

# Linear Models 1

Isfahan University of Technology  
Fall Semester, 2014

## References:



[1] G. A. F., Seber and A. J. Lee (2003). Linear Regression Analysis (2nd ed.). Hoboken, NJ: Wiley.



[2] A. C. Rencher and G. B. Schaalje (2008). Linear Models in Statistics (2nd ed.). John Wiley & Sons, Inc.

[3] R. B., Bapat (2000), Linear Algebra and Linear Models, (2<sup>nd</sup> ed.). Springer-Verlag.

[4] S. R., Searle (1971). Linear Models. New York: Wiley.

The PDF versions of references are available in the link <http://rikhtehgaran.iut.ac.ir/>

## Prerequisites:

- Linear Algebra
- Basic Statistical Inference.

## Grading Policy:

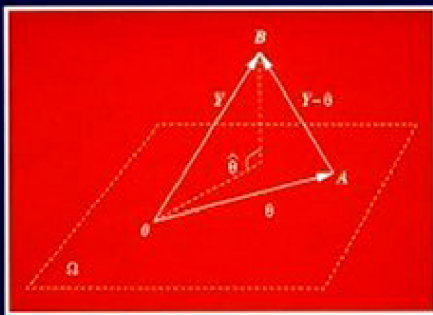
- **Assignments & Quizzes:** *3-5 points.*
- **Midterm Exam:** *5-7 points .*
- **Final Exam:** *10 points.*

Copyrighted Material

WILEY

# Linear Regression Analysis

Second Edition



George A. F. Seber and Alan J. Lee

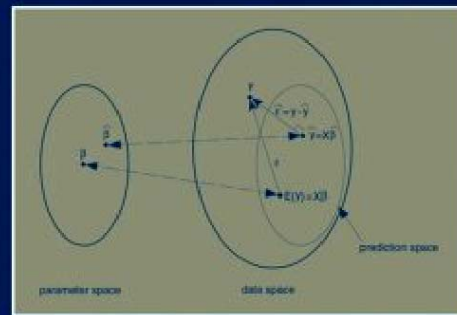
WILEY SERIES IN PROBABILITY AND STATISTICS

Copyrighted Material

WILEY

# Linear Models in Statistics

Second Edition



ALVIN C. RENCHER • G. BRUCE SCHALJE

www.wiley.com



# Contents

## Ch. 2. Matrix Algebra. ( Rencher and Schaalje )

|   |    |
|---|----|
| Preface   | xv |
| 1 Vectors of Random Variables                           | 1  |
| 1.1 Notation  | 1  |
| 1.2 Statistical Models                                  | 2  |
| 1.3 Linear Regression Models                            | 4  |
| 1.4 Expectation and Covariance Operators                | 5  |
| Exercises 1a  | 8  |
| 1.5 <u>Mean and Variance of Quadratic Forms</u>         | 9  |
| Exercises 1b  | 12 |
| 1.6 <u>Moment Generating Functions and Independence</u> | 13 |
| Exercises 1c  | 15 |
| Miscellaneous Exercises 1                               | 15 |
| 2 Multivariate Normal Distribution                      | 17 |
| 2.1 <u>Density Function</u>                             | 17 |
| Exercises 2a  | 19 |
| 2.2 Moment Generating Functions                         | 20 |
| Exercises 2b  | 23 |
| 2.3 Statistical Independence                            | 24 |

