Formulae and Tables

for

Econometrics II

April 02, 2017

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

Part A. Cross Section Econometrics

1 Conditional Maximum Likelihood Estimation

1.1 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,

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

1.2 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,

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

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

1.3 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

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

1.4 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:

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

2 Generalized Method of Moments Estimation

2.1 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:

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

2.2 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

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

2.3 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

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

2.4 Efficient GMM estimation

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

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

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:

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

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

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

2.5 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.

3 Binomial Choice Models

3.1 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:

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

3.2 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

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

3.3 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

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

3.4 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

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

3.5 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

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

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

Part B. Econometrics for Stationary Time Series Processes

4 Stationary Time Series Regression

4.1 Properties of Stochastic Time-Series Processes

Autocovariance of order k:

$$\mathrm{Cov}(y_t, y_{t-k}) = \mathrm{E}\{[y_t - \mathrm{E}(y_t)][y_{t-k} - \mathrm{E}(y_{t-k})]\}$$

Autocorrelation of order k:

$$\mathrm{Corr}(y_t, y_{t-k}) = \frac{\mathrm{E}\{[y_t - \mathrm{E}(y_t)][y_{t-k} - \mathrm{E}(y_{t-k})]\}}{\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(y_{t-k})}}.$$

Cross covariance of order k:

$$\mathrm{Cov}(x_t, y_{t-k}) = \mathrm{E}\{[x_t - \mathrm{E}(x_t)][y_{t-k} - \mathrm{E}(y_{t-k})]\}.$$

Cross correlation of order k:

$$\mathrm{Corr}(x_t, y_{t-k}) = \frac{\mathrm{E}\{[x_t - \mathrm{E}(x_t)][y_{t-k} - \mathrm{E}(y_{t-k})]\}}{\sqrt{\mathrm{Var}(x_t)}\sqrt{\mathrm{Var}(y_{t-k})}}.$$

Weak stationarity: A time series process $\{y_t\}$ is called weakly stationary if the first and second moments are time-invariant and finite, i.e.,

$$\begin{aligned} \mathrm{E}(y_t) &= \mu < \infty \quad \forall t \\ \mathrm{Var}(y_t) &= \sigma^2 < \infty \quad \forall t \\ \mathrm{Cov}(y_t, y_{t-k}) &= \gamma_k < \infty \quad \forall t \end{aligned}$$

Strong stationarity: A time series process $\{y_t\}$ is called strongly stationary if the joint probability distribution of any set of k observations in the sequence $[y_t, y_{t+1}, \ldots, y_{t+k-1}]$ is the same regardless of the origin, t, in the time scale.

Ergodicity: A strongly stationary time-series process, $\{y_t\}$, is ergodic if for any two bounded functions that map vectors in the a and b dimensional real vector spaces to real scalars, $f : \mathbb{R}^a \to \mathbb{R}^1$ and $g : \mathbb{R}^b \to \mathbb{R}^1$,

$$\begin{aligned} &\lim_{k\to\infty} |\mathrm{E}[f(y_t, \ldots, y_{t+a-1})g(y_{t+k}, \ldots, y_{t+k+b-1})]| \\ &\qquad = |\mathrm{E}[f(y_t, \ldots, y_{t+a-1})]|\,|\mathrm{E}[g(y_{t+k}, \ldots, y_{t+k+b-1})]| \end{aligned}$$

Ergodic Theorems for scalar processes: If $\{y_t\}$ is a time-series process that is strongly stationary and ergodic and $E[y_t] = \mu$ exists and is a finite constant, then

$$\bar{y}_T = T^{-1} \sum_{t=1}^{T} y_t \xrightarrow{\text{a.s.}} \mu.$$

If in addition $E[y_t^2] = m$ exists and is finite then

$$T^{-1} \sum_{t=1}^{T} y_t^2 \xrightarrow{\text{a.s.}} m.$$

Ergodic Theorems for vector processes: If $\{\mathbf{y}_t\}$ is a $K \times 1$ vector of strongly stationary and ergodic processes with existing and finite moments $E[\mathbf{y}_t] = \boldsymbol{\mu}$ and $E[\mathbf{y}_t \mathbf{y}_t'] = \mathbf{M}$, then

$$T^{-1} \sum_{t=1}^{T} \mathbf{y}_t \xrightarrow{\text{a.s.}} \boldsymbol{\mu}$$

and

$$T^{-1} \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' \xrightarrow{\text{a.s.}} \mathbf{M}.$$

Martingale sequence: A vector sequence $\mathbf{z}_t$ is a martingale sequence if $\mathrm{E}[\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \ldots] = \mathbf{z}_{t-1}$.

Martingale difference sequence: A vector sequence $\mathbf{z}_t$ is a martingale difference sequence if $\mathrm{E}[\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \ldots] = 0$.

Stationary linear process: A stationary linear process with mean zero is defined as

$$u_t = \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \psi(L) \varepsilon_t,$$

where ε_t is iid white noise with mean zero and variance $\sigma^2 > 0$, if

(C1) $\quad \sum_{j=0}^{\infty} j |\psi_j| < \infty$

(C2) $\quad \psi(1) = \psi_0 + \psi_1 + \psi_2 + \cdots \neq 0$

Then u_t has mean $\mathrm{E}(u_t) = 0$, variance $\mathrm{Var}(u_t) = \sigma^2(\psi_0^2 + \psi_1^2 + \psi_2^2 + \cdots)$ and long-run variance

$$\lambda^2 \equiv \lim_{T \to \infty} \mathrm{Var}(\sqrt{T} \bar{u}_T) = \sigma^2 [\psi(1)]^2 > 0.$$

4.2 Asymptotic Properties of Time Series Regressions

Population model and OLS assuptions: The population model with K regressors is

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + u_t, \qquad t = 1, \ldots, T,$$

where $\{[\mathbf{x}, u]\}$ is a jointly stationary and ergodic process with finite first and second moments and the usual OLS assumptions $\mathrm{E}(\mathbf{x}_t' u_t) = 0$ and $\mathrm{rank}[\mathrm{E}(\mathbf{x}_t'\mathbf{x}_t)] = K$ hold.

