Essays in Panel Data Econometrics

MARC NERLOVE
University of Maryland, College Park

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times Ten 9.75/12 pt. *System* LaTeX 2ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data available

ISBN 0 521 81534 7 hardback

Contents

Preface to the Book *page* xiii

1 The History of Panel Data Econometrics, 1861–1997 1
Preface 1
1. Introduction 4
2. In the Beginning: Astronomy, Agronomy, and Statistics 7
a. Fixed- versus Random-Effects Models 7
b. An Example in Which Mostly Fixed Effects Are Appropriate 8
c. An Example in Which Mostly Random Effects Are Appropriate; Airy's Problem 9
d. Fisher 12
e. Daniels and Eisenhart 13
f. Further Development of the Statistical Literature 14
3. Econometric Development, the Early Years: From Chicago to the Paris Conference of 1977 17
a. Hildreth 17
b. Management Bias and the Estimation of Production Functions: Hoch and Mundlak 18
c. Balestra and Nerlove: Error Components and Dynamic Panel Models 21
d. Highlights of the Ten Years Preceding the First Paris Conference, 1967–1976 26
e. The First Paris Conference, 1977 35
4. Econometric Development, the Later Years: From Paris to Paris 35
a. Problems of Testing Specification: Fixed versus Random Effects; Serially Correlated Effects 36
(1) Hausman Tests 36
(2) Serially Correlated Remainder Effects 39
(3) Serially Correlated Period Effects 41

b. Dynamics: Comparative Properties of Alternative Estimators and the Problem of Initial Conditions 46
(1) Inconsistency of the Pooled-Sample OLS Estimates of the Dynamic Error-Components Model 48
(2) Inconsistency of the OLS Estimators of the Dummy Variable, or Fixed-Effects, Model 49
(3) Generalized Least Squares and Feasible GLS 50
(4) Maximum Likelihood and the Problem of Initial Observations 52
c. Estimation of Limited Dependent, Discrete, and Qualitative Panel Data Models 58
(1) Incidental or Nuisance Parameters 59
(2) Logit and Probit Models 62
(a) Nondynamic Models 62
(b) Dynamic Models 65
(3) Limited Dependent Variables 67
(a) Nondynamic Models 67
(b) Dynamic Models 69
(4) Hazard and Duration Models 69
d. The Second Paris Conference, 1997 70
5. Beyond Paris 70
Appendices to Chapter 1
Appendix A: Decomposition of the Sum of Squares of a Variable Observed in a Two-Dimensional Panel 83
Appendix B: Alternative Formulations of the Concentrated Likelihood Function for the Two-Component Random-Effects Model 86
Appendix C: Introduction to The Econometrics of Panel Data*, Colloque International du CNRS, INSEE, Paris 22–24 August 1977* 92
Appendix D: Derivation of the Likelihood Function for the Three-Component Random-Effects Nondynamic Panel Model with Serially Correlated Period Effects 102
Appendix E: Likelihood for Stationary Panel Data-Generating Processes with Exogenous Variables and Individual-Specific Random Effects 105
Appendix F: Amemiya's Classification of Limited Dependent Variable or Tobit Models 111

2 Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas (with Pietro Balestra) 115
Preface 115
1. The Dynamic Model of Gas Demand 120
2. Estimation by Ordinary Least-Squares Methods 123

3. The Residual Model 127
4. Estimation Procedures 133
5. Conclusions 143
Appendix: Derivation of the Maximum-Likelihood Estimates of ρ and σ^2 144

3 Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections 149
Preface 149
1. Background 153
2. Design of the Experiments 161
a. Generation of the Observations 161
b. Parameter Values 164
c. Methods of Estimation 164
3. The Results 169
a. Generalized Least Squares 171
b. Ordinary Least Squares 171
c. Least Squares with Individual Constant Terms 172
d. Maximum Likelihood 172
e. Two Round 173
f. Comparison of Various Methods 173
4. Conclusions 174
Appendix: Parameter Values 177

