Behavior Genetics, Vol. 32, No. 2, March 2002 (© 2002)

A Limited Dependent Variable Model for Heritability Estimation with Non-Random Ascertained Samples

Timo M. Bechger,[1,4] Dorret I. Boomsma,[2] and Henk Koning[3]

Received 19 Nov. 2001—Final 28 Jan. 2002

In a questionnaire study, a random sample of Dutch families was asked whether they suffered from asthma and related symptoms. From these families, a selected sample was invited to come to the hospital for further phenotyping. Families were selected if at least one family member reported a history of asthma and the twins were 18 years of age or older. Not all families that were thus selected volunteered, leaving us with a fraction of the original sample.

The aim of this paper is to describe a limited dependent variable model that can be used in such situations in order to obtain estimates that are representative of the population from which the sample was originally drawn. The model is a linear (DeFries-Fulker) regression model corrected for sample selection. This correction is possible when (some of) the characteristics that determine whether subjects volunteer (or not) are known for all subjects, including those that did not volunteer.

The questionnaire study is of interest by itself but serves mainly to provide a concrete illustration of our method. The present model is used to analyze the data and the results are compared to those obtained with other methods: raw (or direct) likelihood estimation, multiple imputation, and sample weighting. Throughout, Rubin's general theory of inference with missing data serves as an integrating framework.

KEY WORDS: Sample selection; missing data; volunteering bias; DeFries-Fulker regression; Heckman model; limited dependent variables model; Immunoglobulin E; asthma.

INTRODUCTION

In genetic epidemiological studies, subjects often participate in ways that are beyond the control of the researcher (e.g., they volunteer to participate in the study). This can lead to a sample that is unrepresentative of the population from which it is drawn and unsuited to assess the importance of genetic and shared environmental influences in the population of interest (e.g., Neale *et al.*, 1989; Neale and Eaves, 1993). The aim of this paper is to describe how limited dependent variable models can be used in such situations in order to obtain estimates that are representative of the population from which the sample was originally drawn.

In a questionnaire study (Koopmans *et al.*, 1995; Koopmans and Boomsma, 1996) a sample of Dutch families, consisting of twins (adolescent and young adults) and their parents, reported whether they had ever (versus never) suffered from asthma. When at least one family member reported having asthma, the family was invited to the hospital for further phenotyping. These phenotypes were related to allergy and asthma. About 50% of the families that were approached volunteered, which means that data were obtained for a small percentage (about 4%) of the original sample only. With respect to its design, this study is representative of a large class of studies in genetic epidemiology and it is used as a context to introduce the statistical method that is the focus of this paper.

[1] National Institute for Educational Measurement (Cito).
[2] Department of Experimental Psychology, Vrije Universiteit, Amsterdam, The Netherlands.
[3] Unaffiliated.
[4] CITO, (OPD), PO Box 1034, 6801 MG, Arnhem, The Netherlands. Tel (voice): +31 (0) 26-3521162. Tel (fax): +31 (0) 26-3521356. e-mail: timo.bechger@citogroep.nl

The general missing data theory developed by Rubin (Rubin, 1976; Little and Rubin, 1987) is used throughout this paper as an integrating framework. This theory is briefly introduced in the second section. In the third section, we describe the data, focusing on the way subjects were selected. The fourth section describes a model that is known as the *Heckman model.* We discuss how this model can be used in combination with DeFries-Fulker regression (DeFries and Fulker, 1985) to estimate heritability with selected samples. We briefly describe a relatively simple estimation procedure, but our focus is on the interpretation of the model. In the fifth section, we estimate the heritability of two variables: Immunoglobulin E (IgE) blood levels, and a variable intended as a measure of general health. Our aim is to discuss methodological issues, and the analyses serve as an illustration. In the sixth section, we summarize and discuss our findings.

SAMPLE SELECTION AS A CASE OF MISSING DATA

Sample selection may be viewed as an example of missing data in that the variables of interest are measured on the selected twins only. It is therefore useful to discuss sample selection within the framework of a general theory of missing data, which was developed by Rubin. For an introductory account of this theory and a survey of additional methods to deal with missing data, we refer the reader to Heitjan and Basu (1996) or Schafer (1997).

