Some Equations for Econometrics I

Contents

1 The Linear Regression Model

1.1 Simple Regression Model

$y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, $i = 1, 2, \cdots, n$.
Or, $y = \beta_1 + \beta_2 x + \epsilon$ where y, x and ϵ are $n \times 1$ vector.
• The method of least squares: $\min_{\beta_1,\beta_2} \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i)^2$.
• The normal equations (FOC):

$$\begin{aligned} \sum_{i=1}^n y_i &= n\beta_1 + \beta_2 \sum_{i=1}^n x_i, \\ \sum_{i=1}^n x_i y_i &= \beta_1 \sum_{i=1}^n x_i + \beta_2 \sum_{i=1}^n x_i^2. \end{aligned}$$

• R^2-measure of fit: (Coefficient of determination)

$$\begin{aligned} R^2 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ (Optional) &= \hat{\beta}_2^2 \sum_{i=1}^n (x_i - \bar{x})^2 / \sum_{i=1}^n (y_i - \bar{y})^2 = r_{xy}^2 \end{aligned}$$

• Testing $\beta_1 = 0$.

$$t_{\beta_1} = \frac{\hat{\beta}_1 - 0}{se_{\beta_1}} \sim t_{n-2}$$

1.2 Multiple Regression

• Least Squares

$$\min_\beta S(\beta) \equiv (y - X\beta)'(y - X\beta).$$

• Normal equations: $\frac{\partial S}{\partial \beta} = -2X'y + 2X'X\beta = 0$

$$\bar{R}^2 = 1 - \frac{n-1}{n-K}(1 - R^2)$$

We can write $(X'X)\hat{\beta} = X'y$ as

$$\begin{aligned} X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 &= X_1'y \\ X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 &= X_2'y. \end{aligned}$$

$$\begin{aligned} \hat{\beta}_1 &= (X_1'M_2X_1)^{-1}X_1'M_2y \\ \hat{\beta}_2 &= (X_2'M_1X_2)^{-1}X_2'M_1y. \end{aligned}$$

Thus, $\hat{\beta}_1 = (X_1^{*\prime}X_1^*)^{-1}X_1^{*\prime}y^*$ where $X_1^* = M_2X_1$ and $y^* = M_2y$.

2 Statistical Inferences of Linear Regression Model

2.1 Under the Normality Assumption

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}([(X'X)^{-1}]_{kk})^{1/2}} = \frac{\frac{\hat{\beta}_k - \beta_k}{\sigma([(X'X)^{-1}]_{kk})^{1/2}}}{\sqrt{\frac{(n-K)\frac{\hat{\sigma}^2}{\sigma^2}}{n-K}}} \sim t_{n-k}.$$

$$F = \frac{Wald}{J}\frac{\sigma^2}{\hat{\sigma}^2} = \frac{(R\hat{\beta}-q)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q)/J}{[(n-K)\hat{\sigma}^2/\sigma^2]/(n-K)}.$$

2.2 Classical Asymptotic Theory

Some basic asymptotic concepts and useful results:

Definition: *Convergence in probability* $X_n \xrightarrow{p} X$ or $\text{plim} X_n = X$.

A sequence of random variables $\{X_n\}$ is said to converge to a random variable X *in probability* if

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0, \text{ for all } \epsilon > 0.$$

Definition: *Convergence in distribution* $X_n \xrightarrow{d} X$.

A sequence $\{X_n\}$ is said to converge to X *in distribution* if the distribution function F_n of X_n converges to the distribution function F of X at every continuity point of F. (F is called the limiting distribution of $\{X_n\}$).

Let $\{X_n, Y_n\}$, $n = 1, 2, \cdots$ be a sequence of pairs of random variables. Then

a).

$$X_n \xrightarrow{d} X,\ Y_n \xrightarrow{p} 0 \implies X_n Y_n \xrightarrow{p} 0.$$

b).

$$X_n \xrightarrow{d} X,\ Y_n \xrightarrow{p} c \implies X_n + Y_n \xrightarrow{d} X + c,\ \ X_n Y_n \xrightarrow{d} cX,\ \ X_n/Y_n \xrightarrow{d} X/c, \text{ if } c \neq 0.$$

Law of Large Numbers

Proposition: (Chebyshev's theorem W.L.L.N) Let $E(X_i) = \mu_i$, $V(X_i) = \sigma_i^2$, $Cov(X_i, X_j) = 0$, $i \neq j$. Then,

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = 0 \implies \bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0.$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^n \mu_i$.