|      |  |    |
|------|--|----|
|      | Exercises 2c   | 26 |
| 2.4  | <u>Distribution of Quadratic Forms</u>                       | 27 |
|      | Exercises 2d   | 31 |
|      | Miscellaneous Exercises 2                                    | 31 |
| 3    | <b>Linear Regression: Estimation and Distribution Theory</b> | 35 |
| 3.1  | <u>Least Squares Estimation</u>                              | 35 |
|      | Exercises 3a   | 41 |
| 3.2  | Properties of Least Squares Estimates                        | 42 |
|      | Exercises 3b   | 44 |
| 3.3  | Unbiased Estimation of $\sigma^2$                            | 44 |
|      | Exercises 3c   | 47 |
| 3.4  | Distribution Theory  | 47 |
|      | Exercises 3d   | 49 |
| 3.5  | <u>Maximum Likelihood Estimation</u>                         | 49 |
| 3.6  | Orthogonal Columns in the Regression Matrix                  | 51 |
|      | Exercises 3e   | 52 |
| 3.7  | Introducing Further Explanatory Variables                    | 54 |
|      | 3.7.1 General Theory   | 54 |
|      | 3.7.2 One Extra Variable                                     | 57 |
|      | Exercises 3f   | 58 |
| 3.8  | <u>Estimation with Linear Restrictions</u>                   | 59 |
|      | 3.8.1 Method of Lagrange Multipliers                         | 60 |
|      | 3.8.2 Method of Orthogonal Projections                       | 61 |
|      | Exercises 3g   | 62 |
| 3.9  | <u>Design Matrix of Less Than Full Rank</u>                  | 62 |
|      | 3.9.1 Least Squares Estimation                               | 62 |
|      | Exercises 3h   | 64 |
|      | 3.9.2 Estimable Functions                                    | 64 |
|      | Exercises 3i   | 65 |
|      | 3.9.3 Introducing Further Explanatory Variables              | 65 |
|      | 3.9.4 Introducing Linear Restrictions                        | 65 |
|      | Exercises 3j   | 66 |
| 3.10 | <u>Generalized Least Squares</u>                             | 66 |
|      | Exercises 3k   | 69 |
| 3.11 | Centering and Scaling the Explanatory Variables              | 69 |
|      | 3.11.1 Centering   | 70 |
|      | 3.11.2 Scaling   | 71 |

|  |     |
|--|-----|
| Exercises 3l   | 72  |
| <del>3.12 Bayesian Estimation</del>                          | 73  |
| Exercises 3m   | 76  |
| <del>3.13 Robust Regression</del>                            | 77  |
| 3.13.1 M-Estimates   | 78  |
| 3.13.2 Estimates Based on Robust Location and Scale Measures | 80  |
| 3.13.3 Measuring Robustness                                  | 82  |
| 3.13.4 Other Robust Estimates                                | 88  |
| Exercises 3n   | 93  |
| Miscellaneous Exercises 3                                    | 93  |
| <b>4 Hypothesis Testing</b>                                  | 97  |
| 4.1 Introduction   | 97  |
| 4.2 <u>Likelihood Ratio Test</u>                             | 98  |
| 4.3 <u>F-Test</u>  | 99  |
| 4.3.1 Motivation   | 99  |
| 4.3.2 Derivation   | 99  |
| Exercises 4a   | 102 |
| 4.3.3 Some Examples  | 103 |
| 4.3.4 The Straight Line                                      | 107 |
| Exercises 4b   | 109 |
| 4.4 Multiple Correlation Coefficient                         | 110 |
| Exercises 4c   | 113 |
| <del>4.5 Canonical Form for <math>H</math></del>             | 113 |
| Exercises 4d   | 114 |
| 4.6 <del>Goodness-of-Fit Test</del>                          | 115 |
| 4.7 <del>F-Test and Projection Matrices</del>                | 116 |
| Miscellaneous Exercises 4                                    | 117 |
| <b>5 Confidence Intervals and Regions</b>                    | 119 |
| 5.1 Simultaneous Interval Estimation                         | 119 |
| 5.1.1 Simultaneous Inferences                                | 119 |
| 5.1.2 Comparison of Methods                                  | 124 |
| 5.1.3 Confidence Regions                                     | 125 |
| 5.1.4 Hypothesis Testing and Confidence Intervals            | 127 |
| 5.2 Confidence Bands for the Regression Surface              | 129 |
| 5.2.1 Confidence Intervals                                   | 129 |
| 5.2.2 Confidence Bands                                       | 129 |

