - F(\nu_i^-)}{\nu_i^+ - \nu_i^-} \approx \frac{2/n}{\nu_i^+ - \nu_i^-}$$

where the last step replaces the true distribution function F with the empirical distribution function. This suggests the estimator

$$\frac{1}{\widehat{f}(\nu_i)} = \frac{(\nu_i^+ - \nu_i^-)n}{2}, \qquad \widehat{T}_i = \frac{[D_i - I(V_i \geq 0)]}{\widehat{f}(\nu_i)}$$

See Lewbel and Schennach (2003), Lewbel, Linton, and McFadden (2003).

MORE GENERALIZATIONS

$$D = I(X'\beta + V + \varepsilon \geq 0), \quad E(Z\varepsilon) = 0$$
$$V = g(S) + \nu, \quad \nu \perp S, \varepsilon$$

for unknown function g. Use estimator E, except in step 1, $\widehat{\nu}_i$ is the residuals from a nonparametric regression of V on S.

A further generalization is

$$D = I(X'\beta + V + \varepsilon \geq 0), \quad E(Z\varepsilon) = 0$$
$$V \mid S, \varepsilon \sim V \mid S$$

Now let $\widehat{f_V}(V \mid S)$ be a nonparametric estimator of the conditional density of V given S, construct $\widehat{T}_i$ by

$$\widehat{T}_i = \frac{D_i - I(V_i \geq 0)}{\widehat{f_V}(V_i \mid S_i)}$$

and let $\widehat{\beta}$ be the estimated coefficients of a linear 2SLS regression of $\widehat{T}$ on X, using instruments Z. Still no numerical searches required.

DISCRETE V

Example: D is schooling outcome,
$V = 1$ for older students, 0 younger,
$\widetilde{V}$ is normalized unknown exact age.

$$D = I(X'\beta + \widetilde{V} + \varepsilon \geq 0), \quad E(Z\varepsilon) = 0$$
$$V = I(\widetilde{V} \geq 0), \quad \widetilde{V} \perp S, \varepsilon, \quad \widetilde{V} \sim U[-K, 1-K]$$

D depends on unobserved uniform, very exogenous regressor $\widetilde{V}$, while observed V is binary, discrete. Then

$$T = \frac{D - I(\widetilde{V} \geq 0)}{f_{\widetilde{V}}(\widetilde{V})} = D - V$$

$\widehat{\beta}$ is an ordinary linear 2SLS regression of $D - V$ on X, using instruments Z.

If know birth month, assume births uniform within months, use $(D - V)/(12F)$, where F is fraction of births in the month. Similar for distance ranges or income brackets data. See also Magnac and Maurin (2004).

CHOICE PROBABILITIES

With endogenous regressors prob($D = 1$) conditional on covariates is

$$E(D \mid Y, W) = 1 - F_\varepsilon[-(X'\beta + V) \mid Y, W]$$

Requires estimating high dimensional conditional distribution $F_\varepsilon(\varepsilon \mid Y, W)$.

An alternative is the average structural function (Blundell and Powell 2000) defined as

$$1 - F_\varepsilon[-(X'\beta + V)]$$

Evaluates at marginal $F_\varepsilon(\varepsilon \mid Y, W)$

Define INDEX CHOICE PROBABILITY

$$E(D \mid X'\beta + V) = 1 - F_\varepsilon[-(X'\beta + V) \mid X'\beta + V]$$

This is a middle ground between the previous two choice probability measures (conditioning ε on all covariates versus on none).

When $\varepsilon \perp Y, W$ all three choice probability measures are the same.

Index choice probability equals average structural function when $\varepsilon \perp X'\beta + V$.

Index choice probability equals $E(D \mid Y, W)$ equals when $\varepsilon \mid Y, W = \varepsilon \mid X'\beta + V$.

The index choice probability can be readily estimated by a one dimensional nonparametric regression of D on $X'\beta + V$.

ORDERED CHOICE

$$k^* = \sum_{k=1}^{K} kI\left[\alpha_{k-1} \leq -\left(X'\beta + V + \varepsilon\right) \leq \alpha_k\right]$$

Equals set of binary choices

$$D_k = I(\alpha_k + X'\beta + V + \varepsilon \geq 0)$$

Each $D_k = 1$ if choose $k^* \leq k$, else zero.

WLOG each D_k constant is threshold α_k.

Apply GMM to the collection of Estimator B (control function) or D (very exogenous V) moments corresponding to every D_k model to estimate the parameters β and $\alpha_1, ..., \alpha_K$.

Lewbel (2000) describes an example of this estimator for the very exogenous regressor model with a general estimator for the conditional density of V.

