Vij, A. and **J.L. Walker**, "Hybrid Choice Models: The Identification Problem," in (S. Hess and A. Daly, eds.) *Handbook of Choice Modeling*, Edward Elgar, 2014, pp. 519-564.

Hybrid Choice Models: The Identification Problem

Final version submitted (April 27, 2013)

Akshay Vij
University of California at Berkeley
116 McLaughlin Hall
Berkeley, CA 94720-1720
a.vij@berkeley.edu

Joan L. Walker
University of California at Berkeley
111 McLaughlin Hall
Berkeley, CA 94720-1720
joanwalker@berkeley.edu

1. Introduction

Conventional microeconomic theory has tended to regard individual consumers as rational self-interested actors engaged in a constant process of evaluating the costs and benefits associated with any decision in the marketplace as they strive to maximize their personal well-being. The random utility maximization model has been the model of choice for studies on consumer behavior over the last several decades. Random utility maximization, or discrete choice, models examine potential outcomes from among a set of mutually exclusive alternatives, and have found wide application in fields as diverse as travel demand analysis, marketing, education, labor force participation, etc. Early applications almost exclusively used some model form belonging to the Generalized Extreme Value (GEV) family of models, owing largely to the computational tractability offered by these models. The multinomial logit and nested logit models proved by far to be the most popular (Carrasco and Ortúzar, 2002), earning their colloquial appellation of the workhorses of discrete choice analysis.

Numerous studies have since devoted attention towards improving the specification of the logit model. Extensions include the incorporation of flexible error structures and random taste heterogeneity through the use of either the mixed logit or the multinomial probit model; the inclusion of latent variables representing latent biological, psychological and sociological constructs underlying the formation of individual preferences, such as attitudes, values, norms and affects; the introduction of latent classes to capture latent segments that differ from each other with regards to, for example, the taste parameters; the combination of stated and revealed preference data to capitalize on the benefits offered by either type of data; and the representation of individual decision-making behavior in a dynamic context to capture interdependencies between decisions made at different stages in time. The Hybrid Choice Model (HCM) combines these and other more recent developments in the choice modeling literature under a single unified framework, leading to a statistically more robust and behaviorally richer model of decision-making that obviates many of the limitations of simpler representations.

However, the increased complexity afforded by the HCM raises important questions of identification that remain inadequately addressed in the literature. Any HCM will in general be specified according to some theory of individual behavior. Observable data may then be used to verify the hypothesized theory underlying the model specification. There are two facets to the identification problem: theoretical and empirical. A model specification is said to be theoretically identifiable if no two distinct sets of parameter values generate the same probability distribution of observable data. In most cases, unless restrictions are imposed multiple sets of parameter estimates may generate the same probability distribution for the data. Therefore, the identification problem consists of determining the set of restrictions required to obtain a unique vector of consistent parameter estimates.

The theoretical identification problem as it applies to the family of discrete choice models without latent variables, such as the multinomial logit model, the multinomial probit model and the mixed logit model, has received widespread attention in the literature (see, for example, Ben-Akiva and Lerman, 1985; Walker et al., 2007; and Train, 2009). The theoretical identification problem as it applies to the family of structural equation models with latent variables, such as the confirmatory factor analytic model and the path analytic model, is equally well understood (see, for example, Bollen, 1989). In this chapter, we break apart the HCM into a discrete choice model where the latent variables are treated as observable variables, and a structural equation model with latent variables, assembling the rules of identification that have been developed independently for each of these two constituent pieces elsewhere in the literature, and deriving some of our own for specific cases that haven't been addressed before, into a set of sufficient but not necessary

conditions of theoretical identification for the HCM as a whole. However, in so doing our framework overlooks correlation between the two constituent pieces due to the presence of latent variables in both (Daziano and Bolduc, 2012). It is precisely this correlation that results in situations where the rules of theoretical identification presented in this chapter are sufficient but not necessary. Nevertheless, findings from this chapter represent an important first step in addressing the identification problem as it applies to HCMs, setting the stage for future breakthroughs in this important area of research.

Theoretical identifiability is predicated on the availability of an infinite number of observations. In reality, the analyst will have a finite sample of observations at her disposal that may or may not contain enough variability to support the estimation of a particular model specification. A model is said to be empirically unidentified or underidentified if the model is theoretically identified but cannot be estimated using a sample dataset. The term empirical underidentification was originally introduced by Kenny (1979) in the context of structural equation models. Reasons for empirical unidentification or underidentification may include small sample size, multicollinearity between observable variables, model misspecification, etc.

The objective of this chapter is to provide a general framework for the theoretical and empirical identification of HCMs. We focus on two components of the GEV model that have found immense popularity with recent studies employing discrete choice analysis: the mixed logit model and the choice and latent variable model. That being said, the framework presented in this chapter can be readily broadened to help identify HCMs that employ multinomial probit, latent classes, multiple datasets, etc. The chapter is organized as follows: Section 2 introduces a specific version of the HCM that combines the mixed logit model with the choice and latent variable model. Section 3 examines theoretical identification for each component of the HCM introduced in Section 2, developing a set of sufficient but not necessary conditions for the identification of the model as a whole. Section 3 is accompanied by three appendices at the end of the chapter that illustrate how these conditions may be applied to different model specifications. Section 4 elaborates on the common sources of empirical underidentification. Section 5 discusses estimation tools that may be used to verify theoretical and empirical identification. Section 6 uses a case study to demonstrate the kinds of identification issues that might arise in practice. Section 7 concludes with a discussion of the limitations of the framework presented in this chapter and potential directions for future research.

2. The Hybrid Choice Model

The HCM takes as its kernel the random utility maximization model, adding extensions wherever necessary to relax some of the more limiting assumptions of the kernel. In introducing the different components of the HCM, we begin by summarizing a special case of the HCM as outlined by Walker and Ben-Akiva (2002). Section 2.1 presents the framework of the random utility maximization model. Section 2.2 builds on the framework through the inclusion of more flexible error structures that allow for unrestricted substitution patterns and serial correlation, and random taste heterogeneity to capture unobservable variation in sensitivity to alternative attributes and individual characteristics. Section 2.3 incorporates the influence of latent variables and psychometric data that capture the effects of more abstract psychological constructs on observable behavior.

Walker and Ben-Akiva (2002) take the HCM further through the inclusion of latent classes and the combination of stated and revealed preference data. We desist from including these extensions in our analysis because the set of sufficient conditions for theoretical identification developed

later in Section 3 can be easily broadened to cover these more general cases. Instead, Section 2.4 extends the framework in a different direction that allows for causal relationships between multiple explanatory variables. Causal relationships are commonplace with studies using structural equation models, and are steadily gaining in popularity with studies employing HCMs as well (see, for example, Temme et al., 2008; Zhao, 2009; Tudela et al., 2011). Furthermore, the identification conditions for models with causal relationships between explanatory variables are distinct and deserving of explicit treatment. For these reasons, we include them in our representation of the HCM.

2.1 The Random Utility Maximization Kernel

Consider a decision-maker n $(n = 1, \dots, N)$, faced with a set of mutually exclusive alternatives j $(j = 1, \dots, J)$, where we have assumed, for the sake of notational convenience, that the number of alternatives faced by all individuals is the same. The random utility maximization model states that the chosen alternative is that which provides the greatest utility, and the model is mathematically formulated as:

$$y_{nj} = \begin{cases} 1 & \text{if } u_{nj} \geq u_{nj'} \text{ for } j' = 1, \dots, J \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$u_{nj} = \mathbf{x}'_{\mathbf{nj}}\boldsymbol{\beta} + \varepsilon_{nj} \quad (2)$$

, where y_{nj} is an indicator of the observed choice and u_{nj} is the perceived utility of alternative j for individual n, $\mathbf{x_{nj}}$ is an $(L \times 1)$ vector of all explanatory variables, $\boldsymbol{\beta}$ is an $(L \times 1)$ vector of parameters, and ε_{nj} is the stochastic component. The set of explanatory variables $\mathbf{x_{nj}}$ may represent both characteristics of the decision-maker and attributes of the alternative. Equation (1) is the measurement equation of the choice model, and it links the unobservable utilities u_{nj} to the observed choice indicators y_{nj}. Equation (2) is the structural equation of the choice model, and it links the explanatory variables $\mathbf{x_{nj}}$ to the unobservable utilities u_{nj}.

For random utility maximization models the specification of the absolute levels of utilities is irrelevant; only their differences matter. Nonetheless, we specify the model in level form (that is, $u_{nj}, j = 1, \dots, J$) rather than in difference form (for example, $(u_{nj} - u_{nJ}), j = 1, \dots, (J-1)$). While

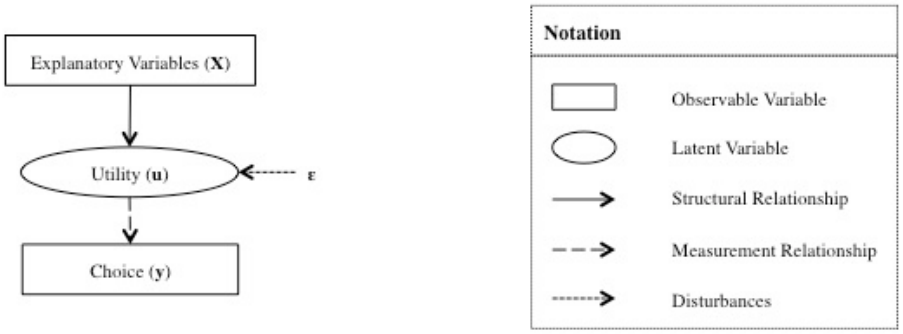

Figure 1: Random utility maximization model framework (Walker and Ben-Akiva, 2002)

working with the difference form would greatly simplify the identification and normalization problem, behaviorally it is often more meaningful to specify and estimate the models at the levels, or structural, form. Employing a more compact vector form, we get:

$$\mathbf{y_n} = [y_{n1}, \dots, y_{nJ}]' \tag{3}$$

$$\mathbf{u_n} = \mathbf{X_n'\beta} + \boldsymbol{\varepsilon_n} \tag{4}$$

, where $\mathbf{y_n}$, $\mathbf{u_n}$ and $\boldsymbol{\varepsilon_n}$ are $(J\times1)$ vectors and $\mathbf{X_n}$ is an $(L\times J)$ matrix. The random utility maximization model forms the kernel of the HCM. In selecting a particular form for the kernel, one of the things to be kept in mind is that it should be computationally tenable. The GEV model has a closed form solution that renders it a natural choice for the kernel. Within the GEV family of models the analyst could choose multinomial logit, the most basic of model forms, and introduce further complexity through the addition of mixture distributions. Alternatively, the analyst could start with a more complex form, such as the nested or cross-nested logit model, as the kernel. From the standpoint of estimation, it isn't always clear which is preferable (Walker and Ben-Akiva, 2002).

The procedure for establishing identification is independent of the kernel model form, and may be used to identify kernel model forms belonging to the GEV family, such as multinomial logit and nested logit, and kernel model forms outside the GEV family, such as multinomial probit. We shall assume throughout the chapter that multinomial logit forms the kernel of the HCM, and the identification conditions will be derived for this special case. It is left to the reader to derive the analogous conditions of identification for other kernel model forms.

2.2 Flexible Disturbances

The GEV family of models, though computationally tractable, is deficient in other ways due largely to the rigidity of the error structure. On the other hand, random utility maximization models such as the multinomial probit that offer more flexible error structures can be computationally burdensome. The specific case of the HCM considered in this chapter combines the advantages of both model forms. The random utility term $\boldsymbol{\varepsilon_n}$ is made up of two components: a probit-like variable with a multivariate distribution, and an i.i.d. Extreme Value random variable corresponding to the logit kernel. The probit-like term allows for a rich covariance structure and the extreme-value term aids computation (Walker and Ben-Akiva, 2002). We specify the random utility term $\boldsymbol{\varepsilon_n}$ using the factor-analytic structure shown below:

$$\boldsymbol{\varepsilon_n} = \mathbf{F_n}\boldsymbol{\xi_n} + \boldsymbol{\nu_n} \tag{5}$$

, where $\boldsymbol{\xi_n}$ is an $(R\times1)$ vector of R multivariate latent factors, $\mathbf{F_n}$ is a $(J\times R)$ matrix of factor loadings that map the factors to the error structure, and $\boldsymbol{\nu_n}$ is a $(J\times1)$ vector of i.i.d. Extreme Value random variables with mean zero and variance g/μ^2, where μ is the scale and g is the variance of a standard Extreme Value random variable (g equals $\pi^2/6$). For estimation purposes, it is desirable to specify the factors as independent, leading us to decompose $\boldsymbol{\xi_n}$ as follows:

$$\boldsymbol{\xi_n} = \boldsymbol{\Upsilon}\boldsymbol{\eta_n^R} \tag{6}$$

, where $\boldsymbol{\eta_n^R}$ is an $(R\times1)$ vector of independent factors with mean zero and variance one, $\boldsymbol{\Upsilon\Upsilon'}$ is the covariance matrix of $\boldsymbol{\xi_n}$, and $\boldsymbol{\Upsilon}$ is an $(R\times R)$ lower triangular matrix that is the Cholesky

factorization of the covariance matrix. Equations (5) and (6) may be combined to get the following factor-analytic form for the error term:

$$\boldsymbol{\varepsilon_n} = \mathbf{F_n}\boldsymbol{\Upsilon}\boldsymbol{\eta}_\mathbf{n}^\mathbf{R} + \boldsymbol{\nu_n} \tag{7}$$

In principal, the distribution of $\boldsymbol{\varepsilon_n}$ is associated with a (J×J) covariance matrix, for which level and scale corrections eventually result in a $(J-1)\times(J-1)$ covariance matrix requiring an additional normalization in terms of scale. The factor-analytic structure was first proposed by McFadden (1984) in the context of multinomial probit models to ease estimation. In the case of the multinomial probit model, the error term $\boldsymbol{\nu_n}$ does not enter equation (7) and the error structure is wholly captured by the factor analytic term $\mathbf{F_n}\boldsymbol{\Upsilon}\boldsymbol{\eta}_\mathbf{n}^\mathbf{R}$. The elements of $\mathbf{F_n}$ are specified by the analyst according to some prior hypotheses about the covariance structure of the sample, whereas the elements of $\boldsymbol{\Upsilon}$ are parameters to be estimated. The only limitation of the factor-analytic structure is that the utility specification should be additively separable into the systematic and the stochastic component, such that the systematic component comprises the expectation of the random utility and the elements of $\boldsymbol{\eta}_\mathbf{n}^\mathbf{R}$ are distributed with mean zero. These restrictions rule out the use of one-sided distributions such as the lognormal or the triangular for elements of $\boldsymbol{\eta}_\mathbf{n}^\mathbf{R}$. For heteroskedastic, nested and cross-nested covariance structures this is rarely a problem, and the factor-analytic specification can suitably capture these covariance structures with relatively few parameters. Bunch (1991) presents a set of comprehensive guidelines for the identification of multinomial probit models, and Walker et al. (2007) adapt these guidelines to the case of the mixed logit model with factor-analytic specifications.

However, one-sided distributions are commonly used when imposing random parameters on alternative attributes and individual characteristics that are expected to have either a positive or a negative effect on the decision-making process, but not both. For models that do employ one-sided distributions, the conditions of identification developed in this chapter cannot be used directly. In such cases the reader may apply the conditions to an analogous model with two-sided distributions that can be represented using a factor-analytic specification as a means of gathering information regarding the identification status of the original model specification. Additionally, the reader may use the estimation methods described in greater detail in Section 5.

Substituting equation (7) into equation (4), we get:

$$\mathbf{u_n} = \mathbf{X_n'}\boldsymbol{\beta} + \mathbf{F_n}\boldsymbol{\Upsilon}\boldsymbol{\eta}_\mathbf{n}^\mathbf{R} + \boldsymbol{\nu_n} \tag{8}$$

The variables $\boldsymbol{\nu_n}$ and $\boldsymbol{\eta}_\mathbf{n}^\mathbf{R}$ are unknown, whereas the variable $\mathbf{X_n}$ is known. The unknown parameters are μ (the scale of $\boldsymbol{\nu_n}$) and the elements of $\boldsymbol{\beta}$ and $\boldsymbol{\Upsilon}$. Though the matrix $\mathbf{F_n}$ may include unknown parameters, these cases are rare in the literature, and we will assume for the remainder of this chapter that $\mathbf{F_n}$ is known. Furthermore, we will be retaining the Extreme Value scale term μ instead of normalizing it to 1. When one arbitrarily sets the scale of one of the elements of $\boldsymbol{\Upsilon}$, the scale of the model (that is, the μ) changes, and this change is reflected in the scale of the estimated parameters in $\boldsymbol{\beta}$. Therefore, it is necessary to retain the μ to interpret the impact of the normalization of $\boldsymbol{\Upsilon}$ on the remaining parameter estimates.

2.3 Latent Variables

Analysts are often interested in the influence exerted by biological, psychological and sociological factors such as attitudes, norms, perceptions, affects, beliefs, etc. on observable

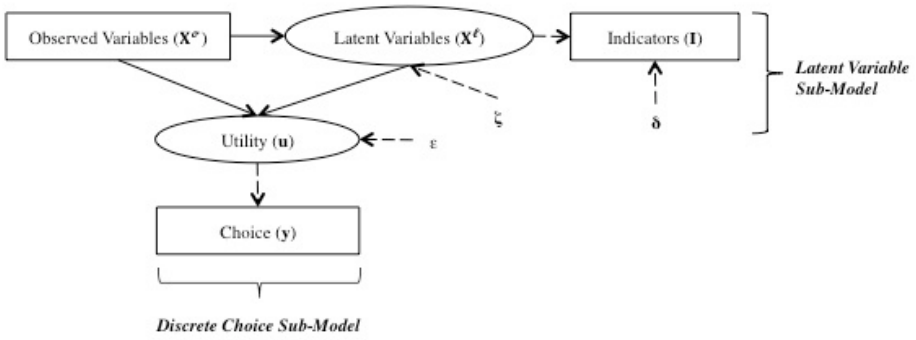

Figure 2: Integrated choice and latent variable model framework
(Adapted from Ben-Akiva et al., 2002)

individual behavior. Unfortunately, most of these constructs are not well defined and cannot be directly measured, and are therefore referred to as latent variables. Just as utility as a latent construct is operationalized with the help of observable choices, the latent variables are operationalized with the help of indicators that most often consist of individual responses to survey questions regarding, for example, the level of agreement with attitudinal statements or satisfaction with alternative attributes. Though the latent variable itself is not observed, its effect on observable variables, or indicators, can be measured, and the nature of the relationship can provide information about the underlying latent variable (Ben-Akiva et al., 2002).