|              |   |     |
|--------------|---|-----|
| 5.3          | Prediction Intervals and Bands for the Response     | 131 |
| 5.3.1        | Prediction Intervals                                | 131 |
| 5.3.2        | Simultaneous Prediction Bands                       | 133 |
| 5.4          | Enlarging the Regression Matrix                     | 135 |
|              | Miscellaneous Exercises 5                           | 136 |
| <b>6</b>     | <b>Straight-Line Regression</b>                     | 139 |
| 6.1          | The Straight Line                                   | 139 |
| 6.1.1        | Confidence Intervals for the Slope and Intercept    | 139 |
| 6.1.2        | Confidence Interval for the $x$ -Intercept          | 140 |
| 6.1.3        | Prediction Intervals and Bands                      | 141 |
| 6.1.4        | Prediction Intervals for the Response               | 145 |
| 6.1.5        | Inverse Prediction (Calibration)                    | 145 |
|              | Exercises 6a  | 148 |
| 6.2          | Straight Line through the Origin                    | 149 |
| 6.3          | <u>Weighted Least Squares for the Straight Line</u> | 150 |
| 6.3.1        | Known Weights                                       | 150 |
| 6.3.2        | Unknown Weights                                     | 151 |
|              | Exercises 6b  | 153 |
| 6.4          | Comparing Straight Lines                            | 154 |
| 6.4.1        | General Model                                       | 154 |
| 6.4.2        | Use of Dummy Explanatory Variables                  | 156 |
|              | Exercises 6c  | 157 |
| 6.5          | <u>Two-Phase Linear Regression</u>                  | 159 |
| 6.6          | Local Linear Regression                             | 162 |
|              | Miscellaneous Exercises 6                           | 163 |
| <del>7</del> | <del>Polynomial Regression</del>                    | 165 |
| 7.1          | Polynomials in One Variable                         | 165 |
| 7.1.1        | Problem of Ill-Conditioning                         | 165 |
| 7.1.2        | Using Orthogonal Polynomials                        | 166 |
| 7.1.3        | Controlled Calibration                              | 172 |
| 7.2          | Piecewise Polynomial Fitting                        | 172 |
| 7.2.1        | Unsatisfactory Fit                                  | 172 |
| 7.2.2        | Spline Functions                                    | 173 |
| 7.2.3        | Smoothing Splines                                   | 176 |
| 7.3          | Polynomial Regression in Several Variables          | 180 |
| 7.3.1        | Response Surfaces                                   | 180 |

|          |   |     |
|----------|---|-----|
| 7.3.2    | Multidimensional Smoothing  | 184 |
|          | Miscellaneous Exercises 7   | 185 |
| <b>8</b> | <b>Analysis of Variance</b>                                       | 187 |
| 8.1      | Introduction  | 187 |
| 8.2      | <u>One-Way Classification</u>                                     | 188 |
| 8.2.1    | General Theory  | 188 |
| 8.2.2    | Confidence Intervals  | 192 |
| 8.2.3    | Underlying Assumptions  | 195 |
|          | Exercises 8a  | 196 |
| 8.3      | <del>Two-Way Classification (Unbalanced)</del>                    | 197 |
| 8.3.1    | Representation as a Regression Model                              | 197 |
| 8.3.2    | Hypothesis Testing  | 197 |
| 8.3.3    | Procedures for Testing the Hypotheses                             | 201 |
| 8.3.4    | Confidence Intervals  | 204 |
|          | Exercises 8b  | 205 |
| 8.4      | <del>Two-Way Classification (Balanced)</del>                      | 206 |
|          | Exercises 8c  | 209 |
| 8.5      | <del>Two-Way Classification (One Observation per Mean)</del>      | 211 |
| 8.5.1    | Underlying Assumptions  | 212 |
| 8.6      | <del>Higher-Way Classifications with Equal Numbers per Mean</del> | 216 |
| 8.6.1    | Definition of Interactions  | 216 |
| 8.6.2    | Hypothesis Testing  | 217 |
| 8.6.3    | Missing Observations  | 220 |
|          | Exercises 8d  | 221 |
| 8.7      | <del>Designs with Simple Block Structure</del>                    | 221 |
| 8.8      | <u>Analysis of Covariance</u>                                     | 222 |
|          | Exercises 8e  | 224 |
|          | Miscellaneous Exercises 8   | 225 |
| <b>9</b> | <b>Departures from Underlying Assumptions</b>                     | 227 |
| 9.1      | Introduction  | 227 |
| 9.2      | Bias  | 228 |
| 9.2.1    | Bias Due to Underfitting  | 228 |
| 9.2.2    | Bias Due to Overfitting   | 230 |
|          | Exercises 9a  | 231 |
| 9.3      | Incorrect Variance Matrix   | 231 |
|          | Exercises 9b  | 232 |