Consistency: The OLS estimator of the population model is consistent,

$$\hat{\boldsymbol{\beta}} \xrightarrow{\mathrm{p}} \boldsymbol{\beta}.$$

Martingale Difference Central Limit Theorem: If $\mathbf{z}_t$ is a vector valued stationary and ergodic martingale difference sequence, with $\mathrm{E}[\mathbf{z}_t\mathbf{z}_t'] = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is a finite positive definite matrix, then

$$\sqrt{T}\bar{\mathbf{z}}_T = T^{-1/2}\sum_{t=1}^{T}\mathbf{z}_t \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Asymptotic distribution of the OLS estimator when u_t is white noise: By the martingale difference CLT,

$$T^{-1/2}\sum_{t=1}^{T}\mathbf{x}_t' u_t \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{B}),$$

where $\mathbf{B} \equiv \mathrm{E}(u_t^2\mathbf{x}_t'\mathbf{x}_t)$. From this follows

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{V}),$$

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

Estimation of the variance of the OLS estimator when u_t is white noise: Using the OLS residuals $\hat{u}_t$, consistent estimators are

$$\hat{\mathbf{A}} = T^{-1}\sum_{t=1}^{T}\mathbf{x}_t'\mathbf{x}_t$$

and

$$\hat{\mathbf{B}} = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2\mathbf{x}_t'\mathbf{x}_t$$

such that

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

Gordin's Central Limit Theorem: If $\{\mathbf{z}_t\}$ is a stationary and ergodic stochastic process of dimension $K \times 1$ that satisfies the following conditions:

(1) Asymptotic uncorrelatedness: $\mathrm{E}[z_t | z_{t-k}, z_{t-k-1}, \ldots]$ converges in mean square to zero as $k \to \infty$.

(2) Summability of autocovariances: the asymptotic variance $\mathbf{\Gamma}^*$ is finite, where

$$\lim_{T\to\infty} \mathrm{Var}(\sqrt{T}\bar{\mathbf{z}}_T) = \sum_{t=0}^{\infty}\sum_{s=0}^{\infty} \mathrm{Cov}(\mathbf{z}_t, \mathbf{z}_s) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(\mathbf{z}_t, \mathbf{z}_{t-k}) = \mathbf{\Gamma}^*.$$

(3) Asymptotic negligibility of innovations: information eventually becomes negligible as it fades far back in time from the current observation.

Then

$$\sqrt{T}\bar{\mathbf{z}}_T = T^{-1/2}\sum_{t=1}^{T} \mathbf{z}_t \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{\Gamma}^*).$$

Asymptotic distribution of the OLS estimator when u_t is temporally dependent: If $\mathbf{x}_t' u_t$ satisfies the conditions of Gordin's CLT,

$$T^{-1/2}\sum_{t=1}^{T} \mathbf{x}_t' u_t \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{B}),$$

where $\mathbf{B} \equiv \sum_{k=-\infty}^{\infty} \mathrm{Cov}(\mathbf{x}_t' u_t, \mathbf{x}_{t-k}' u_{t-k})$. From this follows

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{V}),$$

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

Estimation of the variance of the OLS estimator when u_t is temporally dependent: A consistent estimate is

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

where

$$\hat{\mathbf{A}} = T^{-1}\sum_{t=1}^{T} \mathbf{x}_t'\mathbf{x}_t \xrightarrow{\mathrm{p}} \mathbf{A}.$$

and

$$\hat{\mathbf{B}} = \hat{\mathbf{\Gamma}}_0 + \sum_{k=1}^{q}(\hat{\mathbf{\Gamma}}_k + \hat{\mathbf{\Gamma}}_k'), \qquad \text{with} \qquad \hat{\mathbf{\Gamma}}_k = T^{-1}\sum_{t=k+1}^{T} \hat{u}_t\hat{u}_{t-k}\mathbf{x}_t'\mathbf{x}_{t-k}.$$

Newey-West estimator: A positive semidefinite estimator of $\mathbf{B}$ is

$$\hat{\mathbf{B}} = \hat{\mathbf{\Gamma}}_0 + \sum_{k=1}^{q} \left[1 - \frac{k}{q+1}\right] (\hat{\mathbf{\Gamma}}_k + \hat{\mathbf{\Gamma}}_k').$$

5 Autoregressive Models

5.1 Properties of the Autoregressive Model of Order 1

Autoregressive model of order 1 without constant:

$$u_t = \rho u_{t-1} + e_t, \qquad e_t \overset{\text{iid}}{\sim} (0, \sigma^2)$$

Moving average representation:

$$u_t = e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \cdots = \sum_{i=0}^{\infty} \rho^i e_{t-i}$$

Conditional moments:

$$\begin{aligned} \mathrm{E}(u_t | u_{t-1}) &= \rho u_{t-1} \\ \mathrm{Var}(u_t | u_{t-1}) &= \sigma^2 \end{aligned}$$

Unconditional moments: The mean is

$$\mathrm{E}(u_t) = 0$$

and, if $|\rho| < 1$, the second moments are

$$\begin{aligned} \mathrm{Var}(u_t) &= \sigma^2 \frac{1}{1-\rho^2} \\ \mathrm{Cov}(u_t, u_{t-k}) &= \sigma^2 \frac{\rho^k}{1-\rho^2} \\ \mathrm{Corr}(u_t, u_{t-k}) &= \rho^k. \end{aligned}$$

5.2 Estimation of the Autoregressive Model of Order 1

Consistency of OLS when the disturbance is white noise: The OLS estimator of the AR(1) model without constant,

$$\hat{\rho} = \left(\sum_{t=2}^{T} u_{t-1}^2\right)^{-1} \sum_{t=2}^{T} u_{t-1} u_t,$$

is consistent if e_t is white noise and thus $\mathrm{E}(u_{t-1} e_t) = 0$.

