## Probability theory

Mgf of $X$ is $M_X[t]=\mathrm{E}\,e^{tX}$. Th.: $\mathrm{E}\,X^n=M_X^{(n)}[0]$. If $X,Y$ independent then $M_{X+Y}[t]=M_X[t]M_Y[t]$. Characteristic function: $\phi_X[t]=\mathrm{E}\,e^{itX}$.

Transformations: $Y=g[X]$. If $X$ – discrete, then $f_Y[y]=\sum_j f_X[x_j]\cdot\mathbb{I}[g[x_j]=y]$.

If $X$ – continuous r.v, then partition $\Xi=A_0\sqcup\ldots\sqcup A_k$ ( $\Pr[X\in A_0]=0$ ), define $g_i[x]=g[x],\ x\in A_i$ so that each $g_i$ is monotonous. Then

$f_Y[y]=\sum_{i=1}^k f_X[g_i^{-1}[y]]\left|\frac{d}{dy}g_i^{-1}[y]\right|,\quad y\in\Psi=\{y\mid\exists x\in A_i: y=g_i[x]\}$.

If $\mathbf{X}$ – cont's vector r.v, and $\mathbf{Y}=\mathbf{g}[\mathbf{X}]$, then define partition $\boldsymbol{\Xi}=\mathbf{A}_0\sqcup\ldots\sqcup\mathbf{A}_k$ ( $\Pr[\mathbf{X}\in\mathbf{A}_0]=0$ ) so that $\mathbf{g}_i[\mathbf{x}]=\mathbf{g}[\mathbf{x}],\quad \mathbf{x}\in\mathbf{A}_i$ and each $\mathbf{g}_i$ is one-to-one, then $f_{\mathbf{Y}}[\mathbf{y}]=\sum_{i=1}^k f_{\mathbf{X}}[\mathbf{g}_i^{-1}[\mathbf{y}]]\left|\mathbf{J}_i\right|$, where $\mathbf{J}_i$ is a Jacobian of inverse transformation: $\mathbf{J}_i=\left\|\frac{\partial\mathbf{g}_i^{-1}[\mathbf{y}]}{\partial\mathbf{y}}\right\|$.

Convolution: if $X,Y$ independent, then $f_{X\pm Y}[z]=\int_{-\infty}^{\infty}f_X[w]f_Y[\pm(z-w)]dw$, $f_{XY}[z]=\int_{-\infty}^{\infty}f_X[\tfrac{z}{w}]f_Y[w]\tfrac{1}{|w|}dw$, $f_{X/Y}[z]=\int_{-\infty}^{\infty}f_X[zw]f_Y[w]|w|dw$

If $X\sim N[\mu,\sigma^2]$, $Y\sim N[\nu,\tau^2]$ independent, then $X+Y\sim N[\mu+\nu,\ \sigma^2+\tau^2]$. If $X\sim Poi[\theta]$, $Y\sim Poi[\lambda]$ independent, then $X+Y\sim Poi[\theta+\lambda]$.

LIE: $\mathrm{E}[Y]=\mathrm{E}[\mathrm{E}[Y|X]]$, $\mathrm{Var}[Y]=\mathrm{E}[\mathrm{Var}[Y|X]]+\mathrm{Var}[\mathrm{E}[Y|X]]$.

## Distributions

**Binomial:** $\Pr[X=k]=C_n^k p^k(1-p)^{n-k},\quad k=0,\ldots,n$. $\mathrm{E}\,X=np,\quad \mathrm{Var}\,X=np(1-p),\quad M_X=(pe^t+1-p)^n$

**Poisson**: $\Pr[X=k]=e^{-\lambda}\lambda^k/k!,\quad k=0,1,2,\ldots$. $\mathrm{E}\,X=\lambda,\quad \mathrm{Var}\,X=\lambda,\quad M_X=e^{\lambda(e^t-1)}$

**Uniform**: $f[x]=1/(b-a),\quad x\in[a;b]$. $\mathrm{E}\,X=\frac{1}{2}(b+a),\quad \mathrm{Var}\,X=\frac{1}{12}(b-a)^2$

**Gamma**: $f[x]=\frac{1}{\Gamma[\alpha]\beta^\alpha}x^{\alpha-1}e^{-x/\beta},\quad x>0,\alpha>0,\beta>0$. $\mathrm{E}\,X=\alpha\beta,\quad \mathrm{Var}\,X=\alpha\beta^2,\quad M_X=(1-\beta t)^{-\alpha},\ t<\beta^{-1}$. Cases: $\gamma[\frac{p}{2},2]=\chi^2[p],\quad \gamma[1,\beta]=\exp[\beta]$

**Chi-squared**: $f[x]=2^{-n/2}\Gamma[n/2]e^{-x/2}x^{-n/2-1},\ x>0$. $\mathrm{E}\,X=n,\ \mathrm{Var}\,X=2n$. If $\mathbf{z}\sim N[\mathbf{0},\mathbf{I}_n]$, $\mathbf{A}$ idempotent, then $\mathbf{z}'\mathbf{A}\mathbf{z}\sim\chi^2[\mathrm{rg}\,\mathbf{A}]$. If $\mathbf{z}\sim N[\mathbf{0},\boldsymbol{\Sigma}]$, then $\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}\sim\chi^2[\mathrm{rg}\,\boldsymbol{\Sigma}]$.

**Exponential**: $f[x]=\frac{1}{\lambda}e^{-x/\lambda},\quad x>0,\lambda>0$. $\mathrm{E}\,X=\lambda,\quad \mathrm{Var}\,X=\lambda^2,\quad \mathrm{E}\,X^k=\lambda^k k!$

**Normal**: $f[x]=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$. $\mathrm{E}\,X=\mu,\quad \mathrm{Var}[X]=\sigma^2,\quad \mathrm{E}[(X-\mu)^{2k}]=\sigma^{2k}(2k-1)!!$

**Beta**: $f[x]=\frac{\Gamma[\alpha+\beta]}{\Gamma[\alpha]\Gamma[\beta]}x^{\alpha-1}(1-x)^{\beta-1},\quad x\in(0;1),\alpha>0,\beta>0$. $\mathrm{E}\,X=\frac{\alpha}{\alpha+\beta},\quad \mathrm{Var}\,X=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},\quad \mathrm{E}\,X^n=\frac{\Gamma[\alpha+n]\Gamma[\alpha+\beta]}{\Gamma[a]\Gamma[\alpha+\beta+n]}$

**Lognormal**: $f[x]=\frac{1}{\sqrt{2\pi}\sigma x}e^{-(\ln x-\mu)/(2\sigma^2)},\quad x>0$. $\mathrm{E}\,X=e^{\mu+\frac{1}{2}\sigma^2},\quad \mathrm{Var}\,X=e^{2(\mu+\sigma^2)}-e^{2\mu+\sigma^2}$.

**Exponential family**: $f[x]=h[x]c[\boldsymbol{\theta}]\exp\left[\sum_{i=1}^k w_i[\boldsymbol{\theta}]t_i[x]\right]$, then $\mathrm{E}\left[\sum_{i=1}^k t_i[X]\frac{\partial w_i[\boldsymbol{\theta}]}{\partial\theta_j}\right]=-\frac{\partial}{\partial\theta_j}\ln c[\boldsymbol{\theta}],\quad \mathrm{Var}\left[\sum_{i=1}^k t_i[X]\frac{\partial w_i[\boldsymbol{\theta}]}{\partial\theta_j}\right]=-\frac{\partial^2}{\partial\theta_j^2}\ln c[\boldsymbol{\theta}]-\mathrm{E}\left[\sum_{i=1}^k t_i[X]\frac{\partial^2 w_i[\boldsymbol{\theta}]}{\partial\theta_j^2}\right]$

**Multinomial normal**: $f[\mathbf{x}]=(2\pi)^{-k/2}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$. If $\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\sim N[\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22}\end{pmatrix}]$, then $\mathbf{y}\,|\,\mathbf{x}\sim N[\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1),\ \boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}]$.

