ECONOMETRICS II (ECO 2401S)

University of Toronto. Department of Economics. Spring 2010
Instructor: Victor Aguirregabiria

SOLUTION TO FINAL TEST (March 31, 2010)

QUESTION 1: Consider the following panel data logit model:

$$Y_{it} = 1\{\ X_{it}'\beta + \alpha_i - u_{it} > 0\ \}$$

for $i = 1, 2, ..., N$ and $t = 1, 2$, where N is large. u_{it} is independently and identically distributed over (i, t) with a logistic distribution such that:

$$\Pr(Y_{it} = 1 \mid X_{it}, \alpha_i) = \frac{\exp\{X_{it}'\beta + \alpha_i\}}{1 + \exp\{X_{it}'\beta + \alpha_i\}}$$

The regressors in the vector X are strictly exogenous with respect to the transitory shock u_{it}. However, the individual effects α_i can be correlated with the regressors.

Question 1.1 (20 points): Define $Y_i \equiv \{Y_{i1}, Y_{i2}\}$, $X_i \equiv \{X_{i1}, X_{i2}\}$ and $S_i \equiv Y_{i1} + Y_{i2}$. Obtain the expression of the conditional probability $\Pr(Y_i|X_i, \alpha_i, S_i)$ and show that it does not depend on α_i.

By the law of conditional probability, $\Pr(Y_i, S_i \mid X_i, \alpha_i) = \Pr(Y_i \mid S_i, X_i, \alpha_i)\ \Pr(S_i \mid X_i, \alpha_i)$. Therefore,

$$\Pr(Y_i \mid S_i, X_i, \alpha_i) = \frac{\Pr(Y_i, S_i \mid X_i, \alpha_i)}{\Pr(S_i \mid X_i, \alpha_i)} = \frac{\Pr(Y_i \mid X_i, \alpha_i)}{\Pr(S_i \mid X_i, \alpha_i)}$$

where the last equality comes from the fact that S_i is a deterministic function of $Y_i = (Y_{i1}, Y_{i2})$. With $T = 2$ the statistic S_i can take values in the set $\{0,\ 1,\ 2\}$. Therefore, we should obtain the probabilities $\Pr(Y_i \mid S_i = 0, X_i, \alpha_i)$, $\Pr(Y_i \mid S_i = 1, X_i, \alpha_i)$, and $\Pr(Y_i \mid S_i = 2, X_i, \alpha_i)$.

When $S_i = 0$, the only value of Y_i that is possible is $Y_i = (0, 0)$. Therefore, it is clear that

$$\Pr(Y_i \mid S_i = 0, X_i, \alpha_i) = \begin{cases} 1 & \text{if } \ Y_i = (0,0) \\ \\ 0 & \text{otherwise} \end{cases}$$

Similarly, when $S_i = 2$, the only value of Y_i that is possible is $Y_i = (1, 1)$. Thus,

$$\Pr(Y_i \mid S_i = 2, X_i, \alpha_i) = \begin{cases} 1 & \text{if } \ Y_i = (1,1) \\ \\ 0 & \text{otherwise} \end{cases}$$

We have shown that the probabilities $\Pr(Y_i \mid S_i = 0, X_i, \alpha_i)$ and $\Pr(Y_i \mid S_i = 2, X_i, \alpha_i)$ do not depend on α_i. Unfortunately, these probabilities do not depend on the parameters of interest β either. Therefore, they do not contain any information for the estimation of these parameters.

When $S_i = 1$, there are two possible values for Y_i, i.e., $Y_i = (1,0)$ and $Y_i = (0,1)$. Applying the general expression $\Pr(Y_i \mid S_i, X_i, \alpha_i) = \frac{\Pr(Y_i \mid X_i, \alpha_i)}{\Pr(S_i \mid X_i, \alpha_i)}$, derived above, we have that:

$$\begin{aligned}
\Pr(Y_i = (1,0) \mid S_i = 1, X_i, \alpha_i) &= \frac{\Pr(Y_i = (1,0) \mid X_i, \alpha_i)}{\Pr(S_i = 1 \mid X_i, \alpha_i)} \\
&= \frac{\Pr(Y_i = (1,0) \mid X_i, \alpha_i)}{\Pr(Y_i = (1,0) \mid X_i, \alpha_i) + \Pr(Y_i = (0,1) \mid X_i, \alpha_i)}
\end{aligned}$$