Asymptotic normality of OLS when the disturbance is white noise: By the martingale difference CLT,

$$\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{\mathrm{d}} \mathrm{Normal}(0, \mathbf{V}).$$

If the disturbance e_t is homoscedastic, the asymptotic covariance can be estimated as

$$\hat{\mathbf{V}} = \hat{\sigma}_e^2 \left(T^{-1} \sum_{t=2}^{T} y_{t-1}^2 \right)^{-1} \qquad \Rightarrow \qquad \widehat{\mathrm{Avar}}(\hat{\rho}) = \hat{\sigma}_e^2 \left(\sum_{t=2}^{T} y_{t-1}^2 \right)^{-1}.$$

Inconsistency of OLS when the disturbance is autocorrelated: If the disturbance e_t is autocorrelated, the OLS estimator of the AR(1) model is inconsistent.

Asymptotic bias of OLS when the disturbance follows an AR(1) process: If e_t follows itself an AR(1) process,

$$e_t = \phi e_{t-1} + \varepsilon_t,$$

where ε_t is iid white noise, the asymptotic bias of the OLS estimator of ρ is

$$\operatorname{plim} \hat{\rho} = \frac{\rho + \phi}{1 + \rho\phi}.$$

5.3 The Autoregressive Model of Order p

Lag operator: The lag operator is defined as $Ly_t = y_{t-1}$ and thus

$$L^p y_t = y_{t-p}, \qquad p \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

Lag polynomial: A polynomial in L is called a lag polynomial. The lag polynomial

$$a(L) \equiv a_0 + a_1 L + \ldots + a_p L^p$$

can be applied to a time series variable y_t to yield

$$a(L)y_t = a_0 y_t + a_1 y_{t-1} + \ldots + a_p y_{t-p}.$$

Rules for lag polynomials: Lag polynomials can be multiplied. For example, define $a(L) = a_0 + a_1 L$ and $b(L) = b_0 + b_1 L + b_2 L^2$, then

$$a(L)b(L) = a_0 b_0 + (a_0 b_1 + a_1 b_0)L + (a_0 b_2 + a_1 b_1)L^2 + a_1 b_2 L^3.$$

Lag polynomials can be evaluated at 1,

$$a(1) = a_0 + a_1 + \ldots + a_p = \sum_{i=0}^{p} a_i.$$

A lag polynomial applied to a time invariant quantity μ yields

$$a(L)\mu = a_0\mu + a_1 L\mu + \ldots + a_p L^p \mu = a_0\mu + a_1\mu + \ldots + a_p\mu = a(1)\mu.$$

The AR(p) model: The autoregressive model of order p is

$$y_t = \mu + a_1 y_{t-1} + \ldots + a_p y_{t-p} + e_t,$$

or

$$a(L) y_t = \mu + e_t,$$

where the disturbance process e_t is iid white noise with mean zero.

OLS Estimation of the AR(p) model: The OLS estimator

$$\begin{pmatrix} \hat{\mu} \\ \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{pmatrix} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}.$$

is based on the observation matrices

$$\mathbf{Y} = \begin{pmatrix} y_{p+1} \\ \vdots \\ y_T \end{pmatrix}, \qquad \mathbf{Z} = \begin{pmatrix} 1 & y_p & \cdots & y_1 \\ \vdots & \vdots & & \vdots \\ 1 & y_{T-1} & \cdots & y_{T-p} \end{pmatrix}.$$

If the disturbance is white noise, the OLS estimator is consistent and asymptotically normal.

6 Dynamic Regression

6.1 Autoregressive Distributed Lag Model

Autoregressive distributed lag (ADL) model:

$$y_t = \mu + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + b_0 x_t + b_1 x_{t-1} + \ldots + b_q x_{t-q} + \varepsilon_t$$

or, using lag polynomials,

$$a(L) y_t = \mu + b(L) x_t + \varepsilon_t.$$

OLS estimation of the ADL model: If $\mathrm{E}[\varepsilon_t | y_{t-1}, y_{t-2}, \ldots, x_t, x_{t-1}, x_{t-2}, \ldots] = 0$, the joint process $\{(y_t, x_t)\}$ is stationary and ergodic, and there is no perfect multicollinearity among the regressors $1, y_{t-1}, \ldots, y_{t-p}, x_t, \ldots, x_{t-q}$, then the OLS estimator of the ADL model is consistent and asymptotically normally distributed.