4 Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections 189
Preface 189
1. Introduction 193
2. Design of the Experiments 200
a. Generation of the Observations 200
b. Parameter Values 205
c. Methods of Estimation 205
(1) Generalized Least-Squares Estimates Employing the True Value of ρ Used in the Generation of the Observations 205
(2) Ordinary Least-Squares Estimates of the Relation (2.1) *(i.e., Not Including Individual Constant Terms)* 207
(3) Least-Squares Estimates of the Parameters 207
(4) Instrumental-Variable Estimates Using x_{it-1} as an Instrument for y_{it-1} 207
(5) "Two-Round" Estimates Equivalent to Generalized Least Squares Using Not the True Value of ρ but the Estimate r_c Obtained in (3)*, a Least-Squares Regression with Individual Constant Terms* 208

(6) "Two-Round" Estimates Similar to (5) *but using r_i, an Estimate of ρ Based on the Calculated Residuals from the Instrumental-Variable Estimate of Equation* (2.1) *as Given in* (2.17) 208
(7) Maximum-Likelihood Estimates Obtained by Maximizing the Logarithmic Likelihood Function 208
3. Comparative Properties of Different Estimation Procedures 210
a. Bias 210
b. Relative Mean-Square Errors 213
c. Characteristics of the Maximum-Likelihood Estimates 214
4. Conclusions 216
Appendices for Chapter 4 (Available in pdf Format at http://www.arec.umd.edu/mnerlove/Tables1971.ZIP)

5 A Note on Error-Components Models 219
Preface 219
Errata 221
1. Introduction 222
2. The Characteristic Roots and Vectors of Ω 226
3. The Square Root of Ω 228
4. The Inverse of Ω *and the GLS Estimates of* β 231
5. An Interpretation of $\hat{\beta}$ *and Its Asymptotic Distribution* 233

6 Growth-Rate Convergence, Fact or Artifact? An Essay on Panel Data Econometrics 237
Preface 237
1. Introduction 240
2. The Neoclassical Theory of Growth Convergence 242
a. The Solow/Swan Model 243
b. Growth Convergence in the Neoclassical Model with Exogenous Population Growth and Savings 244
3. Empirical Investigations of Convergence and the Rate of Convergence 246
4. Alternative Methods for Estimating Rates of Convergence 249
a. Inconsistency of the Pooled-Sample OLS Estimates of the Dynamic Error-Components Model 250
b. Inconsistency of the OLS Estimators of the Dummy Variable or Fixed-Effects Model 251
c. Country Means Regression and the Estimation of ρ 252
d. Generalized Least Squares and Feasible GLS 253
e. Bounds for the Coefficient of the Lagged Dependent Variable 254
f. Maximum-Likelihood Conditional on the Initial Value of the Lagged Dependent Variable 255
g. Unconditional Maximum Likelihood 257

5. *Empirical Evidence on Growth-Rate Convergence and the Comparative Performance of Different Panel Data Methods* 260
   a. *Numerical Results* 260
   b. *Graphical Results* 265
6. *Conclusions* 268

Appendix 270

7 Properties of Alternative Estimators of Dynamic Panel Models: An Empirical Analysis of Cross-Country Data for the Study of Economic Growth 273
Preface 273
1. *Introduction* 274
2. *Recent Empirical Investigations of Convergence and the Rate of Convergence* 277
3. *Alternative Methods for Estimation* 281
   a. *Inconsistency of the Pooled-Sample OLS Estimates of the Dynamic Error-Components Model* 282
   b. *Inconsistency of the OLS Estimators of the Dummy Variable or Fixed-Effects Model* 283
   c. *Generalized Least Squares and Feasible GLS* 285
   d. *Bounds for the Coefficient of the Lagged Dependent Variable* 286
   e. *Maximum Likelihood Conditional on the Initial Value of the Lagged Dependent Variable* 287
   f. *Unconditional Likelihood and Unconditional Maximum Likelihood* 289
4. *Empirical Evidence on the Comparative Performance of Different Panel Data Methods* 293
5. *Conclusions* 299

8 Likelihood Inference for Dynamic Panel Models 305
Preface 305
1. *Introduction* 308
2. *The Likelihood Principle* 310
3. *Likelihood Functions for Two Basic Dynamic Panel Models* 314
   a. *The Model in Levels* 316
      (1) *Specification of the Joint Distribution of the Observations, Both Conditional and Unconditional on the Initial Observations* 316
      (2) *The Joint Likelihood Conditional on the Initial Observations and the Exogenous Variables* 321
      (3) *The Joint Likelihood Unconditional on the Initial Observations* 322
   b. *The Model in First Differences* 323