## Inequalities

**Chebychev**: $\Pr[g[X]\ge\varepsilon]\le\varepsilon^{-1}\,\mathrm{E}\,g[X]$

**Vysochanskiĭ-Petunin**: $X\sim f$ unimodal, define $\xi^2=\mathrm{E}[(X-\alpha)^2]$ for any $\alpha$. Then $\Pr[|X-\alpha|>\varepsilon]\le\begin{cases}4\xi^2/9\varepsilon^2, & \varepsilon\ge\xi\sqrt{8/3}\\ 4\xi^2/9\varepsilon^2-\frac{1}{3}, & \varepsilon\le\xi\sqrt{8/3}\end{cases}$

**Stein's** lemma: $X\sim N[\mu,\sigma^2]\quad\Rightarrow\quad \mathrm{E}[g[X](X-\mu)]=\sigma^2\,\mathrm{E}[g'[X]]$

**Hölder**: if $p+q=1$, then $\mathrm{E}|XY|\le(\mathrm{E}|X|^{1/p})^p(\mathrm{E}|Y|^{1/q})^q$. **Cauchy-Schwarz**: $\mathrm{E}|XY|\le\sqrt{\mathrm{E}|X|^2\,\mathrm{E}|Y|^2}$. **Liapounov**: $\left(\mathrm{E}|X|^r\right)^{1/r}\le\left(\mathrm{E}|X|^s\right)^{1/s},\quad 0\le r<s$

**Minkowski**: $\mathrm{E}[|X+Y|^p]^{1/p}\le\mathrm{E}[|X|^p]^{1/p}+\mathrm{E}[|Y|^p]^{1/p},\quad p\ge1$. **(?)**: $\mathrm{E}[|X+Y|^r]\le\max[1,2^{r-1}](\mathrm{E}|X|^r+\mathrm{E}|Y|^r)$.

**Jensen**: if $g[x]$ convex, then $\mathrm{E}\,g[X]\ge g[\mathrm{E}\,X]$.

**Covariance** inequality-II: if $g,h$ both non-increasing or non-decreasing then $\mathrm{E}[g[X]h[X]]\ge\mathrm{E}[g[X]]\mathrm{E}[h[X]]$.

## Convergence

$\{X_n\}$ converges **almost surely** to r.v. $X$ if $\Pr\left[\omega\,\middle|\,\lim_{n\to\infty}X_n[\omega]=X[\omega]\right]=1$.

$\{X_n\}$ converges **in** $L_p$ to r.v. $X$ if $\lim_{n\to\infty}\mathrm{E}[\|X_n-X\|^p]=0$

$\{X_n\}$ converges **in probability** to r.v. $X$ if for $\forall\varepsilon>0\ \ \lim_{n\to\infty}\Pr[\|X_n-X\|>\varepsilon]=0$.

$\{X_n\}$ converges **in distribution** to r.v. $X$ if $\lim_{n\to\infty}\Pr[X_n\le x]=\Pr[X\le x]$ for all $x$ where $F_X[x]$ is continuous.

Relationship: $X_n\xrightarrow{L_p}X\Rightarrow X_n\xrightarrow{L_q}X$ for $p\ge q$, $X_n\xrightarrow{L_1}X\Rightarrow X_n\xrightarrow{p}X\Rightarrow X_n\xrightarrow{d}X$, $X_n\xrightarrow{as}X\Rightarrow X_n\xrightarrow{p}X$.

**Th.**: if for $\forall\boldsymbol{\lambda}\in\mathbb{R}^k$ sequence of scalar r.v's $\boldsymbol{\lambda}'\mathbf{X}_n\xrightarrow{d}\boldsymbol{\lambda}'\mathbf{X}$, then $\mathbf{X}_n\xrightarrow{d}\mathbf{X}$.

**Continuous mapping** theorem: if $X_n\xrightarrow{p}X$ and $h[\cdot]$ is continuous, then $h[X_n]\xrightarrow{p}h[X]$.

**Mann&Wald**: If $g:\mathbb{R}^k\to\mathbb{R}^l$ is continuous, then $\mathbf{X}_n\xrightarrow{as}\mathbf{X}\Rightarrow g[\mathbf{X}_n]\xrightarrow{as}g[\mathbf{X}]$; $\mathbf{X}_n\xrightarrow{p}\mathbf{X}\Rightarrow g[\mathbf{X}_n]\xrightarrow{p}g[\mathbf{X}]$; $\mathbf{X}_n\xrightarrow{d}\mathbf{X}\Rightarrow g[\mathbf{X}_n]\xrightarrow{d}g[\mathbf{X}]$.

**Slutsky**: if $\mathbf{X}_n\xrightarrow{d}\mathbf{X}$, $\mathbf{Y}_n\xrightarrow{p}\boldsymbol{\alpha}$, then $\mathbf{X}_n+\mathbf{Y}_n\xrightarrow{d}\mathbf{X}+\boldsymbol{\alpha}$; $\mathbf{X}_n\mathbf{Y}_n\xrightarrow{d}\mathbf{X}\boldsymbol{\alpha}$, $\mathbf{Y}_n\mathbf{X}_n\xrightarrow{d}\boldsymbol{\alpha}\mathbf{X}$; $\mathbf{Y}_n^{-1}\mathbf{X}_n\xrightarrow{d}\boldsymbol{\alpha}^{-1}\mathbf{X}$, $\mathbf{X}_n\mathbf{Y}_n^{-1}\xrightarrow{d}\mathbf{X}\boldsymbol{\alpha}^{-1}$ when $\Pr[\det\mathbf{Y}_n=0]=0$.

**Delta-method**: if $\mathbf{X}_n\xrightarrow{p}\mathbf{X}=\mathrm{const}$, $\sqrt{n}(\mathbf{X}_n-\mathbf{X})\xrightarrow{d}N[\mathbf{0},\boldsymbol{\Sigma}]$, $g:\mathbb{R}^k\to\mathbb{R}^l$ is cont.diff. at $\mathbf{X}$, then $\sqrt{n}(g[\mathbf{X}_n]-g[\mathbf{X}])\xrightarrow{d}N[\mathbf{0},\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}']$, where $\mathbf{G}=\partial g[\mathbf{X}]/\partial x'$.

**WLLN** (weak law of large numbers): Let $X_1,\ldots,X_n$ be iid with $\mathrm{E}\,X_i=\mu<\infty$. Then $\bar{X}_n\equiv\frac{1}{n}\sum X_i\xrightarrow{p}\mu$.

**LLN Kolmogorov** (iid): Let $\{\mathbf{X}_n\}$ be iid and $\mathrm{E}[\mathbf{X}_i]$ exists. Then $\frac{1}{n}\sum\mathbf{X}_i\xrightarrow{as}\mathrm{E}[\mathbf{X}_i]$.

**LLN Kolmogorov** (niid): Let $\{X_i\}$ be indep. with $\operatorname{Var} X_i = \sigma_i^2$ and $\sum_{i=1}^{\infty} \sigma_i^2 / i^2 < \infty$. Then $\frac{1}{n}\sum X_i - \mathrm{E}\left[\frac{1}{n}\sum X_i\right] \xrightarrow{as} 0$.

**LLN Birkhoff-Khinchin** (ergodic): Let $\{\mathbf{X}_t\}$ be stationary and ergodic. Then $\frac{1}{T}\sum \mathbf{X}_t \xrightarrow{as} \mathrm{E}[\mathbf{X}_t]$.

**CLT Lindeberg-Lévy**: Let $\{X_t\}$ be iid with $\mathrm{E}[X_t] = \mu$ and $\operatorname{Var}[X_t] = \sigma^2$. Then $\sqrt{n}\left(\frac{1}{n}\sum X_t - \mu\right) \xrightarrow{d} N[0,\sigma^2]$.

**CLT Liapounov**: Let $\{X_t\}$ be indep. with $\mathrm{E}[X_t] = \mu_t$, $\operatorname{Var}[X_t] = \sigma_t^2$, and $\mathrm{E}[|X_t - \mu_t|^3] = \nu_t$. If $\lim_{T\to\infty}\left(\sum \sigma_t^2\right)^{-1/2}\left(\sum \nu_t\right)^3 = 0$ then $\frac{\sum (X_t - \mu_t)}{\left(\sum \sigma_t^2\right)^{1/2}} \xrightarrow{d} N[0,1]$.

**CLT Lindeberg-Feller**: Let $\{X_t\}$ be indep. with $\mathrm{E}[X_t] = \mu_t$, $\operatorname{Var}[X_t] = \sigma_t^2$. Define $C_T = \left(\sum \sigma_t^2\right)^{1/2}$. If for $\forall \varepsilon > 0$ $\lim_{T\to\infty} \frac{1}{C_T^2}\int_{|x-\mu_t|\ge \varepsilon C_T} (x-\mu_t)^2 dF_t[x] = 0$, then $\left(\sum \sigma_t^2\right)^{-1/2}\sum (X_t - \mu_t) \xrightarrow{d} N[0,1]$.

**CLT Billingsley**: If $\{X_t\}$ stationary & ergodic martingale difference sequence, $\mathrm{E}[X_t^2] = \sigma^2 < \infty$, then $\frac{1}{\sqrt{T}}\sum X_t \xrightarrow{d} N[0,\sigma^2]$.