Interpretation of the ADL coefficients as partial effects:

$$\frac{\partial\,\mathrm{E}(y_t | 1, y_{t-1}, \ldots, y_{t-p}, x_t, \ldots, x_{t-q})}{\partial y_{t-i}} = a_i \qquad i \in [1, \ldots, p]$$

$$\frac{\partial\,\mathrm{E}(y_t | 1, y_{t-1}, \ldots, y_{t-p}, x_t, \ldots, x_{t-q})}{\partial x_{t-i}} = b_i \qquad i \in [0, \ldots, q]$$

Long-run effect of a permanent shift in x on y:

$$\frac{\partial \operatorname{E}(y|x)}{\partial x} = \frac{b(1)}{a(1)} = \frac{b_0 + b_1 + \ldots + b_q}{1 - a_1 - \ldots - a_p}$$

6.2 Error-Correction Model

Error-correction model (ECM): The ECM is a reparameterization of the ADL model:

$$\Delta y_t = \mu - \alpha(y_{t-1} - \beta x_{t-1}) + \bar{a}_1 \Delta y_{t-1} + \ldots + \bar{a}_{p-1} \Delta y_{t-p+1} + \bar{b}_0 \Delta x_t + \ldots + \bar{b}_{q-1} \Delta x_{t-q+1} + \varepsilon_t$$

where $\Delta = 1 - L$ and

$$\begin{aligned}
\alpha &= a(1) \\
\beta &= b(1)/a(1) \\
\bar{a}_i &= -\sum_{k=i+1}^{p} a_i, \quad \text{for } i = 1, \ldots, p-1 \\
\bar{b}_0 &= b_0 \\
\bar{b}_i &= -\sum_{k=i+1}^{q} b_i, \quad \text{for } i = 1, \ldots, q-1
\end{aligned}$$

Estimation of the ECM by OLS: Writing the ECM as

$$\Delta y_t = \mu + \underbrace{(-\alpha)}_{\gamma_1} y_{t-1} + \underbrace{\alpha\beta}_{\gamma_2} x_{t-1} + \text{ lagged differences } + \varepsilon_t,$$

it can be estimated by OLS. The OLS estimator is, under the conditions stated for the ADL model, consistent and asymptotically normal.

Estimation of long-run parameters of the ECM: Calculate

$$\hat{\beta} = -\frac{\hat{\gamma}_2}{\hat{\gamma}_1}.$$

The asymptotic standard errors of $\hat{\beta}$ can be obtained using the delta method.

7 Tests of Autocorrelation and Lag Order Selection

Durbin-Watson test: Test of the null hypothesis that the regression disturbance ε_t is not autocorrelated against the alternative that it is AR(1). Test statistic based on an estimation sample $t = 1, \ldots, T$:

$$\text{DW} = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}.$$

LM test for autocorrelation (Breusch-Godfrey test): In the model

$$y_t = \mathbf{x}_t\boldsymbol{\gamma} + u_t, \qquad u_t = \rho_1 u_{t-1} + \ldots + \rho_r u_{t-r} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}$$

it tests $H_0 : \rho_1 = \ldots = \rho_r = 0$ against $H_1 : \neg H_0$. The test statistic is

$$LM = T\,\frac{\tilde{\boldsymbol{\varepsilon}}'\mathbf{P_Z}\tilde{\boldsymbol{\varepsilon}}}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}},$$

where $\tilde{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\gamma}}$ are the residuals obtained under H_0, $\mathbf{Z} = [\mathbf{X}, \tilde{\boldsymbol{\varepsilon}}_{-1}, \ldots, \tilde{\boldsymbol{\varepsilon}}_{-r}]$ and $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. Under H_0, $LM \xrightarrow{\text{d}} \chi^2_r$.

Information criteria: Let k be the number of parameters in the ADL model. The following information criteria can be used:

$$\begin{aligned}
\text{AIC}(k) &= \log(\hat{\sigma}^2) + \frac{2k}{T},\\
\text{AICc}(k) &= \log(\hat{\sigma}^2) + \frac{2k}{T} + \frac{2k^2 + 2k}{T - k - 1},\\
\text{HQ}(k) &= \log(\hat{\sigma}^2) + \frac{2k \log\log T}{T},\\
\text{BIC}(k) &= \log(\hat{\sigma}^2) + \frac{k \log T}{T}.
\end{aligned}$$

Part C. Econometrics for Integrated I(1) Time Series Processes

8 Stochastic Trends

8.1 Integrated Processes

A process $\{y_t\}$ is called integrated of order d (I(d)) if it is nonstationary and differencing it at least d times is necessary to obtain a stationary process. In particular, a nonstationary process is I(1) if its first difference, $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$ is I(0).

8.2 Stochastic Trends with iid Increments (Random Walks)

Stochastic trend without drift: Let $\{\varepsilon_t\}$ be a iid white noise process with mean zero and finite variance σ^2. A driftless stochastic trend with iid increments is defined as

$$s_t = s_0 + \varepsilon_1 + \cdots + \varepsilon_t = s_0 + \sum_{t=1}^{t} \varepsilon_t.$$

Assuming $s_0 = 0$, it has mean $\mathrm{E}(s_t) = 0$ and variance $\mathrm{Var}(s_t) = t\sigma^2$.

Stochastic trend with drift: Let $\{\varepsilon_t\}$ be a iid white noise process with mean zero and finite variance σ^2. A stochastic trend with drift and iid increments is defined as

$$s_t = s_0 + \mu t + \varepsilon_1 + \cdots + \varepsilon_t = s_0 + \sum_{t=1}^{t} (\mu + \varepsilon_t).$$

Assuming $s_0 = 0$, it has mean $\mathrm{E}(s_t) = \mu t$ and variance $\mathrm{Var}(s_t) = t\sigma^2$.

