Formulas and Tables

for

Econometric Methods / Econometrics I

December 17, 2021

© Prof. Dr. Kai Carstensen
Institute for Statistics and Econometrics, CAU Kiel

1 Preliminaries

1.1 Conditional expectation and linear projection

Conditional expectation: Let y be a continuous random variable and let $\mathbf{x} \equiv (x_1, x_2, \ldots, x_K)$ be a $1 \times \mathbf{K}$ random vector of conditioning variables. Then the conditional expectation of y given $\mathbf{x}$ is

$$\mathrm{E}\,(y|x_1, \ldots, x_K) = \mathrm{E}\,(y|\mathbf{x}) = \int_{-\infty}^{\infty} y\, f_{y|\mathbf{x}}(y|\mathbf{x})\, dy.$$

Law of iterated expectations: Suppose y is a random variable, $\mathbf{w}$ is a random vector and $\mathbf{x} = \mathbf{f}(\mathbf{w})$ is a function of $\mathbf{w}$. Then

$$\mathrm{E}(y|\mathbf{x}) = \mathrm{E}\left[\mathrm{E}\,(y|\mathbf{w})\,|\mathbf{x}\right] \qquad \text{and} \qquad \mathrm{E}(y|\mathbf{x}) = \mathrm{E}\left[\mathrm{E}\,(y|\mathbf{x})\,|\mathbf{w}\right].$$

An important special case is $\mathrm{E}\left[\mathrm{E}(y|\mathbf{x})\right] = \mathrm{E}(y)$.

Linear projection: The linear projection of the random variable y on 1 and the $(1 \times K)$-dimensional random vector $\mathbf{x}$ is

$$L(y|1, \mathbf{x}) = \beta_0 + \mathbf{x}\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = [\mathrm{Var}(\mathbf{x})]^{-1}\,\mathrm{Cov}(\mathbf{x}', y)$ and $\beta_0 = \mathrm{E}[y] - \mathrm{E}[\mathbf{x}]\boldsymbol{\beta}$.

1.2 Mean value theorems

Mean value theorem (scalar version): Suppose that O is an open interval and $f : O \to \mathbb{R}$ is a differentiable function. Then for any $[a, b] \subset O$, there exists a $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)\ (b - a).$$

Mean value theorem (vector version): Suppose G is an open convex subset of $\mathbb{R}^P$ and $f : G \to \mathbb{R}$ is a differentiable function of P arguments with gradient vector $\nabla_{\mathbf{x}} f(\mathbf{x})$. Then for any vectors $\mathbf{a}, \mathbf{b} \in G$, there exists a vector $\mathbf{c}$ between $\mathbf{a}$ and $\mathbf{b}$ such that

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla_{\mathbf{x}} f(\mathbf{c})\ (\mathbf{b} - \mathbf{a}).$$

1.3 Vector differentiation

Gradient and Hessian: Let $f(\boldsymbol{\theta})$ be a twice continuously differentiable scalar function that depends on the $(n \times 1)$ vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$. The gradient is

$$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \left[\frac{\partial f(\boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_n}\right]$$

and the Hessian is

$$\frac{\partial^2 f(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \nabla_{\boldsymbol{\theta}} \left[\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})\right]' = \begin{pmatrix} \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_n} \\ \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_n} \\ \vdots & \vdots & \ddots & \\ \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_n \partial \theta_1} & \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_n \partial \theta_2} & \cdots & \frac{\partial^2 f(\mathbf{w},\boldsymbol{\theta})}{\partial \theta_n \partial \theta_n} \end{pmatrix}.$$

Differentiation of vector functions: Let $\mathbf{h}(\boldsymbol{\theta}) = [h_1(\boldsymbol{\theta}), \ldots, h_m(\boldsymbol{\theta})]'$ be a $(m \times 1)$ vector function that depends on the $(n \times 1)$ vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$. Then the $(m \times n)$ matrix of first partial derivatives is

$$\frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial h_1}{\partial \theta_1} & \cdots & \frac{\partial h_1}{\partial \theta_n} \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial \theta_1} & \cdots & \frac{\partial h_m}{\partial \theta_n} \end{bmatrix}.$$

Chain rule for vector differentiation: Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be $(m \times 1)$ and $(n \times 1)$ vectors, respectively, and suppose $h(\boldsymbol{\alpha})$ is $(p \times 1)$ and $g(\boldsymbol{\beta})$ is $(m \times 1)$. Further assume $\boldsymbol{\alpha} = g(\boldsymbol{\beta})$ and define $f(\boldsymbol{\beta}) = h(g(\boldsymbol{\beta}))$. Then

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \frac{\partial h(g(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}'} = \frac{\partial h(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}'} \frac{\partial g(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}.$$

Differentiation rules for specific scalar functions: Let $f(\boldsymbol{\theta}) = \boldsymbol{\theta}'\boldsymbol{\theta}$ be a scalar function of the $(n \times 1)$ vector $\boldsymbol{\theta}$. Then

$$\frac{\partial f}{\partial \boldsymbol{\theta}} = 2\boldsymbol{\theta} \qquad \text{and} \qquad \frac{\partial^2 f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = 2I_n.$$

Differentiation rules for linear vector functions: Let $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta} + \mathbf{b}$ be a $(m \times 1)$ vector function, $\mathbf{A}$ a $(m \times n)$ matrix, $\mathbf{b}$ a $(m \times 1)$ vector, and $\boldsymbol{\theta}$ a $(n \times 1)$ vector. Then

$$\frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \mathbf{A}.$$

1.4 Exact distributions

Linear function of normal random variables: Let $\mathbf{W}$ be a $N \times 1$ random vector with normal distribution $\mathbf{W} \sim \text{Normal}(\mathbf{0}, \mathbf{V})$, and let $\mathbf{A}$ be a $K \times N$ matrix of full row rank K and $\mathbf{B}$ a K vector, both of fixed parameters. Then the $K \times 1$ random vector $\mathbf{Z} = \mathbf{AW} + \mathbf{B}$ is normally distributed: $\mathbf{Z} \sim \text{Normal}(\mathbf{B}, \mathbf{AVA}')$.

Quadratic forms of normal random variables: For standard normally distributed random variables, $\mathbf{W} \sim \text{Normal}(\mathbf{0}, \mathbf{I}_N)$, the quadratic form $\mathbf{W}'\mathbf{CW}$, where $\mathbf{C}$ is a symmetric, idempotent matrix of fixed numbers with rank r, is χ^2 distributed with r degrees of freedom.

Student's t distribution: If the random variable Z has a standard normal distribution, the random variable W has a χ^2_r distribution, and Z and W are independently distributed, then the random variable $Z/\sqrt{W/r}$ is t distributed with r degrees of freedom, $\frac{Z}{\sqrt{W/r}} \sim t_r$.

2 Asymptotic theory

2.1 Convergence of deterministic sequences

Convergence and boundedness: (1) A sequence of nonrandom numbers $\{a_N : N = 1, 2, \ldots\}$ converges to a if for all $\varepsilon > 0$, there exists N_ε such that if $N > N_\varepsilon$ then $|a_N - a| < \varepsilon$. We write $a_N \to a$ as $N \to \infty$. (2) A sequence $\{a_N\text{: } N = 1, 2, \ldots\}$ is bounded if and only if there is some $b < \infty$ such that $|a_N| \leq b$ for all $N = 1, 2, \ldots$.

Order: (1) A sequence $\{a_N\}$ is $O(N^\lambda)$ (at most of order N^λ) if $N^{-\lambda}a_N$ is bounded. When $\lambda = 0, \{a_N\}$ is bounded, and we also write $a_N = O(1)$ (big oh one). (2) $\{a_N\}$ is $o(N^\lambda)$ if $N^{-\lambda}a_N \to 0$. When $\lambda = 0, a_N$, converges to zero, and we also write $a_N = o(1)$ (little oh one).

2.2 Convergence in probability

Probability limits: A sequence of random variables $\{x_N : N = 1, 2, \ldots\}$ converges in probability to the constant a, $x_N \xrightarrow{\text{p}} a$, if for all $\varepsilon > 0$, $\Pr[|x_N - a| > \varepsilon] \to 0$ as $N \to \infty$. This can be written x_N: plim $x_N = a$, where a is the probability limit of x_N. If $a = 0$, then $x_N = o_p(1)$.

Boundedness in probability: A sequence of random variables $\{x_N\}$ is bounded in probability, $x_N = \mathbf{O}_p(1)$, if and only if for every $\varepsilon > 0$, there exists a $b_\varepsilon < \infty$ and an integer N_ε such that $\Pr[|x_N| \geq b_\varepsilon] < \varepsilon$ for all $N \geq N_\varepsilon$.

Order in probability: (1) A random sequence $\{x_N\text{: } N = 1, 2, \ldots\}$ is $o_p(a_N)$, $x_N = o_p(a_N)$, where $\{a_N\}$ is a nonrandom, positive sequence, if $x_N/a_N = o_p(1)$. (2) A random sequence $\{x_N\text{: } N = 1, 2, \ldots\}$ is $\mathbf{O}_p(a_N)$, $x_N = \mathbf{O}_p(a_N)$, where $\{a_N\}$ is a nonrandom, positive sequence, if $x_N/a_N = \mathbf{O}_p(1)$.

