Stationary:

SS: $f(X_{t_1},\cdots,X_{t_r}) = f(X_{t_1+\tau},\cdots,X_{t_r+\tau})$

Non stationary

① trend stationary $y_t = \alpha+\beta t+\varepsilon_t$ → Stationary

HAC $(\hat{\alpha},\hat{\beta})$

② AR: $y_t = \alpha + y_{t-1} + \varepsilon_t$

1st order difference: $\hat{\alpha}$

WS: $EX_t = \mu < +\infty$.

$Var(X_t) = \sigma^2 < +\infty$

$Cov(X_t, X_{t-j}) = r_j$ Autocovariance

White noise: $EX_t = 0$. $Var(X_t) = \sigma^2$.

$Cov(X_t, X_{t-j}) = 0 \quad \forall j \neq 0$

Element S $\not\Rightarrow$ Vector S: $\begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ \varepsilon_t \end{bmatrix}$ $\varepsilon_t \overset{iid}{\sim} N(0,1)$

SS $\Rightarrow$ WS if $E(\cdot)$, $Var(\cdot) < +\infty$

Ergodic: $\lim_{n\to+\infty} |Ef(X_t,\cdots,X_{t+k})g(X_{t+n},\cdots,X_{t+l+n})|$

$\Downarrow \; = |Ef(X_t,\cdots,X_{t+k})||Eg(X_{t+n},\cdots,X_{t+l+n})|$

or $\lim_{j\to+\infty} r_j = \lim_{j\to+\infty} Cov(X_t, X_{t-j}) = 0$

Ergodic thm: $\{X_t\}_{t=1}^T$ S.E. $E(X_t) = \mu < +\infty$

$\forall f$ measurable: $\frac{1}{T}\sum_{t=1}^T f(X_t) \xrightarrow{a.s.} f(\mu) \quad (T\to+\infty)$

S.E. process: $Z_t = C + \rho Z_{t-1} + \varepsilon_t \quad |\rho|<1 \quad \varepsilon_t$ IWN

Martingale: $E(X_t|X_{t-1},\cdots) = X_{t-1}$

Random Walk $Z_t = Z_{t-1} + \varepsilon_t$. ε_t IWN

$\{X_t\}$ is martingale $\not\Rightarrow$ $\{f(X_t)\}$ is martingale

Martingale difference sequence: $E(\varepsilon_t|\varepsilon_{t-1},\cdots) = 0$

ARCH Autoregressive conditional heteroskedastic

$\{\varepsilon_t\}_t$: $\varepsilon_t = \sqrt{\sigma_t^2}\, v_t \quad v_t \overset{iid}{\sim} (0,1)$

$\sigma_t^2 = C + \alpha\varepsilon_{t-1}^2 \quad C>0$

$E(\varepsilon_t|\varepsilon_{t-1},\cdots) = 0$

$Var(\varepsilon_t|\varepsilon_{t-1},\cdots) = C + \alpha\varepsilon_{t-1}^2$ (ARCH)

$|\alpha|<1 \Rightarrow$ ARCH is SS & Er

IWN $\overset{①}{\not\Leftarrow}$ stationary mds with finite Var $\overset{②}{\not\Leftarrow}$ WN

① ARCH

② $Z_t = \cos(wt)$ $w\sim unif[0,2\pi]$: $EZ_t = 0$ $Var(Z_t) = \frac{1}{2}$ $Cov(Z_t,Z_s)=0$

CLT: $\{\varepsilon_t\}$ mds, S&E. $E(\varepsilon_t\varepsilon_t^T) = \Sigma < +\infty$

$\sqrt{T}\frac{1}{T}\sum_{t=1}^T \varepsilon_t \xrightarrow{d.} N(0,\Sigma)$

OLS:

A1: Linear DGP: $y_i = X_i^T\beta + \varepsilon_i$

A2: Ergodic stationarity: $\{y_i, X_i\}$ jointly S.E

A3: Moment: $Eg_i = E(X_i(y_i - X_i\beta)) = 0$

A4: Rank(Identification): $rank(E(X_iX_i^T) = \Sigma_{XX}) = K$

A5: $\{g_i\}$ mds with finite Var: $S := Avar(\bar{g}) > 0$

Sufficient condition: $E(\varepsilon_i|\varepsilon_{i-1},\cdots,X_{i-1},\cdots) = 0$

$\hat{\beta}_{OLS} = (\frac{1}{n}\sum_{i=1}^n X_iX_i^T)^{-1}(\frac{1}{n}\sum_{i=1}^n X_iy_i) = \beta + (\frac{1}{n}\sum_{i=1}^n X_iX_i^T)^{-1}\bar{g}_n$

$\Rightarrow \sqrt{n}(\hat{\beta}-\beta) = S_{XX}^{-1}\sqrt{n}\,\bar{g}_n \xrightarrow{d.} N(0, S_{XX}^{-1}SS_{XX}^{-1})$

$s^2 := \frac{1}{n-k}\sum_{i=1}^n \hat{\varepsilon}_i^2$, unbiased $E(s^2) = E(\varepsilon_i^2)$

$H_0: \beta_K = \bar{\beta}_K$.

t statistic: $\hat{t} = \frac{\sqrt{n}(\hat{\beta}_K - \bar{\beta}_K)}{\sqrt{\widehat{Avar}(\hat{\beta})_{KK}}} = \frac{\hat{\beta}_K - \bar{\beta}_K}{\widehat{SE}(\hat{\beta}_K)} \xrightarrow{d.} N(0,1)$

significance level α. critical value $t_{\frac{\alpha}{2}}$

asymptotic size = Type I = α ; power = 1 − Type II

Alternative DGP: $H_1: \beta_K = \bar{\beta}_K + C_0$

Global Power $= P(|\hat{t}| > t_{\frac{\alpha}{2}}|H_1) \to 1 \; (n\to+\infty)$ Consistent!

Local alternatives: $H_1: \beta_K = \bar{\beta}_K + \frac{\gamma}{\sqrt{n}}$ Pitman drift

Local Asymptotic power $= P(|X| > t_{\frac{\alpha}{2}})$, $X\sim N(\frac{\gamma}{\sqrt{Avar(\hat{\beta}_K)}}, 1)$

$H_0: R\beta = r$ or $g(\beta) = r$. $\hat{R} := \frac{\partial g}{\partial\beta}(\hat{\beta})$

Wald: $\hat{W} = n(R\hat{\beta}-r)^T(R\widehat{Avar}(\hat{\beta})R^T)^{-1}(R\hat{\beta}-r) \xrightarrow{d.} \chi^2_r$

Consistent test: Power $\to 1 \; (n\to+\infty) \quad \forall\alpha$

A test is consistent if $\forall\beta\neq\beta_0(H_0)$. Power $\to 1 \; (n\to+\infty) \; \forall\alpha$

Endogenous Bias:

1. Simultaneity bias

$\begin{cases} q_i = \alpha_0 + \alpha_1 p_i + u_i & \text{Demand} \\ q_i = \beta_0 + \beta_1 p_i + v_i & \text{Supply} \end{cases}$ $\quad Eu_i = Ev_i = 0$ $\quad Cov(u_i, v_i) = 0$

$\Rightarrow \begin{cases} p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1} \\ q_i = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 v_i - \beta_1 u_i}{\alpha_1 - \beta_1} \end{cases}$ $\quad Cov(p_i, u_i) = \frac{-Var(u_i)}{\alpha_1 - \beta_1}$ $\quad Cov(p_i, v_i) = \frac{Var(v_i)}{\alpha_1 - \beta_1}$

OLS $q_i \sim p_i$:

$\hat{b}_{OLS} = \frac{Cov(p_i, q_i)}{Var(p_i)} = \alpha_1 + \frac{Cov(p_i, u_i)}{Var(p_i)} = \beta_1 + \frac{Cov(p_i, v_i)}{Var(p_i)}$

