

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

Santos, Andreia Costa; (2008) Demand for Sanitation in Salvador/Brazil. PhD thesis, London School of Hygiene & Tropical Medicine. DOI: <https://doi.org/10.17037/PUBS.00768488>

Downloaded from: <https://researchonline.lshtm.ac.uk/id/eprint/768488/>

DOI: <https://doi.org/10.17037/PUBS.00768488>

Usage Guidelines:

Please refer to usage guidelines at <https://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license. To note, 3rd party material is not necessarily covered under this license: <http://creativecommons.org/licenses/by-nc-nd/3.0/>

<https://researchonline.lshtm.ac.uk>

Demand for Sanitation in Salvador/Brazil

Andreia Costa Santos

**Thesis submitted to the University of London in fulfilment of the
requirements of the degree of Doctor of Philosophy**

Department of Public Health and Policy

Health Services Research Unit

London School of Hygiene and Tropical Medicine



January 2008

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE



Statement of Own Work

All students are required to complete the following declaration when submitting their thesis. A shortened version of the School's definition of Plagiarism and Cheating is as follows (the full definition is given in the Research Degrees Handbook):

The following definition of plagiarism will be used:

Plagiarism is the act of presenting the ideas or discoveries of another as one's own. To copy sentences, phrases or even striking expressions without acknowledgement in a manner which may deceive the reader as to the source is plagiarism. Where such copying or close paraphrase has occurred the mere mention of the source in a biography will not be deemed sufficient acknowledgement; in each instance, it must be referred specifically to its source. Verbatim quotations must be directly acknowledged, either in inverted commas or by indenting. (University of Kent)

Plagiarism may include collusion with another student, or the unacknowledged use of a fellow student's work with or without their knowledge and consent. Similarly, the direct copying by students of their own original writings qualifies as plagiarism if the fact that the work has been or is to be presented elsewhere is not clearly stated.

Cheating is similar to plagiarism, but more serious. Cheating means submitting another student's work, knowledge or ideas, while pretending that they are your own, for formal assessment or evaluation.

Supervisors should be consulted if there are any doubts about what is permissible.

Declaration by Candidate

I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

Signed: [REDACTED] Date: 11/01/2008
 Full name: ANDREIA COSTA SANTOS (please print clearly)

Abstract

Many studies have demonstrated that improvements in infrastructure have been effective in reducing inequalities due to poverty. The Brazilian Government has invested a significant amount of resources to improve access to sanitation facilities in the municipalities in Brazil in the last decade. One of these programmes is the Bahia Azul programme of sanitation, which aimed to supply sanitation for all the population in the city of Salvador and the surrounding areas. In this programme, households have to pay the costs of the sewer connected to household excreta disposal to treatment plants. So far, models applied to sanitation studies were either misspecified, presenting serious bias, or did not demonstrate the causal relationship among variables. The objective of this study is to assess the demand for sanitation in Salvador, with focus on determinants of the choice for types of connections. Sanitation was assessed as a function of the objective variables (socio-economic and demographic, alternative attributes) and non-observed variables, defined in this study as perception and attitude. The Hybrid Choice Model was the theoretical model used in this analysis. A questionnaire was administered to 721 households. The model was estimated using a sequential estimation, associating a latent model (MIMIC) to a mixed logit model. The analysis showed that the inclusion of latent variables in the model increased the magnitude and significance of the estimation of demand. Results indicated that the more educated and wealthy household tended to choose a system of sanitation. The attributes of usefulness, suitability, convenience, and healthy, *not* the latrine and connection themselves, were what the households really cared about. The results of my investigation supported the appropriateness of the Hybrid model for demand evaluation: latent variables incorporated to a discrete choice model improved the explanation of household behaviour, and filled the gap between behavioural theory and discrete choice models applied to sanitation.

Acknowledgments

I am indebted to a great number of people who generously offered advise, encouragement, inspiration, and friendship throughout my time at London School of Hygiene and Tropical Medicine

I would like to thank my supervisor and mentor, Professor Jennifer Roberts, with whom it has been a pleasure to work. She provided me with opportunities to learn more advanced models, and shared her knowledge about economic theories with me.

Thanks to Dr Vassilis Hajivassiliou from London School of Economics for his comments, suggestions and advice on the econometric model. His support was fundamental for my analysis. Thanks to Dr Kara Hanson, Professor Sandy Cairncross and Professor Laura Rodrigues for truly valued comments, suggestions and advice. Thanks to Professor John Cairns for his support and to Professor Mauricio Barreto from Instituto de Saúde Coletiva, Federal University of Bahia, for the financial support to my research.

Thank you to Valeria de Oliveira Cruz for her wonderful support, encouragement and the fun. London will be difficult without your friendship. Thank you to Nicola Lord for her incredible support and friendship. Thank you to Gizelton Alencar for statistical discussion and help.

Thank you to all the friends and colleagues who have made the process as whole more enjoyable: Maria Claudia do Nascimento, Claudia Agresta, Khatia Munguambe, Karina Keilmann, Werner Leber, Nádia Osman, Naomi Toyoshi (thanks for the psychological support), Giuliano Russo, Phursit Prakongsai, Leslie Kelly (thanks for the last editing revisions), Álvaro Alonso, and Fern Terris-Prestholt (our meetings were of a huge psychological and technical support).

Finally, thanks to Sérgio Cunha for his love, support, and friendship during the whole process, and to my beautiful family, my biggest fans, who provide me with unconditional love.