In the present study, we distinguish between two sets of variables: *the variables of interest,* which are only observed in selected families, and a set of *"additional" variables* that are observed in all (selected and unselected) families (see Table I). The general theory of missing data implies that under the assumption of ignorability of the selection mechanism, correct maximum likelihood (ML) estimation can be provided by using the observed data of selected and unselected twins jointly in the estimation. The selection mechanism is *ignorable* when the probability of selection is independent of the variables of interest. This is obtained in two situations. The first situation is when the selected families are a random sample from the target population. In this situation, the families are said to be *selected completely at random* (SCAR). The second situation is when the probability of selection is dependent on the values of the "additional" variables that are related to the variables of interest. In this situation, we say that the families are *selected at random* (SAR). For example, when families were deliberately selected on some of the additional variables, they are SAR. The selection is not ignorable when participation in the study depends in some way on the values of the variables of interest, which are unknown for the twins that were not selected.

Due to volunteering it is unclear whether families are SAR in the asthma study. The assumption of ignorable selection is not clearly defendable and it is impossible to avoid dealing with non-ignorable selection. In the next section we use probit regression to investigate which of the "additional" variables are involved in volunteering. In subsequent sections we discuss how the analysis of the data may proceed, using the results of the probit regression. First, we discuss the design of the study and the data in more detail.

THE DATA

Sample Selection in the First Data Gathering Stage

The subjects were twins, aged between 12 and 24, who participated in the Dutch twin-family study of health-related behavior in the period 1991 to 1993 (Koopmans *et al.*, 1995; Koopmans and Boomsma, 1996). An invitation to participate was sent to a total of 5987 families. The addresses were obtained from the *Netherlands Twin Register (NTR).* The families that are registered with the NTR constitute a random sample of Dutch families with twins, based on city counsel registrations.[5] The families that volunteered received a set of questionnaires, which was returned by 2711 families. It is important that these 2711 families constitute a representative sample, otherwise it is not possible to obtain valid estimates from the subsample subsequently selected from this sample of 2711 families.

A common manifestation of volunteering bias in twin studies is that about two-thirds of the twin pairs are female and about two-thirds are monozygotic, although there are equal numbers of female and male monozygotic (MZ) and dizygotic (DZ) twins in the European populations (Lykken *et al.*, 1987). In the present study, this type of bias was not observed and all five zygosity-sex groups are well represented. In particular, the proportion of opposite-sex twins in our sample (28%) was about equal to the proportion in the total population of twins born between 1970 and 1980 in the Netherlands (Tas, 1990). Furthermore, the sample was representative of the general population with respect to

[5] Information can be obtained from the following internet address: http://www.psy.vu.nl/ntr/

educational level of the twins, religious background, the number of twins that report having smoked, used alcohol, and participated in sports (Koopmans *et al.*, 1995; Koopmans, 1997; Boomsma *et al.*, in press). Hence, in spite of the high rate of non-response we believe that, at this stage, volunteers were SCAR so that the initial sample of 2711 families is fairly representative of the Dutch general population (Little and Schenker, 1995, p. 43).

Deliberate Selection and Volunteering in the Second Data Gathering Stage

Some measurements that assess asthma clinically are expensive, and only those families were tested in which at least one member reported a history of asthma. In addition, the study was limited to twins aged 18 years or over. There were 514 families, or about 19% of the sample, that satisfied these criteria. Of these families, 176 families were not invited because at least one family member had not given a useful answer, because twins already participated in another study, or because the family had not returned an additional questionnaire. This is unfortunate since it means that selection was partly based on known criteria (reported asthma and age) and partly based on unknown variables. Of the (514 − 176 =) 338 families that were invited, 102 volunteered to participate. Hence, in total (514 − 102 =) 412 families were selected, but not tested, which is about 80% of those eligible for selection. As we do not know why some families volunteered while others refused to participate, we should not *a priori* regard the volunteers as SAR, given the twin's age and the number of family members with asthma.

