Laboratoire d'Economie Forestière

Taste and Quality Perception of Attributes in Product Differentiation:
An Econometric Model with Latent Variables

Sandrine COSTA
Serge GARCIA
Lisette IBANEZ

Mai 2007

Document de travail
n° 2007-01

LEF – ENGREF/INRA – 14, rue Girardet – CS 4216 – F-54042 Nancy cedex
☎ 00 33 (0)3 83 39 68 66– 00 33 (0)3 83 37 06 45 – lef@nancy-engref.inra.fr
http://www.lef.nancy-engref.inra.fr

Taste and Quality Perception of Attributes in Product Differentiation:
An Econometric Model with Latent Variables[*]

Sandrine COSTA[1]
Serge GARCIA[2]
Lisette IBANEZ[3]

Mai 2007

Document de travail du LEF n°**2007-01**

Résumé

Goût et perception de la qualité des attributs dans la différenciation de bien :
un modèle économétrique avec variables latentes

Nous étudions les préférences des consommateurs pour un ensemble d'attributs que possède un produit. Nous supposons que le choix des consommateurs n'est pas directement guidé par les caractéristiques observées du produit, mais par la perception que ceux-ci se font sur la qualité du produit. Pour ce faire, nous intégrons des variables latentes dans le modèle et nous nous intéressons à l'impact de l'information possédée par le consommateur sur sa perception de qualité d'un produit. Le modèle considère également les goûts des consommateurs en les supposant être représentés par le poids des caractéristique dans le choix. Une application économétrique sur le choix de matériaux (PVC ou bois) dans le marché des fenêtres français montre la validité de ces hypothèses.

Mots clés : perception de la qualité, information, goût, modèle à variables latentes, bois.

Abstract

We study the consumers' preferences for the various attributes of a product. We consider that the consumers' choices are not guided by observed characteristics of a product, but by the quality perception consumers have on these attributes. Our model integrates this issue by means of latent variables, and is interested in the influence of the information received by the consumer on this perception. It also includes consumers' attitudes by supposing that the weight consumers put on attributes describes their tastes. An econometric application on material choices (wood vs. PVC) in the French window market valid these assumptions.

Key words : quality perception, information, taste, latent variable model, wood.

Classification JEL : C25, D12, L15.

---

[*] The authors would like to thank Michel Simioni for his very helpful comments on a preliminary version of this article and Jean-Marc Rousselle for his technical help on SAS programs.

[1] INRA-ENGREF, LEF, 14 rue Girardet, CS 14216, Nancy, F-54042 France, Email costa@nancy-engref.inra.fr

[2] INRA-ENGREF, LEF, 14 rue Girardet, CS 14216, Nancy, F-54042 France, Email garcia@nancy-engref.inra.fr

[3] INRA-ENGREF, LEF, 14 rue Girardet, CS 14216, Nancy, F-54042 France, Email ibanez@nancy-engref.inra.fr

1 Introduction

In general, a product possesses a large variety of attributes. Firms try to soften price competition by differentiating these attributes. Then the comparison of products and the valuation of consumer preferences is rather complicated. Indeed, it is difficult to distinguish the buying intentions and to value their attachment to attributes.

Product differentiation has been largely studied in the economic literature both through empirical and theoretical angles. Lancaster (1966) supposed that a product is composed of several attributes and consumers attach value to these attributes. The restrictive assumption is the representative consumer. Other theoretical models exist where consumers are differentiated but in that case a product is, in general, reduced to a single attribute. Each attribute can either be horizontally (Hotelling, 1929) or vertically differentiated (Mussa-Rosen, 1978). In empirical models, the main objective is to value consumer preferences. In both cases, the quality level of each attribute is perceived equivalently by consumers. This does not mean that consumers have necessarily same willingness to pay (WTP) for these attributes. Consumer characteristics such as tastes, revenue, localization or other will determine their preferences and so the price they are willing to pay for different product attributes.

However, in reality, consumers possess imperfect information on several product attributes which makes it very difficult to distinguish correctly consumer preferences. Indeed, a lot of product characteristics are not observable before purchase (Nelson, 1974). Thus, consumers have to form beliefs on these characteristics. According to the type of attribute (whether it is associated to a research, experience or credence attribute), the quality perception might differ from one consumer to another. So if a firm wants to put to the fore a specific attribute quality (in order to distinguish its product from its competitors) it needs to correctly inform consumers on these attributes. So the perception of the quality level of each attribute by a consumer depends on the information available for him. Consumers can get the information in different ways. For example, information can be gathered by receiving advertising, contacting professionnals, reading specific brochures, etc. Another way of getting informed on the quality level of attributes is experience. For some attributes, previous purchase and/or consumption is enough to appreciate the real quality level. Then, the consumers having already purchased are better informed on certain product attributes than non-initiated consumers. In any way, consumers possess different information on product attributes and, for this reason, it seems to be justified to assume that they judge differently the quality level of attributes.

Information and quality perception by consumers are crucial in purchase decisions as in situation of imperfect information only low quality products will be demanded (Akerlof, 1970).

In choice modeling, the imperfect information problem is generally ignored and it is supposed that consumers possess perfect information on attributes and that the quality level is exogenous and identically perceived by consumers. Choice models then allow to determine consumer preferences (or equivalently their tastes) for product attributes. In our econometric model we suppose tastes are observed variables. In our survey we have asked each person for weights they attach to each product attributes in purchase decisions. We suppose weights represent perfectly tastes for product attributes.

Besides tastes for product attributes, we focus on the imperfect information consumers possess on product quality. Quality levels of product attributes are not observable variables but consumers form quality perception on these according to their information and experience. And quality perception of products influences highly consumer choices. There exist indicators for these latent variables that can be observed and obtained through survey questions. In our analysis, we aim to understand purchase choices for specific materials (PVC and wood) in window markets. Consumers not necessarily value identically these materials for windows. Indeed, information gathered, previous purchase or experience of materials are factors that influence quality perception by consumers. As most windows attributes (e.g. insulation, maintenance) are not observable before purchase, consumers possess imperfect information and estimate the quality of materials (PVC and wood in our data set) according to their information and experience.

So the aim of our paper is to integrate the imperfect information problem on products and their attributes into a discrete choice model. We suppose that a product is defined by a finite number of attributes and that the quality level of attributes is an endogenous variable, perceived differently by consumers. Then, on the one hand, consumers are supposed to have quality beliefs on these attributes which allow to determine their quality perception of the entire product. So, we suppose that quality perception of a product is a non-observed variable and that it depends on information and consumer's experiences. On the other hand, consumers can be differentiated by the weight they put on product attributes in their purchase decision. We refer to these observed variables as tastes. Our analysis is inspired by McFadden (1986) in integrating psychological factors in choice decision.