and

$$\begin{aligned}
\Pr(Y_i = (0,1) \mid S_i = 1, X_i, \alpha_i) &= \frac{\Pr(Y_i = (0,1) \mid X_i, \alpha_i)}{\Pr(S_i = 1 \mid X_i, \alpha_i)} \\
&= \frac{\Pr(Y_i = (0,1) \mid X_i, \alpha_i)}{\Pr(Y_i = (1,0) \mid X_i, \alpha_i) + \Pr(Y_i = (0,1) \mid X_i, \alpha_i)}
\end{aligned}$$

Given our logit model and that u_{it} is *iid* over time, we have that:

$$\begin{aligned}
\Pr(Y_i = (1,0) \mid X_i, \alpha_i) &= \Pr(Y_{i1} = 1 \mid X_{i1}, \alpha_i) \;\; \Pr(Y_{i2} = 0 \mid X_{i2}, \alpha_i) \\
&= \frac{\exp\{X_{i1}'\beta + \alpha_i\}}{[1 + \exp\{X_{i1}'\beta + \alpha_i\}]\,[1 + \exp\{X_{i2}'\beta + \alpha_i\}]}
\end{aligned}$$

and

$$\begin{aligned}
\Pr(Y_i = (0,1) \mid X_i, \alpha_i) &= \Pr(Y_{i1} = 0 \mid X_{i1}, \alpha_i) \;\; \Pr(Y_{i2} = 1 \mid X_{i2}, \alpha_i) \\
&= \frac{\exp\{X_{i2}'\beta + \alpha_i\}}{[1 + \exp\{X_{i1}'\beta + \alpha_i\}]\,[1 + \exp\{X_{i2}'\beta + \alpha_i\}]}
\end{aligned}$$

Solving these expressions into the previous equations for $\Pr(Y_i = (1,0) \mid S_i = 1, X_i, \alpha_i)$ and $\Pr(Y_i = (1,0) \mid S_i = 1, X_i, \alpha_i)$, we get:

$$\begin{aligned}
\Pr(Y_i = (1,0) \mid S_i = 1, X_i, \alpha_i) &= \frac{\frac{\exp\{X_{i1}'\beta + \alpha_i\}}{[1 + \exp\{X_{i1}'\beta + \alpha_i\}][1 + \exp\{X_{i2}'\beta + \alpha_i\}]}}{\frac{\exp\{X_{i1}'\beta + \alpha_i\}}{[1 + \exp\{X_{i1}'\beta + \alpha_i\}][1 + \exp\{X_{i2}'\beta + \alpha_i\}]} + \frac{\exp\{X_{i2}'\beta + \alpha_i\}}{[1 + \exp\{X_{i1}'\beta + \alpha_i\}][1 + \exp\{X_{i2}'\beta + \alpha_i\}]}} \\
&= \frac{\exp\{X_{i1}'\beta + \alpha_i\}}{\exp\{X_{i1}'\beta + \alpha_i\} + \exp\{X_{i2}'\beta + \alpha_i\}} \\
&= \frac{\exp\{[X_{i1} - X_{i2}]'\beta\}}{1 + \exp\{[X_{i1} - X_{i2}]'\beta\}}
\end{aligned}$$

and

$$\Pr\left(Y_i=(0,1)\mid S_i=1,X_i,\alpha_i\right) = \frac{\frac{\exp\left\{X_{i2}'\beta+\alpha_i\right\}}{\left[1+\exp\left\{X_{i1}'\beta+\alpha_i\right\}\right]\left[1+\exp\left\{X_{i2}'\beta+\alpha_i\right\}\right]}}{\frac{\exp\left\{X_{i1}'\beta+\alpha_i\right\}}{\left[1+\exp\left\{X_{i1}'\beta+\alpha_i\right\}\right]\left[1+\exp\left\{X_{i2}'\beta+\alpha_i\right\}\right]}+\frac{\exp\left\{X_{i2}'\beta+\alpha_i\right\}}{\left[1+\exp\left\{X_{i1}'\beta+\alpha_i\right\}\right]\left[1+\exp\left\{X_{i2}'\beta+\alpha_i\right\}\right]}}$$

$$= \frac{\exp\left\{X_{i2}'\beta+\alpha_i\right\}}{\exp\left\{X_{i1}'\beta+\alpha_i\right\}+\exp\left\{X_{i2}'\beta+\alpha_i\right\}}$$

$$= \frac{1}{1+\exp\left\{[X_{i1}-X_{i2}]'\beta\right\}}$$

It is clear that these probabilities do not depend on α_i. Furthermore, they depend on the vector of parameters β. Therefore, these conditional probabilities can be used to estimate the parameters of interest β.