8.3 Stochastic Trends with Autocorrelated Increments

Stochastic trend without drift: Let $\{u_t\}$ be a stationary linear process with mean zero and long-run variance λ^2. A driftless stochastic trend with autocorrelated increments is defined as

$$\tilde{s}_t = \tilde{s}_0 + u_1 + \cdots + u_t = \tilde{s}_0 + \sum_{t=1}^{t} u_t.$$

Assuming $\tilde{s}_0 = 0$ and normalizing by $\sqrt{t}$, it has mean $\mathrm{E}(\tilde{s}_t/\sqrt{t}) = 0$ and asymptotic variance $\lim_{t\to\infty} \mathrm{Var}(\tilde{s}_t/\sqrt{t}) = \lambda^2$.

Stochastic trend with drift: Let $\{u_t\}$ be a stationary linear process with mean zero and long-run variance λ^2. A stochastic trend with drift and autocorrelated increments is defined as

$$\tilde{s}_t = \tilde{s}_0 + \mu t + u_1 + \cdots + u_t = \tilde{s}_0 + \sum_{t=1}^{t} (\mu + u_t).$$

Assuming $\tilde{s}_0 = 0$ and normalizing by $\sqrt{t}$, it has mean $\mathrm{E}(\tilde{s}_t/\sqrt{t}) = \mu\sqrt{t}$ and asymptotic variance $\lim_{t\to\infty} \mathrm{Var}(\tilde{s}_t/\sqrt{t}) = \lambda^2$.

8.4 Beveridge-Nelson Decomposition

Let $\tilde{s}_t$ be I(1) with drift so that $\Delta\tilde{s}_t$ is a linear I(0) process with long-run variance λ^2,

$$\tilde{s}_t = \tilde{s}_0 + \mu t + \sum_{i=1}^{t} u_i \qquad \Rightarrow \qquad \Delta\tilde{s}_t = \mu + u_t = \mu + \psi(L)\varepsilon_t,$$

where $\mathrm{Var}(\varepsilon_t) = \sigma^2$. Then $\tilde{s}_t$ can be linearly decomposed into (a) a linear deterministic trend, (b) a random walk, (c) a stationary process η, and (d) an initial condition $z_0 = \tilde{s}_0 - \eta_0$:

$$\tilde{s}_t = \underbrace{\mu t}_{\text{divergence at rate } t} + \underbrace{(\lambda/\sigma)\, s_t}_{\text{divergence at rate } \sqrt{t}} + \underbrace{\eta_t}_{\text{bounded in probability}} + \underbrace{z_0}_{\text{bounded in probability}} .$$

9 Unit Root Asymptotics

Standard Brownian motion (Wiener process): The standard Brownian motion $W(t)$, $t \in [0, 1]$ is defined as a stochastic process with the following properties.

- Initialization: $W(0) = 0$.
- Normality: $W(t)$ is distributed as $N(0, t)$.
- Independent increments: For given points in time $0 \leq t_1 < t_2 < \ldots t_k \leq 1$, the increments $(W(t_2) - W(t_1)), (W(t_3) - W(t_2)), \ldots, (W(t_k) - W(t_{k-1}))$ are stochastically independent. Any increment $W(s) - W(t)$, $s > t$, is normally distributed with mean zero and variance $s - t$.
- The process is continuous in t.

Functional central limit theorem: Let $\varepsilon_1, \ldots, \varepsilon_T$ be iid white noise random variables with mean zero and and variance σ^2. Furthermore define $[Tr]$, $0 \leq r \leq 1$, as the largest integer smaller or equal to Tr. Then the step-function

$$X_T(r) = \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t, \quad 0 \leq r \leq 1$$

with $X_T(r) = 0$ for $[Tr] < 1$ converges weakly towards a standard Brownian motion, $X_T(r) \Rightarrow W(r)$, as $T \to \infty$.

Continuous mapping theorem: By the continuous mapping theorem, if $X_T(r) \Rightarrow W(r)$ then $f(X_T(r)) \Rightarrow f(W(r))$ for continuous functions f.

Convergence of sample moments of random walks without drift: Let $\varepsilon_1, \ldots, \varepsilon_T$ be iid white noise random variables with mean zero and and variance σ^2 and define the random walk without drift $y_t = y_{t-1} + \varepsilon_t$, $y_0 = 0$. Then the following convergence results hold:

$$\begin{array}{lll} (1) & T^{-3/2} \sum_{t=1}^{T} y_{t-1} & \Rightarrow \sigma \int_0^1 W(r)dr \\ (2) & T^{-2} \sum_{t=1}^{T} y_{t-1}^2 & \Rightarrow \sigma^2 \int_0^1 (W(r))^2 dr \\ (3) & T^{-1} \sum_{t=1}^{T} \Delta y_t y_{t-1} & \Rightarrow 0.5\sigma^2 (W(1)^2 - 1) \end{array}$$

Convergence of sample moments of a stochastic trend with correlated increments: Let $\{u_t\}$ be a linear I(0) process with mean zero, finite variance γ_0 and finite long-run variance $\lambda^2 > 0$ and define the driftless stochastic trend $\tilde{y}_t = \tilde{y}_{t-1} + u_t$, $u_t = \psi(L)\varepsilon_t$, $\tilde{y}_0 = 0$. Then the following convergence results hold:

$$\begin{array}{lll} (1) & T^{-3/2} \sum_{t=1}^{T} \tilde{y}_{t-1} & \Rightarrow \lambda \int_0^1 W(r)dr \\ (2) & T^{-2} \sum_{t=1}^{T} \tilde{y}_{t-1}^2 & \Rightarrow \lambda^2 \int_0^1 (W(r))^2 dr \\ (3) & T^{-1} \sum_{t=1}^{T} \Delta \tilde{y}_t \tilde{y}_{t-1} & \Rightarrow 0.5\lambda^2 W(1)^2 - 0.5\gamma_0 \end{array}$$