*For Armando, Graça, Adriana, Armandinho,
Fábio e Leonardo – my beautiful family.*

Table of Contents

Chapter 1: Motivation, Background, Scope and Objectives of this Dissertation.....	13
1.1. Motivation for the work described in this dissertation	13
1.2. Situation of sanitation in the world and the necessity for cultural/historical context based evaluations	16
1.3. The Bahia Azul's programme of sanitation	19
1.4. Scope of sanitation demand study	22
1.5. Objectives of this dissertation	23
1.6. Outline of the dissertation	25
Chapter 2 - Studies of Demand for Sanitation: A Critical Review	27
2.1. Introduction.....	27
2.2. Willingness to pay studies	28
2.2.1. <i>Embedding effect</i>	31
2.2.2. <i>Anchoring effect</i>	35
2.2.3. <i>Willingness to pay as an attitude expression</i>	38
2.3. Probabilistic discrete choice models applied to sanitation studies	42
2.4. Attitudinal-Behaviour Studies.....	44
2.4.1. <i>Jenkins' study of latrine adoption in Benin rural: general critique</i>	47
2.4.1.1. <i>Choice set in decision analysis</i>	51
2.4.1.2. <i>Factor analysis, correlation, ordinal variables, and causal relationship</i>	58
2.4.1.3. <i>Final model to forecast demand for latrine in rural Benin: problems with definition of preferences, intentions, and choices</i>	61
2.5. Summary of discussion.....	64
Chapter 3: Economic Choice Behaviour Literature and Contributions to the Development of a Framework of Analysis for Sanitation Demand Studies.....	65
3.1. Introduction.....	65
3.2. Classical economic theory of consumer behaviour: contributions and limitations in understanding the consumer's decision process of choice	66
3.3. The Lancaster model and RUM hypothesis: developments in economic process analysis	77
3.4. The economic framework for causal analysis	87
3.5. The Hybrid Choice Model.....	86
3.5.1. <i>General specification of the model</i>	89
3.5.2. <i>Identification of the integrated model</i>	93
3.6. Summary.....	95

Chapter 4: A Framework of Analysis for Sanitation Demand Studies	96
4.1. Introduction.....	96
4.2. Choice model for sanitation: framework of analysis.....	97
4.2.1. <i>Historical and cultural dimensions of sanitation: the case of Salvador/Brazil before the implementation of Bahia Azul's programme</i>	97
4.2.2. <i>Sanitation: is it a private or a public good?</i>	101
4.2.2.1. Externalities.....	103
4.2.3. <i>Behavioural conceptual framework for sanitation</i>	104
4.2.4. <i>The mixed logit model</i>	107
Chapter 5 – Descriptive Analysis of a Household Survey of Sanitation Adoption	110
5.1. Introduction.....	110
5.2. Methodological issues	112
5.2.1. <i>Study population</i>	112
5.2.2. <i>Estimated sample size for the epidemiological study</i>	112
5.2.3. <i>Effective sample used for demand estimation</i>	113
5.2.4. <i>Previous qualitative study and questionnaire design</i>	113
5.2.5. <i>Topics included in the survey questionnaire</i>	114
5.2.6. <i>Definition of adopter and non-adopter</i>	115
5.2.7. <i>What was evaluated by the questionnaire (questions about attitude and perception)?</i>	115
5.2.8. <i>Possible problems during interviews</i>	116
5.2.9. <i>Researcher assistant training and data collection</i>	116
5.2.10. <i>Missing data</i>	117
5.2.11. <i>Ethics</i>	117
5.2.12. <i>Data entry and analysis</i>	118
5.3. Results	119
5.3.1. <i>General characteristics of households</i>	119
5.3.2. <i>Perception about sanitation</i>	123
5.3.2.1. Non-adopters	123
5.3.2.2. Difference between adopters and non-adopters	125
5.3.3. <i>Attitude toward sanitation</i>	127
5.3.3.1. Non-adopters	127
5.3.3.2. Difference between adopters and non-adopters	128
5.3.4. <i>Non-adopters: preferences, intention for adoption, drives and constraints for installation</i>	130
5.3.5. <i>Adopters: history of latrine (and connection) installation and disadvantages of adoption</i>	133
5.4. Summary of results.....	136

Chapter 6: Methodology for Estimation of the Demand for Sanitation in Salvador/Brazil	137
6.1. Introduction.....	137
6.2. Study Design.....	138
6.2.1. <i>Overview</i>	138
6.2.2. <i>Objectives</i>	139
6.3. Plan of analysis	140
6.3.1. <i>Multinomial logit model</i>	141
6.3.1.1. Goodness of fit.....	142
6.3.1.2. Testing the independence from irrelevant alternatives (IIA) property.....	143
6.3.2. <i>The latent variable model</i>	144
6.3.2.1. Psychometric indicators and latent variables.....	144
6.3.2.2. The multiple indicators, multiple causes (MIMIC) model: a general definition	145
6.3.2.3. Specification of the latent model	147
6.3.2.4. Determination of the latent model	148
6.3.2.5. Goodness-of-fit and reliability coefficients in latent models.....	150
6.3.3. <i>The discrete choice model</i>	152
6.3.3.1. Determination of discrete model.....	153
6.3.3.2. Halton draw	154
6.3.3.3. Goodness-of-fit.....	155
6.3.3.4. Marginal price (cost) elasticities.....	155
6.4. Summary of Analysis	156
Chapter 7 – Results: Estimation of Demand for Sanitation in Salvador/Brazil.....	157
7.1. Introduction.....	157
7.2. Results	158
7.2.1. <i>Multinomial logit (conditional logit): interpretation and results</i>	158
7.2.1.1. A Hausman-McFadden test	161
7.2.2. <i>The latent variable model</i>	162
7.2.3. <i>The discrete choice model with latent variables</i>	165
7.2.3.1. Price (cost) elasticities.....	170
Chapter 8 – Conclusions and Policy Implications for Sanitation Promotion	171
8.1. General comments.....	171
8.2. Summary of the main findings.....	173
8.2.1. <i>Perception and attitude affecting sanitation adoption in Salvador/Brazil</i>	173
8.2.2. <i>Household’s socio-economic and demographic characteristics affecting demand for sanitation</i>	174
8.2.3. <i>Attributes of alternatives</i>	177