Our methodology is applicable to any situation with discrete choice behavior where products are characterized by several attributes for which quality is not directly observable and where there exist indicators, observable variables related to information and product experience, that might determine quality perception.

The paper is organised as follows. In section 2, we state the assumptions of our model and discuss their relevance related to the existing literature. We focus on the quality perception of product attributes by consumers. In section 3, the econometric model is developed. A two-step estimation procedure is adopted: in the first place, we estimate a model for latent variables (quality perceptions), in a second stage, the predicted latent variables are introduced into a discrete choice model. Section 4 proposes an application of the econometric set up. Data on material choices in the French window market are used to show that endogenous quality levels of product attributes significantly improves the choice model. Section 5 concludes and discusses future improvements of the model.

2 Attributes and quality perception in product choice

Consumers purchase decision depends on their preferences for products. The preferences for products have been analysed in different manners in the literature. In the vertical and horizontal differentiation literature, a product is seen as a single attribute. This simple notion of "product" allows to construct sophisticated economic models to analyse consumer and producer behaviour towards product differentiation. However, these types of analyses ignore the complexity of products and the way that consumers perceive them. Indeed, consumers confronted to purchase decision making has to choose between a certain number of products. However, a comparison between products is a complex phenomenon as a product possesses several attributes for which consumers might have different valuations. Lancaster (1966) introduced a model where the consumer attaches importance to attributes rather than the product itself. In Lancaster's approach, consumers choose into a panel of characteristics according to their preferences. The theory is based on objective characteristics, which means that the quality level of the product attribute is exogenous and given, for which consumers might differ in terms of WTP. So what differentiates products are the characteristics that they possess. However, consumers might differ in the value they attach to these characteristics and the way they value them. In the definition of consumer utility, two aspects are crucial. On the one hand the importance consumers attach on product characteristics in their decision making according to their experience. For instance, the colour of the product might be an important feature when buying a t-shirt, whereas it becomes a less important attribute in the purchase decision for a car. We will define this as the attribute weight or consumer tastes or attitude. On the other hand, consumers might judge the quality level of attributes differently for differentiated products according to the information they have. A consumer who chooses between t-shirts might judge the thermal quality, aesthetic value, etc., for a cotton t-shirt differently than for a synthetic one. We will call this the quality perception or beliefs of product's attributes.

In empirical literature, we distinguish two approaches to evaluate consumers' preferences: the hedonic price modelling and discrete choice models. The hedonic price method allows to determine WTP for product attributes. Rosen (1974) states that if the implicit price for an attribute is not significantly different from zero, then either consumers attach a low weight to the attribute or they perceive the attribute to be of low value. The method does not allow to distinguish taste and quality perception. The discrete choice modelling studies the impact of product characteristics on the consumer purchase decision, or rather the probability to choose a particular product. In most models explanatory variables are observable. However, product quality is in a lot of cases difficult to estimate for consumers as this product is composed of several attributes and as its attribute qualities are not necessarily observable before purchase.

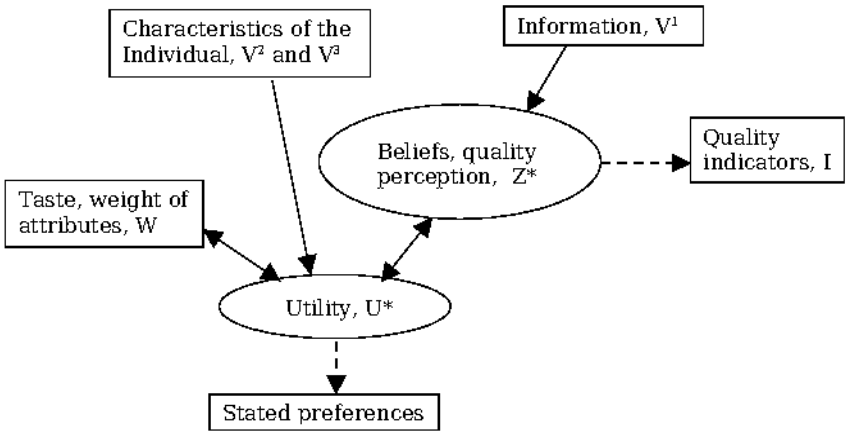
The purpose of this paper is to construct an econometric model that explains consumer purchase decision by integrating simultaneously perceived product quality (latent variables) and tastes (observable variables).

In our framework, product price is not an explanatory variable in consumers' choice. In the survey, products are supposed to be sold at identical prices. The consumers' choice is only based on technical properties of products. This assumption allows us to concentrate mainly on non directly observable product attributes for which consumers might have different quality perceptions. The model aims to link consumer's knowledge on product attributes with his quality perception and tries to see in what way information campaigns could orientate consumers' choices.

In the transportation literature, latent preferences have been introduced in mode choice models (Ben-Akiva et al., 2002). For instance, Johansson et al. (2006) argue that people have different attitudes towards environmental considerations, safety, comfort, convenience and flexibility and that this has an important impact on the transportation mode choice. The introduction of attitudinal and behavioural indicator variables allows to construct a more precise choice model. The marketing literature investigates on the link between perceived value and consumer behaviour. Indeed, the perception of product quality depends on the messages and information transmitted to consumers. And according to the amount and quality of information, consumers might react differently in terms of consumption behaviour. Swait and Sweeney (2000) show that the utility of buying a specific product is conditional on the type of consumer. They propose to segment consumer behaviour into groups possessing each specific behaviour characteristics. Ben-Akiva et al. (2002) propose a general methodology for including latent variables such as attitudes and perceptions, in choice models. The methodology integrates a discrete choice model and a latent variable model consisting each of one or more structural and measurement equations. The basic assumption in their methodology is that observable explanatory variables explain simultaneously utility and latent variables.

We adopt a similar model but also integrate substantial modifications. The set up of our model is outlined in figure 1. We can see that variables related to information possession by

Figure 1: Integrated choice and latent variable model



consumers, which we call V_1, only explain quality perception of products Z^* and not directly utility U^*, differently from Ben-Akiva et al. (2002). Quality perception is a determinant of consumers utility and based on quality indicators I. As in classical choice models, utility is also explained by individual characteristic variables denoted V_2 and V_3 in Figure 1 (see section 4 for description of variables).