Central Limit Theorems

Theorem: (Lindberg-Feller Theorem C.L.T) Let $\{X_n\}$ be a sequence of independent random variables. Let $E(X_i) = \mu_i$, $E(X_i - \mu_i)^2 = \sigma_i^2 \neq 0$ exist. *Denote $C_n = (\sum_{i=1}^n \sigma_i^2)^{1/2}$. If no single dominates the variance average such that $\lim_{n\to\infty} \frac{\max(\sigma_i)}{\sum_{i=1}^n \sigma_i} = 0$, and $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sigma_i$ exists. then $\frac{\sum_{i=1}^n (X_i - \mu_i)}{C_n} \xrightarrow{d} N(0,1)$.*

3 Maximum Likelihood Estimation

3.1 MLE Basics

$$f(y_1, \cdots, y_n|\theta) = \prod_{i=1}^n f(y_i|\theta) = L(\theta|\mathbf{y}),$$

$$\ln L(\theta|y) = \sum_{i=1}^n \ln f(y_i, \theta),$$

Information Identity

$$E\left[\frac{\partial \ln L(\theta|y)}{\partial \theta}\frac{\partial \ln L(\theta|y)}{\partial \theta'}\right] + E\left[\frac{\partial^2 \ln L(\theta|y)}{\partial \theta \partial \theta'}\right] = 0.$$

• Information matrix is $I(\theta)$ where

$$I(\theta) = E\left[\frac{\partial \ln L(\theta|y)}{\partial \theta}\frac{\partial \ln L(\theta|y)}{\partial \theta'}\right].$$

3.2 MLE of Linear Regression Model

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta} &= \frac{1}{\sigma^2}X'(y - X\beta) = 0, \\ \frac{\partial \ln L}{\partial \sigma^2} &= \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)'(y - X\beta) = 0. \end{aligned}$$

MLE: $\hat{\beta}_{ML} = (X'X)^{-1}X'y$; $\hat{\sigma}^2_{ML} = e'e/n$.

• The information matrix of the linear regression model:

$$I(\beta, \sigma^2) = \left(-\frac{1}{n}E\frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'}\right)^{-1} = \begin{pmatrix} \sigma^2(\frac{1}{n}X'X)^{-1} & 0 \\ 0 & 2\sigma^4 \end{pmatrix}.$$

3.3 Asymptotic Tests

• Likelihood ratio test statistic:

$$-2\ln\left[\frac{\max_{h(\theta)=0} L(\theta|y)}{\max_\theta L(\theta|y)}\right] = 2[\ln L(\hat{\theta}|y) - \ln L(\bar{\theta}|y)].$$

• Wald test statistics:

$$h'(\hat{\theta})\left[\frac{\partial h(\hat{\theta})}{\partial \theta'}\left(-\frac{\partial^2 \ln L(\hat{\theta})}{\partial \theta \partial \theta'}\right)^{-1}\frac{\partial h(\hat{\theta})}{\partial \theta}\right]^{-1} h(\hat{\theta}).$$

• (Efficient) Score statistics:

$$\frac{\partial \ln L(\bar{\theta})}{\partial \theta'}\left(-\frac{\partial^2 \ln L(\bar{\theta})}{\partial \theta \partial \theta'}\right)^{-1}\frac{\partial \ln L(\bar{\theta})}{\partial \theta}.$$

3.4 Newton-Raphson Method

This method is applicable for either maximization or minimization problems. Let $Q_n(\theta)$ be the object function for optimization. Let θ_1 be an initial estimate of θ. By a quadratic approximation, define

$$Q_n(\theta) \equiv Q_n(\hat{\theta}_1) + \frac{\partial Q_n(\hat{\theta}_1)}{\partial \theta'}(\theta - \hat{\theta}_1) + \frac{1}{2}(\theta - \hat{\theta}_1)' \frac{\partial Q_n^2(\hat{\theta}_1)}{\partial \theta \partial \theta'}(\theta - \hat{\theta}_1).$$