8.3. Contributions to theory.....	178
8.3.1. <i>The inclusion of latent variables in a model to explain household behaviour to sanitation adoption: the contributions to discrete choice models and behavioural theory applied to sanitation...</i>	178
8.3.2. <i>Limitation of my Hybrid Choice Model to estimate demand for sanitation in Salvador/Brazil...</i>	180
8.3.2.1. Estimation procedure	180
8.3.2.2. The possibility of the influence of the interviewer on households answers	181
8.3.2.3. The historical and cultural contexts as factors underlying households perception and attitude towards sanitation: possibility source of bias.....	182
8.4. Forecasting of demand for sanitation: is the accuracy of prediction the core of analysis?.....	184
8.5. Sanitation promotion	186
8.5.1. <i>Implications of result for sanitation promotion.....</i>	186
8.5.2. <i>Delivery and support to household.....</i>	186
8.5.3. <i>Financing schemes.....</i>	187
8.6. Further researches and conclusion	188
References	190
Appendix 1: Questionnaire.....	200
Appendix 2: Information sheet and consent form.....	224

List of Tables

Table 2.1. Willingness to pay for selected classes of goods and allocations of totals to less inclusive groups	32
Table 2.2. Rank correlations between mean evaluation of 37 issues	40
Table 2.3. The universal choice set for telecommuting (original title: the 'universal' choice set adaptive responses to life-style-related-drives)	54
Table 5.1. Socio-economic, Demographic and Characteristics of Sanitation Choice for a Sample of Households in Salvador/Brazil	119
Table 5.2. Perception of qualities (or non-qualities) of open defecation and latrines among a sample of non-adopter households in Salvador/Brazil	124
Table 5.3. Perception of qualities (or non-qualities) of latrines among a sample of adopter and non-adopter households in Salvador/Brazil	126
Table 5.4. Attitude toward latrine among a sample of non-adopter households in Salvador/Brazil	128
Table 5.5. Attitude related to latrine adoption (advantages) among a sample of adopters and non-adopter households in Salvador/Brazil	129
Table 5.6. Disadvantages related to latrine adoption among a sample of adopter households in Salvador/Brazil	134
Table 6.1. Dependent and independent variables used in the multinomial logit analysis (description)	142
Table 6.2. Indicators Variables and Scales for "Attitude" and "Perception" Latent Variables	145
Table 7.1. Estimation Results for the Multinomial Logit Model	160
Table 7.2. Hausman-McFadden Test Statistic	161
Table 7.3. The η Matrix of Factor Loadings (t -statistic in parentheses)	163
Table 7.4. The δ Matrix (z -statistics in parenthesis)	164
Table 7.5. Estimation Results for the Mixed Logit (MXL) Model: Reference Model (Using 125 Halton Draws)	165
Table 7.6. Estimation Results for the Mixed Logit (MXL) Model: Reference Model with Gender Variables (Using 125 Halton Draws)	166
Table 7.7. Estimation Results for the Mixed Logit (MXL) Model: Reference Model (without Latent Variable) and the Latent Variable Enriched (LVE) Model (Using 125 Halton Draws)	167
Table 7.8. Price (cost) elasticities	170

List of Figures

Figure 2.1. Use of latrine in rural Benin: the choice set	57
Figure 3.1. The Choice Process	73
Figure 3.2. Hybrid Choice Model	88
Figure 4.1. Behavioural Framework for Choice of Sanitation in Salvador Brazil	105
Figure 6.1. A Model with Multiple Indicators and Multiple Causes for Construction of a Latent Variable	146
Figure 6.2. Linking Behavioural Framework for Choice of Sanitation in Salvador Brazil with MIMIC model	147
Figure 6.3. Discrete choice model with observed and latent variables	152
Figure 6.4. General Modelling	156

Acronyms

ISC **Instituto de Saúde Coletiva**

CV **Contingent Valuation**

WTP **Willingness to Pay**

WTA **Willingness to Accept**

RUM **Random Utility Model**

Chapter 1: Motivation, Background, Scope and Objectives of this Dissertation

1.1. Motivation for the work described in this dissertation

In September 2003, I was invited by Professor Mauricio L. Barreto from Instituto de Saúde Coletiva (ISC), Federal University of Bahia, to work in the economic evaluation study of Bahia Azul's programme, a sanitation plan organised and implemented by the Bahia State Government. The programme aimed to supply an external sewage system linked to treatment plants for the population resident in Salvador and surrounding areas. In this programme, it was expected (and it is still expected) that the population would pay to connect their household's internal sewer to the external sewage system (see section 1.3)¹.

The economic evaluation study had two main components: 1) a cost-benefit study using the contingent valuation (willingness to pay) approach, and 2) a cost-benefit analysis assessing the costs of diagnosis and treatment of diarrhoea and some intestinal infectious disease, and the benefits, in terms of costs associated with a reduction of prevalence/incidence of these diseases, to the public health sector and households. My active participation was centred in the last study, and involved study design, fieldwork co-ordination, data analysis, and preparation of a report (Barreto et al., 2006).

This economic study was one component of the Bahia Azul's evaluation programme conducted by ISC and co-ordinated by Professor Barreto. The evaluation was centred on the

¹ The name "Bahia Azul" is a play on words with the name of the State of Bahia and the All Saints' Bay for which it was originally named, plus the universal ecological symbolism of blue, as in the blue agenda, and the positive connotation of blue in the Portuguese language (Andrade, 1997 *op cit* in Teixeira et al., 2002).

epidemiological impact, including, anthropological, geographical, nutritional, and parasitological approaches. This interdisciplinary evaluation permitted the creation of a rich database, which included a pre-, and post-assessment of programme results.