Figure 2 extends the framework of the random utility maximization model to include the additional effect of latent variables. We introduce the superscripts o and ℓ to denote observable and latent variables, respectively. As labeled in the figure, the model comprises two components: the discrete choice sub-model and the latent variable sub-model. The discrete choice sub-model comprises the logit kernel with flexible disturbances built on top of it. The latent variable sub-model maps the indicators onto the latent variables. Indicator responses vary across individuals and, depending upon the latent variable of interest, could also vary across alternatives (and observations, when working with panel data). For the sake of notational convenience we assume that the indicators vary across individuals and alternatives, as in the case of perceptions regarding alternative attributes. In the simplest case, a linear model is appropriate for describing the mapping of the indicators onto the latent variables, leading to the following equation for the measurement model:

$$\mathbf{i_{nj}} = \mathbf{\Lambda x_{nj}^{\ell}} + \mathbf{\delta_{nj}} \tag{9}$$

, where $\mathbf{i_{nj}}$ is a $(Q\times 1)$ vector of observed indicators, $\mathbf{x_{nj}^{\ell}}$ is the $(L^{\ell}\times 1)$ matrix of latent variables, $\mathbf{\delta_{nj}}$ is a $(Q\times 1)$ vector of measurement errors, and $\mathbf{\Lambda}$ is an $(Q\times L^{\ell})$ matrix of coefficients relating the indicators to the latent variables. HCM studies in the literature usually assume that indicator responses are uncorrelated with each other. However, this need not always be the case. For instance, when measuring individual perceptions regarding a specific attribute of each of the alternatives in the choice set, indicator responses for different alternatives for the same individual might be correlated. Similarly, when working with panel data, individual responses to the same indicator might be serially correlated across time. To capture these and other potential sources of correlation, we employ the following factor-analytic representation for the measurement errors:

$$\boldsymbol{\delta_{nj}} = \mathbf{D\Theta\eta_{nj}^{C}} \tag{10}$$

, where $\boldsymbol{\eta_{nj}^{C}}$ is a $(C\times1)$ vector of C independent random factors with mean zero and variance one, $\boldsymbol{\Theta\Theta'}$ is the covariance matrix of $\boldsymbol{\Theta\eta_{nj}^{C}}$ such that $\boldsymbol{\Theta}$ is the $(C\times C)$ lower triangular matrix that is its Cholesky factorization, and $\mathbf{D}$ is a $(Q\times C)$ matrix of factor loadings that map the factors to the error structure. Owing to the Central Limit Theorem, the elements of $\boldsymbol{\eta_{nj}^{C}}$ are usually assumed to be standard normal. However, if the analyst has strong a priori reasons for believing that some other distributional form, such as Extreme Value or the Laplace, might be a better fit then that too can be employed.

2.4 Causal Relationships

The framework shown in Figure 2 can be generalized further to include causal relationships within and across observable and latent variables. The literature on HCMs is dominated by studies employing some variation of the Multiple Indicator Multiple Cause (MIMIC) model where each latent variable is measured by multiple indicator responses and explained by multiple observable variables. The use of more complex causal relationships between explanatory variables in HCMs has been limited thus far, with only a handful of studies looking beyond the MIMIC model for a representation of individual behavior (see, for example, Temme et al., 2008; Zhao, 2009; Tudela et al., 2011). However, these methods are employed widely by studies in psychology and the social sciences to test theoretical relationships between multiple variables, observable or latent. They are powerful tools for analyzing the mediating influence of intervening constructs on observable individual behavior, and are a natural extension to the MIMIC models currently being used by most studies.

Figure 3 shows the generalized framework of the HCM with causal relationships between the variables. It is helpful to differentiate between exogenous and endogenous explanatory variables. A variable whose value is determined by the states of other variables in the model is an endogenous variable, superscripted d, whereas a variable whose value is independent

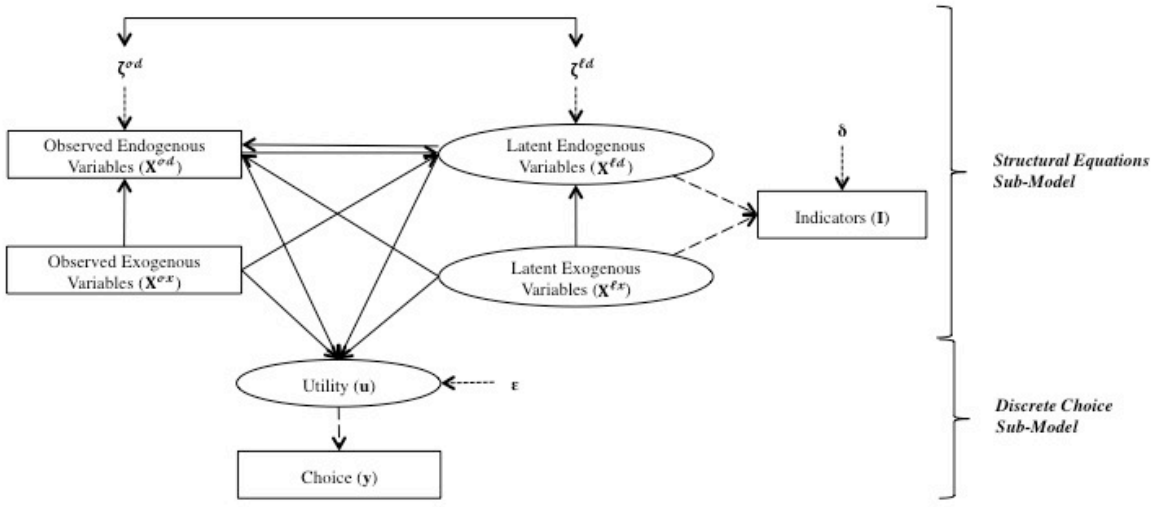

Figure 3: HCM framework
(The bi-directional arrow denotes correlation)

of the states of other variables is an exogenous variable, superscripted $\boldsymbol{x}$. Both latent and observable variables can be either exogenous or endogenous, resulting in the four way stratification of explanatory variables shown in Figure 3.

As can further be seen from Figure 3, two components to the HCM can be distinguished: a discrete choice sub-model and a structural equations sub-model. The discrete choice sub-model is the same as before: equations (1) and (2) are still the measurement and structural components of the sub-model, respectively. For the structural equation sub-model, equations (9) and (10) which correspond to the latent variable sub-model from Figure 2 form the measurement component of the structural equations sub-model. The structural component for the sub-model corresponds to the causal relationships between the explanatory variables themselves, the equations for which may be written as follows:

$$\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}} = \mathbf{B}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}} + \boldsymbol{\Gamma}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}} + \boldsymbol{\zeta}_{\mathbf{nj}} \tag{11}$$

, where $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}$ is the $(\mathrm{L}^{\boldsymbol{d}}\times 1)$ vector of endogenous variables, $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}$ is the $(\mathrm{L}^{\boldsymbol{x}}\times 1)$ vector of exogenous variables, $\boldsymbol{\zeta}_{\mathbf{nj}}$ is an $(\mathrm{L}^{\boldsymbol{d}}\times 1)$ vector of random errors, $\mathbf{B}$ is an $(\mathrm{L}^{\boldsymbol{d}}\times \mathrm{L}^{\boldsymbol{d}})$ matrix of coefficients for the endogenous variables, and $\boldsymbol{\Gamma}$ is an $(\mathrm{L}^{\boldsymbol{d}}\times \mathrm{L}^{\boldsymbol{x}})$ matrix of coefficients for the exogenous variables.

Among the studies mentioned in a previous paragraph that have incorporated causal relationships within the HCM framework, all have assumed that the endogenous variables are uncorrelated. However, studies in psychology and the social sciences that use structural equation models routinely introduce correlation between endogenous variables to help distinguish correlation from causation. We propose to do the same within the HCM framework through the use of the following factor-analytic representation for the covariance structure of the measurement errors:

$$\boldsymbol{\zeta}_{\mathbf{nj}} = \mathbf{G}\boldsymbol{\Psi}\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}} \tag{12}$$

, where $\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}}$ is an $(\mathrm{S}\times 1)$ vector of S independent random factors with mean zero and variance one, $\boldsymbol{\Psi}\boldsymbol{\Psi}'$ is the covariance matrix of $\boldsymbol{\Psi}\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}}$ such that $\boldsymbol{\Psi}$ is the $(\mathrm{S}\times\mathrm{S})$ lower triangular matrix that is its Cholesky factorization, and $\mathbf{G}$ is an $(\mathrm{L}^{\boldsymbol{d}}\times\mathrm{S})$ matrix of factor loadings that map the factors to the error structure. As was the case in Section 2.3, the elements of $\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}}$ are usually assumed to be standard normal, but if the analyst so desires other distribution forms may also be used.

3. Theoretical Identification

A model is theoretically identifiable if it is possible to infer the true underlying parameter values given an infinitely large number of observations. Identifiability precludes observational equivalence - if a model is identifiable then no two sets of parameter values result in the same probability distribution of observable variables. As with any complex econometric model, identification is an issue with HCM. Though the identification problem has been explored in substantial detail for special cases, a more general framework remains lacking, largely due to the complexity of HCMs that renders infeasible any monolithic examination of the model form. A more tractable approach is to break apart the model into smaller sub-models that can then be examined independently and more fruitfully. The approach adopted in this chapter consists of breaking the HCM into a discrete choice model where the latent variables are treated as observable variables, and a structural equation model with latent variables. The normalizations

and restrictions that apply to a discrete choice model without latent variables also apply here, as do the identification rules that apply to a traditional structural equations model with latent variables. Therefore, a sufficient but not necessary condition for identification can be obtained by extending the Two Step Rule used for structural equation models with latent variables (Bollen, 1989) to a Three Step Rule for HCMs (Ben-Akiva et al., 2002):

1. Confirm that the measurement component of the structural equations sub-model is identified, reformulating the equations as a confirmatory factor analysis.

2. Confirm that the structural component of the structural equations sub-model is identified, reformulating the equations as a structural equations model with observable variables and treating each latent variables like an observed variable that is perfectly measured.

3. Confirm that the structural component of the discrete choice sub-model is identified, treating each explanatory variable like an exogenous observed variable that is perfectly measured.

For example, consider the HCM shown in Figure 4A. It may be broken apart into three constituent sub-models. The confirmatory factor analytic model (Figure 4B) comprises the two correlated latent variables $\mathbf{X}_1^{\ell d}$ and $\mathbf{X}_2^{\ell x}$ and the five indicators used to measure them. In pulling apart the latent variables from the HCM and reformulating the model as a confirmatory factor analytic model, the structural relationships between the latent variables need to be ignored and additional relationships that capture correlation between each pair of latent variables must be introduced (even if the latent variables aren't structurally related to each other). The structural equations model (Figure 4C) is similar to the corresponding sub-model in Figure 4A, except that the two latent variables $\mathbf{X}_1^{\ell d}$ and $\mathbf{X}_2^{\ell x}$ may now be treated as observable variables and additional relationships need to be introduced that capture correlation between each pair of exogenous variable, as would be the case with a structural equations model with observable variables. Lastly, the discrete choice model (Figure 4D) is relatively straightforward in that all explanatory variables may be treated as correlated exogenous observed variables. If identification can be established for each of the three sub-models, then the Three Step Rule states that the HCM as a whole is identifiable as well.

A limitation to the Three Step Rule is that it provides a set of sufficient but not necessary conditions for theoretical identification. Conditional on the latent variables, the discrete choice sub-model and the measurement component of the structural equations sub-model can be examined independently of each another. The challenge to the researcher then is to ensure that the structural component of the structural equations sub-model can be identified from either or both components. In the Three Step Rule, the structural component of the structural equations sub-model relies solely upon the measurement component of the structural equations sub-model, reformulated as a confirmatory factor analytic model, for identification, and ignores the information available to the analyst about the latent variables through the discrete choice sub-model (Daziano and Bolduc (2012) provide an excellent discussion on the additional insights that can be had in terms of the identification problem by a joint examination of the discrete choice sub-model and the structural equations sub-model). As a consequence, if one or more components of the sub-models resulting from the application of the Three Step Rule fail identification, the HCM may still be identified. In such cases, unless the analyst can verify that the HCM is indeed identified, it might be better to impose appropriate constraints to ensure that each of the three sub-models is identifiable and that the HCM satisfies the Three Step Rule.

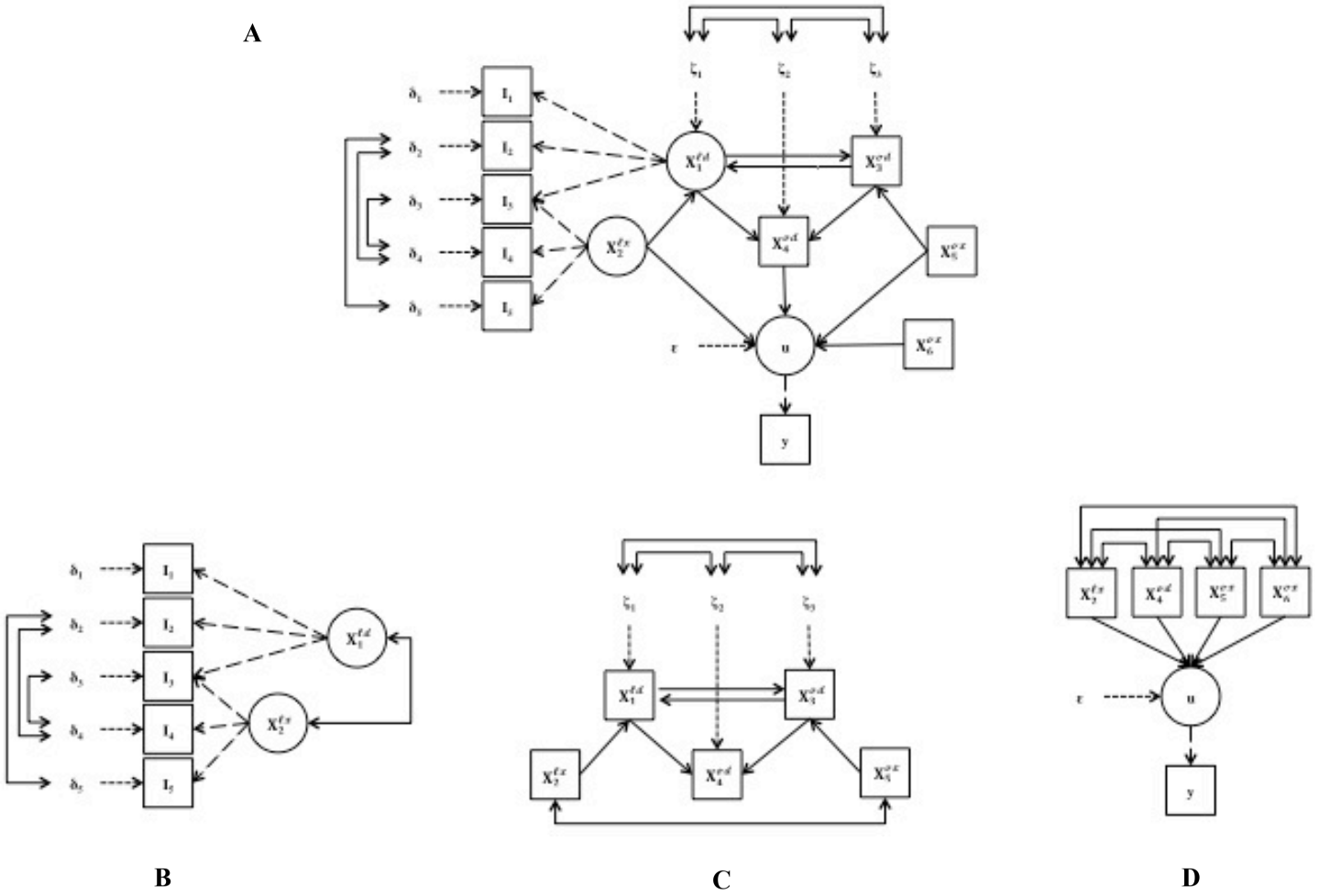

Figure 4: An example of an HCM and how it might be broken apart for the purpose of theoretical identification: (**A**) The HCM; (**B**) Confirmatory factor analytic sub-model; (**C**) Structural equations sub-model with observable variables; and (**D**) Discrete choice sub-model

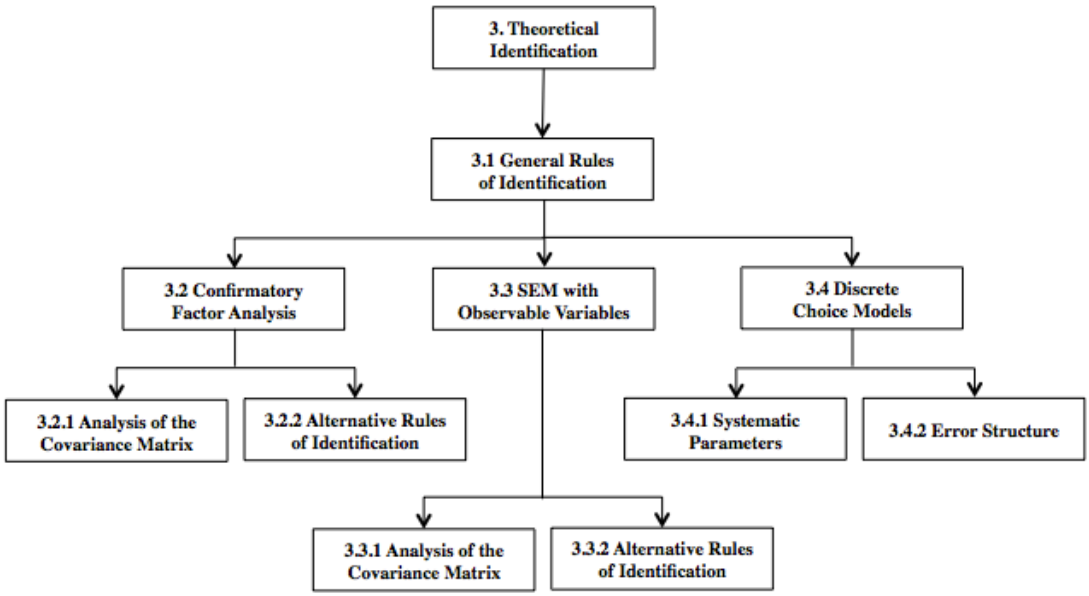

Figure 5: Roadmap to Section 3

The objective of this section is to provide a repository of information regarding theoretical identification of the different pieces that comprise an HCM. In the spirit of structural equation models, Figure 5 shows a roadmap to Section 3 in the form of a path diagram to help the reader negotiate the section. Section 3.1 derives rules of identification for a general econometric model form through the analysis of its covariance structure, and Sections 3.2, 3.3 and 3.4 apply these rules to confirmatory factor analytic models (as in Figure 4B), structural equation models with observable variables (as in Figure 4C), and discrete choice models (as in Figure 4D), respectively. Sections 3.2 and 3.3 also present a selection of alternative rules of identification for confirmatory factor analytic models and structural equation models with observable variables, respectively. Section 3.4 addresses discrete choice models with flexible covariance structures.

Section 3 need not be read in its entirety, and the reader is encouraged to use Figure 5 to take whichever path appears best suited to the problem at hand. For example, for a discrete choice model with random parameters and cross-sectional data the reader may only read Sections 3.1, 3.4, 3.4.1 and 3.4.2. Similarly, for a relatively simple confirmatory factor analytic model the reader might be better served by one of the alternative rules of identification, and should refer first to Sections 3.2 and 3.2.2. If none of the rules apply, the reader may switch to the other branch and refer to Sections 3.1, 3.2 and 3.2.1.