Maximizing (or minimizing) the right-hand side approximation provides a second-round estimator $\hat{\theta}_2$,

$$\hat{\theta}_2 = \hat{\theta}_1 - \left[\frac{\partial Q_n^2(\hat{\theta}_1)}{\partial \theta \partial \theta'}\right]^{-1} \frac{\partial Q_n(\hat{\theta}_1)}{\partial \theta}.$$

The iteration is to be repeated until the sequence $\{\hat{\theta}_j\}$ converges. The step sizes in the iteration can also be modified as

$$\hat{\theta}_2 = \hat{\theta}_1 - \lambda \left[\frac{\partial Q_n^2(\hat{\theta}_1)}{\partial \theta \partial \theta'}\right]^{-1} \frac{\partial Q_n(\hat{\theta}_1)}{\partial \theta'},$$

4 Discrete Choice Models

4.1 Introduction

- **Pobit and Logit**

$$F(u) = \int_{-\infty}^{u} \phi(w) dw,$$

where ϕ is the standard normal density $\phi(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$, then this model is called *probit* model.

$$F(u) = \frac{e^u}{1 + e^u}$$

where $f(u) = \frac{\exp(u)}{(1+\exp(u))^2}$, it is known as the *logit* model.

4.2 Probit and Logit Models

$$\begin{aligned}
\ln L(\beta) &= \text{sum of} \begin{cases} \ln[1 - F(x_i\beta)] & \text{if } y_i = 0 \\ \ln F(x_i\beta) & \text{if } y_i = 1 \end{cases} \\
&= \sum_{i=1}^{n} [y_i \ln F(x_i\beta) + (1 - y_i) \ln(1 - F(x_i\beta))].
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L}{\partial \beta} &= \sum_{i=1}^{n} y_i \frac{f(x_i\beta)x_i'}{F(x_i\beta)} + \sum_{i=1}^{n} (1 - y_i) \frac{-f(x_i\beta)x_i'}{1 - F(x_i\beta)} \\
&= \sum_{i=1}^{n} \frac{y_i [1 - F(x_i\beta)] - (1 - y_i)F(x_i\beta)}{F(x_i\beta)[1 - F(x_i\beta)]} f(x_i\beta)x_i' \\
&= \sum_{i=1}^{n} \frac{y_i - F(x_i\beta)}{F(x_i\beta)(1 - F(x_i\beta))} f(x_i\beta)x_i',
\end{aligned}$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = \left\{ -\sum_{i=1}^{n} \left[\frac{y_i}{F^2(x_i\beta)} + \frac{1 - y_i}{[1 - F(x_i\beta)]^2} \right] f^2(x_i\beta) + \sum_{i=1}^{n} \frac{y - F(x_i\beta)}{F(x_i\beta)[1 - F(x_i\beta)]} f' \right\} x_i x_i'.$$

5 Truncation, Censoring and Sample Selection

5.1 Truncation

$$y_i = x_i'\beta + \epsilon_i, \quad \epsilon_i|x \sim N(0, \sigma^2)$$

$$f(y|x_i, c_i) = \frac{\phi(y|x_i'\beta, \sigma^2)}{\Phi(c_i|x_i'\beta, \sigma^2)} = \frac{\frac{1}{\sigma}\phi(\frac{yi - x_i'\beta}{\sigma})}{\Phi(\frac{ci - x_i'\beta}{\sigma})}, \; y \leq c_i, \tag{5.1}$$

5.2 Censoring

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \tag{5.2}$$

$$P(y_i = 0|x_i) = P(y_i^* < 0|x_i) = P(x_i'\beta + \epsilon_i < 0|x_i) = P(\frac{\epsilon_i}{\sigma} < -\frac{x_i'\beta}{\sigma}|x_i) = \Phi\left(-\frac{x_i'\beta}{\sigma}\right).$$

$$f(y|x_i) = \Phi\left(-\frac{x_i'\beta}{\sigma}\right)^{1(y=0)} \cdot \left(\frac{1}{\sigma}\phi\left(\frac{y_i - x_i\beta}{\sigma}\right)\right)^{1(y=1)}$$

$$\begin{aligned} \log L(\beta, \sigma) &= \textstyle\sum_{i=1}^{n} \log f(y_i|x_i) \\ &= \textstyle\sum_{y_i=0} \log \Phi\left(-\frac{x_i'\beta}{\sigma}\right) + \sum_{y_i=1} \log\left(\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right). \end{aligned}$$

