Contents

1 Introduction and Overview 1

2 Detecting Influential Observations and Outliers 6

2.1 Theoretical Foundations, 9
Single-Row Effects, 12
Deletion, 12
Coefficients and Fitted Values. The Hat Matrix. Residuals. Covariance Matrix.
Differentiation, 24
A Geometric View, 26
Criteria for Influential Observations, 27
External Scaling. Internal Scaling. Gaps.
Partial-Regression Leverage Plots, 30
Multiple-Row Effects, 31
Deletion, 31
Studentized Residuals and Dummy Variables, 33
Differentiation, 35
Geometric Approaches, 37
Final Comments, 38

2.2 Application: an Intercountry Life-Cycle Savings Function, 39
A Diagnostic Analysis of the Model, 39
The Model and Regression Results, 40
Single-Row Diagnostics, 42
Residuals. Leverage and Hat-Matrix Diagonals. Coefficient Sensitivity. Covariance Matrix Sensitivity. Change in Fit. Internal Scaling. A Provisional Summary.

Multiple-Row Diagnostics, 51
Partial-Regression Leverage Plots: a Preliminary Analysis. Using Multiple-Row Methods. Deletion. Residuals. Differentiation. Geometry.
Final Comments, 63

Appendix 2A: Additional Theoretical Background, 64
Deletion Formulas, 64
Differentiation Formulas, 65
Theorems Related to the Hat Matrix, 66
Size of the Diagonal Elements. Distribution Theory. Dummy Variables and Singular Matrices.

Appendix 2B: Computational Elements, 69
Computational Elements for Single-Row Diagnostics, 69
Orthogonal Decompositions, the Least-Squares Solution, and Related Statistics. The Diagonal Elements of the Hat Matrix. Computing the DFBETA.
Computational Elements for Multiple-Row Diagnostics, 75
Notation and the Subset Tree. An Algorithm for the Geometric Measure, Wilks' Λ Dummy Variables, Sequential Choleski Decomposition, and the Andrews-Pregibon Statistic. Further Elements Computed from the Triangular Factors. Inequalities Related to MDFFIT.

3 Detecting and Assessing Collinearity **85**

3.1 Introduction and Historical Perspective, 85
Overview, 91
Historical Perspective, 92
A Basis for a Diagnostic, 96

3.2 Technical Background, 98
The Singular-Value Decomposition, 98
Exact Linear Dependencies: Rank Deficiency, 99
The Condition Number, 100
Near Linear Dependencies: How Small is Small?, 104
The Regression-Coefficient Variance Decomposition, 105
Two Interpretive Considerations, 107

Near Collinearity Nullified by Near Orthogonality, 107
At Least Two Variates Must Be Involved, 108
An Example, 110
A Suggested Diagnostic Procedure, 112
The Diagnostic Procedure, 112
Examining the Near Dependencies, 113
What is "Large" or "High," 114
The Ill Effects of Collinearity, 114
Computational Problems, 114
Statistical Problems, 115
Harmful Versus Degrading Collinearity, 115

3.3 Experimental Experience, 117
The Experimental Procedure, 117
The Choice of the X's, 119
Experimental Shortcomings, 119
The Need for Column Scaling, 120
The Experimental Report, 121
The Individual Experiments, 121
The Results, 125

3.4 Summary Interpretation, and Examples of Diagnosing Actual Data for Collinearity, 152
Interpreting the Diagnostic Results: a Summary of the Experimental Evidence, 152
Experience with a Single Near Dependency, 153
Experience with Coexisting Near Dependencies, 154
Employing the Diagnostic Procedure, 156
The Steps, 157
Forming the Auxiliary Regressions, 159
Software, 160
Applications with Actual Data, 160
The Bauer Matrix, 161
The Consumption Function, 163
The Friedman Data, 167
An Equation of the IBM Econometric Model, 169

Appendix 3A: The Condition Number and Invertibility, 173

Appendix 3B: Parameterization and Scaling, 177
The Effects on the Collinearity Diagnostics Due to Linear Transformations of the Data, 177

Each Parameterization is a Different Problem, 178
A More General Analysis, 180
Column Scaling, 183

Appendix 3C: The Weakness of Correlation Measures in Providing Diagnostic Information, 185

Appendix 3D: The Harm Caused by Collinearity, 186
The Basic Harm, 187
The Effect of Collinearity, 190

4 Applications and Remedies **192**

4.1 A Remedy for Collinearity: the Consumption Function with Mixed-Estimation, 193
Corrective Measures, 193
Introduction of New Data, 193
Bayesian-Type Techniques, 194
Pure Bayes. Mixed-Estimation. Ridge Regression.
Application to the Consumption-Function Data, 196
Prior Restrictions, 197
Ignored Information, 199
Summary of Prior Data, 200
Regression Results and Variance-Decomposition Proportions for Mixed-Estimation Consumption-Function Data, 200

4.2 Row-Deletion Diagnostics with Mixed-Estimation of the U.S. Consumption Function, 204
A Diagnostic Analysis of the Consumption-Function Data, 204
Single-Row Diagnostics, 205
Residuals. Leverage and Hat-Matrix Diagonals. Coefficient Sensitivity.
Summary, 207
A Reanalysis after Remedial Action for Ill Conditioning, 207
The Row Diagnostics, 208
A Suggested Research Strategy, 210

4.3 An Analysis of an Equation Describing the Household Demand for Corporate Bonds, 212

An Examination of Parameter Instability and Sensitivity, 215
Tests for Overall Structural Instability, 215
Sensitivity Diagnostics, 217
Residuals. Leverage and Coefficient Sensitivity.
The Monetary Background, 219
A Use of Ridge Regression, 219
Summary, 228

4.4 Robust Estimation of a Hedonic Housing-Price Equation, 229
The Model, 231
Robust Estimation, 232
Partial Plots, 235
Single-Row Diagnostics, 237
Multiple-Row Diagnostics, 241
Summary, 243

Appendix 4A: Harrison and Rubinfeld Housing-Price Data, 245

5 Research Issues and Directions for Extensions 262

5.1 Issues in Research Strategy, 263

5.2 Extensions of the Diagnostics, 266
Extensions to Systems of Simultaneous Equations, 266
Influential-Data Diagnostics, 266
Collinearity Diagnostics, 268
Extensions to Nonlinear Models, 269
Influential-Data Diagnostics, 269
Collinearity Diagnostics, 272
Additional Topics, 274
Bounded-Influence Regression, 274
Multiple-Row Procedures, 274
Transformations, 275
Time Series and Lags, 276

Bibliography 277

Author Index 285

Subject Index 287