3.1 General Rules of Identification

In this section, we present a framework for identification based on the analysis presented in Bollen (1989) and Walker et al. (2007) that can be applied to any general econometric model of the form:

$$\mathbf{y_n} = \mathbf{\Xi y_n} + \mathbf{\Pi x_n} + \mathbf{R T \eta_n^W} \tag{13}$$

, where $\mathbf{y_n}$ is a $(G\times 1)$ vector of endogenous variables, $\mathbf{x_n}$ is a $(H\times 1)$ vector of exogenous variables, $\mathbf{\Xi}$ is a $(G\times G)$ matrix of coefficients of the endogenous variables, $\mathbf{\Pi}$ is a $(G\times H)$ matrix

of coefficients of the exogenous variables, $\boldsymbol{\eta}_{\mathbf{n}}^{\mathbf{W}}$ is a $(\mathrm{W}\times 1)$ vector of W independent random factors with mean zero and variance one, $\mathbf{TT}'$ is the covariance matrix of $\mathbf{T}\boldsymbol{\eta}_{\mathbf{n}}^{\mathbf{G}}$ such that $\mathbf{T}$ is the $(\mathrm{W}\times\mathrm{W})$ lower triangular matrix that is its Cholesky factorization, and $\mathbf{R}$ is a $(\mathrm{G}\times\mathrm{W})$ matrix of factor loadings that map the factors to the error structure. The variables $\mathbf{y_n}$ and $\mathbf{x_n}$ are observed without any measurement error; the elements of $\mathbf{R}$ are specified by the analyst based on some prior hypothesis about the covariance structure of the dependent variable $\mathbf{y_n}$; and the elements of the matrices $\boldsymbol{\Xi}$, $\boldsymbol{\Pi}$ and $\mathbf{T}$ are parameters to be estimated. The procedure to check that equation (13) is identifiable involves the following steps:

1. Hypothesize the model of interest: Select the endogenous variables $\mathbf{y_n}$ that are relevant to the study, the exogenous variables $\mathbf{x_n}$ that are expected to influence the outcome of $\mathbf{y_n}$, the structure of the coefficient matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Pi}$, and an a priori specification of the covariance structure as denoted by $\mathbf{R}$.

2. Formulate the covariance matrix of the observed variables $\mathbf{y_n}$ and $\mathbf{x_n}$ as a function of the unknown parameters in $\boldsymbol{\Xi}$, $\boldsymbol{\Pi}$ and $\mathbf{T}$. The identification conditions are based on the hypothesis that the covariance matrix of the observed variables is a function of a set of parameters. If the hypothesized model is correct and the analyst knows the parameters in $\boldsymbol{\Xi}$, $\boldsymbol{\Pi}$ and $\mathbf{T}$, then the population covariance matrix can be exactly reproduced such that:

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\theta}) \tag{14}$$

   , where $\boldsymbol{\Omega}$ is the $(\mathrm{G}+\mathrm{H})\times(\mathrm{G}+\mathrm{H})$ sample covariance matrix of the observed variables $\mathbf{y_n}$ and $\mathbf{x_n}$, such that the element ω_{ij} represents the covariance between the i^{th} and j^{th} observable variables in the sample population; the vector $\boldsymbol{\theta}$ is the set of all unknown parameters in $\boldsymbol{\Xi}$, $\boldsymbol{\Pi}$ and $\mathbf{T}$; and $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is the covariance matrix as predicted by the model specification, and is a function of $\boldsymbol{\theta}$. Equation (14) results in a system of simultaneous equations where the left-hand side contains the observables (as calculated from the data) and the right-hand side contains the unknowns (as predicted by the model specification).

3. Apply the order condition, which states that for a covariance matrix $\boldsymbol{\Omega}(\boldsymbol{\theta})$ of dimension $(\mathrm{G}+\mathrm{H})\times(\mathrm{G}+\mathrm{H})$, the number of estimable parameters S should satisfy:

$$\mathrm{S} \leq \frac{(\mathrm{G}+\mathrm{H})(\mathrm{G}+\mathrm{H}+1)}{2} \tag{15}$$

   , where the upper bound is equal to the number of unique elements in the covariance matrix $\boldsymbol{\Omega}(\boldsymbol{\theta})$, or the maximum potential number of independent equations available from equation (14). The order condition is a necessary but insufficient condition of identification, and depending on the hypothesized model structure the number of parameters that can be estimated is often less than that suggested by the order condition. Nonetheless, the order condition does provide for a quick check to avoid major blunders, and there are models that have been published that do not pass this test.

4. Determine whether the system of equations represented by (14) can be solved for all of the unknown parameters in $\boldsymbol{\theta}$. If a parameter in $\boldsymbol{\theta}$ can be written as a function solely of one or more elements of $\boldsymbol{\Omega}$ and none of the elements of $\boldsymbol{\theta}$, that parameter is identified. If all unknown parameters in $\boldsymbol{\theta}$ are identified, then the model as a whole is identified; if not, necessary constraints need to be imposed to ensure identifiability.

A model specification is said to be just-identified if all the parameters are identified and the system of equations represented by (14) results in an equal number of independent equations and unknown parameters. A model specification is said to be overidentified if all the parameters are identified and the system of equations represented by (14) results in more independent equations than unknown parameters. Just-identified models yield a trivially perfect fit and are uninteresting from the standpoint of analysis. Since overidentified models do not always fit the observed data very well, when one does the analyst may take that to mean that the model is a reasonable representation of the behavior under study. A model is said to be underidentified if at least one of the model parameters cannot be identified. In general, the identified parameters in an underidentified model can be consistently estimated.

5. When the conclusion from steps 3 and 4 is that further identifying restrictions are required, the equality condition is used to determine the set of acceptable normalizations. The equality condition states that any normalization must satisfy:

$$\boldsymbol{\Omega}(\boldsymbol{\theta}_{\mathbf{N}}) = \boldsymbol{\Omega}(\boldsymbol{\theta}) \tag{16}$$

, where $\boldsymbol{\theta}_{\mathbf{N}}$ is the vector of parameters from the normalized model. It is necessary to verify that the imposed normalization does not otherwise restrict the model; that is, the covariance matrix must remain the same as before the restriction is imposed. The equality condition assumes particular importance for HCMs with the logit kernel for reasons that shall become clear in Section 3.4 and discussed in greater detail in Walker et al. (2007).

The identification steps described above apply to any general econometric model of the form shown in equation (13). Over the course of the following sections, we use the steps outlined above to ascertain identifiability of the different components of the HCM as outlined by the Three Step Rule.

3.2 Confirmatory Factor Analytic Model

Step 1 of the Three Step Rule requires that the measurement component of the structural equations sub-model, when reformulated as a confirmatory factor analytic model, be identifiable. Prior to model estimation, the indicators are usually processed. For each indicator response, the analyst calculates the deviation from the respective mean, and it is these deviations that are used directly during model estimation. Combining equations (9) and (10) from Section 2.3, the measurement component of the structural equations sub-model may be restated as:

$$\mathbf{i}_{\mathbf{nj}} = \boldsymbol{\Lambda}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{\ell}} + \mathbf{D}\boldsymbol{\Theta}\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{C}} \tag{17}$$

, where $\mathbf{i}_{\mathbf{nj}}$ is a $(\mathrm{Q}\times 1)$ vector of observed indicators representing deviations from the sample mean, $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{\ell}}$ is an $\left(\mathrm{L}^{\boldsymbol{\ell}}\times 1\right)$ matrix of latent variables, $\boldsymbol{\Lambda}$ is a $\left(\mathrm{Q}\times \mathrm{L}^{\boldsymbol{\ell}}\right)$ matrix of coefficients, $\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{C}}$ is a $(\mathrm{C}\times 1)$ vector of independent random factors with mean zero and variance one, $\boldsymbol{\Theta}$ is a $(\mathrm{C}\times\mathrm{C})$ lower triangular matrix that is the Cholesky factorization of the covariance structure between the indicators, and $\mathbf{D}$ is a $(\mathrm{Q}\times\mathrm{C})$ matrix of factor loadings that map the random factors $\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{C}}$ to the covariance structure. In reformulating the measurement model as a confirmatory factor analytic model, we introduce the additional term $\boldsymbol{\Phi} = \mathbf{E}\left(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{\ell}}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{\ell}\,\prime}\right)$, the $\left(\mathrm{L}^{\boldsymbol{\ell}}\times\mathrm{L}^{\boldsymbol{\ell}}\right)$ covariance matrix of the latent factors that captures correlation between each pair of latent variables. The parameters to be identified are the elements of $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$.

Section 3.2.1 applies the general rules of identification presented in Section 3.1 to the confirmatory factor analytic model of equation (17), and works through an example to demonstrate how the rules might be applied in practice. Unfortunately, with growing model complexity the general rules of identification can often prove unwieldy. In such cases, alternative rules can be more useful. Section 3.2.2 reviews some of the commonly employed rules in the literature to determine identifiability. These rules cover most confirmatory factor analytic models found in the literature, and readers uninterested in the mathematical details pertaining to the general case may skip ahead to Section 3.2.2.

3.2.1 Analysis of the Covariance Matrix

This section is based on Bollen (1989), and for more details the reader is referred to the original text. For confirmatory factor analytic models, the indicators $\mathbf{i_{nj}}$ are the only observable variables. Assuming that the vector of observable variables $\mathbf{i_{nj}}$ in equation (17) does not denote absolute values but deviations from the mean, the covariance matrix may be parameterized as follows:

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \mathbf{E}\left(\mathbf{i_{nj}}\mathbf{i_{nj}'}\right) = \mathbf{E}\left[\left(\boldsymbol{\Lambda}\mathbf{x_{nj}^{\ell}} + \mathbf{D}\boldsymbol{\Theta}\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{C}}\right)\left(\mathbf{x_{nj}^{\ell}}'\boldsymbol{\Lambda}' + \boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{C}}{}'\boldsymbol{\Theta}'\mathbf{D}'\right)\right]$$

$$\Rightarrow \boldsymbol{\Omega}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\mathbf{E}\left(\mathbf{x_{nj}^{\ell}}\mathbf{x_{nj}^{\ell}}'\right)\boldsymbol{\Lambda}' + \mathbf{D}\boldsymbol{\Theta}\mathbf{E}\left(\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{C}}\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{C}}{}'\right)\boldsymbol{\Theta}'\mathbf{D}'$$

$$\Rightarrow \boldsymbol{\Omega}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \mathbf{D}\boldsymbol{\Theta}\boldsymbol{\Theta}'\mathbf{D}' \tag{18}$$

, where $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is the (Q×Q) parameterized covariance matrix (and is independent of the observed data). Therefore, the covariance matrix of observable variables $\mathbf{i_{nj}}$ may be parameterized in terms of the elements of $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, $\mathbf{D}$ and $\boldsymbol{\Theta}$. The elements of $\mathbf{D}$ comprise zeros and ones, and must be set by the analyst based on prior hypotheses. The elements of $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ are unknown parameters. Combining equations (14) and (18), the identification problem may be restated as finding constraints that ensure a solution to the following system of nonlinear equations:

$$\Rightarrow \boldsymbol{\Omega} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \mathbf{D}\boldsymbol{\Theta}\boldsymbol{\Theta}'\mathbf{D}' \tag{19}$$

, where $\boldsymbol{\Omega}$ here is the (Q×Q) sample covariance matrix of observable indicator responses $\mathbf{i_{nj}}$ to each of the Q indicator constructs. The left-hand side of equation (19) is a function of the observed data and the right-hand side is a function of the hypothesized model specification.

The rules of identification presented in Section 3.1 may now be applied to equation (19) to verify identifiability. The general approach for any confirmatory factor analytic model requires the analyst to be able to express all of the unknown parameters in $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ as some function of the elements of the sample covariance matrix $\boldsymbol{\Omega}$. Appendix A uses this approach to demonstrate why the analyst needs to impose constraints to set the scale of the latent variables, and how this might be accomplished, and subsequently applies equation (19) to evaluate identifiability of the confirmatory factor analytic model of Figure 4B.

3.2.2 Alternative Rules of Identification

Ensuring an algebraic solution to equation (19) can often be tedious and error-prone (Bollen, 1989). For these reasons, researchers have developed alternative rules for some of the more popularly employed model forms. Over the following paragraphs, we review some of the rules commonly cited in the literature and useful to the identification of HCMs. There are perhaps as

many rules of identification as there are model forms. The rules presented here are by no means an exhaustive set and apply only to models with a factor complexity of one, i.e. models where each indicator loads on a single latent variable. Notable among the rules not included here are the set of necessary and sufficient conditions for identification of models with factor complexity one developed by Reilly (1995), and the set of sufficient but not necessary conditions for identification of models with arbitrary factor complexity developed by Reilly and O'Brien (1996). Lastly, each of the following rules assumes that the scale of the latent factors has already been set through the imposition of appropriate constraints on $\mathbf{\Phi}$, the covariance matrix between the latent factors (readers interested in more details on how or why to set the scale of the latent factor should refer to Appendix A).

1. *The Three Indicator Rule (Bollen, 1989)*: A model with one or more factors is identified when it has (1) three or more indicators per latent factor; (2) a factor complexity of one; and (3) uncorrelated measurement errors between the indicators. The Three Indicator Rule places no additional restrictions on $\mathbf{\Phi}$, the covariance matrix between the latent factors, other than those required to set the scale of the latent factors. It is a sufficient but not necessary condition for identification.

2. *The Two Indicator Rule (Bollen, 1989)*: This is an alternative sufficient condition which states that a model with one or more latent factors is identified when it has (1) two or more indicators per latent factor; (2) a factor complexity of one; (3) uncorrelated measurement errors between the indicators; and (4) each latent factor is correlated with at least one other latent factor.

3. *Single Factor One Indicator Rule (O'Brien, 1994):* For a single factor model with two uncorrelated indicators, if the factor loading of one of the indicators is identified (based on some other rule) then the factor loading of the other indicator is also identified.

4. *Single Factor Error Variance Rule (O'Brien, 1994):* For a single factor model with any number of indicators, if the factor loading of a particular indicator is identified then the variance of the measurement error of the same indicator is also identified.

5. *Single Factor Error Covariance Rule (O'Brien, 1994):* For a single factor model with two correlated indicators, if the factor loadings of both indicators are identified then the covariance between the measurement errors of the two indicators is also identified.

6. *Multifactor Two Indicator One Indicator Rule (O'Brien, 1994):* If a model has two latent factors and three uncorrelated indicators such that two indicators load on one factor and one indicator on the other factor, then the loadings of the two indicators loading on the same factor are identified.

7. *Latent Variable Covariance Rule (O'Brien, 1994):* For a model with two correlated latent factors and two uncorrelated indicators such that the each indicator loads on one of the two latent factors, if the loadings on both indicators are identified then the covariance between the corresponding latent factors is also identified.

8. *Multifactor One Indicator Rule (O'Brien, 1994):* For a model with two correlated latent factors and two uncorrelated indicators such that the each indicator loads on one of the two latent factors, if the covariance between the two latent factors and one of the factor loadings are both identified, then the other factor loading is also identified.

9. *Multifactor Error Covariance Rule (O'Brien, 1994):* For a model with two correlated latent factors and two correlated indicators such that the each indicator loads on one of the two latent factors, if the covariance between the two latent factors and both factor loadings are identified, then the covariance between the measurement errors of the two corresponding indicators is also identified.

The different rules together may read much like a cookbook, but the proof for any of these rules can be derived fairly straightforwardly through the analysis of the covariance matrix of the analogous model using the methods described in Section 3.2.1. Furthermore, through repeated application of the rules in the right sequential order, the reader can address the identification problem for fairly complex confirmatory factor analytic models. For more details on how these rules may be used in practice, the reader is referred to O'Brien (1994).

3.3 Structural Equation Models with Observable Variables

Step 2 of the Three Step Rule requires that the structural component of the structural equations sub-model, when reformulated as a structural equations model with observable variables, be identifiable. For the purposes of identification, we assume throughout this section that the endogenous and exogenous variables have been processed such that they represent deviations from the sample mean. Note that this is purely for notational convenience, and does not change the identification problem for structural equation models with observable variables. Combining equations (11) and (12) from Section 2.4, the structural component of the structural equations sub-model may be stated as:

$$\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}} = \mathbf{B}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}} + \boldsymbol{\Gamma}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}} + \mathbf{G}\boldsymbol{\Psi}\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}} \tag{20}$$

, where $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}$ is the $(\mathrm{L}^{\boldsymbol{d}}\times 1)$ vector of endogenous variables representing deviations from the sample mean, $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}$ is the $(\mathrm{L}^{\boldsymbol{x}}\times 1)$ vector of exogenous variables representing deviations from the sample mean, $\mathbf{B}$ is an $(\mathrm{L}^{\boldsymbol{d}}\times\mathrm{L}^{\boldsymbol{d}})$ matrix of coefficients, $\boldsymbol{\Gamma}$ is an $(\mathrm{L}^{\boldsymbol{d}}\times\mathrm{L}^{\boldsymbol{x}})$ matrix of coefficients, $\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}}$ is an $(\mathrm{S}\times 1)$ vector of independent random factors with mean zero and variance one, $\boldsymbol{\Psi}$ is an $(\mathrm{S}\times\mathrm{S})$ lower triangular matrix that is the Cholesky factorization of the covariance structure between the endogenous variables, and $\mathbf{G}$ is an $(\mathrm{L}^{\boldsymbol{d}}\times\mathrm{S})$ matrix of factor loadings that map the random factors $\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}}$ to the covariance structure. In reformulating the structural component as a structural equations model with observed variables, we assume that both $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}$ and $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}$ are observed, and reintroduce the symbol $\boldsymbol{\Phi}$, but on this occasion to represent the $(\mathrm{L}^{\boldsymbol{x}}\times\mathrm{L}^{\boldsymbol{x}})$ covariance matrix of the exogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}$, i.e. $\boldsymbol{\Phi} = \mathbf{E}\left(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}\,\prime}\right)$. The unknown parameters are the elements of $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$.

Section 3.3.1 applies the general rules of identification presented in Section 3.1 to the structural equations model with observable variables of equation (20). Section 3.3.2 presents a set of alternative rules that cover two cases most commonly employed in the literature on HCMs: linear regression and recursive models. Readers uninterested in the general conditions may skip ahead to Section 3.3.2.