$q_i = \beta_0 + \beta_1 p_i + \beta_2 x_i + \zeta_i$ Supply shifter. $Cov(x_i, \zeta_i) = 0$

$\Rightarrow \begin{cases} p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1}x_i + \frac{\zeta_i - u_i}{\alpha_1 - \beta_1} \\ q_i = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}x_i + \frac{\alpha_1\zeta_i - \beta_1 u_i}{\alpha_1 - \beta_1} \end{cases}$ $\quad Cov(p_i, x_i) \neq 0$

IV: $\alpha_1 = \frac{Cov(x_i, q_i)}{Cov(x_i, p_i)}$

2SLS: ① $p_i \sim x_i$ get $\hat{p}_i = Ep_i + \frac{Cov(x_i, p_i)}{Var(x_i)}(x_i - Ex_i)$

② $q_i \sim \hat{p}_i$: $q_i = \alpha_0 + \alpha_1\hat{p}_i + [u_i + \alpha_1(p_i - \hat{p}_i)]$

$\alpha_1 = \frac{Cov(q_i, \hat{p}_i)}{Var(\hat{p}_i)} = \frac{Cov(x_i, q_i)}{Cov(x_i, p_i)}$

2. Measurement errors

$C_i^* = kY_i^*$, $\begin{cases} C_i = C_i^* + c_i & Ec_i = 0 & c_i \perp\!\!\!\perp y_i \\ Y_i = Y_i^* + y_i & Ey_i = 0 & (c_i, y_i) \perp\!\!\!\perp (C_i^*, Y_i^*) \end{cases}$

$\Rightarrow C_i = kY_i + u_i$. $u_i = c_i - ky_i$ $\quad E(Y_iu_i) = -kE(y_i^2)$

IV: $x_i = 1$ $\quad Ex_iu_i = 0$. $Ex_iY_i = EY_i^* \neq 0$

$k = \frac{Ex_iC_i}{Ex_iY_i} = \frac{EC_i}{EY_i}$

3. Variable chosen by the agent taking into account error component unobservable to the econometrician.

$\ln Q_i = \ln A_i + \phi_1\ln L_i + v_i$. (A_i, v_i) unobservable.

$= E\ln A_i + \phi_1\ln L_i + v_i + u_i$, $u_i := \ln A_i - E\ln A_i$

$L_i = \underset{L}{\arg\max}\; pA_iL^{\phi_1}B - wL$, $B = E\exp\{v_i\}$.

$\Rightarrow \ln L_i = \frac{1}{\phi_1 - 1}[\ln\frac{w}{p} - \ln(B\phi_1) - E\ln A_i] + \frac{1}{1-\phi_1}u_i$

GMM (linear model)

A1: Linear DGP: $y_i = x_i^T \delta + \varepsilon_i$ unique & nonconstant elements

A2: Ergodic stationarity: $\{y_i, x_i, z_i\} \to \{w_i\}$ jointly S.E

$LW_i = \delta_1 + \delta_2 S_i + \delta_3 EXPR_i + \delta_4 IQ_i + \varepsilon_i$

$y_i = LW_i$, $x_i = [1\ S_i\ EXPR_i\ IQ_i]^T$ predetermined vs. endogenous

$z_i = [1\ S_i\ EXPR_i\ AGE_i\ MED_i]^T$

$w_i = [LW_i\ S_i\ EXPR_i\ IQ_i\ AGE_i\ MED_i]^T$

A3: Moment: $E g_i = E(z_i(y_i - x_i^T\beta)) = 0$

A4: Identification: $\mathrm{rank}(E z_i x_i^T = \Sigma_{zx}) = k$

$q_i = \alpha_0 + \alpha_1 p_i + u_i$

$y_i = q_i$, $x_i = [1\ p_i]^T$ $z_i = [1\ x_i]^T$

$E z_i x_i^T = \begin{bmatrix} 1 & Ep_i \\ Ex_i & Ex_i p_i \end{bmatrix}$ full column rank $\Leftrightarrow Cov(x_i, p_i) \neq 0$

δ is identified if $\tilde{\delta} = \delta$ is the only solution to

$E g(w_i, \tilde{\delta}) = 0$

$\Rightarrow$ order condition $K \le L$

A5: $\{g_i\}$ mds with 2nd moment $< +\infty$ $S := Avar(\bar{g}) > 0$

Sufficient condition: $E(\varepsilon_i \mid \varepsilon_{i-1}, \cdots, z_{i-1}, \cdots) = 0$

$\hat{\delta}(\hat{W}) = \underset{\delta}{\arg\min}\ J(\delta, \hat{W})$, $J(\delta, \hat{W}) = n\, g_n(\delta)^T \hat{W} g_n(\delta)$

Linear model: $g_n(\delta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\delta) = S_{zy} - S_{zx}\delta$

FOC $\Rightarrow \hat{\delta}(\hat{W}) = (S_{zx}^T \hat{W} S_{zx})^{-1} S_{zx}^T \hat{W} S_{zy}$

$= \delta + (S_{zx}^T \hat{W} S_{zx})^{-1} S_{zx}^T \hat{W} g_n(\delta)$

Consistency: $\hat{\delta}(\hat{W}) - \delta \xrightarrow{p} 0$

Asymptotic normality: $\sqrt{n}(\hat{\delta}(\hat{W}) - \delta) \xrightarrow{d} N(0, Avar(\hat{\delta}(\hat{W})))$

$Avar(\hat{\delta}(\hat{W})) = (\Sigma_{zx}^T W \Sigma_{zx})^{-1} \Sigma_{zx}^T W S W \Sigma_{zx} (\Sigma_{zx}^T W \Sigma_{zx})^{-1}$

$S = Avar(g_n(\delta)) \overset{\text{mds } g_i}{=} E g_i g_i^T$: $\hat{S} = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 z_i z_i^T \xrightarrow{p} S$

Efficient GMM: $\hat{\delta}(\hat{W})$ s.t. $\hat{W} \xrightarrow{p} S^{-1}$

Hausman Equality:

$Avar(\hat{\delta}(\hat{W}) - \hat{\delta}(\hat{S}^{-1})) = Avar(\hat{\delta}(\hat{W})) - Avar(\hat{\delta}(\hat{S}^{-1})) \geq 0$

1. 2-step ① $\hat{W} = S_{zz}^{-1} \xrightarrow{GMM} \hat{\delta}(S_{zz}^{-1})$. $\hat{\varepsilon}_i = y_i - x_i^T\hat{\delta}$, $\hat{S}$
   ② $\hat{W} = \hat{S}^{-1}$
2. Iteration: $\|\hat{\delta}(\hat{S}_p^{-1}) - \hat{\delta}(\hat{S}_{p+1}^{-1})\| < \varepsilon$
3. Continuous updating:
   $\hat{\delta} = \underset{\delta}{\arg\min}\ n\, g_n^T(\delta) (\frac{1}{n}\sum_{i=1}^{n} g_i^*(\delta) g_i^*(\delta)^T)^{-1} g_n(\delta)$
   $g_i^*(\delta) = g_i(\delta) - \frac{1}{n}\sum_{i=1}^{n} g_i(\delta)$

Global power: power $\to 1\ (n \to +\infty)$ $\forall \hat{W}$

Local power: $Avar(\hat{\delta}) \downarrow$ Apower $\uparrow$ maximized at $\hat{S}^{-1}$

Overidentification Test: $H_0: E g_i(\delta_0) = 0$ $L > K$

J statistic: $J(\hat{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) = n\, g_n(\hat{\delta})^T \hat{S}^{-1} g_n(\hat{\delta}) \xrightarrow{d} \chi^2_{L-K}$

$L = K \Leftrightarrow J(\hat{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) = 0$

$g_n(\hat{\delta}) = \hat{B} g_n(\delta_0)$, $\hat{B} = I_L - S_{zx}(S_{zx}^T \hat{S}^{-1} S_{zx})^{-1} S_{zx}^T \hat{S}^{-1}$