In our model, we suppose that consumer heterogeneity is observable, contrary to models that focus only on consumer tastes through latent classes.[1] The observed variable is the weight consumers value to attributes into their purchase decision, indicated W in Figure 1. We assume that weights consumers put on attributes describe perfectly the tastes of consumers. Making this assumption it is assumed that stated tastes are real tastes, as our data set includes the statements of consumers about their attitudes. It is of course a strong assumption. However, all our data are based on statements of consumers, as for instance product choice or perception indicators. So, there is a coherence between all our data. Moreover, the statements of consumers about their product choice give approximatively the same market shares to different products as observed in the market.

3 Econometric methodology

The integrated model described in the previous section is composed of two parts: in addition to the standard discrete choice model, a latent variable model is defined to carry out the treatment

[1] See Greene (2001) for a review of literature.

and the prediction of latent perception variables. The latent variable model (or structural equation model, SEM) consists of structural latent variable equations and measurement equations. The discrete choice model is composed of a structural equation and a decision rule equation.

3.1 The latent variable model

The latent variable model has been popularized by the program LISREL – Linear Structural Relationship (Jöreskog, 1979; Jöreskog and Sörbom, 1982). The latent variables are modeled by specifying a structural model and a measurement model. In general, the structural latent variable equations specify the relationship between endogenous and exogenous latent variables. However, it is also possible, as in our analysis, to include exogenous observed variables as part of these structural equations. The measurement equations specify the relationship between the measured indicators and the latent variables.[2]

3.1.1 Structural latent variable equations

A linear specification of the structural latent variable equations can be written:

$$Z^* = V^1\beta + \epsilon_Z, \tag{1}$$

where Z^* is a vector of latent (or non observed) variables of quality perception and believes. V^1 is a vector of (observed) explanatory variables on the information hold by the consumer, and β the matrix of unknown parameters to be estimated and ϵ_Z an error vector. As said above, contrary to Ben-Akiva et al. (2002), the variables V^1 are specifically and only factors explaining Z^*.

3.1.2 Measurement equations

The set of measurement equations is defined as follows:

$$I = Z^*\gamma + \epsilon_I, \tag{2}$$

where I is a vector of measured indicators of perceived quality and γ the matrix of unknown parameters to be estimated. ϵ_I is the measurement error vector.

[2] In structural equation modeling, non-observed variables are called latent variables and observed variables are called manifest variables.

3.2 The discrete choice model

As in many empirical works, the structural utility function is assumed to be linear:

$$U^* = X\alpha_X + Z^*\alpha_Z + W\alpha_W + \epsilon_U, \tag{3}$$

where U^* is the latent utility concerning the choice of a commodity. It is standard that the utility depends on observed explanatory variables (X) including socio-demographic variables (V^2) and behavorial variables related to housing (V^3) in our analysis. However, utility also depends on latent variables of quality perception (Z^*) and taste factors (W). W is a vector of (observed) weights allocated to the commodity characteristics, and capturing tastes and attitudes of consumers. The introduction of such variables are the other important difference with respect to the model of Ben-Akiva et al. (2002). α_X, α_Z, α_W are the parameter vectors associated to these variables. ϵ_U is the disturbance term.

The decision rule consists of a stated choice depending on the latent utilityU^*:

$$y = \begin{cases} 1 & \text{if} \quad U^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $y = 1$ if the respondent chooses wood as the material for her/his windows and$y = 0$ if she/he prefers PVC.

3.3 Estimation method

There exist several methods to estimate the complete (or integrated) model composed of the latent variable model and the discrete choice model. Ideally (and efficiently), the complete model could be solved by simultaneously estimating the system of equations (1)–(4). This involves to build the joint likelihood function that includes complex multi-dimensional integrals, one corresponding to the choice variable and the other ones to each latent variable. One estimation method can be implemented from a simultaneous numerical approach. But if the dimensionality (the number of integrals) of the likelihood function increases too much, it can be preferable to use a method of estimation by simulation.

However, computing simulated maximum likelihood estimation for a very complex likelihood may be prohibitively expensive. As Johansson et al. (2006), we have chosen to estimate the interest parameters in two steps. First, the latent variable model, equations (1) and (2), and the associated parameters are estimated. In particular, it allows us to compute the predicted values of latent variables (perception variables). The framework for modelling and estimation of the variable latent model is based on the program LISREL. It is an hybrid technique that encompasses aspects of confirmatory factor analysis, path analysis and regression. This technique is explained in Appendix B. Second, we report the predicted perception variables into the choice model, equations (3) and (4), and the resulting model is then estimated conditional on the perception variables (as well as other explaining variables and taste factors).

We decide to adopt a two-step method and specifically analyse the latent variable model for several reasons. First, SEM in general and associated programs to solve it, are powerful tools that allow us to model interactions, correlated independent variables, measurement error, correlated error terms, multiple latent variable each measured by multiple indicators and so on. Second, a two-step method allows a good adjustment of the latent variable model. Finally, it allows us to specifically study the reliability of indicators for latent perception and to better understand how the information influences quality perception for materials which is the focus of our paper.

As noted in Ben-Akiva and al. (2002), the two-step estimation procedure that consists in reporting the fitted latent variables in the utility function may introduce measurement error and result in inconsistent estimates of the parameters. However, if the variance of the latent variable's random error is small, then a sufficiently large size of sample may reduce the measurement error and result in acceptable parameter estimates, see Train et al. (1986).

4 Application: Material choice in window market

4.1 Data

In the French building markets, the share of wood is relatively low when compared to other European countries. One objective of the French government is to increase this share by 25% in value terms.[3] Different ways to favor the use of wood have been chosen: an increase of wood purchase in public procurements, architect training to the use of wood, final consumer information with TV spots and advertisement in newspapers. Our objective here is to analyse if information policies are a good tool to increase wooden products demand. Our data concerns one segment of building markets: the windows market. In the case of windows, consumers have the choice between four materials: wood, PVC, aluminium and steel. We concentrate in our study on the choice between wood and PVC,[4] as PVC covers the largest market share and is the most competing material for wood in the window market. All the data used here come from a specific survey on tastes and beliefs of consumers about windows. The inquiry was conducted by a marketing agency[5] in November 2003, and concerned a set of 968 consumers representative of

[3]Today, the wood represents 10% in value of all materials used in the building sector.
[4]Our data set includes also data on aluminium and steel, but we do not analyse these data here.
[5]ED Institut, marketing studies institute, www.edinstitut.com.

French population. There remain 940 observations after dropping individuals with missig data.