3.3.1 Analysis of the Covariance Matrix

This section is based on findings presented in Fisher (1976) and Bollen (1989), and readers interested in more details are referred to the original texts. For structural equation models with

observed variable, both the vector of endogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}$ and the vector of exogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}$ are observed, and the $(\mathrm{L}^{\boldsymbol{d}}+\mathrm{L}^{\boldsymbol{x}})\times(\mathrm{L}^{\boldsymbol{d}}+\mathrm{L}^{\boldsymbol{x}})$ covariance matrix of observed variables may be given by:

$$\boldsymbol{\Omega}(\boldsymbol{\theta})=\begin{bmatrix}\mathrm{Var}(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}) & \\ \mathrm{Cov}(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}},\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}) & \mathrm{Var}(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}})\end{bmatrix} \tag{21}$$

Rearranging equation (20), we get $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}=\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}}-\mathbf{B}\right)^{-1}\left(\boldsymbol{\Gamma}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}+\mathbf{G}\boldsymbol{\Psi}\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}}\right)$, where $\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}}$ denotes the $(\mathrm{L}^{\boldsymbol{d}}\times\mathrm{L}^{\boldsymbol{d}})$ identity matrix. Then, the variance of the endogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}$ can be calculated as follows:

$$\begin{aligned}\mathrm{Var}(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}) &= \mathbf{E}\left(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}\,\prime}\right)\\ &= \mathrm{E}\left[\left(\mathcal{I}_{\mathrm{L}^{d}}-\mathrm{B}\right)^{-1}\left(\Gamma\mathrm{x}_{\mathrm{nj}}^{x}+\mathrm{G}\Psi\eta_{\mathrm{nj}}^{\mathrm{S}}\right)\left(\mathrm{x}_{\mathrm{nj}}^{x\,\prime}\Gamma'+\eta_{\mathrm{nj}}^{\mathrm{S}\,\prime}\Psi'\mathrm{G}'\right)\left(\mathcal{I}_{\mathrm{L}^{d}}-\mathrm{B}\right)^{-1\prime}\right]\\ &= \left(\mathcal{I}_{\mathrm{L}^{d}}-\mathrm{B}\right)^{-1}\left(\Gamma\mathrm{E}\left[\mathrm{x}_{\mathrm{nj}}^{x}\mathrm{x}_{\mathrm{nj}}^{x\,\prime}\right]\Gamma'+\mathrm{G}\Psi\mathrm{E}\left[\eta_{\mathrm{nj}}^{\mathrm{S}}\eta_{\mathrm{nj}}^{\mathrm{S}\,\prime}\right]\Psi'\mathrm{G}'\right)\left(\mathcal{I}_{\mathrm{L}^{d}}-\mathrm{B}\right)^{-1\prime}\\ &= \left(\mathcal{I}_{\mathrm{L}^{d}}-\mathrm{B}\right)^{-1}(\Gamma\Phi\Gamma'+\mathrm{G}\Psi\Psi'\mathrm{G}')\left(\mathcal{I}_{\mathrm{L}^{d}}-\mathrm{B}\right)^{-1\prime}\end{aligned} \tag{22}$$

, where it is assumed that the exogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}$ are uncorrelated with the measurement errors $\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}}$. Next, by definition $\boldsymbol{\Phi}=\mathrm{Var}(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}})$. And last, the covariance term $\mathrm{Cov}(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}},\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}})$ is given by:

$$\begin{aligned}\mathrm{Cov}(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}},\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}) &= \mathbf{E}\left[\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}\,\prime}\right]\\ &= \mathbf{E}\left[\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}\left(\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}\,\prime}\boldsymbol{\Gamma}'+\boldsymbol{\eta}_{\mathbf{nj}}^{\mathbf{S}\,\prime}\boldsymbol{\Psi}'\mathbf{G}'\right)\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}}-\mathbf{B}\right)^{-1\prime}\right]\\ &= \mathbf{E}\left[\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}\,\prime}\right]\boldsymbol{\Gamma}'\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}}-\mathbf{B}\right)^{-1\prime}\\ &= \boldsymbol{\Phi}\boldsymbol{\Gamma}'\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}}-\mathbf{B}\right)^{-1\prime}\end{aligned} \tag{23}$$

Combining equations (22) and (23) with equations (14) and (21), the identification problem for a structural equations model with observed variables may be stated as finding solutions to the following system of equations:

$$\begin{bmatrix}\boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{d}}\mathbf{X}^{\boldsymbol{d}}} & \\ \boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{x}}\mathbf{X}^{\boldsymbol{d}}} & \boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{x}}\mathbf{X}^{\boldsymbol{x}}}\end{bmatrix}=\begin{bmatrix}\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}}-\mathbf{B}\right)^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}'+\mathbf{G}\boldsymbol{\Psi}\boldsymbol{\Psi}'\mathbf{G}')\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}}-\mathbf{B}\right)^{-1\prime} & \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}'\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}}-\mathbf{B}\right)^{-1\prime} & \boldsymbol{\Phi}\end{bmatrix} \tag{24}$$

, where $\boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{d}}\mathbf{X}^{\boldsymbol{d}}}$ is the $(\mathrm{L}^{\boldsymbol{d}}\times\mathrm{L}^{\boldsymbol{d}})$ sample covariance matrix of the endogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}$; $\boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{x}}\mathbf{X}^{\boldsymbol{x}}}$ is the $(\mathrm{L}^{\boldsymbol{x}}\times\mathrm{L}^{\boldsymbol{x}})$ sample covariance matrix of the exogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{x}}$; and $\boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{x}}\mathbf{X}^{\boldsymbol{d}}}$ is the $(\mathrm{L}^{\boldsymbol{x}}\times\mathrm{L}^{\boldsymbol{d}})$ sample covariance matrix between the endogenous and exogenous variables. The parameters to be identified are $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. First, the reader should observe that $\boldsymbol{\Phi}$, the covariance matrix of the exogenous variables, is fully identified from the equation $\boldsymbol{\Phi}=\boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{x}}\mathbf{X}^{\boldsymbol{x}}}$. Substituting the expression for $\boldsymbol{\Phi}$ in equation (23) and combining with equation (24), we get:

$$\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}} - \mathbf{B}\right)^{-1}\boldsymbol{\Gamma} = \boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{d}}\mathbf{X}^{\boldsymbol{x}}}\boldsymbol{\Omega}^{-\mathbf{1}}_{\mathbf{X}^{\boldsymbol{x}}\mathbf{X}^{\boldsymbol{x}}} \tag{25}$$

$$\Rightarrow \boldsymbol{\Gamma} + \mathbf{B}\boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{d}}\mathbf{X}^{\boldsymbol{x}}}\boldsymbol{\Omega}^{-\mathbf{1}}_{\mathbf{X}^{\boldsymbol{x}}\mathbf{X}^{\boldsymbol{x}}} = \boldsymbol{\Omega}_{\mathbf{X}^{\boldsymbol{d}}\mathbf{X}^{\boldsymbol{x}}}\boldsymbol{\Omega}^{-\mathbf{1}}_{\mathbf{X}^{\boldsymbol{x}}\mathbf{X}^{\boldsymbol{x}}} \tag{26}$$

With regards to set of equations in the endogenous variables $\mathbf{x}^{\boldsymbol{d}}_{\mathbf{nj}}$ given by (20), we examine the identification of the elements of $\mathbf{B}$ and $\boldsymbol{\Gamma}$ one equation at a time or, with regards to the matrices $\mathbf{B}$ and $\boldsymbol{\Gamma}$ itself, one row at a time. Any row of the matrix on the left hand side of equation (26) may comprise a maximum of $\left(\mathrm{L}^{\boldsymbol{x}} + \mathrm{L}^{\boldsymbol{d}} - 1\right)$ unknown parameters, $\mathrm{L}^{\boldsymbol{x}}$ from the corresponding row in $\boldsymbol{\Gamma}$ and $\left(\mathrm{L}^{\boldsymbol{d}} - 1\right)$ from the corresponding row in $\mathbf{B}$ (minus one because the diagonal elements of $\mathbf{B}$ are constrained to zero). However, the dimension of the matrix on the right hand side is $\left(\mathrm{L}^{\boldsymbol{d}}\times\mathrm{L}^{\boldsymbol{x}}\right)$, i.e. the number of elements in any row is $\mathrm{L}^{\boldsymbol{x}}$. Therefore, for each of the $\mathrm{L}^{\boldsymbol{d}}$ equations corresponding to the endogenous variables $\mathbf{x}^{\boldsymbol{d}}_{\mathbf{nj}}$, we have a maximum of $\left(\mathrm{L}^{\boldsymbol{x}} + \mathrm{L}^{\boldsymbol{d}} - 1\right)$ unknown parameters and $\mathrm{L}^{\boldsymbol{x}}$ equations. The order condition may be restated as the requirement that for each of the $\mathrm{L}^{\boldsymbol{d}}$ equations corresponding to the endogenous variables $\mathbf{x}^{\boldsymbol{d}}_{\mathbf{nj}}$, at least $\left(\mathrm{L}^{\boldsymbol{d}} - 1\right)$ of the endogenous and exogenous variables $\mathbf{x}^{\boldsymbol{d}}_{\mathbf{nj}}$ and $\mathbf{x}^{\boldsymbol{x}}_{\mathbf{nj}}$ must be excluded from the equation.

When $\mathbf{B}$ is large, $\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}} - \mathbf{B}\right)^{-1}$ is tedious to compute and equation (25) can be hard to solve for the unknown parameters in $\mathbf{B}$ and $\boldsymbol{\Gamma}$. An alternative approach, proposed by Fisher (1976), begins by constructing the following matrices:

$$\mathbf{A} = \begin{bmatrix}\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}} - \mathbf{B}\right) & -\boldsymbol{\Gamma}\end{bmatrix}, \text{ and } \mathbf{W} = \begin{bmatrix}\left(\boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{d}}} - \mathbf{B}\right)^{-1}\boldsymbol{\Gamma} \\ \boldsymbol{\mathcal{I}}_{\mathrm{L}^{\boldsymbol{x}}}\end{bmatrix}$$

$$\Rightarrow \mathbf{AW} = \mathbf{0} \Rightarrow \mathbf{a_t W} = \mathbf{0} \tag{27}$$

, where $\mathbf{A}$ is an $\mathrm{L}^{\boldsymbol{d}}\times\left(\mathrm{L}^{\boldsymbol{d}} + \mathrm{L}^{\boldsymbol{x}}\right)$ matrix and $\mathbf{a_t}$ is the t^{th} row of the matrix $\mathbf{A}$; and $\mathbf{W}$ is an $\left(\mathrm{L}^{\boldsymbol{d}} + \mathrm{L}^{\boldsymbol{x}}\right)\times\mathrm{L}^{\boldsymbol{x}}$ matrix. Let J^{t} be the number of parameters constrained to zero in the t^{th} equation (not including the corresponding diagonal element of $\mathbf{B}$), and $\boldsymbol{\emptyset}^{\mathbf{t}}$ denote the $\left(\mathrm{L}^{\boldsymbol{d}} + \mathrm{L}^{\boldsymbol{x}}\right)\times\mathrm{J}^{\mathrm{t}}$ matrix whose element $\emptyset^{\mathrm{t}}_{\mathrm{ij}}$ equals one if the j^{th} constraint on equation t states that the i^{th} parameter in $\mathbf{a_t}$ should be zero. In other words, $\mathbf{a_t}\boldsymbol{\emptyset}^{\mathbf{t}} = \mathbf{0}$. Combining with equation (27), we get:

$$\mathbf{a_t}[\boldsymbol{\emptyset}^{\mathbf{t}}|\mathbf{W}] = \mathbf{0} \Rightarrow \mathbf{a_t C^t} = \mathbf{0} \tag{28}$$

, where $\mathbf{C^t} = [\boldsymbol{\emptyset}^{\mathbf{t}}|\mathbf{W}]$ is the $\left(\mathrm{L}^{\boldsymbol{d}} + \mathrm{L}^{\boldsymbol{x}}\right)\times(\mathrm{J}^{\mathrm{t}} + \mathrm{L}^{\boldsymbol{x}})$ matrix formed by adjoining the matrices $\boldsymbol{\emptyset}^{\mathbf{t}}$ and $\mathbf{W}$. Since equation (28) captures all that is known about the vector of parameters $\mathbf{a_t}$, the equation is identifiable if and only if any vector that satisfies the equation is a scalar multiple of the true underlying $\mathbf{a_t}$. But $\mathbf{a_t'}$ belongs to the null space of $\mathbf{C^{t'}}$. Hence, the equation is identifiable if and only if the dimension of the null-space of $\mathbf{C^{t'}}$ is equal to one. Applying the rank-nullity theorem, the dimension of the null-space of $\mathbf{C^{t'}}$ is equal to one if and only if the rank of $\mathbf{C^{t'}}$, and hence the rank of $\mathbf{C^t}$, is equal to $\left(\mathrm{L}^{\boldsymbol{d}} + \mathrm{L}^{\boldsymbol{x}}\right) - 1$. It can further be shown that the rank of $\mathbf{C^t}$ is equal to $\left(\mathrm{L}^{\boldsymbol{d}} + \mathrm{L}^{\boldsymbol{x}}\right) - 1$ if and only if the rank of $\mathbf{A}\boldsymbol{\emptyset}^{\mathbf{t}} = \mathrm{L}^{\boldsymbol{d}} - 1$. Hence the parameters in the t^{th} equation in $\mathbf{x}^{\boldsymbol{d}}_{\mathbf{nj}}$, or the t^{th} row of $\mathbf{B}$ and $\boldsymbol{\Gamma}$, are identified if and only if the rank of the matrix $\mathbf{A}\boldsymbol{\emptyset}^{\mathbf{t}}$ is $\mathrm{L}^{\boldsymbol{d}} - 1$. This is known as the rank condition. Note that the matrix $\mathbf{A}\boldsymbol{\emptyset}^{\mathbf{t}}$ can be obtained from removing the columns of $\mathbf{A}$ that do not have a zero in the t^{th} row of $\mathbf{A}$.

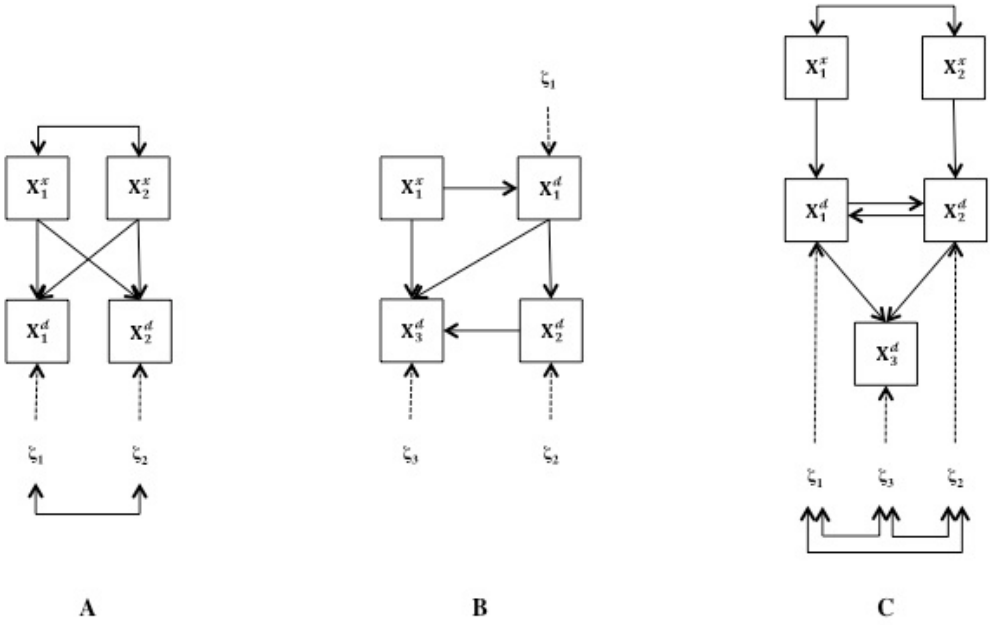

Figure 6: Examples of structural equation models with observable variables: (**A**) Seemingly unrelated regression; (**B**) Recursive model; and (**C**) Nonrecursive model

Once the parameters in **B** and **Γ** have been identified, equations (22) and (24) can be used to show that the full covariance matrix between the measurement errors of the endogenous variables, as represented by the parameters **Ψ** and the factor loadings **G**, is also identified. The order condition is a necessary but not sufficient condition of identification, whereas the rank condition is a necessary and sufficient condition of identification. Appendix B illustrates how the order and rank conditions might be applied to establish identification of the nonrecursive model shown in Figure 6C, taken from Hanneman (2000).

3.3.2 Alternative Rules of Identification

Most studies on HCMs employ fairly simple representations of the structural relationships between explanatory variables. In the following paragraphs, we briefly review two rules that cover most model forms found in the literature. For a more thorough treatment of the rules, the reader is referred to Bollen (1989).

1. *The Null B Rule*: The rule is a sufficient condition of identification that states that a model with endogenous variables that do not affect one another, i.e. a model with a null **B** matrix, is identified. The rule places no restriction on the covariance structure between the vector of endogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}$. If the endogenous variables are uncorrelated, then each variable may be treated separately as a regression equation. If the analyst has reasons to hypothesize correlation between the endogenous variables, then the model specification is reduced to a system of seemingly unrelated regression equations. The model drawn in Figure 6A is an example where the Null B rule may be used to determine identification.

2. *The Recursive Rule*: A structural equations model with observed variables is said to be recursive if the system of equations given by (20) contain no reciprocal or causation loops, and it is possible to write the matrix **B** denoting the influence of the endogenous variables on

each other as a lower triangular matrix. The Recursive Rule is a sufficient condition of identification that states that a recursive model with multiple endogenous and exogenous variables is identified if the vector of endogenous variables $\mathbf{x}_{\mathbf{nj}}^{\boldsymbol{d}}$ is uncorrelated. Figure 6B shows a model that can be identified using the recursive rule.

If the structural component of the structural equations sub-model can be written as either a regression model or a recursive model, then the null B or recursive rule provide sufficient conditions for identification. For other model forms, such as a recursive model with correlated endogenous variables, the rank and order conditions presented in Section 3.3.1 provide a set of necessary and sufficient conditions for identification.

3.4 Discrete Choice Models

Step 3 of the Three Step Rule requires that the structural component of the discrete choice model be identifiable, treating each explanatory variable as an exogenous observed variable with no measurement error. We restate the structural component of the discrete choice model:

$$\mathbf{u_n} = \mathbf{v_n} + \boldsymbol{\varepsilon_n} \tag{29}$$

$$\Rightarrow \mathbf{u_n} = \mathbf{X_n'}\boldsymbol{\beta} + \mathbf{F_n}\boldsymbol{\Upsilon}\boldsymbol{\eta}_{\mathbf{n}}^{\mathbf{R}} + \boldsymbol{\nu_n} \tag{30}$$

, where $\mathbf{u_n}$ is a $(\mathrm{J}\times 1)$ vector of random utilities; $\mathbf{v_n}$ and $\boldsymbol{\varepsilon_n}$ are $(\mathrm{J}\times 1)$ vectors that comprise the systematic and stochastic component of $\mathbf{u_n}$, respectively; $\mathbf{X_n}$ is an $(\mathrm{L}\times\mathrm{J})$ matrix of explanatory variables; $\boldsymbol{\beta}$ is an $(\mathrm{L}\times 1)$ vector of parameters; $\boldsymbol{\eta}_{\mathbf{n}}^{\mathbf{R}}$ is an $(\mathrm{R}\times 1)$ vector of independent random factors with mean zero and variance one; $\boldsymbol{\Upsilon}$ is an $(\mathrm{R}\times\mathrm{R})$ lower triangular matrix that is the Cholesky factorization of the covariance structure of the utilities; $\mathbf{F_n}$ is a $(\mathrm{J}\times\mathrm{R})$ matrix of factor loadings that map the random factors to the covariance structure; and $\boldsymbol{\nu_n}$ is a $(\mathrm{J}\times 1)$ vector of i.i.d. Extreme Value random variables with mean zero and variance g/μ^2, where μ is the scale and g is the variance of a standard Extreme Value random variable.

To address the identification problem for discrete choice models, we reframe equation (30) so that it resembles the form of structural equation models with observable variables discussed in Section 3.3. As was the case in Section 3.3, we will assume that the explanatory variables have been processed such that they represent deviations from the sample mean. To reiterate, this is purely for notational convenience, and does not change the identification problem. Let $\mathbf{x_n}$ be the $(\mathrm{JL}\times 1)$ vector of explanatory variables constructed as shown below:

$$\mathbf{x_n} = \begin{bmatrix} \mathbf{x_{n1}} \\ \vdots \\ \mathbf{x_{nJ}} \end{bmatrix}$$

, where $\mathbf{x_{nj}}$ is the $(\mathrm{L}\times 1)$ vector of explanatory variables corresponding to the j^{th} alternative, i.e. $\mathbf{x_{nj}}$ is the j^{th} column of $\mathbf{X_n}$. We reintroduce the symbol $\boldsymbol{\Phi}$ to represent the $(\mathrm{JL}\times\mathrm{JL})$ covariance matrix of the explanatory variables $\mathbf{x_n}$, i.e. $\boldsymbol{\Phi} = \mathbf{E}(\mathbf{x_n}\mathbf{x_n'})$. Similarly, let $\mathbf{B}$ be the $(\mathrm{J}\times\mathrm{JL})$ block diagonal matrix of parameters constructed as follows:

$$\mathbf{B} = \begin{bmatrix} \beta' & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \beta' \end{bmatrix}$$

Then, equation (30) may be restated as:

$$\mathbf{u_n} = \mathbf{B x_n} + \mathbf{F_n \Upsilon \eta_n^R} + \boldsymbol{\nu_n} \tag{31}$$

There are two sets of relevant parameters to be considered for identification: the matrix $\mathbf{B}$ that enters the systematic component of the utility specification, and the unrestricted parameters of the distribution of the stochastic component $\boldsymbol{\varepsilon}$. Utility as a construct is a latent variable whose location and scale need to be fixed. With regards to location, only the difference in utilities is observable from the measurement component of the discrete choice sub-model, and the parameters $\mathbf{B}$ and $\boldsymbol{\varepsilon}$ must be normalized accordingly. The scale can be set either by constraining the variance of the error (for example, $\mu = 1$) or by constraining one of the systematic parameters (for example, a particular $\beta = 1$). The assumption made throughout this section is that the scale is normalized through the error variance.

