Part 1:

The Central Account

II

Latent Variable Models

1. *Introduction*

It is often said that the purpose in employing a latent variable model is to "account" for the association that exists in a set of manifest variates. DeLeeuw (1996, p.viii), for example, states that "...we have come to realize that conceptualizing in terms of latent variables is a particular way of looking at the relationships between observables." Indeed, the very fact that the equations of such models contain both manifest and latent variates is what makes them both unique and controversial. Exactly what makes a latent variate *latent* is an issue often given short shrift, experts making haste to the topics of estimation and statistical hypothesis testing. The standard explanation, which will be considered in Chapter III's description of the Central Account, makes mention of the unobservability or unmeasurability of latent variates, and the observability of manifest variates. Save for the fact that latent variable models involve latent variates, these models are structurally akin to ordinary regression models, in that they, like ordinary regression models, make claims about certain aspects of the distribution of a set of manifest variates through restrictions placed on the conditional distribution of the manifest, given the latent variates. In this chapter, an overview of the mathematics of a select few latent variable models is provided, and the evolution of some of the more important ideas of latent variable modeling is traced.

2. *Structure of latent variable models*

2a. The manifest variates

The input into a latent variable analysis is typically an N by p matrix X of real numbers whose ijth element, x_{ij}, is the score of the ith object under study, on the jth *manifest variate*. Latent variable models have been invented to handle a wide range of manifest variate types, including continuous, dichotomous, ordered categorical, and nominal.[1] It is usually the wont of the social and behavioural scientist to view the N objects under study as a random sample from a larger population, *P*, of "similar" objects. In this book, this "sampling issue", and the related need to describe compactly the distribution of the manifest variates in *P*, is handled through the employment of random variates and the idealization of *P* as infinite sized. Several points must be made in regard this choice: i) As noted by Rozeboom (1996), the employment of random variates often does more to obscure than clarify. This is especially true of latent variable modeling, the employment of random variates often being taken as license to forego explanation of the meanings of the symbols that occur in the equations of latent variable models. Random variates will be employed because this is the typical currency of psychometric analysis; ii) The employment of a latent variable model should include a statement detailing the set of properties an object must possess in order to be judged a member of *P*. In published applications, the

[1] Because the aim is not to undertake a comprehensive survey of latent variable models, most of the attention in this book is given to models for continuous and dichotomous manifest variates.

objects under study are almost always humans, and only very rarely is mention made of the rule for inclusion in *P*. In truth, we seldom, in practice, have much of an idea as to how to go about constructing a convincing description of *P*; iii) Taking *P* as infinite sized, while in line with contemporary psychometric practice, has, admittedly, no other function than to simplify the mathematics of the models herein considered.

If the researcher were to be completely honest, he would have to admit that the claim that "the rows of X are a size N random sample from *P*" usually just means that: i) He has gone ahead and collected the scores comprising X according to more or less explicit principles of data collection, and very often convenience has played the greatest role; ii) He would like to employ a particular latent variable model whose use necessitates that the manifest variates have some particular distribution, $F_{\underline{\mathbf{X}}}(\underline{t})$, in an infinite sized population; iii) He is hopeful that it is not too far-fetched to view the objects in his study as having been drawn from a larger collection of "similar" objects (for which the "infinite sized" stipulation is but a mathematical convenience) in which the joint distribution of the manifest variates is approximately as described by $F_{\underline{\mathbf{X}}}(\underline{t})$. Having made these apologies and disclaimers, it must be emphasized that these issues are peripheral to the chief topic of this book, i.e., the Central Account, the lens through which the practice of latent variable modeling is viewed.

Exactly how *meaning* is invested in the scores x_{ij}, and, correlatively, the grounds for attaching a concept-label (e.g., *self-esteem*, *anxiety* item 1, *depression* item 2, etc.) to each variate, is, as will be seen, a topic of profound complexities. It is also a topic which receives little attention in the latent variable modeling literature. For the moment, it will simply be noted that prevailing explanations of the concept of *manifest variate*, and the realizations on these variates that constitute the data in a latent variable analysis, exhaust themselves with a few comments about the form of $F_{\underline{\mathbf{X}}}(\underline{t})$ and assurances that such variates are "observable", "uniquely determined", or "measurable." What these terms mean turns out to be far more problematic than their ubiquitous employment would suggest. Nevertheless, there does exist a received account of these, and related, terms, and this account will be reviewed in Chapter III.

2b. The latent variate

It is customary to distinguish between latent variable models in which the latent variates are random variates, and those in which they are a set of *scores/person parameters* (i.e., a set of degenerate random variates with point distributions). In the former case, the model produced is called a *random latent variable model*, and, in the latter, a *fixed-score latent variable model*. It has been suggested in the literature (e.g., Bartholomew, 1981) that a certain degree of confusion has been visited upon the theory of latent variable models, notably linear factor analysis, by a careless use of terms such as *latent variate*, *factor score*, *factor score estimate*, *regression estimate*, etc., and a corresponding absence of clarity in the specification of the statistical model under consideration. In this book, attention is given only to the random latent variable models. To align the current treatment with prevailing thought on the matter, the following conventions will, therefore, be employed: i) Estimation is an issue only with respect to the structural parameters of the model (i.e., the "item parameters"), and not with respect to the scores that comprise the distribution of the latent variate. That is, in this context there are no such things as "factor scores", "person parameters", or "factor score estimates"; ii) Because the latent variate is jointly distributed with the manifest variates in population *P* under study, prediction (e.g., of the latent variate from the manifest variates), dependency, correlation, etc., are the relevant statistical notions when considering the latent variate itself.

Let there be a (q+1)th random variate, $\boldsymbol{\theta}$, distributed in *P*, called the latent variate, which, then, has a joint distribution with $\underline{\mathbf{X}}$. Let the joint density of $\boldsymbol{\theta}$ and $\underline{\mathbf{X}}$ be symbolized as $f_{\underline{\mathbf{X}},\boldsymbol{\theta}}(\underline{t},s)$. Whereas the $\mathbf{X}_i$ are said to be "observable" or "uniquely defined", the latent variate is said to be "unobservable". Certainly, there do not appear in matrix X, which contains the data to be analyzed in a latent variable analysis, realizations taken on $\boldsymbol{\theta}$. In the absence of the data generating context in which are embedded the manifest variates, the issue of how the meaning of $\boldsymbol{\theta}$, and, correlatively, the scores that comprise its distribution, is determined, becomes a key issue. Here too the Central Account provides the latent variable modeller with ready answers.

2c. The model equations

While there are in use a diverse array of latent variable models, they have in common a characteristic representation, a feature which was, according to Bartholomew (1995), first noted by Anderson (1959), and later by Fielding (1977). In a series of articles, Bartholomew has extended and popularized this representation. The treatment provided herein borrows from both Bartholomew (e.g., Bartholomew & Knott,1999) and Holland (1990). Any latent variable model can be represented as

(2.1)
$$f_{\underline{\mathbf{X}},\boldsymbol{\theta}}(\underline{\mathbf{X}}=\underline{x},\boldsymbol{\theta}=\theta) = f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x} \mid \boldsymbol{\theta}=\theta) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=\theta)$$

in which $f_{\underline{\mathbf{X}},\boldsymbol{\theta}}(\underline{\mathbf{X}}=\underline{x},\boldsymbol{\theta}=\theta)$ is the joint density of the p manifest variates and the latent variate $\boldsymbol{\theta}$, $f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta}=\theta)$ is the conditional density of $\underline{\mathbf{X}}$ given $\boldsymbol{\theta}$, and $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=\theta)$ is the unconditional density of the latent variate (Bartholomew & Knott, 1999). Particular models are generated through the making of particular choices of $f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta}=\theta)$ and $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=\theta)$, and though the choice of restrictions placed on the parameters of these densities. By integrating over the distribution of the unobservable variate $\boldsymbol{\theta}$, one derives a model implied density for the manifest variates:

(2.2)
$$f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x}) = \int_{-\infty}^{\infty} f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x} \mid \boldsymbol{\theta}=t) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=t)dt$$

Hence, a latent variable model is a set of equations relating features of $f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x})$, the unconditional density of the manifest variates, to features of $f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta}=\theta)$ and $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=\theta)$.