$\mathrm{rank}(\hat{B}) = L - K$

Specification test: $\hat{J}$ is large : any A1-A5 may fail

Testing Subsets of Orthogonality Conditions

$H_0: E g_{i2}(\delta_0) = 0$ assuming $E g_{i1}(\delta_0) = 0$ $L_1 > K$

C statistic: $C = J - J_1 \xrightarrow{d} \chi^2_{L_2}$ $g_{n1} = R g_n$

$g_n(\hat{\delta}) = B g_n$ $g_n(\hat{\delta}_1) = RCR g_n$

$J = n g_n^T \hat{S}^{-1} B g_n$ $J_1 = n g_n^T R^T R^T S_{11}^{-1} RCR g_n$

$J - J_1 = n g_n^T A g_n$, $\mathrm{rank}(A) = tr(A) = tr - tr = (L-K) - (L_1 - K)$

Likelihood-Ratio Principle. $H_0: a(\delta) = 0$

$LR = J(\bar{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) - J(\hat{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) \xrightarrow{d} \chi^2_{\#a}$

$\bar{\delta}(\hat{S}^{-1}) = \underset{\delta}{\arg\min}\ J(\delta, \hat{S}^{-1})$ s.t. $a(\delta) = 0$

$\hat{\delta}(\hat{S}^{-1}) = \underset{\delta}{\arg\min}\ J(\delta, \hat{S}^{-1})$

$W = n\, a(\hat{\delta})^T [A(\hat{\delta}) Avar(\hat{\delta}) A(\hat{\delta})^T]^{-1} a(\hat{\delta}) \xrightarrow{d} \chi^2_{\#a}$

$LR - W \xrightarrow{p} 0 \mid H_0$, For linear $a(\cdot)$: $LR = W$

GMM (non-linear model)

$\hat{\delta}(\hat{W}) = \underset{\delta}{\arg\min}\ J(\delta, \hat{W}) = n\, g_n(\delta)^T \hat{W} g_n(\delta)$

FOC: $G^T(\hat{\delta}) \hat{W} g_n(\hat{\delta}) = 0$ $G(\delta) = \frac{\partial g_n}{\partial \delta}(\delta)$

$G^T(\hat{\delta}) \hat{W} \{g_n(\delta_0) + G(\delta_0)(\hat{\delta} - \delta_0)\} = 0$ 1st Taylor

$\Rightarrow \hat{\delta}(\hat{W}) - \delta_0 = -(R^T \hat{W} R)^{-1} R^T \hat{W} g_n(\delta_0)$

Consistency: $\hat{\delta}(\hat{W}) \xrightarrow{p} \delta_0$

$|J(\hat{\delta}, W) - J(\delta_0, W)|$

$\le |J(\hat{\delta}, W) - J(\hat{\delta}, \hat{W})| + \underbrace{|J(\hat{\delta}, \hat{W}) - J(\delta_0, W)|}$

$\le \max\{|J(\delta_0, \hat{W}) - J(\delta_0, W)|, |J(\hat{\delta}, \hat{W}) - J(\hat{\delta}, W)|\}$

$J(\delta, \hat{W}) \xrightarrow{p} J(\delta, W)$ uniformly $\forall \delta$

Asymptotic normality: $\sqrt{n}(\hat{\delta}(\hat{W}) - \delta_0) \xrightarrow{d} N(0, Avar(\hat{\delta}))$

$Avar(\hat{\delta}) = (R^T W R)^{-1} R^T W S W R (R^T W R)^{-1}$

Conditional Moment $E(\rho_i(\delta) \mid z_i) = 0$

Generalized IV: $F(z_i)$ Optimal $F_i^0 = -R_i^T \Sigma_i^{-1}$

$R_i = E[\frac{\partial \rho_i}{\partial \delta}(\delta) \mid z_i)$

$\Sigma_i = E(\rho_i(\delta) \rho_i^T(\delta) \mid z_i)$

$E(g_i g_i^T) = E(R_i^T \Sigma_i^{-1} \rho_i \rho_i^T \Sigma_i^{-1} R_i)$

$= E(R_i^T \Sigma_i^{-1} \Sigma_i \Sigma_i^{-1} R_i) = E(R_i^T \Sigma_i^{-1} R_i)$

To prove $n X^T A^{-1} X \xrightarrow{d} \chi^2_{\# X}$:

① $\sqrt{n} A^{-\frac{1}{2}} X \xrightarrow{d} N(0, \Pi)$

② $\Pi^T = \Pi$, $\Pi^2 = \Pi$ $\mathrm{rank}(\Pi) = tr(\Pi) = \# X$

Take derivatives on matrix:

Chain rule: $\frac{d g_n^T}{d\theta}_{K\times 1} = \frac{d g_n^T}{d\theta}_{K\times L} \frac{d g_n}{d g_n}_{L\times 1}$

Total derivative: $d(X^T A X) = tr(d(X^T A X))$

$= tr(dX^T A X + X^T A dX) = X^T A^T dX + X^T A dX$

$\Rightarrow \frac{d(X^T A X)}{dX^T} = X^T A^T + X^T A$

$\frac{d(X^T A X)}{dX} = AX + A^T X$

Moment Selection

$G = \{c \in \mathbb{R}^{L_{max}} : c_j = 1\{j\text{-th condition is included}\}, |c| \geq k\}$

$\hat{c} = \arg\min_{c \in G} J_n(c) - h(|c|)k_n$. $\quad \hat c \xrightarrow{p} c_0$

AR1: $y_t = c + \phi y_{t-1} + \varepsilon_t = x_t^T \beta + \varepsilon_t$.

OLS estimator of AR(p) is consistent & asymptotically N.

ARp: $y_t = c + \sum_{\tau=1}^{p} \phi_\tau y_{t-\tau} + \varepsilon_t$ stationary.

$\{\varepsilon_t\} \sim IWN(0,\sigma^2)$

$y_t = x_t^T\beta + \varepsilon_t$. $\quad \hat\beta = (\sum_t x_t x_t^T)^{-1}(\sum_t x_t y_t)$

$Avar(\hat\beta) = (E x_t x_t^T)^{-1}\sigma^2$

$\widehat{Avar}(\hat\beta) = (\frac1n \sum_{t=1}^n x_t x_t^T)^{-1} s^2$, $\quad s^2 = \frac{1}{n-p-1}\sum_{t=1}^n \hat\varepsilon_t^2$

BIC: Bayesian Information Criteria

$\hat p_{BIC} = \arg\min_{p \le p_{max}} \ln \frac{SSR_p}{n} + h(p) g(n) \qquad \hat p_{BIC} \xrightarrow{p} p_0$

Gordin CLT: ① $E(y_t y_t^T) < +\infty$

$Ey_t = 0 \Leftarrow$ ② $E(y_t | y_{t-j}, y_{t-j-1}, \dots) \xrightarrow{m.s.} 0 \quad (j \to +\infty)$

$y_t = \sum_{j=0}^{+\infty} r_{tj} \Leftarrow$ ③ $\sum_{j=0}^{+\infty} (E\, r_{tj}^T r_{tj})^{\frac12} < +\infty$

$r_{tj} := E(y_t|y_{t-j},\dots) - E(y_t|y_{t-j-1},\dots)$

Then $E(y_t)=0 \quad \sum_{j=-\infty}^{+\infty} |\Gamma_j| < +\infty$. $\quad \Gamma_j := E(y_t y_{t-j}^T)$

$\sqrt n (\frac1n \sum_{t=1}^n y_t) \xrightarrow{d} N(0, \sum_{j=-\infty}^{+\infty}\Gamma_j)$

$Var(\sqrt n \bar y) = \frac1n Var(\sum_{t=1}^n y_t) = \frac1n \sum_{j=1-n}^{n-1}(n-|j|)\Gamma_j$

AR(1): $y_t = \phi y_{t-1} + \varepsilon_t$, $|\phi|<1$, $\varepsilon_t \sim IWN(0,\sigma^2)$