Question 1.2 (20 points): Based on the result in Question 1.1, obtain the expression of the conditional log-likelihood function $l(\beta)=\sum_{i=1}^{N}\log\Pr(Y_i|X_i,S_i;\beta)$. Is the conditional maximum likelihood estimator consistent in this model? Explain why.

The probabilities $\Pr\left(Y_i\mid S_i=0,X_i\right)$ and $\Pr\left(Y_i\mid S_i=2,X_i\right)$ do not depend on β, and so we can ignore them in the log-likelihood function. Then,

$$l(\beta) = \sum_{Y_i=(1,0)}\log\left(\frac{\exp\left\{[X_{i1}-X_{i2}]'\beta\right\}}{1+\exp\left\{[X_{i1}-X_{i2}]'\beta\right\}}\right)+\sum_{Y_i=(0,1)}\log\left(\frac{1}{1+\exp\left\{[X_{i1}-X_{i2}]'\beta\right\}}\right)$$

This is the log likelihood function of a standard logit model for the sub-sample of observations with $S_i=1$ where the dependent variable is $1\{Y_i=(1,0)\}$ and the vector of explanatory variables is $X_{i1}-X_{i2}$. This function is globally concave in β.

The Conditional Maximum Likelihood estimator (CMLE) is the value of β that maximizes $l(\beta)$. The function $l(\beta)$ is a well-defined log-likelihood function that satisfies all the necessary and sufficient regularity conditions for the CMLE to be consistent and asymptotically normal.

QUESTION 2: Consider the following model:

$$Y = \begin{cases} X\ \beta_0 + \varepsilon_0 & \text{if} \quad D = 0 \\ \\ X\ \beta_1 + \varepsilon_1 & \text{if} \quad D = 1 \end{cases}$$

with

$$D = 1\{Z\ \delta - u > 0\}$$

where the unobservable variables $(\varepsilon_0, \varepsilon_1, u)$ **are independent of** X **and** Z **and they have a joint normal distribution with zero means. The researcher observes a random sample of** $\{Y_i, D_i, X_i, Z_i : i = 1, 2, ..., N\}$**.**

Question 2.1 (10 points): Show that

$$E(\varepsilon_0 \mid X, Z, D = 0) = \frac{\sigma_{0u}}{\sigma_u}\ \phi\left(\frac{Z\delta}{\sigma_u}\right) / \left[1 - \Phi\left(\frac{Z\delta}{\sigma_u}\right)\right]$$

and

$$E(\varepsilon_1 \mid X, Z, D = 1) = \frac{-\sigma_{1u}}{\sigma_u}\ \phi\left(\frac{Z\delta}{\sigma_u}\right) / \Phi\left(\frac{Z\delta}{\sigma_u}\right)$$

where $\sigma_{0u} \equiv cov(\varepsilon_0, u)$**,** $\sigma_{1u} \equiv cov(\varepsilon_1, u)$**, and** $\sigma_u^2 \equiv var(u)$**.**

Since $(\varepsilon_0, \varepsilon_1, u)$ are jointly normal, we have that $E(\varepsilon_0|u) = \frac{\sigma_{0u}}{\sigma_u^2}u$, and $E(\varepsilon_1|u) = \frac{\sigma_{1u}}{\sigma_u^2}u$. Furthermore, for any set A within the real line:

$$\begin{aligned} E(\varepsilon_0 \mid u \in A) &= \int E(\varepsilon_0|u)\ pdf(u|\ u \in A)\ du \\ &= \int_{u \in A} \frac{\sigma_{0u}}{\sigma_u^2} u\ \frac{pdf(u)}{\Pr(u \in A)}\ du \\ &= \frac{\sigma_{0u}}{\sigma_u^2} \frac{1}{\Pr(u \in A)} \int_{u \in A} \frac{u}{\sigma_u}\ \phi\left(\frac{u}{\sigma_u}\right)\ du \end{aligned}$$

where $\phi\,(.)$ is the PDF of the standard normal distribution. Similarly,

$$E(\varepsilon_1 \mid u \in A) = \frac{\sigma_{1u}}{\sigma_u^2} \frac{1}{\Pr(u \in A)} \int_{u \in A} \frac{u}{\sigma_u}\ \phi\left(\frac{u}{\sigma_u}\right)\ du$$