To determine whether there are additional variables that relate to the probability of selection we conducted a probit regression analysis. We started with a model including the deliberate selection variables: the twins' age and the number of family members with asthma. Both variables were found to have a significant effect on the probability of selection. Then, 10 additional variables were included in the regression analysis in a stepwise fashion. These additional variables included personality traits, general health, age, and variables related to socio-economic status (see Table I). Of these additional variables, only two were found to have an independent, significant effect on the probability of being in the sample: the twins' present level of education had a positive effect, and the mean social conformity score of the family had a negative effect. Social conformity is measured with the social conformity scale of the Amsterdam Biographical Questionnaire (Wilde, 1970). Goldberger's R^2 (Arminger, 1995, p. 149) equals 0.11 after correction for sampling fluctuations, which indicates that there is considerable variation in the probability in the sample that is left unexplained.

Table I. Ten Additional Variables That Might Be Relevant to Participation in the Study

1. Trait anxiety averaged over the family
2. Neurosis averaged over the family
3. Experience seeking averaged over the family
4. Neurotic-somatic complaints averaged over the family
5. Social conformity averaged over the family
6. Mean age of the parents
7. Mean education of the parents
8. Mean education of the twins
9. Number of family members that belong to the reformed church[a]
10. General health averaged over the family

[a] The Free University was originally founded by the reformed church.

DEFRIES-FULKER REGRESSION AND LIMITED DEPENDENT VARIABLE MODELS

Introduction

The DeFries-Fulker regression method (DF-regression) regresses the scores of one twin on the scores of the other twin. The first twin may be a *proband* and it is assumed that his or her ascertainment does not depend on the ascertainment of the other twin, called the *cotwin.* In cases where there is no clear distinction between probands and cotwins, a doubly-entry procedure of the data is sometimes used. The DF regression equation is:

$$Y = \beta_0 + \beta_1 R + \beta_2 P + \beta_3 PR + \varepsilon_i, \qquad (1)$$

where P denotes the score of the proband, and Y denotes the cotwin score. The character R is the coefficient of relationship for the pair (i.e., 1/2 for DZ twins and 1 for MZ twins), β_0 is the regression constant, and ε_i, is a disturbance term. The variables Y and P are assumed to be distributed normally with the same variance. DeFries and Fulker (1985) demonstrated that β_3 is an estimate of narrow sense heritability (h^2), and that β_2 is an estimate of the proportion of variance due to shared environment, (c^2). The DF regression model may be written more compactly as:

$$y_i = \beta^t \mathbf{x}_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma_\varepsilon^2), \qquad (2)$$

where y_i and $\mathbf{x}_i = (1, r_i, p_i, pr_i)^t$ contain the observed values for the i-th twin pair, and $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^t$.

The residuals are assumed to be independent and normally distributed with mean 0 and variance σ_ε^2. Note that we use capital letters for the "name" of a random variable and lowercase letters for its realizations. Superscripted *t* denotes transposition.

By definition (or by assumption) $E(\varepsilon_i \mid \mathbf{X} = \mathbf{x}_i) = 0$ for every value of **X**. Therefore, if the data are *proband selected*, the regression model is not violated in the selected sample and we may obtain correct estimates of the parameters. This is not true when both twins are non-randomly selected, which implies that the OLS estimator of β is biased and inconsistent (Heckman, 1979, p. 155). The situation becomes more intricate when twins are selected as pairs. Methods to obtain correct estimates in this situation are discussed in the econometric literature (e.g., Greene, 2000) under the heading *limited dependent variable models*, usually abbreviated to LIMDEP models. LIMDEP models provide solutions to a wide variety of sample selection problems. On the basis of a review of the literature (Bechger, 1997) we decided to use the so-called Heckman model in the present study.

The Heckman Model

It is assumed that families participate in a study only when their value on a *selection variable* Z* exceeded a threshold, which is arbitrarily assumed to be zero. When families are not selected on the variables of interest, this kind of selection is called *indirect selection* (e.g., families are selected based on self-reported asthma and the variable of interest is the level of IgE). The selection variable is assumed to be a linear function of the "additional" variables, that is,

$$z^*_i = \gamma^t \mathbf{w}_i + u_i, \tag{3}$$

where γ denotes a vector of fixed regression coefficients, and $\mathbf{w}_i$ is the value of the i-th family on the additional variables that were found to affect the probability of selection. U designates a random error term representing unmeasured variables that influence the liability to volunteer. The presence of this random disturbance causes random variation in the selection variable among families with the same value of $\gamma^t \mathbf{w}_i$ so that the Heckman model explicitly allows for "*indirect, soft selection*" (Martin and Wilson, 1982; Neale *et al.*, 1989). The selection variable is related to the DF regression through the error terms. To be more specific, it is assumed that ε_i and u_i have a bivariate normal distribution with zero means and correlation ρ. A non-zero correlation indicates that, in the population, there are unmeasured influences on selection that affect the characters of interest. To be more specific, if ρ is positive, values of Y in the selected group are likely to be larger than those in the unselected group, and vice versa is ρ is negative.