4. An Empirical Example: A Study of Cross-Country Economic Growth Using Panel Data 326
a. Recent Empirical Investigations 327
b. New Estimates Based on Maximizing the Conditional and Unconditional Likelihood Functions 331

Index 349

ONE

The History of Panel Data Econometrics, 1861–1997

Preface

In his famous and influential monograph, *The Probability Approach in Econometrics*, Haavelmo (1944) laid the foundations for the formulation of *stochastic* econometric models and an approach that has dominated our discipline to this day. He wrote:

> *. . . we shall find that two individuals, or the same individual in two different time periods, may be confronted with exactly the same set of specified influencing factors [and, hence, they have the same y^*, . . .], and still the two individuals may have different quantities y, neither of which may be equal to y. We may try to remove such discrepancies by introducing more "explaining" factors, x. But, usually, we shall soon exhaust the number of factors which could be considered as common to all individuals, and which, at the same time, were not merely of negligible influence upon y. The discrepancies $y - y^*$ for each individual may depend upon a great variety of factors, these factors may be different from one individual to another, and they may vary with time for each individual.* (Haavelmo, 1944, p. 50).

And further that:

> *. . . the class of populations we are dealing with does not consist of an infinity of different individuals, it consists of an infinity of possible decisions which might be taken with respect to the value of y.*
>
> *. . . we find justification for applying them [stochastic approximations] to economic phenomena also in the fact we usually deal only with – and are interested only in – total or average effects of many individual decisions, which are partly guided by common factors, partly by individual specific factors . . .* (Haavelmo, 1944, pp. 51 and 56).

Marschak (1950) and (1953) further amplified Haavelmo's themes in his introduction to Cowles Commission Monographs 10 and 14, observing that: The numerous causes that determine the error incurred . . . are not listed separately; instead their joint effect is represented by the probability distribution of the error, a random variable (1950, p. 18) [, which] . . . is called 'disturbance' or 'shock,' and can be regarded as the joint effect of numerous separately insignificant variables that we are unable or unwilling to specify but presume to be independent of observable exogenous variables. (1953, p. 12).

Since the early work of Mundlak (1961) and Balestra and Nerlove (1966), panel or longitudinal data have become increasingly important in econometrics, and methods for the analysis of such data have generated a vast literature, the history of which is selectively recounted in the essay that follows. A recurrent theme in this historical essay is the interpretation of what is not observed; that is, the disturbances in the relationships about which we wish to draw inferences and the proper interpretation of these disturbances. In the beginning, Sir George Biddell Airy's 1861 monograph on astronomical observations made essentially the same point.

The stochastic elements in the analysis, which are reflected in the unobserved variables characterizing individual heterogeneity and heterogeneity of individual decisions, lie at the heart of econometric analysis. Some conclusions that may be drawn from the historical overview presented in this essay are as follows:

(a) One of the main reasons for being interested in panel data is the unique possibility of uncovering disaggregate dynamic relationships using such data sets.
(b) In a dynamic context, one of the primary reasons for heterogeneity among individuals is the different history that each has.
(c) If the relevant "population" is, following Haavelmo, the space of possible decisions, different past histories take the form of individual specific random variables that are generally correlated with all of the variables taken as explanatory, not just the lagged values of the endogenous variable. The former, therefore, cannot be conditioned upon in the usual way.

History is important, not only because

> "Whereof what's past is prologue, what's to come, . . ."
> *The Tempest, II, i*, 247

but also because

> "Those who cannot remember the past are condemned to repeat it."
> George Santayana, *The Life of Reason, Vol. 1*, 1905.

References

Airy, G. B. (1861), *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations*, Cambridge: Macmillan.

Balestra, P., and M. Nerlove (1966), "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Economic Model: The Demand for Natural Gas," *Econometrica*, 34: pp. 585–612.

Haavelmo, T. (1944), "The Probability Approach in Econometrics," *Supplement to Econometrica*, 12.

Marschak, J. (1950), "Statistical Inference in Economics: An Introduction," in *Statistical Inference in Dynamic Economic Models*, T. C. Koopmans (Ed.), New York: John Wiley, pp. 1–50.

Marschak, J. (1953), "Economic Measurements for Policy and Prediction," pp. 1–26 in *Studies in Econometric Method*, W. C. Hood and T. P. Koopmans, Eds., New York: John Wiley.