The event $D = 0$ is equivalent to the event $u \geq Z\delta$. Therefore, $E(\varepsilon_0|X, Z, D = 0) = E(\varepsilon_0|X, Z, u \geq Z\delta)$, and:

$$E(\varepsilon_0|X, Z, D = 0) = \frac{\sigma_{0u}}{\sigma_u^2} \frac{1}{\Pr(u \geq Z\delta)} \int_{Z\delta}^{\infty} \frac{u}{\sigma_u}\ \phi\left(\frac{u}{\sigma_u}\right)\ du$$

To solve the integral, we use the change in variable $u^* = u/\sigma_u$ that implies $du = \sigma_u \; du^*$. Then,

$$E(\varepsilon_0|X,Z,D=0) = \frac{\sigma_{0u}}{\sigma_u} \frac{1}{1-\Phi(Z\delta/\sigma_u)} \int_{Z\delta//\sigma_u}^{\infty} u^* \; \phi\left(u^*\right) \; du^*$$

Finally, taking into account that $\phi\left(u\right) = (2\pi)^{-1/2}\exp\{-u^2/2\}$, it is simple to show that $u^*\phi\left(u^*\right) = -\dfrac{d\phi\left(u^*\right)}{du^*}$. Thus, we have:

$$\begin{aligned}
E(\varepsilon_0|X,Z,D=0) &= \frac{\sigma_{0u}}{\sigma_u} \frac{1}{1-\Phi(Z\delta/\sigma_u)} \int_{Z\delta//\sigma_u}^{\infty} - d\phi\left(u^*\right) \\
&= \frac{\sigma_{0u}}{\sigma_u} \frac{1}{1-\Phi(Z\delta/\sigma_u)} \left[-\phi\left(\infty\right) + \phi\left(Z\delta/\sigma_u\right)\right] \\
&= \frac{\sigma_{0u}}{\sigma_u} \frac{\phi\left(Z\delta/\sigma_u\right)}{1-\Phi(Z\delta/\sigma_u)}
\end{aligned}$$

Similarly, the event $D=1$ is equivalent to the event $u < Z\delta$. Therefore, $E(\varepsilon_1|X,Z,D=1) = E(\varepsilon_1|X,Z,u<Z\delta)$, and:

$$\begin{aligned}
E(\varepsilon_1|X,Z,D=1) &= \frac{\sigma_{1u}}{\sigma_u^2} \frac{1}{\Pr(u<Z\delta)} \int_{-\infty}^{Z\delta} \frac{u}{\sigma_u} \; \phi\left(\frac{u}{\sigma_u}\right) \; du \\
&= \frac{\sigma_{1u}}{\sigma_u} \frac{1}{\Phi(Z\delta/\sigma_u)} \int_{-\infty}^{Z\delta/\sigma_u} - d\phi\left(u^*\right) \\
&= -\frac{\sigma_{1u}}{\sigma_u} \frac{\phi(Z\delta/\sigma_u)}{\Phi(Z\delta/\sigma_u)}
\end{aligned}$$

Question 2.2 (10 points): Based on the model, we can write the regression equation

$$Y = (1-D)X \; \beta_0 + DX \; \beta_1 + e$$

where the error term e is $(1-D)\varepsilon_0 + D\varepsilon_1$. Obtain the expression of $E(e|X,Z)$. Show that the OLS estimation of this equation provides inconsistent estimates. Explain why.