**CLT Anderson**: Let $\{X_t\}$ be stationary & ergodic with $w_z = \sum_{j=-\infty}^{+\infty} \operatorname{Cov}[X_t, X_{t-j}] < \infty$. Then $\sqrt{T}\left(\frac{1}{T}\sum X_t - \mathrm{E}[X_t]\right) \xrightarrow{d} N[0, w_z]$.

## Data reduction

Parametric model: $\boldsymbol{P} = \{P_\theta : \theta \in \Theta\}$. Parameter: any mapping $\nu : \boldsymbol{P} \to \boldsymbol{N} \subset \mathbb{R}^k$ (natural parameterization: $\nu[P_\theta] = \theta$). Parameter $\nu[P_\theta]$ is identifiable if $P_{\theta_1} = P_{\theta_2} \Rightarrow \nu[P_{\theta_1}] = \nu[P_{\theta_2}]$. Parameter $\theta$ is **identifiable** if $\theta_1 \ne \theta_2 \Rightarrow P_{\theta_1} \ne P_{\theta_2}$, or equivalently $P_{\theta_1} = P_{\theta_2} \Rightarrow \theta_1 = \theta_2$.

**Statistic** – any measurable function of the data $T[\mathbf{X}]$. Statistic $T[\mathbf{X}]$ is **sufficient** for $\theta$ if $f_{\mathbf{X}|T[\mathbf{X}]=t}$ does not depend on $\theta$. Statistic $S[\mathbf{X}]$ is **ancillary** for $\theta$ if $f_{S[\mathbf{X}]}$ does not depend on $\theta$. Statistic $S[\mathbf{X}]$ is **1st order ancillary** for $\theta$ if $\mathrm{E}_\theta S[\mathbf{X}]$ does not depend on $\theta$. Statistic $T[\mathbf{X}]$ is **minimal sufficient** if for any other sufficient statistic $S[\mathbf{X}]$ we can find function $r$ such that $T[\mathbf{X}] = r[S[\mathbf{X}]]$. Statistic $T[\mathbf{X}]$ is **complete** if $\exists! g \equiv 0$ such that $g[T[\mathbf{X}]]$ is first-order ancillary.

**Factorization th.**: statistic $T : \boldsymbol{X} \to \boldsymbol{T}$ is <u>sufficient</u> for $\theta$ if $\exists g : \boldsymbol{T}\times\boldsymbol{\Theta} \to \mathbb{R}$ and $\exists h : \boldsymbol{X} \to \mathbb{R}$ such that $f[x,\theta] = g[T[x],\theta]\cdot h[x]$ for $\forall x \in \boldsymbol{X}, \theta \in \boldsymbol{\Theta}$.

**Th.**: if $T[\mathbf{X}]$ is such that $\forall \mathbf{x},\mathbf{y} \in \boldsymbol{X}$ ratio $f[\mathbf{x}\,|\,\theta] / f[\mathbf{y}\,|\,\theta]$ does not depend on $\theta$ iff $T[\mathbf{x}] = T[\mathbf{y}]$, then $T[\mathbf{X}]$ is <u>minimal sufficient</u> for $\theta$.

**Th.**: for exp. family $f[x,\theta] = h[x]\exp\left[\sum_{j=1}^{k} h_j[\boldsymbol{\theta}] T_j[x] - B[\boldsymbol{\theta}]\right]$ statistic $T[\mathbf{X}] = \left(\sum_{i=1}^{n} T_1[X_i], \ldots, \sum_{i=1}^{n} T_k[X_i]\right)$ is <u>complete</u> if $\boldsymbol{\eta}[\Theta]$ contains an open set in $\mathbb{R}^k$.

**Basu th.**: if $T[\mathbf{X}]$ is complete and minimal sufficient, then it is independent of any ancillary statistic.

**Th.**: if minimal sufficient statistic exists, then any complete statistic will also be minimal sufficient.

## Estimator properties

**Estimator** – any measurable function of the data $\phi[X]$. Estimator $\phi[\cdot]$ is **unbiased** for parameter $g[\theta]$ if for $\forall \theta \in \Theta$ $\mathrm{E}_\theta[\phi[X]] = g[\theta]$. Unbiased estimator $\phi[\cdot]$ is a **UMVUE** (uniformly minimum variance unbiased estimator) if $\operatorname{Var}_\theta[\phi] < \infty$ and for any other unbiased $\delta[X]$ we have $\operatorname{Var}_\theta \phi[X] \le \operatorname{Var}_\theta \delta[X]$ $\forall \theta$.

**Cramér-Rao ineq.**: Let $\{X_i\}_{i=1}^{n} \sim f[\mathbf{x}\,|\,\theta]$ and $\phi[\mathbf{X}]$ be unbiased for $g[\theta]$ s.t. in $\frac{d}{d\theta}\mathrm{E}_\theta[\phi]$ order of integr'n & diff'n is interchangeable. Then

$\boxed{\operatorname{Var}_\theta[\phi[\mathbf{X}]] \ge \frac{d\psi[\theta]}{d\theta}\,\Im[\theta]^{-1}\,\frac{d\psi[\theta]}{d\theta'}} \equiv CRLB[\theta]$, where $\Im[\theta] = \mathrm{E}_\theta[-\frac{\partial^2 \ln f[\mathbf{x}|\theta]}{\partial\theta\partial\theta'}] = \mathrm{E}_\theta[\frac{\partial \ln f[\mathbf{x}|\theta]}{\partial\theta'}\frac{\partial \ln f[\mathbf{x}|\theta]}{\partial\theta}]$ (**Fisher Information** matrix).

**CRLB attainment**: Let $\{X_i\}_{i=1}^{n} \sim iid$ and $W$ is unbiased for $\tau[\theta]$. Then $W$ attains CRLB iff $a[\theta]\left(W[\mathbf{x}] - \tau[\theta]\right) = \frac{\partial}{\partial\theta}\ln L[\theta\,|\,\mathbf{x}]$ for some $a[\theta]$.

**Hausman principle**: $W$ is UMVUE of $\tau[\theta]$ iff $W$ is uncorrelated with all unbiased estimators of 0.

**Rao-Blackwell th.**: Let $W$ be unbiased est'r of $\tau[\theta]$ and $T$ be sufficient statistic for $\theta$. Then $\phi[T] = \mathrm{E}[W\,|\,T]$ is UMVUE of $\tau[\theta]$.

**Lehman-Scheffé th.**: Let $T$ be a complete sufficient statistic for $\theta$. Then $\phi[T]$ based only on $T$ is the unique UMVUE of $\mathrm{E}\,\phi[T]$.

## Hypotheses testing

**Hypothesis** – any statement about model parameter. **Null** hypothesis: $\theta \in \Theta_0$, **alternative** hypothesis: $\theta \in \Theta_1$, where $\Theta_0 \cap \Theta_1 = \varnothing$. **Action space** $\mathfrak{A} = \{0,1\}$, where "1" is rejection of null. **Loss function**: $l[\theta,a] = \mathbb{I}[\theta \notin \Theta_a]$. **Test function**: $\delta : \mathbf{X} \to \{0,1\}$. **Critical region**: $C = \{\mathbf{x} \in \boldsymbol{X} : \delta[\mathbf{x}] = 1\}$. **Type-I error** – to reject $H_0$ when $\theta \in \Theta_0$. **Type-II error** – to accept $H_0$ when $\theta \in \Theta_1$. **Power function**: $\beta_\delta[\theta] = \Pr_\theta[\delta[\mathbf{X}] = 1], \theta \in \Theta_1$ (probability to correctly reject $H_0$ when $\theta \in \Theta_1$). **Size** of test: $size = \sup_{\theta\in\Theta_0} \beta_\delta[\theta]$. **Level** of test is $\alpha$ if $\sup_{\theta\in\Theta_0} \beta_\delta[\theta] \le \alpha$. **P-value** of test: $\hat{p}[\mathbf{X}] = \inf_{\alpha\in(0;1):\mathbf{X}\in C_\alpha} \alpha$. If test is $\delta[\mathbf{X}] = \mathbb{I}[T[\mathbf{X}] \ge c]$, then define $\alpha[c] = \sup_{\theta\in\Theta_0} \Pr_\theta[T[\mathbf{X}] \ge c]$ and p-value is $\alpha[T[\mathbf{x}]]$. Test $\phi$ is **unbiased** of level $\alpha$ if $\beta_\phi[\theta] \le \alpha$ $\forall \theta \in \Theta_0$ and $\beta_\phi[\theta] \ge \alpha$ $\forall \theta \in \Theta_1$. Test $c^* \in C$ is **uniformly most powerful** in class $C$ if $\beta_*[\theta] \ge \beta_c[\theta]$ for $\forall c \in C, \theta \in \Theta_1$. Family $\{P_\theta\,|\,\theta \in \Theta \subset \mathbb{R}\}$ is **monotone likelihood ratio family** if for $\forall \theta_1 > \theta_2$ $P_{\theta_1} \ne P_{\theta_2}$ and $\frac{f[x|\theta_1]}{f[x|\theta_2]}$ is a monotone function of some $T[\mathbf{x}]$. Test $\phi$ is $\alpha$-similar on $\Theta^* \subset \Theta$ if $\beta_\phi[\theta] = \alpha$ $\forall \theta \in \Theta^*$. Mapping $S : \boldsymbol{X} \to 2^N$ is $(1-\alpha)$ **confidence region** for parameter $\nu[\theta]$ if $\Pr_\theta[S[\mathbf{X}] \supset \{\nu[\theta]\}] \ge 1-\alpha$.