Gordin: ① $E y_t^2 = Var(y_t) = \frac{\sigma^2}{1-\phi} < +\infty$

② $E(y_t|y_{t-j},\dots) = \phi^j y_{t-j} \xrightarrow{m.s.} 0 \quad (j\to+\infty)$

③ $r_{tj} = \phi^j y_{t-j} - \phi^{j+1} y_{t-j-1} = \phi^j \varepsilon_{t-j}$

$\sum_{j=0}^{+\infty} E(r_{tj}^2)^{\frac12} = \sum_{j=0}^{+\infty}|\phi|^j \sigma = \frac{\sigma}{1-|\phi|} < +\infty$

Thus $\sqrt n\, \bar y \xrightarrow{d} N(0, \sum_{j=-\infty}^{+\infty}\gamma_j)$

GMM: A5' $\{g_i\}$ satisfies Gordin's condition. S invertible

How to estimate $S = Avar(\bar g_n) = \sum_{j=-\infty}^{+\infty}\Gamma_j = \Gamma_0 + \sum_{j=1}^{+\infty}(\Gamma_j + \Gamma_j^T)$

1. Parametric approach: $g_t = \alpha g_{t-1} + \varepsilon_t$. $\quad \varepsilon_t \sim IWN(0,\sigma^2)$

OLS: $\hat\alpha$; Plug in: $\hat\Gamma_0 = \frac{\sigma^2}{1-\hat\alpha^2}$. $\quad \hat\Gamma_j = \hat\alpha^j \frac{\sigma^2}{1-\hat\alpha^2}$

$\hat S = \sum_{j=-\infty}^{+\infty}\hat\Gamma_j \xrightarrow{p} S$

2. Nonparametric: $\hat\Gamma_j = \frac1T \sum_{t=1}^T g_t g_{t-j}^T \xrightarrow{p} \Gamma_j \quad \forall j$

2.1: $\hat S = \sum_{j=1-n}^{n-1}\hat\Gamma_j \not\xrightarrow{p} S$: $\quad Avar(\sum_{j=1-n}^{n-1}\hat\Gamma_j - S) \to +\infty \quad (n\to+\infty)$

2.2: HAC: Heterogeneity autocovariance

$\hat S^{HAC} = \sum_{j=1-n}^{n-1} k(\frac jp)\hat\Gamma_j \qquad$ k: kernel $\quad$ p: bandwidth

① K: Uniform K: $K(u) = 1\{|u|\le 1\}$

$\Rightarrow$ p-lag estimator: $\hat S^{HAC} = \sum_{j=-p}^{p}\hat\Gamma_j$ could be n.d.

Bartlette K: $K(u) = (1-|u|)1\{|u|\le1\}$

$\Rightarrow$ Newey-West estimator $\hat S^{HAC}_{NW}$ is p.s.d.

② p: $\hat p = \arg\min_p MSE = (E\hat S - S)^2 + Var(\hat S) = O(p^{-2l} + \frac pn)$

l: order of k: $l = \inf_l \{\lim_{u\to0} \frac{k(u)-k(0)}{|u|^l} < +\infty\}$

Rule of thumb: $p \propto n^{1/(2l+1)}$

Plug in method: $p_p = C n^{1/(2l+1)} \qquad p_p \xrightarrow{p} p_0$

2 stage: ①: $\forall C$, get p_p, $\Rightarrow \hat S \Rightarrow \hat S^{(2)} = \sum_{j=-p_p}^{p_p} j^2 \Gamma_j$

② $C = 1.32\, \hat S^{(2)}/\hat S$, get p_p

Leave-1-out CV: $\hat S_{(i)} = \sum_{j=1-n,\, j\neq i}^{n-1} k(\frac jp)\hat\Gamma_j$

$p_{CV} = \arg\min_p \sum_{i=1}^n (\hat S - \hat S_{(i)})^2 \not\xrightarrow{p} p_0$

Panel Data Model $\{y_{it}, x_{it}\}$ $i\le n.\ t\le T$

$y_{it} = x_{it}^T \delta + \varepsilon_{it}$ iid across i.

1. Pooling. $E(x_{it}\varepsilon_{it})=0$ contempory exogeneity

OLS: $\hat\delta = (\sum_{i=1}^n\sum_{t=1}^T x_{it}x_{it}^T)^{-1}(\sum_{i=1}^n\sum_{t=1}^T x_{it}y_{it})$

$= \delta + (\sum_{i=1}^n\sum_{t=1}^T x_{it}x_{it}^T)^{-1}(\sum_{i=1}^n\sum_{t=1}^T x_{it}\varepsilon_{it})$

$\hat\delta_{OLS}-\delta \xrightarrow{P} (\sum_{t=1}^T E x_{it}x_{it}^T)^{-1}(\sum_{t=1}^T E(x_{it}\varepsilon_{it}))=0$

$\sqrt n(\hat\delta_{OLS}-\delta)\xrightarrow{d}(\sum_{t=1}^T E x_{it}x_{it}^T)^{-1} N(0, \sum_{t=1}^T\sum_{s=1}^T E(x_{it}x_{is}^T\varepsilon_{it}\varepsilon_{is}))$

$Avar(\hat\delta_{OLS}) = (E X_i^TX_i)^{-1}\sum_{t=1}^T\sum_{s=1}^T E(x_{it}x_{is}^T\varepsilon_{it}\varepsilon_{is})(EX_i^TX_i)^{-1}$

$=(EX_i^TX_i)^{-1}(EX_i^T\Omega_iX_i)(EX_i^TX_i)^{-1}$

if $E(x_{it}x_{is}^T\varepsilon_{it}\varepsilon_{is})=0$ $\forall t\ne s$.

$E(\varepsilon_i\varepsilon_i^T|X_i)=\Omega_i$ (If $\Omega_i=\sigma^2 I$, BLUE)

Error component: $\varepsilon_{it}=\alpha_i+\eta_{it}$

2. Random effect: $E(x_{it}\alpha_i)=0$ $E(x_{it}\eta_{it})=0$

$\hat\delta_{RE}=\hat\delta_{GLS}=(\sum_{i=1}^n X_i^T\hat\Omega_i^{-1}X_i)^{-1}(\sum_{i=1}^n X_i^T\hat\Omega_i^{-1}y_i)$

$=\delta+(\sum_{i=1}^n X_i^T\hat\Omega_i^{-1}X_i)^{-1}(\sum_{i=1}^n X_i^T\hat\Omega_i^{-1}\varepsilon_i)$

$Avar(\hat\delta_{RE})=(EX_i^T\Omega_i^{-1}X_i)^{-1}$

$S=E(X_i^T\varepsilon_i\varepsilon_i^TX_i)$. $\hat S=\frac1n\sum_{i=1}^n X_i^T\hat\varepsilon_i\hat\varepsilon_i^TX_i$

With $E(\varepsilon_i\varepsilon_i^T|X_i)=\Omega$: $\hat\delta_{RE}=\hat\delta_{GMM}(\{\hat\Omega\otimes\frac1n\sum_{i=1}^n x_ix_i^T\}^{-1})$

$\hat\delta_{OLS}$ is not BLUE. $\hat\delta_{GLS}=\hat\delta_{RE}$ is BLUE.