First, we obtain expression for $E(e|X,Z)$.

$$\begin{aligned}
E(e|X,Z) &= E\left((1-D)\varepsilon_0 + D\varepsilon_1 \mid X,Z\right) \\
&= \Pr\left(D=0|X,Z\right) \; E(\varepsilon_0|X,Z,D=0) + \Pr\left(D=1|X,Z\right) \; E(\varepsilon_1|X,Z,D=1)
\end{aligned}$$

Given the expressions that we have derived in Question 2.1, and given that $\Pr(D=0|X,Z) = 1-\Phi(Z\delta/\sigma_u)$ and $\Pr(D=1|X,Z) = \Phi(Z\delta/\sigma_u)$, we have that:

$$E(e|X,Z) = \left(\frac{\sigma_{0u}}{\sigma_u} - \frac{\sigma_{1u}}{\sigma_u}\right) \; \phi\left(Z\delta/\sigma_u\right)$$

The error term e is mean independent of X and Z if and only if $E(e|X,Z)$ is zero for any value of (X,Z). Given the previous expression, it is clear that e is mean independent of X and Z if and only if $\sigma_{0u} = \sigma_{1u}$. That is, the model should be one where $\varepsilon_0 = \pi u + v_0$ and $\varepsilon_1 = \pi u + v_1$ where π is a constant and v_0 and v_1 are not correlated with u.

The OLS estimator in the regression of Y on $(1-D)X$ and DX is consistent if the error term e is not correlated with D and X. The expression $E(e|X,Z) = \left(\frac{\sigma_{0u}}{\sigma_u} - \frac{\sigma_{1u}}{\sigma_u}\right) \phi\left(Z\delta/\sigma_u\right)$ shows that e is correlated with Z. In general, the observables in Z and X will be correlated. Therefore, e is correlated with X. Furthermore, e is correlated with D. To see this, note that:

$$
\begin{aligned}
E\left(D\ e\right) &= E\left(D\ \left[(1-D)\varepsilon_0 + D\varepsilon_1\right]\right) = E\left(D\ \varepsilon_1\right) = E_{X,Z}\left(E\left(D\ \varepsilon_1 \mid X,Z\right)\right) \\
&= E_{X,Z}\left(\Pr\left(D=1 \mid X,Z\right)\ E\left(\varepsilon_1 \mid X,Z,D=1\right)\right) \\
&= E_{X,Z}\left(-\frac{\sigma_{1u}}{\sigma_u}\ \phi(Z\delta/\sigma_u)\right)
\end{aligned}
$$

And, in general, $E_{X,Z}\left(-\frac{\sigma_{1u}}{\sigma_u}\ \phi(Z\delta/\sigma_u)\right)$ is different to zero.

Question 2.3 (20 points): Consider again the regression equation $Y = (1-D)X\beta_0 + DX\beta_1 + e$, where $e = (1-D)\varepsilon_0 + D\varepsilon_1$. Suppose that the vector Z contains variables not included in X such that we have exclusion restrictions. Let $\hat{D}(Z)$ be the fitted value of D from the probit model of D on Z, i.e., $\hat{D}(Z) = \Phi(Z'\hat{\delta}/\sigma_u)$. Suppose that we run an OLS estimation for Y on $(1-\hat{D}(Z))X$ and $\hat{D}(Z)X$. Show that this estimator of β_0 and β_1 is inconsistent. Explain why. What if we assume that $\varepsilon_1 = \varepsilon_0$? What if we assume that $\varepsilon_1 = \lambda\varepsilon_0$ for a constant parameter λ?

This OLS estimator is in the spirit of an IV or 2SLS estimator. The first step is a "probit regression" of the endogenous variable D on the vector of "instruments" Z. In the second step, we run an OLS regression of Y on $(1-\hat{D}(Z))X$ and $\hat{D}(Z)X$. Note that $(1-\hat{D}(Z))X$ and $\hat{D}(Z)X$ are consistent estimators of $E((1-D)X|X,Z)$ and $E(DX|X,Z)$, respectively. Therefore, the OLS estimation in the second step (i.e., the IV estimator) is consistent if the orthogonality conditions

$$
E\left(Y - (1-D)X\ \beta_0 + DX\ \beta_1 \mid X,Z\right) = 0
$$

are true in this model. That is, in the regression of Y on $(1-\hat{D}(Z))X$ and $\hat{D}(Z)X$, the OLS estimator is consistent if the error term e is mean independent of Z and X, i.e., if $E(e|X,Z) = 0$. However, we have shown in Question 2.2 that $E(e|X,Z) = \left(\frac{\sigma_{0u}}{\sigma_u} - \frac{\sigma_{1u}}{\sigma_u}\right) \phi\left(Z\delta/\sigma_u\right)$, that in general is not zero. Therefore, without further restrictions, this estimator is not consistent.