My contact with discrete choice models and behavioural literature started at the same time: in the beginning of 2003. As an economist, I have a continuous desire to understand and apply demand studies, mainly new approaches that include psychometric variables as part of the explanation of consumer behaviour. I also had a strong desire to move from the traditional cost-benefit and cost-effectiveness analysis to a demand-side approach. At that time, I had some contact with behavioural literature and discrete choice models. I was following some readings in these areas, with special attention placed on Kahneman's and McFadden's papers on consumer behaviour (both Nobel Laureate). The results reported from the interdisciplinary studies of Bahia Azul's programme seemed a very good opportunity to start my research in the demand/behavioural field.

The anthropological study, where population opinion and perception about the new sanitation programme were evaluated, revealed interesting results. I began considering some questions: How can one measure, model and interpret population perceptions and attitudes related to a specific intervention? Are discrete choice models able to give me a robust answer to predict consumer behaviour related to sanitation intervention?

I received excellent comments and suggestions from my advisory committee on the first draft of my project, which included helpful suggestions of papers and theses that I should read. In particular, the thesis by Jenkins (1999) suggested by Professor Sandy Cairncross (London School of Hygiene and Tropical Medicine), gave me an excellent perspective from which to start my investigation. Based on her exploratory work to investigate demand for latrines in rural Benin, I developed a project of study. Using a similar questionnaire, I started my fieldwork at the end of 2004.

A thorough examination of the econometric literature (doing a course in Advanced Econometrics at the London School of Economics), a continuous contact with demand studies using discrete choice models, and critiques and contributions from psychological studies, have formed my understanding and provided the basis for the development of a framework of analysis and a model for estimating the demand for sanitation in Salvador/Brazil, which is the focus of this dissertation.

1.2. Situation of sanitation in the world and the necessity for cultural/historical context based evaluations

In 2000, the World Health Organisation (WHO) estimated that nearly 18% of the world population still remained without access to safe sources of water, and 40% had no access to any form of improved sanitation services. This represented about two billion people living in urban and rural areas in developing countries without access to those services. As a consequence, these people, most of them children, are exposed to severe health risks, and it is estimated that about 10% of them die every year, especially of the diarrhoeal diseases, associated with a lack of potable water and sanitation services (WHO, 2000).

The relation between poverty, lack of potable water and sanitation, and health problems is clear, and has been discussed in several studies (Kothari, 1987; Jaswal, 1991; Wichmann, 1995; Stephens, 1996, Soares et al., 2002, Checkley et al., 2004). In urban areas, poor people without access to safe water and sanitation are exposed to severe health risks because lack of sanitation leads to water contamination with human faeces, and associated to the absence of safe water and hygiene education, food is contaminated (Esrey et al., 1985). Thus, a vicious poverty circle is formed: poor people cannot pay to have access to safe water and sanitation, therefore, they are exposed to severe health risks; and people with high health risks have difficulty in accessing the labour market or being self-employed, due to health problems, so they have low productive capacity, and consequently they receive low salaries or income, and become even poorer.

Sanitation facilities interrupt the transmission of faecal-oral diseases at its most important source, by preventing human faeces contaminating the water and soil. Epidemiological evidence suggests that sanitation is at least as effective in preventing disease as improvements in the water supply (WHO, 2000). Esrey et al.(1985, 1991) and Esrey and Habicht (1986) reviewed several

studies about the impact of water and sanitation on diarrhoea. Their conclusions were that “*excreta disposal appears to consistently play a more important role in determining children’s health in developing areas than water supplies, especially where the prevalence of diarrhoea is high. This effect was seen for all types of outcomes: morbidity, children growth, and mortality*” (Esrey and Habicht, 1986, page 125). Nevertheless, researchers agree that access to water, sewage system, and hygiene education are actions that should be implemented together with the aim of reducing diarrhoeal diseases (Curtis and Cairncross, 2003).

The United Nations Task Force on Water and Sanitation identified and analysed areas where current low levels of access to sanitation and drinking water (less than 50%) are related to the prevalence of diarrhoeal diseases. Most of these areas include countries in Africa, Afghanistan and Bangladesh. It was estimated that US\$ 2 billion was the amount necessary to improve living conditions for these populations. This would allow 500 million people to have access to safe water and sanitation. The benefits resulting from the improvements to sanitation were estimated as 8 times higher than the total cost of the implementation of these improvements. However, funds were still insufficient for these investments. The report also concluded that most of current financing for water and sanitation (80-90% of current funds) comes from domestic public and private sources, not international ones (UN, 2004; Rijsberman, 2004). These results identify two main important factors: 1) households are paying directly to access sanitation, and 2) understanding what determines their choice relating to sanitation is fundamental.

The traditional contingent valuation approach seems to fail to give an accurate prediction of consumer behaviour. Studies using this methodology have proved poorly suited, biased, and unreliable in providing answers to policy questions facing planners. In addition, the marketing messages promoted by planners based only on faecal-oral transmission of diseases did not seem to match with household perceptions of sanitation benefits in some areas in the world.

In rural Benin, Jenkins (1999) demonstrated that the few latrines installed in that area were paid for the households, and that the determinants of household choice for sanitation were associated with factors other than the health messages disseminated by governmental and non-governmental agencies. Her findings indicated the necessity for deeper investigation of the demand-side aspects, understanding the cultural and historical context, needs, perception and preferences of households, as a way of improving coverage and sustainability. Cultural and historical contexts play an important role in the sanitation market, influencing households' perceptions and attitudes towards sanitation, and should not be ignored when planning interventions².