The survey includes 6 kinds of data:

- Information individuals possess on attributes and their quality (variables denoted V^1): windows material of their lodging (wood, PVC, aluminium etc.), age of the windows, how they appreciate the insulation (e.g. thermal, acoustic) and brightness of their housing. Moreover, the respondent is asked to answer to the following questions: "Do you feel informed on windows?" and "Do you feel informed on materials?" As typically for this kind of question, a five-point Likert scale is used and people have to choose in the following set of answers: not at all good, not so good, indifferent, good or very good
- Socio-demographic characteristics of individuals (variables denoted V^2): age, gender, socio-professional group (SPG), family composition.
- Behaviour of individuals towards housing (variables denoted V^3): the kind of lodging (e.g. house, apartment), building style of their lodging (classic, modern, typical), importance (with a five-point Likert scale) attached to home.
- Tastes for product attributes (weight W): as in Lancaster, we assume that consumer choices depend on the attributes of the product rather than the product itself. We consider several attributes of a window: thermal insulation, acoustic insulation, maintenance, lifetime, aesthetics, environmental properties, fire resistance and safety. For each of them, consumers were asked: "when you choose a window, how important is this attribute?" Proposed answers were: not at all important, not very important, indifferent, important or very important (five-point Likert scale).
- Quality perception of product attributes (quality indicators I), which was analysed for all attributes and for each material (wood and PVC). So the question was, for instance: "What do you think about the thermal insulation of a wooden window?". Again a five-point Likert scale was used: not at all good, not so good, indifferent, good or very good.
- Stated choice y based on the question: "Suppose that wooden windows and PVC windows are sold at the same price, which material do you choose?"

Table A.1 in Appendix A presents summary statistics and description of variables V^1, V^2, V^3 and y. All answers concerning judgements are rated on a 1 to 5 scale such that 1 corresponds to a low appreciation and 5 to a high one, 3 indicates indifference. Table 1 gives summary statistics concerning the quality perception of windows attributes I according to the material (wood or PVC) as well as the weight of each attribute W in the choice of a window.

Table 1: Descriptive statistics on quality indicators and taste

Variable	Mean	Std.	Min.	Max.
Indicators I				
Wood				
I_{11} Global quality	3.8191	1.1207	1.0000	5.0000
I_{12} Thermic insulation	3.7340	1.1701	1.0000	5.0000
I_{13} Acoustic insulation	3.6149	1.1651	1.0000	5.0000
I_{14} Maintenance	2.4415	1.1882	1.0000	5.0000
I_{15} Lifetime	3.5617	1.2260	1.0000	5.0000
I_{16} Aesthetics	4.6277	0.7567	1.0000	5.0000
I_{17} Environment	3.6362	1.2304	1.0000	5.0000
I_{18} Fire resistance	2.1011	1.1377	1.0000	5.0000
I_{19} Safety	2.7979	1.2561	1.0000	5.0000
PVC				
I_{21} Global quality	4.1234	1.0025	1.0000	5.0000
I_{22} Thermic insulation	4.1521	0.9431	1.0000	5.0000
I_{23} Acoustic insulation	4.0479	0.9406	1.0000	5.0000
I_{24} Maintenance	4.4585	0.8200	1.0000	5.0000
I_{25} Lifetime	4.0457	0.9719	1.0000	5.0000
I_{26} Aesthetics	3.6628	1.2079	1.0000	5.0000
I_{27} Environment	3.1032	1.1384	1.0000	5.0000
I_{28} Fire resistance	2.5287	1.1758	1.0000	5.0000
I_{29} Safety	3.0957	1.2208	1.0000	5.0000
Weight W				
Thermic insulation	4.8170	0.5620	1.0000	5.0000
Acoustic insulation	4.6181	0.8559	1.0000	5.0000
Lifetime	4.6213	0.7521	1.0000	5.0000
Aesthetics	4.4904	0.8553	1.0000	5.0000
Fire resistance	4.1351	1.2383	1.0000	5.0000
Maintenance	4.5319	0.8720	1.0000	5.0000
Environment	4.1138	1.1458	1.0000	5.0000
Safety	4.4532	1.0010	1.0000	5.0000

Notes: $N = 940$.

Data results on quality indicators reveal that PVC has a better image than wood for all attributes except for aesthetics and environmental properties. Data results on attribute preferences indicate that people attach a great importance to technical properties of windows such as thermal and acoustic insulation or lifetime and maintenace, as well as to aesthetics. Moreover, people are rather indifferent concerning attributes such as environmental properties, fire resistance or safety.[6] Results also show that beliefs about technical properties of wood are sometimes false for some of its characteristics, which is less frequently the case for PVC. Indeed, Table 1 shows that beliefs on some technical properties of wood such as thermal insulation or lifetime compared to PVC were not scientifically justified.[7] One of the explications of this result can be experience

[6] A simple test shows that rates are not significantly different from 3 (i.e. indifference) for these attributes.
[7] See for example www.sustainability-ed.org/pages/example5-2.htm or
www.brown.edu/Departments/Brown_Is_Green/greenarch/winintro.html.

of respondents on rather old (not necessarily well maintained) wooden windows which in most cases do not possess double glasses.

4.2 Results

4.2.1 The latent variable model

The latent variable model consists of a structural model and a measurement model. Both models are simultaneously estimated by a LISREL-type structural equation modeling program. The principle of the model estimation is to minimize the difference between the sample covariance matrix and the covariance matrix depending on the parameters to be estimated. Different estimation methods can be used in practice. Due to the specificity of our variables (dichotomous and ordinal variables) implying the non-normality of data, we choose the generalized least squares (GLS) method implemented with tetrachoric and polychoric correlations used as imput for the model, see Appendix B for details.[8] The first part of the latent variable model is composed of two latent variables Z_1^* (quality perception of wood) and Z_2^* (quality perception of PVC), and eight observed exogenous variables (see Table 2 for exogenous variables retained in the empirical application). The second part of the latent variable model deals with the measurement equations. Two sets of indicators are used in our model, each one corresponding to one latent variable. The complete system can be written:

[8] We use the CALIS procedure (Covariance Analysis of Linear Structural Equations) from SAS system for windows (version 9.1) to estimate parameters and test the appropriateness of the linear structural equation model using covariance structure analysis. As we have a set of linear structural equations to describe our model, we use the LINEQS statement. In particular, it is possible to specify variances and covariances in the model, to choose between different estimation methods and to enter correlation matrices instead of raw data. We construct tetrachoric/polychoric correlation matrices thanks to the FREQ procedure of SAS.