Calculation rules: If $w_N = o_p(1), x_N = o_p(1), y_N = \mathbf{O}_p(1)$, and $z_N = \mathbf{O}_p(1)$, then (1) $w_N + x_N = o_p(1)$, (2) $y_N + z_N = \mathbf{O}_p(1)$, (3) $x_N + z_N = \mathbf{O}_p(1)$, (4) $w_N \cdot x_N = o_p(1)$, (5) $y_N \cdot z_N = \mathbf{O}_p(1)$, (6) $x_N \cdot z_N = o_p(1)$.

Slutsky's theorem: Let $\mathbf{g} : \mathbb{R}^K \to \mathbb{R}^J$ be a function continuous at some point $\mathbf{c} \in \mathbb{R}^K$. Let $\{\mathbf{x}_N : N = 1, 2, \ldots\}$ be a sequence of $K \times 1$ random vectors such that $\mathbf{x}_N \xrightarrow{\text{p}} \mathbf{c}$. Then $\mathbf{g}(\mathbf{x}_N) \xrightarrow{\text{p}} \mathbf{g}(\mathbf{c})$ as $N \to \infty$. Equivalently, plim $\mathbf{g}(\mathbf{x}_N) = \mathbf{g}(\text{plim } \mathbf{x}_N)$ if $\mathbf{g}(\cdot)$ is continuous at $\text{plim}\, \mathbf{x}_N$.

Uniform convergence: Let $\mathbf{w}$ be a random vector taking values in $\mathcal{W} \subset \mathbb{R}^M$, let $\mathbf{\Theta}$ be a subset of $\mathbb{R}^P$, and let $q : \mathcal{W} \times \mathbf{\Theta} \to \mathbb{R}$ be a real-valued function. Assume that (a) $\mathbf{\Theta}$ is compact; (b) for each $\boldsymbol{\theta} \in \mathbf{\Theta}$, $q(\cdot, \boldsymbol{\theta})$ is Borel measurable on $\mathcal{W}$; (c) for each $\mathbf{w} \in \mathcal{W}$, $q(\mathbf{w}, \cdot)$ is continuous on $\mathbf{\Theta}$; and (d) $|q(\mathbf{w}, \boldsymbol{\theta})| \leq b(\mathbf{w})$ for all $\boldsymbol{\theta} \in \mathbf{\Theta}$, where b is a nonnegative function on $\mathcal{W}$ such that $\mathrm{E}[b(\mathbf{w})] < \infty$. Then uniform convergence in probability holds:

$$\max_{\boldsymbol{\theta} \in \mathbf{\Theta}} \left| N^{-1} \sum_{i=1}^{N} q(\mathbf{w}_i, \boldsymbol{\theta}) - \mathrm{E}[q(\mathbf{w}, \boldsymbol{\theta})] \right| \xrightarrow{\text{p}} 0.$$

2.3 Convergence in distribution

Convergence in distribution: A sequence of random variables $\{x_N : N = 1, 2, \ldots\}$ converges in distribution to the continuous random variable x, $x_N \xrightarrow{\text{d}} x$, if and only if $F_N(\xi) \to F(\xi)$ as $N \to \infty$ for all $\xi \in \mathbb{R}$, where F_N is the cumulative distribution function (c.d.f.) of x_N and F is the (continuous) c.d.f. of x.

Continuous mapping theorem: Let $\{\mathbf{x}_N\}$ be a sequence of $K \times 1$ random vectors such that $\mathbf{x}_N \xrightarrow{\text{d}} \mathbf{x}$. If $\mathbf{g}\colon \mathbb{R}^K \to \mathbb{R}^J$ is a continuous function, then $\mathbf{g}(\mathbf{x}_N) \xrightarrow{\text{d}} \mathbf{g}(\mathbf{x})$.

The asymptotic normal distribution: Let $\{\mathbf{z}_N\}$ be a sequence of $K \times 1$ random vectors with asymptotic normal distribution, $\mathbf{z}_N \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{V})$. Then, for any $K \times M$ nonrandom matrix $\mathbf{A}$, $\mathbf{A}'\mathbf{z}_N \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A})$, and $\mathbf{z}_N'\mathbf{V}^{-1}\mathbf{z}_N \xrightarrow{\text{d}} \chi^2_K$.

Asymptotic equivalence lemma: Let $\{\mathbf{x}_N\}$ and $\{\mathbf{z}_N\}$ be sequences of $K \times 1$ random vectors. If $\mathbf{z}_N \xrightarrow{\text{d}} \mathbf{z}$ and $\mathbf{x}_N - \mathbf{z}_N \xrightarrow{\text{p}} \mathbf{0}$, then $\mathbf{x}_N \xrightarrow{\text{d}} \mathbf{z}$.

Cramer's theorem: Let $\{x_N\}$ and $\{b_N\}$ be sequences of scalar random variables. If $x_N \xrightarrow{\text{d}} x$ and $b_N \xrightarrow{\text{p}} b$, where x is a random variable and b is a constant, then (1) $x_N + b_N \xrightarrow{\text{d}} x + b$, (2) $x_N \cdot b_N \xrightarrow{\text{d}} x \cdot b$, and (3) $x_N / b_N \xrightarrow{\text{d}} x/b$, provided $b \neq 0$.

2.4 Limit theorems for random samples

Weak law of large numbers (WLLN): Let $\{\mathbf{w}_i : i = 1, 2, \ldots\}$ be a sequence of independent, identically distributed $G \times 1$ random vectors such that $\mathrm{E}(|w_{ig}|) < \infty, g = 1, \ldots, G$. Then the sequence satisfies the weak law of large numbers: $N^{-1} \sum_{i=1}^{N} \mathbf{w}_i \xrightarrow{\text{p}} \boldsymbol{\mu}_w \equiv \mathrm{E}(\mathbf{w}_i)$.

Lindeberg-Levy central limit theorem: Let $\{\mathbf{w}_i : i = 1, 2, \ldots\}$ be a sequence of independent, identically distributed $G \times 1$ random vectors such that $\mathrm{E}(w_{ig}^2) < \infty, g = 1, \ldots, G$, and $\mathrm{E}(\mathbf{w}_i) = \mathbf{0}$. Then $N^{-1/2} \sum_{i=1}^{N} \mathbf{w}_i \xrightarrow{\text{d}} \mathsf{Normal}(\mathbf{0}, \mathbf{B})$, where $\mathbf{B} = \mathrm{Var}(\mathbf{w}_i) = \mathrm{E}(\mathbf{w}_i \mathbf{w}_i')$ is necessarily positive semidefinite.

2.5 Asymptotic properties of estimators

Consistency: Let $\{\hat{\boldsymbol{\theta}}_N : N = 1, 2, \ldots,\}$ be a sequence of estimators of the $P \times 1$ vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where N indexes the sample size. If $\hat{\boldsymbol{\theta}}_N \xrightarrow{\text{p}} \boldsymbol{\theta}$ for any value of $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\theta}}_N$ is a consistent estimator of $\boldsymbol{\theta}$.

Asymptotic normality: Let $\{\hat{\boldsymbol{\theta}}_N \colon N = 1, 2, \ldots\}$ be a sequence of estimators of the $P \times 1$ vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Suppose that $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\text{d}}$ Normal$(\mathbf{0}, \mathbf{V})$, where $\mathbf{V}$ is a $P \times P$ positive semidefinite matrix. Then $\hat{\boldsymbol{\theta}}_N$ is $\sqrt{\mathbf{N}}$-asymptotically normally distributed and $\mathbf{V}$ is the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$, $\mathrm{Avar}[\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})] = \mathbf{V}$.

Approximate normality: Suppose that $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\text{d}}$ Normal$(\mathbf{0}, \mathbf{V})$. Then $\hat{\boldsymbol{\theta}}_N$ is said to be, for N not too small, approximately normally distributed, $\hat{\boldsymbol{\theta}}_N \overset{a}{\sim}$ Normal$(\boldsymbol{\theta}, \mathbf{V}_{\hat{\boldsymbol{\theta}}})$, where $\mathbf{V}_{\hat{\boldsymbol{\theta}}} = \mathbf{V}/N$.

Asymptotic efficiency: Let $\hat{\boldsymbol{\theta}}_N$ and $\tilde{\boldsymbol{\theta}}_N$ be two estimators of $\boldsymbol{\theta}$ satisfying, respectively, $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\text{d}}$ Normal$(\mathbf{0}, \mathbf{V})$ and $\sqrt{N}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\text{d}}$ Normal$(\mathbf{0}, \mathbf{D})$. Then $\hat{\boldsymbol{\theta}}_N$ is asymptotically efficient relative to $\tilde{\boldsymbol{\theta}}_N$ if $\mathbf{D} - \mathbf{V}$ is positive semidefinite for all $\boldsymbol{\theta}$.

2.6 Asymptotic properties of tests

Asymptotic size of a test: The asymptotic size of a testing procedure is defined as the limiting probability of rejecting H_0 when it is true, asymptotic size $= \lim_{N\to\infty} \mathrm{Pr}_N(\text{reject } H_0 | H_0)$, where the N subscript indexes the sample size.

Consistency of a test: A test is consistent against the alternative H_1 if the null hypothesis is rejected with probability approaching one when H_1 is true: $\lim_{N\to\infty} \mathrm{Pr}_N(\text{reject } H_0 | H_1) = 1$.