Mundlak, Y. (1961), "Empirical Production Functions Free of Management Bias," *Journal of Farm Economics, 43*: pp. 44–56.

The History of Panel Data Econometrics, 1861–1997[1]

Whereof what's past is prologue, what's to come, . . .
The Tempest, II, i, 247

1. INTRODUCTION

I was asked a while ago to discuss the future of panel data econometrics. Being rather bad at forecasting, I took my cue from Prospero's line in *The Tempest*, quoted above, and reviewed instead the history of panel data econometrics from Hildreth (1950) down to Mátyás and Sevestre's monumental handbook (1996) and the Paris Conference of 1997. This essay is the fruit of that endeavor. Alain Trognon (2000) discusses much more fully more recent developments emphasizing more than me the internal methodological momentum of the subject. Our conclusions regarding the future, such as they are, are not greatly at variance. The future of panel data econometrics I hope for is much like its past, viewed in a long-term perspective. Our concern has been, and I hope will continue to be, with the best way to formulate statistical models for inference

[1] An earlier version of this essay was presented at the Ninth International Conference on Panel Data, June 22–23, 2000, Geneva, Switzerland, at the session on "The Future of Panel Data Econometrics." The research on which it is based was supported by the Maryland Agricultural Experiment Station.

The essay is dedicated to the memory of Zvi Griliches (1930–1999) and G. S. Maddala (1933–1999), who both contributed greatly to the development of panel data econometrics. It also marks the fiftieth anniversary of the first paper ever in panel data econometrics (Hildreth 1950), regrettably unpublished to this day.

A portion of the present survey is freely adapted from Nerlove (Chapter 19, "Analysis of Panel Data," 1999a). I am indebted to John Aldrich, Ramon Lopez, and Patrick Sevestre for helpful comments on earlier partial drafts. Anke Meyer read and commented on several earlier drafts, as well as the present one. Appendix D is based on an unpublished paper by Karlsson and Skoglund and on the work of Pietro Balestra contained in a personal communication to me.

motivated and shaped by substantive problems and our understanding of the processes generating the data at hand to resolve them. If the present trend toward increasing internalization pointed out by Trognon continues, however, these hopes may be unrealized. The principal factors in the research environment affecting the future course of panel data econometrics, in contrast to its past, are the phenomental growth in the computational power available to the individual researcher at his or her desktop and the ready availability of data sets, both large and small, via the Internet. Whether these factors will lead to increasing proliferation of "special" methods applied to a few "illustrative" data sets or to a broader, more comprehensive analytical framework and the analysis of a greater variety of data, I cannot say. I also hope that increased understanding of panel data econometrics will lead to more sophisticated design of surveys for the collection of panel data, and thus to a greater variety of data appropriate for the analysis of important and relevant policy issues, although by no means a certain outcome given the present academic climate in which we work.

Observations on many individual economic units (firms, households, geographical areas, and the like) over a period of time are said to be a "panel data set." For example, in Balestra and Nerlove (1966), data on thirty-six U.S. states over a thirteen-year period were used in the analysis. Panel data may be contrasted with pure *cross-section data*, observations on individual units at a point in time, and with pure time-series data, observations, usually of an aggregate nature, over time without any "longitudinal" dimension. For some purposes, it may be useful to view a cross section as a panel with time dimension 1. Panel data are sometimes treated as "cross sections over time" or "pooled" cross-section time-series data, but this terminology leaves open the question of whether or not the cross sections over time refer to identical individuals.[2]

Panel data offer several important advantages over data sets with only a temporal or a longitudinal dimension. First, more observations are generally available than with conventional time-series data, although cross-section data sets are often very large.

Second, because panel data are not so highly aggregated as typical time-series and because, in the best of circumstances, we observe the same individual units through time, more complicated dynamic and behavioral hypotheses can be tested than those that can be tested using unidimensional data. In the next section, I argue that economic behavior is inherently dynamic so that most econometrically interesting relationships are explicitly or implicitly dynamic.[3]

[2] Verbeek (1996) gives a useful survey of methods used to treat repeated cross sections, not necessarily involving the same individuals, which is apropos in this context.