**Neyman-Pearson th.**: consider $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ and <u>likelihood-ratio</u> test function $\phi_k[\mathbf{x}] = \{1, \text{ if } \frac{f[x|\theta_1]}{f[x|\theta_0]} > k, \quad 0 \text{ if } \frac{f[x|\theta_1]}{f[x|\theta_0]} < k, \quad \forall \in [0;1] \text{ if } \frac{f[x|\theta_1]}{f[x|\theta_0]} = k\}$. Then 1) $\phi_k$ is MP in class of all level $\alpha = \mathrm{E}_{\theta_0}\phi_k[\mathbf{x}]$ tests; 2) for $\forall \alpha$ exists MP level $\alpha$ of the form $\phi_k$; 3) if a test $\tilde{\phi}$ is MP, then it has form of $\phi_k$ a.s.

**Karlin-Rubin th.**: suppose $\{P_\theta\,|\,\theta \in \Theta\}$ is MLR increasing in $T[\mathbf{x}]$. Define $\delta_t[\mathbf{x}] = \mathbb{I}[T[\mathbf{x}] > t]$. Then 1) $b_{\delta_t}[\theta]$ is increasing in $\theta$; 2) $\delta_t$ is UMP level $\alpha = \mathrm{E}_{\theta_0}[\delta_t[\mathbf{x}]]$ for testing $H_0 : \theta \le \theta_0$ against $H_1 : \theta > \theta_0$.

**Th.**: consider an exp. family $f[x\,|\,\theta] = e^{\theta T[x] - A[\theta]}$ and a test $\phi^*[x] = \{1, \text{ if } T[x] < c_1 \vee T[x] > c_2, \quad 0 \text{ if } T[x] \in (c_1, c_2), \quad \gamma_i[x] \text{ if } T[x] = c_i\}$, then this test is UMPU level $\alpha$ for testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \ne \theta_0$ iff $\mathrm{E}_{\theta_0}[\phi] = \alpha$ and $\mathrm{E}_{\theta_0}[\phi T] = \alpha\,\mathrm{E}_{\theta_0}[T]$.

**Duality th.**: Let $\delta_{\theta_0}$ be level $\alpha$ test of $H_0 : \theta = \theta_0$, and $A[\theta_0] = \{\mathbf{x} \in \boldsymbol{X} : \delta_{\theta_0}[\mathbf{x}] = 1\}$. Define $S[\mathbf{x}] = \{\theta_0 \in \Theta : \mathbf{x} \in A[\theta_0]\}$. Then $S[\mathbf{X}]$ is $(1-\alpha)$ confidence set. Conversely, if $S[\mathbf{X}]$ is $(1-\alpha)$ confidence set then $A[\theta_0] = \{\mathbf{x} \in \boldsymbol{X} : \theta_0 \in S[\mathbf{x}]\}$ is acceptance region of a level $\alpha$ test of $H_0 : \theta = \theta_0$.

**Duality th. for MLR**: suppose $\{P_\theta \mid \theta \in \Theta\}$ is MLR increasing in $T[\mathbf{x}]$, and $F_\theta[t]$ is cont's. If $F_\theta[t] = 1-\alpha$ has solution $\theta_l[\alpha,t] \in \Theta$ and $F_\theta[t] = \alpha$ has solution $\theta_u[\alpha,t] \in \Theta$, then $\forall \alpha_1, \alpha_2 : \alpha_1 + \alpha_2 < 1$ interval $[\theta_l[\alpha_1, T[\mathbf{X}]], \theta_u[\alpha_2, T[\mathbf{X}]]]$ is a $(1-\alpha_1-\alpha_2)$ confidence interval for $\theta$.

## OLS

Model: $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$, $i = 1,\dots,n$; in stacked form: $\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times k}\boldsymbol{\beta}_{k\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$. Assumptions: $\mathrm{E}[\varepsilon_i \mid \mathbf{X}] = 0$ (strict exogeneity); $\mathrm{E}[\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \mid \mathbf{X}] = \sigma^2 \mathbf{I}_n$ (homoscedasticity); $\Pr[\mathrm{rg}\,\mathbf{X} = k] = 1$ (no multicollinearity). OLS estimators: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\hat{\sigma}^2 = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$, $s^2 = \frac{1}{n-k}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$, where $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{y}$, $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{M}\mathbf{y}$, $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{M} = \mathbf{I} - \mathbf{P}$. Partitioned regression: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, then $\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}$, $\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y}$.

**Frisch-Waugh th.**: $\hat{\boldsymbol{\beta}}_2$ from $\mathbf{y} \rhd \mathbf{X}_1, \mathbf{X}_2$ is the same as from regression $\tilde{\mathbf{y}} \rhd \tilde{\mathbf{X}}_2$, where $\tilde{\mathbf{y}}$ are residuals in $\mathbf{y} \rhd \mathbf{X}_1$ and $\tilde{\mathbf{X}}_2$ are residuals in $\mathbf{X}_2 \rhd \mathbf{X}_1$.
**Corollary**: if regression contains intercept, you can first demean it and then carry out regression.
**Th.**: if $z$ is one of regressors, then partial correlation $r_{yz}^* \equiv \tilde{\mathbf{z}}'\tilde{\mathbf{y}} \Big/ \sqrt{\tilde{\mathbf{z}}'\tilde{\mathbf{z}} \cdot \tilde{\mathbf{y}}'\tilde{\mathbf{y}}} = \left(1 + \#df / t_z^2\right)^{-1/2} \mathrm{sgn}[t_z]$, where $t_z$ is t-statistic for $z$.

RSS=regression sum of squares, TSS=total --"--, ESS=errors --"--. If $\mathbf{X}$ contains $\boldsymbol{\iota}$ then $\mathbf{L} = \mathbf{I} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}'$ and $R^2 \equiv \frac{RSS}{TSS} = 1 - \frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{\mathbf{y}'\mathbf{L}\mathbf{y}}$. Adjusted $R^2$: $\tilde{R}^2 = 1 - \frac{n-1}{n-k}(1-R^2)$. When $z$ is added to regression, then $R^2_{\mathbf{x},z} = R^2_{\mathbf{x}} + (1-R^2_{\mathbf{x}})r_{yz}^{*2}$ and $\tilde{R}^2$ will increase only if $t_z^2 > 1$.

### *Finite sample properties*

$\mathrm{E}[\hat{\boldsymbol{\beta}} \mid \mathbf{X}] = \boldsymbol{\beta}$, $\mathrm{Var}[\hat{\boldsymbol{\beta}} \mid \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, $\mathrm{E}[s^2 \mid \mathbf{X}] = \sigma^2$, $\mathrm{Cov}[\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\varepsilon}} \mid \mathbf{X}] = 0$. **Gauss-Markov th.**: $\hat{\boldsymbol{\beta}}$ is BLUE (best linear unbiased estimator).
Under normality assumption: $\hat{\boldsymbol{\beta}} \mid \mathbf{X} \sim N[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$, $\frac{n-k}{\sigma^2}s^2 \sim \chi^2[n-k]$ ( $\mathrm{Var}[s^2 \mid \mathbf{X}] = \frac{2}{n-k}\sigma^4$ ) and $\hat{\boldsymbol{\beta}}$, $s^2$ are independent. Test $H_0 : \beta_j = b$ using $t = (\hat{\beta}_j - b) \Big/ \sqrt{s^2(\mathbf{X}'\mathbf{X})^{-1}_{jj}} \sim t[n-k]$. Test $H_0 : \boldsymbol{\beta} = \mathbf{0}$ (except intercept) using $F = \frac{1}{k-1}R^2 \Big/ \frac{1}{n-k}(1-R^2) = \frac{n-k}{k-1}\frac{RSS}{ESS} \sim F[k-1, n-k]$. Test $H_0 : \mathbf{R}_{q\times k}\boldsymbol{\beta} = \mathbf{r}$ using $F = \frac{1}{qs^2}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \frac{n-k}{q}\frac{ESS_R - ESS}{ESS} = \frac{n-k}{q}\frac{R^2 - R_R^2}{1-R^2} \sim F[q, n-k]$. Constrained estimator: $\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda}$, $\boldsymbol{\lambda} = (\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}})$.
Prediction: **best predictor**: $BP[y \mid \mathbf{x}] = \mathrm{E}[y \mid \mathbf{x}]$, **best linear predictor**: $BLP[y \mid \mathbf{x}] = \mathbf{x}'\,\mathrm{E}[\mathbf{x}\mathbf{x}']^{-1}\,\mathrm{E}[\mathbf{x}y]$.