Asymptotic distribution of a t statistic: Let $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{\text{d}}$ Normal$(0, v^2)$. Under $H_0 : \theta = \theta_0$, $t = \sqrt{N}(\hat{\theta}_N - \theta_0)/\hat{v} \xrightarrow{\text{d}}$ Normal$(0, 1)$.

Asymptotic distribution of a Wald statistic: Suppose that $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{V})$. Under $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$, where $\boldsymbol{\theta}$ is a $P \times 1$ vector, $\mathbf{R}$ is a nonrandom $Q \times P$ matrix with $Q \leq P$ and $\text{rank}(\mathbf{R}) = Q$ and $\mathbf{r}$ is a nonrandom $Q \times 1$ vector, the Wald statistic $W_N \equiv [\mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r}]'[\mathbf{R}(\hat{\mathbf{V}}_N/N)\mathbf{R}']^{-1}[\mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r}] \xrightarrow{\text{d}} \chi^2_Q$, where $\hat{\mathbf{V}}_N$ is a consistent estimator of $\mathbf{V}$.

2.7 Delta method

Suppose the P-dimensional random variable $\hat{\boldsymbol{\theta}}$ has asymptotic normal distribution $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{\text{d}} \text{Normal}(0, \mathbf{V})$. Let $\mathbf{c} : \boldsymbol{\Theta} \to \mathbb{R}^Q$, $Q \leq P$, be a continuously differentiable function and assume that $\boldsymbol{\theta}$ is in the interior of the parameter space. Define the $Q \times P$ Jacobian $\mathbf{C}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\mathbf{c}(\boldsymbol{\theta})$. Then $\sqrt{N}[\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{c}(\boldsymbol{\theta}_o)] \xrightarrow{\text{d}} \text{Normal}(0, \mathbf{C}(\boldsymbol{\theta})\mathbf{V}\mathbf{C}(\boldsymbol{\theta})')$.

3 The multiple regression model

3.1 OLS estimator

Regression model in vector form:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i, \qquad i = 1, \ldots, N,$$

where $\mathbf{x}_i \equiv (x_{1i}, \ldots, x_{Ki})$ is a $(1 \times K)$ vector of explanatory variables and $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_K)'$ is a $(K \times 1)$ vector of parameters. Assume $x_{1i} \equiv 1$.

Regression model in matrix form: Stacking all N observations yields

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U},$$

where $\mathbf{Y} \equiv (y_1, \ldots, y_N)'$ is a $(N \times 1)$ vector, $\mathbf{X} \equiv (\mathbf{x}_1', \ldots, \mathbf{x}_N')'$ is a $(N \times K)$ matrix, and $\mathbf{U} \equiv (u_1, \ldots, u_N)'$ is a $(N \times 1)$ vector.

OLS estimator:

$$\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'y_i\right).$$

OLS residuals:

$$\hat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{LS} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{M}_\mathbf{X}\mathbf{Y} = \mathbf{M}_\mathbf{X}\mathbf{U},$$

where $\mathbf{M}_\mathbf{X} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is a symmetric, idempotent matrix of rank $N - K$.

Estimator of the error variance:

$$\hat{\sigma}^2 = \frac{1}{N-K}\sum_{i=1}^{N}\hat{u}_i^2 = \frac{1}{N-K}\mathbf{U}'\mathbf{M}_{\mathbf{X}}\mathbf{U}.$$

Measure of fit: Defining $\tilde{\mathbf{Y}}$ as the mean-adjusted observations, the R-squared is

$$R^2 = \frac{\hat{\tilde{\mathbf{Y}}}'\hat{\tilde{\mathbf{Y}}}}{\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}} = 1 - \frac{\hat{\mathbf{U}}'\hat{\mathbf{U}}}{\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}} = 1 - \frac{\sum_{i=1}^{N}\hat{u}_i^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2}.$$

Estimator of variance of the OLS estimator:

$$\widehat{\mathrm{Var}(\hat{\beta}_{LS})} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}.$$

3.2 Inference if X is fixed and U is normal

Assumptions: The regressor matrix $\mathbf{X}$ is a matrix of fixed numbers and has full column rank K. The disturbances u_i, $i = 1, \ldots, N$, are independently and identically normally distributed with $\mathrm{E}[u_i] = 0$ and $\mathrm{Var}[u_i] = \sigma^2$. Hence,

$$\mathbf{U} \sim \mathrm{Normal}(\mathbf{0}, \sigma^2\mathbf{I}_N).$$

Mean and variance of the OLS estimator:

$$\mathrm{E}[\hat{\boldsymbol{\beta}}_{LS}] = \boldsymbol{\beta}, \qquad \mathrm{Var}[\hat{\boldsymbol{\beta}}_{LS}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Distribution of $\hat{\sigma}^2$:

$$\frac{N-K}{\sigma^2}\times\hat{\sigma}^2 \sim \chi^2_{N-K}.$$

Distribution of the OLS estimator:

$$t_i = \frac{\hat{\beta}_{LS,i}-\beta_i}{\widehat{\mathrm{se}(\hat{\beta}_{LS,i})}} \sim t_{N-K},$$

where $\widehat{\mathrm{se}(\hat{\beta}_{LS,i})}$ is the i-th diagonal element of the root of the estimated variance of $\hat{\boldsymbol{\beta}}_{LS}$.

Confidence interval: A $(1-\alpha)100\%$ two-sided confidence interval for β_i is given by the set

$$\left[\hat{\beta}_i - \widehat{\mathrm{se}(\hat{\beta}_{LS,i})}t_{N-K,1-\alpha/2},\ \hat{\beta}_i + \widehat{\mathrm{se}(\hat{\beta}_{LS,i})}t_{N-K,1-\alpha/2}\right],$$

where $t_{N-K,1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the t distribution with $N-K$ degrees of freedom.

***t* test:** A t test with significance level α of the two-sided hypothesis $H_0 : \beta_i = \beta_{i,0}$ is decided by comparing the t statistic t_i and the $(1-\alpha/2)100\%$ quantile of the t_{N-K} distribution.

***p*-value:** The p-value for a two-sided t test is p-value $= 2\Pr(t_i > |t_i^{act}|)$, where t_i^{act} is the actual realization of the t statistic.

***F* test:** An F test with significance level α of the null hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, where $\mathbf{R}$ and $\mathbf{r}$ are fixed restriction matrices, is decided by comparing the F statistic

$$F = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})/q}{\hat{\sigma}^2}$$

and the $(1-\alpha)100\%$ quantile of the $F_{q,N-K}$ distribution.

3.3 Inference if y and $\mathbf{x}$ are sampled randomly

Assumptions: (OLS.1) $\mathrm{E}(\mathbf{x}'u) = \mathbf{0}$, (OLS.2) rank $E(\mathbf{x}'\mathbf{x}) = K$, (OLS.3) $\mathrm{E}(u^2\mathbf{x}'\mathbf{x}) = \sigma^2\,\mathrm{E}(\mathbf{x}'\mathbf{x})$, where $\sigma^2 = \mathrm{E}(u^2)$.

Consistency: Under Assumptions OLS.1 and OLS.2, OLS on a random sample is consistent: $\operatorname{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$.

Asymptotic normality: Under Assumptions OLS.1 and OLS.2 the OLS estimator is asymptotically normally distributed,

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathrm{p}} \mathrm{Normal}(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ with $\mathbf{A} = \mathrm{E}(\mathbf{x}'\mathbf{x})$ and $\mathbf{B} = \mathrm{E}(u^2\mathbf{x}'\mathbf{x})$.

Estimator of the asymptotic variance matrix: $\hat{\mathbf{V}} = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}$, where

$$\hat{\mathbf{A}} = N^{-1} \sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i \xrightarrow{\mathrm{p}} \mathbf{A}$$

and

$$\hat{\mathbf{B}} = (N-K)^{-1} \sum_{i=1}^{N} \hat{u}_i^2 \mathbf{x}_i'\mathbf{x}_i \xrightarrow{\mathrm{p}} \mathbf{B}.$$

Approximate standard errors for $\hat{\boldsymbol{\beta}}$: Take the square roots of the elements on the main diagonal of

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}} = \widehat{\mathrm{Avar}}(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{V}}/N = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}/N.$$

Wald test of linear hypotheses: To test the linear hypotheses Q $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ against $H_1 : \mathbf{R}\theta \neq \mathbf{r}$, the Wald statistic is (distribution under H_0)

$$W_N \equiv \left[\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right]' \left[\mathbf{R}(\hat{\mathbf{V}}/N)\mathbf{R}'\right]^{-1} \left[\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right] \overset{a}{\sim} \chi^2_Q.$$

3.4 Omitted variable bias

Population model including unobserved variable:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \gamma q + v$$

where $x_1, \ldots, x_K$ are observed, q is unobserved, and $\mathrm{E}\left(v | x_1, x_2, \ldots, x_K, q\right) = 0$.

Population model excluding unobserved variable:

$$y = \left(\beta_0 + \gamma\delta_0\right) + \left(\beta_1 + \gamma\delta_1\right) x_1 + \cdots + \left(\beta_K + \gamma\delta_K\right) x_K + v + \gamma r,$$

where δ_i, $i = 1, \ldots, K$, and r are defined by the linear projection of q onto the observable explanatory variables, $q = \delta_0 + \delta_1 x_1 + \ldots + \delta_K x_K + r$.