[3] The dynamic component is not always obvious. For example, any relationship that involves a stock or a flow based on a stock, which is, in turn, a result of the past decisions we are trying to explain or related to them, will constitute a dynamic element of greater or lesser importance to the behavior modeled. A particularly insidious example is family size and composition

Finally, the use of panel data may also provide a means for analyzing more fully the nature of the latent, or unobserved, disturbance terms in the econometric relationships. These disturbances are supposed to measure the effects of all sorts of left-out factors and, as such, may frequently be subject to the objection that some of them are correlated with the included explanatory variables. Not only do panel data frequently provide the opportunity for introducing many more explanatory variables and more complicated dynamics, but they also permit us to model more explicitly the latent disturbances themselves as components common to all individuals at a point in time and as time-persistent components. The problem of latent individual heterogeneity is the central problem in panel data econometrics.

Panel data need not be restricted to two dimensions, one of which is time: for example, many types of cross-sectional survey data are obtained through "cluster" sampling. Certain geographical units are first selected (e.g., villages), then individuals are sampled within each village. Thus, the village from which an individual observation comes may be thought of as one dimension of the data, just as in a traditional panel the time period associated with an observation on an individual is a dimension of the data. Thus, panel data methods are of special importance in research in developing countries, which may not have a long tradition of statistical data collection and for which, therefore, it is generally necessary to obtain original survey data to answer many significant and important questions. But it undoubtedly remains true that the most important use of panel data and methods is in the analysis of dynamic models of behavior over time.

In the following section, I discuss the early development of panel data statistical methods from their introduction by Airy in 1861 for the analysis of astronomical data, through the work on human heredity of Galton and Fisher, development of fixed-effects ANOVA by Fisher and his disciples, down to recent work on variance-components or random-effects models in the analysis of animal breeding experiments. In Section 3, I take up the thread in econometrics, focusing on the earliest work of Hildreth, Hoch, Mundlak, and Balestra and Nerlove, culminating in the First Paris Conference on Panel Data Econometrics of 1977. In Section 4, I continue the story more selectively dealing with the work on specification tests, dynamic models, and estimation of limited dependent panel data models during the period leading up to the twentieth anniversary conference held in Paris in 1997. Finally, I draw a somewhat pessimistic conclusion about the future of panel data econometrics based on my reading of its past history.

in cross-section studies of farm households, in which past fertility decisions and decisions to invest in the health and nutrition of children partly determine demographic variation across families at a given time.

2. IN THE BEGINNING: ASTRONOMY, AGRONOMY, AND STATISTICS

"Those who cannot remember the past are condemned to repeat it."
George Santayana, *The Life of Reason, Vol. 1*, 1905.

a. Fixed- versus Random-Effects Models

Both so-called fixed-effects models and random-effects models have a long history in statistics.[4] It is a theme that runs through the history of the subject that I emphasize in this essay.

The origins of least squares in the astronomical work of Gauss (1809) and Legendre (1805) are well known.[5] And the relation of least squares to the analysis of variance as developed by R. A. Fisher (1918, 1925) is widely appreciated. In the Gauss-Legendre formulation, the independent or explanatory variables are treated as fixed and the dependent variable as subject to error. The conventional interpretation of Fisher's formulation of the analysis of variance is as an extension of least-squares theory, but, as Eisenhart (1947) points out, this was not the only interpretation Fisher placed on his analysis. In (1925, Chapter 7, especially Section 40), Fisher interprets the *intraclass correlation* in analysis of variance terms and, in this discussion, implies a random-effects formulation. Eisenhart (1947, pp. 3–5) is the classic locus of the distinction. He writes:

> ... analysis of variance can be, and is, used to provide solutions to problems of two fundamentally different types. These two distinct classes of problems are:
>
> *Class I: Detection and Estimation of Fixed (Constant) Relations Among Means of Sub-Sets of the Universe of Objects Concerned.* This class includes all of the usual problems of estimating and testing to determine whether to infer the existence of, true differences among "treatment" means, among "variety" means, and, under certain conditions, among "place" means. Included in this class are all the problems of univariate and multivariate regression and of harmonic analysis. With respect to problems of estimation belonging to this class, analysis of variance is simply a form of the method of least squares....
>
> *Class II: Detection and Estimation of Components of (Random) Variation Associated with a Composite Population.* This class includes all problems of estimating, and testing to determine whether to infer the existence of components of variance ascribable to random deviation of the characteristics of individuals of a particular generic type from the mean values of these characteristics in the 'population' of all individuals of that generic type, etc. In this sense, *this is the true analysis of variance*, and the estimation of the respective components of the overall variance of a single observation requires further steps beyond the evaluations of the entries of

[4] Accounts are given, *inter alios*, by Scheffé (1956); Anderson (1978); and Searle, Casella, and McCulloch (Chapter 2, 1992).