Since only the differences in utilities are observable, and not the absolute levels themselves, we rewrite the structural component of the discrete choice model as follows:

$$\Delta \mathbf{u_n} = \Delta \mathbf{v_n} + \Delta \boldsymbol{\varepsilon_n} \tag{32}$$

$$\Rightarrow \Delta \mathbf{u_n} = \Delta \mathbf{B x_n} + \Delta \mathbf{F_n \Upsilon \eta_n^R} + \Delta \boldsymbol{\nu_n} \tag{33}$$

, where Δ is the linear operator that transforms the J utilities into $(J-1)$ utility differences taken with respect to the J^{th} alternative. We have assumed here, for the sake of simplicity, that all individuals face the same choice set. For a discussion on mapping from the deviation with respect to the last alternative available to decision-maker n with heterogeneous choice sets, see Bolduc (1999). Δ is a $(J-1)\times J$ matrix that consists of a $(J-1)\times(J-1)$ identity matrix with a column vector of -1's appended as the J^{th} column. Though Δ performs the differences with respect to the last alternative for each choice situation, the identification problem is invariant with respect to which alternative is used as the base.

Equation (14) can now be used to examine identification of equation (33), where the observable variables are $\Delta \mathbf{u_n}$ and $\mathbf{x_n}$, and the unknown parameters are $\mathbf{B}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Upsilon}$ and μ (the scale of $\boldsymbol{\nu_n}$). The $(J-1+JL)\times(J-1+JL)$ covariance matrix of observed variables may be given by:

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \begin{bmatrix} \mathrm{Var}(\Delta \mathbf{u_n}) & \\ \mathrm{Cov}(\mathbf{x_n}, \Delta \mathbf{u_n}) & \mathrm{Var}(\mathbf{x_n}) \end{bmatrix} \tag{34}$$

The variance of the differences in utilities $\mathrm{Var}(\Delta \mathbf{u_n})$ can be calculated as follows:

$$\begin{aligned} \mathrm{Var}(\Delta \mathbf{u_n}) &= \mathbf{E}(\Delta \mathbf{u_n} \mathbf{u_n'} \Delta') \\ &= \Delta \mathrm{E}\left[\left(\mathbf{B x_n} + \mathbf{F_n \Upsilon \eta_n^R} + \boldsymbol{\nu_n}\right)\left(\mathbf{x_n'} \mathbf{B'} + \boldsymbol{\eta_n^{R'}} \boldsymbol{\Upsilon'} \mathbf{F_n'} + \boldsymbol{\nu_n'}\right)\right] \Delta' \\ &= \Delta \mathbf{B} \mathrm{E}[\mathbf{x_n x_n'}] \mathbf{B'} \Delta' + \Delta \mathbf{F_n \Upsilon} \mathrm{E}\left[\boldsymbol{\eta_n^R \eta_n^{R'}}\right] \boldsymbol{\Upsilon'} \mathbf{F_n'} \Delta' + \Delta \mathrm{E}[\boldsymbol{\nu_n \nu_n'}] \Delta' \end{aligned}$$

$$= \Delta\mathbf{B}\boldsymbol{\Phi}\mathbf{B}'\Delta' + \Delta\mathbf{F_n}\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'\mathbf{F_n}'\Delta' + \left(\frac{g}{\mu^2}\right)\Delta\Delta' \tag{35}$$

, where it is assumed that the explanatory variables $\mathbf{x_n}$ are uncorrelated with any of the measurement errors. Next, by definition, $\mathrm{Var}(\mathbf{x_n}) = \mathbf{E}(\mathbf{x_n}\mathbf{x_n}') = \boldsymbol{\Phi}$. Lastly, the covariance term $\mathrm{Cov}(\mathbf{x_n}, \Delta\mathbf{u_n})$ is calculated as follows:

$$\begin{aligned}\mathrm{Cov}(\mathbf{x_n}, \Delta\mathbf{u_n}) &= \mathbf{E}(\mathbf{x_n}\mathbf{u_n}'\Delta') \\ &= \mathrm{E}\left[\mathbf{x_n}\left(\mathbf{x_n}'\mathbf{B}' + \boldsymbol{\eta}_\mathbf{n}^{\mathbf{R}'}\boldsymbol{\Upsilon}'\mathbf{F_n}' + \boldsymbol{\nu}_\mathbf{n}'\right)\Delta'\right] \\ &= \mathrm{E}[\mathbf{x_n}\mathbf{x_n}']\mathbf{B}'\Delta' \\ &= \boldsymbol{\Phi}\mathbf{B}'\Delta'\end{aligned} \tag{36}$$

Combining equations (35) and (36) with equations (14) and (34), the identification problem for discrete choice models may be stated as finding solutions to the following system of equations:

$$\begin{bmatrix}\boldsymbol{\Omega}_{\Delta\mathbf{u_n},\Delta\mathbf{u_n}} & \\ \boldsymbol{\Omega}_{\mathbf{x_n},\Delta\mathbf{u_n}} & \boldsymbol{\Omega}_{\mathbf{x_n},\mathbf{x_n}}\end{bmatrix} = \begin{bmatrix}\Delta\mathbf{B}\boldsymbol{\Phi}\mathbf{B}'\Delta' + \Delta\mathbf{F_n}\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'\mathbf{F_n}'\Delta' + \left(\frac{g}{\mu^2}\right)\Delta\Delta' & \\ \boldsymbol{\Phi}\mathbf{B}'\Delta' & \boldsymbol{\Phi}\end{bmatrix} \tag{37}$$

, where $\boldsymbol{\Omega}_{\Delta\mathbf{u_n},\Delta\mathbf{u_n}}$ is the $(J-1)\times(J-1)$ sample covariance matrix of the difference in utilities; $\boldsymbol{\Omega}_{\mathbf{x_n},\mathbf{x_n}}$ is the $(JL\times JL)$ sample covariance matrix of the explanatory variables; and $\boldsymbol{\Omega}_{\mathbf{x_n},\Delta\mathbf{u_n}}$ is the $JL\times(J-1)$ sample covariance matrix between the explanatory variables and the differences in utilities. It should be apparent that $\boldsymbol{\Phi}$, the covariance matrix of the explanatory variables, is fully identified from the equation $\boldsymbol{\Phi} = \boldsymbol{\Omega}_{\mathbf{x_n},\mathbf{x_n}}$. Section 3.4.1 will use the equation $\boldsymbol{\Phi}\mathbf{B}'\Delta' = \boldsymbol{\Omega}_{\mathbf{x_n},\Delta\mathbf{u_n}}$ to establish the identification conditions for the unknown parameter $\mathbf{B}$ in the systematic component of the choice sub-model. Once $\boldsymbol{\Phi}$ and $\mathbf{B}$ have both been identified, Section 3.4.2 uses equation (35) to identify the unknown parameters $\boldsymbol{\Upsilon}$ and μ contained in the stochastic component.

3.4.1 The Systematic Parameters

The matrix of systematic parameters $\mathbf{B}$ is identified if the following equation can be solved for each element of $\mathbf{B}$:

$$\boldsymbol{\Phi}\mathbf{B}'\Delta' = \boldsymbol{\Omega}_{\mathbf{x_n},\Delta\mathbf{u_n}}$$

$$\Rightarrow \Delta\mathbf{B} = \boldsymbol{\Omega}_{\Delta\mathbf{u_n},\mathbf{x_n}}\boldsymbol{\Omega}_{\mathbf{x_n},\mathbf{x_n}}^{-1} \tag{38}$$

, where $\Delta\mathbf{B}$ is a $(J-1)\times JL$ matrix. Identification of the unknown parameters in $\Delta\mathbf{B}$ may be examined one row at a time. If the sample covariance matrix of the explanatory variables $\boldsymbol{\Omega}_{\mathbf{x_n},\mathbf{x_n}}$ is non-singular, then all of the parameters in $\Delta\mathbf{B}$ are theoretically identified. However, if any of the explanatory variables in $\mathbf{x_n}$ are linearly dependent then the matrix $\boldsymbol{\Omega}_{\mathbf{x_n},\mathbf{x_n}}$ is singular. Since the utility specification is linear in parameters, it can be additively separated into the linearly independent component and the linearly dependent component, and the expression in equation (36) may be derived separately for the two components. The sub-matrix of $\boldsymbol{\Omega}_{\mathbf{x_n},\mathbf{x_n}}$ corresponding to the linearly independent variables allows identification of the corresponding parameters in a particular row of $\Delta\mathbf{B}$, whereas the sub-matrix of $\boldsymbol{\Omega}_{\mathbf{x_n},\mathbf{x_n}}$ corresponding to the linearly dependent variables is singular and the parameters in a particular row of $\Delta\mathbf{B}$ corresponding to these variables cannot be identified using the equations available from that row.

In general, it helps to make a distinction between continuous variables and categorical variables, and between alternative attributes and individual characteristics. The rules on how to include each of the four variable types in the utility specification to maintain identifiability are summarized below:

1. *Continuous attributes*, such as travel cost and travel time in a travel mode choice model, can enter the utility specification for each alternative, as long as there is some heterogeneity in the values taken by the attributes across different alternatives and choice situations.

2. *Categorical attributes*, such as vehicle make in a vehicle ownership model, require that a reference level be selected. For example, for a categorical attribute with C levels, a binary variable might be introduced for $C-1$ levels, excluding the reference level, in the utility for all J_n alternatives.

3. *Continuous characteristics*, such as age and income, may be included in the utilities of $J-1$ alternatives, one alternative being used as a reference.

4. *Categorical characteristics*, such as gender or education, require that both a reference level and a reference alternative be selected. A binary variable might then be introduced for each level of each characteristic, except the reference level for that characteristic, in the utilities of the $J-1$ alternatives, excluding the reference alternative.

The selection of the reference level and reference alternative has no effect on the model other than to shift the values of the parameters, preserving their differences, and this property holds even when the choice set varies across observations. Characteristics interacted with attributes result in variables that must also be treated as attributes, and depending on whether the resulting variable is continuous or categorical the appropriate attribute-specific rule may be used to verify identification. For more details on the specification of the systematic component, the reader is referred to Ben-Akiva and Lerman (1985).

3.4.2 The Error Structure

Once the analyst has identified $\mathbf{\Phi}$, the covariance matrix of the explanatory variables, and $\mathbf{B}$, the matrix of the unknown parameters in the systematic component of the utility specification, equation (35) may be rearranged as follows to help identify the unknown parameters $\mathbf{\Upsilon}$ and μ that define the error structure:

$$\Delta\mathbf{F_n}\mathbf{\Upsilon}\mathbf{\Upsilon}'\mathbf{F_n}'\Delta' + \left(\frac{g}{\mu^2}\right)\Delta\Delta' = \mathbf{\Omega}_{\Delta\mathbf{u_n},\Delta\mathbf{u_n}} - \Delta\mathbf{B}\mathbf{\Phi}\mathbf{B}'\Delta' = \mathbf{\Omega}_{\Delta\mathbf{u_n},\Delta\mathbf{u_n}} - \mathbf{\Omega}_{\Delta\mathbf{u_n},\mathbf{x_n}}\mathbf{\Omega}^{-\mathbf{1}}_{\mathbf{x_n},\mathbf{x_n}}\mathbf{\Omega}_{\mathbf{x_n},\Delta\mathbf{u_n}} \quad (39)$$

$$\Rightarrow \mathbf{\Omega}(\boldsymbol{\theta}) = \mathbf{\Omega} \quad (40)$$

, where the left hand side of the equation contains the function $\mathbf{\Omega}(\boldsymbol{\theta})$ of the unknown parameters $\boldsymbol{\theta} = \{\mathbf{\Upsilon}, \mu\}$, and the right hand side comprises the $(J-1)\times(J-1)$ symmetric matrix $\mathbf{\Omega}$ of known values. We persist in denoting the right-hand side of equation (39) by $\mathbf{\Omega}$ even though it isn't technically a sample covariance matrix, and the left-hand side by $\mathbf{\Omega}(\boldsymbol{\theta})$ even though it isn't a parameterized covariance matrix either. The identification problem for the error structure of the discrete choice model may be stated as finding a solution to each of the unknown parameters contained in equation (39).

Since the scale of the utility is typically normalized through the error structure, the system of equations given by (39) will contain one fewer independent equation than the general case. Therefore, the order condition from Section 3.1 must be restated. The number of estimable parameters S in the vector of unknown parameters $\boldsymbol{\theta}$ must satisfy the inequality:

$$S \leq \frac{J(J-1)}{2} - 1 \tag{41}$$

If $\boldsymbol{\Upsilon}$ is diagonal, as is often the case, the system of equations given by (39) is linear in the unknown parameters g/μ^2 and σ_i^2, where σ_i is the i^{th} diagonal element of $\boldsymbol{\Upsilon}$. Then, the number of estimable parameters S must also satisfy the following equality:

$$S = \text{Rank}\left(\text{Jacobian}\left(\text{vecu}\left(\boldsymbol{\Omega}(\boldsymbol{\theta})\right)\right)\right) - 1 \tag{42}$$

, where $\text{vecu}(.)$ is a function that vectorizes the unique elements of $\boldsymbol{\Omega}(\boldsymbol{\theta})$ into a column vector, and the Jacobian is equal to the derivatives of the elements of $\boldsymbol{\Omega}(\boldsymbol{\theta})$ with respect to the unknown parameters g/μ^2 and σ_i^2 contained in $\boldsymbol{\theta}$, where we redefine $\boldsymbol{\theta} = \{g/\mu^2, \sigma_i^2 \forall i = 1, \dots, R\}$. Since (39) results in a system of simultaneous linear equations, the rank of the Jacobian equals the number of independent equations in (39), minus one to set the scale of the utilities. The rank condition is more restrictive than the order condition, and is sufficient to ensure that there is a solution to (39). The order condition simply counts cells and ignores the internal structure of $\boldsymbol{\Omega}(\boldsymbol{\theta})$. The rank condition, however, counts the number of linearly independent equations available in $\boldsymbol{\Omega}(\boldsymbol{\theta})$ that can be used to estimate the parameters of the model. Together, (41) and (42) form a set of necessary and sufficient conditions for identification of the error structure.

The objective of the procedure outlined above is to find conditions for a discrete choice sub-model specified in levels under which the error structure can be properly identified and normalized. It is important to emphasize the implications of imposing restrictions on the covariance matrix at the levels ($\boldsymbol{\Upsilon}$) rather than on differences in utilities ($\boldsymbol{\Omega}(\boldsymbol{\theta})$), because this is the root cause of the complexity. Technically, only utility differences are estimable from the data. Once an arbitrary constraint has been selected for $\boldsymbol{\Omega}(\boldsymbol{\theta})$, one is done with identification. However, restrictions in the discrete choice sub-model are typically imposed in levels instead of on differences of utilities. This is because the structural parameters (i.e. the elements of $\boldsymbol{\Upsilon}$) are interpretable, whereas the parameters in the difference model (i.e. the elements of $\boldsymbol{\Omega}(\boldsymbol{\theta})$) are not. Therefore, our aim is to impose and possibly test restrictions on $\boldsymbol{\Upsilon}$ and verify that the model is identified. If the model is unidentified, some restriction will need to be imposed. Since we are working with the levels form, we want to impose the constraints on $\boldsymbol{\Upsilon}$. The choice of constraint on $\boldsymbol{\Upsilon}$ isn't always arbitrary (as it is on $\boldsymbol{\Omega}(\boldsymbol{\theta})$), and one has to make sure that it does not impose additional restrictions on $\boldsymbol{\Omega}(\boldsymbol{\theta})$. The equality condition described in Section 3.1 is necessary to ensure that the constraints do not change $\boldsymbol{\Omega}(\boldsymbol{\theta})$, and this is necessary due to the mixing with an Extreme Value error term that has already been normalized.

Walker et al. (2007) provides a comprehensive discussion of how the rules of identification may be applied to mixed logit models with heteroskedastic, nested and cross-nested error structures. Appendix C extends these findings to include models with random parameters on explanatory variables and demonstrates how the identification conditions may be established when working with panel data sets and agent effects. We summarize the important results below, and the reader may refer to the citations for more details:

1. *The Heteroskedastic Model:* In the Heteroskedastic model, the random error of each alternative has a different variance. The model allows for situations in which the systematic portion of the utility better represents the utility of some alternatives more than others. For $J = 2$, neither of the alternative-specific variances can be identified. For $J \geq 3$, $J - 1$ of the alternative-specific variances can be identified, and normalization must be imposed on the parameter corresponding to the minimum variance alternative. However, in practice there is no prior knowledge of the minimum variance alternative. For the general case with J alternatives, a brute force solution is to estimate J versions of the model, each with a different variance term normalized; the model with the best fit is the one with the correct normalization. This is both cumbersome and time consuming. Alternatively, one can estimate the unidentified model with all J variance terms. Although this model is not identified, a software estimation program will produce maximum likelihood parameter estimates (but not standard errors) that reflect the true covariance structure of the model. Therefore, the variance term with minimum estimated variance in the unidentified model is the minimum variance alternative, thus eliminating the need to estimate J different models.

2. *Nested and Cross-Nested Models:* In nested and cross-nested models, the stochastic component of the utility specification can be correlated across alternatives to allow for more flexible substitution patterns. The alternatives are partitioned into nests such that alternatives within a nest are correlated, and alternatives that do not share a nest are uncorrelated. Nested models refer to cases where the nests are mutually exclusive, i.e. an alternative can only belong to one nest. Cross-nested models relax this assumption and allow for overlapping. There are no general rules for the identification of nested and cross-nested models, and the analyst has to check the rank and order conditions on a case-by-case basis.

3. *The Random Parameters Model:* If the random parameter is imposed on a continuous attribute, there are no identification issues per se. Data permitting, the full covariance structure can be estimated. For a categorical attribute with two levels, independently distributed generic random parameters can be imposed on only one of the two binary variables corresponding to the two levels. For a categorical attribute with three or more levels, the variance term can be identified for independently distributed generic random parameters on the binary variables corresponding to each of the levels. In the case of independently distributed alternative-specific random parameters and a categorical attribute with C levels, where C can be two or more, a reference level must be chosen for the disturbances and only $J(C - 1)$ of the variance terms corresponding to the random parameters are estimable.

   If a random parameter is placed on a characteristic of the decision-maker that is continuous, it necessarily must be interacted with an alternative-specific variable (otherwise it will cancel out when the differences in utility are taken). The normalization of such parameters then depends on the type of variable with which it interacts. In general, if the characteristic interacts with alternative-specific or nest-specific binary variables, then at most one additional disturbance might be identified over the analogous model form without the interaction with the characteristic variable. For example, if the characteristic interacts with alternative-specific dummy variables, then the model is similar to the heteroskedastic case, except that for $J \geq 3$ a variance term can be identified for all J alternatives. For characteristics that are categorical variables, irrespective of the interaction structure a reference level must be chosen for the disturbances, and only $(C - 1)$ of the random parameters can be identified per interaction, where C denotes the number of levels to the categorical variable.