$$Z_1^* = \beta_{11}V_1^1 + \beta_{12}V_2^1 + \beta_{13}V_3^1 + \beta_{14}V_4^1 + \beta_{15}V_5^1 + \beta_{16}V_6^1 + \beta_{17}V_7^1 + \beta_{18}V_8^1 + \epsilon_1 \tag{5}$$

$$Z_2^* = \beta_{21}V_1^1 + \beta_{22}V_2^1 + \beta_{23}V_3^1 + \beta_{24}V_4^1 + \beta_{25}V_5^1 + \beta_{26}V_6^1 + \beta_{27}V_7^1 + \beta_{28}V_8^1 + \epsilon_2 \tag{6}$$

$$\begin{bmatrix} I_{11} \\ I_{12} \\ I_{13} \\ I_{14} \\ I_{15} \\ I_{16} \\ I_{17} \\ I_{18} \\ I_{19} \\ I_{21} \\ I_{22} \\ I_{23} \\ I_{24} \\ I_{25} \\ I_{26} \\ I_{27} \\ I_{28} \\ I_{29} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \gamma_{12} & 0 \\ \gamma_{13} & 0 \\ \gamma_{14} & 0 \\ \gamma_{15} & 0 \\ \gamma_{16} & 0 \\ \gamma_{17} & 0 \\ \gamma_{18} & 0 \\ \gamma_{19} & 0 \\ 0 & 1 \\ 0 & \gamma_{22} \\ 0 & \gamma_{23} \\ 0 & \gamma_{24} \\ 0 & \gamma_{25} \\ 0 & \gamma_{26} \\ 0 & \gamma_{27} \\ 0 & \gamma_{28} \\ 0 & \gamma_{29} \end{bmatrix} \begin{bmatrix} Z_1^* \\ Z_2^* \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{15} \\ \epsilon_{16} \\ \epsilon_{17} \\ \epsilon_{18} \\ \epsilon_{19} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{25} \\ \epsilon_{26} \\ \epsilon_{27} \\ \epsilon_{28} \\ \epsilon_{29} \end{bmatrix} \tag{7}$$

The covariance of the error terms in the structural latent equations and in the measurement equations are constrained to be equal to zero.

Estimation results of latent variable equations One objective of this paper is to better understand how quality beliefs are formed, and more precisely if information may change the quality beliefs. Table 2 shows the estimation results for the two latent variables, quality perception of wood and quality perception of PVC.[9]

First, the significance of the coefficient associated with the variable on received information on windows shows that information, conveyed by way of advertising for instance, can modify quality perception. In this case, recent information received by respondents have a negative impact on her/his perception of wood. Second, the degree of information respondents feel to possess both on windows and materials have a significant impact on the perceived quality of window materials. On the one hand, more individuals feel to be informed on windows, more the quality perception of wood will be good. But information on windows has no impact on perceived quality of PVC. On the other hand, the coefficient associated with the information on materials is significantly

[9]In order to have the best predictions of the latent variables (i.e. minimizing the variance of their random error), we only retain the statistical significant factors.

Table 2: Estimation results of latent variable equations

	Z_1^* (Wood)		Z_2^* (PVC)	
Variables V^1	Estimate	t-stat	Estimate	t-stat
Received information	-0.1404^{**}	-2.1298		
Informed on windows (feeling)	0.0961^{***}	4.4447		
Informed on materials (feeling)			0.0905^{***}	4.5238
Do-it-yourself	0.0519^{***}	3.1352		
Very good thermal insulation × Home wooden windows	0.8989^{***}	12.9582	-0.7233^{***}	-10.4848
Good acoustic insulation × Home wooden windows	0.4463^{***}	6.3015	-0.7372^{***}	-9.6640
Bad Brightness × Home wooden windows			0.4124^{***}	3.4590
Purchase of windows × Home wooden windows	1.1016^{***}	6.9362	0.5664^{***}	3.6546
R^2	0.3773		0.1962	

Notes: $N = 940$.
***: significant at 1%, **: significant at 5%, *: significant at 10%.

positive for PVC perceived quality and not significantly different from 0 for perception of wood. Third, another way of information getting for the consumers is experience. We have introduced some variables that can proxy this experience. Estimation results show significant impact of these variables on perceived quality. If people have the practice of do-it-yourself, they have a better opinion on wooden windows while this has no effect on perception of PVC. According to the material of their windows at their home, the observed level of insulation and brightness at their home influences the opinion of individuals for quality perception of materials. For instance, a good or very good thermal and acoustic insulation associated with wooden windows at home improve the perceived quality of wood and decrease the perception of PVC. In the same way, a bad brightness associated with wooden windows increases the perception of PVC while this has no impact on perception of wood.

It is important to check that our estimates of the parameters are acceptable. Indeed as precised in the description of the estimation method, the two-step procedure may result in inconsistency of parameters due to the introduction of measurement error by reporting fitted latent variables in the discrete choice model. However, the estimated variances of the latent variable random errors are small (respectively 0.44 and 0.53 for wood perceived quality and PVC perceived quality). Thus, a large size of sample (around 1,000 observations) associated with small variances of errors means that an individual's true value of the latent variable is not too far off from its expected value.

Estimation results of measurement equations Results concerning the measurement model indicate the level of reliability and validity of the indicators of the latent variables. These results are reported in Table 3.

Table 3: Estimation results of measurement equations

		Z_1^* (Wood)			Z_2^* (PVC)		
	Indicator variable	Estimate	t-stat	R^2	Estimate	t-stat	R^2
Wood	I_{11} Global quality	1.0000	–	0.6082			
	I_{12} Thermal insulation	0.9258	19.8331	0.5601			
	I_{13} Accoustic insulation	0.8460	19.0715	0.4732			
	I_{14} Maintenance	0.7118	15.5035	0.3131			
	I_{15} Lifetime	0.9215	19.7174	0.5147			
	I_{16} Aesthetics	0.5396	18.5920	0.4448			
	I_{17} Environment	0.5261	10.5409	0.1448			
	I_{18} Fire	0.3953	8.8896	0.1641			
	I_{19} Safety	0.5181	12.4842	0.3330			
PVC	I_{21} Global quality				1.0000	–	0.7105
	I_{22} Thermal insulation				0.9277	27.3029	0.7323
	I_{23} Acoustic insulation				0.8380	22.7859	0.5986
	I_{24} Maintenance				0.6517	21.0451	0.4666
	I_{25} Lifetime				0.6514	17.6991	0.3371
	I_{26} Aesthetics				0.8318	18.8835	0.3798
	I_{27} Environment				0.4922	10.9605	0.1766
	I_{28} Fire				0.3275	7.3378	0.0927
	I_{29} Safety				0.4928	12.5280	0.2671

Notes: $N = 940$.
All parameter estimates are significant at the 1% level.