SELECTION MODELS

Without endogenous regressors, use GMM form of Heckit estimator (Proposition 1) for efficiency correct standard errors.

WITH ENDOGENOUS REGRESSORS

Same variable definitions, but also outcome P only observed with $D = 1$, else $P = 0$.

$$\begin{aligned} P &= (X'\beta + V\gamma + \varepsilon)D, \quad E(Z\varepsilon) = 0 \\ D &= I(a_0 \leq M(S, e) \pm V \leq a_1) \\ V &= S'b + \nu, \quad \nu \perp S, \varepsilon, e, \quad \nu \sim N(0, \sigma_\nu^2) \end{aligned}$$

Function M, scalars a_0 and a_1, conditional distribution $e, \varepsilon \mid S$ all unknown. If a_0 and a_1 are finite then

$$E\left[Z\frac{D}{f(v)}(P - X'\beta + V\gamma)\right] = 0$$

For infinite a_0 or a_1, do asymptotic trimming or have tiny bias of order max $1/|V|$.

Example: P wage, D employment, $a_0 = 0$, $a_1 = \infty$, M linear, X training, education, demographics, experience; endogenous Y spouse or parent income, V age or access (cost, distance) to schooling, ε, e are correlated, contain ability, drive.

Example: Lewbel (2003) provides an empirical application in which P is factory investment rate, D indicates nonzero investment which occurs when returns to investment are positive, V is plant size, and the endogenous regressor Y is the factory profit rate, which proxies for Tobin's Q.

Example: Ordered selection models. Both a_0 and a_1 are finite. Latent $M(S, e) \pm V$ is desired schooling, D is high shool, not college. Latent $< a_0$ is drop out, $> a_1$ is college grad, P is earnings for high school grads without college degrees.

$$
\begin{aligned}
P &= (X'\beta + V\gamma + \varepsilon)D, \quad E(Z\varepsilon) = 0 \\
D &= I(a_0 \leq M(S, e) \pm V \leq a_1) \\
V &= S'b + \nu, \quad \nu \perp S, \varepsilon, e, \quad \nu \sim N(0, \sigma_\nu^2)
\end{aligned}
$$

ESTIMATOR E

1. For each i, construct $\widehat{\nu}_i = V_i - S_i'\widehat{b}$ as residuals of an OLS regression of V on S.

2. For each i, let $f(\widehat{\nu}_i, \widehat{\sigma}_\nu^2)$ be the pdf of $\widehat{\nu}_i$,

$$
f(\widehat{\nu}_i, \widehat{\sigma}_\nu^2) = \frac{1}{\sqrt{2\pi\widehat{\sigma}_\nu^2}} \exp\left(\frac{-\widehat{\nu}_i^2}{2\widehat{\sigma}_\nu^2}\right)
$$

where $\widehat{\sigma}_\nu^2$ is the sample mean of $\widehat{\nu}^2$.

3. For each observation i, construct instruments $\widehat{Z}_i = Z_i D_i / f(\widehat{\nu}_i, \widehat{\sigma}_\nu^2)$.

4. Let $\widehat{\beta}$ and $\widehat{\gamma}$ the estimated coefficients of an ordinary linear two stage least squares regression of P on X and V, using instruments $\widehat{Z}$

$$
\begin{aligned}
P &= (X'\beta + V\gamma + \varepsilon)D, \quad E(Z\varepsilon) = 0 \\
D &= I(a_0 \leq M(S,e) \pm V \leq a_1) \\
V &= S'b + \nu, \quad \nu \perp S, \varepsilon, e, \quad \nu \sim N(0, \sigma_\nu^2)
\end{aligned}
$$

ESTIMATOR F

Use GMM to estimate the parameters β, γ , b, σ_ν^2 based on the moment conditions

$$
\begin{aligned}
E[S(V - S'b)] &= 0 \\
E[\sigma_\nu^2 - (V - S'b)^2] &= 0
\end{aligned}
$$

$$
E\left[ZD\left(P - X'\beta + V\gamma\right)\exp\left(\frac{(V - S'b)^2}{2\sigma_\nu^2}\right)\right] = 0
$$

These estimators do not require estimating models for selection D, endogenous regressors Y, or the joint distribution of errors in the models of D, Y, and outcome P. outcome, selection, and endogenous regressors is not specified or estimated.

PANEL BINARY CHOICE

Fixed T asymptotics, variant of Honore and Lewbel (2002).

$$
\begin{aligned}
D_{it} &= I(X_{it}'\beta + V_{it} + \alpha_i + \varepsilon_{it} \geq 0) \\
V_{it} &= S_{it}'b_t + \nu_{it}, \quad \nu_{it} \perp S_{it}, \alpha_i + \varepsilon_{it} \\
\nu_{it} &\sim N(0, \sigma_{t\nu}^2), \quad E[Z_{it}(\varepsilon_{it} - \varepsilon_{it-1})] = 0
\end{aligned}
$$

e.g., V_{it} = cost shocks or transitory income.