3. Fixed effect $E(x_{it}\alpha_i)\ne0$ $E(x_{it}\eta_{is})=0$

① Demean: $y_{it}-\bar y_i=(x_{it}-\bar x_i)^T\delta+(\eta_{it}-\bar\eta_i)$

$E\{(x_{it}-\bar x_i)(\eta_{it}-\bar\eta_i)\}=0$

$\hat\delta_{FE}=(\sum_{i=1}^n\hat X_i^T\hat X_i)^{-1}(\sum_{i=1}^n\hat X_i^T\hat y_i)$ $\widehat{Var}_i = Var_i - \overline{Var}_i$

② Dummy: $y_{it}=x_{it}^T\delta+d_i^T\alpha+\eta_{it}$ $\alpha: n\times1,\ d_i: n\times1$

$E x_{it}\eta_{it}=0$ $E d_i\eta_{it}=0$

$\hat\delta_{DM}$: FWL: $y_{it}=d_i^T\gamma_y+e_{it}^y$. $\hat e_{it}^y=y_{it}-\bar y_i$

$x_{it}=\Gamma_x d_i+E_{it}^x$. $\hat E_{it}^x=x_{it}-\bar x_i$

$\hat e_{it}^y=\hat E_{it}^{xT}\delta+e_{it}$ $\hat\delta_{DM}=\hat\delta_{FE}$

$\sqrt n(\hat\delta_{FE}-\delta)\xrightarrow{d}(E\hat X_i^T\hat X_i)^{-1}N(0,E(\hat X_i^T\hat\eta_i\hat\eta_i^T\hat X_i))$

$Avar(\hat\delta_{FE})=\sigma_\eta^2(E\hat X_i^T\hat X_i)^{-1}$ with conditional homosk..

$\hat\sigma_\eta^2=\frac{1}{(T-1)n-K}\sum_{i=1}^n(\hat y_i-\hat X_i\hat\delta_{FE})^T(\hat y_i-\hat X_i\hat\delta_{FE})$

α (time invariant parameter) is not identified!

$\bar y_i=\bar x_i^T\hat\delta_{FE}+\hat\alpha_i$: $\hat\alpha_i-\alpha_i\xrightarrow{P}\bar\eta_i=\frac1T\sum_{t=1}^T\eta_{it}\ne0$ (Fix T)

$y_{it}=x_{it}^T\delta_1+B_i^T\delta_2+\eta_{it}$: δ_2 is not identified!

$e_{it}=y_{it}-x_{it}^T\hat\delta_{1FE}=B_i^T\delta_2+\eta_{it}$ $\begin{cases}EB_i\eta_{it}=0 & \text{reg}\\ EB_i\eta_{it}\ne0 & \text{instrument}\end{cases}$

3.1 Intercept in Fixed Effect Model

$y_{it}=\mu+x_{it}^T\delta+\alpha_i+\eta_{it}$ ① FE: $\hat y_{it}=\hat x_{it}^T\delta+\hat\eta_{it}$. $\hat\delta_{FE}$

$\hat\mu-\mu\xrightarrow{P}0$ ② Mean: $\bar y=\hat\mu+\bar x^T\hat\delta_{FE}$

3.2. 2-way fixed effect. $y_{it}=x_{it}^T\delta+\alpha_i+\lambda_t+\eta_{it}$

$E(x_{it}\alpha_i)\ne0$ $E(x_{it}\lambda_t)\ne0$ $E(x_{it}\eta_{is})=0$ $\forall t, \forall s$

Demean: $\bar y_i=\bar x_i^T\delta+\alpha_i+\bar\lambda_t+\bar\eta_i$

$\bar y_t=\bar x_t^T\delta+\bar\alpha_i+\lambda_t+\bar\eta_t$

$\bar y=\bar x^T\delta+\bar\alpha_i+\bar\lambda_t+\bar\eta$

$y_{it}-\bar y_i-\bar y_t+\bar y=(x_{it}-\bar x_i-\bar x_t+\bar x)^T\delta+(\eta_{it}-\bar\eta_i-\bar\eta_t+\bar\eta)$

Hausman Specification Test.

$H_0: E x_{it}\alpha_i=0$

	$\hat\beta_{FE}$	$\hat\beta_{RE}$
$Ex_{it}\alpha_i\ne0$	$\hat\beta\xrightarrow{P}\beta_0$	$\hat\beta\not\xrightarrow{P}\beta_0$
$Ex_{it}\alpha_i=0$	$\hat\beta\xrightarrow{P}\beta_0$ Not Efficient	$\hat\beta\xrightarrow{P}\beta_0$ Efficient

Hausman statistic:

$H=n(\hat\beta_{FE}-\hat\beta_{RE})^T(\widehat{Avar}(\hat\beta_{FE})-\widehat{Avar}(\hat\beta_{RE}))^{-1}(\hat\beta_{FE}-\hat\beta_{RE})\xrightarrow{d}\chi^2_{\#\beta}$

If $\widehat{Avar}$ is singular. psudo inverse. Hausman equaility: $Cov(\hat\theta_{eff}, \hat\theta-\hat\theta_{eff})=0$

Unbalanced Panel (Zeroing out)

Selection variable: $d_{it}=1\{$observation t exists in sample $i\}$

Fixed Effect: $y_{it}=x_{it}^T\delta+\alpha_i+\eta_{it}$.

$E(x_{it}\alpha_i)\ne0$. No selection bias: $E(x_{it}\eta_{is}|d_i)=0$

Dynamic Panel

AR1: $y_{it}=\rho y_{it-1}+\alpha_i+\eta_{it}$

Assumption: A2: Stationary & ergodic $|\rho|<1$

A3: Moment condition $E(\eta_{it}|y_{i,t-1},\dots,\alpha_i)=0$

A5: mds: $\eta_{it}\overset{iid}{\sim}(0,\sigma_\eta^2)$

$\alpha_i\overset{iid}{\sim}(0,\sigma_\alpha^2)$

1st order difference: $y_{it}-y_{it-1}=\rho(y_{it-1}-y_{it-2})+\eta_{it}-\eta_{it-1}$

$\Delta y_{it}=\rho\Delta y_{i,t-1}+\Delta\eta_{i,t}$

IV: $y_{i,t-2}$: $E(y_{i,t-2}\Delta y_{i,t-1})\ne0$. $E(y_{i,t-2}\Delta\eta_{i,t})=0$

$\hat\rho_{AH}=(\frac{1}{n(T-1)}\sum_{i=1}^n\sum_{t=2}^T y_{i,t-2}\Delta y_{i,t-1})^{-1}(\frac{1}{n(T-1)}\sum_{i=1}^n\sum_{t=2}^T y_{i,t-2}\Delta y_{it})$

GMM: $\forall s\le t-2$ $E(y_{is}\Delta\eta_{it})=0$ & Full column rank

$g_i(\rho)=[y_{is}\Delta\eta_{it}]_{s\le t-2,\ 2\le t\le T}$ $g_i(\rho): \frac{T(T-1)}{2}\times1$

$\hat\rho_{AB}=\underset{\rho}{\operatorname{argmin}}\ n\bar g_n^T(\rho)\hat S^{-1}\bar g_n(\rho)$

With pre-determined variables: $y_{it}=\rho y_{i,t-1}+x_{it}^T\beta+\alpha_i+\eta_{it}$

Assumption: A3: Moment $E(\eta_{it}|y_{i(t-1)},\dots,\alpha_i,x_{it},\dots)=0$

1st order difference: $\Delta y_{it}=\rho\Delta y_{i(t-1)}+\Delta x_{it}^T\beta+\Delta\eta_{it}$

IV: $(y_{i(t-2)},\ x_{it-1})$

GMM: $\forall s\le t-2$ $E(y_{is}\Delta\eta_{it})=0$

$\forall s\le t-1$ $E(x_{is}\Delta\eta_{it})=0$

Diff-in-diff. $y_{it}=\theta D_{it}+x_{it}^T\beta+u_i+v_t+\varepsilon_{it}$

Demean twice: $\tilde y_{it}=\theta\tilde D_{it}+\tilde x_{it}^T\beta+\tilde\varepsilon_{it}$. $\tilde z_{it}=[\tilde D_{it}\ \tilde x_{it}^T]^T$

$\widetilde{Var}_{it}=Var_{it}-\overline{Var}_i-\overline{Var}_t+\overline{Var}$

Assumption: A4 Identification $E\tilde z_{it}\tilde z_{it}^T>0$ (full column rank)