Using the properties of the bivariate, truncated normal distribution (e.g., Maddala, 1983, pp. 266–267), Heckman (1979) shows that the following model applies to the twins in the selected sample:

$$E[y_i | y_i \text{ is observed}] = \beta^t \mathbf{x}_i + \rho\sigma_\varepsilon \lambda_i \tag{4}$$

where

$$\lambda_i = \frac{\phi(\gamma^t \mathbf{w}_i)}{\Phi(\gamma^t \mathbf{w}_i)} \tag{5}$$

$\phi(\gamma^t \mathbf{w}_i)$ and $\Phi(\gamma^t \mathbf{w}_i)$ are the density and distribution functions of the standard normal distribution, respectively, evaluated at $\gamma^t \mathbf{w}_i$. Compared to equation (**1**), the DF regression model is misspecified in the selected sample because a variable is omitted. Based upon this observation, Heckman (1979) proposes a simple estimation procedure. The first step is to use probit regression to obtain an estimate of γ, and compute an estimate of λ_i for each observation in the selected sample. In addition, we must compute

$$\hat{\delta}_i = \hat{\lambda}_i(\hat{\lambda}_i + \hat{\gamma}^t \mathbf{w}_i). \tag{6}$$

The second step is to estimate β and $\rho\sigma_\varepsilon \equiv \beta_\lambda$ by ordinary least squares regression. The third step, finally, is to calculate the asymptotic standard errors for β and β_λ:

$$\begin{aligned} \mathbf{V} = \hat{\sigma}_\varepsilon^2 (\mathbf{x}_*{}^t \mathbf{x}_*)^{-1} [\mathbf{x}_*{}^t (\mathbf{I} - \hat{\rho}^2 \Delta) \mathbf{x}_* \\ + \hat{\rho}^2 (\mathbf{x}_*{}^t \Delta \mathbf{w}) \Sigma (\mathbf{w}^t \Delta \mathbf{x}_*)] (\mathbf{x}_*{}^t \mathbf{x}_*)^{-1}, \end{aligned} \tag{7}$$

where $\mathbf{x}_*$ denotes the augmented matrix $[\mathbf{x}, \hat{\lambda}]$, where **x** is the data matrix (including the constant) for the selected subjects, and $\hat{\lambda}$ is a column vector with all the values $\hat{\lambda}_i$ as elements. $\Delta = (\hat{\delta}_i)$ is a diagonal matrix, **I** is an identity matrix of appropriate dimensionality, and Σ denotes the asymptotic covariance matrix of the parameters of the probit equation. Estimates of σ^2 and ρ are calculated as:

$$\sigma_\varepsilon^2 = \frac{\text{SSE}}{\text{nr. of sel. cases}} + \bar{\delta}\hat{\beta}_\lambda^2, \text{ and } \hat{\rho} = \frac{\hat{\beta}_\lambda}{\hat{\sigma}_\varepsilon}, \tag{8}$$

where SSE denotes the residual sum-of-squares from the linear regression, and $\bar{\delta}$ denotes the mean of $\hat{\delta}$ taken over the selected subjects. Further details appear in

Greene (1981; 2000). Software is available at *http://www.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm.*[6]

Heckman's estimation procedure is widely used. In simulation studies, it is found to perform well when compared to full ML estimation, unless there is a high degree of multicollinearity between λ_i and $\mathbf{x}_i$ as evidenced by an $R^2 > 0.80$ when $\gamma^t\mathbf{w}_i$ is regressed on $\mathbf{x}_i$ (Nawata, 1993; Nawata and Nagase, 1996).

