

ECN726

MIDTERM

Dilsat Dalkiran Ozel

November 2013

PART 1: PROPERTIES OF ESTIMATORS

A) 2SLS

State 1: Regress each endogenous variable on all instruments and exogenous variables

$$\begin{aligned}x_1 &= z\gamma + r \\ \hat{x}_1 &= z\hat{\gamma}\end{aligned}$$

where

$$\hat{\gamma} = (z'z)^{-1}z'x_1$$

Stage 2: Regress Y on \hat{x}_1 and x_2 to get $\hat{\beta}_{2SLS}$.

$$\begin{aligned}y &= \hat{x}_1\beta_1 + x_2\beta_2 + \mu \\ \hat{\beta}_{12SLS} &= (\hat{x}_1'\hat{x}_1)^{-1}(\hat{x}_1'y) \\ &= (\hat{\gamma}'z'z\hat{\gamma})^{-1}(\hat{\gamma}'z'y) \\ &= (\hat{\gamma}'z'z\hat{\gamma})^{-1}[\hat{\gamma}'z'(\hat{x}_1\beta_1 + x_2\beta_2 + \mu)] \\ &= \beta_1 + \beta_2(\hat{\gamma}'z'z\hat{\gamma})^{-1}\hat{\gamma}'z'x_2 + (\hat{\gamma}'z'z\hat{\gamma})^{-1}\hat{\gamma}'z'\mu\end{aligned}$$

Now, derive the probability limit of the 2SLS estimator for β_1 .

$$\begin{aligned}plim(\hat{\beta}_{12SLS}) &= \beta_1 + \beta_2 plim\left(\frac{1}{N}x_1'P_zx_1\right)^{-1} plim\left(\frac{1}{N}x_1'P_zx_2\right) \\ &\quad + plim\left(\frac{1}{N}x_1'P_zx_1\right)^{-1} plim\left(\frac{1}{N}x_1'P_z\mu\right) \\ &= \beta_1\end{aligned}\tag{1}$$

Consistency result at (1) follows from the assumptions below. Note that $P_z = z(z'z)^{-1}z'$. Assumptions which are necessary for consistency:

$$plim \left(\frac{1}{N} x_1' P_z \mu \right) = 0 \quad (2)$$

as $Cov(z, \mu) = 0$ by validity of z as an instrument.

By Law of Large Numbers,

$$plim \left(\frac{1}{N} x_1' P_z x_1 \right)^{-1} = 0 \quad (3)$$

as $x_1' P_z x_1$ is finite. Lastly and most importantly, to get the consistency result (1), we should have $Cov(z, x_2) = 0$ so that

$$plim \left(\frac{1}{N} x_1' P_z x_2 \right) = 0 \quad (4)$$

B) Probit Model

i) Demonstration

First of all, we know that Probit model arises if the error term is standard normal. In this model, the first thing we realize that the error term is not standard. Hence, a researcher should convert this distribution to standard normal to use Probit estimation. For that we can use Theorem A17 at the appendix which is called "Product Limit Normal Rule". However, as the question asks us to demonstrate the result, I derived the result by first finding the MLE estimator with the given distribution. With some change of variables, I showed that foc is exactly the same as the foc we have when we had standard normal distribution. Analytically,

$$\begin{aligned} \ln L &= \sum_{i=0}^N y_i \ln \int_{-\infty}^{X_i \beta} \frac{1}{\kappa \sqrt{2\pi}} \exp \left(-\frac{1}{2} \frac{z^2}{\kappa^2} \right) dz \\ &+ \sum_{i=0}^N (1 - y_i) \ln \left(1 - \int_{-\infty}^{X_i \beta} \frac{1}{\kappa \sqrt{2\pi}} \exp \left(-\frac{1}{2} \frac{z^2}{\kappa^2} \right) dz \right) \end{aligned}$$

Maximize this by choosing β :

$$\begin{aligned}
foc(\beta) &= \sum_{i=0}^N \frac{y_i}{\int_{-\infty}^{X_i\beta} \frac{1}{\kappa\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{z^2}{\kappa^2}\right) dz} \frac{1}{\kappa\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(X_i\beta)^2}{\kappa^2}\right) X_i \\
&- \sum_{i=0}^N \frac{(1-y_i)}{1 - \int_{-\infty}^{X_i\beta} \frac{1}{\kappa\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{z^2}{\kappa^2}\right) dz} \frac{1}{\kappa\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(X_i\beta)^2}{\kappa^2}\right) X_i \\
&= \left[y_i - \int_{-\infty}^{X_i\beta} \frac{1}{\kappa\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{z^2}{\kappa^2}\right) dz \right] X_i \\
&\quad * \frac{\frac{1}{\kappa\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(X_i \frac{\beta}{\kappa}\right)^2\right)}{\int_{-\infty}^{X_i\beta} \frac{1}{\kappa\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{z^2}{\kappa^2}\right) dz \left(1 - \int_{-\infty}^{X_i\beta} \frac{1}{\kappa\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{z^2}{\kappa^2}\right) dz\right)}
\end{aligned}$$

Now, make change of variables: Let $\tilde{z} = \frac{z}{\kappa}$. Then, $d\tilde{z} = \frac{dz}{\kappa}$ and $X_i\beta = \tilde{z}\kappa$. Also recall that $\kappa \neq 1$.