### *Large sample properties*

Denote $\mathbf{Q}_{xx} = \mathrm{E}[\mathbf{x}_i\mathbf{x}_i'] > 0$, $m_4 = \mathrm{E}[\varepsilon_i^4]$. Then asy.distributions: $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N[\mathbf{0}, \sigma^2\mathbf{Q}_{xx}^{-1}]$, $\sqrt{n}\left(\hat{\sigma}^2 - \sigma^2\right) \xrightarrow{d} N[0, m_4 - \sigma^4]$. Test $H_0 : \beta_j = b$ using $t = \sqrt{n}(\hat{\beta}_j - b) \Big/ \sqrt{\hat{\sigma}^2(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}_{jj}} \to N[0,1]$. Test $H_0 : g[\boldsymbol{\beta}] = \mathbf{0}_{q\times 1}$ using Wald's test: $W = n\frac{1}{\hat{\sigma}^2}g[\hat{\boldsymbol{\beta}}]'(\mathbf{G}[\hat{\boldsymbol{\beta}}](\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}[\hat{\boldsymbol{\beta}}]')^{-1}g[\hat{\boldsymbol{\beta}}] \to \chi^2[q]$, Likelihood ratio test: $LR = 2(\ln\mathfrak{L} - \ln\mathfrak{L}_R) \to \chi^2[q]$, Lagrange multiplier test: $LM = n\frac{1}{\hat{\sigma}^2}\boldsymbol{\lambda}'(\mathbf{G}[\hat{\boldsymbol{\beta}}_R](\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}[\hat{\boldsymbol{\beta}}_R]')^{-1}\boldsymbol{\lambda} \to \chi^2[q]$, where $\mathbf{G}_{q\times k} = \frac{\partial g[\boldsymbol{\beta}]}{\partial \boldsymbol{\beta}}$.
Heteroscedasticity case: denote $\mathbf{Q}_{xx\varepsilon^2} = \mathrm{E}[\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i']$. Then $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N[\mathbf{0}, \mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xx\varepsilon^2}\mathbf{Q}_{xx}^{-1}]$. If $\mathrm{E}[(x_{ij}x_{ik})^2]$ exists and finite for $\forall i,j,k$, then HCSE (**hetero-scedasticity-consistent standard errors**): $\widehat{\mathrm{AVar}}[\hat{\boldsymbol{\beta}}] = \hat{\mathbf{Q}}_{xx}^{-1}\hat{\mathbf{Q}}_{xx\varepsilon^2}\hat{\mathbf{Q}}_{xx}^{-1}$, $\hat{\mathbf{Q}}_{xx} = \frac{1}{T}\sum \mathbf{x}_t\mathbf{x}_t'$, $\hat{\mathbf{Q}}_{xx\varepsilon^2} = \frac{1}{T}\sum \hat{\varepsilon}_t^2\mathbf{x}_t\mathbf{x}_t' \prec \frac{1}{T-k}\sum \hat{\varepsilon}_t^2\mathbf{x}_t\mathbf{x}_t' \prec \frac{1}{T}\sum \frac{\hat{\varepsilon}_t^2}{(1-\mathbf{x}_t(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_t')^d}\mathbf{x}_t\mathbf{x}_t'$, $d = 1$ or $2$.
**White's heteroscedasticity test**: regress $\hat{\varepsilon}_i^2 \rhd 1, \boldsymbol{\psi}_i$, where $\boldsymbol{\psi}_i$ contains unique non-constant elements of $\mathbf{x}_i\mathbf{x}_i'$; then $nR^2 \xrightarrow{d} \chi^2[\dim\boldsymbol{\psi}_i]$ under $H_0$.

## GLS, WLS

Model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathrm{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = 0$, $\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}] = \boldsymbol{\Sigma}_{n\times n}$ (known). OLS estimator has properties $\mathrm{E}[\hat{\boldsymbol{\beta}}_{OLS} \mid \mathbf{X}] = \boldsymbol{\beta}$ and $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{OLS} \mid \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$.
Generalized least squares (GLS) estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$ with $\mathrm{E}[\hat{\boldsymbol{\beta}} \mid \mathbf{X}] = \boldsymbol{\beta}$, $\mathrm{Var}[\hat{\boldsymbol{\beta}}] = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$; this estimator is BLUE for this model.
**Th.**: OLS ~ GLS if (❖) $\boldsymbol{\Sigma}\mathbf{X} = \mathbf{X}\mathbf{B}$ for some non-singular $\mathbf{B}$; (❖) $\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z} = 0$ for $\forall \mathbf{Z} : \mathbf{Z}'\mathbf{X} = \mathbf{0}$; (❖) $\boldsymbol{\Sigma} = \mathbf{X}\boldsymbol{\Gamma}\mathbf{X}' + \mathbf{Z}\boldsymbol{\Theta}\mathbf{Z}' + \sigma^2\mathbf{I}$ for some $\sigma^2, \boldsymbol{\Gamma}, \boldsymbol{\Theta}, \mathbf{Z} : \mathbf{Z}'\mathbf{X} = \mathbf{0}$.

### *Conditional heteroscedasticity*

Assumption: $\boldsymbol{\Sigma} \mid \mathbf{X} = \mathrm{diag}[\sigma^2[\mathbf{x}_1], \dots, \sigma^2[\mathbf{x}_n]]$; $\mathbf{Q}_{xx\sigma^{-2}} \equiv \mathrm{E}[\mathbf{x}_i\mathbf{x}_i' / \sigma^2[\mathbf{x}_i]] > 0$. Then $\hat{\boldsymbol{\beta}}_{GLS} = \left(\frac{1}{n}\sum \frac{\mathbf{x}_i\mathbf{x}_i'}{\sigma^2[\mathbf{x}_i]}\right)^{-1}\left(\frac{1}{n}\sum \frac{\mathbf{x}_i y_i}{\sigma^2[\mathbf{x}_i]}\right)$, $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) \xrightarrow{d} N[\mathbf{0}, \mathbf{Q}_{xx\sigma^{-2}}^{-1}]$. Feasible GLS: either estimate $\sigma^2[\mathbf{x}]$ non-parametrically or run auxiliary regression $\hat{\boldsymbol{\varepsilon}}^2 \rhd \mathbf{Z}$ and then use $\hat{\sigma}^2[\mathbf{x}_i] = \mathbf{z}_i'\hat{\boldsymbol{\gamma}}_i$: $\hat{\boldsymbol{\beta}}_{FGLS} = \left(\frac{1}{n}\sum \frac{\mathbf{x}_i\mathbf{x}_i'}{\hat{\sigma}^2[\mathbf{x}_i]}\right)^{-1}\left(\frac{1}{n}\sum \frac{\mathbf{x}_i y_i}{\hat{\sigma}^2[\mathbf{x}_i]}\right)$.

## IV

Model: $\mathbf{y}_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t$, where $\{y_t, \underset{k\times 1}{\mathbf{x}_t}, \underset{l\times 1}{\mathbf{z}_t}\}$ is stationary&ergodic, $\mathrm{E}[\mathbf{z}_t\varepsilon_t] = 0$, $\mathbf{Q}_{zx} = \mathrm{E}[\mathbf{z}_t\mathbf{x}_t']$, rank condition for ID: $\mathrm{rg}\,\mathbf{Q}_{zx} = k$, order condition for ID: $l \geq k$.
Estimator would be asy.normal when $\{\mathbf{z}_t\varepsilon_t\}$ is mds and $\mathbf{Q}_{zz\varepsilon^2} = \mathrm{E}[\varepsilon_t^2\mathbf{z}_t\mathbf{z}_t'] > 0$.