10 Unit Root Tests

10.1 Dickey-Fuller (DF) Tests

Dickey-Fuller test without intercept and trend: Estimate by OLS the population model

$$y_t = \rho y_{t-1} + \varepsilon_t,$$

where ε_t is iid white noise with variance σ^2 and compute the test statistics

$$\text{DF-}\rho = T(\hat{\rho} - 1) = T\left(\frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} - 1\right)$$
$$\text{DF-}t = \frac{\hat{\rho} - 1}{SE(\hat{\rho})} = \frac{\hat{\rho} - 1}{\hat{\sigma}(\sum_{t=1}^{T} y_{t-1}^2)^{-1/2}}.$$

Under the null hypothesis $\rho = 1$, the test statistics have limiting distribution

$$\text{DF-}\rho \Rightarrow \text{DF}_\rho \equiv \frac{0.5(W(1)^2 - 1)}{\int_0^1 (W(r))^2 dr}$$
$$\text{DF-}t \Rightarrow \text{DF}_t \equiv \frac{0.5(W(1)^2 - 1)}{\sqrt{\int_0^1 (W(r))^2 dr}}.$$

Dickey-Fuller test with intercept: Estimate by OLS the population model

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t,$$

where ε_t is iid white noise with variance σ^2 and compute the test statistics

$$\text{DF-}\rho^\mu = T(\hat{\rho}^\mu - 1)$$
$$\text{DF-}t^\mu = \frac{\hat{\rho}^\mu - 1}{SE(\hat{\rho}^\mu)}.$$

Under the null hypothesis $\rho = 1$, the test statistics have limiting distribution

$$\text{DF-}\rho^\mu \Rightarrow \text{DF}_\rho^\mu \equiv \frac{0.5([W^\mu(1)]^2 - W^\mu(0)^2 - 1)}{\int_0^1 (W^\mu(r))^2 dr}$$
$$\text{DF-}t^\mu \Rightarrow \text{DF}_t^\mu \equiv \frac{0.5([W^\mu(1)]^2 - W^\mu(0)^2 - 1)}{\sqrt{\int_0^1 (W^\mu(r))^2 dr}}.$$

Dickey-Fuller test with intercept and trend: Estimate by OLS the population model

$$y_t = \alpha_0 + \alpha_1 t + \rho y_{t-1} + \varepsilon_t,$$

where ε_t is iid white noise with variance σ^2 and compute the test statistics

$$\text{DF-}\rho^\tau = T(\hat{\rho}^\tau - 1)$$
$$\text{DF-}t^\tau = \frac{\hat{\rho}^\tau - 1}{SE(\hat{\rho}^\tau)}.$$

Under the null hypothesis $\rho = 1$, the test statistics have limiting distribution

$$\text{DF-}\rho^{\tau} \Rightarrow \text{DF}^{\tau}_{\rho} \equiv \frac{0.5([W^{\tau}(1)]^2 - W^{\tau}(0)^2 - 1)}{\int_0^1 (W^{\tau}(r))^2 dr}$$
$$\text{DF-}t^{\tau} \Rightarrow \text{DF}^{\tau}_{t} \equiv \frac{0.5([W^{\tau}(1)]^2 - W^{\tau}(0)^2 - 1)}{\sqrt{\int_0^1 (W^{\tau}(r))^2 dr}}.$$

10.2 Phillips-Perron (PP) Test

Phillips-Perron test without intercept and trend: Estimate the DF-regression

$$y_t = \rho y_{t-1} + u_t,$$

where u_t is potentially autocorrelated. In addition, estimate the autocovariances of u_t, γ_k, up to order q, from which an estimate of the long-run variance, λ^2, is computed. Then, under the null of $\rho = 1$,

$$\text{PP-}\rho \equiv T(\hat{\rho} - 1) - 0.5 \left(T^2 (SE(\hat{\rho}))^2 / \hat{\sigma}_u^2\right) (\hat{\lambda}^2 - \hat{\sigma}_u^2)$$

converges to DF_ρ and

$$\text{PP-}t \equiv \sqrt{\frac{\hat{\sigma}_u^2}{\hat{\lambda}^2}} \frac{\hat{\rho} - 1}{SE(\hat{\rho})} - 0.5 \frac{\hat{\lambda}^2 - \hat{\sigma}_u^2}{\hat{\lambda}} (T \times SE(\hat{\rho}) / \hat{\sigma}_u)$$

converges to DF_t.

10.3 Augmented Dickey-Fuller (ADF) Tests

Augmented Dickey-Fuller test without intercept and trend: Estimate by OLS the augmented model

$$y_t = \rho y_{t-1} + \phi_1 \Delta \tilde{y}_{t-1} + \cdots + \phi_p \Delta \tilde{y}_{t-p} + v_t$$

and compute the t-statistic

$$\text{ADF-}t = \frac{\hat{\rho} - 1}{SE(\hat{\rho})}.$$

Under the null hypothesis $\rho = 1$, the ADF-t statistic has limiting distribution

$$\text{ADF-}t \Rightarrow \text{DF}_t.$$

Augmented Dickey-Fuller test with intercept: Estimate by OLS the augmented model

$$y_t = \alpha + \rho y_{t-1} + \phi_1 \Delta \tilde{y}_{t-1} + \cdots + \phi_p \Delta \tilde{y}_{t-p} + v_t$$