Within the general theory on missing data, Heckman's model is classified as a model for *unknown, nonignorable missingness* (Little and Rubin, 1987). The missingness is called unknown because the selection variable Z* is only known as a stochastic function of other variables, known and unknown. Missingness cannot be ignored because the missing data mechanism depends on data that are missing. The missing data mechanism is the probability of being selected into the sample given **X** and Y. This distribution is Bernoulli with probability of response for the i-th twin pair

$$\Pr(z^*_i > 0 \,|y_i, x_i) = \Pr(u_i < \gamma^t\mathbf{w}_i|\varepsilon_i) = \Phi\left[\frac{1}{\sqrt{1-\rho^2}}[\gamma^t\mathbf{w}_i - \rho\sigma_\varepsilon^{-1}(y_i - \beta^t\mathbf{x}_i)]\right]. \quad (9)$$

According to the general missing data theory, the sample is SAR if this distribution is independent of the missing data. Equation (9) shows that this requirement obtains if $\rho = 0$. Hence the sample is SAR if $\rho = 0$, which implies that the selection mechanism is independent of the characters of interest.

APPLICATIONS

Immunoglobulin E Levels in Twins

To illustrate the practical application of the Heckman model we used the model to estimate the genetic and environmental influences on IgE blood levels measured on the selected sample of twins. IgE is produced as a reaction to inhaling allergens and is believed to be associated with allergy and asthma (e.g., Burrows *et al.,* 1989). Since the measurements are counts or quantities, we expect them to be distributed log-normally (Aitchison and Browne, 1969). We therefore analyzed the natural logarithm of the observed measurements, which are expected to be distributed normally before selection (Dudewicz and Mishra, 1989, example 7.2.17 and example 7.3.11). Heckman's procedure was used to obtain parameter estimates. As predictors of selection we included the variables that were demonstrated to affect on the probability of selection: number of family members with asthma, the twin's age, the twin's education, and the average social conformity of the family.

The results in Table II reveal that the estimated value of β_λ was not significantly different from zero, which indicates that the DF-regression was not biased due to selection. This is confirmed by the fact that the results with and without the correction for sample selection were virtually the same. The estimate of c^2 was found to be small and nonsignificant, and c was left out of the model. The heritability was found to be 0.78 (approximate standard error = 0.14).

Related studies by Russell *et al.* (1984) and Hanson *et al.* (1991) report heritabilities of 0.61 and 0.56, respectively, which are within 1.96 times the standard error from our estimate. Hanson *et al.* (1991) compared twins raised together with twins raised apart and found no evidence for an effect of shared environment, thus supporting the present findings.

General Health

Data on general health were gathered for all subjects by survey. This provides a good opportunity to compare the performance of Heckman's model to other procedures: structural equation modeling (SEM) using *raw-data ML* (Wilks, 1932; Lange *et al.*, 1976; Finkbeiner, 1979; Arbuckle, 1996), multiple imputation (Rubin, 1986), and sample weighting based on the inverse of the predicted probabilities from a probit (or logistic) regression (Heath, Madden, and Martin, 1998). These alternative procedures are all based on the assumption that the data are SAR, given the additional variables (Bechger, 1997). ML estimates are obtained with the ubiquitous Mx program (Neale, 1997), which is available at http://views.vcu.edu/mx. Multiple

Table II. The Heckman Model Applied to the Blood Levels of IgE. N = 53 (Cases with Missing Responses Were Listwise Deleted)

DF-regression without correction	DF-regression with correction
$h^2 = 0.78$ (s.e. 0.15)	0.78 (s.e. 0.14)
$\beta_\lambda =$	−0.36 (s.e. 0.37)
$\sigma_\varepsilon =$	1.29
$\rho =$	−0.27

[6] We have used our own software. Our present aim is to provide (just) enough detail for people to write their own program to do the analysis.

imputations are calculated with the WINNORM program kindly made available by Joe Schafer (Schafer, 1997) at http://www.stat.psu.edu/~jls/misoftwa.html#top.

The results in Table III indicate that all procedures give similar results, suggesting that the sample is indeed SAR given **W**. Simply ignoring the sample selection also gave good estimates, and we tentatively conclude that the sample was SCAR with respect to general health, which was also suggested by the probit regression analysis conducted earlier. Note that, although it is often seen that different ways to deal with missing data give similar results, this should not be taken for granted.