## GMM

Model:
Special cases: OLS: $\mathrm{E}[\mathbf{x}_i\varepsilon_i] = 0$, WLS: $\mathrm{E}[\mathbf{x}_i\,\sigma^{-2}[\mathbf{x}_i]\varepsilon_i] = 0$

## SUR

Model: $\underset{T\times 1}{\mathbf{y}_i} = \underset{T\times k_i}{\mathbf{X}_i}\underset{k_i\times 1}{\boldsymbol{\beta}_i} + \underset{T\times 1}{\mathbf{u}_i}$, $i = 1,\dots,N$, where $\mathrm{E}[u_{it} \mid \mathbf{X}] = 0$, $\mathrm{E}[u_{it}u_{jt} \mid \mathbf{X}] = \sigma_{ij}$; in stacked form: $\underset{TN\times 1}{\mathbf{y}} = \underset{TN\times K}{\underline{\mathbf{X}}}\underset{K\times 1}{\boldsymbol{\beta}} + \underset{TN\times 1}{\mathbf{u}}$, where $\underline{\mathbf{X}} = \mathrm{diag}[\mathbf{X}_1, \dots, \mathbf{X}_N]$, $K = k_1 + \dots + k_N$.
GLS estimator: $\hat{\boldsymbol{\beta}} = (\underline{\mathbf{X}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\underline{\mathbf{X}})^{-1}\underline{\mathbf{X}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{y}$ (if $\boldsymbol{\Sigma}$ known). FGLS estimator builds upon $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i$ (here $\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i$), $\hat{\sigma}_{ij} = \frac{1}{T}\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j \Rightarrow \hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Sigma}} \otimes \mathbf{I}_T$. Asymptotic distribution: $\sqrt{T}(\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}) \xrightarrow{d} \sqrt{T}(\hat{\boldsymbol{\beta}}_{FGLS} - \boldsymbol{\beta}) \xrightarrow{d} N[\mathbf{0}, \left(\mathrm{plim}\,\frac{1}{T}\underline{\mathbf{X}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\underline{\mathbf{X}}\right)^{-1}]$.
**Th.**: OLS ~ GLS if (❖) $\boldsymbol{\Sigma}$ is diagonal; (❖) $\mathbf{X}_1 = \mathbf{X}_2 = \dots = \mathbf{X}_N$.

## SEM

Model: $\underset{1\times N}{\mathbf{y}'_{(t)}}\underset{N\times N}{\mathbf{\Gamma}} = \underset{1\times K}{\mathbf{x}'_{(t)}}\underset{K\times N}{\mathbf{B}} + \underset{1\times N}{\mathbf{u}'_{(t)}}$ ; in stacked form: $\underset{T\times N}{\mathbf{Y}}\underset{N\times N}{\mathbf{\Gamma}} = \underset{T\times K}{\mathbf{X}}\underset{K\times N}{\mathbf{B}} + \underset{T\times N}{\mathbf{U}}$ ; assumptions: $\mathbf{u}_{(t)} \sim iid[\mathbf{0}, \underset{N\times N}{\mathbf{\Sigma}}]$ , $\Pr[\operatorname{rg}\mathbf{X} = K] = 1$ , $\exists \operatorname{plim} \frac{1}{T}\mathbf{X}'\mathbf{X} > 0$ , $\det \mathbf{\Gamma} \neq 0$ , $\mathbf{\Gamma}_{ii} = 1$ .

Reduced form: $\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}$ , where $\mathbf{\Pi} = \mathbf{B\Gamma}^{-1}$ , $\mathbf{V} = \mathbf{U\Gamma}^{-1}$ , $\mathbf{\Lambda} = \mathbf{\Gamma}'^{-1}\mathbf{\Sigma\Gamma}^{-1}$ . Denote $\boldsymbol{\gamma}_i = \operatorname{grep}\{-\mathbf{\Gamma}_{ji} : j \neq i \wedge \mathbf{\Gamma}_{ji} \neq 0\}_{j=1}^{N}$ , $\mathbf{Y}_i = \operatorname{grep}\{\mathbf{y}_j : j \neq i \wedge \mathbf{\Gamma}_{ji} \neq 0\}_{j=1}^{N}$ ,

$\boldsymbol{\beta}_i = \operatorname{grep}\{\mathbf{B}_{ji} : \mathbf{B}_{ji} \neq 0\}_{j=1}^{K}$ , $\mathbf{X}_i = \operatorname{grep}\{\mathbf{x}_j : \mathbf{B}_{ji} \neq 0\}_{j=1}^{K}$ , $\mathbf{Z}_i = (\mathbf{Y}_i \;\; \mathbf{X}_i)$ , $\boldsymbol{\alpha}_i = (\boldsymbol{\gamma}'_i \;\; \boldsymbol{\beta}'_i)'$ , $L_i = N_i + K_i$ , $\mathbf{y} = (\mathbf{y}'_1 \; \ldots \; \mathbf{y}'_N)'$ , $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1 \; \ldots \; \boldsymbol{\alpha}'_N)'$ , $\underline{\mathbf{Z}} = \operatorname{diag}[\mathbf{Z}_1, \ldots, \mathbf{Z}_N]$ .

Conventional form: $\underset{T\times 1}{\mathbf{y}_i} = \underset{T\times N_i}{\mathbf{Y}_i}\underset{N_i\times 1}{\boldsymbol{\gamma}_i} + \underset{T\times K_i}{\mathbf{X}_i}\underset{K_i\times 1}{\boldsymbol{\beta}_i} + \underset{T\times 1}{\mathbf{u}_i} \equiv \underset{T\times L_i}{\mathbf{Z}_i}\underset{L_i\times 1}{\boldsymbol{\alpha}_i} + \mathbf{u}_i$ ; in stacked form: $\underset{NT\times 1}{\mathbf{y}} = \underset{NT\times \Sigma L_i}{\mathbf{Z}}\underset{\Sigma L_i\times 1}{\boldsymbol{\alpha}} + \underset{NT\times 1}{\mathbf{u}}$ , where $\mathbf{\Omega} \equiv \mathrm{E}[\mathbf{uu}'] = \mathbf{\Sigma} \otimes \mathbf{I}_T$ .

### *Identification*

Write $\mathbf{\Pi\Gamma} = \mathbf{B}$ as $\underset{K_i\times 1}{\boldsymbol{\pi}^{i1}} - \underset{K_i\times N_i}{\mathbf{\Pi}^{i1}}\underset{N_i\times 1}{\boldsymbol{\gamma}_i} = \underset{K_i\times 1}{\boldsymbol{\beta}_i} \wedge \underset{K_{(i)}\times 1}{\boldsymbol{\pi}^{i0}} - \underset{K_{(i)}\times N_i}{\mathbf{\Pi}^{i0}}\underset{N_i\times 1}{\boldsymbol{\gamma}_i} = \underset{K_{(i)}\times 1}{\mathbf{0}}$ , where $\boldsymbol{\pi}^{i(0/1)} = \operatorname{grep}\{\mathbf{\Pi}_{ji} : \mathbf{B}_{ji}(= / \neq)0\}_{j=1}^{K}$ , $\mathbf{\Pi}^{i(0/1)} = \operatorname{grep}\{\mathbf{\Pi}_{jk} : \mathbf{B}_{ji}(= / \neq)0 \wedge k \neq i \wedge \mathbf{\Gamma}_{ki} \neq 0\}_{j=1}^{K}{}_{k=1}^{N}$ .

In words, $\mathbf{\Pi}^{i0}$ consists lies in intersection of those columns of $\mathbf{\Pi}$ which correspond to included (in i$^{\text{th}}$ equation) endogenous variables and those rows which correspond to excluded exogenous variables.
Order condition: $K_{(i)} \geq N_i$ , rank condition: $\operatorname{rg}\mathbf{\Pi}^{i0} = N_i$ .

### *Full Information Model*

$\ln \mathfrak{L} = -\frac{1}{2}NT\ln[2\pi] + T\ln\|\mathbf{\Gamma}\| - \frac{1}{2}T\ln|\mathbf{\Sigma}| - \frac{1}{2}\operatorname{tr}\left[\mathbf{\Sigma}^{-1}(\mathbf{Y\Gamma} - \mathbf{XB})'(\mathbf{Y\Gamma} - \mathbf{XB})\right]$ , concentrated log-l: $\ln \mathfrak{L}^* = -\frac{1}{2}T\ln\left|(\mathbf{Y} - \mathbf{XB\Gamma}^{-1})'(\mathbf{Y} - \mathbf{XB\Gamma}^{-1})\right| \Rightarrow \hat{\boldsymbol{\alpha}}_{FIML}$ .