[5] In Nerlove (Chapter 1, *The Likelihood Principle*, 1999), I retell the story emphasizing the relation of least squares to likelihood methods for the optimal combination of observations.

the analysis-of-variance table itself. Problems of this class have received considerably less attention in the literature of the analysis of variance than have problems of Class I. . . .

. . . the mathematical models appropriate to problems of Class I differ from the mathematical models appropriate to problems of Class II and, consequently, so do the questions to be answered by the data.

The typical problem addressed by models of Class I is the analysis of experimental data such as occur in agronomic investigations, while the typical problem addressed by models of Class II is the analysis of nonexperimental, observational data such as are the norm in astronomical or economic investigations. Scheffé (1956) calls Class I "Model I" or the fixed-effects model, and Class II "Model II" or the "random-effects model." Often, if the effects are assumed to be independent of one another and random, Model II is the basis for an analysis of *variance components*. In the random-effects model, all effects are assumed to have zero mean, which can be enforced by assuming some fixed effects such as an overall mean. And, as usual in regression formulations, any mean may be regarded as a function of observed variables with unknown parameters. Moreover, even in a purely fixed-effects model, there is always at least one random effect called the error. In an experimental context, randomness of this error and independence from any fixed effects included is often enforced by randomizing aspects of the experiment reflecting uncontrolled variation. But in any model, some effects are always assumed random and others fixed. Scheffé (1956, pp. 254–255) writes: "We see that in formulating a model one must ask for each factor whether one is interested individually in the particular levels occurring in the experiment or primarily in a population from which the levels in an experiment can be regarded as a sample: the main effects are accordingly treated as fixed or as random. (It is conceivable that for two different purposes the same data might be analyzed according to two different models in which the same main effects are regarded as fixed or as random effects.) Interactions between several factors are naturally treated as fixed if all these factors have fixed effects and as random if one or more of these factors have random effects." It is clear that Scheffé is thinking primarily in terms of data generated by experimentation. To make the distinction between fixed and random effects clear, however, it is useful to consider two extreme examples in which the contrast between experimental and nonexperimental data is made clear.

b. An Example in Which Mostly Fixed Effects Are Appropriate

Suppose we are evaluating two varieties of high-yielding rice. We want to know how each variety responds to fertilizer application and to water availability, so we design an experiment in which each variety is planted several times over and is subjected to various determined and accurately measured levels of fertilizer and water application. At the end of the day, we measure the yield of each variety on each plot and for each combination of fertilizer and water application. If we have designed the experiment well, varieties are allocated to plots and treatments in a random manner. Clearly, there are a great many unobserved factors affecting the yields of each variety observed besides water availability and level of fertilizer application, most of which have to do with the particular plot. Suppose that we distinguish three levels of fertilizer application: low, medium, and high; and three levels of water application: low, medium, and high. The standard fixed-effects ANOVA model consists of an overall mean, a main effect for each of the factors: variety, fertilizer, and water, represented respectively by one, two, and two parameters; three bivariate interaction effects; and one trivariate interaction.[6] The treatment levels and varieties can be represented by dummy variables with appropriate restrictions, so that this ANOVA problem can be treated as a regression problem in which rice yield is the dependent variable and the observed independent variables are the dummies and various products thereof, the disturbance is assumed to be a random variable, independent of variety and treatment levels, which represents all the left-out variables associated with plot. This is the kind of problem Fisher (1925) considered in detail. The important thing to note is that variety and fertilizer and water treatment levels are fixed by the experimenter; there is no thought that they might have been selected from a larger, possibly unknown, population of varieties or levels. On the other hand, the plot effects can be considered random draws from an unknown population of unobserved plot-specific factors. In an experimental context, these effects are "controlled" by randomization.[7]

c. An Example in Which Mostly Random Effects Are Appropriate; Airy's Problem

For my next example, I turn to a quintessentially nonexperimental science, astronomy (at least it used to be so!). It is perhaps no accident that much of the early work of Gauss, Legendre, Laplace, and others of those who founded statistics was done in an astronomical context. As remarked, fixed-effect

[6] If there are Q variables, there are, in general, $\binom{Q}{k}$, $k = 1, \ldots, Q$ main and interaction effects. If all of them are present, the model is called saturated. If each variable is categorical, as is the case in the example, it does not require the number of parameters equal to the product of the number of categories for each variable included in an interaction to represent that effect, but a considerably lesser number since the ANOVA restrictions imply that the unconstrained parameter values sum to zero over any index. In the case discussed, for example, there are only two parameters required for each main effect, but four for each bivariate interaction, and eight for the single trivariate interaction. See Nerlove and Press (1978, 1986).

