Contents

A Linear algebra basics **1**
A.1 Basic operations 1
A.2 Fundamental theorem of linear algebra 4
A.2.1 Part one 4
A.2.2 Part two 5
A.3 Nature of the solution 7
A.3.1 Exactly-identified 7
A.3.2 Under-identified 8
A.3.3 Over-identified 10
A.4 Matrix decomposition and inverse operations 12
A.4.1 LU factorization 12
A.4.2 Cholesky decomposition 16
A.4.3 Eigenvalues and eigenvectors 17
A.4.4 Singular value decomposition 23
A.4.5 Spectral decomposition 28
A.4.6 quadratic forms, eigenvalues, and positive definiteness 29
A.4.7 similar matrices, Jordan form, and generalized eigenvectors 29
A.5 Gram-Schmidt construction of an orthogonal matrix 32
A.5.1 QR decomposition 34
A.5.2 Gram-Schmidt QR algorithm 34
A.5.3 Accounting example 35
A.5.4 The Householder QR algorithm 36
A.5.5 Accounting example 36

A.6 Computing eigenvalues 39
A.6.1 Schur's lemma 39
A.6.2 Power algorithm 40
A.6.3 QR algorithm 40
A.6.4 Schur decomposition 42
A.7 Some determinant identities 46
A.7.1 Determinant of a square matrix 46
A.7.2 Identities 47
A.8 Matrix exponentials and logarithms 49

B Iterated expectations 51
B.1 Decomposition of variance 53
B.2 Jensen's inequality 54

C Multivariate normal theory 55
C.1 Conditional distribution 57
C.2 Special case of precision 61
C.3 Truncated normal distribution 63

D Projections and conditional expectations 71
D.1 Gauss-Markov theorem 71
D.2 Generalized least squares (GLS) 74
D.3 Recursive least squares 76

E Two stage least squares IV (2SLS-IV) 79
E.1 General case 79
E.2 Special case 81

F Seemingly unrelated regression (SUR) 83
F.1 Classical 84
F.2 Bayesian 84
F.3 Bayesian treatment effect application 85

G Maximum likelihood estimation of discrete choice models 87

H Optimization 89
H.1 Linear programming 89
H.1.1 basic solutions or extreme points 90
H.1.2 fundamental theorem of linear programming 91
H.1.3 duality theorems 91
H.1.4 example 92
H.1.5 complementary slackness 92
H.2 Nonlinear programming 93
H.2.1 unconstrained 93
H.2.2 convexity and global minima 93
H.2.3 example 94

H.2.4 constrained — the Lagrangian 94
H.2.5 Karash-Kuhn-Tucker conditions 95
H.2.6 example 96
H.3 Theorem of the separating hyperplane 97

I Quantum information **99**
I.1 Quantum information axioms 99
I.1.1 The superposition axiom 99
I.1.2 The transformation axiom 100
I.1.3 The measurement axiom 100
I.1.4 The combination axiom 101
I.2 Summary of quantum "rules" 103
I.3 Observables and expected payoffs 104
I.4 Density operators and quantum entropy 105

J Common distributions **109**

Appendix G

Maximum likelihood estimation of discrete choice models

The most common method for estimating the parameters of discrete choice models is maximum likelihood. The likelihood is defined as the joint density for the parameters of interest θ conditional on the data X_t. For binary choice models and $D_t = 1$ the contribution to the likelihood is $F\left(X_t^T\theta\right)$, and for $D_t = 0$ the contribution to the likelihood is $1-F\left(X_t^T\theta\right)$ where these are combined as binomial draws and $F\left(X_t^T\theta\right)$ is the cumulative distribution function evaluated at $X_t^T\theta$. Hence, the likelihood is

$$L\left(\theta|X\right) = \prod_{t=1}^{n} F\left(X_t^T\theta\right)^{D_t}\left[1-F\left(X_t^T\theta\right)\right]^{1-D_t}$$

The log-likelihood is

$$\ell\left(\theta|X\right) \equiv logL\left(\theta|X\right) = \sum_{t=1}^{n} D_t log\left(F\left(X_t^T\theta\right)\right) + \left(1-D_t\right) log\left(1-F\left(X_t^T\theta\right)\right)$$

Since this function for binary response models like probit and logit is globally concave, numerical maximization is straightforward. The first order conditions for a maximum, $\max_{\theta} \ell\left(\theta|X\right)$, are

$$\sum_{t=1}^{n} \frac{D_t f\left(X_t^T\theta\right)X_{it}}{F\left(X_t^T\theta\right)} - \frac{\left(1-D_t\right) f\left(X_t^T\theta\right)X_{ti}}{1-F\left(X_t^T\theta\right)} = 0 \quad i = 1,\ldots,k$$

where $f\left(\cdot\right)$ is the density function. Simplifying yields

$$\sum_{t=1}^{n} \frac{\left[D_t - F\left(X_t^T\theta\right)\right] f\left(X_t^T\theta\right)X_{ti}}{F\left(X_t^T\theta\right)\left[1-F\left(X_t^T\theta\right)\right]} = 0 \quad i = 1,\ldots,k$$

Estimates of θ are found by solving these first order conditions iteratively or, in other words, numerically.

A common estimator for the variance of $\hat{\theta}_{MLE}$ is the negative inverse of the Hessian matrix evaluated at $\hat{\theta}_{MLE}$, $\left[-H\left(D,\hat{\theta}\right)\right]^{-1}$. Let $H\left(D,\theta\right)$ be the Hessian matrix for the log-likelihood with typical element $H_{ij}\left(D,\theta\right) \equiv \frac{\partial^2 \ell_t(D,\theta)}{\partial\theta_i \partial\theta_j}$.[1]

[1] Details can be found in numerous econometrics references and chapter 4 of *Accounting and Causal Effects: Econometric Challenges*.