Previous studies conducted by ISC to evaluate the Bahia Azul's programme indicated the importance of cultural and historical contexts in assessing demand for sanitation. Environmental concerns and perception of sanitation as a basic social right of citizens were some of the aspects mentioned by the households in Salvador, during interviews regarding sanitation conducted by the anthropological team. These results showed the necessity of changing from the traditional economic analysis, focused only on socio-economic and demographic variables to explain behaviour, to the incorporation of more qualitative explanations for household choices (Ben-Akiva, 1999). In this sense, the Bahia Azul's programme as an intervention plan provided an excellent opportunity to explore the power of qualitative variables in the consumer choice evaluation, in view that the model's results could be validated, comparing the periods pre- and post-intervention.

² Other studies also pointed out the necessity of a demand-side approach to understand and improve the access to sanitation in the world (see Cairncross, 1992, Water Supply and Sanitation Collaborative Council, 1994).

1.3. The Bahia Azul's programme of sanitation

In Brazil, sanitation programmes are financed using public funds, from federal and state sources. One of these programmes is the Bahia Azul's programme of sanitation in Salvador, State of Bahia, Northeast's Brazil. The majority of the costs of Bahia Azul's programme were covered by resources from Bahia's government with support of international loans from the Inter-American Development Bank (BID), and the World Bank. Evaluation of the results from this intervention on the health status of the population and other scientific evaluations were conducted by the ISC (Barreto, 2006).

The Programme of Sanitation of the Todos os Santos' Bay (Bahia Azul's Programme) is a programme that aimed to minimise the lack of sanitation and water supply to people living in the municipalities around Todos os Santos' Bay, especially in Salvador, Brazil's third most populous city. This programme started in 1995. It included projects directed at improving the collection and final disposal of solid wastes, pollution control, and environmental and sanitary education. The programme intended to expand the sanitation coverage from 25% to 80% of households. Most of areas in Salvador are now covered by the Bahia Azul's programme. However, there are some areas that are still without intervention. According to the Bahia State Government, those areas will be part of a second stage intervention that aims to cover areas not covered by the first stage of the programme (SEDUR, 2002, 2005).

Salvador has a predominantly service, commercial, and industrial economy, with a nominal GDP of US\$ 5.7 billion, per capita annual income of US\$ 2,300, and a Gini index of 0.587, in 2004 (IBGE, 2004). The median income for the 25% most poor was US\$ 80 per month and average household expenses on food, transport and rent was 88% of total income. The educational level was low: 19% of the total population was illiterate, and 72% spent less than 9 years at school. About 93% of the total population of Salvador had access to water, and 91% to external sanitation

connections (Bahia Azul's programme or other); however, only 70% of the total population had internal excreta disposal connected to a sanitation system - Bahia Azul, septic tank, or other (Barreto, 2006; IGBE, 2004).

Bahia Azul's sanitation programme had two stages: the first (the larger one), aimed to construct external connections for excreta disposal to be linked with the treatment stations. A rigorous environmental control of this excreta disposal was also conducted. The second stage consisted of the linkages between external connections of excreta disposal and internal connection within households providing links from the toilet disposal, bath area, and kitchen sink. This stage should be carried out and paid by each household. The cost of the internal connection to Bahia Azul's sanitation programme was about US\$ 20 (SEDUR, 2002).

The first stage of this programme, the external connection, was concluded. There was some evidence from the epidemiological study that, in Salvador, about 91% of the population have access to some type of external link to sanitation after the implementation of the Bahia Azul's sanitation programme: 75.4% of households with connection to Bahia Azul's programme, and 15.5% with a traditional system of sanitation³. However, only 70% of the population had linkages between internal and external connections (Bahia Azul or traditional system) – Barreto (2006).

In Salvador the cost of construction and connecting the internal system to the Bahia Azul's sanitation programme was estimated to be US\$ 210 (at 2006 prices): US\$ 150 to construct the internal sewage facilities, US\$ 40 to buy a latrine and flush toilet system, and US\$ 20 to link the internal system to Bahia Azul's programme⁴. Considering that the median income of majority of population in Salvador was less than US\$ 80 per month, and these people spend most of it on food, transport and rent, a question regarding the success of the Bahia Azul's programme can be raised: will the poor population of the Salvador be able to bear the costs to implement sanitation and/or

³ Traditional system of sanitation refers to different forms of external connection such as drainage systems, septic tanks and the municipal system.

⁴ Costs included payment of labour costs, when households did not do the work themselves, and excluded maintenance.

connect to Bahia Azul's programme? What are the determinants of households' choice? The answer to this question involves different aspects that are related, not only to the socio-economic and demographic characteristics of individuals, but also, their perception, beliefs, and attitude towards sanitation, influenced by their historical and cultural background.

We need to understand that part of the population in Salvador was able to isolate their faeces; around 70% of population had a latrine inside the home. As revealed by the anthropological studies (see Chapters 4 and 5), the population's perception of sanitation as an important service in preventing diseases and protecting the social environment was strong in Salvador. With a positive perception of sanitation and attitudes towards it, it is expected that, in spite of low income, households have a willingness to adopt the new sanitation programme. Also, it is possible that many of the households that already have a type of external connection may be willing to change their connection to the Bahia Azul's programme. However in this latter case, some incentives should be available as they already are able to isolate their sewer⁵.