If we assume that $\varepsilon_1 = \varepsilon_0$, we have that $\sigma_{1u} = \sigma_{0u}$ and $E(e|X,Z) = 0$. Therefore, under this assumption this OLS estimator is consistent. Note that we could consider a weaker condition: if $\varepsilon_0 = \pi u + v_0$ and $\varepsilon_1 = \pi u + v_1$ where π is a constant and v_0 and v_1 are not correlated with u, then $\sigma_{1u} = \sigma_{0u}$.

If we assume that $\varepsilon_1 = \lambda \varepsilon_0$, then we have that $\sigma_{1u} = \lambda \sigma_{0u}$ and

$$E(e|X, Z) = \frac{(1-\lambda)\sigma_{0u}}{\sigma_u} \phi\left(Z\delta/\sigma_u\right)$$

Therefore, if $\lambda \neq 0$, this OLS estimator is inconsistent under this assumption.

Question 2.4 (20 points): Based on the result in Question 2.1, describe a two-step method to obtain consistent estimates of the parameters β_0, β_1, $\frac{\sigma_{0u}}{\sigma_u}$, and $\frac{\sigma_{1u}}{\sigma_u}$. Explain how this estimator can be generalized to a semiparametric model where we relax the normality assumption on the distribution of the unobservables.

Heckman's two-step parametric method. Based on the results in Question 2.1, we have that using the subsample of observations with $D = 0$, we have the regression model:

$$Y = X\ \beta_0 + \frac{\sigma_{0u}}{\sigma_u}\ \lambda_0(Z) + e_0 \qquad \text{if } D = 0$$

where $\lambda_0(Z) \equiv \phi\left(\frac{Z\delta}{\sigma_u}\right) / \left[1 - \Phi\left(\frac{Z\delta}{\sigma_u}\right)\right]$, and $e_0 \equiv e - \lambda_0(Z)$. By construction, we have that $E(e_0|X, Z, D = 0) = 0$. Therefore, if we knew the selection term $\lambda_0(Z)$, we could run an OLS estimation of Y on X and $\lambda_0(Z)$ using the subsample of observations with $D = 0$ to obtain a consistent estimator of β_0 and $\frac{\sigma_{0u}}{\sigma_u}$. Though the selection term $\lambda_0(Z)$ is unknown, we can estimate it consistently using a probit regression for D on Z. Thus, the two-step method proceeds as follows.

Step 1. Probit regression for D on Z to obtain a consistent estimator of δ/σ_u. Using this estimator, $\widehat{\delta/\sigma_u}$, we can estimate consistently the selection term $\lambda_0(Z)$ for every value Z in the sample: $\hat{\lambda}_0(Z) = \phi\left(Z\ (\widehat{\delta/\sigma_u})\right) / \left[1 - \Phi\left(Z\ (\widehat{\delta/\sigma_u})\right)\right]$.

Step 2. We use the subsample of observations with $D = 0$ to run an OLS regression of Y on X and $\hat{\lambda}_0(Z)$.

Though the OLS in Step 2 provides consistent estimates of β_0 and $\frac{\sigma_{0u}}{\sigma_u}$, the OLS standard errors are not correct because they ignore the estimation error in $\hat{\lambda}_0(Z)$. Amemiya (1985) provides the expression for the correct standard errors. A simple approach to obtain correct standard errors is to apply one BHHH iteration for the maximization of the likelihood function using the two-step estimator as the initial value. As usual in MLE, the inverse of the outer-product of the scores is a consistent estimator of the variance-covariance matrix.