This model allows:
endogenous regressors
weakly exogenous regressors
lagged dependent variables, dynamics
autocorrelated errors, fixed effects
unknown error distribution

COROLLARY 1:

$$T_{it} = \frac{D_{it} - I(V_{it} \geq 0)}{f_t(\nu_{it})}$$

$$E\left[Z_{it}\left(T_{it} - T_{it-1} - (X_{it} - X_{it-1})'\beta\right)\right] = 0$$

$$
\begin{aligned}
D_{it} &= I(X_{it}'\beta + V_{it} + \alpha_i + \varepsilon_{it} \geq 0) \\
V_{it} &= S_{it}'b_t + \nu_{it}, \quad \nu_{it} \perp S_{it}, \alpha_i + \varepsilon_{it} \\
\nu_{it} &\sim N(0, \sigma_{t\nu}^2), \quad E[Z_{it}(\varepsilon_{it} - \varepsilon_{it-1}) = 0
\end{aligned}
$$

ESTIMATOR G

1. Let $\widehat{\upsilon}_{it} = V_{it} - S_{it}'\widehat{b_t}$, time t OLS.
2. Let $\widehat{\sigma}_t^2$ = time t mean of $\widehat{\nu}_{it}^2$,

$$
f(\widehat{\nu}_{it}) = \frac{1}{\sqrt{2\pi\widehat{\sigma}_t^2}} \exp\left(\frac{-\widehat{\nu}_{it}^2}{2\widehat{\sigma}_t^2}\right)
$$

3. For each it construct

$$
\widehat{T}_{it} = \frac{D_{it} - I(V_{it} \geq 0)}{f(\widehat{\nu}_{it}, \widehat{\sigma}_t^2)}
$$

4. Let $\widehat{\beta}_t$ be estimated coefficients of linear two stage least squares regression of $\widehat{T}_{it} - \widehat{T}_{it-1}$ on $X_{it} - X_{it-1}$, using instruments Z_{it}. Let $\widehat{\beta}$ be an average of $\widehat{\beta}_t$.

ESTIMATOR H

Use GMM to estimate the parameters β, b_1, ..., b_T, σ_1^2, ..., σ_T^2 based on the moment conditions:

$$E[S_t(V_t - S_t' b_t)] = 0$$

$$E[\sigma_t^2 - (V_t - S_t' b_t)^2] = 0$$

$$E\left[Z_t \left(\begin{array}{c} \frac{[D_t - I(V_t \geq 0)]\sigma_t}{\exp\left(-\frac{\left(V_t - S_t' b_t\right)^2}{2\sigma_t^2}\right)} - \frac{\left[D_{t-1} - I(V_{t-1} \geq 0)\right]\sigma_{t-1}}{\exp\left(-\frac{\left(V_{t-1} - S_{t-1}' b_{t-1}\right)^2}{2\sigma_{t-1}^2}\right)} \\ -(X_t - X_{t-1})\beta \end{array} \right) \right] = 0$$

for $t = 1, ..., T$, where V_t denotes a draw of the random variable V in time period t, and similarly for S_t, X_t, D_t, and Z_t.

Panels of sample selection similar.

LARGE T DYNAMIC BINARY CHOICE MODELS WITH ENDOGENOUS REGRESSORS

$$
\begin{aligned}
D_t &= I(X_t'\beta + V_t + \varepsilon_t \geq 0), \quad E(Z_t\varepsilon_t) = 0 \\
V_t &= S_t'b + \exp(S_t'c)\nu_t, \quad \nu_t \perp S_t, \varepsilon_t, \ \nu_t \sim N(0,1)
\end{aligned}
$$

Regressors X_t can include endogenous, mismeasured, and lagged dependent variables. Essentially an ARCH model for V_t. The S_t variables can include lags, may need ν_t iid.

Theorem 1 holds, so can apply estimators C' and D.'

CONCLUSIONS

1. Multistep estimators easily made efficient with right standard errors by writing as GMM. Modern GMM applies, e.g., GEL, weak instrument GMM, etc.,.

2. Simple estimators for many hard limited dependent variable problems exist.

3. When developing new estimators, it's worth noting how one might sacrifice a bit of generality if the result is a much simpler estimator.

1. Smoothers affect finite sample performance.
2. Simpler variance estimation, and bootstrapping feasible.
3. Fewer coding errors, ease of use.
4. Avoid numerical failures.
5. Provides good starting values for hard.