Because latent variable models are often used to "account" for the association in the manifest variates, the focal feature of $f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x})$ is very often its set of association parameters. More particularly, since $\boldsymbol{\theta}$ is viewed as being "responsible for" the association amongst the manifest variates, it is standard practice to assert that

(2.3)
$$f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta}=\theta) = \prod_{j=1}^{p} f_{\mathbf{X}_j}(\mathbf{X}_j = x_j \mid \boldsymbol{\theta}=\theta)$$

the property called conditional (local) independence, or that

(2.4) $C(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta}=\theta)= \Psi_\theta$,

in which Ψ_θ is diagonal and positive definite. The property of conditional independence, (2.3), claims that, when the latent variate is held constant, the manifest variates are statistically independent, while (2.4) claims that they are merely conditionally uncorrelated.

2d. The model parameters

A given latent variable model is characterized by a set of m model parameters. Letting the m-vector $\underline{\Pi}$ contain these parameters, $\underline{\Pi}$ then ranges over a subspace, π, of R^m.

2e. The modeled parameters of $f(\underline{\mathbf{X}})$

A latent variable model is ultimately a claim about a subset of t of the parameters of the joint density of the manifest variates. Letting this set of parameters be contained in the vector $\underline{\Omega}$, $\underline{\Omega}$ is then contained within a subspace, Δ, of R^t. Some (e.g., Holland, 1990) have called Δ the data space of latent variable models, but this terminology will be avoided because of its potential to generate confusions of Δ with the data upon which estimation is based. The vector $\underline{\Omega}$ must also be distinguished from the value $\underline{\Omega}_T$ it assumes in a particular population P_T at a particular time.

2f. The model (Claim about $\underline{\Omega}_T$)

A given latent variable model maps points in π, via the model equations, into points in Δ. This "model implied" subset of points in Δ will be symbolized as $\underline{\Omega}(\underline{\Pi})$. Hence, as $\underline{\Pi}$ ranges over all possible values in π, a given model then traces out a surface or manifold, $M:\{\underline{\Omega}(\underline{\Pi})\}\subset \Delta$ (Holland, 1990). In the case of a particular population, P_T, a given latent variable model (a particular choice of $f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta}=\theta)$ and $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=\theta)$, and set of restrictions on the parameters of these densities) can then be viewed as a claim that $\underline{\Omega}_T$ is a point on the manifold M. That is to say, that $\underline{\Omega}_T\subset M$.

2g. Conditional distribution of $\boldsymbol{\theta}$ given $\underline{\mathbf{X}}$

From (2.1), the conditional density of $\boldsymbol{\theta}$ given $\underline{\mathbf{X}}=\underline{x}$, $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=\theta|\underline{\mathbf{X}}=\underline{x})$, is

(2.5) $$\frac{f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x}\mid\boldsymbol{\theta}=\theta)f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=\theta)}{\int_{-\infty}^{\infty} f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x}\mid\boldsymbol{\theta}=t)f_{\boldsymbol{\theta}}(\boldsymbol{\theta}=t)dt}$$

The conditional mean and variance of $\boldsymbol{\theta}$ given $\underline{\mathbf{X}}$=$\underline{x}$, are then, respectively:

(2.6) $$E(\boldsymbol{\theta}|\underline{\mathbf{X}}=\underline{x}) = \int_{-\infty}^{\infty} \boldsymbol{\theta} f_{\boldsymbol{\theta}}(\boldsymbol{\theta} \mid \underline{\mathbf{X}} = \underline{x}) d\boldsymbol{\theta}$$

and

(2.7) $$V(\boldsymbol{\theta}|\underline{\mathbf{X}}=\underline{x}) = \int_{-\infty}^{\infty} (\boldsymbol{\theta} - E(\boldsymbol{\theta} \mid \underline{\mathbf{X}} = \underline{x})^2 f_{\boldsymbol{\theta}}(\boldsymbol{\theta} \mid \underline{\mathbf{X}} = \underline{x}) d\boldsymbol{\theta}$$

3. *A survey of models*

3a. The unidimensional linear common factor (ulcf) model

There exist many versions of the "unidimensional (single-factor) linear factor model." A basic formulation, which does not make a complete statement about the form of f($\boldsymbol{\theta}$) and f($\underline{\mathbf{X}}$|$\boldsymbol{\theta}$), is as follows:

(2.8) $$\underline{\mathbf{X}}=\underline{\Lambda}\boldsymbol{\theta}+\Psi^{1/2}\underline{\boldsymbol{\delta}}$$

and

(2.9) $$E\begin{bmatrix}\boldsymbol{\theta}\\ \underline{\boldsymbol{\delta}}\end{bmatrix} = \underline{0}, \ E\begin{bmatrix}\boldsymbol{\theta}\\ \underline{\boldsymbol{\delta}}\end{bmatrix}\left[\boldsymbol{\theta} : \underline{\boldsymbol{\delta}}'\right] = I,$$ [2]

in which $\boldsymbol{\theta}$, the latent variate, is called the "common factor", $\underline{\boldsymbol{\delta}}$ contains the "specific factors", $\underline{\Lambda}$ is a p×1 vector of "factor loadings", Ψ is a p×p diagonal, positive definite matrix of "specific variances", σ^2_u, and $\underline{\mathbf{X}}$ is a p×1 vector of random, manifest (observed) variates. The model places the following constraints on the parameters of the conditional distribution of $\underline{\mathbf{X}}$ given $\boldsymbol{\theta}=\theta_o$: i) $E(\underline{\mathbf{X}}|\boldsymbol{\theta}=\theta_o) = \underline{\Lambda}\theta_o$; and ii) $C(\underline{\mathbf{X}}|\boldsymbol{\theta}=\theta_o) = \Psi$, in which Ψ is diagonal and positive definite. The latter restriction claims that the latent variate "explains" fully the linear associations amongst the manifest variates in the sense that, if it is held constant, the manifest variates are uncorrelated.

The model parameters are $\underline{\Lambda}$ and Ψ. Hence, $\underline{\Pi}$ is, in this case, the 1×2p vector $[\underline{\Lambda}',\sigma^2_{\zeta 1},...,\sigma^2_{\zeta p}]$. Since the unique variances must be positive, $\underline{\Pi}$ is contained within a subspace,

[2] Specification (2.10) is called, by McDonald (1981) the "weak form" of the principle, as it employs (2.4) rather than (2.3). McDonald (1981) claims that researchers intend (2.3), but, for convenience, employ the weaker counterpart. Given the multinormality of $\underline{\mathbf{X}}$, the two principles are equivalent.

π, of R^{2p}. The only element of the distribution of $\underline{\mathbf{X}}$ of relevance is its covariance matrix Σ. Thus, $\underline{\Omega}$ is, in this case, the $1\times\frac{1}{2}p(p+1)$ vector containing the non-redundant elements of Σ. Because the diagonal elements of Σ must be positive, Ω is contained within a subspace, Δ, of $R^{\frac{1}{2}p(p+1)}$.

Stipulation (2.8)-(2.9) implies that

$$(2.10)\qquad E(\underline{\mathbf{X}})=\underline{0},$$

and that, for some choice of $\underline{\Lambda}$ and Ψ, Σ is equal to

$$(2.11)\qquad \tilde{\tilde{\Sigma}} = E(\underline{\Lambda}\boldsymbol{\theta}+\Psi^{1/2}\underline{\boldsymbol{\delta}})(\underline{\Lambda}\boldsymbol{\theta}+\Psi^{1/2}\underline{\boldsymbol{\delta}})=\underline{\Lambda}\underline{\Lambda}'+\Psi.$$

Thus, $\underline{\Omega}(\underline{\Pi})$ is equal to $Vec(\tilde{\tilde{\Sigma}})$.

Consequence (2.11) is a necessary condition for a set of manifest variates to be described by the model. The "reduced matrix" $\Sigma^{*}=\tilde{\tilde{\Sigma}}-\Psi$ is of rank unity. The model equations (2.11) are a set of $\frac{1}{2}p(p+1)$ equations which map $\underline{\Pi}$ into the unique elements of $\tilde{\tilde{\Sigma}}$. As $\underline{\Pi}$ ranges over π, the model traces out a manifold M_{ulcf} that is a subspace of Δ. Without any further restrictions on $\underline{\Pi}$, the model is only identified for the case of $p\geq3$.