The estimated factor loadings in the measurement equations are all positive and statistically significant at a 1% level of significance. The results confirm that the question on the global quality of windows gives the best indicator of perceived quality among all indicators, with R-square respectively equal to 0.61 and 0.71 for wood and PVC perceived quality. Moreover, thermal insulation seems to be the second best indicator (respectively 0.56 and 0.73 for wood and PVC), followed by acoustic insulation, maintenance and lifetime. Instead, fire resistance, safety and environmental considerations are worser factor for the perception of windows quality.[10]

Finally, fit indices for the latent variable model are presented in Table 4, see Appendix B for definitions of these goodness-of-fit indices. A χ^2 statistics built on the discrepancy function gives a good indication of the model fit. For our model, the test statistics is equal to 1800.88 with 268 degrees of freedom. Its p-value is less than 0.0001. The goodness of fit index (GFI) and the

[10] We also tested dependence between quality perception and price in order to be sure that consumers did not integrate price in judgement for particular materials. Correlation coefficients between the price and the latent variable (quality perception) turned to be very low, respectively 0.0017 and 0.0193 for wood and PVC. Complete estimation results are available from authors upon request.

adjusted GFI are close to 1 (respectively 0.8525 and 0.8068). Moreover, the root mean square residual (RMR) is equal to 0.1283 while the root mean square error of approximation (RMSEA) is estimated to 0.0780. These measures of fit indicate that the model fits the data well.

Table 4: Goodness of fit statistics

Fit function	χ^2	GFI	AGFI	RMR	RMSEA
1.9179	1800.88 (Pr<0.0001)	0.8525	0.8068	0.1283	0.0780

Notes: $N = 940$, degrees of freedom $df = 268$.

4.2.2 The discrete choice model

In the discrete choice model, we estimate the probability to choose wood rather than PVC as material for windows. Our model includes, along with "usual" explaining variables, the predicted latent variables (i.e. predicted quality perceptions of wood and PVC) and taste (or attitudes) by way of weight of attributes. Table 6 presents the results of the choice model estimation.

Specification tests Before analysing parameter estimates in Table 6, we proceed to several specification tests on our model. Indeed, we must check that the integrated model described in figure 1 is the good specification for explaining the role of information in the choice decision. In particular, a latent variable model has been attached to the choice model in order to take into account non observed perception variables that could influence the choice of consumers. Moreover, our model supposes that the preferences and tastes are different between consumers and that the weights consumers put on attributes (assumed to represent the tastes of consumers) are observed factors that affect their utility. We report results of (likelihood ratio) tests on these assumptions in Table 5.

We first test a model without weights of attributes. There are 8 restrictions based on the nullity of coefficients associated with the weights for the 8 defined attributes of a window. This model is not accepted at the 5% level showing the importance of tastes in the choice decision. Second, the hypothesis that the two parameters of the predicted latent variables are jointly equal to 0 is also rejected. This result confirms that the integration of latent variable model is necessary and that the quality perception variables provides a meaningful explanation of the choice decision of consumers. Finally, we test for the null hypothesis (gathering the two previous ones) that both coefficients associated with weights and coefficients associated with latent variables are equal to 0. The test statistics is also greater than the critical value implying the reject of this hypothesis.

Table 5: Specification tests

Null hypothesis	$-2\ln Lr$	LR statistics	Critical value	Decision
Model without weights of attributes W H_0 : coef. related to W null	1110.74	52.77	$\chi^2_{0.05}(8) = 15.51$	Reject H_0
Model without quality perception Z^* H_0 : coef. related to $\hat{Z}_1^*$ and $\hat{Z}_2^*$ null	1125.15	67.19	$\chi^2_{0.05}(2) = 5.99$	Reject H_0
Model without Z^* and W H_0 : coef. related to W, $\hat{Z}_1^*$ and $\hat{Z}_2^*$ null	1186.66	128.70	$\chi^2_{0.05}(10) = 18.31$	Reject H_0

Notes: $N = 940$.
For the unrestricted (complete) model, $-2\ln Lu = 1057.96$.

This proves that the integrated model is the best specification.[11]

Analysis of results Now several comments can be drawn from the estimate results in Table 6. First, as announced by the results of previous tests, estimated parameters associated with each of two latent variables, quality perception of wood and quality perception of PVC are highly significantly different from 0 (at the 1% level). As expected, a higher quality perception of wood leads to a larger probability of choosing this material for windows. In the same way, a lower quality perception of PVC also implies a choice of wood rather than PVC.

Second, tastes have significative effects on the material choice for three of the eight characteristics of a window: thermal insulation, aesthetics and maintenance. For these indicators, quality beliefs differ a lot between wooden windows and PVC windows (see Table 1). Furthermore, the thermal-insulation property has a negative effect on the choice. This would mean that one believes that wood has a worser power of insulation than PVC, which is confirmed in results of Table 1. We can check in Table 3 that the thermal-insulation property influences highly the quality perception of either wooden or PVC windows. Moreover, people who think that aesthetics is important or very important when choosing a window, will choose wooden windows rather than PVC ones. Indeed, people believe on average that wooden windows have a more beautiful aesthetics than PVC windows (see Table 1). PVC windows are believed to be more easy to maintain than wooden windows, on average. As a result, when people attach importance to maintenance, they will choose PVC windows rather than wooden windows. It should be noticed that both parameters of indicators aesthetics and maintenance are significantly different from 0 at the 1% level while the one of thermal insulation at the 5% level.

[11] Significant gains with respect to simpler models can be also measured in terms of correctly predicted observations and pseudo R^2. For the integrated model, 72% of observations are correctly predicted by the model and the R^2 is equal to 0.18. Without latent variables, 68% of observations are correctly predicted by the model and R^2 falls to 0.13. If weights and latent variables are not included in the model, the percentage of correctly predicted observations only reaches 65% and the pseudo R^2 0.10.