By changing of variables and after some simplifications we get exactly the same foc of $\tilde{\beta} = \frac{\beta}{\kappa}$ of a Probit estimation with standard normal distribution, which is

$$\sum_{i=0}^N \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{X_i\tilde{\beta}}{2}\right)}{\int_{-\infty}^{X_i\tilde{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \tilde{z}^2\right) d\tilde{z} \left(1 - \int_{-\infty}^{X_i\tilde{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \tilde{z}^2\right) d\tilde{z}\right)} \left[y_i - \int_{-\infty}^{X_i\tilde{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \tilde{z}^2\right) d\tilde{z} \right] X_i$$

ii) Identification

Identification of a single-index model requires some restrictions on the variance of error term as the single-index model can only identify β up to a scale (page 476, CT). Hence, for example, placing a restriction that $v(u) = 1$ secures uniqueness of β . However, the "problem" is not the unobserved variable itself. If there is an unobserved variable in the model, Probit model is a method to estimate these kind of models. So, when we look at the foc of beta above, the fact that $foc = 0$ implies that residual are uncorrelated with the covariates. Hence, we can estimate β 's from the foc although it is hard.

C) Minimum Years for Consistency

Notes and Assumptions:

- α : unobserved, time-constant

Then, $E[\alpha_i | d_{it}] \neq 0$ implies that it is either c (constant) or $f(d_{it})$

- $E[\epsilon_{it} | w_{it-1}, d_{it}, x_{it}, \alpha_i] = 0$

As our main aim is to find minimum years required for consistency, I am choosing FD estimator. Then, consistency requires zero covariance between first difference in error term with first difference in covariates. Analytically,

$$w_{it} - w_{it-1} = (w_{it-1} - w_{it-2})\beta + (d_{it} - d_{it-1})\gamma + (x_{it} - x_{it-1})\delta + (\epsilon_{it} - \epsilon_{it-1})$$

As α_i is the time invariant, it cancels out once we take the first difference. Now, define $a_t - a_{t-1} = \Delta a_t$ for every a which is a generic term. Then,

$$\Delta w_{it} = \Delta w_{it-1}\beta + \Delta d_{it}\gamma + \Delta x_{it}\delta + \Delta \epsilon_{it}$$

Hence, the FD estimators are as follows:

$$\begin{aligned} \hat{\beta}_{FD} &= \left[\sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta w_{it-1} \right]^{-1} \sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta w_{it} \\ &= \beta + \gamma \left[\sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta w_{it-1} \right]^{-1} \sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta d_{it} \\ &+ \delta \left[\sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta w_{it-1} \right]^{-1} \sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta x_{it} \\ &+ \left[\sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta w_{it-1} \right]^{-1} \sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta \epsilon_{it} \\ \text{plim}(\hat{\beta}_{FD}) &= \beta + \gamma \text{plim} \left[\frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta w_{it-1} \right]^{-1} \text{plim} \left[\sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta d_{it} \right] \\ &+ \delta \text{plim} \left[\frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta w_{it-1} \right]^{-1} \text{plim} \left[\sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta x_{it} \right] \\ &+ \text{plim} \left[\frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta w_{it-1} \right]^{-1} \text{plim} \left[\sum_{i=1}^N \sum_{t=2}^T \Delta w'_{it-1} \Delta \epsilon_{it} \right] \quad (5) \end{aligned}$$

As the error term is serially uncorrelated and $E[\epsilon_{it}|w_{it-1}, d_{it}, x_{it}, \alpha_i] = 0$ the last term on equation (5) is zero. I assume there is no multicollinearity among the covariates hence second and third term are also equal to zero. Hence, by Slutsky Thm and LLN, the estimator is consistent.

Same assumptions also guarantees the consistency of $\hat{\gamma}_{FD}$ and $\hat{\delta}_{FD}$. If I have time I am going to derive them too but for now I will skip this step as it is exactly what I did above. Now, the minimum number of years that is necessary for consistency is 3. Why? Because, it is necessary to take the first difference of all variables.

PART 2: ECONOMETRIC RESEARCH DESIGN

Notes:

- Panel data
- FRR (Fatal risk rate) does not change over the time horizon a lot
does not necessarily mean that it is not changing at all
- There is individual heterogeneity which is job skill
sign for fixed effect!
mechanism: high skill workers may choose low FRR jobs
- Most workers remain in the same ind (industry) and occ (occupation) throughout the study period
- effect of change in FRR on w (wages)

The equation I plan to estimate is the following:

$$w_{isgt} = edu_{isgt}\beta + skill_i + DFRR_{isgt}\alpha + FRR_{isgt}\psi + exp_{isgt}\gamma + ind_{gi}\theta + occ_{is}\phi + u_{isgt}$$

where,

$$DFRR = \begin{cases} 1 & \text{if there is a change in FRR over the period} \\ 0 & \text{if there is no change in FRR} \end{cases}$$

DFRR is a dummy variable as defined above. i is an index for individuals, g is index for occupation, s is index for industry. The reason that that I used different indices for both ind. and occ is that FRR changes within the industries. w is wages as expected, edu is education, $skill$ is individually heterogeneous job skill, exp is experience and lastly u is the error term.