[7] This was not always so. Fisher's battles with the experimental establishment to introduce randomization into experimental design, which had been heretofore systematic, are described in detail in Box (1978, pp. 140–166).

models have their origin in the work on least squares of Gauss and Legendre, who were concerned with the optimal combination of astronomical observations, but the random-effects or variance-components models also originated in the attempts of nineteenth-century astronomers to make sense of their observations. In a monograph published in 1861, George Biddell Airy makes explicit use of a variance-component model for the analysis of astronomical panel data.[8] Here is how (1861, p. 92) Airy puts the problem (note that what Airy calls a *Constant Error*, we would call a random day effect):

When successive series of observations are made, day after day, of the same measurable quantity, which is either invariable ... or admits of being reduced by calculation to an invariable quantity ...; and when every known instrumental correction has been applied (as for zero, for effect of temperature upon the scale, etc.); still it will sometimes be found that the result obtained on one day differs from the result obtained on another day by a larger quantity than could have been anticipated, the idea then presents itself, that possibly there has been on some one day or on every day, some cause, special to the day, which has produced a *Constant Error* in the measures of that day. It is our business now to consider the evidence for, and the treatment of, such constant error.

Continuing (pp. 93–94), Airy writes:

First, it ought, in general, to be established that there is possibility of error, constant on one day but varying from day to day .... suppose ... that we have measured the apparent diameter of Jupiter. It is evident that both atmospheric and personal circumstances may sensibly alter the measure; and here we may admit the possibility of the error.... Now let us take the observations of each day separately, and ... investigate from each separate day the probable error of a single measure. We may expect to find different values (the mere paucity of observations will sufficiently explain the difference); but as the different observations on the different days either are equally good, or (as well as we can judge) have such a difference in merit that we can approximately assign the proportion of their probable errors, we can define the value of error for observations of the standard quality as determined from the observations of each day; and combining these with greater weight for the deductions from the more numerous observations, we shall have the final value of the probable error of each observation not containing the effects of the Constant Error.

[8] Reference to and a brief discussion of Airy's work (1861) are found in Scheffé (1956), who credits Churchill Eisenhart for the reference. George Biddell Airy was born July 27, 1801, at Alnwick, Northumberland, England, and died six months short of his ninety-first birthday at Greenwich, England, on January 2, 1892. He went down to Cambridge to study mathematics, becoming successively Senior Wrangler in 1823, Fellow of Trinity College in 1824, and Lucasian Professor of Mathematics in 1826, a professorship once held by Isaac Newton. He was appointed Astronomer Royal of England and Director of the Royal Observatory at Greenwich in 1835, a post he held until 1881. He was knighted by Queen Victoria in 1872. Although the Greenwich meridian, Longitude 0°, had been used by seafarers since 1767 as a reference point for time and longitude (Sobel, 1995, p. 166), it was Airy's precise measurement of the location of the meridian by means of an instrument he invented that made it the universal standard.

Airy goes on, on subsequent pages, to develop verbally the following model: Let us observe the phenomenon, say the apparent diameter of Jupiter, on I nights, with J_i observations being made the ith night. Let the measurement be y_{ij}; then

$$y_{ij} = \mu + \delta_i + \varepsilon_{ij}, \quad j = 1, \dots, J_i, \quad i = 1, \dots, I, \tag{1}$$