OLS omitted variables inconsistency: An OLS regression of y on $1, x_1, \ldots, x_K$ yields estimators $\hat{\beta}_i$, $i = 0, \ldots, K$, where

$$\operatorname{plim} \hat{\beta}_i = \beta_i + \gamma\delta_i.$$

3.5 Measurement error in an explanatory variable

Model (classical error-in-variables):

$$y = \beta_0 + \beta_1 x_1^* + v,$$

where v satisfies assumptions (OLS.1) and (OLS.2) but x_1^* is measured with error, $x_1 = x_1^* + e_1$ and $\mathrm{Cov}\left(x_1^*, e_1\right) = 0$.

Error-in-variables bias: An OLS regression of y on 1 and x_1 yields estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, where

$$\operatorname{plim} \hat{\beta}_1 = \beta_1 \, \frac{\mathrm{Var}(x_1^*)}{\mathrm{Var}(x_1^*) + \mathrm{Var}(e_1)}.$$

4 Instrumental variables estimation of a single equation

4.1 The baseline IV estimator

Population model: $y = \mathbf{x}\boldsymbol{\beta} + u$, where $\mathbf{x}$ is a $1 \times K$ vector of explanatory variables.

Assumptions: The $1 \times K$ vector of instruments satisfies $\mathrm{E}\left(\mathbf{z}'u\right) = \mathbf{0}$ and $\operatorname{rank}\left[\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\right] = K$.

IV estimator:

$$\hat{\boldsymbol{\beta}}_{IV} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'y_i\right)$$

Consistency: $\operatorname{plim}\hat{\boldsymbol{\beta}}_{IV} = \boldsymbol{\beta}$.

Asymptotic normality: $\sqrt{N}(\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}')^{-1})$, where $\mathbf{A} \equiv \mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)$ and $\mathbf{B} \equiv \mathrm{E}(u^2\mathbf{z}'\mathbf{z})$.

4.2 The 2SLS estimator

Population model: $y = \mathbf{x}\boldsymbol{\beta} + u$, where $\mathbf{x}$ is a $1 \times K$ vector of explanatory variables.

Assumptions: The $1 \times L$ vector of instruments satisfies $\mathrm{E}\left(\mathbf{z}'u\right) = \mathbf{0}$, $\operatorname{rank}\left[\mathrm{E}\left(\mathbf{z}'\mathbf{z}\right)\right] = L$, and $\operatorname{rank}\left[\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\right] = K$.

2SLS estimator:

$$\hat{\boldsymbol{\beta}}_{2SLS} = \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}.$$

Consistency: $\operatorname{plim}\hat{\boldsymbol{\beta}}_{2SLS} = \boldsymbol{\beta}$.

Asymptotic normality:

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} \mathrm{Normal}\left(\mathbf{0}, \left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1}\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{B}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1}\right),$$

where $\mathbf{A}_{xz} \equiv \mathrm{E}(\mathbf{x}'\mathbf{z})$, $\mathbf{A}_{zx} \equiv \mathrm{E}(\mathbf{z}'\mathbf{x})$, $\mathbf{A}_{zz} \equiv \mathrm{E}(\mathbf{z}'\mathbf{z})$, and $\mathbf{B} \equiv \mathrm{E}(u^2\mathbf{z}'\mathbf{z})$.

5 Generalized least squares

Population model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$.

Assumptions: The usual OLS assumptions such as $\mathrm{E}[\mathbf{u}|\mathbf{X}] = 0$ apply. However, it is allowed that $\mathrm{Var}[\mathbf{u}|\mathbf{X}] = \mathrm{E}[\mathbf{u}\mathbf{u}'|\mathbf{X}] = \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ might be nondiagonal and a function of $\mathbf{X}$.

GLS estimator: $\hat{\boldsymbol{\beta}}_{GLS} = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y}.$

FGLS estimator: $\hat{\boldsymbol{\beta}}_{FGLS} = \left(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Y}.$

6 M-estimation

6.1 The M-estimator

M-estimator: Consider the population parameter $\boldsymbol{\theta}_o$ that minimizes the population function

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathrm{E}\left[q(\mathbf{w}, \boldsymbol{\theta})\right].$$

It is estimated by the M-estimator $\hat{\boldsymbol{\theta}}$ that minimizes the sample function

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} N^{-1}\sum_{i=1}^{N} q(\mathbf{w}_i, \boldsymbol{\theta}).$$

Score: Suppose the objective function is once continuously differentiable with respect to $\boldsymbol{\theta}$. Then the score is

$$\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = \nabla'_{\boldsymbol{\theta}} q(\mathbf{w}, \boldsymbol{\theta}) = \left[\frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \theta_p}\right]'$$

First order condition:

$$\sum_{i=1}^{N} \mathbf{s}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) = \mathbf{0}.$$

Hessian: Suppose the objective function is twice continuously differentiable with respect to $\boldsymbol{\theta}$. Then the Hessian is

$$\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) = \frac{\partial^2 q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = \nabla_{\boldsymbol{\theta}} \mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = \begin{pmatrix} \nabla_{\boldsymbol{\theta}} s^{(1)}(\mathbf{w}, \boldsymbol{\theta}) \\ \vdots \\ \nabla_{\boldsymbol{\theta}} s^{(P)}(\mathbf{w}, \boldsymbol{\theta}) \end{pmatrix}$$

6.2 Asymptotic Properties

Consistency: Suppose the M-estimator is identified and uniform convergence of the sample average of $q(\mathbf{w}_i, \boldsymbol{\theta})$ holds. Then the M-estimator is consistent,

$$\hat{\boldsymbol{\theta}} \xrightarrow{\text{p}} \boldsymbol{\theta}_o.$$

Asymptotic normality: Suppose the M-estimator is identified and some regularity conditions are satisfied. Then the M-estimator is asymptotically normal,

$$N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}),$$

where $\mathbf{A}_o = \text{E}[\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}_o)]$ and $\mathbf{B}_o = \text{E}[\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)'] = \text{Var}[\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)]$.

6.3 Estimators of the Asymptotic Variance

Estimator of $\mathbf{A}_o$: Either use

$$\hat{\mathbf{A}} = N^{-1}\sum_{i=1}^{N}\mathbf{H}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) = N^{-1}\sum_{i=1}^{N}\hat{\mathbf{H}}_i$$

or

$$\hat{\mathbf{A}} = N^{-1}\sum_{i=1}^{N}\mathbf{A}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) = N^{-1}\sum_{i=1}^{N}\hat{\mathbf{A}}_i,$$

where $\mathbf{A}(\mathbf{x}_i, \boldsymbol{\theta}_o) = \text{E}[\mathbf{H}(\mathbf{w}_i, \boldsymbol{\theta}_o)|\mathbf{x}]$.

Estimator of $\mathbf{B}_o$:

$$\hat{\mathbf{B}} = N^{-1}\sum_{i=1}^{N}\mathbf{s}(\mathbf{w}_i, \hat{\boldsymbol{\theta}})\mathbf{s}(\mathbf{w}_i, \hat{\boldsymbol{\theta}})' = N^{-1}\sum_{i=1}^{N}\hat{\mathbf{s}}_i\hat{\mathbf{s}}_i'.$$

Estimator of $\mathbf{V}_o$:

$$\hat{\mathbf{V}} = \widehat{\text{Avar}}[\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)] = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}.$$

6.4 Generalized information matrix equality

The generalized information matrix equality means

$$\text{E}[\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)'] = \sigma_o^2\,\text{E}[\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}_o)], \qquad \sigma_o^2 > 0.$$

6.5 Nonlinear Least Squares

Assumptions. NLS.1: For some $\boldsymbol{\theta}_o \in \boldsymbol{\Theta}$, $\mathrm{E}(y|\mathbf{x}) = m(\mathbf{x}, \boldsymbol{\theta}_o)$. NLS.2: $\mathrm{E}\left\{[m(\mathbf{x}, \boldsymbol{\theta}_o) - m(\mathbf{x}, \boldsymbol{\theta})]^2\right\} > 0$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_o$. NLS.3: $\mathrm{Var}(y|\mathbf{x}) = \mathrm{Var}(u|x) = \sigma_o^2$.

NLS estimator: Suppose $\mathrm{E}(y|\mathbf{x}) = m(\mathbf{x}, \boldsymbol{\theta}_o)$ and thus $y = m(\mathbf{x}, \boldsymbol{\theta}_o) + u$, where $\mathrm{E}(u|\mathbf{x}) = 0$. The NLS estimator minimizes

$$N^{-1} \sum_{i=1}^{N} [y_i - m(\mathbf{x}_i, \boldsymbol{\theta})]^2/2.$$

In terms of the M-estimator, the objective function specializes to $q(\mathbf{w}, \boldsymbol{\theta}) = [y - m(\mathbf{x}, \boldsymbol{\theta})]^2/2$.