Therefore, to evaluate these different aspects of sanitation and forecast demand, it is necessary to use a model where consumer behaviour can be evaluated based also in his/her perception and attitude towards sanitation, and the cultural and historical context to which these concepts are inserted. The traditional demand model is not able to predict behaviour, especially in non-market field, where price and quantity are not directly observed, as in the case of the sanitation market. A model where perception and attitude towards sanitation can be included, and historical and cultural aspects can be discussed, appears to give a more robust answer to forecast demand in special markets. Or, as pointed out by Persson (2001), *"it is not enough to supply a wide range of alternatives unless households appreciate their benefits. Thus, the potential gains from improved understanding of demand side aspects of water and sanitation are substantial"*

⁵ Bahia State Government implemented a Law that aimed to connect all households to Bahia Azul programme. This question is discussed in Chapter 4.

1.4. Scope of sanitation demand study

The subject of sanitation is broad: it is concerned with access to sewage systems (and treatment plants), water supply, garbage collection, environmental education, hygiene behaviour etc. In this dissertation, I concentrated on the demand by households for sewage system facilities (including latrines). Households make their decisions based on their needs and environment, and their decisions do not change frequently over time. In other words, the choice of sanitation is not an intertemporal behavioural decision: my analysis corresponded to a moment in time where households were able to choose from a set of mutually exclusive alternatives.

Household choice for sanitation involves different products and services. It involves a choice of latrines (and equipment, such as a flush toilet) and a choice of a sewage system (Bahia Azul or other). In Salvador, the most common type of latrine is made of ceramic. There are several toilet seats in different styles, most are made of plastic of different quality. High, middle and part of the lower classes have an water closet (W.C.)⁶ inside the home (some households have more than two). Most of them combine the W.C. with the bath area.

As well as Bahia Azul's programme, there are the municipal systems (most of them externally substituted by Bahia Azul's programme), and septic tanks. Also, there are several other combinations of external disposal, but these usually are considered unacceptable for the population. Hence, my demand study for sanitation was related to latrines and different types of connection.

⁶ Toilet and sanitary facilities.

1.5. Objectives of this dissertation

According to Domencich and McFadden (1975), there is a need to develop models of demand which can satisfy the end objectives of planners: 1) maximise social benefits based on a given budget; 2) estimate benefits of alternative designs of a new system; 3) provide a simulation of an urban economy and a projection of a long-term public service needs. To achieve these objectives, it is necessary to develop demand models that are sensitive to policy changes. It requires that the model be causal, establishing the behavioural link between the attributes of the alternatives and the decisions of individuals.

Also, they suggested that the model be flexible, allowing application to a variety of planning problems without major data collection and calibration costs; be transferable from one urban setting to another, allowing reuse without expensive re-estimation in each setting; and be efficient, in terms of providing maximum forecasting accuracy per dollar spent on data collection (Domencich and McFadden, 1975, page 3).

For this dissertation, the model adopted aimed to incorporate those requirements. The main objective of this thesis was to define a framework of analysis that allows the incorporation of psychometric data aiming to estimate demand for sanitation in Salvador/Brazil, explicating the causal relationship among alternative attributes and household characteristics. Therefore, this is essentially a methodological thesis, exploring an econometric model to explain household behaviour for choice of sanitation, testing and illustrating a modelling approach rather than an attempt to develop models for immediate planning purpose, although this topic is briefly explored in the final chapter. Additional objectives included the estimation of household elasticities, and a discussion of the results of a household survey to understand perception, attitude, beliefs and constraints influencing demand for sanitation in Salvador.

The proposed model is intended to be flexible enough to be implemented in other urban contexts (and eventually, rural areas), and efficient enough to give robust answers for forecasting demand for sanitation in developing countries.

No pretence is made of having developed a definitive model for any of the choices being examined. However, the statistical results of the empirical analysis are highly encouraging, signalling that the proposed model (the Hybrid Choice Model) could be a good instrument for sanitation demand estimation.

1.6. Outline of the dissertation

The dissertation is organised as follows:

Chapter 2 presents a critical review on demand studies for sanitation. There is a general discussion on contingent valuation method, without a specific focus on sanitation studies using this method. I opted for this approach because the general critique is applicable to any field where this methodology is used. My main conclusion about this method was that contingent valuation approach is an inefficient and biased method, and that it does not evaluate what it intends to evaluate. In the chapter 2, I also conducted critical reviews for studies applied to sanitation. My special attention was focused on the Jenkins thesis (1999). Her study offered an important way to evaluate sanitation, exploring psychometric data for evaluation. However, her model did not test and demonstrate a causal relationship among variables, and so, cannot be considered a model to predict behaviour.

In *Chapter 3*, I review the traditional economic consumer theory and its contribution to the development of discrete models. I believe that, in spite of several limitations of this theory, it launched the basis for the recent development in economic behaviour theory, with models that (now) allow the incorporation of variables that give more realistic and robust answers to questions about consumer behaviour. The concept of indirect utility maximisation, developed by Lancaster (1966 a,b), is presented and conjunctly with the economic framework for causal analysis will form the core of the proposed framework of analysis for assessing the demand for sanitation.

In *Chapter 4*, I discuss a framework of analysis for assessing the demand for sanitation. My framework is based on the Hybrid model. Findings from the anthropological study contribute to the discussion and construction of my assumptions for the household's choice of sanitation in Salvador.

The *Chapter 5* presents the results from a household survey of sanitation adoption, exploring the differences between adopters and non-adopters. The history of latrine adoption, constraints, and beliefs related to sanitation, perception and attitude of households are discussed. The focus of

analysis was the exploration of the households understanding about sanitation, and their perception and attitude towards it. Both these latent variables will be part of the demand analysis.

The methodological details for estimation of demand for sanitation are presented in *Chapter 6*. Here, the specification and determination of both models, latent and discrete choice models are discussed, as well as the goodness of fit of these models.