An equivalent formulation of the model is as follows. Let $E(\underline{\mathbf{X}})=\underline{0}$. The unidimensional linear factor model then describes $\underline{\mathbf{X}}$ if there exists a random variate $\boldsymbol{\theta}$ ($E(\boldsymbol{\theta})=0$, $V(\boldsymbol{\theta})=1$), jointly distributed with $\underline{\mathbf{X}}$, such that the residuals of the linear regressions of the $\mathbf{X}_i$ on $\boldsymbol{\theta}$ are mutually uncorrelated. That is

$$(2.12)\qquad E(\underline{\mathbf{X}}|\boldsymbol{\theta}=\theta_o)=\underline{\Lambda}\theta_o,$$

$$(2.13)\qquad E((\underline{\mathbf{X}}-\underline{\Lambda}\theta_o)((\underline{\mathbf{X}}-\underline{\Lambda}\theta_o)'|\boldsymbol{\theta}=\theta_o)=\Psi,\ \Psi \text{ diagonal and positive definite.}$$

Because $\Sigma=C(E(\underline{\mathbf{X}}|\boldsymbol{\theta}))+E(C(\underline{\mathbf{X}}|\boldsymbol{\theta}))$, it follows from (2.12) and (2.13), that, once again, $\tilde{\tilde{\Sigma}}=\underline{\Lambda}\underline{\Lambda}'+\Psi$. Certain applications of the model begin by taking the distribution of $\boldsymbol{\theta}$ to be normal, the model then being stated as follows,

$$(2.14)\qquad \boldsymbol{\theta}\sim N(0,1),$$

and

(2.15) $\underline{\mathbf{X}}|\boldsymbol{\theta}=\theta_o \sim N_p(\underline{\Lambda}\theta_o,\Psi)$,

from which follows, once again, that $\tilde{\Sigma}=C(E(\underline{\mathbf{X}}|\boldsymbol{\theta}))+E(C(\underline{\mathbf{X}}|\boldsymbol{\theta}))=\underline{\Lambda\Lambda}'+\Psi$.

Given that it is a necessary condition for the model to describe a set of manifest variates, it is not surprising that the representation $\tilde{\Sigma}=\underline{\Lambda\Lambda}'+\Psi$ has traditionally been the chief focus of linear factor analysis. For a set of four manifest variates, this representation may be spelled out, element-by-element, as

(2.16)

$$\begin{pmatrix} \sigma^2{}_1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma^2{}_2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma^2{}_3 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma^2{}_4 \end{pmatrix} = \begin{pmatrix} \lambda^2{}_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4 \\ \lambda_2\lambda_1 & \lambda^2{}_2 & \lambda_2\lambda_3 & \lambda_2\lambda_4 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda^2{}_3 & \lambda_3\lambda_4 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda^2{}_4 \end{pmatrix} + \begin{pmatrix} \sigma^2{}_{\zeta 1} & 0 & 0 & 0 \\ 0 & \sigma^2{}_{\zeta 2} & 0 & 0 \\ 0 & 0 & \sigma^2{}_{\zeta 3} & 0 \\ 0 & 0 & 0 & \sigma^2{}_{\zeta 4} \end{pmatrix}.$$

This covariance structure has traditionally been called *hierarchy*, and yields the famous "tetrad differences" of zero:

(2.17) $$\frac{\sigma_{ik}}{\sigma_{il}} - \frac{\sigma_{jk}}{\sigma_{jl}} = \frac{\lambda_i\lambda_k}{\lambda_i\lambda_l} - \frac{\lambda_j\lambda_k}{\lambda_j\lambda_l} = 0, \quad k \neq l; i \neq j.$$

Early linear factor analytic algorithms generally sought to find a diagonal, positive definite matrix, M, such that Σ-M was gramian and of as small a rank as possible, after which, principal component analysis was applied to the reduced matrix Σ-M. The optimal choice of the matrix M was the subject of much early work in factor analysis (cf. the communalities problem). More statistically transparent means of estimating the parameters of the model including, under model (2.14)-(2.15), maximum likelihood, are now generally employed.

Because $\boldsymbol{\theta}$ is taken to be unobservable, it is said that "factor scores cannot be determined precisely, but only estimated" (Pawlik, 1968; McDonald & Burr, 1967). Given the random linear factor model, this view would be expressed analogously as "an individual's score with respect to the common factor cannot be determined, but only *predicted*." Many different predictors have been invented, and McDonald and Burr (1967) review the statistical properties of four of these. Bartholomew (1980) notes that the tone and substance of many facets of the theory of "factor score estimation" are ill-conceived, as they incorrectly treat the random common factor as a set of "person parameters", thus recasting the problem as a matter of estimation. In fact, $\boldsymbol{\theta}$ and $\underline{\mathbf{X}}$ have a joint distribution, and, hence, $\boldsymbol{\theta}$ has a distribution conditional on $\underline{\mathbf{X}}$. What has traditionally been called the "least-squares estimator" (e.g., as in McDonald &

Burr, 1967), is, under the random, unidimensional linear factor model, merely the conditional expectation of $\boldsymbol{\theta}$ given $\underline{\mathbf{X}}$ (Bartholomew, 1981), i.e.,

(2.18) $$E(\boldsymbol{\theta}|\underline{\mathbf{X}}=\underline{x})=\underline{\Lambda}'\Sigma^{-1}\underline{x}.$$

The conditional variance of $\boldsymbol{\theta}$ given $\underline{\mathbf{X}}$ is

(2.19) $$1-\underline{\Lambda}'\Sigma^{-1}\underline{\Lambda}.$$

3b. The LISREL model

LISREL, an acronym for *LI*near *S*tructural *REL*ations, is both an accepted alternative label for the general structural equation model (the JKW model) of Joreskog (1973), Keesling (1972), and Wiley (1973), and the name of a marketed structural equation modeling program (Joreskog and Sorbom, 1979). LISREL is, essentially, a conjoining of multivariate regression and factor analysis. Let $\underline{\boldsymbol{\rho}}'=[\underline{\boldsymbol{\eta}}',\underline{\boldsymbol{\xi}}',\underline{\boldsymbol{\zeta}}',\underline{\boldsymbol{\delta}}',\underline{\boldsymbol{\varepsilon}}']$ be a set of jointly distributed random, latent variates, with $E(\underline{\boldsymbol{\rho}}')=\underline{0}'$, and

(2.20) $$\underline{\boldsymbol{\eta}}=B\underline{\boldsymbol{\eta}}+\Gamma\underline{\boldsymbol{\xi}}+\underline{\boldsymbol{\zeta}}.$$

In the above, B is an m×m matrix of coefficients with zeros on the diagonal and with the property that (I-B) is non-singular; Γ is an m×n matrix of coefficients; $\underline{\boldsymbol{\eta}}$ is an m×1 vector of "latent dependent" or "endogenous" (Joreskog and Sorbom, 1993) variates; $\underline{\boldsymbol{\xi}}$ is an n×1 vector of "latent independent", or "exogenous" (Joreskog and Sorbom, 1993) variates; $\underline{\boldsymbol{\zeta}}$ is an m×1 vector of regression residuals (Joreskog and Sorbom, 1993); $\underline{\boldsymbol{\delta}}$ is a q×1 vector of "measurement errors"; and $\underline{\boldsymbol{\varepsilon}}$ is a p×1 vector of "measurement errors." The following stipulations are made regarding the distribution of $\underline{\boldsymbol{\rho}}'$:

(2.21) $$\Sigma_{\eta,\varepsilon}=\varnothing;\ \Sigma_{\xi,\delta}=\varnothing;\ \Sigma_{\xi,\zeta}=\varnothing;\ \Sigma_{\zeta,\delta}=\varnothing;\ \Sigma_{\zeta,\varepsilon}=\varnothing;\ \Sigma_{\delta,\varepsilon}=\varnothing.$$

The pair of (2.20) and (2.21) imply that Σ_{ρ} has the form:

(2.22)
$$\begin{pmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma'+\Psi)(I-B)^{-1}{}' & (I-B)^{-1}\Gamma\Phi & (I-B)^{-1}\Psi & \circ & \circ \\ \Phi'\Gamma'(I-B)^{-1}{}' & \Phi & \circ & \circ & \circ \\ \Psi(I-B)^{-1}{}' & \circ & \Psi & \circ & \circ \\ \circ & \circ & \circ & \Theta_\delta & \circ \\ \circ & \circ & \circ & \circ & \Theta_\varepsilon \end{pmatrix}$$

in which Φ is the n×n covariance matrix of $\underline{\boldsymbol{\xi}}$, Ψ the m×m covariance matrix of $\underline{\boldsymbol{\zeta}}$, Θ_δ the q×q covariance matrix of $\underline{\boldsymbol{\delta}}$, and Θ_ε the p×p covariance matrix of $\underline{\boldsymbol{\varepsilon}}$.