4. *Extensions to panel data:* For heteroskedastic, nested and cross-nested models, the use of panel data and a model with agent effects can result in the identification of at most one additional parameter over an equivalent model with cross-sectional data and alternative-specific effects. For the random parameters model, the use of panel data and agent effects does not change the identification problem: continuous attributes are theoretically always identifiable, and the same conditions hold for categorical attributes, and continuous and categorical characteristics as with cross-sectional data.

For multinomial probit models where the error structure is specified using the factor analytic form, the identification problem can be reduced to finding a solution to each of the unknown parameters contained in the following equation:

$$\Delta \mathbf{F_n} \boldsymbol{\Upsilon} \boldsymbol{\Upsilon}' \mathbf{F_n}' \Delta' = \boldsymbol{\Omega}_{\Delta \mathbf{u_n}, \Delta \mathbf{u_n}} - \boldsymbol{\Omega}_{\Delta \mathbf{u_n}, \mathbf{x_n}} \boldsymbol{\Omega}^{-1}_{\mathbf{x_n}, \mathbf{x_n}} \boldsymbol{\Omega}_{\mathbf{x_n}, \Delta \mathbf{u_n}} \tag{43}$$

$$\Rightarrow \boldsymbol{\Omega}(\boldsymbol{\theta}) = \boldsymbol{\Omega} \tag{44}$$

, where as before the left hand side of the equation contains the function $\boldsymbol{\Omega}(\boldsymbol{\theta})$ of the unknown parameters $\boldsymbol{\theta} = \boldsymbol{\Upsilon}$, and the right hand side comprises the $(J-1)\times(J-1)$ symmetric matrix $\boldsymbol{\Omega}$ of known values. The same rules of identification hold as for the mixed logit model. For multinomial probit models not specified using the factor analytic specification, as is often the case in the literature, the reader is referred to Bunch (1991) and Train (2009, Chapter 5) for a discussion on how to establish identifiability. In general, these different forms have served as prototypes for most error structures commonly employed in the literature on HCMs and the reader should be able to use these findings to ascertain identification of more general forms that combine one or more features from these prototypes.

4. Empirical Identification

Since the definition of theoretical identification rests on the availability of an infinite number of observations, it has its limitations. A model that is theoretically identified may often be empirically unidentified due to insufficient variability in the observed data. The flexibility offered by HCMs should be used with caution. If the dataset is not rich enough to support models with a high degree of complexity, multiple model specifications can result in the same improvement in fit (McFadden and Train, 2000), and in some cases this can even result in empirically unidentified models (Walker, 2001). When working with HCMs, it is helpful to have a prior idea of the sample size required to support models of a particular degree of complexity. One of the ways in which a reasonable sample size can be determined for any hypothesized model form is through a Monte Carlo experiment. For more details, the reader is referred to Muthén and Muthén (2002).

A second source of empirical unidentification is multicollinearity. Multicollinearity occurs when two or more explanatory variables in the model are strongly correlated and provide redundant information about the behavior of interest, e.g. travel times and travel costs in travel demand models. Any data sample will always contain some degree of multicollinearity, and it is up to the analyst to decide a tolerable limit. Though a high degree of multicollinearity can lessen the reliability of parameter estimates and the accompanying statistical inference, the exclusion of partially redundant variables from the analysis can also compromise the objectives of the study, and finding a balance between the two isn't always straightforward.

Lastly, the models that we have so far examined have made strong assumptions about linearity, additivity and, in the case of the structural equations sub-model, normality. Violations of these

assumptions, or omission and/or incorrect inclusion of important factors, variables or causal paths, may result in empirically unidentified model forms. One of the ways in which any hypothesized model form can be checked for misspecifications is through an outlier analysis. Outliers are data points that deviate markedly from other data points in an observed sample (Grubbs, 1969). To detect outliers, the hypothesized model specification is estimated on the complete sample. The probability of observing each data point in the observed sample is subsequently calculated assuming that the estimated model is the true underlying model. Data points for which the predicted probabilities lie below some predetermined threshold are labeled outliers. Outliers can often occur randomly due to chance deviations in natural populations. In some circumstances though, the outliers may exhibit a systematic trend, and the analyst should check that the theoretical assumptions underlying the model specification are credible. The distinction between systematic and random errors is not always clear, and the analyst should have valid reasons for excluding any data points. For a recent discussion on how to deal with outliers in discrete choice models, the reader is referred to Campbell et al. (2010).

5. Estimation Methods

Unfortunately, a general framework of theoretical and empirical identification that is readily practicable remains elusive. Throughout this chapter, we've addressed the identification problem for a select subset of model forms that are linear and additive, and conform to certain distributional assumptions. Establishing identifiability even under these restrictive assumptions can be a challenge. As analysts start to relax some of these restrictions, model specifications can grow increasingly complex, to the point where it is virtually impossible to analyze the covariance structure to determine whether the parameters are identifiable, or to predict a priori whether a particular dataset will be able to support such complexity.

Estimation methods can provide insights into the identification status of a model that would otherwise be unavailable from more theoretical procedures. For instance, if the estimation routine for a given model specification fails to converge to a solution, and the Hessian matrix at the optimum is singular or ill conditioned (resulting in absurdly large standard errors), the model may be theoretically or empirically unidentified. If the model is estimated but the parameters lie outside the range of reasonable values, the model may again be unidentified. One of the ways in which an analyst can check for identification is to estimate the model multiple times, employing different starting values for the parameters for each estimation run. If the estimation routine consistently converges to the same solution, the analyst can be reasonably confident that the model is identified and, just as importantly, that the solution is a global maxima[1]. Alternatively, if the analyst is interested in the identification status of a specific parameter, it might be helpful to fix the parameter value to some arbitrary, often unreasonable, value (Hayduk, 1988). If the log-likelihood at convergence does not change with the addition of the constraint, then the parameter is probably unidentified and the log-likelihood is flat along the direction of that parameter.

In most cases, the likelihood function for HCMs comprises a multi-dimensional integral that does not have a closed form solution and cannot be approximated using Gaussian quadrature methods, and estimation routines usually rely on Monte Carlo simulation to numerically approximate the integral. Though simulation allows for the estimation of more flexible model forms, simulation

[1] This latter result is particularly useful for HCMs that often exhibit irregularly shaped likelihood functions and multiple local maxima. Of course, the analyst can never be absolutely certain that the solution is a global maxima, but the probability that it is a global maxima is certainly higher if repeated runs converge to the same set of values.

noise leads to biased parameter estimates (Walker, 2001) and may mask identification problems inherent in the model (Chiou and Walker, 2007). This is particularly a problem in the case of HCMs because these models usually require additional or more extensive simulation routines.

With regards to simulation bias, the number of draws must rise with sample size at a sufficiently fast rate for the parameters to asymptotically converge to their true values (Train, 2009). Since the appropriate number of draws is a function of the model structure and observed data, there is no way to know a priori what an appropriate number might be. It has been suggested that the analyst estimate the model multiple times, using different starting values and increasing the number of draws with each subsequent run (Hensher and Greene, 2003). If the parameters remain relatively stable, then the analyst can be fairly confident that the estimation routine has converged to the true solution. The definition of what constitutes stable is of course subjective, but an oft-used thumb rule requires that parameter estimates lie within one standard error of each other over subsequent runs with increasing number of draws. For a recent discussion on the issue of simulation bias in discrete choice models, the reader is referred to Bastin and Cirillo (2010).

The number of draws also plays an important role in masking identification. Often, unidentified models estimated with a small number of draws appear to be identified in that the Hessian is non-singular and well conditioned. As the dimension of the problem increases, the number of draws required to adequately cover the dimension space also increases. Consequently, for unidentified models the estimation routine may break down only when the number of draws is high enough, where high enough could be any number between 100 and 10000, and maybe even higher. For more details on the masking effect of simulation noise on identification, the reader is referred to Chiou and Walker (2007).

To summarize, estimation methods can provide an additional source of information on the identifiability of HCMs. Run-time symptoms of unidentified models include high standard errors, unstable and/or unreasonable parameter estimates with increasing number of draws, singular or ill-conditioned Hessian, etc. However, due to the confounding effect of simulation noise, they can also be misleading at times. In general, it is good practice to establish identification using one of the techniques presented in the previous sections, and estimation methods should only be used as supplementary tools.

6. Case Study

In this section, we present estimation results for a stated preference dataset of travel mode choice to illustrate some of the issues that can arise in practice. The dataset for our analysis was collected as part of a series of experiments conducted at the Experimental Social Sciences Laboratory (XLAB) in the Haas Business School at the University of California, Berkeley. The experiments sought to assess the impact of information provision on various aspects of travel behavior. The kinds of information offered ranged from service reliability and greenhouse gas emissions to health benefits and peer behavior, and the dimensions of travel behavior studied included vehicle ownership, route choice and travel mode choice. More details on the experiments can be found in Gaker et al. (2011).

The particular dataset that we use here corresponds to the travel mode choice experiment. Survey respondents were asked to choose a travel mode for some hypothesized trip given the travel times and travel costs of the different modal alternatives, and the greenhouse gas emissions associated with each mode. The original dataset comprised 1670 observations made by 334 undergraduate

Model:	1-1	1-2	1-3	1-4
Identification Status:	Unidentified		Identified	
Parameter	Est. (SE)	Est. (SE)	Est. (SE)	Est. (SE)
Utility Specification				
Alt. specific constants				
Auto	0.00 (-)	0.00 (-)	0.00 (-)	0.00 (-)
Bus	0.71 (0.20)	0.71 (0.20)	0.71 (0.20)	0.71 (0.20)
Train	0.64 (0.18)	0.64 (0.18)	0.64 (0.18)	0.64 (0.18)
Bike	0.40 (0.30)*	0.40 (0.30)*	0.40 (0.30)	0.40 (0.30)
Emissions (lbs. of CO_2)	-0.02 (0.05)*	-0.02 (0.05)*	-0.02 (0.05)*	-0.02 (0.05)*
Environmental Attitudes x Emissions (lbs. of CO_2) (β)	-0.07 (na)	-0.07 (0.03)	-0.04 (0.02)	**-1.00 (-)**
Latent Variable – Pro-Environmental Attitudes				
Standard deviation (ϕ)	1.00 (na)	**1.00 (-)**	1.66 (0.10)	0.06 (0.03)
Indicator - We should raise the price of gasoline to reduce congestion and air pollution				
Factor loading (λ)	1.66 (na)	1.66 (0.10)	**1.00 (-)**	25.10 (11.3)
Standard deviation	0.56 (0.27)	0.56 (0.28)	0.56 (0.28)	0.56 (0.28)
$\lambda\phi$	1.66	1.66	1.66	1.66
$\beta\phi$	-0.07	-0.07	-0.07	-0.07
Simulated log likelihood	-1522.40	-1522.40	-1522.40	-1522.40

* Insignificant (5% level of significance)
1000 pseudo random draws

Table 1: Discrete choice model of travel model choice with a latent variable denoting attitudes towards the environment

students from the university, such that each respondent was presented with five different scenarios and the alternatives for any single scenario included three of seven pre-defined travel modes. We will be restricting attention to a subsample of 501 observations made by 306 respondents, where the number of observations for any single respondent in the subsample may vary between one and three, and the alternatives for any single scenario may include any three of the following four travel modes: auto, bus, train and bike. We excluded choice situations that featured any one of the three other alternatives in the original dataset to keep the model specification deliberately sparse and to more clearly emphasize potential identification issues.

The application concerns an HCM with a multinomial logit kernel and a latent characteristic denoting pro-environmental attitudes. The latent characteristic is operationalized via a single response asking for agreement on a scale of 1 to 7 with the attitudinal statement, "We should raise the price of gasoline to reduce congestion and air pollution," where a higher response indicates stronger agreement. The latent variable enters the choice model through an interaction with greenhouse gas emissions for each travel mode. The structural component of the structural equations sub-model comprises a normally distributed random factor with mean zero and standard deviation that needs to be estimated.

Table 1 enumerates the estimation results for models with different sets of constraints. Model 1-1 is the unconstrained partially identified model. The scale of the latent variable has not been fixed, which results in a singular hessian and unreasonable standard errors for the factor loading (λ), the standard deviation of the latent variable (ϕ) and the coefficient on the latent variable in the utility specification (β). The literature on HCMs prescribes two general methods for setting the scale of the latent variable, covered in Appendix A: either by constraining the standard deviation of the latent variable or by constraining one of the factor loadings on the indicators (Raveau et al.; 2012). Model 1-2 sets the scale of the latent variable by constraining the standard deviation of the latent variable ϕ to 1 and Model 1-3 by constraining the factor loading λ to 1. There is in fact a third way of setting the scale of the latent variable that isn't mentioned in the literature or covered by Appendix A: by constraining the coefficient on the latent variable β, as demonstrated by Model 1-3, which fixes it to -1.

The framework of identification presented in this chapter breaks the HCM into three sub-models, and analyzes the covariance matrix for each of these sub-models in isolation. While such an approach is algebraically convenient, it is oblivious to the additional information that would be available from an analysis of the covariance matrix of observable variables from different sub-models. For instance, for the example discussed here the measurement component of the structural equations sub-model, when reformulated as a confirmatory factor analytic model, comprises a single equation and two unknowns, and thereby fails the order condition. And yet the model is identifiable. This is because the covariance between the indicator and the differences in utilities provides an additional independent equation in the term $\beta\phi$ that allows identification of Models 1-2 through 1-4 (and a third way for setting the scale on the latent variable). In fact, only the terms $\lambda\phi$ and $\beta\phi$ are identifiable and not the three parameters λ, β and ϕ. Had the additional equation not been available, Models 1-2 through 1-4 would not have been identified and the analyst would be required to impose an additional constraint. This example serves to demonstrate the limitations of the set of sufficient but not necessary conditions of identification developed in this chapter, and offers an exciting direction for future research on the subject.

7. Conclusions

The HCM has gained currency over the last decade with empirical studies examining different aspects of individual behavior. The HCM combines the simplicity of random utility maximization, or discrete choice, models that belong to the GEV family, such as the multinomial logit and nested logit models, with the flexibility offered by mixed logit models and the behavioral richness of latent variable models. Notwithstanding the popularity of the HCM, questions concerning its identification remain outstanding in the literature. In particular, the identification problem has been explored in detail for many special cases but a general framework of identification has been found wanting.

In this chapter, we combined literature from the fields of discrete choice analysis and structural equation models to develop a set of sufficient conditions for theoretical identification of HCMs. The procedure for establishing identification began by decomposing the HCM into three constituent sub-models: a confirmatory factor analytic model, a structural equations model with observable variables, and a discrete choice model with exogenous observable variables. We employed a general framework of identification based on the analysis of the covariance matrix of observable data, and applied this framework to each of the three components. Wherever applicable, alternative rules that can provide quicker checks on identification were also presented. Though we focused our attention on HCMs that combine mixed logit models with choice and latent variable models, the framework of theoretical identification can be extended to incorporate multinomial probit, latent classes, multiple datasets, and dynamic choice models. Next, we looked at the issue of empirical under identification, highlighting problems with the model and/or the dataset that may result in empirically unidentified models. In some cases, estimation methods can provide a useful supplement to more rigorous theoretical procedures. We discussed some of the more popular estimation techniques for determining model identifiability, and their limitations. Finally, we looked at a case study on travel mode choice to demonstrate how identification issues may manifest themselves in practice, and how they might be suitably addressed.

One of the limitations of the framework of theoretical identification developed in this chapter is that it provides a set of sufficient but not necessary conditions of identification that are based on separating the model into smaller components. While such an approach is mathematically more practicable, it ignores additional information offered by the covariances between observable variables from different components of the model. Consequentially, models that may fail these conditions may still be identified, as demonstrated in Section 6. Future research should attempt to develop a set of sufficient and necessary conditions for identification based on an analysis of the covariance matrix of the HCM as a whole.

The exponential growth in computational power has engendered a commensurate explosion in the complexity of models being employed by studies on discrete choice analysis. The models examined explicitly in this chapter comprise but a small subset of the full range of choice models at the analyst's disposal, but the methods described in the chapter can be used to verify identification of any general model form. However, as models grow increasingly complex so does the identification problem, and establishing identification needn't always be straightforward. Nevertheless, it is imperative that the analyst verify that a model is identified before proceeding forward with estimation and inference. If used appropriately, HCMs can be powerful tools for studying individual behavior.

Acknowledgements

This research was funded by the National Science Foundation, the University of California Transportation Center Dissertation Grant and the Multi-Campus Research Programs and Initiatives. We wish to thank David Gaker and Dave Vautin for allowing us access to the data that we used as part of our case study. We would also like to express our thanks to an anonymous referee whose comments and suggestions helped us improve the quality of our presentation.

References

Bastin, F., and Cirillo, C. (2010), "**Reducing simulation bias in mixed logit model estimation**," *Journal of Choice Modelling*, Vol. 3, No. 2, pp. 71-88.

Ben-Akiva, M., and Lerman, S. (1985), "**Discrete choice analysis: Theory and application to travel demand**," MIT Press: Cambridge, MA.

Ben-Akiva, M., Walker, J. L., Bernardino, A. T., Gopinath, D. A., Morikawa, T., and Polydoropoulou, A. (2002), "**Integration of choice and latent variable models**," in: Hani S. Mahmassani (ed.): *In perpetual motion: Travel behavior research opportunities and application challenges*, Elsevier, Amsterdam, pp. 431–470.

Bolduc. D. (1999), "**A practical technique to estimation of multinomial probit models in transportation**," *Transportation Research B: Methodological*, Vol. 33, pp. 63–79.

Bollen, K. A. (1989), "**Structural Equations with Latent Variables**," *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons.

Campbell, D., and Hess, S. (2009), "**Outlying sensitivities in discrete choice data: Cause, consequences and remedies**," *European Transport Conference*, Amsterdam

Carrasco, J. A., and Ortúzar, J. de D. (2002), "**Review and assessment of the nested logit model,**" *Transport Reviews: A Transnational Transdisciplinary Journal*, Vol. 22, No. 2, pp. 197-218.

Chiou, L., and Walker, J. L. (2007), "**Masking identification of discrete choice models under simulation methods**," *Journal of Econometrics*, Vol. 141, pp. 683-703.

Daziano, R. A., and Bolduc, D. (2012), "**Covariance, identification, and finite sample performance of the MSL and Bayes estimators of a logit model with latent variables**," *Transportation*.

Fisher, F. M. (1976), "**The identification problem in econometrics**," New York: McGraw-Hill.

Gaker, D., Vautin, D., Vij, A., and Walker, J. L. (2011), "**The power and value of green in promoting sustainable transport behavior**," *Environmental Research Letters,* Vol. 6, 034010.

Grubbs, F. E. (1969), "**Procedures for detecting outlying observations in samples**," *Technometrics*, Vol. 11, pp. 1–21.

Hannemann, R. A. (2000), "**Structural equation models: Identification issues**," *Sociology 203B*, http://faculty.ucr.edu/~hanneman/soc203b/lectures/identify.html.

Hayduk, L. A. (1988), "**Structural equation modeling with LISREL: Essentials and advances**," Johns Hopkins University Press, Baltimore, MD, 1987.

Hensher, D., and Greene, W. H. (2003), "**The mixed logit model: The state of practice**," *Transportation*, Vol. 30, No. 2, pp. 133–176.