In the *Chapter 7*, I present the results of demand model. The multinomial logit is applied to test for multicollinearity. The indicators chosen to form the perception and attitude latent constructs are integrated, and causal relationships with household's socio-economic and demographic characteristics, and alternative's attributes are tested. The final discrete choice model with and without latent constructs is presented. Clearly, the introduction of latent variables improves the magnitude and significance of estimated coefficients.

The *Chapter 8* provides a general discussion on results, limitations of study, recommendations and directions for further research.

Chapter 2 - Studies of Demand for Sanitation: A Critical Review

2.1. Introduction

The objective of this Chapter is not to present an exhaustive literature review about the demand for sanitation (in particular the adoption of latrines and sewers), but to present and discuss a summary of the methodology used in these studies, focusing on their limitations and possible contributions to the development of models to forecast demand for sanitation.

Some methods of evaluating demand were considered misspecified, including serious bias problems, as is the case of contingent valuation (willingness to pay surveys). Critiques for this method were not specifically directed to sanitation studies, but to the method in general. Nonetheless, all critiques are relevant to research on sanitation, as the methodology employed in these investigations is similar.

Other studies, such as the probabilistic discrete choice models, especially Jenkins' research (1999) of latrine adoption in rural Benin, present an important contribution to the development of consumer behaviour examination of demand for sanitation. The results of Jenkins' research influenced several studies in marketing sanitation, and it was the inspiration behind the determination of variables for the model presented in this thesis (Chapter 4 and 6). However, some limitations in this research prevent it from being considered as a true behavioural model to forecast demand for sanitation. Their limitations and contributions will be discussed in the following paragraphs. This discussion launches the basis to the proposed framework of analysis of demand for sanitation presented in the Chapter 4.

2.2. Willingness to pay studies

In September 2006 I conducted a search of PUBMED, COCHRANE, and LILACS databases, Google scholar, and JSTOR business and economics journals, interrelating the words “willingness to pay”, “contingent valuation”, “sanitation”, “sewage”, “sewer”, and “latrine”, and found 1,154 citations, books, articles, and reports. My objective was to update a previous search I conducted between January and February 2006 about willingness to pay (WTP) studies and sanitation (latrines and sewers)⁷.

The application of contingent valuation methodology to evaluate demand for sanitation has been disseminated over the last decade (Whittington et al., 1993; Whittington, 1998; Griffin et al., 2004). As individuals do not generally purchase sanitation directly, because it does not have a direct market price analogue, economic techniques which evaluate non-market goods have become quite popular. Among those, contingent valuation is the most celebrated methodology used to evaluate sanitation interventions. Other techniques for evaluating non-market goods include the travel-cost method (Clawson and Knetsch, 1966) and the hedonic-method (Rosen, 1974).

Contingent valuation uses survey research procedures to measure (or try to measure) the WTP of consumers, their preferences, and to predict adoption. Individuals are asked how much money they would be willing to pay for an incremental change in the provision of sanitation using a contingent market (Thayer, 1981; Regens, 1991). Supporters of contingent valuation have argued that properly elicited statements of WTP reveal genuine economic preferences related to economic consumer theory (Hoehn and Randall, 1987; Mitchell and Carson, 1989; Smith, 1992).

⁷ Most of those articles, books and reports were related to water rather than latrines and sewage/sewer, as general definition of the word “sanitation” also includes access (or lack of access) to water and garbage collection (see Chapter 1). As the majority of studies assume a similar framework of analysis, I chose a general critique of these studies, focusing on results of experiments which confronted contingent valuation findings rather than a particular discussion of sewage intervention studies.

The presupposition of this method is that the measures of the maximum amounts people will pay to avoid a loss (willingness to pay) and the minimum compensation necessary for them to accept it (willingness to accept - WTA) are expected to be equivalent (Knetsch and Sinden, 1984). However, several experiments have questioned the credibility of this methodology to demonstrate this equivalence and to describe preferences (the economic value of a good), and forecast demand, and so contribute to policy programmes.

In defence of this technique, some economists have argued that discrepancies between willingness to pay and willingness to accept are due to income or wealth effects which are the limits of available income or wealth that an individual could be willing to offer to pay (based on their wealth) in contrast with non-commensurate and non-restricted value to receive as a compensation (Krutilla and Fisher, 1975). Nevertheless, different experiments have presented evidence that, aside from possible income (or wealth) effects, contingent valuation studies are inconsistent with economic theory and they do not measure the preferences/utilities they attempt to measure (Diamond and Hausman, 1994). In addition, conclusions from the literature state that income and wealth effects do not appear to make much difference in that analysis (Freeman, 1979).

Defenders of the contingent valuation method also argue that most of the experiments carried out to demonstrate failures of that approach are not based on real transactions, and so respondents could feel free to give unreliable answers with a view to obtaining personal benefits (Dwyer and Bowes, 1978; Brookshire et al., 1980). To verify this possibility, Knetsch and Sinden (1984) conducted five tests using real transactions, and concluded that the two measures of assessment (willingness to pay and willingness to accept) were not as equivalent as it was supposed they should be. On the contrary, the difference between both measures demonstrated a high disparity in their values.

In one of their experiments with 76 students at the University of New England, they distributed lottery tickets of two colours after students entered a classroom. Tickets were distributed randomly among participants. The prize consisted of either \$70 worth of merchandise vouchers to be spent at a local shop chosen by the winner, or \$50 in cash. One half of the randomly selected participants were asked to pay \$2 to keep their ticket in the draw for the prize. Their options were paying for the ticket or refusing payment. The other half were allowed to take part in the raffle for free, but they were offered \$2 to forgo their chance (willingness to accept). No single participant was presented with both options. Discussion of chances and plans among participants was actively discouraged, and negotiations were done individually to minimise (or eliminate) informal influence and information flows among them. To control income or wealth effects, each individual participant was required, one by one, stop at the cash desk to discuss the options; thus, transaction costs were made equal (or nearly equal) for all of them.