5.3 Truncated Data and Censored Data: Tobit Models

$$\begin{aligned} y_i^* &= x_i'\beta + \epsilon_i, \quad \epsilon_i \text{ is i.i.d.} \\ y_i &= \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases}, \end{aligned} \tag{5.3}$$

$$E(y|y > 0) = E(y^*|y^* > 0) = x\beta + E(\epsilon|\epsilon > -x\beta) = x\beta + \sigma_\epsilon \cdot \frac{\phi\left(\frac{x\beta}{\sigma_\epsilon}\right)}{\Phi\left(\frac{x\beta}{\sigma_\epsilon}\right)}. \tag{5.4}$$

$$\begin{aligned} E(y|x) &= \Pr(y = 0) \cdot 0 + \Pr(y > 0|x) \cdot E(y|y > 0, x) \\ &= \Pr(\epsilon \leq -x\beta) \cdot 0 + \Pr(\epsilon > -x\beta)E(y^*|x; \epsilon > -x\beta) \\ &= \Pr(\epsilon > -x\beta)[x\beta + E(\epsilon|\epsilon > -x\beta)]. \end{aligned} \tag{5.5}$$

The likelihood function of the truncated data is equal to

$$L_1 = \textstyle\prod_1 [\Pr(y_i > 0|x_i)]^{-1} f(y_i),$$

The likelihood function for the censored data is equal to

$$\begin{aligned} L_2 &= \textstyle\prod_0 \Pr(y_i = 0|x_i) \prod_1 f(y_i) \\ &= \{\textstyle\prod_0 \Pr(y_i = 0|x_i) \prod_1 \Pr(y_i > 0|x_i)\} \\ &\quad \times \{\textstyle\prod_1 [\Pr(y_i > 0|x_i)]^{-1} f(y_i)\}, \end{aligned} \tag{5.6}$$

$$\begin{aligned} y_i &= E(y_i|x_i; y_i > 0) + \eta_i \\ &= x_i\beta + \sigma_\epsilon \frac{\phi\,(x\delta)}{\Phi\,(x\delta)} + \eta_i \text{ for those } i \,\text{such that } y_i > 0, \end{aligned}$$

where $E(\eta_i|x_i) = 0$, $var(\eta_i|x_i) = \sigma_\epsilon^2[1 - (x_i\delta)\lambda_i - \lambda_i^2]$, and $\lambda_i = \frac{\phi(x\delta)}{\Phi(x\delta)}$.

5.4 Sample Selection: Type II Tobit Model

$$\begin{aligned} y_i^* &= x_i'\beta + \epsilon_{1i} \\ d_i &= 1(z_i'\gamma + \epsilon_{0i} > 0) \end{aligned}$$

$$E(y_i|d_i = 1) = x_i\beta + E(\epsilon_{1i}|\epsilon_{0i} > -z_i'\gamma)\,.$$

$$\begin{aligned} E(\epsilon_{1i}|d_i &= 1, z_i) \\ &= E(\epsilon_{1i}|\{\epsilon_{0i} > -z_i'\gamma\}, z_i) \\ &= \rho E(\epsilon_{0i}|\{\epsilon_{0i} > -z_i'\gamma\}, z_i) + E(v_i|\{\epsilon_{0i} > -z_i'\gamma\}, z_i) \\ &= \rho\lambda(z_i'\gamma) \end{aligned}$$

$$y_i = x_i'\beta + \rho\lambda(z_i'\gamma) + u_i$$

is a valid regression equation for the observations for which $T_i = 1$.