A3 Moment: $E(x_{it}\varepsilon_{is})=0$ $\forall t,s$

$D_{it}\perp\!\!\!\perp\varepsilon_{is}|x_{i1},\dots,x_{iT}$

③ First-differencing: $y_{it}-y_{it-1}=(x_{it}-x_{it-1})^T\beta+(\varepsilon_{it}-\varepsilon_{it-1})$

$C := \begin{bmatrix}-1 & 1 & 0 & \cdots & \\ 0 & -1 & 1 & 0 & \cdots\\ 0 & 0 & -1 & 1 & \cdots\end{bmatrix}_{T-1\times T}$ difference matrix

$\hat\beta_{FD}=(\sum_{i=1}^n X_i^TC^TCX_i)^{-1}\sum_{i=1}^n X_i^TC^TCY_i$ (Not BLUE)

① Demean: $Q:=I_T-\frac1T 1_T1_T^T$. $QQ=Q$. $Q^T=Q$

$\hat\delta_{FE}=(\sum_{i=1}^n X_i^TQX_i)^{-1}(\sum_{i=1}^n X_i^TQY_i)$ (Not BLUE)

Extremum Estimator $\hat{\theta} \in \underset{\theta\in\Theta}{\operatorname{argmax}} Q_n(\theta)$

① M-Estimator: $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} m(w_t;\theta)$

e.g. ML: $m(w_t;\theta) = f(w_t;\theta)$ density

NLS: $m(w_t;\theta) = [y_t - g(x_t;\theta)]^2$ least square

② GMM: $Q_n(\theta) = -\frac{1}{2} g_n(\theta)^T \hat{W} g_n(\theta)$, $g_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(w_t;\theta)$

③ Classical minimum distance estimator

$g_n(\theta)$ is not necessarily a sample mean

Maximum Likelihood Estimation (MLE)

iid w_t: $\hat{\theta}_{ML} \in \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{t=1}^{n} \ln f(w_t;\theta)$

GMM: $E(w_t - \mu(\theta_0)) = 0$, $\mu(\theta_0) = E(w_t;\theta_0)$

Ex. Estimate (μ,σ^2): $\{w_t\}_{t=1}^{n} \overset{iid}{\sim} N(\mu,\sigma^2)$

MLE: $f(w_t;\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(w_t-\mu)^2}{2\sigma^2}\}$

$\ln f(w_t;\mu) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(w_t-\mu)^2$

$Q_n = \frac{1}{n}\sum_{t=1}^{n} \ln f(w_t;\mu) = -\frac{1}{2n\sigma^2}\sum_{t=1}^{n}(w_t-\mu)^2 - \frac{1}{2}\ln(2\pi\sigma^2)$

FOC: $\frac{\partial Q_n(\mu)}{\partial \mu} = +\frac{1}{2n\sigma^2}\sum_{t=1}^{n} 2(w_t-\mu) = 0$

$\frac{\partial Q_n(\sigma^2)}{\partial \sigma^2} = \frac{1}{2n(\sigma^2)^2}\sum_{t=1}^{n}(w_t-\mu)^2 - \frac{1}{2\sigma^2} = 0$

$\Rightarrow \begin{cases} \hat{\mu}_{ML} = \frac{1}{n}\sum_{t=1}^{n} w_t \\ \hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{t=1}^{n}(w_t-\bar{w})^2 \end{cases}$

GMM: $E(w_t-\mu) = 0$ $\quad g_n = \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} w_t-\mu \\ (w_t-\mu)^2-\sigma^2 \end{bmatrix}$

$E(w_t-\mu)^2 - \sigma^2 = 0$

$Q_n = -\frac{1}{2} g_n^T \hat{W} g_n$

FOC: $\frac{dQ_n}{d\theta}_{2\times1} = \frac{dg_n^T}{d\theta}_{2\times2} \cdot \frac{dQ_n}{dg_n}_{2\times1}$

$= \begin{bmatrix} -1 & -2(\bar{w}-\mu) \\ 0 & -1 \end{bmatrix} (-\hat{W} g_n) = 0_{2\times1}$

Let $\hat{W} = I$: $(\hat{\mu}_{GMM}, \hat{\sigma}_{GMM}) = (\hat{\mu}_{ML}, \hat{\sigma}_{ML})$

Conditional MLE:

iid $\{y_t, x_t^T\}$: $\hat{\theta}_{CML} \in \underset{\theta}{\operatorname{argmax}} \frac{1}{n}\sum_{t=1}^{n} f(y_t|x_t;\theta)$

$Avar(\hat{\theta}_{CML}) - Avar(\hat{\theta}_{ML}) \geq 0$.

Ex. Linear Reg with Conditional Homoskedasticity.

iid $\{y_t, x_t^T\}$ $y_i = x_i^T\beta_0 + \varepsilon_t$. $\varepsilon_t|x_t \sim N(0,\sigma_0^2)$

$\ln f(y_t|x_t;\beta,\sigma^2) = -\frac{1}{2\sigma^2}(y_t - x_t^T\beta)^2 - \frac{1}{2}\ln 2\pi\sigma^2$

$Q_n = \frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|x_t;\beta,\sigma^2)$

FOC: $\frac{\partial Q_n}{\partial \beta} = \frac{1}{n}\sum_{t=1}^{n} 2\frac{1}{2\sigma^2} x_t(y_t - x_t^T\beta) = 0$

$\frac{\partial Q_n}{\partial \sigma^2} = \frac{1}{n}\sum_{t=1}^{n}\{\frac{1}{2\sigma^4}(y_t - x_t^T\beta)^2 - \frac{1}{2\sigma^2}\} = 0$

$\Rightarrow \hat{\beta}_{CMLE} = S_{xx}^{-1} S_{xy} = \beta_0 + S_{xx}^{-1}\frac{1}{n}\sum_{t=1}^{n} x_t\varepsilon_t$

$\hat{\sigma}^2_{CMLE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - x_t^T\hat{\beta}_{CMLE})^2$

GMM: $E(\varepsilon_t|x_t) = 0 \Rightarrow E(x_t(y_t - x_t^T\beta)) = 0$

$E(\varepsilon_t^2|x_t) = \sigma_0^2 \Rightarrow E(x_t[(y_t - x_t^T\beta)^2 - \sigma^2]) = 0$

Ex. Probit $\quad f(y_t=1|x_t;\theta_0) = \Phi(x_t^T\theta_0)$

$f(y_t=0|x_t;\theta_0) = 1 - \Phi(x_t^T\theta_0)$

$f(y_t|x_t;\theta_0) = \Phi(x_t^T\theta_0)^{y_t}[1-\Phi(x_t^T\theta_0)]^{1-y_t}$

$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} y_t\Phi(x_t^T\theta) + \frac{1}{n}\sum_{t=1}^{n}(1-y_t)(1-\Phi(x_t^T\theta))$

ML is invariant to reparameterizing:

$\tau: \Theta \to \Lambda$ bijecture: $\hat{\pi}_{ML} = \underset{\pi\in\Lambda}{\operatorname{argmax}} \tilde{Q}_n(\pi) = \hat{\theta}_{ML}$

$\Leftrightarrow \tilde{Q}_n(\pi) = Q_n(\tau^{-1}(\pi))$ $\forall \pi\in\Lambda$

GMM is not invariant $\quad \pi = \frac{1}{\theta}$ $\quad \tilde{Q}_n(\pi) \neq Q_n(\tau^{-1}(\pi))$

$y_t = \theta_0 x_t + \varepsilon_t$. $g(w_t,\theta_0) = z_t(y_t - \theta_0 x_t)$

$x_t = \pi_0 y_t - \pi_0\varepsilon_t$ $\quad \tilde{g}(w_t,\theta_0) = z_t(x_t - \pi_0 y_t)$