and compute the t-statistic

$$\text{ADF-}t^\mu = \frac{\hat{\rho}^\mu - 1}{SE(\hat{\rho}^\mu)}.$$

Under the null hypothesis $\rho = 1$, the ADF-t^μ statistic has limiting distribution

$$\text{ADF-}t^\mu \Rightarrow \text{DF}_t^\mu.$$

Augmented Dickey-Fuller test with intercept and trend: Estimate by OLS the augmented model

$$y_t = \alpha_0 + \alpha_1 t + \rho y_{t-1} + \phi_1 \Delta \tilde{y}_{t-1} + \cdots + \phi_p \Delta \tilde{y}_{t-p} + v_t$$

and compute the t-statistic

$$\text{ADF-}t^\tau = \frac{\hat{\rho}^\tau - 1}{SE(\hat{\rho}^\tau)}.$$

Under the null hypothesis $\rho = 1$, the ADF-t^τ statistic has limiting distribution

$$\text{ADF-}t^\tau \Rightarrow \text{DF}_t^\tau.$$

Choice of the lag order: Information criteria such as the AIC or BIC can be used to determine the lag order p. The maximum lag order $p_{\max}$ in this search can be chosen as the integer part of $12 \times (T/100)^{1/4}$.

11 Cointegration

11.1 Multivariate Beveridge-Nelson decomposition

Linear vector I(0) process: A linear n dimensional zero-mean vector I(0) process is

$$\mathbf{u}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \ldots = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\Psi}_0 = \mathbf{I},$$

where $\boldsymbol{\varepsilon}_t$ is iid with mean zero and positive definite variance matrix $\boldsymbol{\Omega}$, if

(C1) the sequence $\mathbf{I}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots$ is one-summable so that $\boldsymbol{\Psi}(1) = \mathbf{I} + \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2 + \cdots$ is finite,

(C2) at least one element of $\boldsymbol{\Psi}(1)$ is nonzero.

Long-run variance matrix: The long-run variance matrix of a linear vector I(0) process is

$$\boldsymbol{\Lambda} \equiv \lim_{T\to\infty} \text{Var}(\sqrt{T}\bar{\mathbf{u}}_T) = \boldsymbol{\Psi}(1)\boldsymbol{\Omega}\boldsymbol{\Psi}(1)'.$$

Vector I(1) process: A vector I(1) process of dimension n is

$$\Delta \mathbf{y}_t = \boldsymbol{\delta} + \mathbf{u}_t = \boldsymbol{\delta} + \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t,$$

where $\mathbf{u}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t$ is a linear zero-mean vector I(0) process.

Multivariate Beveridge-Nelson decomposition: A vector I(1) process of dimension n can be decomposed into

$$\mathbf{y}_t = \boldsymbol{\delta} \cdot t + \boldsymbol{\Psi}(1) \sum_{i=1}^{t} \boldsymbol{\varepsilon}_i + \boldsymbol{\eta}_t + \mathbf{z}_0,$$

where $\boldsymbol{\delta} \cdot t$ is a deterministic trend, $\boldsymbol{\Psi}(1)[\boldsymbol{\varepsilon}_1 + \cdots + \boldsymbol{\varepsilon}_t]$ is a stochastic trend, $\boldsymbol{\eta}_t$ is a linear vector I(0) process, and $\mathbf{z}_0$ is an initial condition.

11.2 Cointegration and common stochastic trends

Cointegration: Let $\mathbf{y}_t$ be an n-dimensional vector I(1) process with stochastic increment $\mathbf{u}_t$ that is a linear vector I(0) process. Then $\mathbf{y}_t$ is cointegrated with cointegration vector $\boldsymbol{\beta}_1$ if $\boldsymbol{\beta}_1 \neq \mathbf{0}$ and $\boldsymbol{\beta}_1'\mathbf{y}_t$ is stationary. (In a strict sense, stationarity of $\boldsymbol{\beta}_1'\mathbf{y}_t$ requires a suitable choice of initial conditions.)

Cointegration rank: The cointegration rank is the number, r, of linearly independent cointegration vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$, and the cointegration space is the space spanned by the cointegration vectors.

Cointegration matrix: The $(n \times r)$ cointegration matrix $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r]$ is the matrix of all cointegration vectors. The cointegration matrix has rank r.

Rank of $\boldsymbol{\Psi}(1)$: The matrix $\boldsymbol{\Psi}(1)$ of the multivariate Beveridge-Nelson decomposition has rank $k \equiv n - r$.

Common stochastic trends: A cointegrated n-dimensional vector I(1) process with cointegration rank r is driven by $k = n - r$ common stochastic trends.

11.3 Engle-Granger test of the null of no cointegration

Spurious regression: Suppose $\mathbf{y}_t$ is an n-dimensional vector I(1) process and not cointegrated. Then the regression

$$y_{1,t} = \mu + \gamma_2 y_{2,t} + \cdots + \gamma_n y_{n,t} + z_t^*$$

is called spurious. The disturbance z_t^* is I(1) and the OLS estimator has a non-standard asymptotic distribution.

Cointegrating regression: Suppose $\mathbf{y}_t$ is an n-dimensional vector I(1) process and cointegrated with cointegration rank $r = 1$. Then the regression

$$y_{1,t} = \mu + \gamma_2 y_{2,t} + \cdots + \gamma_n y_{n,t} + z_t^*,$$

is called cointegrating regression. The disturbance z_t^* is I(0). The OLS estimator is superconsistent but has a non-standard asymptotic distribution.