3SLS: 1,2) obtain $\hat{\boldsymbol{\alpha}}_{i,2SLS}$ for $i = 1, \ldots, N$ , estimate $\hat{\sigma}_{ij} = \frac{1}{T}\hat{\mathbf{u}}'_i\hat{\mathbf{u}}_j$ , where $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{Z}_i\hat{\boldsymbol{\alpha}}_{i,2SLS}$ , 3) find $\hat{\boldsymbol{\alpha}}_{3SLS} = \left(\hat{\mathbf{Z}}'(\hat{\mathbf{\Sigma}}^{-1} \otimes \mathbf{I})\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'(\hat{\mathbf{\Sigma}}^{-1} \otimes \mathbf{I})\mathbf{y} \overset{A}{=} \hat{\boldsymbol{\alpha}}_{FIML}$ .

### *Limited Information Model*

Model: $\mathbf{y}_j = \mathbf{Y}_j\boldsymbol{\gamma}_j + \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{u}_j \equiv \mathbf{Z}_j\boldsymbol{\alpha}_j + \mathbf{u}_j$ , $\mathbf{Y}_j = \mathbf{X}\mathbf{\Pi}_j + \mathbf{V}_j$ , where $(u_{jt}, \mathbf{V}_{jt}) \sim iidN[\mathbf{0}, \mathbf{\Sigma} = \{\sigma_j^2 \vdots \mathbf{\Sigma}_{j2} \vdots \mathbf{\Sigma}_{22}\}]$ , and $rg\mathbf{\Pi}^{j0} = N_j$ .

LIML estimator: $\hat{\boldsymbol{\alpha}}_{j,LIML} = \left(\mathbf{Z}'_j(\mathbf{I} - \lambda\mathbf{M})\mathbf{Z}_j\right)^{-1}\mathbf{Z}'_j(\mathbf{I} - \lambda\mathbf{M})\mathbf{y}_j$ , where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , $\mathbf{M}_j = \mathbf{I} - \mathbf{X}_j(\mathbf{X}'_j\mathbf{X}_j)^{-1}\mathbf{X}'_j$ , $\mathbf{W} = (\mathbf{y}_j \;\; \mathbf{Y}_j)'\mathbf{M}(\mathbf{y}_j \;\; \mathbf{Y}_j)$ ,

$\mathbf{W}_j = (\mathbf{y}_j \;\; \mathbf{Y}_j)'\mathbf{M}_j(\mathbf{y}_j \;\; \mathbf{Y}_j)$ , and $\lambda$ is smallest characteristic root of $\mathbf{W}_j\mathbf{W}^{-1}$ . 2SLS estimator: $\hat{\boldsymbol{\alpha}}_{j,2SLS} = (\mathbf{Z}'_j\mathbf{P}\mathbf{Z}_j)^{-1}\mathbf{Z}'_j\mathbf{P}\mathbf{y}_j$ , where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ .

Denote $\mathbf{A} = \begin{pmatrix} \mathbf{\Pi}^{j1\prime} & \mathbf{\Pi}^{j0\prime} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \operatorname{plim}\left[\frac{1}{T}\mathbf{X}'\mathbf{X}\right] \begin{pmatrix} \mathbf{\Pi}^{j1} & \mathbf{I} \\ \mathbf{\Pi}^{j0} & \mathbf{0} \end{pmatrix}$ , then $\sqrt{T}\left(\hat{\boldsymbol{\alpha}}_{LIML} - \boldsymbol{\alpha}_j\right) \xrightarrow{d} \sqrt{T}\left(\hat{\boldsymbol{\alpha}}_{2SLS} - \boldsymbol{\alpha}_j\right) \xrightarrow{d} N[\mathbf{0}, \sigma_j^2\mathbf{A}^{-1}]$ . Interpretation of 2SLS: 1) $\mathbf{Y}_j \rhd \mathbf{X} \Rightarrow \hat{\mathbf{Y}}_j$ , 2) $\mathbf{y}_j \rhd \hat{\mathbf{Y}}_j, \mathbf{X}_j \Rightarrow \hat{\boldsymbol{\alpha}}_{2SLS}$ . $\widehat{\mathrm{AVar}}[\hat{\boldsymbol{\alpha}}_{j,2SLS}] = \left(\frac{1}{T}(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\alpha}}_j)'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\alpha}}_j)\right)\left(\frac{1}{T}\mathbf{Z}'_j\mathbf{P}\mathbf{Z}_j\right)^{-1}$ .

## NLLS

Model:

## Binary choice models

Model: $\Pr[y_i = 1] = F[\mathbf{x}'_i\boldsymbol{\beta}]$ , where $F'[x] = f[x] > 0 \; \forall x$ , $\exists f'$ , $\mathbf{x}_i \sim iid$ , $\mathrm{E}[\mathbf{x}_i\mathbf{x}'_i] > 0$ . Special cases: linear probability model: $F[x] = x$ , probit model: $F[x] = \Phi[x] = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$ , logit model: $F[x] = \Lambda[x] = (1 + e^{-x})^{-1}$ . Log-likelihood: $\ln \mathfrak{L} = \sum_{i=1}^{n}\left(y_i\ln F[\mathbf{x}'_i\boldsymbol{\beta}] + (1 - y_i)\ln[1 - F[\mathbf{x}'_i\boldsymbol{\beta}]]\right)$ . This function is globally concave for logit and probit specifications. $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N[\mathbf{0}, \mathfrak{I}^{-1}]$ , where $\mathfrak{I} = \lim_{n\to\infty}\mathrm{E}[-\frac{1}{n}\frac{\partial^2\ln\mathfrak{L}[\boldsymbol{\beta}]}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}] = \lim_{n\to\infty}\frac{1}{n}\sum\frac{f[\mathbf{x}'_i\boldsymbol{\beta}]^2}{F[\mathbf{x}'_i\boldsymbol{\beta}](1 - F[\mathbf{x}'_i\boldsymbol{\beta}])}\mathbf{x}_i\mathbf{x}'_i$ .