Kenny, D. A. (1979), "C**orrelation and causality**," New York: Wiley.

Manski, C. F. (1995), "**Identification Problems in the Social Sciences**," Harvard University Press.

McFadden, D. L. (1986), "**The choice theory approach to marketing research**," *Marketing Science*, Vol. 5, No. 4, pp. 275–297.

McFadden, D., and Train, K. (2000), "**Mixed MNL models for discrete response**," *Journal of Applied Econometrics,* Vol. 15, pp. 447-470.

Muthén, L. K, and Muthén, B. (2002), "**How to use a Monte Carlo study to decide on sample size and determine power**," *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 9, No. 4, pp. 599-620.

O'Brien, R. M. (1994), "**Identification of Simple Measurement Models with Multiple Latent Variables and Correlated Errors**," *Sociological Methodology*, Vol. 24, pp. 137-170

Raveau, S., Yáñez, M. F., and Ortúzar, J. de D. (2012), "**Practical and empirical identifiability of hybrid discrete choice models**," *Transportation Research Part B: Methodological*, Vol. 46, No. 10, pp. 1374-1383.

Reilly, T. (1995), "**A necessary and sufficient condition for identification of confirmatory factor analysis models of factor complexity one**," *Sociological Methods and Research*, Vol. 23, pp. 421-441.

Reilly, T., and O'Brien, R. M. (1996), "**Identification of confirmatory factor analysis model of arbitrary complexity: The side-by-side rule**," *Sociological Methods and Research*, Vol. 23, pp. 473-491.

Rigdon, E. E. (1995), "**A necessary and sufficient identification rule for structural models estimated in practice**," *Multivariate Behavioral Research*, Vol. 30, No. 3, pp. 359-383.

Temme, D., Paulssen, M., and Dannewald, T. (2008), "**Incorporating latent variables into discrete choice models – A simultaneous estimation approach using SEM software**", *BuR - Business Research*, Vol. 1, No. 2, pp. 220-237.

Tudela, A., Habib, K. M. N., Carrasco, J. A., and Osman, A. O. (2011), "**Incorporating the explicit role of psychological factors on mode choice: A hybrid mode choice model by using data from an innovative psychometric survey**," Second International Choice Modeling Conference, July 2011, Leeds, UK.

Train, K. E. (2009), "**Discrete choice models with simulation**," Cambridge University Press, Cambridge.

Walker, J. L. (2001), "**Extended discrete choice models: integrated framework, flexible error structures, and latent variables**," *PhD dissertation*, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Walker, J., Ben-Akiva, M. (2002), "**Generalized random utility model**," *Mathematical Social Sciences*, Vol. 43, pp. 303-343.

Walker, J. L., Ben-Akiva, M., and Bolduc, D. (2007), "**Identification of Parameters in Normal Error Component Logit-Mixture (NECLM) Models**," *Journal of Applied Econometrics*, Vol. 22, pp. 1095-1125.

Zhao, J. (2009), "**Preference accommodating and preference shaping: Incorporating traveler preferences into transportation planning**," *PhD Dissertation*, Massachusetts Institute of Technology.

Appendix A: Applying the General Rules of Identification to a Confirmatory Factory Analytic Model

In Section 3.2.1, we stated that the identification problem for a confirmatory factor analytic model can be reduced to finding constraints that ensure a solution to the following system of nonlinear equations:

$$\Rightarrow \boldsymbol{\Omega} = \boldsymbol{\Lambda\Phi\Lambda'} + \mathbf{D\Theta\Theta'D'}$$

The rules of identification presented in Section 3.1 may now be applied to the above equation to verify identifiability. The general approach requires the analyst to be able to express all of the unknown parameters in $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ as some function of the elements of the sample covariance matrix $\boldsymbol{\Omega}$. Section A.1 uses this approach to demonstrate why the analyst needs to impose constraints to set the scale of the latent variables, and how this might be accomplished. Section A.2 applies equation (19) to evaluate identifiability of the confirmatory factor analytic model shown in Figure 1B.

A.1 The Location and Scale of the Latent Variable

One of the first steps to ensuring identifiability of any confirmatory factor analytic model is to establish the location and scale of each latent variable included in the model specification. Since the indicator responses are usually normalized around the mean, the location of the latent variables is implicitly set to zero. This still leaves the analyst the task of imposing constraints that set the scale of the latent variables. To illustrate why this is necessary, we consider the confirmatory factor analytic model shown in Figure 1A. The model consists of a single latent

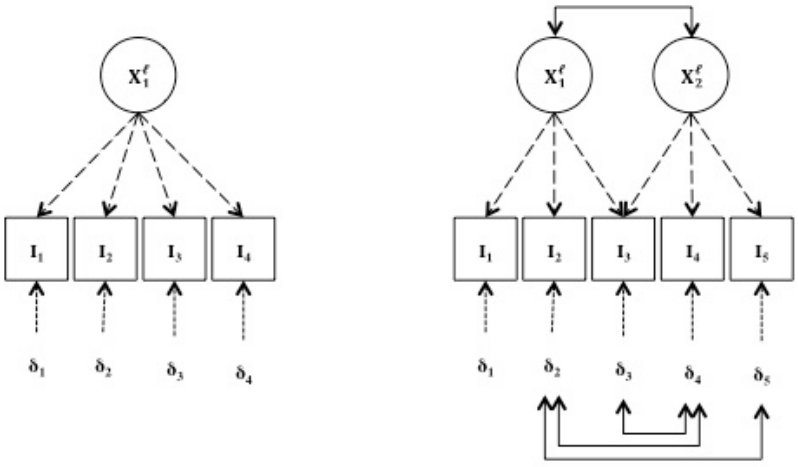

Figure 1: Examples of confirmatory factor-analytic models: **(A)** A model with a single latent variable loaded on by four uncorrelated indicators; **(B)** A model with two latent variables, five partially correlated indicators and factor complexity two

variable $\mathbf{X_1^{\ell}}$ loaded on by four uncorrelated indicators $\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}$ and $\mathbf{I_4}$. The parameters for the model are given as follows:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix}, \ \mathbf{\Phi} = \begin{bmatrix} \phi_{11} \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathbf{\Theta} = \begin{bmatrix} \varphi_1 & 0 & 0 & 0 \\ 0 & \varphi_2 & 0 & 0 \\ 0 & 0 & \varphi_3 & 0 \\ 0 & 0 & 0 & \varphi_4 \end{bmatrix}$$

, where the parameter λ_{i1} denotes the factor loading of indicator $\mathbf{I_i}$ on the latent variable $\mathbf{X_1^{\ell}}$; the parameter ϕ_{11} denotes the variance of the latent variable $\mathbf{X_1^{\ell}}$; and the parameters along the diagonal of $\mathbf{\Theta}$ denote the standard deviations of the measurement errors corresponding to each of the four indicators. Substituting the expressions for $\mathbf{\Lambda}$, $\mathbf{\Phi}$, $\mathbf{D}$ and $\mathbf{\Theta}$ in equation (18), we get:

$$\mathbf{\Omega}(\theta) = \begin{bmatrix} \lambda_{11}^2\phi_{11} + \varphi_1^2 & & & \\ \lambda_{11}\lambda_{21}\phi_{11} & \lambda_{21}^2\phi_{11} + \varphi_2^2 & & \\ \lambda_{11}\lambda_{31}\phi_{11} & \lambda_{21}\lambda_{31}\phi_{11} & \lambda_{31}^2\phi_{11} + \varphi_3^2 & \\ \lambda_{11}\lambda_{41}\phi_{11} & \lambda_{21}\lambda_{41}\phi_{11} & \lambda_{31}\lambda_{41}\phi_{11} & \lambda_{41}^2\phi_{11} + \varphi_4^2 \end{bmatrix}$$

Holding $\mathbf{\Theta}$ constant, for any $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ that result in a particular outcome for $\mathbf{\Omega(\theta)}$, $\mathbf{\Lambda}/2$ and $4\mathbf{\Phi}$ result in the same outcome. Therefore, the model is theoretically unidentified and some constraints need to be imposed to set the scale of the latent variable. The nature of the identification problem is such that the scale can be set in one of multiple ways: by setting the variance of the latent variable to a constant such as one (by constraining the appropriate diagonal element of $\mathbf{\Phi}$, the covariance matrix of the latent variables), or by scaling it to any one of the observed indicators by constraining some λ_{ij} coefficient, usually to one. In most cases, the choice of constraint is trivial, and the analyst is free to choose whichever constraint is most convenient from the standpoint of estimation. Usually, the scale is set by constraining the factor loading of an indicator that is strongly related to the latent variable. An advantage of this approach is that the latent variable has the same units as the indicator, and is easier to interpret. Once the scale for each latent variable has been set, the analyst should verify that the other model parameters are also identifiable.

A.2 A More Complicated Example

Consider, for the sake of illustration, the confirmatory factor analytic model shown in Figure 1B (same as the model from Figure 4B). As we shall show in this Section, the model is just identifiable. The model consists of two correlated latent factors: $\mathbf{X_1^{\ell}}$ and $\mathbf{X_2^{\ell}}$, five partially correlated indicator measures: $\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}, \mathbf{I_4}$ and $\mathbf{I_5}$, and has a factor complexity of two, where

factor complexity is defined as the maximum number of latent variables loaded on by a single indicator. The parameters for the model are given as follows:

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & \lambda_{32} \\ 0 & \lambda_{42} \\ 0 & 1 \end{bmatrix}, \boldsymbol{\Phi} = \begin{bmatrix} \phi_{11} & \phi_{21} \\ \phi_{21} & \phi_{22} \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\Theta} = \begin{bmatrix} \varphi_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \varphi_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \varphi_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \varphi_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \varphi_5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \varphi_6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \varphi_7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \varphi_8 \end{bmatrix}$$

, where the parameter λ_{ij} belonging to $\boldsymbol{\Lambda}$ denotes the factor loading of indicator i on latent variable j; the parameter ϕ_{ij} belonging to $\boldsymbol{\Phi}$ denotes the covariance between latent variables i and j; the parameters along the diagonal of $\boldsymbol{\Theta}$ denote the standard deviations of the independent factors in $\boldsymbol{\eta}$; the first five columns of $\mathbf{D}$, and the corresponding parameters $\varphi_1, \varphi_2, \varphi_3, \varphi_4$ and φ_5 belonging to $\boldsymbol{\Theta}$, allow for heteroscedastic error components across the five indicator constructs; and the last three columns of $\mathbf{D}$, along with the corresponding parameters φ_6, φ_7 and φ_8 belonging to $\boldsymbol{\Theta}$, capture correlation between the three pairs of indicators $(\mathbf{I_2}, \mathbf{I_4})$, $(\mathbf{I_3}, \mathbf{I_4})$ and $(\mathbf{I_2}, \mathbf{I_5})$, respectively. The scale of the latent variables $\mathbf{X_1^{\ell}}$ and $\mathbf{X_2^{\ell}}$ is set by constraining $\lambda_{11} = 1$ and $\lambda_{52} = 1$, respectively. In all, there are 15 unknown parameters. Substituting the expressions for $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, $\mathbf{D}$ and $\boldsymbol{\Theta}$ in equation (19), we get:

$$\boldsymbol{\Omega} = \begin{bmatrix} \phi_{11} + \varphi_1^2 & & & & \\ \lambda_{21}\phi_{11} & \lambda_{21}^2\phi_{11} + \varphi_2^2 + \varphi_6^2 + \varphi_8^2 & & & \\ \lambda_{31}\phi_{11} + \lambda_{32}\phi_{21} & \lambda_{21}\lambda_{31}\phi_{11} + \lambda_{21}\lambda_{32}\phi_{21} & \lambda_{31}^2\phi_{11} + 2\lambda_{31}\lambda_{32}\phi_{21} + \lambda_{32}^2\phi_{22} + \varphi_3^2 + \varphi_7^2 & & \\ \lambda_{42}\phi_{21} & \lambda_{21}\lambda_{42}\phi_{21} + \varphi_6^2 & \lambda_{31}\lambda_{42}\phi_{21} + \lambda_{32}\lambda_{42}\phi_{22} + \varphi_7^2 & \lambda_{42}^2\phi_{22} + \varphi_4^2 + \varphi_6^2 + \varphi_7^2 & \\ \phi_{21} & \lambda_{21}\phi_{21} + \varphi_8^2 & \lambda_{31}\phi_{21} + \lambda_{32}\phi_{22} & \lambda_{42}\phi_{22} & \phi_{22} + \varphi_5^2 + \varphi_8^2 \end{bmatrix}$$

Let ω_{ij} denote the element of the population covariance matrix $\boldsymbol{\Omega}$ that lies in row i and column j, i.e. $\omega_{ij} = \mathrm{Cov}(\mathbf{I_i}, \mathbf{I_j})$. If equation (19) allows for each of the 15 parameters to be expressed as some function of the ω_{ij}'s, then the model as a whole is identifiable. There are 15 unknown parameters and at most 15 independent equations. Therefore, the model satisfies the order condition. To start, note that $\phi_{21} = \omega_{51}$, and so ϕ_{21} is identifiable (and it equals $\omega_{51} = \mathrm{Cov}(\mathbf{I_5}, \mathbf{I_1})$). Then λ_{42} and ϕ_{22} may be calculated as follows:

$$\lambda_{42}\phi_{21} = \omega_{41} \Rightarrow \lambda_{42} = \omega_{41}/\omega_{51}$$

$$\lambda_{42}\phi_{22} = \omega_{54} \Rightarrow \phi_{22} = \omega_{54}\omega_{51}/\omega_{41}$$

Similarly, λ_{21} can be identified from the following pair of equations:

$$\lambda_{21}\lambda_{31}\phi_{11} + \lambda_{21}\lambda_{32}\phi_{21} = \omega_{32} \text{ and } \lambda_{31}\phi_{11} + \lambda_{32}\phi_{21} = \omega_{31} \Rightarrow \lambda_{21} = \omega_{32}/\omega_{31}$$

$$\Rightarrow \phi_{11} = \omega_{21}\omega_{31}/\omega_{32}$$

Lastly, the parameters λ_{31} and λ_{32} may be identified using the following pair of equations:

$$\lambda_{31}\phi_{11} + \lambda_{32}\phi_{21} = \omega_{31}$$

$$\lambda_{31}\phi_{21} + \lambda_{32}\phi_{22} = \omega_{53}$$

Once all of the elements of $\boldsymbol{\Phi}$ and $\boldsymbol{\Lambda}$ have been identified, we can turn our attention to $\boldsymbol{\Theta}$. The covariances between the indicators φ_6, φ_7 and φ_8 can be solved using the three equations:

$$\lambda_{21}\lambda_{42}\phi_{21} + \varphi_6^2 = \omega_{42}$$

$$\lambda_{31}\lambda_{42}\phi_{21} + \lambda_{32}\lambda_{42}\phi_{22} + \varphi_7^2 = \omega_{43}$$

$$\lambda_{21}\phi_{21} + \varphi_8^2 = \omega_{52}$$

We skip enumerating the equations, but the remaining φ's can also be identified from the five elements in $\boldsymbol{\Omega}(\boldsymbol{\theta})$ along the diagonal. Therefore, all of the parameters are identifiable. In fact, the model is "just-identified" in that the number of unknown parameters exactly equals the number of equations, and the model has zero degrees of freedom.

Appendix B: Applying the Rank and Order Conditions of Identification to a Structural Equations Model with Observable Variables

To illustrate how the order and rank conditions might be applied, we examine the nonrecursive model of Figure 1, taken from Hanneman (2000). The reader should recognize that the model is the same as the one shown in Figure 4C. The model parameters are as follows:

$$\mathbf{B} = \begin{bmatrix} 0 & \beta_{12} & \beta_{13} \\ \beta_{21} & 0 & \beta_{23} \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \\ 0 & 0 \end{bmatrix}, \mathbf{\Phi} = \begin{bmatrix} \phi_{11} & \\ \phi_{21} & \phi_{22} \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}, \mathbf{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_6 \end{bmatrix}$$

, where the parameter β_{ij} belonging to **B** denotes the coefficient of the j^{th} endogenous variable in the equation for the i^{th} endogenous variable; the parameter γ_{ij} belonging to **Γ** denotes the coefficient of the j^{th} exogenous variable in the equation for the i^{th} endogenous variable; the parameter ϕ_{ij} belonging to **Φ** denotes the covariance between exogenous variables i and j; the parameters along the diagonal of **Ψ** denote the standard deviations of the independent factors in

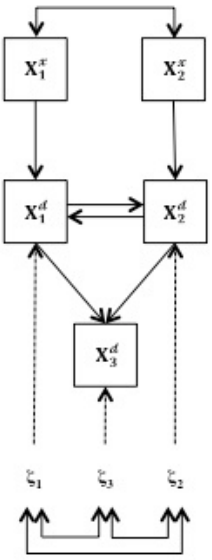

Figure 1: A nonrecursive structural equations model with observable variables

$\boldsymbol{\eta}$; the first three columns of $\mathbf{G}$, and the corresponding parameters ψ_1, ψ_2 and ψ_3 belonging to $\boldsymbol{\Psi}$, allow for heteroscedastic error components across the three endogenous variables; and the last three columns of $\mathbf{G}$, along with the corresponding parameters ψ_4, ψ_5 and ψ_6 belonging to $\boldsymbol{\Psi}$, capture correlation between the three pairs of endogenous variables $\left(\mathbf{X_1^d}, \mathbf{X_2^d}\right)$, $\left(\mathbf{X_2^d}, \mathbf{X_3^d}\right)$ and $\left(\mathbf{X_1^d}, \mathbf{X_3^d}\right)$, respectively.

To check the order condition: $\mathrm{L}^{\boldsymbol{d}} - 1 = 2$, and the order condition is satisfied if each of the equations corresponding to the three endogenous variables excludes at least two of the remaining four exogenous and endogenous variables. It can be seen from the path diagram itself that the equation for $\mathbf{X_1^d}$ excludes $\mathbf{X_3^d}$ and $\mathbf{X_2^x}$; the equation for $\mathbf{X_2^d}$ excludes $\mathbf{X_3^d}$ and $\mathbf{X_1^x}$; and the equation for $\mathbf{X_3^d}$ excludes $\mathbf{X_1^x}$ and $\mathbf{X_2^x}$. Therefore, the order condition is satisfied. To check the rank condition, we first construct the matrix $\mathbf{A}$ for the model:

$$\mathbf{A} = \begin{bmatrix} 1 & -\beta_{12} & 0 & -\gamma_{11} & 0 \\ -\beta_{21} & 1 & 0 & 0 & -\gamma_{22} \\ -\beta_{31} & -\beta_{32} & 1 & 0 & 0 \end{bmatrix}$$

Then, the matrices $\boldsymbol{\emptyset^t}$ and $\mathbf{A}\boldsymbol{\emptyset^t}$ corresponding to each of the three equations in the endogenous variables may be written as:

$$\boldsymbol{\emptyset^1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{\emptyset^2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \boldsymbol{\emptyset^3} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A}\boldsymbol{\emptyset^1} = \begin{bmatrix} 0 & 0 \\ 0 & -\gamma_{22} \\ 1 & 0 \end{bmatrix}, \mathbf{A}\boldsymbol{\emptyset^2} = \begin{bmatrix} 0 & -\gamma_{11} \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathbf{A}\boldsymbol{\emptyset^3} = \begin{bmatrix} -\gamma_{11} & 0 \\ 0 & -\gamma_{22} \\ 0 & 0 \end{bmatrix}$$

The rank of each of the three matrices $\mathbf{A}\boldsymbol{\emptyset^1}, \mathbf{A}\boldsymbol{\emptyset^2}$ and $\mathbf{A}\boldsymbol{\emptyset^3}$ is two, and the rank condition is also satisfied. Therefore, the model is identified.