Table 6: Choice model estimation results

Variable	Estimate	t-stat
Constante	1.5929***	2.6245
Age	0.1722***	3.8203
Not sensitive to green	−0.3036	−1.1437
Not very sensitive to green	−0.2280	−1.4242
Rather sensitive to green	−0.2674***	−2.6280
Very good finances	0.7230***	3.0703
very few importance to home	−1.0598*	−1.9012
very big importance to home	−0.3022***	−3.1005
SPG1	0.2364	1.1863
SPG3	0.1106	0.8112
SPG4	−0.1301	−0.8467
Com	−0.0406	−1.0332
Reg1	−0.1514	−0.9420
Reg2	−0.3556***	−2.6491
Reg3	−0.2474*	−1.8872
Reg4	−0.2061	−1.3040
Single	0.3890***	2.8105
Family	0.1910*	1.8019
Woman	0.0744	0.7885
Owner	−0.4410***	−3.9572
Modern	−0.2857**	−2.0935
Classic	−0.3041***	−2.7362
Rooms	0.0349**	2.0044
Quality perception of wood	0.3566***	2.8885
Quality perception of PVC	−0.6480***	−4.3831
Acoustic insulation	0.0066	0.1067
Thermal insulation	−0.2062**	−2.2256
Lifetime	−0.0394	−0.5970
Aesthetic	0.2009***	3.4918
Fire resistance	−0.0219	−0.4722
Maintenance	−0.3177***	−5.2107
Environment	0.0255	0.5558
Safety	−0.0256	−0.4633
$-2\ln l$	1057.96	
McFadden's Pseudo R^2	0.1774	
Correctly predicted	72.23%	

Notes: $N = 940$.
***: significant at 1%, **: significant at 5%, *: significant at 10%.

Third, among variables describing behavior towards housing or socio-economic factors, building style contributes to explain choice of window material. When the house is a classic or modern one, one chooses less wood than PVC with respect to a typical style. PVC will be chosen more frequently by owners rather than tenants, perhaps because owners think that they will have to maintain their windows on the contrary of tenants. Wood is more frequently chosen when age increases, perhaps because one has more time to spend for their home and one is more attached to the aesthetics of wood. The region where one lives is also a factor explaining the choice of material for windows but not the SPG. However a very good state of finances for the consumer explains the choice of wood rather than PVC. A family with children has a light preference for wood whereas this preference is definitely marked for the single people. Surprisingly, consumers rather sensitive to green opinion would prefer PVC windows than wooden windows, certainly because they think that harvesting and thus wood as a material is harmful for forests.

Our results show that information that consumers possess on windows and materials influences highly their consumption choice. In the case of windows, quality perception is determined by the level of knowledge consumers have either by (external) information or experience. The choice model shows that the consumption decision is highly dependent on the quality perception of consumers. This means that information policies might be an efficient tool in the objective to increase the market share for wood.

5 Conclusion and perspectives

We have set up an econometric model where the quality level of product attributes is an endogenous variable and depend on the information level of consumers. We show that the introduction of quality perceptions and tastes for different attributes significantly improves a discrete choice model. This result is not very surprising as several attributes of windows are not observable before purchase. The situation of imperfect information in purchase decisions makes that consumers do not value identically the objective quality level of attributes. So, depending on the information they possess on attributes they form different quality perceptions. The weight on some attributes in the purchase decision by consumers (tastes) such as thermal insulation, aesthetics and maintenance also influences significantly their product choice.

The quality level of attributes is a determining factor in the purchase decision, especially for the attributes for which consumers have high tastes. This means that if producers can influence consumers' perception favorably, demand would increase. For example, higher evaluation for thermal insulation (which is not observable before purchase) influences positively quality perception and thus could influence highly purchase decisions. This would mean that if wood producers can persuade that wooden windows are as isolating as PVC ones (or even better), demand for wooden windows would increase. Information campaigns might then be an efficient tool.

So our results show that information policies can modify consumer purchase choice, especially when consumers are imperfectly informed on the properties of products. In our application on windows we use stated choice data. It could be useful to apply the model to observed choices and stated preferences (taste indicators and quality perception of products). Studying consumers beliefs about product quality could also be possible with experimental data. More sophisticated analyses on the impact of different kinds of information provision (e.g. general information before going into the store, information available in the store on the product) could also enrich the analysis and allow interesting insights for policy making.

References

AKERLOF, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* 84(3): 488–500.

BEN-AKIVA, M., WALKER, J., BERNARDINO, A. T., GOPINATH, D. A., MORIKAWA, T. AND POLYDOROPOULOU, A. (2002). Integration of Choice and Latent Variable Models. in *Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges*, (Elsevier Science, Ed. Mahmassani), Chapter 21, 431-470.

BROWNE, M. W. (1984). Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology* 37(1): 62–83.

GREENE, W. (2001). Fixed and Random Effects in Nonlinear Models. Working papers New York University, Leonard N. Stern School of Business, Department of Economics.

HOTELLING, H. (1929). Stability in Competition. *Economic Journal* 39(1): 41–57.

JOHANSSON, M. V., HELDT, T. AND JOHANSSON, P. (2006). The Effects of Attitudes and Personality Traits on Mode Choice. *Transportation Research Part A* 40: 507–525.

JÖRESKOG, K. G. (1979). Structural Equation Models in Social Science. in Karl G. Jöreskog and Dag Sörbom (eds), *Advances in Factor Analysis and Structural Equation Models*, ch. 4, Cambridge, MA: Alt Books.

JÖRESKOG, K. G. AND SÖRBOM, D. (1982). Recent Developments in Structural Equation Modeling. *Journal of Marketing Research* 19(4): 404–465.

LANCASTER, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy* 74(2): 132–157.

MCFADDEN, D. (1986). The Choice Theory Approach to Market Research. *Marketing Science* 5(4): 275–297 Special Issue on Consumer Choice Models.

MUSSA, M. AND ROSEN, S. (1978). Monopoly and Product Quality. *Journal of Economic Theory* 18(2): 301–317.

MUTHÉN, B. (1984). A General Structural Equation Model with Ordered Categorical, and Continuous Latent Variables Indicators. *Psychometrica* 49(1): 115–132.

NELSON, P. (1974). Advertising as Information. *Journal of Political Economy* 82(4): 729–754.

ROSEN, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* 82(1): 34–55.

SWAIT, J. AND SWEENEY, J. C. (2000). Perceived Value and its Impact on Choice Behavior in a Retail Setting. *Journal of Retailing and Consumer Services* 7(2): 77–88.

TRAIN, K., MCFADDEN, D. AND GOETT, A. (1986). The Incorporation of Attitudes in Econometric Models of Consumer Choice. Cambridge Systematics Working Paper.

ZHANG, G. AND BROWNE, M. W. (2006). Bootstrap Fit Testing, Confidence Intervals, and Standard Error Estimation in the Factor Analysis of Polychoric Correlation Matrices.*Behaviormetrika* 33(1): 61–74.