|           |  |            |
|-----------|--|------------|
| 9.4       | Effect of Outliers   | 233        |
| 9.5       | Robustness of the $F$ -Test to Nonnormality                | 235        |
| 9.5.1     | Effect of the Regressor Variables                          | 235        |
| 9.5.2     | Quadratically Balanced $F$ -Tests                          | 236        |
|           | Exercises 9c   | 239        |
| 9.6       | Effect of Random Explanatory Variables                     | 240        |
| 9.6.1     | Random Explanatory Variables Measured without Error        | 240        |
| 9.6.2     | Fixed Explanatory Variables Measured with Error            | 241        |
| 9.6.3     | Round-off Errors   | 245        |
| 9.6.4     | Some Working Rules   | 245        |
| 9.6.5     | Random Explanatory Variables Measured with Error           | 246        |
| 9.6.6     | Controlled Variables Model                                 | 248        |
| 9.7       | Collinearity   | 249        |
| 9.7.1     | Effect on the Variances of the Estimated Coefficients      | 249        |
| 9.7.2     | Variance Inflation Factors                                 | 254        |
| 9.7.3     | Variances and Eigenvalues                                  | 255        |
| 9.7.4     | Perturbation Theory  | 255        |
| 9.7.5     | Collinearity and Prediction                                | 261        |
|           | Exercises 9d   | 261        |
|           | Miscellaneous Exercises 9                                  | 262        |
| <b>10</b> | <b>Departures from Assumptions: Diagnosis and Remedies</b> | <b>265</b> |
| 10.1      | Introduction   | 265        |
| 10.2      | Residuals and Hat Matrix Diagonals                         | 266        |
|           | Exercises 10a  | 270        |
| 10.3      | Dealing with Curvature                                     | 271        |
| 10.3.1    | Visualizing Regression Surfaces                            | 271        |
| 10.3.2    | Transforming to Remove Curvature                           | 275        |
| 10.3.3    | Adding and Deleting Variables                              | 277        |
|           | Exercises 10b  | 279        |
| 10.4      | Nonconstant Variance and Serial Correlation                | 281        |
| 10.4.1    | Detecting Nonconstant Variance                             | 281        |
| 10.4.2    | Estimating Variance Functions                              | 288        |
| 10.4.3    | Transforming to Equalize Variances                         | 291        |
| 10.4.4    | Serial Correlation and the Durbin–Watson Test              | 292        |
|           | Exercises 10c  | 294        |
| 10.5      | Departures from Normality                                  | 295        |
| 10.5.1    | Normal Plotting  | 295        |



|        |  |     |
|--------|--|-----|
| 10.5.2 | Transforming the Response                                    | 297 |
| 10.5.3 | Transforming Both Sides                                      | 299 |
|        | Exercises 10d  | 300 |
| 10.6   | Detecting and Dealing with Outliers                          | 301 |
| 10.6.1 | Types of Outliers  | 301 |
| 10.6.2 | Identifying High-Leverage Points                             | 304 |
| 10.6.3 | Leave-One-Out Case Diagnostics                               | 306 |
| 10.6.4 | Test for Outliers  | 310 |
| 10.6.5 | Other Methods  | 311 |
|        | Exercises 10e  | 314 |
| 10.7   | Diagnosing Collinearity                                      | 315 |
| 10.7.1 | Drawbacks of Centering                                       | 316 |
| 10.7.2 | Detection of Points Influencing Collinearity                 | 319 |
| 10.7.3 | Remedies for Collinearity                                    | 320 |
|        | Exercises 10f  | 326 |
|        | Miscellaneous Exercises 10                                   | 327 |
| 11     | <del>Computational Algorithms for Fitting a Regression</del> | 329 |
| 11.1   | Introduction   | 329 |
|        | 11.1.1 Basic Methods   | 329 |
| 11.2   | Direct Solution of the Normal Equations                      | 330 |
|        | 11.2.1 Calculation of the Matrix $X'X$                       | 330 |
|        | 11.2.2 Solving the Normal Equations                          | 331 |
|        | Exercises 11a  | 337 |
| 11.3   | QR Decomposition   | 338 |
|        | 11.3.1 Calculation of Regression Quantities                  | 340 |
|        | 11.3.2 Algorithms for the QR and WU Decompositions           | 341 |
|        | Exercises 11b  | 352 |
| 11.4   | Singular Value Decomposition                                 | 353 |
|        | 11.4.1 Regression Calculations Using the SVD                 | 353 |
|        | 11.4.2 Computing the SVD                                     | 354 |
| 11.5   | Weighted Least Squares                                       | 355 |
| 11.6   | Adding and Deleting Cases and Variables                      | 356 |
|        | 11.6.1 Updating Formulas                                     | 356 |
|        | 11.6.2 Connection with the Sweep Operator                    | 357 |
|        | 11.6.3 Adding and Deleting Cases and Variables Using QR      | 360 |
| 11.7   | Centering the Data   | 363 |
| 11.8   | Comparing Methods  | 365 |