Appendix C: Applying the General Rules of Identification to Mixed Logit Models

In the following subsections, we apply the rules of identification presented in Section 3.4.2 to two special cases of the mixed logit model that haven't been addressed previously in the literature. Section C.1 examines the conditions of identification as they apply to models with random parameters on explanatory variables. Section C.2 demonstrates how the identification conditions may be established when working with panel data sets and agent effects.

C.1 The Random Parameters Model

The random parameters model allows the vector of coefficients $\boldsymbol{\beta}$ to be randomly distributed across decision-makers in the sample, and is used when the analyst has reason to believe that tastes in the sample population vary with unobservable variables or purely randomly. The model formulation with normally distributed random taste parameters can be written as:

$$\mathbf{u_n} = \mathbf{X_n}\boldsymbol{\beta_n} + \boldsymbol{\nu_n}, \text{ where } \boldsymbol{\beta_n} \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}')$$

$\boldsymbol{\beta_n}$ is an $(L \times 1)$ random normal vector with mean $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'$. Substituting $\boldsymbol{\beta_n} = \boldsymbol{\beta} + \boldsymbol{\Upsilon}\boldsymbol{\eta}_L^n$, where $\boldsymbol{\Upsilon}$ is the lower triangular Cholesky decomposition of the covariance matrix of $\boldsymbol{\beta_n}$, leads to a general factor-analytic specification with $\mathbf{F_n} = \mathbf{X_n}$:

$$\mathbf{u_n} = \mathbf{X_n}\boldsymbol{\beta} + \mathbf{X_n}\boldsymbol{\Upsilon}\boldsymbol{\eta}_L^n + \boldsymbol{\nu_n}$$

The parameters to be identified are μ and the elements of $\boldsymbol{\Upsilon}$. Though the specification $\mathbf{F_n} = \mathbf{X_n}$ does slightly change the form of $\mathrm{Var}(\Delta\mathbf{u_n})$, the identification of equation (39) continues to be a sufficient condition for identification of the unknown parameters μ and the elements of $\boldsymbol{\Upsilon}$. The matrix $\boldsymbol{\Upsilon}$ is usually specified as diagonal, but it does not have to be. Also, the random distribution needn't always be normal. Alternative distributions popularly employed in the literature include lognormal, triangular, uniform, truncated normal, etc. In analyzing the covariance matrix of utility differences, we have so far assumed that the systematic portion of the utility is linearly separable from the error structure. However, with distributions such as the lognormal, the mean and the variance of the random parameter are both a function of the two disturbance parameters, and linear separability does not exist. In such a case, the covariance matrix of utility differences is no longer given by equation (39), and must be derived on a case-by-case basis.

In the two special cases analyzed so far, the matrix $\mathbf{F_n}$ was held constant across decision-makers, allowing us to restrict our attention to the covariance matrix of utility difference for a single decision-maker. However, for the random parameters model $\mathbf{F_n}$ varies across observations, and the number of independent rows in $\boldsymbol{\Omega}(\boldsymbol{\theta})$ can be as large as NJ. For these same reasons, the order condition is rarely restrictive, and in applying the rank condition one need only look at the column rank of the Jacobian. Through the following paragraphs, we work through the rules of identification for models where a random normal distribution is imposed on continuous and categorical alternative attributes and individual characteristics.

1. *Continuous attributes:* There are no identification issues per se. Data permitting, the full covariance structure (i.e. variances for each parameter as well as covariances across parameters) can be estimated.

2. *Categorical attributes:* An interesting and unintuitive identification issue arises when a categorical attribute is specified with independently distributed generic random parameters.

Say there are C categories for the variable. Assuming no correlation, there is theoretically a β_c and σ_c for each category $c = 1, \ldots, C$. However, as was mentioned in Section 3.3.1, for the systematic component β_c a reference level needs arbitrarily to be selected and only $(C-1)$ β_c's can be identified. However, this is not necessarily true for the disturbances. To illustrate this, we shall consider a two alternative example (since the number of alternatives for a random parameters model does not matter) and a categorical variable with 2 levels first, and then with 3 levels. Let x_{njp} be the p^{th} binary variable for alternative j and individual n such that x_{njp} equals 1 if the categorical variable equals p, and zero otherwise. For the two levels case, adopting the scalar notation, we specify the utility of alternative j and individual n as follows:

$$u_{nj} = \beta_{n1} x_{nj1} + \beta_{n2}\left(1 - x_{nj1}\right) + \nu_{nj}$$

$$\Rightarrow \mathbf{F} = \begin{bmatrix} x_{1n} & 1 - x_{1n} \\ x_{2n} & 1 - x_{2n} \end{bmatrix}, \text{ and } \Upsilon = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

$$\Rightarrow \boldsymbol{\Omega}_n(\theta) = \left(x_{1n} - x_{2n}\right)^2 \left(\sigma_1^2 + \sigma_2^2\right) + 2g/\mu^2$$

From above, it should be apparent that only the sum $\sigma_1^2 + \sigma_2^2$ is estimable, and not the independent parameters themselves. Either parameter can be normalized to zero, or the parameters can be constrained to be the same. For the three levels case:

$$u_{nj} = \beta_{n1} x_{nj1} + \beta_{n2} x_{nj2} + \beta_{n3}\left(1 - x_{nj1} - x_{nj2}\right) + \nu_{nj}$$

$$\Rightarrow \mathbf{F_n} = \begin{bmatrix} x_{n11} & x_{n12} & 1 - x_{n11} - x_{n12} \\ x_{n21} & x_{n22} & 1 - x_{n21} - x_{n22} \end{bmatrix}, \text{ and } \Upsilon = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$$

$$\Rightarrow \boldsymbol{\Omega}_n(\theta) = \left(x_{n11} - x_{n21}\right)^2 \sigma_1^2 + \left(x_{n12} - x_{n22}\right)^2 \sigma_2^2 + \left(x_{n11} + x_{n12} - x_{n21} - x_{n22}\right)^2 \sigma_3^2 + 2g/\mu^2$$

Therefore, there is one linearly independent equation for each σ in $\boldsymbol{\Omega}_\mathbf{n}(\boldsymbol{\theta})$, i.e. all three σ parameters are identified. The reader should verify that the condition holds for all $J \geq 2$ and $C \geq 3$, i.e. a random parameter for each of the categories is theoretically identified for all $C \geq 3$.

When the categorical attribute is specified with independently distributed alternative-specific random parameters, a reference level must be chosen for the disturbances, and only $J(C-1)$ σ_{cj}'s are estimable. For example, for a model with two alternatives, and alternative-specific random parameters on a categorical attribute with 3 levels:

$$u_{nj} = \beta_{nj1} x_{nj1} + \beta_{nj2} x_{nj2} + \beta_{nj3}\left(1 - x_{nj1} - x_{nj2}\right) + \nu_{nj}$$

$$\mathbf{F_n} = \begin{bmatrix} x_{n11} & x_{n12} & 1-x_{n11}-x_{n12} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{n21} & x_{n22} & 1-x_{n21}-x_{n22} \end{bmatrix}$$

$$\Rightarrow \boldsymbol{\Omega}_n(\theta) = x_{n11}^2\sigma_1^2 + x_{n12}^2\sigma_2^2 + (1-x_{n11}-x_{n12})^2\sigma_3^2 + x_{n21}^2\sigma_4^2 + x_{n22}^2\sigma_5^2 + (1-x_{n21}-x_{n22})^2\sigma_6^2 + 2g/\mu^2$$

$$\Rightarrow \boldsymbol{\Omega}_n(\theta) = x_{n11}\sigma_1^2 + x_{n12}\sigma_2^2 + (1-x_{n11}-x_{n12})\sigma_3^2 + x_{n21}\sigma_4^2 + x_{n22}\sigma_5^2 + (1-x_{n21}-x_{n22})\sigma_6^2 + 2g/\mu^2$$

, where the last equality holds because the variables x_{njp} are binary. Therefore, there are four linearly independent equations in the six σ's, and normalization needs to be imposed on one of $(\sigma_1, \sigma_2, \sigma_3)$ and one of $(\sigma_4, \sigma_5, \sigma_6)$. The reader should apply the equality condition to see that the normalization can be arbitrary.

The results derived here were for a model with a single categorical variable. However, all of the results hold true for models with multiple categorical variables as well, as long as the variables themselves and the random parameters imposed on them are both independent. The extrapolation to multiple categorical variables follows from the independence assumption, which allows the analyst to break the covariance matrix into smaller pieces such that each piece corresponds to a different categorical variable. The analyst can then apply the rules of identification to the covariance matrix for each categorical variable separately.

3. *Continuous characteristics:* If a random parameter is placed on a characteristic of the decision-maker (for example, years employed), it necessarily must be interacted with an alternative-specific variable (otherwise it will cancel out when the differences are taken). The normalization of such parameters then depends on the type of variable with which it interacts. In general, if the characteristic interacts with alternative-specific or nest-specific binary variables, then at most one additional disturbance might be identified over the analogous model form without the interaction with the characteristic variable. This is because $\mathbf{F_n} = x_n\mathbf{F}$, where x_n is a characteristic variable of individual n, and $\mathbf{F}$ is the matrix containing alternative-specific and nest-specific binary variables. Then, the covariance matrix of utility differences can be given by:

$$\boldsymbol{\Omega_n}(\boldsymbol{\theta}) = \Delta\mathbf{F_n}\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'\mathbf{F_n}'\Delta' + \left(\frac{g}{\mu^2}\right)\Delta\Delta' = x_n^2\Delta\mathbf{F}\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'\mathbf{F}'\Delta' + \left(\frac{g}{\mu^2}\right)\Delta\Delta'$$

$$\Rightarrow \boldsymbol{\Omega_n}(\boldsymbol{\theta}) = x_n^2\boldsymbol{\Omega_{ec}}(\boldsymbol{\theta}) + (1-x_n^2)\left(\frac{g}{\mu^2}\right)\Delta\Delta'$$

, where $\boldsymbol{\Omega_{ec}}(\boldsymbol{\theta}) = \Delta\mathbf{F}\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'\mathbf{F}'\Delta' + (g/\mu^2)\Delta\Delta'$ is the covariance matrix of the error structure for the analogous error components model. Therefore, the column rank of the Jacobian for random parameters on individual characteristics can be at most one more than the column rank of the Jacobian for the analogous error components case. Since Δ is a function only of the number of alternatives J and is independent of the parameters, the only additional linearly independent equation can be with regards to the unknown parameter (g/μ^2). For example, if the characteristic interacts with alternative-specific dummy variables, then the model is similar to the heteroskedastic case, (see Walker et al., 2007) except that for $J \geq 3$ a variance term can be identified for all J alternatives.

4. *Categorical characteristics:* For characteristics that are categorical variables (for example, low income, medium income, high income), irrespective of the interaction structure a reference level must be chosen for the disturbances, and only $(C-1)$ σ's can be identified per interaction. We omit the details of the proof, but it is nearly identical to the example where a categorical attribute was specified with independently distributed alternative-specific random parameters.

C.2 Extensions to Panel Data

Panel data refers to multi-dimensional data that contains observations over different time periods for each decision-maker in the sample. Identification for panel data is different from the cross-sectional case because the error components from the mixing distribution take the same value for all choice situations for a given decision-maker, whereas the Extreme Value terms are i.i.d. across decision-makers and choice situations. The mixed errors create correlation over choice situations for a given individual, which can be used for identification.

Typically, equation (1) would be modified for a panel context by adding subscripts k to denote the time period of the choice and the explanatory variables for that choice. Since identification is determined via the covariance structure, we will focus on this aspect of the formulation and modify equation (1) such that the covariance structure is a function of all utilities faced by an individual over all time periods. For simplicity, we assume the same number of time periods $(k = 1, \dots, K)$ observed for each person and a universal choice set across individuals and time periods. The covariance structure of interest for a given individual is then a function of all JK utilities that the individual faces:

$$\mathbf{u_{n,pl}} = \mathbf{X_{n,pl}}\boldsymbol{\beta} + \mathbf{F_{n,pl}}\boldsymbol{\Upsilon}\boldsymbol{\eta}_{\mathbf{n}}^{\mathbf{R}} + \boldsymbol{\nu}_{\mathbf{n,pl}}$$

$$\boldsymbol{\Omega}_{\mathbf{n,pl}}(\boldsymbol{\theta}) = \Delta_{\mathbf{pl}}\mathbf{F_{n,pl}}\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'\mathbf{F}'_{\mathbf{n,pl}}\Delta'_{\mathbf{pl}} + \Delta_{\mathbf{pl}}\left(\frac{g}{\mu^2}\right)\boldsymbol{\mathcal{I}}_{\mathbf{JK}}\Delta'_{\mathbf{pl}}$$

, where pl denotes panel data, $\mathbf{u_{n,pl}}$ and $\boldsymbol{\nu}_{\mathbf{n,pl}}$ are (JK×1) vectors, $\mathbf{X_{n,pl}}$ is a (JK×L) matrix of observed and latent explanatory variables, $\boldsymbol{\beta}$ is (L×1), $\mathbf{F_{n,pl}}$ is (JK×R), $\boldsymbol{\Upsilon}$ is (R×R), $\boldsymbol{\eta}_{\mathbf{n}}^{\mathbf{R}}$ is (R×1), $\boldsymbol{\mathcal{I}}_{\mathbf{JK}}$ is a (JK×JK) identity matrix, and $\Delta_{\mathbf{pl}}$ is $(J-1)$K×JK. The key in terms of identification is that the covariance matrix of utility differences is now of dimension $(JK-1)\times(JK-1)$, which incorporates the added correlation over choice situations for a given individual, referred to as the agent effect. The idea of an agent effect is that what is unobserved for one individual in one time period is likely the same as what is unobserved for the same individual in another time period. This is implemented by having alternative- and individual-specific covariances that are repeated in all time periods for any given individual.

For heteroskedastic, nested and cross-nested models, the use of panel data and a model with agent effects can result in the identification of at most one additional parameter over an equivalent model with cross-sectional data and alternative-specific effects. The proof is very similar to that in Section A.1 for a model with a random parameter on a continuous characteristic such that the characteristic interacts with alternative-specific or nest-specific binary variables. To illustrate this, we consider a dataset containing two observations for each individual $(K = 2)$, and the same number of alternatives J across all observations and individuals. Dropping the subscript n, the matrices $\Delta_{\mathbf{pl}}$ and $\mathbf{F_{pl}}$ can then be expressed in terms of their analogs $\Delta_{\mathbf{cs}}$ and $\mathbf{F_{cs}}$ from the equivalent cross-sectional model as follows:

$$\Delta_{\mathbf{pl}}=\begin{bmatrix}\Delta_{\mathbf{cs}} & \emptyset_{\mathbf{J-1,J}}\\ \emptyset_{\mathbf{J-1,J}} & \Delta_{\mathbf{cs}}\end{bmatrix}, \text{ and } \mathbf{F_{pl}}=\begin{bmatrix}\mathbf{F_{cs}}\\ \mathbf{F_{cs}}\end{bmatrix}$$

, where $\emptyset_{\mathbf{MN}}$ is an $(\mathrm{M}\times\mathrm{N})$ matrix of zeros, $\Delta_{\mathbf{pl}}$ is a $(\mathrm{J}-1)\mathrm{K}\times\mathrm{JK}$ block diagonal matrix formed by stacking the matrix $\Delta_{\mathbf{cs}}$ along the diagonal K times, and $\mathbf{F_{pl}}$ is a $\mathrm{JK}\times\mathrm{R}$ matrix formed from stacking $\mathbf{F_{cs}}$ vertically K times, where $\mathrm{K}=2$ in this case. Then, the covariance matrix of utility differences can be calculated as shown below:

$$\boldsymbol{\Omega}_{\mathbf{pl}}(\boldsymbol{\theta})=\begin{bmatrix}\Delta_{\mathbf{cs}}\mathbf{F_{cs}}\boldsymbol{\Upsilon}\\ \Delta_{\mathbf{cs}}\mathbf{F_{cs}}\boldsymbol{\Upsilon}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Upsilon}'\mathbf{F}_{\mathbf{cs}}'\Delta_{\mathbf{cs}}' & \boldsymbol{\Upsilon}'\mathbf{F}_{\mathbf{cs}}'\Delta_{\mathbf{cs}}'\end{bmatrix}+\left(\frac{\mathrm{g}}{\mu^2}\right)\begin{bmatrix}\Delta_{\mathbf{cs}}\Delta_{\mathbf{cs}}' & \emptyset_{\mathbf{j-1,j-1}}\\ \emptyset_{\mathbf{j-1,j-1}} & \Delta_{\mathbf{cs}}\Delta_{\mathbf{cs}}'\end{bmatrix}$$

$$\Rightarrow \boldsymbol{\Omega}_{\mathbf{pl}}(\boldsymbol{\theta})=\begin{bmatrix}\boldsymbol{\Omega}_{\mathbf{cs}}(\boldsymbol{\theta}) & \\ \boldsymbol{\Omega}_{\mathbf{cs}}(\boldsymbol{\theta})-(\mathrm{g}/\mu^2)\Delta_{\mathbf{cs}}\Delta_{\mathbf{cs}}' & \boldsymbol{\Omega}_{\mathbf{cs}}(\boldsymbol{\theta})\end{bmatrix}$$

$$\Rightarrow \mathrm{vecu}\left(\boldsymbol{\Omega}_{\mathbf{pl}}(\boldsymbol{\theta})\right)=\mathrm{vecu}\left(\begin{bmatrix}\boldsymbol{\Omega}_{\mathbf{cs}}(\boldsymbol{\theta}) & \\ \boldsymbol{\Omega}_{\mathbf{cs}}(\boldsymbol{\theta})-(\mathrm{g}/\mu^2)\Delta_{\mathbf{cs}}\Delta_{\mathbf{cs}}' & \end{bmatrix}\right)=\begin{bmatrix}\mathrm{vecu}\left(\boldsymbol{\Omega}_{\mathbf{cs}}(\boldsymbol{\theta})\right)\\ \mathrm{vecu}(\boldsymbol{\Omega}_{\mathbf{cs}}(\boldsymbol{\theta})-(\mathrm{g}/\mu^2)\Delta_{\mathbf{cs}}\Delta_{\mathbf{cs}}')\end{bmatrix}$$

Note that the same expression for $\mathrm{vecu}\left(\boldsymbol{\Omega}_{\mathbf{pl}}(\boldsymbol{\theta})\right)$ holds for all $\mathrm{K}\geq 2$. Therefore, the row rank of the Jacobian for panel data can be at most $\mathrm{J}(\mathrm{J}-1)/2$ more than the row rank for the cross-sectional data (the maximum number of unique elements in $\boldsymbol{\Omega}_{\mathbf{cs}}(\boldsymbol{\theta})-(\mathrm{g}/\mu^2)\Delta_{\mathbf{cs}}\Delta_{\mathbf{cs}}'$). It should further be apparent that the column rank of the Jacobian for panel data can be at most one more than the column rank for cross-sectional data. Thus, the use of panel data and agent effects can result in at most one additional linearly independent equation (in g/μ^2) available in $\boldsymbol{\Omega}(\boldsymbol{\theta})$. This may only be true in the case of heteroskedastic, nested and cross-nested specifications, and even then not always. For the random parameters model, the use of panel data and agent effects does not change the identification problem: continuous attributes are theoretically always identifiable, and the same conditions hold for categorical attributes, and continuous and categorical characteristics as with cross-sectional data.