Let $\underline{\mathbf{z}}'=[\underline{\mathbf{y}}',\underline{\mathbf{x}}']$ be a set of manifest (observed) variates or "indicators" related to the latent variates as described by the equations

(2.23)
$$\begin{pmatrix} \underline{\mathbf{y}} \\ \underline{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} \Lambda_y & \\ & \Lambda_x \end{pmatrix} \begin{pmatrix} \underline{\boldsymbol{\eta}} \\ \underline{\boldsymbol{\xi}} \end{pmatrix} + \begin{pmatrix} \underline{\boldsymbol{\varepsilon}} \\ \underline{\boldsymbol{\delta}} \end{pmatrix}.$$

In (2.23), Λ_y is a p×m matrix of coefficients ("factor loadings") and Λ_x is a q×n matrix of coefficients ("factor loadings"). From (2.22)-(2.23), it follows that the LISREL models claims that Σ_z has the form:

(2.24)
$$\overset{\approx}{\Sigma} = \begin{pmatrix} \Lambda_y(I-B)^{-1}(\Gamma\Phi\Gamma'+\Psi)(I-B)^{-1}{}'\Lambda_y{}'+\Theta_\varepsilon & \Lambda_y(I-B)^{-1}\Gamma\Phi\Lambda_x{}' \\ \Lambda_x\Phi\Gamma'(I-B)^{-1}{}'\Lambda_y{}' & \Lambda_x\Phi\Lambda_x{}'+\Theta_\delta \end{pmatrix}$$

The model parameters are the $((p\times m)+(q\times n)+q^2+p^2+n(n-1)+m^2+m(m-1)+mn)$ elements of the matrices $[\Lambda_x,\Lambda_y,\Theta_\delta,\Theta_\varepsilon,\Phi,\Psi,B,\Gamma]$, and, hence, these are the elements that are contained in $\underline{\Pi}$. On the other hand, $\underline{\Omega}$ is the $1\times(.5(p+q)(p+q+1))$ vector containing the unique elements of Σ_z. The model equations are the set of $.5(p+q)(p+q+1)$ equations $\underline{\Omega}(\underline{\Pi})=\mathrm{Vec}(\overset{\approx}{\Sigma})=\Sigma_z$ that link the elements of $\underline{\Pi}$ to the elements of Σ_z. Without further restrictions on $\underline{\Pi}$ the model is not identified, and the estimation of its parameters is not possible. Which restrictions should be placed is both a matter of substantive consideration, and identifiability conditions which have been theoretically deduced for various values of $(.5(p+q)(p+q+1))$, various types of model, and various brands of restriction. Commonly employed restrictions include: i) requiring that Ψ, Θ_δ, and Θ_ε each be diagonal; ii) requiring that a certain number of zeros be included in each of the columns of Λ_x and Λ_y; iii) Requiring that B be lower triangular (i.e, that the model be recursive). For purposes of statistical inference, e.g., estimation via maximum likelihood, $\underline{\mathbf{z}}$ is often asserted to have a multivariate normal distribution.

3c. Non-linear factor models

McDonald (1983) discusses a class of latent variable models that are non-linear extensions of the linear factor model. In such models,

(2.25) $\boldsymbol{\theta}\sim N(0,1)$, and

(2.26) $\underline{\mathbf{X}}|\boldsymbol{\theta}=\theta_o \sim N_p(\Lambda\underline{f}(\theta_o),\Psi)$,

in which $\underline{f}(\boldsymbol{\theta})$ is a vector of r functions (in general, nonlinear) of $\boldsymbol{\theta}$, Λ a p×r matrix of real coefficients, Ψ a p×p diagonal, positive definite matrix of residual variances, and $r<p$. Hence, the regressions of the manifest variates on $\boldsymbol{\theta}$ are linear functions of the parameters Λ, but nonlinear functions of $\boldsymbol{\theta}$. It follows from (2.25) and (2.26) that the model implied covariance matrix is equal to

(2.27) $\tilde{\Sigma}=C(E(\underline{\mathbf{X}}|\boldsymbol{\theta}))+E(C(\underline{\mathbf{X}}|\boldsymbol{\theta}))=\Lambda C(\underline{f}(\boldsymbol{\theta}),\underline{f}(\boldsymbol{\theta})')\Lambda'+\Psi=\Lambda\Sigma_{\underline{f}(\boldsymbol{\theta})}\Lambda'+\Psi$,

in which $\Sigma_{\underline{f}(\boldsymbol{\theta})}=C(\underline{f}(\boldsymbol{\theta}),\underline{f}(\boldsymbol{\theta})')$. Let $\Sigma_{\underline{f}(\boldsymbol{\theta})}=GG'$, and

(2.28) $\underline{\varphi}(\boldsymbol{\theta})=G^{-1}\underline{f}(\boldsymbol{\theta})$.

It follows then that $\tilde{\Sigma}=AA'+\Psi$ in which $A=\Lambda G$ and $E(\underline{\mathbf{X}}|\boldsymbol{\theta}=\theta_o)=A\underline{\varphi}(\boldsymbol{\theta})$.

McDonald (1967, 1983) has considered a number of special cases of this class of latent variable models. For example, model (2.25)-(2.26) with the choice

(2.29) $\underline{f}(\boldsymbol{\theta})=[\boldsymbol{\theta},\dfrac{(\boldsymbol{\theta}-1)^2}{\sqrt{2}}]$,

produces a unidimensional, quadratic factor model. That is, since $E(\mathbf{X}_i|\boldsymbol{\theta}=\theta_o)=a_{i1}\theta_o+a_{i2}\dfrac{(\theta_o-1)^2}{\sqrt{2}}$, each item/factor regression can be either linear or quadratic, depending on whether or not a_{i2} is, or is not, equal to zero. Notice that with choice (2.29), $C(\underline{f}(\boldsymbol{\theta}))=I$, so that $\Sigma=AA'+\Psi$ in which A is a p×2 matrix. This implies that at the level of implied covariance structure, the unidimensional, quadratic factor model cannot be distinguished from the two-dimensional linear factor model (McDonald, 1967). A unidimensional, cubic factor model is created by adding a third element to $\underline{f}(\boldsymbol{\theta})$, and, at the level of implied covariance structure, is identical to a three-dimensional linear factor model. McDonald initially employed graphical plots of the distribution of Bartlett factor predictors to distinguish between linear and such polynomial cases, but, later, was able to invent confirmatory procedures for such decision making.

3d. Item response models

Item response models were invented to deal with the problem of the inappropriateness of linear factor models in analyzing dichotomous manifest variates (usually test items). The classical item response models were generalizations of the linear common factor model which answered the need for manifest/latent variate regressions to have a lower asymptote of zero and an upper asymptote of unity. Subsequently, they were further developed so as to accomodate categorical manifest variate types other than dichotomies. Here, the classical 2-parameter item response models for dichotomous variates are reviewed.

For a set of q dichotomous (0/1) manifest variates, let $\underline{x}=(x_1,x_2,...,x_q)$ be a given *response pattern* (i.e., a set of 0's and 1's). There are $t=2^q$ such response patterns and let these be contained in set S. Each object in population *P* under study is described by such a pattern. Define $P(\underline{\mathbf{X}}=\underline{x})$ to be the proportion of objects in *P* with response pattern $\underline{x}$. There are t such proportions, one corresponding to each of the t response patterns, and these proportions satisfy $P(\underline{\mathbf{X}}=\underline{x})\geq 0$ and $\sum_{\underline{x}\in S} P(\underline{\mathbf{X}}=\underline{x})=1$. In this case, $\underline{\Omega}$ contains the $P(\underline{\mathbf{X}}=\underline{x})$ (ordered in some particular manner), and is an element of a (t-1)-dimensional subspace (probability simplex), Δ, of R^t.

The model equations for any unidimensional item response model map a particular set of values of the model parameters into a particular value of $\underline{\mathbf{\Omega}}$, and are of the form

(2.30) $$P(\underline{\mathbf{X}}=\underline{x})=\int_A P(\underline{\mathbf{X}}=\underline{x}\mid\boldsymbol{\theta})f(\boldsymbol{\theta})d\theta .$$

Specific models are specified through restrictions on $P(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta})$ and $f(\boldsymbol{\theta})$. With regard $P(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta})$, Holland and Rosenbaum (1986) note that all classical item response models assert that

(2.31) $$P(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta}=\theta)=\prod_{j=1}^{p}P(\mathbf{X}_j=x_j\mid\boldsymbol{\theta}=\theta)$$

$$=\prod_{j=1}^{p}P(\mathbf{X}_j=1\mid\boldsymbol{\theta}=\theta)^{x_j}(1-P(\mathbf{X}_j=1\mid\boldsymbol{\theta}=\theta))^{1-x_j}$$

which claims that, conditional on the latent variate, the manifest variates are statistically independent[3]. That is, all of the association amongst the manifest variates is "explained" by their dependence on the latent variate. The conditional probabilities $P(\mathbf{X}_j=x_j|\boldsymbol{\theta})$ are called "item characteristic curves" or "item trace lines", and, with regard these manifest variate/latent variate regressions, one of two possibilities is generally adopted:

[3] Holland (1981) considers item response models that are founded on weaker senses of dependency.