A Descriptive Statistics

Table A.1: Descriptive statistics

Variable	Description	Mean	Std.	Min.	Max.
Stated Choice y	=1 if wood is chosen	0.4330	0.4958	0.0000	1.0000
Variables V^1					
Home thermal insulation	Assessment of insulation	3.8553	1.3393	1.0000	5.0000
Home acoustic insulation	Assessment of insulation	3.8298	1.3514	1.0000	5.0000
Home brightness	Assessment of brightness	4.4319	0.9337	1.0000	5.0000
Home Wooden windows	=1 if windows in wood	0.5372	0.4989	0.0000	1.0000
Home PVC windows	=1 if windows in PVC	0.3128	0.4639	0.0000	1.0000
Change of windows	=1 if recent change	0.0649	0.2465	0.0000	1.0000
Received information	=1 if recent information	0.2543	0.4357	0.0000	1.0000
Informed on windows	Assessment of information	3.0787	1.3152	1.0000	5.0000
Informed on materials	Assessment of information	2.8819	1.3288	1.0000	5.0000
Do-it-yourself	Do-it-yourself rate	3.3202	1.5007	1.0000	5.0000
Purchase of windows	=1 if recent purchase	0.0840	0.2776	0.0000	1.0000
Variables V^2					
Home fitting	Importance of fitting	4.5489	0.7170	1.0000	5.0000
Apartment	=1 if apartment	0.3160	0.4651	0.0000	1.0000
Rooms	Number of rooms	5.2883	3.1098	0.0000	73.000
Owner	=1 if owner (0 if tenant)	0.6660	0.4719	0.0000	1.0000
Modern	=1 if modern building	0.2287	0.4202	0.0000	1.0000
Classic	=1 if classic building	0.5298	0.4994	0.0000	1.0000
Typical	=1 if typical building	0.2298	0.4209	0.0000	1.0000
Variables V^3					
Age	Class of ages	3.2755	1.2068	1.0000	5.0000
Woman	=1 if woman	0.5840	0.4931	0.0000	1.0000
Single	=1 if single	0.1564	0.3634	0.0000	1.0000
One-parent	=1 if one-parent family	0.0372	0.1894	0.0000	1.0000
Couple	=1 if couple	0.2915	0.4547	0.0000	1.0000
Family	=1 if couple with kids	0.4957	0.5002	0.0000	1.0000
Green	Green opinion	3.9734	1.0495	1.0000	5.0000
Finances	Assessment of own finances	3.0021	1.1611	1.0000	5.0000
SPG1	Farmer, artisan, merchant	0.0596	0.2368	0.0000	1.0000
SPG2	Executive, liberal profession	0.1074	0.3098	0.0000	1.0000
SPG3	Intermediate profession	0.1447	0.3520	0.0000	1.0000
SPG4	Employee	0.1106	0.3139	0.0000	1.0000
SPG5	Worker, personnel	0.1787	0.3833	0.0000	1.0000
SPG6	Unemployed, retired	0.3989	0.4899	0.0000	1.0000
City size	City size	2.8777	1.4221	1.0000	5.0000
Reg1	Paris region	0.1660	0.3722	0.0000	1.0000
Reg2	Western region	0.2245	0.4175	0.0000	1.0000
Reg3	North-Eastern region	0.2426	0.4289	0.0000	1.0000
Reg4	South-Western region	0.1213	0.3266	0.0000	1.0000
Reg5	South-Eastern region	0.2457	0.4308	0.0000	1.0000

Notes: $N = 940$.

B Estimation of the latent variable model

B.1 Method

The latent variable model is a structural equation model composed of structural equations and measurement equations expressing relations among latent (non observed) variables and manifest (observed) variables. Structural equation modeling is a multivariate technique using covariance structure analysis. The parameter vector θ is estimated iteratively by an algorithm that minimizes the difference between the sample covariance matrix S (based on the Pearson correlations) and the estimated covariance matrix of estimated parameters $\widehat{\Sigma}(\hat{\theta})$.

Maximum likelihood estimation (MLE) is the most common method used to estimate structural parameters of this model. Multivariate normal distribution and continuous variables are some of key assumptions of MLE. However, the manifest variables are often not continuous (dichotomous or polytomous variables). This is the case in our model. In particular, the measured indicators and some observed variables are based on a five-point Likert scale (ordinal variables). In this case, the assumption of multivariate normality of data does not hold. The parameter estimates of the latent variable model are still convergent but the estimated standard errors are underestimated and the fit measures based on χ^2 values are not good.

In practice, when the number of Likert categories is 4 or higher and skew and kurtosis are within normal limits, use of MLE may be justified. However as recalled by Zhang and Browne (2006), another approach can be to estimate tetrachoric and polychoric correlations[12] respectively from the dichotomous and ordinal variables and to use generalized least squares (GLS) where the weight matrix is the inverse of the matrix of tetrachoric/polychoric correlations, see Muthén (1984). When the model is correctly specified, GLS gives asymptotically valid test statistics and standard error estimates, see Browne (1984).

The discrepancy function to minimize is noted $F = F(S, \Sigma(\theta))$. For GLS, this function is:

$$F_{GLS} = \frac{1}{2} tr(W^{-1}(S - \Sigma)^2),$$

where W is the weight matrix and tr indicates the trace of a matrix.

B.2 Goodness-of-fit indices

An indication of the overall model fit can be given by the closeness of the minimum value of the function F to 0. The overall χ^2 measure is this minimum value multiplied by $N-1$, with N the

[12] Tetrachoric and polychoric correlations extrapolate what the categorical variables' distributions would be if continuous, adding tails to the distribution. As such it is an estimate based on the assumption of an underlying continuous bivariate normal distribution.

sample size

$\chi^2 = (N-1) \times F$

There exist several other indices for assessing model fit:

- The Goodness-of-Fit Index (GFI) represents the overall degree of fit
  $GFI = 1 - \dfrac{Tr((W^{-1}(S-\Sigma))^2)}{Tr((W^{-1}S)^2)}$
- The Adjusted Goodness-of-Fit Index (AGFI) is the GFI adjusted for the degrees of freedom (df) of the model
  $AGFI = 1 - \dfrac{n(n+1)}{2df}(1-GFI)$, with n the number of manifest variables
- The Root Mean Square Residual (RMR) is the mean od the squared residuals
  $RMR = \sqrt{\dfrac{2}{(n+1)} \sum_i^n \sum_j^i (s_{ij} - \sigma_{ij})^2}$
- The Root Mean Square Error of Approximation (RMSEA) index is measure of discrepancy per degree of freedom
  $RMSEA = \sqrt{\max\left(\dfrac{F}{df} - \dfrac{1}{N-1}, 0\right)}$