NLS score: Suppose $m(\mathbf{x}, \boldsymbol{\theta}_o)$ is once continuously differentiable with respect to $\boldsymbol{\theta}$. Then the score is

$$\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = \nabla'_{\boldsymbol{\theta}} q(\mathbf{w}, \boldsymbol{\theta}) = -u \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

NLS Hessian: Suppose $m(\mathbf{x}, \boldsymbol{\theta}_o)$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$. Then the Hessian is

$$\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) = -u \frac{\partial^2 m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} + \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}.$$

NLS variance matrix: The asymptotic variance matrix is

$$\mathbf{V}_o = \mathbf{A}_o^{-1} \mathbf{B}_o \mathbf{A}_o^{-1},$$

where the matrix $\mathbf{A}_o$ becomes

$$\mathbf{A}_o = \mathrm{E}[\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}_o)] = \mathrm{E}\left[\frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]$$

and the matrix $\mathbf{B}_o$ becomes

$$\mathbf{B}_o = \mathrm{E}[\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)'] = \mathrm{E}\left[u^2 \frac{\partial m(\mathbf{x}, \boldsymbol{\theta}_o)}{\partial \boldsymbol{\theta}} \frac{\partial m(\mathbf{x}, \boldsymbol{\theta}_o)}{\partial \boldsymbol{\theta}'}\right].$$

Homoscedasticity: Under homoscedasticity, i.e., under Assumption NLS.3, $\mathrm{Var}(y|\mathbf{x}) = \mathrm{Var}(u|x) = \sigma_o^2$, the generalized information matrix equality holds. Hence, $\mathbf{V}_o = \sigma_o^2 \mathbf{A}_o^{-1}$.

6.6 Inference

Wald test of linear hypotheses: To test the linear hypotheses Q $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ against $H_1 : \mathbf{R}\theta \neq \mathbf{r}$, the Wald statistic is (distribution under H_0)

$$W_N \equiv \left[\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right]' \left[\mathbf{R}(\hat{\mathbf{V}}/N)\mathbf{R}'\right]^{-1} \left[\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right] \overset{a}{\sim} \chi_Q^2.$$

Wald test of nonlinear hypotheses: To test the Q nonlinear hypotheses $H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ against $H_1 : \mathbf{c}(\boldsymbol{\theta}) \neq \mathbf{0}$, the Wald statistic is (distribution under H_0)

$$W_N \equiv \mathbf{c}(\hat{\boldsymbol{\theta}})' \left[\mathbf{C}(\hat{\boldsymbol{\theta}})(\hat{\mathbf{V}}/N)\mathbf{C}(\hat{\boldsymbol{\theta}})'\right]^{-1} \mathbf{c}(\hat{\boldsymbol{\theta}}) \overset{a}{\sim} \chi^2_Q.$$

6.7 Optimization Methods

Newton-Raphson method: To find the coefficient vector $\hat{\boldsymbol{\theta}}$ that is the root of the FOC

$$\sum_{i=1}^{N} \mathbf{s}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) = \mathbf{0},$$

the Newton-Raphson method uses the iteration

$$\boldsymbol{\theta}^{\{g+1\}} = \boldsymbol{\theta}^{\{g\}} - \left[\sum_{i=1}^{N} \mathbf{H}(\mathbf{w}_i, \boldsymbol{\theta}^{\{g\}})\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{s}(\mathbf{w}_i, \boldsymbol{\theta}^{\{g\}})\right].$$

Generalized Gauss-Newton method: The generalized Gauss-Newton method uses the iteration

$$\boldsymbol{\theta}^{\{g+1\}} = \boldsymbol{\theta}^{\{g\}} - r \left[\sum_{i=1}^{N} \mathbf{A}(\mathbf{x}_i, \boldsymbol{\theta}^{\{g\}})\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{s}(\mathbf{w}_i, \boldsymbol{\theta}^{\{g\}})\right].$$

BHHH method: The Berndt, Hall, Hall, and Hausman (BHHH) algorithm uses the iteration

$$\boldsymbol{\theta}^{\{g+1\}} = \boldsymbol{\theta}^{\{g\}} - r \left[\sum_{i=1}^{N} \mathbf{s}(\mathbf{w}_i, \boldsymbol{\theta}^{\{g\}})\mathbf{s}(\mathbf{w}_i, \boldsymbol{\theta}^{\{g\}})'\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{s}(\mathbf{w}_i, \boldsymbol{\theta}^{\{g\}})\right].$$

7 Conditional Maximum Likelihood Estimation

7.1 The conditional maximum likelihood estimator

Conditional log likelihood for observation i:

$$\ell_i(\boldsymbol{\theta}) \equiv \ell(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}) = \log f(\mathbf{y}_i|\mathbf{x}_i; \boldsymbol{\theta}),$$

where $f(\mathbf{y}_i|\mathbf{x}_i; \boldsymbol{\theta}_o)$, $\boldsymbol{\theta}_o \in \boldsymbol{\Theta}$, is the conditional density for the random vector $\mathbf{y}_i$ given the random vector $\mathbf{x}_i$.

Sample log likelihood function:

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log f(\mathbf{y}_i|\mathbf{x}_i;\boldsymbol{\theta})$$

Conditional maximum likelihood estimator: the CMLE of $\boldsymbol{\theta}_o$ is the vector $\hat{\boldsymbol{\theta}}$ that solves

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} N^{-1}\mathcal{L}(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} N^{-1}\sum_{i=1}^{N} \ell_i(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} N^{-1}\sum_{i=1}^{N} \log f(\mathbf{y}_i|\mathbf{x}_i;\boldsymbol{\theta}).$$

Score: Suppose the conditional log likelihood function is once continuously differentiable with respect to $\boldsymbol{\theta}$. Then the score is

$$\mathbf{s}_i(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\ell_i(\boldsymbol{\theta})' = \left[\frac{\partial \ell_i}{\partial\theta_1}(\boldsymbol{\theta}),\ldots,\frac{\partial \ell_i}{\partial\theta_P}(\boldsymbol{\theta})\right]'.$$

First order condition:

$$\sum_{i=1}^{N}\mathbf{s}_i(\hat{\boldsymbol{\theta}}) = \mathbf{0}.$$

Hessian: Suppose the conditional log likelihood function is twice continuously differentiable with respect to $\boldsymbol{\theta}$. Then the Hessian is

$$\mathbf{H}_i(\boldsymbol{\theta}) = \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \nabla_{\boldsymbol{\theta}}\mathbf{s}_i(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_1} & \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_P} \\ \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_2\partial\theta_1} & \frac{\partial^2\ell_i(\boldsymbol{\theta}))}{\partial\theta_2\partial\theta_2} & \cdots & \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_2\partial\theta_P} \\ \vdots & \vdots & \ddots & \\ \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_P\partial\theta_1} & \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_P\partial\theta_2} & \cdots & \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_P\partial\theta_P} \end{pmatrix}$$

Expectation of the Hessian:

$$\mathbf{A}_o = -\operatorname{E}[\mathbf{H}_i(\boldsymbol{\theta}_o)]$$

Fisher information matrix:

$$\mathbf{B}_o = \operatorname{Var}[\mathbf{s}_i(\boldsymbol{\theta}_o)] = \operatorname{E}[\mathbf{s}_i(\boldsymbol{\theta}_o)\mathbf{s}_i(\boldsymbol{\theta}_o)'].$$

Conditional information matrix equality (CIME): Under fairly general conditions, in the maximum likelihood context the conditional information matrix equality holds

$$-\operatorname{E}[\mathbf{H}_i(\boldsymbol{\theta}_o)|\mathbf{x}_i] = \operatorname{E}[\mathbf{s}_i(\boldsymbol{\theta}_o)\mathbf{s}_i(\boldsymbol{\theta}_o)'|\mathbf{x}_i]$$

Unconditional information matrix equality (UIME): By the law of iterated expectations,

$$\mathbf{A}_o = -\,\mathrm{E}[\mathbf{H}_i(\boldsymbol{\theta}_o)] = \mathrm{E}[\mathbf{s}_i(\boldsymbol{\theta}_o)\mathbf{s}_i(\boldsymbol{\theta}_o)'] = \mathbf{B}_o.$$

7.2 Asymptotic Properties

Consistency: Suppose conditions equivalent to those for the M-estimator are satisfied. Then the CMLE is consistent,

$$\hat{\boldsymbol{\theta}} \xrightarrow{\mathrm{p}} \boldsymbol{\theta}_o.$$

Asymptotic normality: Suppose conditions equivalent to those for the M-estimator are satisfied. Then the CMLE is asymptotically normal,

$$N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{V}_o),$$

where $\mathbf{V}_o = \mathbf{A}_o^{-1}$

7.3 Estimators of the Asymptotic Variance

(a) Direct estimate:

$$\hat{\mathbf{V}} = \hat{\mathbf{A}}^{-1} = \left[N^{-1}\sum_{i=1}^{N} -\mathbf{H}_i(\hat{\boldsymbol{\theta}})\right]^{-1}.$$

(b) Using the conditional expectation:

$$\hat{\mathbf{V}} = \hat{\mathbf{A}}^{-1} = \left[N^{-1}\sum_{i=1}^{N} \mathbf{A}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\right]^{-1},$$

where $\mathbf{A}(\mathbf{x}_i, \boldsymbol{\theta}_o) \equiv -\,\mathrm{E}[\mathbf{H}(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}_o)|\mathbf{x}_i)]$.