The main reason that I choose the use DID estimation is the following: At first, I thought about using Fixed Effect estimation in order to get rid of omitted variable problem. However, for that FRR should change over time. What we know is it is changing but the rate of change is near zero. Hence, in order to grasp the exact effect of FRR on wages, DID estimation looks more plausible. Here the main question is, whether FRR is changing, if yes, how it affects wage differentials. There may be a case that although FRR is not changing over the time periods, it may change within the industries due to some exogenous shock. This can affect the wage rate within the industries. Or else, FRR may vary through time which bring about a change in the wages between industries. All these different sources of variation can easily be grasped by using DID estimation. In other words, by using the DID estimation, I am going to catch both within-groups and over-time-horizon effects.

The main caveat about using DID estimation is the assumption of consistency of DID estimator.

PART 3: MULTI-SAMPLE GMM

A) Demonstration

$$\begin{aligned}\hat{\theta}_{2SGMM} &= \underset{\theta \in \Theta}{\operatorname{argmin}} \frac{N_1}{N_1 + N_2} Q_1(\theta) + \frac{N_2}{N_1 + N_2} Q_2(\theta) \\ Q_1(\theta) &= \left[\frac{1}{N_1} \sum_{i=1}^{N_1} h_1(v_i, \theta) \right]' W_1 \left[\frac{1}{N_1} \sum_{i=1}^{N_1} h_1(v_i, \theta) \right] \\ Q_2(\theta) &= \left[\frac{1}{N_2} \sum_{i=1}^{N_2} h_2(v_i, \theta) \right]' W_1 \left[\frac{1}{N_2} \sum_{i=1}^{N_2} h_2(v_i, \theta) \right]\end{aligned}$$

Given the definitions of g_j and G_j in the question sheet, equation (3) in the question sheet follows.

By mean value theorem,

$$g_1(\hat{\theta}) = g_1(\hat{\theta}_0) + G_1(\theta^+)(\hat{\theta} - \theta_0)$$

$$g_2(\hat{\theta}) = g_2(\hat{\theta}_0) + G_2(\theta^+)(\hat{\theta} - \theta_0)$$

$$N_1 G_1(\hat{\theta})' W_1 g_1(\hat{\theta}) = N_1 G_1(\hat{\theta})' W_1 g_1(\hat{\theta}_0) + N_1 G_1(\hat{\theta})' W_1 G_1(\theta^+)(\hat{\theta} - \theta_0) \quad (6)$$

$$N_2 G_2(\hat{\theta})' W_2 g_2(\hat{\theta}) = N_2 G_2(\hat{\theta})' W_2 g_2(\hat{\theta}_0) + N_2 G_2(\hat{\theta})' W_2 G_2(\theta^+)(\hat{\theta} - \theta_0) \quad (7)$$

(6) + (7) = 0 by foc of $Q_N(\hat{\theta})$ implies that

$$\sqrt{N}(\hat{\theta} - \theta_0) = \left[N_1 G_1(\hat{\theta})' W_1 G_1(\theta^+) + N_2 G_2(\hat{\theta})' W_2 G_2(\theta^+) \right]^{-1} \sqrt{N} \left[N_1 G_1(\hat{\theta})' W_1 g_1(\hat{\theta}_0) + N_2 G_2(\hat{\theta})' W_2 g_2(\hat{\theta}_0) \right]$$

Assuming the data are independent over i and $E[h_i(v, \theta)] = 0$, by CLT $\sqrt{N}g_i(\theta_0)$ converges in distribution to $N(0, S)$.

By assumptions (A1) – (A5) + (A7) + (A9) + (A11) – (A12) in the handout "Asymptotic Properties of GMM Estimators", $G_j(\hat{\theta})$ and $G_j(\theta^+)$ converges to G_0

After some simplifications, the result follows. (I will write later it if I have time and energy.)

B)

iii) Discussion

As we update the weights matrix, it converges to Ω^{-1} . Hence, we can say that, the estimates become more efficient from (i) to (iii). In the graph, there is no difference between (ii) and (iii). This is because, in my calculations, I have reached the asymptotically efficient estimators faster. The reason may be the iteration restrictions I put on *fminunc* command or the moment conditions I used. Additionally, the main reason that I used Quasi-Newton method is the fact that my objective function is differentiable. We have already learned that Quasi-Newton model gives more information about the curvature properties of the objective function.

	(i)	(ii)	(iii)
Parameters	1st Step	2nd Step	Iterated
α	64.2729	64.2879	64.2879
γ	0.9393	1.0457	1.0457
β	-0.0295	-0.9731	-0.9731
δ	-0.2921	0.5600	0.5600