where μ is the "true" value, and $\{\delta_i\}$ and $\{\varepsilon_{ij}\}$ are random effects with the following interpretation: δ_i is what Airy calls the Constant Error associated with day i, what we would call the "day effect"; that is, the atmospheric and personal circumstances peculiar to the ith night, and ε_{ij} is all the rest, or the errors about the conditional mean, $\mu + \delta_i$, on the ith night. He assumes that the ε_{ij} and δ_i are each independently and identically distributed and independent of each other and have zero means. Let the variances of δ and of ε be σ_δ^2 and σ_ε^2, respectively, and suppose, for simplicity, J equals numbers of observations each night (a balanced panel). To make his point, Airy wants to reject the hypothesis that $\sigma_\delta^2 = 0$. He computes an estimate of the "within" variance for each night i as

$$\hat{\sigma}_{\varepsilon,i}^2 = \frac{1}{J-1}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{i.})^2,$$

and then takes the arithmetic mean of the square roots to estimate the root of σ_ε^2:

$$\hat{\sigma}_\varepsilon^2 = \left(\frac{1}{I}\sum_{i=1}^{I}\sqrt{\hat{\sigma}_{\varepsilon,i}^2}\right)^2.$$

To estimate σ_δ^2 Airy uses not the between-nights sum of squares, but rather the corresponding mean absolute deviation:

$$d = \frac{1}{I}\sum_{i=1}^{I}|\bar{y}_{i.} - \bar{y}_{..}|.$$

He then calculates an approximate probable error for d from a standardized normal by replacing σ_ε^2 by $\hat{\sigma}_\varepsilon^2$ and μ by $\bar{y}_{..}$. The calculated value of d being larger than this value, Airy rejects the hypothesis of no night effect. If the details of Airy's analysis seem a bit clumsy from a modern point of view, the spirit of his model and calculations are surprisingly up-to-date.

Only a few years later, William Chauvenet (1863) published the first edition of his two-volume text in spherical astronomy, which became the standard reference work until the end of the century.[9] His calculations of the probable

[9] William Chauvenet (1820–1870) was professor of mathematics at the U.S. Naval Academy in Annapolis from its founding in 1845 until his departure for Washington University in St. Louis in 1859, where he ultimately became Chancellor of the University. His book, *A Manual of*

error of transit observations (1863; fifth edition, 1889, pp. 194–200) uses the estimate

$$\mathrm{Var}(\bar{y}..) = \frac{1}{I}\left(\hat{\sigma}_\delta^2 + \frac{1}{J}\hat{\sigma}_\varepsilon^2\right).$$

Clearly the random-effects model for the analysis of panel data was well established long before Fisher wrote about the *intraclass* correlation in 1925. Indeed, Francis Galton (1889) introduced the concept, although under another name, and used a variance-components model in his work on human inheritance and his anthropometric investigations. See Stigler (1999, p. 182).[10]

d. Fisher

The terms *variance* and *Analysis of Variance* were both introduced by R. A. Fisher in his famous and seminal papers on quantitative genetics (1918a) and (1918b).[11] The concepts and methods of both fixed-effects and random-effects models were elaborated greatly in Fisher (1925), especially in Chapters 7 and 8, “Intraclass Correlations and the Analysis of Variance” and “Further Applications of the Analysis of Variance.” But Fisher was never clear on the distinction between the fixed-effects model and the random-effects model. In Sec. 40, Chapter 7 (page references are to the 1970 reprint of the 14th edition), Fisher (1925, reprinted 1970, p. 234) writes the usual ANOVA table for assessing the significance of the variation of the heights of brothers from the same family across families (i.e., the table appropriate for the question can the family “effect” account for a significant part of the total variation in heights). He then goes on to interpret the problem in terms of the proportion of variance attributable to the “family effect,” with a clear “random-effect” flavor (pp. 225–226):

> Let a quantity be made up of two parts, each normally and independently distributed; let the variance of the first part be A, and that of the second part B; then it is easy to see that the variance of the total quantity is $A + B$. Consider a sample of n' values of the first part, and to each of these add a sample of k values of the second part, taking a fresh sample of k in each case. We then have n' families of values with k in each family. *In the infinite population from which these are drawn* [italics supplied] the correlation between pairs of members of the same family will be
>
> $$\rho = \frac{A}{A+B}.$$

Theoretical and Practical Astronomy, went through many editions, the fifth and last, to which I have had access, being published in 1889.

[10] Hald (1988, p. 675) mentions two additional precursors of Fisher: Edgeworth (1885) and Thiele (1903).

[11] See also Moran and Smith (1966). Fisher (1918b) was the paper submitted first to *Biometrika* that Pearson rejected as editor. Relations between the two men were never the same after that!