6 Generalized Linear Regression Model

6.1 Generalized Least Squares

$$\hat{\beta}_G = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

• If $\Omega X = X\Gamma$ where Γ is a nonsingular matrix, then $\hat{\beta}_G = \hat{\beta}_{LS}$, i.e., GLS=OLS. (e.g., the SUR)
Proof: The equality $\Omega X = X\Gamma$ implies that $X\Gamma^{-1} = \Omega^{-1}X$ and hence $X'\Omega^{-1} = \Gamma'^{-1}X'$. Therefore,

$$(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = (\Gamma'^{-1}X'X)^{-1}\Gamma'^{-1}X' = (X'X)^{-1}\Gamma'\Gamma'^{-1}X' = (X'X)^{-1}X'.$$

6.2 WLS

$$\hat{\beta}_G = [\sum_{i=1}^{n} \omega_i^{-1}x_ix_i']^{-1}\sum_{i=1}^{n}\omega_i^{-1}x_iy_i.$$

6.3 Autocorrelated Disturbance

$$\sigma_u^2\Omega = \frac{\sigma_u^2}{(1-\rho^2)}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \cdots & \cdots & \cdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{pmatrix}.$$

$$P = \begin{pmatrix} \sqrt{(1-\rho^2)} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=2}^{T} e_t^2}.$$

6.4 Set of Regression Equations

$$\begin{aligned} Y_1 &= X_1\beta_1 + U_1 \\ &\vdots \\ Y_m &= X_m\beta_m + U_m \end{aligned}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} + U,$$

or $Y = X\beta + U$, where

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_m \end{pmatrix}.$$

- The GLS estimator is

$$\begin{aligned} \hat{\beta}_G &= (X'V^{-1}X)^{-1}X'V^{-1}Y \\ & \begin{pmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 & \cdots & \sigma^{1m}X_1'X_m \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{m1}X_m'X_1 & \sigma^{m1}X_m'X_2 & \cdots & \sigma^{mm}X_m'X_m \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{m}\sigma^{1j}X_1'Y_j \\ \vdots \\ \sum_{j=1}^{m}\sigma^{mj}X_m'Y_j \end{pmatrix}. \end{aligned}$$

7 Simultaneous-Equations Models and IV Estimation

7.1 IV Estimation

$$y_j = Y_j\gamma_j + X_j\beta_j + \epsilon_j = Z_j\delta_j + \epsilon_j,$$

$$\hat{\delta}_{j,IV} = [W_j'Z_j]^{-1}W_j'y_j.$$

$$Asy.Var[\hat{\delta}_{j,IV}] = \frac{\sigma_{jj}}{T}\text{plim}\left[\frac{W_j'Z_j}{T}\right]^{-1}\left[\frac{W_j'W_j}{T}\right]\left[\frac{Z_j'W_j}{T}\right]^{-1}.$$

$$
\begin{aligned}
\hat{\delta}_{j,2SLS} &= [\hat{Z}_j'\hat{Z}_j]^{-1}\hat{Z}_j'y_j \\
&= [Z_j'X(X'X)^{-1}X'Z_j]^{-1}Z_j'X(X'X)^{-1}X'y_j.
\end{aligned}
$$

$$
\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} Z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_m \end{pmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix},
$$

3SLS:
Step 1: regress Z_j on X and get $\hat{Z}_j$;
Step 2: get 2SLS, $\hat{\delta}_{j,2sls} = [\hat{Z}_j'Z_j]^{-1} \times [\hat{Z}_j'y_j]$, get $\hat{\varepsilon}_j$, estimate $\hat{\sigma}_{ij} = \frac{1}{T}\hat{\varepsilon}_i'\hat{\varepsilon}_j$,
hence an estimate of $\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix}$;
Step 3, feasible GLS, $\delta_{3sls} = [\hat{\mathbf{Z}}(\hat{\Sigma}^{-1} \otimes I_T)\mathbf{Z}]^{-1} \times [\hat{\mathbf{Z}}(\hat{\Sigma}^{-1} \otimes I_T)\mathbf{y}]$.

7.2 Hausman's Specification Test

1. $\hat{\theta}$ is asymptotically efficient under H_0 but is inconsistent under H_1,
2. $\bar{\theta}$ is asymptotically inefficient under H_0 but is consistent under both H_0 and H_1.

$$
(\hat{\theta} - \bar{\theta})'\hat{V}^{-1}(\hat{\theta} - \bar{\theta}) \xrightarrow{d} \chi^2(k).
$$

$$
\mathcal{HS} = \frac{1}{\hat{\sigma}^2}\{(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})'[(X'Z_1(Z_1'Z_1)^{-1}Z_1'X)^{-1} - (X'X)^{-1}]^{-}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})\}
$$