Existence & Consistency

① $\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmax}} Q_n(\theta)$. $\xrightarrow{P}$ $\theta_0 = \underset{\theta\in\Theta}{\operatorname{argmax}} Q_0(\theta)$

i Θ is compact $\quad$ i $\theta_0 \in int(\Theta)$. Θ convex

ii $Q_n(\theta)$ cont in θ $\forall\{w_t\}$ $\quad$ (ii) $Q_n(\theta)$ concave in θ

iii $Q_n(\theta)$ measurable in $\{w_t\}$ $\forall\theta$

(a) Identification: $Q_0(\theta)$ is uniquely maximized at θ_0

(b) Uniform $Q_n(\theta) \xrightarrow{P} Q_0(\theta)$ $\quad$ (b) Pointwise $Q_n(\theta) \xrightarrow{P} Q_0(\theta)$

② $\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{t=1}^{n} m(w_t;\theta) \xrightarrow{P} \theta_0 = \underset{\theta\in\Theta}{\operatorname{argmax}} Em(w_t;\theta)$

$\{w_t\}$ Ergodic & stationary

i Θ is compact $\quad$ i $\theta_0 \in int(\Theta)$. Θ convex

ii m cont. in θ $\forall\{w_t\}$. $\quad$ ii m concave in θ

iii m measurable in $\{w_t\}$ $\forall\theta$

(a) Identification: $Em(w_t;\theta)$ is uniquely maximized at θ_0

(b) $E(\sup_{\theta\in\Theta}|m(w_t;\theta)|) < +\infty$ $\quad$ (b) $E(|m(w_t;\theta)|) < +\infty$

Identification:

① NLS: $m(w_t;\theta) = -(y_t - \varphi(x_t;\theta))^2$, $E(y_t|x_t) = \varphi(x_t;\theta_0)$

$-Em(w_t;\theta) = E[(y_t - \varphi(x_t;\theta_0)) + (\varphi(x_t;\theta_0) - \varphi(x_t;\theta))]^2$

$= E(y_t - \varphi(x_t;\theta_0))^2 + E(\varphi(x_t;\theta_0) - \varphi(x_t;\theta))^2$

Identification $\Leftrightarrow \forall\theta\neq\theta_0$: $Prob(\varphi(x_t;\theta_0) \neq \varphi(x_t;\theta)) > 0$

② MLE: $m(w_t;\theta) = \ln f(y_t|x_t;\theta)$

$Em(w_t;\theta) = E\ln\frac{f(y_t|x_t;\theta)}{f(y_t|x_t;\theta_0)} + E\ln f(y_t|x_t;\theta_0)$

Identification $\Leftrightarrow \forall\theta\neq\theta_0$.

$E\ln\frac{f(y_t|x_t;\theta)}{f(y_t|x_t;\theta_0)} > 0 = \ln E\frac{f(y_t|x_t;\theta)}{f(y_t|x_t;\theta_0)}$

$\Leftrightarrow Prob(f(y_t|x_t;\theta) \neq f(y_t|x_t;\theta_0)) > 0$

Ex. Linear Regression: $y_t = x_t^T\beta + \varepsilon_t$, $\varepsilon_t|x_t \sim N(0,\sigma^2)$

$m = \ln f(y_t|x_t;\theta) = -\frac{1}{2\sigma^2}(y_t - x_t^T\beta)^2 - \frac{1}{2}\ln 2\pi\sigma^2$

If $\sigma^2 \neq \sigma_0^2$: $Prob(m(w_t;\theta) \neq m(w_t;\theta)) > 0$

If $\beta \neq \beta_0$: Identification $\Leftarrow Prob(x_t^T\beta \neq x_t^T\beta_0) > 0$

$\Leftarrow E(\beta-\beta_0)^T x_t x_t^T(\beta-\beta_0) > 0 \Leftarrow E x_t x_t^T$ non-singular.

Ex. Probit model: $f(y_t|x_t;\theta) = \Phi(x_t^T\theta)^{y_t}[1-\Phi(x_t^T\theta)]^{1-y_t}$

Identification $\Leftarrow E x_t x_t^T$ non-singular.

③ GMM: $Q_0(\theta) = -\frac{1}{2} Eg_t^T(w_t;\theta) W Eg_t(w_t;\theta)$

Identification $\Leftrightarrow \forall\theta\neq\theta_0$. $Eg_t(w_t;\theta) \neq 0$

Asymptotic Normality

M-Estimator: $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} m(w_t;\theta)$

Score: $S(w_t;\theta)_{k\times 1} = \frac{\partial m(w_t;\theta)}{\partial\theta}$

Hessian: $H(w_t;\theta)_{k\times k} = \frac{\partial^2 m(w_t;\theta)}{\partial\theta\,\partial\theta^T} = \frac{\partial S(w_t;\theta)}{\partial\theta^T}$

$\hat\theta = \arg\max_{\theta} Q_n(\theta)$: FOC $\frac{1}{n}\sum_{t=1}^{n} S(w_t;\hat\theta) = 0$

Mean Value Thm: $\frac{1}{n}\sum_{t=1}^{n} S(w_t;\theta_0) + \frac{1}{n}\sum_{t=1}^{n} H(w_t;\bar\theta)(\hat\theta-\theta_0) = 0$

$\bar\theta$ is between $\hat\theta$ & θ_0

$$\Rightarrow \sqrt{n}(\hat\theta-\theta_0) = \Big[\frac{1}{n}\sum_{t=1}^{n} H(w_t;\bar\theta)\Big]^{-1}\frac{\sqrt{n}}{n}\sum_{t=1}^{n} S(w_t;\theta_0)$$
$$\xrightarrow{d} E H(w_t;\theta_0)^{-1}\, N(0,\Sigma)$$

① S MDS $\Rightarrow \Sigma = E(S(w_t;\theta_0)S^T(w_t;\theta_0)) = -EH(w_t;\theta_0)$

$Avar(\hat\theta) = -EH(w_t;\theta_0)^{-1} = E(S(w_t;\theta_0)S^T(w_t;\theta_0))^{-1}$

② Otherwise : $\Sigma = \sum_{j=-\infty}^{+\infty} E(S(w_t;\theta_0)S^T(w_{t-j};\theta_0))$

Conditional MLE: $m(w_t;\theta) = \ln f(y_t|x_t;\theta)$

$$E(S(w_t;\theta_0)|x_t) = \int \frac{\partial \ln f(y_t|x_t;\theta_0)}{\partial\theta} f(y_t|x_t;\theta_0)\,dy_t = \int \frac{\partial}{\partial\theta} f(y_t|x_t;\theta_0)\,dy_t = \frac{\partial}{\partial\theta}1 = 0$$

$$\begin{aligned}
&E(S(w_t;\theta_0)S^T(w_t;\theta_0)|x_t) + E(H(w_t;\theta_0)|x_t) \\
&= \int S(w_t;\theta_0)\frac{\partial \ln f(y_t|x_t;\theta_0)}{\partial\theta^T}\exp\{\ln f(y_t|x_t;\theta_0)\} + \frac{\partial S(w_t;\theta_0)}{\partial\theta^T}\exp\{\ln f(y_t|x_t;\theta_0)\}\,dy_t \\
&= \int \frac{\partial}{\partial\theta^T}\{S(w_t;\theta_0)\exp\{\ln f(y_t|x_t;\theta_0)\}\}\,dy_t \\
&= \frac{\partial}{\partial\theta^T}\int\frac{\partial}{\partial\theta} f(y_t|x_t;\theta_0)\,dy_t = \frac{\partial}{\partial\theta^T}\frac{\partial}{\partial\theta}1 = 0
\end{aligned}$$

Ex. $\ln f(y_t|x_t;\theta) = -\frac{(y_t-x_t^T\beta)^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)$

$$S(w_t;\theta_0) = \begin{bmatrix} \frac{1}{\sigma^2}x_t(y_t-x_t^T\beta_0) \\ \frac{(y_t-x_t^T\beta_0)^2}{2\sigma_0^4} - \frac{1}{2\sigma_0^2}\end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_0^2}x_t\varepsilon_t \\ -\frac{1}{2\sigma_0^2} + \frac{\varepsilon_t^2}{2\sigma_0^4}\end{bmatrix}$$

$$E\,S(w_t;\theta_0)S^T(w_t;\theta_0) = \begin{bmatrix} \frac{1}{\sigma_0^2}Ex_tx_t^T & 0 \\ 0 & \frac{1}{2\sigma_0^2}\end{bmatrix}$$

Ex. $\ln f(y_t|x_t;\theta) = y_t\ln\Phi(x_t^T\theta) + (1-y_t)\ln(1-\Phi(x_t^T\theta))$

$$S(w_t;\theta_0) = \frac{\phi(x_t^T\theta)\,[y_t-\Phi(x_t^T\theta)]}{\Phi(x_t^T\theta)\,[1-\Phi(x_t^T\theta)]}x_t$$

iid $\{w_t\}$: ML is more efficient than GMM.