(c) Outer product of the score:

$$\hat{\mathbf{V}} = \hat{\mathbf{A}}^{-1} = \left[N^{-1}\sum_{i=1}^{N} \mathbf{s}_i(\hat{\boldsymbol{\theta}})\mathbf{s}_i(\hat{\boldsymbol{\theta}})'\right]^{-1}.$$

Asymptotic standard errors for $\hat{\boldsymbol{\theta}}$: Take the square roots of the elements on the main diagonal of

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} = \widehat{\mathrm{Avar}}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{V}}/N = \hat{\mathbf{A}}^{-1}/N.$$

7.4 Inference

Wald test of linear hypotheses: To test the linear hypotheses Q $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ against $H_1 : \mathbf{R}\theta \neq \mathbf{r}$, the Wald statistic is (distribution under H_0)

$$W_N \equiv \left[\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right]' \left[\mathbf{R}(\hat{\mathbf{V}}/N)\mathbf{R}'\right]^{-1} \left[\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right] \overset{a}{\sim} \chi^2_Q.$$

Wald test of nonlinear hypotheses: To test the Q nonlinear hypotheses $H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ against $H_1 : \mathbf{c}(\boldsymbol{\theta}) \neq \mathbf{0}$, the Wald statistic is (distribution under H_0)

$$W_N \equiv \mathbf{c}(\hat{\boldsymbol{\theta}})' \left[\mathbf{C}(\hat{\boldsymbol{\theta}})(\hat{\mathbf{V}}/N)\mathbf{C}(\hat{\boldsymbol{\theta}})'\right]^{-1} \mathbf{c}(\hat{\boldsymbol{\theta}}) \overset{a}{\sim} \chi^2_Q.$$

Likelihood ratio (LR) test: To test the Q nonlinear hypotheses $H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ against $H_1 : \mathbf{c}(\boldsymbol{\theta}) \neq \mathbf{0}$, the LR statistic is (distribution under H_0)

$$LR \equiv 2[\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\tilde{\boldsymbol{\theta}})] \overset{a}{\sim} \chi^2_Q,$$

where $\tilde{\boldsymbol{\theta}}$ is the restricted estimator (estimated under H_0) and $\hat{\boldsymbol{\theta}}$ is the unrestricted estimator (estimated under H_1).

Lagrange multiplier (LM) or score test: To test the Q nonlinear hypotheses $H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ against $H_1 : \mathbf{c}(\boldsymbol{\theta}) \neq \mathbf{0}$, the LM statistic is (distribution under H_0)

$$LM \equiv \left(N^{-1/2}\sum_{i=1}^{N} \tilde{\mathbf{s}}_i\right)' \tilde{\mathbf{A}}^{-1} \left(N^{-1/2}\sum_{i=1}^{N} \tilde{\mathbf{s}}_i\right) \overset{a}{\sim} \chi^2_Q,$$

where $\tilde{\mathbf{s}}_i = \mathbf{s}_i(\tilde{\boldsymbol{\theta}})$ is the $P \times 1$ score evaluated at the restricted estimate $\tilde{\boldsymbol{\theta}}$ and $\tilde{\mathbf{A}}$ is an estimator of $\mathbf{A}_o$. One of the following estimators can be used:

$$\tilde{\mathbf{A}} = N^{-1}\sum_{i=1}^{N} -\mathbf{H}_i(\tilde{\boldsymbol{\theta}}) \quad \text{or} \quad \tilde{\mathbf{A}} = N^{-1}\sum_{i=1}^{N} \mathbf{A}(\mathbf{x}_i, \tilde{\boldsymbol{\theta}}) \quad \text{or} \quad \tilde{\mathbf{A}} = N^{-1}\sum_{i=1}^{N} \tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i'.$$

8 Generalized Method of Moments Estimation

8.1 The Generalized Method of Moments Estimator

Moment restrictions: Let $\{\mathbf{w} \in \mathbb{R}^M \colon i = 1, 2, \ldots\}$ denote a set of independent, identically distributed random vectors, where some feature of the distribution of $\mathbf{w}_i$ is indexed by the $P \times 1$ parameter vector $\boldsymbol{\theta}$. It is assumed that for some function $\mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) \in \mathbb{R}^L$, the parameter $\boldsymbol{\theta}_o \in \boldsymbol{\Theta} \subset \mathbb{R}^P$ satisfies

$$\mathrm{E}\left[\mathbf{g}\left(\mathbf{w}_i, \boldsymbol{\theta}_o\right)\right] = \mathbf{0}.$$

Generalized method of moments (GMM) estimator: The GMM estimator $\hat{\boldsymbol{\theta}}$ minimizes

$$\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q_N(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left[N^{-1} \sum_{i=1}^{N} \mathbf{g}\left(\mathbf{w}_i, \boldsymbol{\theta}\right) \right]' \hat{\boldsymbol{\Xi}} \left[N^{-1} \sum_{i=1}^{N} \mathbf{g}\left(\mathbf{w}_i, \boldsymbol{\theta}\right) \right],$$

where $\hat{\boldsymbol{\Xi}}$ is an $L \times L$ symmetric, positive semidefinite weighting matrix.

First order condition:

$$\left[\sum_{i=1}^{N} \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) \right]' \hat{\boldsymbol{\Xi}} \left[\sum_{i=1}^{N} \mathbf{g}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) \right] \equiv \mathbf{0}.$$

Expected gradient of the moment condition:

$$\mathbf{G}_o = \mathrm{E}\left[\nabla_{\theta} \mathbf{g}\left(\mathbf{w}_i, \boldsymbol{\theta}_o\right)\right].$$

Variance of the moment condition:

$$\boldsymbol{\Lambda}_o = \mathrm{E}\left[\mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}_o)\mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}_o)'\right] = \mathrm{Var}\left[\mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}_o)\right].$$

8.2 Asymptotic Properties

Consistency: Suppose conditions similar to those for the M-estimator are satisfied and $\hat{\boldsymbol{\Xi}} \xrightarrow{\mathrm{p}} \boldsymbol{\Xi}_o$, where $\boldsymbol{\Xi}_o$ is an $L \times L$ positive definite matrix. Then the GMM estimator is consistent,

$$\hat{\boldsymbol{\theta}} \xrightarrow{\mathrm{p}} \boldsymbol{\theta}_o.$$

Asymptotic normality: Suppose conditions equivalent to those for the M-estimator are satisfied, $\hat{\boldsymbol{\Xi}} \xrightarrow{\mathrm{p}} \boldsymbol{\Xi}_o$, where $\boldsymbol{\Xi}_o$ is an $L \times L$ positive definite matrix, and $\mathbf{G}_o$ has rank P. Then the GMM estimator is asymptotically normal,

$$N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{V}_o),$$

where

$$\mathbf{V}_o = \mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}$$

with

$$\mathbf{A}_o \equiv \mathbf{G}_o'\boldsymbol{\Xi}_o\mathbf{G}_o$$

and

$$\mathbf{B}_o \equiv \mathbf{G}_o'\boldsymbol{\Xi}_o\boldsymbol{\Lambda}_o\boldsymbol{\Xi}_o\mathbf{G}_o.$$

8.3 Estimation of the Variance

Estimator of $\boldsymbol{\Lambda}_o$:

$$\hat{\boldsymbol{\Lambda}} \equiv N^{-1} \sum_{i=1}^{N} \mathbf{g}_i(\hat{\boldsymbol{\theta}})\, \mathbf{g}_i(\hat{\boldsymbol{\theta}})'$$

Estimator of $\mathbf{G}_o$:

$$\hat{\mathbf{G}} \equiv N^{-1} \sum_{i=1}^{N} \nabla_{\boldsymbol{\theta}} \mathbf{g}_i(\hat{\boldsymbol{\theta}}).$$

Estimator of $\mathbf{A}_o$:

$$\hat{\mathbf{A}} = \hat{\mathbf{G}}' \hat{\boldsymbol{\Xi}} \hat{\mathbf{G}}$$

Estimator of $\mathbf{B}_o$:

$$\hat{\mathbf{B}} = \hat{\mathbf{G}}' \hat{\boldsymbol{\Xi}} \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\Xi}} \hat{\mathbf{G}}$$

Estimator of $\mathbf{V}_o$:

$$\hat{\mathbf{V}} = \hat{\mathbf{A}}^{-1} \hat{\mathbf{B}} \hat{\mathbf{A}}^{-1}$$

Asymptotic standard errors for $\hat{\boldsymbol{\theta}}$: Take the square roots of the elements on the main diagonal of

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} = \widehat{\mathrm{Avar}}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{V}}/N = \hat{\mathbf{A}}^{-1} \hat{\mathbf{B}} \hat{\mathbf{A}}^{-1}/N.$$

8.4 Efficient GMM estimation

Optimal weighting matrix: The optimal weighting matrix is chosen such that $\hat{\boldsymbol{\Xi}}_{\mathrm{opt}} \xrightarrow{\mathrm{p}} \boldsymbol{\Lambda}_o^{-1}$, e.g.,

$$\hat{\boldsymbol{\Xi}}_{\mathrm{opt}} = \hat{\boldsymbol{\Lambda}}^{-1}.$$

Efficient GMM estimator: The asymptotically efficient GMM estimator solves

$$\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left[N^{-1} \sum_{i=1}^{N} \mathbf{g}_i(\boldsymbol{\theta}) \right]' \hat{\boldsymbol{\Lambda}}^{-1} \left[N^{-1} \sum_{i=1}^{N} \mathbf{g}_i(\boldsymbol{\theta}) \right].$$

Asymptotic distribution of the efficient GMM estimator:

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o\right) \xrightarrow{\mathrm{d}} \mathrm{Normal}\left(\mathbf{0}, \mathbf{V}_o\right),$$

where

$$\mathbf{V}_o = \left[\mathbf{G}_o'\boldsymbol{\Lambda}_o^{-1}\mathbf{G}_o\right]^{-1}.$$

Asymptotic standard errors for the efficient GMM estimator: Take the square roots of the elements on the main diagonal of

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} = \widehat{\mathrm{Avar}}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{V}}/N = \left[\hat{\mathbf{G}}'\hat{\boldsymbol{\Lambda}}^{-1}\hat{\mathbf{G}}\right]^{-1}/N.$$

8.5 Inference

Test of the validity of the moment conditions: Hansen's J statistic is (distribution under H_0)

$$J = N\,Q_N(\hat{\boldsymbol{\theta}}) = N\left[N^{-1}\sum_{i=1}^{N}\mathbf{g}_i(\hat{\boldsymbol{\theta}})\right]'\hat{\boldsymbol{\Lambda}}^{-1}\left[\sum_{i=1}^{N}N^{-1}\mathbf{g}_i(\hat{\boldsymbol{\theta}})\right] \overset{a}{\sim} \chi^2_{L-P},$$

where L is the number of moment conditions and P is the number of parameters.