|                            |   |     |
|----------------------------|---|-----|
| 11.8.1                     | Resources   | 365 |
| 11.8.2                     | Efficiency  | 366 |
| 11.8.3                     | Accuracy  | 369 |
| 11.8.4                     | Two Examples  | 372 |
| 11.8.5                     | Summary   | 373 |
| Exercises 11c              |   | 374 |
| 11.9                       | Rank-Deficient Case                                 | 376 |
| 11.9.1                     | Modifying the QR Decomposition                      | 376 |
| 11.9.2                     | Solving the Least Squares Problem                   | 378 |
| 11.9.3                     | Calculating Rank in the Presence of Round-off Error | 378 |
| 11.9.4                     | Using the Singular Value Decomposition              | 379 |
| 11.10                      | Computing the Hat Matrix Diagonals                  | 379 |
| 11.10.1                    | Using the Cholesky Factorization                    | 380 |
| 11.10.2                    | Using the Thin QR Decomposition                     | 380 |
| 11.11                      | Calculating Test Statistics                         | 380 |
| 11.12                      | Robust Regression Calculations                      | 382 |
| 11.12.1                    | Algorithms for $L_1$ Regression                     | 382 |
| 11.12.2                    | Algorithms for M- and GM-Estimation                 | 384 |
| 11.12.3                    | Elemental Regressions                               | 385 |
| 11.12.4                    | Algorithms for High-Breakdown Methods               | 385 |
| Exercises 11d              |   | 388 |
| Miscellaneous Exercises 11 |   | 389 |
| <br>                       |   |     |
| 12                         | <del>Prediction and Model Selection</del>           | 391 |
| 12.1                       | Introduction  | 391 |
| 12.2                       | Why Select?   | 393 |
| Exercises 12a              |   | 399 |
| 12.3                       | Choosing the Best Subset                            | 399 |
| 12.3.1                     | Goodness-of-Fit Criteria                            | 400 |
| 12.3.2                     | Criteria Based on Prediction Error                  | 401 |
| 12.3.3                     | Estimating Distributional Discrepancies             | 407 |
| 12.3.4                     | Approximating Posterior Probabilities               | 410 |
| Exercises 12b              |   | 413 |
| 12.4                       | Stepwise Methods                                    | 413 |
| 12.4.1                     | Forward Selection                                   | 414 |
| 12.4.2                     | Backward Elimination                                | 416 |
| 12.4.3                     | Stepwise Regression                                 | 418 |
| Exercises 12c              |   | 420 |

|        |  |     |
|--------|--|-----|
| 12.5   | Shrinkage Methods                                    | 420 |
| 12.5.1 | Stein Shrinkage                                      | 420 |
| 12.5.2 | Ridge Regression                                     | 423 |
| 12.5.3 | Garrote and Lasso Estimates                          | 425 |
|        | Exercises 12d  | 427 |
| 12.6   | Bayesian Methods                                     | 428 |
| 12.6.1 | Predictive Densities                                 | 428 |
| 12.6.2 | Bayesian Prediction                                  | 431 |
| 12.6.3 | Bayesian Model Averaging                             | 433 |
|        | Exercises 12e  | 433 |
| 12.7   | Effect of Model Selection on Inference               | 434 |
| 12.7.1 | Conditional and Unconditional Distributions          | 434 |
| 12.7.2 | Bias   | 436 |
| 12.7.3 | Conditional Means and Variances                      | 437 |
| 12.7.4 | Estimating Coefficients Using Conditional Likelihood | 437 |
| 12.7.5 | Other Effects of Model Selection                     | 438 |
|        | Exercises 12f  | 438 |
| 12.8   | Computational Considerations                         | 439 |
| 12.8.1 | Methods for All Possible Subsets                     | 439 |
| 12.8.2 | Generating the Best Regressions                      | 442 |
| 12.8.3 | All Possible Regressions Using QR Decompositions     | 446 |
|        | Exercises 12g  | 447 |
| 12.9   | Comparison of Methods                                | 447 |
| 12.9.1 | Identifying the Correct Subset                       | 447 |
| 12.9.2 | Using Prediction Error as a Criterion                | 448 |
|        | Exercises 12h  | 456 |
|        | Miscellaneous Exercises 12                           | 456 |
|        | <b>Appendix A Some Matrix Algebra</b>                | 457 |
| A.1    | Trace and Eigenvalues                                | 457 |
| A.2    | Rank   | 458 |
| A.3    | Positive-Semidefinite Matrices                       | 460 |
| A.4    | Positive-Definite Matrices                           | 461 |
| A.5    | Permutation Matrices                                 | 464 |
| A.6    | Idempotent Matrices                                  | 464 |
| A.7    | Eigenvalue Applications                              | 465 |
| A.8    | Vector Differentiation                               | 466 |
| A.9    | Patterned Matrices                                   | 466 |