$Avar(\hat\theta_{GMM}) \geq E[g(w_t;\theta_0)g^T(w_t;\theta_0)]^{-1} \geq E[S(w_t;\theta_0)S^T(w_t;\theta_0)]^{-1}$

$g(w_t;\theta) = S(w_t;\theta) = \frac{\partial \ln f(w_t;\theta)}{\partial\theta}$

$$\sqrt{n}(\hat\theta-\theta_0) = -(\Psi)^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} + o_p$$

	ML	GMM
$Q_n(\theta_0)$	$\frac{1}{n}\sum_{t=1}^{n} m(w_t;\theta_0)$	$-\frac{1}{2}g_n^T(\theta_0)\hat W g_n(\theta_0)$
$\frac{\partial Q_n}{\partial\theta}(\theta_0)$	$\frac{1}{n}\sum_{t=1}^{n} S(w_t;\theta_0)$	$-\frac{\partial g_n^T(\theta_0)}{\partial\theta}\hat W g_n(\theta_0)$
Ψ	$-E\,H(w_t;\theta_0)$	$G \leftarrow \frac{\partial g^T(\theta_0)}{\partial\theta} W \frac{\partial g(\theta_0)}{\partial\theta^T}$
$Avar(\frac{\partial Q_n(\theta_0)}{\partial\theta})$	$Avar(\frac{1}{n}\sum_{t=1}^{n} S_t)$	$G^T W\, Avar(\frac{1}{n}\sum_{t=1}^{n} g_t)\, W G$

Hypothesis Testing (If $Avar(\hat\theta) = \Sigma = -\Psi$)

$H_0: a(\theta_0) = 0$. $A(\theta) = \frac{\partial a(\theta)}{\partial\theta^T}$ $(r\times K)$

Wald: $W = n\, a^T(\hat\theta)\,\widehat{Avar}(a(\hat\theta))^{-1} a(\hat\theta) \xrightarrow{d} \chi^2_r$

$\sqrt{n}(a(\hat\theta)-a(\theta_0)) = A(\theta_0)\sqrt{n}(\hat\theta-\theta_0) + o_p \to N_r$

$Avar(a(\hat\theta)) = A(\theta_0)\,Avar(\hat\theta)\,A^T(\theta_0)$

Lagrangian Multiplier: $LM = n\,\gamma_n^T\,\widehat{Avar}(\gamma_n)^{-1}\gamma_n \xrightarrow{d} \chi^2_r$

"Score test" $= n\left(\frac{\partial Q_n(\hat\theta)}{\partial\theta}\right)^T\Sigma^{-1}\left(\frac{\partial Q_n(\hat\theta)}{\partial\theta}\right)$

$\tilde\theta = \arg\max_{\theta} Q_n(\theta)$ s.t. $a(\theta) = 0$.

$$\begin{cases} \frac{\partial Q_n(\tilde\theta)}{\partial\theta} + \gamma_n A^T(\tilde\theta) = 0 & \Rightarrow S(\theta_0) - \Psi(\tilde\theta-\theta_0) + \gamma_n A^T(\theta_0) = 0 \\ a(\tilde\theta) = 0 & \quad a(\theta_0) + A(\theta_0)(\tilde\theta-\theta_0) = 0 \end{cases}$$

$$\Rightarrow \begin{bmatrix} \Psi & A^T(\theta_0) \\ A(\theta_0) & 0 \end{bmatrix}\begin{bmatrix} \sqrt{n}(\tilde\theta-\theta_0) \\ \sqrt{n}\gamma_n \end{bmatrix} = \begin{bmatrix} -\frac{1}{n}\sum_{t=1}^{n} S_t(\theta_0) \\ 0 \end{bmatrix} + o_p$$

$$\Rightarrow \begin{bmatrix} \sqrt{n}(\tilde\theta-\theta_0) \\ \sqrt{n}\gamma_n \end{bmatrix} = \begin{bmatrix} \Psi^{-1}(I-A^T(A\Psi^{-1}A^T)^{-1}A\Psi^{-1}) & \cdots \\ (A\Psi^{-1}A^T)^{-1}A\Psi^{-1} & \cdots \end{bmatrix}\begin{bmatrix} -\frac{1}{n}\sum_{t=1}^{n} S_t(\theta_0) \\ 0 \end{bmatrix}$$

$Avar(\gamma_n) = [A(\theta_0)\,\Sigma^{-1}A^T(\theta_0)]^{-1}$

Likelihood Ratio: $LR = 2n[Q_n(\hat\theta) - Q_n(\tilde\theta)] \xrightarrow{d} \chi^2_r$

$$Q_n(\hat\theta) = Q_n(\theta_0) + \frac{1}{n}\sum_{t=1}^{n} S(w_t;\theta_0)(\hat\theta-\theta_0) + \frac{1}{2}(\hat\theta-\theta_0)^T\frac{1}{n}\sum_{t=1}^{n} H(w_t;\theta_0)(\hat\theta-\theta_0)$$

$$2n[Q_n(\hat\theta) - Q_n(\tilde\theta)] = 2\frac{\sqrt{n}}{n}\sum_{t=1}^{n} S(w_t;\theta_0)\sqrt{n}(\hat\theta-\tilde\theta) + \sqrt{n}(\hat\theta-\tilde\theta)^T\Psi\sqrt{n}(\hat\theta-\tilde\theta) + o_p$$

$$\begin{cases} \sqrt{n}(\tilde\theta-\theta_0) = -[\Psi^{-1}-\Psi^{-1}A^T(A\Psi^{-1}A^T)^{-1}A\Psi^{-1}]\frac{\sqrt{n}}{n}\sum_{t=1}^{n} S(w_t;\theta_0) + o_p \\ \sqrt{n}(\hat\theta-\theta_0) = -\Psi^{-1}\frac{\sqrt{n}}{n}\sum_{t=1}^{n} S(w_t;\theta_0) + o_p \end{cases}$$

$$\Rightarrow \sqrt{n}(\hat\theta-\tilde\theta) = -\Psi^{-1}A^T(A\Psi^{-1}A^T)^{-1}A\Psi^{-1}\frac{\sqrt{n}}{n}\sum_{t=1}^{n} S(w_t;\theta_0) + o_p$$
$$\xrightarrow{d} N(0, \Psi^{-1}A^T(A\Psi^{-1}A^T)^{-1}A\Psi^{-1}) = N(0,\Psi^{-1})$$

$W - LM \xrightarrow{p} 0$. $LM - LR \xrightarrow{p} 0$

$\{w_t\}$ Ergodic Stationary:

① Quasi-ML: $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\ln f(w_t;\theta)$

$\hat\theta_{QML} \xrightarrow{p} \theta_0$.

Estimated by HAC

$Avar(\hat\theta) = (EH(w_t;\theta_0))^{-1}\,Avar(S(w_t;\theta_0))\,(EH(w_t;\theta_0))^{-1}$

② Exact ML: $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\ln f(y_t|y_{t-1},\cdots,y_0;\theta) + \frac{1}{n}\ln f(y_0;\theta)$

③ Conditional ML: $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\ln f(y_t|y_{t-1},\cdots,y_0;\theta)$