DISCUSSION

The general theory of missing data implies that selection can be ignored when the probability of selection is independent of the values that have not been observed. Under ignorability, ML estimation can be performed using the observed data only.

With volunteer samples, ignorable selection is not clearly defendable, so that it is not possible to avoid dealing with the problem of nonignorable selection. Our aim was to demonstrate that LIMDEP models could be used to estimate heritability in the face of such selection. We focused on a questionnaire study involving deliberate stratification as well as volunteering. With respect to its design, this study is representative of many studies in genetic epidemiology.

We have found that the DF-Heckman regression model provides an easy solution to deal with samples of the present type. It provides both a correct estimate of the heritability parameters and their asymptotic standard errors, as well as an assessment of the degree to which the selection mechanism was ignorable. When DF-regression is used to detect *Quantitative Trait Loci* (QTL), as proposed by Fulker *et al.* (1991), subjects are deliberately selected to increase power. The Heckman procedure provides a way to estimate the importance of the QTL in the unselected group, even when some selected subjects refuse to volunteer.

Finally, we have assumed that there was no additional nonresponse among the selected subjects. In fact, in the analysis of the IgE data selected twins with incomplete data were excluded. Future (simulation) studies are necessary to determine the appropriate course of action when we have to deal with nonresponse due to selection and item nonresponse at the same time. In closing, we mention two other topics for future research. First, the procedure described in this paper should be extended to the multivariate case. Second, we wish to know how to correct for sample selection with ordinal data. When the probability of participa-

Table III. Genetic and Environmental Parameters of General Health Estimated with the Complete Data (N = 2635 Twin Pairs) and with Various Methods for Dealing with Sample Selection (There were 82 Twin Pairs in the Selected Sample; the Numbers in Parentheses Represent Asymptotic Standard Errors)

A model with common environment and additive genetic influences							
	DF-regression		SEM				
	Correction						
Complete sample	No	Yes	Whole sample	Selected sample	Raw data likelihood	Sample weights	Multiple imputation
c^2 = 0.14 (0.05)	0.10	0.08	0.06	0.02	0.07	<0.001	0.03
h^2 = 0.27 (0.07)	0.36	0.38	0.39	0.47	0.40	0.43	0.39
β_λ =		−0.02 (0.14)					
A model with only additive genetic influences							
	DF-regression		SEM				
	Correction						
Complete sample	No	Yes	Whole sample	Selected sample	Raw data likelihood	Sample weights	Multiple imputation
h^2 = 0.45 (0.02)	0.49	0.48	0.47	0.49	0.48	0.43	0.42
β_λ =		−0.02 (0.15)					

tion of a twin pair is the average of the probabilities that each twin participates one may use a procedure suggested by Heath *et al.* (1998), which is based on weighting, but no other procedure is known to us. We suspect that readers more knowledgeable on recent developments in econometrics might immediately see ways to improve upon the present procedure.

REFERENCES

Aitchison, J., and Brown, J. (1969). *The Lognormal Distribution with Special Reference to Its Uses in Econometrics.* New York: Cambridge University Press.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In Schumacker, R., and Marcoulides, G. A. (eds.), *Advanced Structural Equation Modeling.* Mahuah, NJ: Lawrence-Earlbaum.

Arminger, G. H. (1995). Specification and estimation of mean structures: Regression models. In Arminger, G., Clogg, C., and Sobel, M. (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences.* New York: Plenum Press.

Bechger, T. M. (1997). A case in heritability estimation with complex non-random ascertained samples. Unpublished report, Faculty of Psychology, Free University Amsterdam.

Boomsma, D. I., Geus, E. J. C. de, Baal, G. C. M. van, and Koopmans, J. R. (1999). Religious upbringing reduces the influence of genetic factors on dishibition: Evidence for interaction between genotype and environment. *Twin Research*, in press.

Burrows, B., Hasan, F. M., Barbee, R. A., Halonene, M., and Lebowitz, M. D. (1989). Association of asthma with serum IgE levels and skin-test reactivity to allergens. *N. Engl. J. Med.* **320**: 271–277.

DeFries, J. C., and Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavior Genetics* **15**:467–473.

Dudewicz, E. J., and Mishra, S. N. (1988). *Modern Mathematical Statistics.* New York: Wiley.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika* **44**(4):409–420.