GMM distance statistic: To test the Q nonlinear hypotheses $H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ against $H_1 : \mathbf{c}(\boldsymbol{\theta}) \neq \mathbf{0}$, the GMM distance statistic is (distribution under H_0)

$$\left\{\left[\sum_{i=1}^{N}\mathbf{g}_i(\tilde{\boldsymbol{\theta}})\right]'\hat{\Lambda}^{-1}\left[\sum_{i=1}^{N}\mathbf{g}_i(\tilde{\boldsymbol{\theta}})\right] - \left[\sum_{i=1}^{N}\mathbf{g}_i(\hat{\boldsymbol{\theta}})\right]'\hat{\boldsymbol{\Lambda}}^{-1}\left[\sum_{i=1}^{N}\mathbf{g}_i(\hat{\boldsymbol{\theta}})\right]\right\}/N \xrightarrow{\mathrm{d}} \chi^2_Q,$$

where $\tilde{\boldsymbol{\theta}}$ is the restricted estimator (estimated under H_0), $\hat{\boldsymbol{\theta}}$ is the unrestricted estimator (estimated under H_1), and $\hat{\boldsymbol{\Lambda}}$ is obtained from an initial unrestricted estimator.

9 Binomial Choice Models

9.1 Model setup

Latent variable representation: The observable variable y_i takes the values 0 and 1 according to

$$y_i = \begin{cases} 1 & \text{if } \; {y_i}^* > 0 \\ 0 & \text{if } \; {y_i}^* \leq 0, \end{cases}$$

where y_i^* is a continuous latent variable that is determined by

$$y_i^* = \mathbf{x}_i\boldsymbol{\theta} + e_i.$$

Distribution of e_i: The error e_i is assumed to be distributed according to the twice continuously differentiable distribution function (cdf) $G(\cdot)$ that has symmetric first derivative (pdf) $g(\cdot)$. Moreover, $\mathrm{E}(e_i) = 0$ (inclusion of an intercept in the latent model).

Conditional probability that $y_i = 1$:

$$p(\mathbf{x}_i) = \Pr(y_i = 1|\mathbf{x}_i) = G(\mathbf{x}_i\boldsymbol{\theta}).$$

Conditional expectation:

$$\mathrm{E}(y|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\theta}).$$

9.2 Conditional Maximum Likelihood Estimation

Log likelihood function: for observation i

$$\ell_i(\boldsymbol{\theta}) = \log f(y_i|\mathbf{x}_i;\boldsymbol{\theta}) = y_i \log[G(\mathbf{x}_i\boldsymbol{\theta})] + (1-y_i)\log[1-G(\mathbf{x}_i\boldsymbol{\theta})].$$

and for the full sample of size N

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\theta}) = \sum_{i=1}^{N} \left\{ y_i \log[G(\mathbf{x}_i\boldsymbol{\theta})] + (1-y_i)\log[1-G(\mathbf{x}_i\boldsymbol{\theta})] \right\}.$$

Score:

$$\mathbf{s}_i(\boldsymbol{\theta}) = \left[\frac{y_i}{G(\mathbf{x}_i\boldsymbol{\theta})} - \frac{1-y_i}{1-G(\mathbf{x}_i\boldsymbol{\theta})}\right] g(\mathbf{x}_i\boldsymbol{\theta})\mathbf{x}_i'$$

or, defining $u_i \equiv y_i - \mathrm{E}(y_i|\mathbf{x}_i) = y_i - G(\mathbf{x}_i\boldsymbol{\theta}_o)$,

$$\mathbf{s}_i(\boldsymbol{\theta}) = \frac{g(\mathbf{x}_i\boldsymbol{\theta})}{G(\mathbf{x}_i\boldsymbol{\theta})[1-G(\mathbf{x}_i\boldsymbol{\theta})]}\,\mathbf{x}_i' u_i.$$

Hessian:

$$\mathbf{H}_i(\boldsymbol{\theta}) = -\left[\frac{y_i g_i}{G_i^2} + \frac{(1-y_i)g_i}{[1-G_i]^2}\right] g_i \mathbf{x}_i'\mathbf{x}_i + \left[\frac{y_i}{G_i} - \frac{1-y_i}{1-G_i}\right] g'(\mathbf{x}_i\boldsymbol{\theta})\mathbf{x}_i'\mathbf{x}_i,$$

where $G_i \equiv G(\mathbf{x}_i\boldsymbol{\theta})$ and $g_i \equiv g(\mathbf{x}_i\boldsymbol{\theta})$.

Conditional expectation of the Hessian:

$$\mathbf{A}(\mathbf{x}_i,\boldsymbol{\theta}_o) = -\,\mathrm{E}[\mathbf{H}_i(\boldsymbol{\theta}_o)|\mathbf{x}] = \frac{g(\mathbf{x}_i\boldsymbol{\theta}_o)^2}{G(\mathbf{x}_i\boldsymbol{\theta}_o)[1-G(\mathbf{x}_i\boldsymbol{\theta}_o)]}\mathbf{x}_i'\mathbf{x}_i.$$

Estimator of the asymptotic variance:

$$\hat{\mathbf{V}} = \left[N^{-1} \sum_{i=1}^{N} \frac{g(\mathbf{x}_i \hat{\boldsymbol{\theta}})^2}{G(\mathbf{x}_i \hat{\boldsymbol{\theta}})[1 - G(\mathbf{x}_i \hat{\boldsymbol{\theta}})]} \mathbf{x}_i' \mathbf{x}_i \right]^{-1}.$$

Asymptotic standard errors: Take the square roots of the elements on the main diagonal of

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} = \widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{V}}/N.$$

9.3 Probit

Probit model: standard normal distribution for e_i,

$$\Pr(y = 1|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\theta}) = \int_{-\infty}^{\mathbf{x}\boldsymbol{\theta}} \phi(t) dt$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf, respectively, of the standard normal distribution.

FOC:

$$\sum_{i=1}^{N} \mathbf{s}_i(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{N} \frac{\phi(\mathbf{x}_i \hat{\boldsymbol{\theta}})}{\Phi(\mathbf{x}_i \hat{\boldsymbol{\theta}})[1 - \Phi(\mathbf{x}_i \hat{\boldsymbol{\theta}})]} \mathbf{x}_i'[y_i - \Phi(\mathbf{x}_i \hat{\boldsymbol{\theta}})] = \mathbf{0}$$

Asymptotic standard errors: Take the square roots of the elements on the main diagonal of

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}) = \left[\sum_{i=1}^{N} \frac{\phi(\mathbf{x}_i \hat{\boldsymbol{\theta}})^2}{\Phi(\mathbf{x}_i \hat{\boldsymbol{\theta}})[1 - \Phi(\mathbf{x}_i \hat{\boldsymbol{\theta}})]} \mathbf{x}_i' \mathbf{x}_i \right]^{-1}.$$

9.4 Logit

Logit model: logistic distribution for e_i,

$$\Pr(y = 1|\mathbf{x}) = \Lambda(\mathbf{x}\boldsymbol{\theta}) = \frac{\exp(\mathbf{x}\boldsymbol{\theta})}{1 + \exp(\mathbf{x}\boldsymbol{\theta})},$$

where $\Lambda(\cdot)$ is the cdf of a standard logistic distribution with pdf

$$\lambda(z) = \frac{\exp(z)}{[1 + \exp(z)]^2} = \Lambda(z)[1 - \Lambda(z)].$$

FOC:

$$\sum_{i=1}^{N} \mathbf{s}_i(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{N} \mathbf{x}_i'[y_i - \Lambda(\mathbf{x}_i\hat{\boldsymbol{\theta}})] = \mathbf{0}.$$

Asymptotic standard errors: Take the square roots of the elements on the main diagonal of

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}) = \left[\sum_{i=1}^{N} \lambda(\mathbf{x}_i\hat{\boldsymbol{\theta}})\mathbf{x}_i'\mathbf{x}_i\right]^{-1}.$$