8 Generalized Method of Moments

- *Orthogonality condition*

$$
Ef(x, \beta_0) = 0
$$

$$
g_T(x, \beta) = \frac{1}{T}\sum_{t=1}^{T} f(x_t, \beta)
$$

$$
\hat{\beta} = \arg.\min_{\beta} g_T'(x, \beta)a_T'a_T g_T(x, \beta).
$$

$$
\frac{\partial g_T'(\hat{\beta}_T)}{\partial \beta}a_T'a_T g_T(\hat{\beta}_T) = 0.
$$

Let a_T^* denote $\frac{\partial g_T'(\omega)}{\partial \beta}a_T'a_T$.

$$
0 = a_T^* g_T(\hat{\beta}_T) = a_T^* g_T(\beta_0) + a_T^* \frac{\partial g_T(\bar{\beta})}{\partial \beta}(\hat{\beta}_T - \beta_0),
$$

which implies that

$$
\begin{aligned}
\sqrt{T}(\hat{\beta}_T - \beta_0) &= -\left(a_T^* \frac{\partial g_T(\bar{\beta})}{\partial \beta}\right)^{-1} \sqrt{T}a_T^* g_T(\beta_0) \\
&= -\left(a_T^* \frac{\partial g_T(\bar{\beta})}{\partial \beta}\right)^{-1} a_T^* \frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(x_t, \beta_0).
\end{aligned}
$$

It follows that $\sqrt{T}(\hat{\beta}_T - \beta_0)$ converges in distribution to a normally distributed random vector with mean zero and covariance matrix

$$(d_0' a_0' a_0 d_0)^{-1} d_0' a_0' a_0 S_w a_0' a_0 d_0 (d_0' a_0' a_0 d_0)^{-1},$$

where $S_w \equiv \lim_{T\to\infty} var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T f(x_t,\beta_0)\right) = \sum_{t=-\infty}^{\infty} E[(f(x_t,\beta_0)f(x_t,\beta_0)]$, $a_0 = p\lim a_T$ and $d_0 = p\lim \frac{\partial g_T(\beta_0)}{\partial\beta}$.

Optimal a_0, the GMM method corresponds to

$$\min_\beta g_T'(\beta)\hat{S}_w^{-1} g_T(\beta),$$

9 Panel Data Models

9.1 Fixed Effects Model

$$y_{it} = \alpha_i + x_{it}\beta + \epsilon_{it}, \qquad i = 1,\cdots,n; t = 1,\cdots,T,$$

$$\min_{\alpha_i,\beta} Q(\alpha_1,\cdots,\alpha_n,\beta) = \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \alpha_i - x_{it}\beta)^2$$

$$\begin{aligned} \sum_{t=1}^T (y_{it} - \hat{\alpha}_i - x_{it}\hat{\beta}) &= 0, \;\; i = 1,\cdots,n; \\ \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{\alpha}_i - x_{it}\hat{\beta})x_{it}' &= 0, \end{aligned}$$

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i\hat{\beta}, \text{ and } \hat{\beta} = [\sum_{i=1}^n \sum_{t=1}^T (x_{it} - \bar{x}_i)'(x_{it} - \bar{x}_i)]^{-1} \sum_{i=1}^n \sum_{t=1}^T (x_{it} - \bar{x}_i)'(y_{it} - \bar{y}_i). \tag{9.1}$$

9.2 Random Effects Model (Error Component Model) (or Variance Component Model)

$$y_{it} = x_{it}\beta + u_{it}, \qquad u_{it} = \alpha_i + \epsilon_{it}$$

$$E(u_i u_i') = E((\alpha_i l_T + \epsilon_i)(\alpha_i l_T + \epsilon_i)') = \sigma_\alpha^2 l_T l_T' + \sigma_\epsilon^2 I_T = \sigma_u^2 A,$$

where $\sigma_u^2 = \sigma_\alpha^2 + \sigma_\epsilon^2$, and

$$A = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} = (1-\rho)I_T + \rho l_T l_T'$$

with $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\epsilon^2)$.