Fulker, D. W., Cardon, L. R., DeFries, J. C., Kimberling, W. F., Pennington, B. F., and Smith, S. D. (1991). Multiple regression analysis of sib-pair data on reading to detect quantitative trait loci. *Reading and Writing: An Interdisciplinary Journal* **3**: 299–313.

Greene, W. H. (1981). Sample selection as a specification error: Comment. *Econometrica* **49**:795–798.

Greene, W. H. (2000). *Econometric Analysis.* (4th edition). Upper Saddle River, NJ: Prentice Hall.

Hanson, B., McGue, M., Roitman-Johnson, B., Segal, N. L., Bouchard, T. J., and Blumenthal, M. N. (1991). Atopic disease and Immunoglobulin E in twins reared apart and together. *American Journal of Human Genetics* **48**:873–879.

Heath, A. C., Madden P. A. F., and Martin, N. G. (1998). Assessing the effects of cooperation bias and attrition in behavioral genetic research using data-weighting. *Behavior Genetics* **28**(6):415–428.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* **47**:153–161.

Heitjan, D. F., and Basu, S. (1996). Distinguishing between "missing at random" and "missing completely at random." *The American Statistician* **50**(3):207–212.

Koopmans, J. R., Boomsma, D. I., Heath, A. C., and Doornen, L. J. P. (1995). A multivariate genetic analysis of sensation seeking. *Behavior Genetics* **25**:349–356.

Koopmans, J. R., and Boomsma, D. I. (1996). Familial resemblance in alcohol use: Genetic or cultural transmission? *Journal of Studies on Alcohol* **57**:19–28.

Lange, K., Westlake, J., and Spence, M. (1976). Extensions to pedigree analysis: III. Variance components by the scoring method. *Annals of Human Genetics* **39**:485–491.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Little, R. J. A., and Schenker, N. (1995). Missing data. In Arminger, G., Clogg, C., and Sobel, M. (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences.* New York: Plenum Press.

Lykken, D. T., McGue, M., and Tellegen, A. (1987). Recruitment Bias in Twin Research: The Rule of Two-Thirds Reconsidered. *Behavior Genetics* **17**(4):343–362.

Maddala, G. S. (1983). Limited dependent and qualitative variables in econometrics. Econometric society monographs nr. 3, Cambridge University Press.

Nawata, K. (1993). Estimation of sample selection bias models by the maximum likelihood estimator and Heckman's two-step estimator. *Economics Letters* **45**:33–40.

Nawata, K., and Nagase, N. (1996). Estimation of sample selection bias models. *Econometric Reviews* **15**:387–400.

Neale, M. C. (1997). Mx. Statistical Modeling with Mx. Box 126 MCV, Richmond, VA 23298: Department of Psychiatry.

Neale, C. M., Eaves, L. J., Kendler, K. S., and Hewitt, J. K. (1989). Bias in correlations from selected samples of relatives: The effects of soft selection. *Behavior Genetics* **19**(2):163–169.

Neale, M. C., and Eaves, L. J. (1993). Estimating and controlling for the effect of volunteer bias within pairs of relatives. *Behavior Genetics* **23**(3):271–277.

Rubin, D. B. (1976). Inference and Missing data. *Biometrika* **63**: 581–592.

Rubin, D. D. (1986). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Russell, J., Hopp, D. O., Againdra, K., Bewtra, M. D., Gavin, D., Watt, M. P. H., Nicki, M., Hair, M. D., and Townley, R. G. (1984). Genetic analysis of allergic disease in twins. *Clinical Immunology* **73**:265–270.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data.* New York: Chapman and Hall.

Tas, R. F. J. (1990). *Meerlingen in Nederland, 1900–1988 [Multiple Births in the Netherlands, 1990–1988].* Maanstatistiek Bevolking: CBS (Central Bureau of Statistics).

Wilde, A. G. J. S. (1970). *Neurotische labiliteit gemeten volgens de vragenlijstmethode [The Questionnaire Method as a Means of Measuring Neurotic Instability].* Amsterdam: van Rossen.

Wilks, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples, *Annals of Mathematical Statistics* **3**:163–195.

Edited by Stacey Cherny