(2.32a) $$\frac{\exp(g(\underline{\Pi}_j;\boldsymbol{\theta}))}{1+\exp(g(\underline{\Pi}_j;\boldsymbol{\theta}))}$$

or

(2.32b) $$\Phi(g(\underline{\Pi}_j;\boldsymbol{\theta}))$$

in which $\underline{\Pi}_j$ is a vector containing the model parameters associated with variate j, g(·;·) is a function, and Φ(·) is the standard normal distribution dunction. Finally, there exists a number of standard choices with respect $g(\underline{\Pi}_j;\boldsymbol{\theta})$:

(2.33a) $$a_j(\boldsymbol{\theta}-b_j);$$

(2.33b) $$\frac{\lambda_j\boldsymbol{\theta}-\tau_j}{\sqrt{1-\lambda_j^2}};$$

(2.33c) $$\alpha_j\boldsymbol{\theta}+c_j.$$

The parameters a_j are called slope or discrimination parameters, the b_j, difficulty parameters, the λ_j, loadings, and the τ_j, thresholds. Choice (2.32b) in conjunction with (2.33a) is described in Lord and Novick (1968), and (2.33b), by, e.g., Muthen (1978). These three parameterizations are included in the output of the TESTFACT program of Wilson, Wood, and Gibbons (1991). There exists a one-one correspondence between each pair of parametrizations: By knowing, e.g., the values of a_j and b_j in (2.33a), the values of λ_j and τ_j in (2.33b) can be derived. For these models, $\underline{\Pi}$ is a 1×2p vector $\text{Vec}(\underline{\Pi}_1,\underline{\Pi}_2,...,\underline{\Pi}_p)$ whose elements will be either λ's and τ's, a's and b's, or α's and c's, one pair for each variate.

Classical treatments have often further assumed that $f_{\boldsymbol{\theta}}(\theta)$ is standard normal, while, in recent times, semi- and non-parametric treatments have been derived in which no specification of the density of $\boldsymbol{\theta}$ is made. In fact, certain treatments assume nothing more than that the $P(\mathbf{X}_j=1|\boldsymbol{\theta}=\theta)$ are monotone increasing functions of θ, thus producing a class of models which Holland and Rosenbaum call Unidimensional, Monotone, Latent Variable (UMLV) models. Regardless, for a given choice of one of (2.33a) to (2.33c), one of (2.32a) or (2.32b), and an $f_{\boldsymbol{\theta}}(\theta)$, the model equations (2.30) map values of $\underline{\Pi}$ into points in Δ, tracing out a manifold, M (a subspace of Δ). For various further restrictions on these choices, Holland (1990) describes the geometry of some of these manifolds.

3f. The latent class and profile models

This class of models is a simple generalization of the models covered in (2.2ii) and (2.2iv). The difference is that the distribution of $\boldsymbol{\theta}$ is not continuous, but is rather an m-point discrete distribution, each point corresponding to a "class" or "type" of object contained in *P*. Let $\boldsymbol{\theta}$ assume the m integer values unity to m, so that

$$(2.34) \qquad P(\boldsymbol{\theta}=i)=\pi_i \text{ and } \sum_{i=1}^{m}\pi_i = 1 .$$

Each of the integer values then represents a "class" of individuals, and π_i is the proportion of individuals in population *P* who belong to class i. If the manifest variates are dichotomous, the model has traditionally been called a latent class model, while if they are continuous, a latent profile model. In the former case, the local independence stipulation is that

$$(2.35) \qquad P(\underline{\mathbf{X}}=\underline{x}|\boldsymbol{\theta}=i) = \prod_{j=1}^{p} P(\mathbf{X}_{ij} = x_j \mid \boldsymbol{\theta} = i)^{x_j}(1-P(\mathbf{X}_{ij} = x_j \mid \boldsymbol{\theta} = i))^{1-x_j}$$

which implies the unconditional probability mass function

$$(2.36) \qquad P(\underline{\mathbf{X}}=\underline{x})= \sum_{i=1}^{m}\pi_i \prod_{j=1}^{p} P(\mathbf{X}_{ij} = x_j \mid \boldsymbol{\theta} = i)^{x_j}(1-P(\mathbf{X}_{ij} = x_j \mid \boldsymbol{\theta} = i))^{1-x_j} .$$

In the latter case, the local independence claim is

$$(2.37) \qquad f_{(\underline{\mathbf{X}}|\boldsymbol{\theta}=i)} = \prod_{j=1}^{p} f_{(\mathbf{X}_j|\boldsymbol{\theta}=i)}$$

whereby,

$$(2.38) \qquad f_{\underline{\mathbf{X}}}= \sum_{i=1}^{m}\pi_i \prod_{j=1}^{p} f_{(\mathbf{X}_j|\boldsymbol{\theta}=i)}$$

Often, the further stipulation is made that $f_{(\mathbf{x}_j|\boldsymbol{\theta}=i)}$ are normal densities.

The model parameters of the latent class model are the m class proportions π_i and the mp conditional probabilities $P(\mathbf{X}_{ij}=x_j|\boldsymbol{\theta}=i)$. However, since, from (2.34), there are only (m-1) unique π_i, $\underline{\Pi}$ will be an nm+(m-1) vector. The vector $\underline{\Omega}$, on the other hand, contains the $t=2^p$ probabilities $P(\underline{\mathbf{X}}=\underline{x})$ (ordered in some particular manner) and is an element of a (t-1)-dimensional subspace, Δ, of R^t. For the latent profile model with $f_{(\mathbf{x}_j|\boldsymbol{\theta}=i)}$ normal densities, $\underline{\Pi}$ will contain (m-1) linearly indepedent π_i, mp conditional means $E(\mathbf{X}_j|\boldsymbol{\theta}=i)=\mu_{ij}$, and mp conditional variances σ_{ij}. Since $\underline{\Omega}$ will include only the p means and $\frac{1}{2}p(p+1)$ unique elements of the covariance matrix of the distribution of the $\mathbf{X}_j$, the model, without further restrictions, is unidentified.

4. *Practice*

The following is a pared down version of what actually goes on in a latent variable analysis. The employments of latent variable models in attempts to study the psychometric properties of psychological inventories, for example, are multifaceted and diverse. It is also true that distinct latent variable models are very often situated within their own unique analytic practices. Nevertheless, the sketch provided is, in all respects pertinent to the topics addressed in this book, accurate.

i. The investigator takes a "random sample" of size N of the objects contained in population P_T, each member of the sample possessing a score on each of the p manifest variates, and this data stored in an N×p data matrix, X. When a particular unidimensional model, lvm, is "fit" to this data, this means something along the lines of the following:

ii. Since $\underline{\Omega}_T$ is a parameter of $f_{\underline{\mathbf{X}}}(\underline{\mathbf{X}}=\underline{x})$, the joint distribution of the $\mathbf{X}_j$, j=1..p, in population P_T, the researcher employs the data X to estimate it. Symbolize this estimate as $\hat{\underline{\Omega}}_T$.

iii. The researcher chooses a "fit function", $\text{FIT}((\underline{\Omega};\underline{\Omega}(\underline{\Pi}))$, which quantifies the discrepancy between any particular value of $\underline{\Omega}$ and any particular value of $\underline{\Omega}(\underline{\Pi})$, a point in M_{lvm} (the manifold traced out by lvm).

iv. The parameter vector, $\underline{\Pi}$, is estimated by choosing $\hat{\underline{\Pi}}$ to be that value that makes $\text{FIT}(\hat{\underline{\Omega}}_T;\underline{\Omega}(\hat{\underline{\Pi}}))$ as small as possible, say, κ.

v. Judgment is made as to whether κ is, or is not, small enough to support the claim that $\underline{\Omega}_T \subset M_{lvm}$, i.e., that, in P_T, the manifest variates are described (at least to close approximation) by lvm.

vi. If the judgment rendered in (v) is in the affirmative, $\hat{\underline{\Pi}}$ is taken seriously as an estimate of $\underline{\Pi}$, and the results are "interpreted", the interpretation resting heavily on those elements of the Central Account favoured by the researcher who carried out the analysis (Chapter III is devoted to an overview of the Central Account). If the answer is in the negative, any of a number of secondary strategies are employed, including the employment of an analogous multidimensional model.