9.3 Dynamic Panel Data

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + v_{it}, \; t = 1,\cdots,T.$$

$$\hat{\gamma} = [\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\bar{y}_{i,-1})'(y_{i,t-1}-\bar{y}_{i,-1})]^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\bar{y}_{i,-1})'(y_{it}-\bar{y}_i) \quad (9.2)$$

$$= \gamma + [\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\bar{y}_{i,-1})'(y_{i,t-1}-\bar{y}_{i,-1})]^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\bar{y}_{i,-1})'(v_{it}-\bar{v}_i) \quad (9.3)$$

As $y_{i,t-1}-\bar{y}_{i,-1}$ and $v_{it}-\bar{v}_i$ are correlated (because $\bar{y}_{i,-1}$ and $\bar{v}_i$ are correlated), $\hat{\gamma}$ is biased. Unless T is large, $\hat{\gamma}$ is inconsistent.

$$\sum_{t=1}^{T}(y_{i,t-1}-\bar{y}_{i,-1})'(v_{it}-\bar{v}_i) = \sum_{t=1}^{T} y_{i,t-1}v_{it} - T\cdot\bar{y}_{i,-1}\bar{v}_i$$

• We make the first difference such that

$$\Delta y_{it} = \gamma\Delta y_{i,t-1} + \Delta v_{it},\ t = 2,\cdots,T.$$

We can use the lagged value of y_{is} where $s \leq t-2$ as the IVs.

$$E(y_{i,t-s}\Delta v_{it}) = 0,\ \text{ for } t = 2,\cdots,T \text{ and } s \geq 2.$$

Denote $(T-1)\times m$ matrix $Z_i = \begin{bmatrix} y_0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_0 & y_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & y_0 & \cdots & y_{T-2} \end{bmatrix}$ and $T-1$ vector $\Delta u_i = (\Delta v_{i2}, \Delta v_{i3}, \cdots, \Delta v_{iT})'$.

$$g_i = Z_i'\Delta u_i, \text{ so that } g = Z'u$$

we have GMM estimator γ as

$$\hat{\gamma}_{dif} = (y_{-1}'ZA_nZ'y_{-1})^{-1}(y_{-1}'ZA_nZ'y)$$

$$A_n = \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\hat{u}_i\hat{u}_i'Z_i\right)^{-1}$$

10 Spatial Econometrics

$$Y_n = \lambda W_nY_n + X_n\beta + \epsilon_n$$

1): MLE

$$\ln L_n(\lambda,\beta,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}(S_n(\lambda)Y_n - X_n\beta)'(S_n(\lambda)Y_n - X_n\beta),$$

2): **Method of Moments** (MOM) for SAR process:

$$\min_{\theta} g_n'(\theta)g_n(\theta).$$

The moment equations are based on three moments:

$$E(\epsilon_n'\epsilon_n) = n\sigma^2, \qquad E(\epsilon_n'W_n'W_n\epsilon_n) = \sigma^2 tr(W_n'W_n), \qquad E(\epsilon_n'W_n\epsilon_n) = 0.$$

These correspond to

$$g_n(\theta) = (Y_n' S_n'(\lambda) S_n(\lambda) Y_n - n\sigma^2,\ Y_n' S_n'(\lambda) W_n' W_n S_n(\lambda) Y_n - \sigma^2 tr(W_n' W_n),\ Y_n' S_n'(\lambda) W_n S_n(\lambda) Y_n)'.$$

3): **2SLS Estimation** for the MRSAR model

Denote $Z_n = (W_n Y_n, X_n)$ and $\delta = (\lambda, \beta')'$.

$$\hat{\delta}_{n,2sls} = \{Z_n' H_n (H_n' H_n)^{-1} H_n' Z_n\}^{-1} \times Z_n' H_n (H_n' H_n)^{-1} H_n' Y_n.$$

4) **Moments for GMM estimation**

Now consider constant $n \times n$ matrix $P_{1n}, \cdots, P_{mn}$ each with a zero diagonal. The moment functions $(P_{jn}\epsilon_n(\theta))'\epsilon_n(\theta)$ can be used in addition to $Q_n'\epsilon_n(\theta)$. These moment functions form a vector

$$g_n(\theta) = (P_{1n}\epsilon_n(\theta), \cdots, P_{mn}\epsilon_n(\theta), Q_n)'\epsilon_n(\theta) = \begin{pmatrix} \epsilon_n'(\theta) P_{1n} \epsilon_n(\theta) \\ \vdots \\ \epsilon_n'(\theta) P_{mn} \epsilon_n(\theta) \\ Q_n' \epsilon_n(\theta) \end{pmatrix}$$