9.5 Partial Effects

Partial effect of a continuous variable:

$$\frac{\partial\,\mathrm{E}(y_i|\mathbf{x}_i)}{\partial x_{i,k}} = g(\mathbf{x}_i\boldsymbol{\theta})\theta_k.$$

Partial effect of a discrete variable: The partial effect of a dummy variable, $x_{i,P}$, i.e., the effect a change in $x_{i,P}$ from 0 to 1 has on $\Pr(y_i = 1|\mathbf{x}_i)$, is

$$\Delta P_i = \Pr(y_i = 1|x_{i,P} = 1) - \Pr(y_i = 1|x_{i,P} = 0).$$

where the $x_{i,1}, \ldots, x_{i,P-1}$ are as observed. The probabilities are computed as follows:

$$\Delta P_i = G([x_{i,1}, \ldots, x_{i,P-1}, 1]\boldsymbol{\theta}) - G([x_{i,1}, \ldots, x_{i,P-1}, 0]\boldsymbol{\theta}).$$

Partial effect of the average (PEA): For a continuous explanatory variable x_k, this is in population

$$PEA = g\left(\mathrm{E}[\mathbf{x}_i]\boldsymbol{\theta}\right)\theta_k$$

which is estimated as

$$\widehat{PEA} = g\left(\bar{\mathbf{x}}\hat{\boldsymbol{\theta}}\right)\hat{\theta}_k.$$

Average partial effect (APE): For a continuous explanatory variable x_k, this is in population

$$APE = \mathrm{E}\left[g(\mathbf{x}_i\boldsymbol{\theta})\right]\theta_k$$

which is estimated as

$$\widehat{APE} = N^{-1}\sum_{i=1}^{N} g(\mathbf{x}_i\boldsymbol{\theta})\hat{\theta}_k.$$

PEA and APE for discrete variables: For a discrete explanatory variable x_P, one computes

$$\widehat{PEA} = G([\bar{x}_1, \ldots, \bar{x}_{P-1}, 1]\hat{\boldsymbol{\theta}}) - G([\bar{x}_1, \ldots, \bar{x}_{P-1}, 0]\hat{\boldsymbol{\theta}})$$

$$\widehat{APE} = N^{-1} \sum_{i=1}^{N} \Big[G([x_{i,1}, \ldots, x_{i,P-1}, 1]\hat{\boldsymbol{\theta}}) - G([x_{i,1}, \ldots, x_{i,P-1}, 0]\hat{\boldsymbol{\theta}}) \Big] .$$

Appendix: Tables

CDF of the standard normal distribution

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Percentiles of the χ^2–distribution

F_{χ^2}	0.0100	0.0250	0.0500	0.1000	0.9000	0.9500	0.9750	0.9900
$r = 1$	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349
2	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2103
3	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449
4	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767
5	0.5543	0.8312	1.1455	1.6103	9.2364	11.0705	12.8325	15.0863
6	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916	14.4494	16.8119
7	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671	16.0128	18.4753
8	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073	17.5345	20.0902
9	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190	19.0228	21.6660
10	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070	20.4832	23.2093
11	3.0535	3.8157	4.5748	5.5778	17.2750	19.6751	21.9200	24.7250
12	3.5706	4.4038	5.2260	6.3038	18.5493	21.0261	23.3367	26.2170
13	4.1069	5.0088	5.8919	7.0415	19.8119	22.3620	24.7356	27.6882
14	4.6604	5.6287	6.5706	7.7895	21.0641	23.6848	26.1189	29.1412
15	5.2293	6.2621	7.2609	8.5468	22.3071	24.9958	27.4884	30.5779
16	5.8122	6.9077	7.9616	9.3122	23.5418	26.2962	28.8454	31.9999
17	6.4078	7.5642	8.6718	10.0852	24.7690	27.5871	30.1910	33.4087
18	7.0149	8.2307	9.3905	10.8649	25.9894	28.8693	31.5264	34.8053
19	7.6327	8.9065	10.1170	11.6509	27.2036	30.1435	32.8523	36.1909
20	8.2604	9.5908	10.8508	12.4426	28.4120	31.4104	34.1696	37.5662
21	8.8972	10.2829	11.5913	13.2396	29.6151	32.6706	35.4789	38.9322
22	9.5425	10.9823	12.3380	14.0415	30.8133	33.9244	36.7807	40.2894
23	10.1957	11.6886	13.0905	14.8480	32.0069	35.1725	38.0756	41.6384
24	10.8564	12.4012	13.8484	15.6587	33.1962	36.4150	39.3641	42.9798
25	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3141
26	12.1981	13.8439	15.3792	17.2919	35.5632	38.8851	41.9232	45.6417
27	12.8785	14.5734	16.1514	18.1139	36.7412	40.1133	43.1945	46.9629
28	13.5647	15.3079	16.9279	18.9392	37.9159	41.3371	44.4608	48.2782
29	14.2565	16.0471	17.7084	19.7677	39.0875	42.5570	45.7223	49.5879
30	14.9535	16.7908	18.4927	20.5992	40.2560	43.7730	46.9792	50.8922
40	22.1643	24.4330	26.5093	29.0505	51.8051	55.7585	59.3417	63.6907
50	29.7067	32.3574	34.7643	37.6886	63.1671	67.5048	71.4202	76.1539
60	37.4849	40.4817	43.1880	46.4589	74.3970	79.0819	83.2977	88.3794
70	45.4417	48.7576	51.7393	55.3289	85.5270	90.5312	95.0232	100.4252
80	53.5401	57.1532	60.3915	64.2778	96.5782	101.8795	106.6286	112.3288
90	61.7541	65.6466	69.1260	73.2911	107.5650	113.1453	118.1359	124.1163
100	70.0649	74.2219	77.9295	82.3581	118.4980	124.3421	129.5612	135.8067

Percentiles of the t–distribution

F_t	0.9000	0.9500	0.9750	0.9900	0.9950
$r = 1$	3.0777	6.3138	12.7062	31.8205	63.6567
2	1.8856	2.9200	4.3027	6.9646	9.9248
3	1.6377	2.3534	3.1824	4.5407	5.8409
4	1.5332	2.1318	2.7764	3.7469	4.6041
5	1.4759	2.0150	2.5706	3.3649	4.0321
6	1.4398	1.9432	2.4469	3.1427	3.7074
7	1.4149	1.8946	2.3646	2.9980	3.4995
8	1.3968	1.8595	2.3060	2.8965	3.3554
9	1.3830	1.8331	2.2622	2.8214	3.2498
10	1.3722	1.8125	2.2281	2.7638	3.1693
11	1.3634	1.7959	2.2010	2.7181	3.1058
12	1.3562	1.7823	2.1788	2.6810	3.0545
13	1.3502	1.7709	2.1604	2.6503	3.0123
14	1.3450	1.7613	2.1448	2.6245	2.9768
15	1.3406	1.7531	2.1314	2.6025	2.9467
16	1.3368	1.7459	2.1199	2.5835	2.9208
17	1.3334	1.7396	2.1098	2.5669	2.8982
18	1.3304	1.7341	2.1009	2.5524	2.8784
19	1.3277	1.7291	2.0930	2.5395	2.8609
20	1.3253	1.7247	2.0860	2.5280	2.8453
21	1.3232	1.7207	2.0796	2.5176	2.8314
22	1.3212	1.7171	2.0739	2.5083	2.8188
23	1.3195	1.7139	2.0687	2.4999	2.8073
24	1.3178	1.7109	2.0639	2.4922	2.7969
25	1.3163	1.7081	2.0595	2.4851	2.7874
26	1.3150	1.7056	2.0555	2.4786	2.7787
27	1.3137	1.7033	2.0518	2.4727	2.7707
28	1.3125	1.7011	2.0484	2.4671	2.7633
29	1.3114	1.6991	2.0452	2.4620	2.7564
30	1.3104	1.6973	2.0423	2.4573	2.7500
∞	1.2816	1.6449	1.9600	2.3263	2.5758

Percentiles of the F–distribution

The Table shows the values k, for which $P(v \leq k) = F(k) = 0.95$ holds. (r_1 = degrees of freedom of the nominator,r_2 = degrees of freedom of the denominator)

$F(k) = 0.95$	$\mathbf{r_1} = 1$	2	3	4	5	6	7	8	9	10	20	120
$\mathbf{r_2} = 1$	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	248.0131	253.2529
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4458	19.4874
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.6602	8.5494
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.8025	5.6581
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.5581	4.3985
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.8742	3.7047
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.4445	3.2674
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.1503	2.9669
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	2.9365	2.7475
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.7740	2.5801
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.6464	2.4480
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.5436	2.3410
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.4589	2.2524
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.3879	2.1778
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.3275	2.1141
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.2756	2.0589
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.2304	2.0107
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.1906	1.9681
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.1555	1.9302
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.1242	1.8963
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.0960	1.8657
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.0707	1.8380
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.0476	1.8128
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.0267	1.7896
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.0075	1.7684
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	1.9898	1.7488
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	1.9736	1.7306
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	1.9586	1.7138
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	1.9446	1.6981
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	1.9317	1.6835
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	1.8389	1.5766
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.7480	1.4673
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.6587	1.3519
∞	3.8416	2.9958	2.6050	2.3720	2.2142	2.0987	2.0097	1.9385	1.8800	1.8308	1.5706	1.2216