5. *The concept latent variate to* $\underline{\mathbf{X}}$

In discussing latent variable models, researchers often speak in general terms about "latent variates", and psychometricians, when discussing the mathematical details of latent variable models, standardly employ the symbol $\boldsymbol{\theta}$ to stand for such variates. It will be essential to distinguish this general usage from that involved when, for particular model, lvm, and population, P_{T}, it is actually the case that $\underline{\Omega_{\mathrm{T}}}\subset \mathrm{M}_{\mathrm{lvm}}$. In such instances the general terminology will be replaced by the concept *latent variate to* $\underline{\boldsymbol{X}}$ (with respect to lvm and P_{T}). The distribution of the random variate $\boldsymbol{\theta}$ will, similarly, be renamed the distribution of the random variate $\boldsymbol{\theta}$ to $\underline{\mathbf{X}}$ (with respect to lvm and P_{T}). What is signified by the concept *latent variate to* $\underline{\boldsymbol{X}}$ (hence, what is the *meaning* of the scores that comprise the distribution of $\boldsymbol{\theta}$ to $\underline{\mathbf{X}}$) is the chief issue of this book. The Central Account makes a very specific set of claims with respect this issue, claims which, in this book, will be disputed. The CA claims that the concept *latent variate to* $\underline{\boldsymbol{X}}$ signifies a property/attribute (cause) of the phenomena represented by the manifest variates. Because this referent is unobservable, its "identity" is claimed to be unknowable, and, hence, the subject of inference. This inference is made when the analyst "interprets the latent variate." The scores which comprise the distribution of $\boldsymbol{\theta}$ to $\underline{\mathbf{X}}$ are taken to be measurements with respect this detected property/attribute (causal source).

6. *Historical context of terms and ideas*

Latent variable modeling has a long and complicated history. Since, in this book, it will be argued that the place of latent variable models in scientific work is misportrayed by the received account, the Central Account, it will be useful to sketch out some of the historical developments from which modern practice has evolved. This is the aim of the final section of this chapter.

6a. Precursors

The roots of latent variable modeling are found in developments that took place in the late 1800s and early 1900s, which include: i) The invention of psychophysical models; ii) The development of the theory of association and correlation; iii) The growing recognition of the importance of collecting data that contained as little measurement error as possible, and the consequent need to quantify the amount of measurement error in a wide variety of data collection settings. Each of these developments played a role in setting the stage for both the technical innovations that were to occurr, and the growth of the unique brand of thinking that came to be closely tied to these innovations.

a) Empirical studies in psychophysics had led researchers to conclude that stimulus discrimination was related to the physical magnitudes of the discriminated stimuli. The phi-gamma law, roughly 1880, claimed that the probability that an observer would detect a stimuli was a monotonic function of its intensity. Fechner called the difference in stimuli magnitudes that was detectable 50% of time, the "just noticeable difference" (jnd). The important point to note, here, is that this work made salient the idea that the probability of an individual's responding to a stimuli possessing amount *t* of some property could be expressed as a smooth function of *t*. This idea was the seed from which would grow the notion of the item characteristic curve (icc), which represents responding to a set of manifest variates to the latent variable, and which came to be a foundational concept of latent variable modeling.

b) Bravais (1846) began development of what Galton would later call the coefficient of correlation. Galton (1875) took over the development of the coefficient of correlation, and used it to quantify association in a series of experiments on sweet pea seeds, making this work known in an 1877 address, *Typical Laws of Heredity in Man*. Just before the turn of the century, Karl Pearson refined and extended Galton's work on the correlation coefficient and invented the multiple correlation coefficient. From approximately 1900 to 1915, Pearson and Udny Yule debated the correct theoretical foundations for correlation theory, and, in particular, the appropriate theory for the treatment of contingency tables. Yule developed coefficients which quantified association based only on the cell and marginal probabilities that characterize such tables. Pearson, on the other hand, prefered to found the treatment on the the idea that contingency tables arose from the discretization of continuous joint distributions. This led to his invention of the biserial and tetrachoric correlation coefficients, among others, and, importantly, offered the conceptualization of observed data as "arising" from "underlying", inaccessible structures, a conceptualization that would ultimately grow into the unobservability talk of modern latent variable modeling.

c) A third major influence on the development of latent variable modeling was the growing recognition that the presence of measurement error was detrimental to the drawing of valid scientific conclusions. The "personal equation" derived by Bessel in 1821 aimed to account for variations in astronomical measurements resulting from individual differences of observers, and the Gaussian distribution was developed to model the distribution of errors of measurement within astronomy. This heightened awareness set the stage for attempts within the behavioural sciences to quantify measurement error, leading to true-score theory, and, eventually, structural equation modeling.

6b. History of some important ideas of latent variable modeling

Classical true-score theory

True-score models were invented to address the issue of measurement error as it was believed to arise in the social and behavioural sciences. While not mathematically identical, latent variable and true-score models share a close kinship, and their developments were intertwined. The true score plus error conception, on which virtually all test theory has since been based, was suggested to Charles Spearman in a letter from Udny Yule (see Spearman, 1910). In his 1904 landmark paper, *Proof and measurement of Association between Two Things*, Spearman discusses the need to consider the effect of measurement error in empirical psychological work, and, especially, in regard its impact on the correlation coefficient. He also: i) Considers various measures of correlation, including that of Bravais-Pearson, those of ranks, and the views of Yule and Pearson on the measurement of association in contingency tables; ii) Distinguishes between systematic and accidental deviations that arise in the measurement of an object; iii) Introduces the disattenuated correlation; iv) Presents a formula for the partial correlation coefficient, a coefficient whose invention he admits is due to Yule; v) Presents a formula that he claims, without proof, relates the lengths and reliabilities of two "tests", to their correlation. According to Levy (1995, pp. 224-225), Spearman's *Demonstration of formulae for* *true measurement of correlation* (1907) quite possibly contains the first correlation of a manifest and latent variate.

The sociologist Louis Guttman (1945, *A basis for analyzing test-retest reliability*) formalized Spearman's rather informally stated conceptions of reliability, his treatment founded on the notion of the "propensity distribution". The propensity distribution of a given object under study, with respect a property Y, contains all of the measurements of the object with respect Y that *could* possibly be taken under essentially identical conditions. Hence, each member of a population *P* of objects under study possesses its own propsensity distribution. The object's "true-score" with respect property Y is the mean of its propensity distribution. From this, Guttman derived definitions of all of the core concepts of classical reliability, and six different lower bounds to reliability, including L_3, which would come to be known as Cronbach's α.

In his review of test theory, Lewis (1986, p.12) claims that the publication of Lord and Novick's *Statistical theories of mental test scores* (1968) was an event that "...clearly had a major impact on the field." In the opinion of Lewis, the book represented "A comprehensive and integrated coverage of test theory topics...in which the fundamental assumptions and essentially stochastic nature of the subject were stressed. For the first time, researchers in test theory had available the common language of modern statistical inference with which to discuss all aspects of their field" (1986, p.12). Not noted by Lewis is the fact that Lord and Novick's treatment was drenched in empirical realist philosophy, a feature that, as will be seen, was influential in regard the way in which psychometricians portrayed the nature of measurement, science, and model. Lord and Novick's treatment, along with the 1955 publication of Cronbach and Meehl's account of construct validation theory, offered a view of measurement in which unobservability was equated with latency, and in which statistical models were required to make inferences about unobservable or latent entities. As will be seen, the treatment given to the topic by Lord and Novick was what is called, herein, the CAM. Lord and Novick's book contained perhaps the first systematic discussion of the role of item response models in the building of cases about the measurement of psychological phenomena.