Similarly, using the subsample of observations with $D = 1$, we have the regression model:

$$Y = X\ \beta_1 + \frac{\sigma_{1u}}{\sigma_u}\ \lambda_1(Z) + e_1 \qquad \text{if } D = 1$$

where $\lambda_1(Z) \equiv \phi\left(\frac{Z\delta}{\sigma_u}\right) / \Phi\left(\frac{Z\delta}{\sigma_u}\right)$, and $e_1 \equiv e - \lambda_1(Z)$. By construction, we have that $E(e_1|X, Z, D = 1) = 0$. We can apply a two-step procedure similar to the one described above to estimate consistently β_1 and $\frac{\sigma_{1u}}{\sigma_u}$. The only differences with respect to the procedure described above are the different expression for the function $\lambda_1(Z)$, and that in the second step now we use the subsample of observations with $D = 1$.

Heckman's two-step semiparametric method. Under certain conditions, the previous two-step method can be extended to a semiparametric model where the unobservables $(\varepsilon_0, \varepsilon_1, u)$ can have any probability distribution. The following are sufficient conditions that allow for that semiparametric estimation: (1) $(\varepsilon_0, \varepsilon_1, u)$ are independent of X and Z; (2) the vector Z contains a continuous variable with full support over the real line; and (3) the vector Z contains at least one variable that is not included in the vector X, i.e., exclusion restriction.

Under conditions (1) and (2) we can use Klein-Spady method to estimate a semiparametric binary choice model of D on Z. More specifically, we can estimate consistently the vector of parameters δ and the CDF of u, $F_u(.)$. Of course, that means that we can estimate consistently the "propensity score" $\Pr(D = 1|Z) = F_u(Z\delta)$ for every value of Z in the sample.

Under condition (1), it is possible to show that, for any distribution of $(\varepsilon_0, \varepsilon_1, u)$, the selection terms $E(\varepsilon_0|X, Z, D = 0)$ and $E(\varepsilon_1|X, Z, D = 1)$ are functions ONLY of the propensity score $F_u(Z\delta)$, i.e.,

$$\begin{aligned} E(\varepsilon_0|X, Z, D = 0) &= s_0(F_u(Z\delta)) \\ E(\varepsilon_1|X, Z, D = 1) &= s_1(F_u(Z\delta)) \end{aligned}$$

The form of the functions $s_0(.)$ and $s_1(.)$ depend on the form of the distribution of $(\varepsilon_0, \varepsilon_1, u)$. Since we do not know the distribution of $(\varepsilon_0, \varepsilon_1, u)$, we should specify the functions $s_0(.)$ and $s_1(.)$ nonparametrically. For instance, we may consider a polynomial series approximation in $F_u(Z\delta)$ to approximate these functions:

$$\begin{aligned} s_0(F_u(Z\delta)) &= \sigma_{00} F_u(Z\delta) + \sigma_{01}[F_u(Z\delta)]^2 + ... + \sigma_{0J}[F_u(Z\delta)]^J \\ s_1(F_u(Z\delta)) &= \sigma_{10} F_u(Z\delta) + \sigma_{11}[F_u(Z\delta)]^2 + ... + \sigma_{1J}[F_u(Z\delta)]^J \end{aligned}$$

Under condition (3), the terms of these polynomial, $\{F_u(Z\delta)$, $[F_u(Z\delta)]^2$, ..., $[F_u(Z\delta)]^J\}$ are not perfectly collinear with X. Therefore, an OLS regression of Y on X and $\{F_u(Z\delta)$, $[F_u(Z\delta)]^2$, ..., $[F_u(Z\delta)]^J\}$ can identify the parameters of interest together with the σ parameters.

According to this semiparametric model:

$$Y = X\ \beta_0 + \sigma_{00} F_u(Z\delta) + ... + \sigma_{0J}[F_u(Z\delta)]^J + e_0 \qquad \text{if } D = 0$$

where $e_0 \equiv e - s_0(F_u(Z\delta))$, and by construction we have that $E(e_0|X, Z, D = 1) = 0$.

Step 1. We apply Klein-Spady method to the binary choice model for D on Z to obtain consistent estimators of δ and of the CDF of u, $F_u(.)$. Then, we obtain estimates of "propensity score" $\hat{F}_u(Z\hat{\delta})$ for any value Z in the sample.

Step 2. We use the subsample of observations with $D = 0$ to run an OLS regression of Y on X and $\{\hat{F}_u(Z\hat{\delta})$, $[\hat{F}_u(Z\hat{\delta})]^2$, ..., $[\hat{F}_u(Z\hat{\delta})]^J\}$.