Engle-Granger test: Suppose $\mathbf{y}_t$ is an n-dimensional vector I(1) and either cointegrated with cointegration rank $r = 1$ or not cointegrated. The hypotheses H_0: $r = 0$ versus H_1: $r = 1$ can be tested as follows.

Stage 1. Estimate by OLS the regression

$$y_{1,t} = \hat{\mu} + \hat{\gamma}_2 y_{2,t} + \cdots + \hat{\gamma}_n y_{n,t} + \hat{z}_t^*.$$

Stage 2. Estimate the ADF regression without intercept and trend on the first-stage residuals

$$\hat{z}_t^* = \rho \hat{z}_{t-1}^* + \phi_1 \Delta \hat{z}_{t-1}^* + \cdots + \phi_p \Delta \hat{z}_{t-p}^* + v_t.$$

Stage 3. Construct the ADF statistic

$$\text{ADF-}t = \frac{\hat{\rho} - 1}{SE(\hat{\rho})}$$

and compare it with the appropriate critical values.

Part D. Tables

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

Critical values for the DF tests

Table 1: Empirical cumulative distribution of the DF-ρ statistic under $H_0 : \rho = 1$

Sample size (T)	Probability that the statistic is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
(a) No intercept, no trend								
25	-11.8	-9.3	-7.3	-5.3	1.01	1.41	1.78	2.28
50	-12.8	-9.9	-7.7	-5.5	0.97	1.34	1.69	2.16
100	-13.3	-10.2	-7.9	-5.6	0.95	1.31	1.65	2.09
250	-13.6	-10.4	-8.0	-5.7	0.94	1.29	1.62	2.05
500	-13.7	-10.4	-8.0	-5.7	0.93	1.28	1.61	2.04
∞	-13.8	-10.5	-8.1	-5.7	0.93	1.28	1.60	2.03
(b) Intercept, no trend								
25	-17.2	-14.6	-12.5	-10.2	-0.76	0.00	0.65	1.39
50	-18.9	-15.7	-13.3	-10.7	-0.81	-0.07	0.53	1.22
100	-19.8	-16.3	-13.7	-11.0	-0.83	-0.11	0.47	1.14
250	-20.3	-16.7	-13.9	-11.1	-0.84	-0.13	0.44	1.08
500	-20.5	-16.8	-14.0	-11.2	-0.85	-0.14	0.42	1.07
∞	-20.7	-16.9	-14.1	-11.3	-0.85	-0.14	0.41	1.05
(c) Intercept and trend								
25	-22.5	-20.0	-17.9	-15.6	-3.65	-2.51	-1.53	-0.46
50	-25.8	-22.4	-19.7	-16.8	-3.71	-2.60	-1.67	-0.67
100	-27.4	-23.7	-20.6	-17.5	-3.74	-2.63	-1.74	-0.76
250	-28.5	-24.4	-21.3	-17.9	-3.76	-2.65	-1.79	-0.83
500	-28.9	-24.7	-21.5	-18.1	-3.76	-2.66	-1.80	-0.86
∞	-29.4	-25.0	-21.7	-18.3	-3.77	-2.67	-1.81	-0.88

Source: Hayashi (2000), p. 576.

Table 2: Empirical cumulative distribution of the DF-t statistic under $H_0 : \rho = 1$

Sample size (T)	Probability that the statistic is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
(a) No intercept, no trend								
25	-2.65	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.15
50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
100	-2.60	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.04
250	-2.58	-2.24	-1.95	-1.62	0.89	1.28	1.63	2.02
500	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.01
∞	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.01
(b) Intercept, no trend								
25	-3.75	-3.33	-2.99	-2.64	-0.37	0.00	0.34	0.72
50	-3.59	-3.23	-2.93	-2.60	-0.41	-0.04	0.28	0.66
100	-3.50	-3.17	-2.90	-2.59	-0.42	-0.05	0.26	0.63
250	-3.45	-3.14	-2.88	-2.58	-0.42	-0.06	0.24	0.62
500	-3.44	-3.13	-2.87	-2.57	-0.44	-0.07	0.24	0.61
∞	-3.42	-3.12	-2.86	-2.57	-0.44	-0.08	0.23	0.60
(c) Intercept and trend								
25	-4.38	-3.95	-3.60	-3.24	-1.14	-0.81	-0.50	-0.15
50	-4.15	-3.80	-3.50	-3.18	-1.19	-0.87	-0.58	-0.24
100	-4.05	-3.73	-3.45	-3.15	-1.22	-0.90	-0.62	-0.28
250	-3.98	-3.69	-3.42	-3.13	-1.23	-0.92	-0.64	-0.31
500	-3.97	-3.67	-3.42	-3.13	-1.24	-0.93	-0.65	-0.32
∞	-3.96	-3.66	-3.41	-3.12	-1.25	-0.94	-0.66	-0.33

Source: Hayashi (2000), p. 578.

Critical values for the ADF test for cointegration (Engle-Granger test)

Number of regressors, excluding constant	1%	2.5%	5%	10%
A. Regressors have no drift				
1	-3.96	-3.64	-3.37	-3.07
2	-4.31	-4.02	-3.77	-3.45
3	-4.73	-4.37	-4.11	-3.83
4	-5.07	-4.71	-4.45	-4.16
5	-5.28	-4.98	-4.71	-4.43
B. Some regressors have drift				
1	-3.96	-3.67	-3.41	-3.13
2	-4.36	-4.07	-3.80	-3.52
3	-4.65	-4.39	-4.16	-3.84
4	-5.04	-4.77	-4.49	-4.20
5	-5.36	-5.02	-4.74	-4.46

Source: Hayashi (2000), Table 10.1, p. 646.