|   |     |
|---|-----|
| A.10 Generalized Inverse  | 469 |
| A.11 Some Useful Results  | 471 |
| A.12 Singular Value Decomposition   | 471 |
| A.13 Some Miscellaneous Statistical Results                                 | 472 |
| A.14 Fisher Scoring   | 473 |
| Appendix B Orthogonal Projections   | 475 |
| B.1 Orthogonal Decomposition of Vectors                                     | 475 |
| B.2 Orthogonal Complements  | 477 |
| B.3 Projections on Subspaces  | 477 |
| Appendix C Tables   | 479 |
| C.1 Percentage Points of the Bonferroni $t$ -Statistic                      | 480 |
| C.2 Distribution of the Largest Absolute Value of $k$ Student $t$ Variables | 482 |
| C.3 Working–Hotelling Confidence Bands for Finite Intervals                 | 489 |
| Outline Solutions to Selected Exercises                                     | 491 |
| References  | 531 |
| Index   | 549 |

## Why statistical models?

It is in human nature to try and understand the physical and natural phenomena that occur around us. **When observations on a phenomenon can be quantified, such an attempt at understanding often takes the form of building a **mathematical model****, even if it is only a simplistic attempt to capture the essentials. Either because of our ignorance or in order to keep it simple, many relevant factors may be left out. Also models need to be validated through measurement, and such measurements often come with error. In order to account for the measurement or observational errors as well as the factors that may have been left out, one needs a **statistical model which incorporates some amount of uncertainty**.

## Why a linear model?

Some of the reasons why we undertake a detailed study of the linear model are as follows.

- (a) Because of its simplicity, the linear model is better understood and easier to interpret than most of the other competing models, and the methods of analysis and inference are better developed. Therefore, if there is no particular reason to presuppose another model, the linear model may be used at least as a first step.
- (b) The linear model formulation is useful even for certain nonlinear models which can be reduced to the linear form by means of a transformation.
- (c) Results obtained for the linear model serve as a stepping stone for the analysis of a much wider class of related models such as mixed effects model, state-space and other time series models.
- (d) Suppose that the response is modelled as a nonlinear function of the explanatory variables plus error. In many practical situations only a part of the domain of this function is of interest. For example, in a manufacturing process, one is interested in a narrow region centered around the operating point. If the above function is reasonably smooth in this region, a linear model serves as a good first order approximation to what is globally a nonlinear model.

## Many statistical concepts can be viewed in the framework of linear models

suppose that we wish to compare the means of two populations, say,  $\mu_i = E[U_i]$  ( $i = 1, 2$ ). Then we can combine the data into the single model

$$\begin{aligned} E(Y) &= \mu_1 + (\mu_2 - \mu_1)x \\ &= \beta_0 + \beta_1x, \end{aligned}$$



where  $x = 0$  when  $Y$  is a  $U_1$  observation and  $x = 1$  when  $Y$  is a  $U_2$  observation. Here  $\mu_1 = \beta_0$  and  $\mu_2 = \beta_0 + \beta_1$ , the difference being  $\beta_1$ . We can extend this idea to the case of comparing  $m$  means using  $m - 1$  dummy variables.

In a similar fashion we can combine two straight lines,

$$U_j = \alpha_j + \gamma_j x_1 \quad (j = 1, 2),$$

and 1 otherwise. The combined model is

$$\begin{aligned} E(Y) &= \alpha_1 + \gamma_1 x_1 + (\alpha_2 - \alpha_1) x_2 + (\gamma_2 - \gamma_1) x_1 x_2 \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \end{aligned}$$