## Tobit

Type-I model: $\{y_i^* = \mathbf{x}'_i\boldsymbol{\beta} + u_i, \; y_i = \max[y_i^*, 0]\}, \; i = 1, \ldots, n$ ; assumptions: $u_i \sim iidN[0, \sigma^2]$ , observed: $\{y_i, \mathbf{x}_i\}$ , $\mathbf{x}_i \sim iid$ , $\mathrm{E}[\mathbf{x}_i\mathbf{x}'_i] > 0$ . Likelihood function: $\mathfrak{L} = \prod_0\left(1 - \Phi[\mathbf{x}'_i\boldsymbol{\beta}/\sigma]\right)\prod_1\sigma^{-1}\phi[(y_i - \mathbf{x}'_i\boldsymbol{\beta})/\sigma]$ . Truncated model: data for $y_i^* < 0$ unobserved $\Rightarrow \mathfrak{L} = \prod_1\Phi[\mathbf{x}'_i\boldsymbol{\beta}/\sigma]^{-1}\sigma^{-1}\phi[(y_i - \mathbf{x}'_i\boldsymbol{\beta})/\sigma]$ . Denote $\lambda[z] = \frac{\phi[z]}{\Phi[z]}$ .
**Heckman two-step**: 1) estimate $\boldsymbol{\alpha} = \boldsymbol{\beta}/\sigma$ in probit $\Pr[y_i > 0] = \Phi[\mathbf{x}'_i\boldsymbol{\alpha}]$ by MLE, 2) regress $y_i \rhd \mathbf{x}_i, \lambda[\mathbf{x}'_i\hat{\boldsymbol{\alpha}}]$ using sample $y_i > 0$ . Std.errors must be computed with White's HSCE formula. **NLS**: apply to $y_i = \mathbf{x}'_i\boldsymbol{\beta} + \sigma\lambda[\mathbf{x}'_i\boldsymbol{\beta}/\sigma] + \boldsymbol{\varepsilon}_i$ . **NLWLS**: apply to same eq'n with $\mathrm{Var}[\varepsilon_i \,|\, \mathbf{x}_i] = \sigma^2\left(1 - \mathbf{x}'_i\boldsymbol{\alpha}\lambda[\mathbf{x}'_i\boldsymbol{\alpha}] - \lambda[\mathbf{x}'_i\boldsymbol{\alpha}]^2\right)$ . **MLE**: log-likelihood globally concave in terms of $\boldsymbol{\alpha}$ and $\sigma^{-1}$ . All estimators are consistent if data serially correlated, but inconsistent under heteroscedasticity or non-normality in error term.
Type-II model: $\{y_{1i}^* = \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + u_{1i}, \; y_{2i} = y_{1i}^* > 0 \, ? \, \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + u_{2i} : 0\}, \; i = 1, \ldots, n$ ; assumptions: $(u_{1i} \;\; u_{2i}) \sim N[\mathbf{0}, (\sigma_1^2 \vdots \sigma_2^2 \vdots \rho\sigma_1\sigma_2)]$ , observed: $\{y_{2i}, \operatorname{sign} y_{1i}^*, \mathbf{x}_{1i}, \mathbf{x}_{2i}\}$ .
Likelihood function: $\mathfrak{L} = \prod_{y_{2i}=0}\left(1 - \Phi[\mathbf{x}'_{1i}\boldsymbol{\beta}_1/\sigma_1]\right)\prod_{y_{2i}\neq 0}\Phi\left[\frac{1}{\sqrt{1-\rho^2}}\left(\mathbf{x}'_{1i}\boldsymbol{\beta}_1/\sigma_1 + \rho(y_{2i} - \mathbf{x}'_{2i}\boldsymbol{\beta}_2)/\sigma_2\right)\right]\sigma_2^{-1}\phi\left[(y_{2i} - \mathbf{x}'_{2i}\boldsymbol{\beta}_2)/\sigma_2\right]$ . If there are no constraints on parameters, then $\sigma_1$ is unidentified $\boldsymbol{\alpha}_1 = \boldsymbol{\beta}_1/\sigma_1$ . **Heckman two-step**: $y_{2i} = \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + \rho\sigma_2\lambda[\mathbf{x}'_{1i}\boldsymbol{\alpha}_1] + \varepsilon_{2i}$ ( $y_{2i} \neq 0$ ), $\mathrm{Var}\,\varepsilon_{2i} = \sigma_2^2\left(1 - \rho^2\mathbf{x}'_{1i}\boldsymbol{\alpha}_1\lambda[\mathbf{x}'_{1i}\boldsymbol{\alpha}_1] - \rho^2\lambda[\mathbf{x}'_{1i}\boldsymbol{\alpha}_1]^2\right)$ .
Type-III model: $\{y_{1i}^* = \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + u_{1i}, \; y_{1i} = \max[y_{1i}^*, 0], \; y_{2i} = y_{1i}^* > 0 \, ? \, \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + u_{2i} : 0\}, \; i = 1, \ldots, n$ , $\mathfrak{L} = \prod_{y_{1i}=0}\Pr[y_{1i}^* \leq 0] \cdot \prod_{y_{1i}>0} f[y_{1i}, y_{2i}]$ .

Type-IV model: $\{y_{1i}^* = \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + u_{1i}, \; y_{1i} = \max[y_{1i}^*, 0], \; y_{2i} = y_{1i}^* > 0 \, ? \, \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + u_{2i} : 0, \; y_{3i} = y_{1i}^* \leq 0 \, ? \, \mathbf{x}'_{3i}\boldsymbol{\beta}_3 + u_{3i} : 0\}$ , $\mathfrak{L} = \prod_{y_{1i}=0}\int_{-\infty}^{0} f_{13}[y_{1i}^*, y_{3i}]dy_{1i}^* \cdot \prod_{y_{1i}>0} f_{12}[y_{1i}, y_{2i}]$ .

Type-V model: $\{y_{1i}^* = \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + u_{1i}, \; y_{2i} = y_{1i}^* > 0 \, ? \, \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + u_{2i} : 0, \; y_{3i} = y_{1i}^* \leq 0 \, ? \, \mathbf{x}'_{3i}\boldsymbol{\beta}_3 + u_{3i} : 0\}, \; i = 1, \ldots, n$ , $\mathfrak{L} = \prod_{y_{2i}\neq 0}\int_{0}^{\infty} f_{12}[y_{1i}^*, y_{2i}]dy_{1i}^* \cdot \prod_{y_{3i}\neq 0}\int_{-\infty}^{0} f_{13}[y_{1i}^*, y_{3i}]dy_{1i}^*$ .

## Time series

Process $\{\mathbf{z}_t\}_{t=-\infty}^{\infty}$ is **strictly stationary** if $\forall t_0, \ldots, t_k$ $f[\mathbf{z}_{t_0}, \ldots, \mathbf{z}_{t_k}]$ depends only on $t_1 - t_0, \ldots, t_k - t_0$ but not on $t_0$ . It is **weakly stationary** (or **2$^{\text{nd}}$-order stationary**) if $\mathrm{E}\,\mathbf{z}_t = \boldsymbol{\mu} = \text{const}$ and $\mathrm{Cov}[\mathbf{z}_t, \mathbf{z}_{t-s}] = \mathbf{\Gamma}_s \equiv \mathbf{\Gamma}'_{-s}$ for $\forall t, s$ . For scalar processes, **autocorrelation function:** $\rho_k = \gamma_k/\gamma_0$ . Second-order stationary process is **white noise** if $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{\Gamma}_s = 0$ for $\forall s \neq 0$ . Process $\{\mathbf{z}_t\}$ is called **martingale** if $\mathrm{E}[\mathbf{z}_t \,|\, \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \ldots] = \mathbf{z}_{t-1}$ . Stationary process $\{\mathbf{z}_t\}$ is **ergodic** if for $\forall f : \mathbb{R}^{km} \to \mathbb{R}, \forall g : \mathbb{R}^{kl} \to \mathbb{R}$ : $\lim_{n\to\infty}\left|\mathrm{E}\left[f[\mathbf{z}_t, \ldots, \mathbf{z}_{t+m}] \cdot g[\mathbf{z}_{t+n}, \ldots, \mathbf{z}_{t+n+l}]\right]\right| = \left|\mathrm{E}\left[f[\mathbf{z}_t, \ldots, \mathbf{z}_{t+m}]\right] \cdot \mathrm{E}\left[g[\mathbf{z}_t, \ldots, \mathbf{z}_{t+l}]\right]\right|$ .

Sample autocovariance: $\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}(z_t - \bar{z})(z_{t-j} - \bar{z})$ , sample ACF: $\hat{\rho}_j = \hat{\gamma}_j / \hat{\gamma}_0$ . **Th.**: If $z_t = \mu + \varepsilon_t$ where $\varepsilon_t$ is stationary mds with $\mathrm{E}[\varepsilon_t^2 \mid \varepsilon_{t-1}, \varepsilon_{t-2}, ...] = \sigma^2$ , then: $\sqrt{T}\hat{\boldsymbol{\gamma}} \xrightarrow{d} N[\mathbf{0}, \sigma^4 \mathbf{I}_p]$ , $\sqrt{T}\hat{\boldsymbol{\rho}} \xrightarrow{d} N[\mathbf{0}, \mathbf{I}_p]$ where $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1 \ ... \ \hat{\gamma}_p)'$ , $\hat{\boldsymbol{\rho}} = (\hat{\rho}_1 \ ... \ \hat{\rho}_p)'$ . **Box-Pierce Q**: $T\sum_{j=1}^{p}\hat{\rho}_j^2 \xrightarrow{d} \chi^2[p]$ , **Ljung-Box Q**: $T\sum_{j=1}^{p}\frac{T+2}{T-j}\hat{\rho}_j^2 \xrightarrow{d} \chi^2[p]$ . Suppose $y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t$ , $\{y_t, \mathbf{x}_t\}$ is stationsry&ergodic, $\varepsilon_t \mid \varepsilon_{t-1}, ..., \mathbf{x}_{t-1}, ... \sim r.v[0, \sigma^2]$ and $\mathrm{E}[\mathbf{x}_t \mathbf{x}_t'] > 0$ . If we calculate: $\hat{\gamma}_j = \frac{1}{T}\sum_{t=1+j}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-j}$ , $\hat{\rho}_j = \hat{\gamma}_j / \hat{\gamma}_0$ then $\sqrt{T}\hat{\boldsymbol{\gamma}} \xrightarrow{d} N[\mathbf{0}, \sigma^4(\mathbf{I}_p - \boldsymbol{\Phi})]$ , $\sqrt{T}\hat{\boldsymbol{\rho}} \xrightarrow{d} N[\mathbf{0}, \mathbf{I}_p - \boldsymbol{\Phi}]$ , where $\boldsymbol{\Phi}_{jk} = \sigma^{-2}\,\mathrm{E}[\mathbf{x}_t\varepsilon_{t-j}]'\mathbf{Q}_{xx}^{-1}\,\mathrm{E}[\mathbf{x}_t\varepsilon_{t-k}]$ .