Linear factor analysis

According to Cyril Burt (1940, p.42), Sully was the first to use the term "factor" to designate a principle of classification within the domain of psychological investigation. Spearman's invention of linear factor analysis occurred during a period in which there was intense competition between the theory of intellectual organization offered by Spencer and Ward, on the one hand, and that of Cattell and Wissler, on the other. Spencer and Ward had put forth the "Theorem of intellectual unity" in opposition to the views of Pearson and Galton, who believed that the universe of mental activities were underscored by a variety of distinct "general" and "special" abilities. Spearman sided with Spencer and Ward. His "two-factor theory" of intellectual functioning posited that an individual's standing on any intellectual capacity was expressible as a linear combination of his standing on: i) *general intelligence* (*g*); and ii) a factor *specific* to the capacity. In roughly 1901, Spearman had become interested in Galton and Pearson's work on correlation and had deduced how this work could be used to test his ideas about the nature of intelligence. In 1902, he obtained data suitable to the the testing of his theory, and, in 1904, published *General intelligence, objectively determined and measured*, which contained the first use of the term *factor* in its technical, factor analytic sense. However, as Bartholomew (1995, p.211) has noted, "The modern reader turning to the original paper of 1904 may not immediately recognize that it has anything at all to do with factor analysis." Although earlier models, e.g., the models of psychophysics, contained what would come to be known as latent variates, linear factor analysis is often seen as the first true latent variable model. Spearman, of course, did not employ the term *latent variate*, but spoke in terms of *factor*, *g*, *general intelligence*.

Thurstone renamed Spearman's "two-factor theory" the "single-factor model" (Bartholomew, 1995, p.214). Garnett (1919) invented multiple factor analysis, and Thurstone (1931-) developed and popularized it, invented the term *factor analysis* (Horst, 1965, p.3), and invented the simple structure principle to resolve the rotational indeterminacy inherent to the model. Lederman (1937) derived an inequality concerning the smallest rank to which a correlation matrix could be reduced by adjusting its diagonal. Guttman (1940) proved that the limiting value of the multiple correlation of a manifest variate with the remaining manifest variates, as number of variates increases without bound, is equal to the communality of the variate.

Lawley (1940) and Lawley and Maxwell (1963) brought factor analysis into the fold of modern statistical theory, restating it in terms of random variates, and showing how the factor model's parameters could be estimated by maximum likelihood. Ledyard Tucker (1955) invented the distinction between exploratory and confirmatory factor analysis, and Howe (1955) developed one of the first successfully convergent algorithms for maximum likelihood estimation of the parameters of the basic exploratory model. Guttman (1956) wrote a number of foundational papers, including one on the "best possible" systematic estimates for communalities. Joreskog (1967) derived the maximum likelihood equations of the exploratory model, developed an efficient algorithm for parameter estimation using the Fletcher-Powell method, and succeeded in inventing the first usable confirmatory factor analytic strategy (Joreskog, 1969).

Item response theory

In 1904, Müller developed a method of fitting the normal ogive to data using weights. Urban showed, in 1910, that the Müller weights were incorrect, and derived the correct weighting scheme. In around 1915, Binet and Simon appear to have first employed a primitive "item characteristic curve", their curve representing the relationship between "probability of passing an item" and "age". In 1925 and 1927 Thurstone published detailed accounts of his application of psychophysical methods to the analysis of attitude items. These techniques were based on the idea of a latent "attitude" continuum and a function relating the continuum to the probability of responding in a particular way. However, Thurstone and Chave (1929) claim that Cattell, in his study of "degree of eminence of scientific me", was the first to extend psychophysical methods to stimuli other than simple sensory values.

According to Baker (1965), Richardson (1936) was the first to apply the constant method to test data. Certainly, Richardson (1936) discusses the *difficulty* of each of a set of items, and defines this property as the proportion of individuals passing the item. Using the standard normal curve, he estimates item parameters by minimizing the sum of squared differences between observed and model implied proportions of passing. His discussion of the conditionality of a test's validity on a particular point on the ability continuum presages the later development of the item and test information curves that are now central to item response theory. In his article *Item selection by the constant method* (1942), Ferguson comments that "Within recent years an increasing tendency has become apparent among psychometricians to bring the methods of mental measurement more directly in line with the psychophysical methods of experimental psychology" (p.19). According to Ferguson (1942, p.19), Guilford (1936) "furnished a complete formulation of the problem...but offered no solution." Ferguson notes that "If one could establish a scale of difficulty in psychological units, it would be possible to identify any test item whatsoever by giving its median value and its 'precision' value in terms of h as in the method of constant stimuli" (1942, p.19). According to Guilford (1936) the use of the constant process to estimate the limen value in psychophysics rests on the phi-gamma process: the probability of detection as a function of intensity of the signal to be detected. Ferguson states that "In the application of the constant process to item selection, we define the limen as that point measured in "σ-units" of ability where the probability of a person of that ability either passing or failing the item is one half" (1942, p.20). On page 21, he employs the standard normal distribution to model the probability of passing as a function of ability.

In 1943, D.N. Lawley took up Ferguson's work: "The method of approach we shall adopt has been suggested by an article of Ferguson (1942), and is based on the elementary theory of probability" (1943, p.273). Based on the hypothesis of Ferguson, that the probability of passing an item may be modelled by the standard normal curve, Lawley provides a method whereby the two item parameters *a* (the limen, or point on the ability scale at which the probability of passing is .5) and σ (which determines how well the item discriminates between individuals of high and low ability) are estimated. He also derives formulae for the expected value and standard error of the test score of individuals, and employs what later will be called conditional maximum likelihood, his approach recquiring the formation of homogeneous groups with respect to ability.

In 1944, D.J. Finney showed how developments in probit analysis could be used to solve Ferguson's problem regarding estimation of limina and precisions of items. He states that "In such a test each item admits of only a quantal response from each person attempting it: either he passes or he fails. The problem considered by Ferguson is thus closely analogous to those arising in connection with other material giving quantal responses, and in particular with much toxicological experimentation" (1944, p.31). Moreover, "When the problem of the toxicologist is compared with the psychometrical problems discussed by Ferguson, it is seen that the analogue of the *dosage* of poison is the *ability* of a group of subjects, and the analogue of the *poison* itself is the *item* under investigation" (1944, p.31). Finney criticized Lawley's solution to the estimation problem.

According to Baker (1965), Tucker (1946) introduced the term *item characteristic curve* in his article *Maximum validity of a test with equivalent items*. However, while, in this article, Tucker certainly works with such curves, no mention of the term itself is made. Lord's 1952 Psychometrika monograph is often cited as the birth-parent of item response theory proper. While this monograph is likely the first systematic treatment of the topic, Lord himself (wrongly) attributes the use of the normal ogive curve to model test data to Tucker (1946). Lord also claims that Tucker (1946) was the source of the term *item characteristic curve*. Lazarsfeld (1950) was the source of the analogous concept *trace line*. The invention of this concept is evident in the following: "This fact can be formulated more precisely in the following way: We assume that the probability *y*, that respondents check the first alternative in the *i*th question, is a function $f_i(x)$ of their position the continuum *x*... The graphical picture of our functions $f_i(x)$ we shall call the *trace line* of item *i*" (p.364); "The problem is how one can, with the help of these accounting equations, learn the nature of the trace lines and the population distribution after an empirical test has provided the frequencies of all possible response patterns which are made by a sample of people" (p.371); "It turns out that *the appropriate degree for the hypothetical trace lines can be derived from the manifest material*" (p.372).

Birnbaum, in Lord and Novick (1968), provides a formal treatment of the two-parameter item response model which includes a discussion of the estimation of its parameters by "unconditional maximum likelihood." In their article *Fitting a response model for n dichotomously scores items*, Bock and Lieberman (1970) provided a means for the efficient estimation of the parameters of the two-parameter item response model through the use of the maximum likelihood principle. Christoffersson (1975) employed weighted least squares to estimate the parameters of the two-parameter model, and Muthen (1978) employed weighted least squares in a method which linked the older method of the linear factor analysis of tetrachoric correlations to modern item response modeling. Bock and Aitkin (1981) provided a major advancement with their marginal maximum likelihood estimation by the EM algorithm.

LISREL

Through his invention of path analysis, the biometrician Sewall Wright was a forefather of structural equation modeling. His 1918 article *On the nature of size factors*, which appeared in the journal *Genetics*, was an analysis of the size components of bone measurements, and, according to Bollen (1989, pp.5-6), contained an independent derivation of factor analysis. The relevant material is the following: "Let X and Y be two characters whose variations are determined in part by certain causes A, B, C, etc., which act on both and in part by causes which apply to only one or the other, M and N respectively. These causes are assumed to be independent of each other...The extent to which a cause determines the variation in an effect is measured by the proportion of the squared standard deviation of the latter for which it responsible" (Wright, 1918, p.370). Unobserved variates also appeared in some of his other statistical inventions. The origins of structural equation modeling can be seen in his 1921 *J. of Agricultural Research* paper. In discussing the simple correlation coefficient he states: "But it would often be desirable to use a method of analysis by which the knowledge that we have in regard to causal relations may be combined with the knowledge of the degree of relationship furnished by the coefficients of correlation" (1921, p.559). Wright's essential contributions were to invent the path diagram which pictorially represents hypothesized causal relations amongst a set of of variates, and derive a systematic approach to the conversion of such diagrams into claims about the variances and covariances of the variates. On page 560 of the 1921 paper, Wright presents an early path diagram for the variates "weight at birth", "weight at 33 days", and "size of litter", of guinea pigs.

Frisch (1934) analyzed certain problems in the mathematical modeling of errors of measurement and disturbances in equations. Blalock (1963) argued that sociologists should use causal models containing both indicators and underlying variates to make inferences about causal relations among the underlying variates on the basis of the covariances of the observed indicators: "...the variables actually measured may not be the ones in which we are primarily interested. For example, the variables being correlated may be demographic factors such as race, sex, income...Causal thinking, however, may center around such variables as exposure to discrimination, social status, psychological maturity, or anomie. The measured variables may be considered theoretically unimportant in their own right. They are merely taken as indicators of other underlying variables" (p.54). This is an early version of an idea that will become commonplace within latent variable modeling, namely, that a concept's "abstractness" implies its underlyingness/unobservability, and that the scientist can only come to know such unobservables via their relations to observables. The causal analysis angle is stated clearly: "H.A. Simon has introduced a method for making causal inferences from correlational data, provided we are willing to make certain assumptions about outside variables that might possibly operate to disturb the pattern of intercorrelations (p.53).

Costner (1969) studied multiple indicators of correlated unobservable variables and Land (1970) advised that there will be "two general cases for which sociologists will be interested in utilizing unmeasured variables in a causal model...The first of these arises in the study of measurement error...The common characteristic of all of these applications of path analysis is that the hypothetical (unmeasured) variables enter the path models only as causes of the observed variables. A second case in which sociologists will utilize unmeasured variables is as variables which intervene between measured variables in a causal model" (p.507). Hence, Costner indicates the existence of what are called, herein, the measurement and causality pictures of the Central Account.

Hauser and Goldberger (1971) provided a comprehensive treatment of the estimation problem in linear equation systems with unobservable variates, and, as did Lord and Novick (1968), made reference to the notion of variables that cannot be "directly measured." They state that "In this chapter our purpose will be to draw attention to the recurrent problem of estimating the parameters of overidentified path models which feature unobservable variables. Such variables have not been, or perhaps cannot be, measured directly, and the structural coefficients pertaining to their causes and effects must be inferred from the available measurements and the postulated causal structure" (Hauser & Goldberger, 1971, p.82). Joreskog (1973), Keesling (1972), and Wiley (1973) proposed the general model that would come to be known as LISREL. Joreskog (1973), for example, states that "We shall describe a general method for estimating the unknown coefficients in a set of linear structural equations. In its most general form the method will allow for both errors in equations (residuals, disturbances) and errors in variables (errors of measurement, observational errors) and will yield estimates of the disturbance variance-covariance matrix and the measurement-error variances as well as estimates of the unknown coefficients in the structural equations, provided that all these parameters are identified" (p.85). Wiley (1973) introduces his 1973 effort as follows: "This paper has four goals. The first is to present a general framework for the consideration of models with error in the measured variables and/or constructs which were not measured directly" (p.69). He explains that "'Unmeasured' is used here to indicate that there is no operational variable which was intended to directly measure a construct which is symbolized in the model. The model presented is a version of that discussed by Hauser & Goldberger (1971)" (Wiley, 1973, p.77).

Nonlinear factor models

As early as 1938, Thurstone (1938, p.89) had discussed the need for non-linear factor methods. Lazarsfeld (1950) considered polynomial trace lines, making him perhaps the first to consider non-linear factor analysis in the sense of McDonald (1967). McDonald (1967) provided a comprehensive treatment of the topic and derived both piece-meal and efficient methods for the estimation of the parameters of various non-linear factor models.

Latent class and profile models

Lazarsfeld's (1950) invention of latent structure analysis was the origin of latent class and profile analysis. Goodman (1974) developed efficient ways to estimate the parameters of these models, and Haberman (1974) explained the relationship between latent class and loglinear analysis. In the general statistical literature, latent profile analysis is referred to as finite mixture modeling (see, e.g., McLachlan & Peel, 2000), and many technical innovations have come from this area of research. In his taxometrics approach to latent class and profile analysis, Meehl (1965; Waller & Meehl, 1998) provided a forceful statement of the CAC.

Origins of the manifest/latent variate distinction

The latent variates appearing in the equations of a latent variable model have gone by many names, including *g*, *functional unities*, *factors*, *hypothetical variates*, *unobservable variates*, to name but a few. The general mathematical form of the latent variable model is now well understood, and the "observable" and "unobservable" variates of the model are almost uniformly called *manifest* and *latent* variates, respectively. Early references to this dichotomy are found in Reiersol (1950) and Lazarsfeld (1950). Reiersol's background was in econometrics, while Lazarsfeld was a sociologist, and the inventor of latent structure analysis. In his 1950 paper, Lazarsfeld comprehensively details the reasoning behind latent structure analysis in particular, and latent variable modeling in general: "The investigator assumes that a one-dimensional continuum exists, such as soldier morale, anti-Semitism or intelligence" (p.363); "The purpose of giving a test is obviously the following one: From the distribution of "answers" to the items we can, it is hoped, make inferences as to the position of the people on the assumed continuum" (p.363).

The important passage with regard the origins of the manifest/latent distinction is the following: "Because the "response patterns" of the people are obtained from actual experiments, we shall call them *manifest data*; all the information inferred as to the nature of the continuum or the position of people thereon we shall call *latent*. Nothing more is implied in this terminology than the distinction between information obtained by direct observation and information obtained by using additional assumptions and/or by making inferences from the original data" (pp.363-364). The tone of this comment and the fact that the employment of these terms is unreferenced suggests that Lazarsfeld believes that he is contributing original terminology. Furthermore, in J. Royce's 1958 article *The Development of Factor Analysis*, latent structure analysis is defined in a glossary, the definition being: "A new analytical method developed by the sociologist Lazarsfeld which is concerned with the identification of "latent" or "hidden" factors..." (p.159). The use of quotations seems to imply novelty in regard the term "latent."

Origins of conditional independence as a fundamental concept

A number of mid-century papers identified the principle of *local* or *conditional independence* as a building block of latent variable models. The discussions of latent structure analysis by Lazarsfeld (1950, 1959) was certainly important in this regard. In his 1950 account, he states: "We shall now call a *pure test* of a continuum *x* an aggregate of items which has the following properties: *All interrelationships between the items should be accounted for by the way in which each item alone is related to the latent continuum*" (p.367); "The crucial point is to obtain a precise formulation for the idea that the interrelationship between items should be completely explained by the existence of one underlying continuum. Scrutiny of research practice suggests that this be formulated as a negative statement. If a group of people has the same *x*-value...then, with [it] held constant, nothing else "holds the two questions together" (pp.367-368); "In terms of trace lines, for an x=x_c, the probability of a joint positive answer $f_{12}(x)$ is the product of the independent probabilities for the two items under discussion..." (p.368); "The term *accounting equations* [expresses] the postulate that the joint occurences and therefore the interrelationships between the test items are accounted for by the existence of one underlying continuum" (p.370).

McDonald (1981, p.115) claims that Anderson (1959) is the origin of the general principle of local independence, but, given Lazarsfeld (1950,1959), this does not seem correct. In an early analysis of the latent class model, McDonald (1962) himself provides the general representation (2.1) along with (2.2), as does Meredith (1965), who cites McDonald (1962). Meredith discusses models in which it is "...assumed that an individual will "endorse" or "pass" a dichotomous item can be represented by some function of an underlying, "latent" attribute (possibly multivariate) called the item trace line or item characteristic curve" (1965, p.419). According to Meredith, "All of these models for test behavior are unified by two basic assumptions. One is the stochastic assumption of a probability of passing or endorsing an item...the other assumption common to this class of models is the assumption of *local independence*" (p.419).

7. *Summary*

Latent variable modeling is a complex practice, and, of course, its historical and mathematical development is more complicated than is, herein, described. For example, at any given time, the thinking of applied modellers and psychometricians has been shaped in part by the dominant philosophical orientations of the social and behavioural sciences. As will be later discussed, empirical realist philosophy, and the version of this philosophy espoused in the construct validation program of test validation of Cronbach and Meehl (e.g., 1955), has played a powerful role in the growth of the Central Account. In particular, it has legitimized those theses of the CA pertaining to unobservability. The Central Account, a description of which is provided in the next chapter, is the picture that binds together the various mathematical, philosophical, and applied facets of latent variable modeling.