Based on contingent valuation presuppositions, it was expected that the percentage of people paying \$2 should be the same as the percentage refusing \$2. However, experimental results showed a large variance between the measures: of the 38 people offered the opportunity to pay \$2 to participate in the raffle, 50% were willing to pay for the prize opportunity. Of those offered compensation, 76% were willing to accept the \$2, and refuse to participate in the prize opportunity. This difference was statistically significant at the 5% level. Similar results were also observed in the other four experiments conducted by the authors.

Their experimental results reinforce evidence from previous studies in suggesting a wide disparity between the two measures of contingent valuation (WTP and WTA), and negate the income and wealth effect as being an adequate explanation of the variation between the two measures (Knetsch and Sinden, 1984).

Other unusual features have been observed in experiments using contingent valuation analysis, putting serious concerns on the reliance of that approach as a demand evaluation assessment method (and government decision making tool). In general, these features were named as (1) an embedding effect, and (2) an anchoring effect (Kahneman and Ritov, 1994; McFadden, 1994; Diamond and Hausman, 1994; Diamond, 1996; Kahneman et al., 1999). From this point, the terms “contingent valuation” and “willingness to pay” will be used as synonyms.

2.2.1. Embedding effect

The embedding effect is the most important and discussed anomaly in contingent valuation studies, and it is observed when the WTP for the same good varies widely depending on whether it is assessed on its own or embedded as part of a more inclusive good. In this manner, the problem for the interpretation of contingent valuation studies is: which measure is the correct one? As the discussions in the literature do not provide a satisfactory answer and experimental control against the embedding problem, the results of contingent valuation method become arbitrary, presenting different measures of the same object (Kahneman and Knetsch, 1992).

One of Kahneman and Knetsch’s (1992) experiments evaluated the WTP for the increased availability of equipment and trained personnel for rescue operations in disasters. Three samples of adults living in Vancouver, Canada, were interviewed by telephone⁸. Their investigation occurred (coincidentally) within weeks of the San Francisco earthquake of 1989, which enhanced the relevance of the topic.

For the first sample, respondents were presented with the environmental services which included: preserving wilderness areas, protecting wildlife, providing parks, preparing for disasters, controlling air pollution, ensuring water quality, and routine treatment and disposal of industrial

⁸ There is no evidence that psychometric biases are more severe in telephone surveys than in personal surveys (McFadden, 1994).

waste. They were then asked for their willingness to pay (*per year*) for improved environmental services. If the respondent's answer was zero, his/her interview was concluded. Other respondents were asked in particular about improved preparedness for disasters in terms of his/her willingness to pay. Subjects were allowed to answer by stating a dollar amount, a fraction, or a percentage. Then, they were asked what part of the total amount to improve preparedness they thought should be allocated specifically to improve the availability of equipment and trained personnel for rescue operations.

For the second sample, an initial question referred to a special fund to improve preparedness for disaster with a subsequent allocation to go specifically to improve the availability of equipment and trained personnel for rescue operations. In the third sample, they answered on preparedness for disasters and were asked to state their willingness to pay into a special fund to improve availability of equipment and trained personnel for rescue operations. Table 2.1 presents the authors' findings.

Table 2.1. Willingness to pay for selected classes of goods and allocations of totals to less inclusive groups

		Sub-Sample		
		Group 1 (N = 66) (\$)	Group 2 (N = 78) (\$)	Group 3 (N = 74) (\$)
Public Good				
Environmental services	Mean	135.91		
	Median	50.00		
Improve disaster preparedness	Mean	29.06	151.60	
	Median	10.00	50.00	
Improve rescue equipment, Personnel	Mean	14.12*	74.65**	122.64
	Median	1.00	16.00	25.00

*Two respondents did not answer this question, reducing N to 64; **Four respondents did not answer this question, reducing N to 74 (authors' notes).

Source: Kahneman and Knetsch, 1992, page 61

Results demonstrated a large embedding effect: estimates of WTP for the same good differed depending on the scope of the initial question. The difference factor was 16 for medians or 8 for means. The questions posed to respondents were hardly affected by the inclusiveness of this good. The percentages of positive contributions were 61% for improvements in “the availability of equipment and trained personnel for rescue operations”, 63% for “preparedness for disaster”, and 65% for “all environmental services”. The lowest level of willingness to pay was \$25, and the highest level was \$50, but the difference was not significant.

As the data included extremely high values from some respondents⁹, the means of WTP were affected by those extremes. To correct this problem, estimates were based on medians, using all responses, and according to the authors, the qualitative conclusions were not affected by this. They also considered the zero responses to the calculation, from those respondents who answered zero for the questions (Kahneman and Knetsch, 1992).

It is standard practice in contingent valuation analysis to eliminate the zeros (the so-called “protest zero”) and high values (viewed as a non-true willingness to pay), according to WTP users. This practice is frequently adopted without any criteria (Diamond and Hausman, 1994).

Recent studies have attempted to address this with, for example, a dichotomous-choice contingent-valuation methodology. This method was proposed by Arrow et al. (1995) aiming to solve the problem with outliers. The dichotomous format asks for a “yes” or “no” response for the individual’s WTP (a randomly assigned amount) for a good or service, instead of asking for a maximum amount that he/she is willing to pay. The result is estimated through the cumulative frequency distribution of positive responses. However, this approach seems to exacerbate the problem of eliminating zero and large magnitude responses, affecting the mean value estimated (Green et al., 1998). According to Werner (1999), “*to date, dichotomous-choice contingent*

⁹ Misrepresentation interferes with the correct comprehension of the question for the respondent, which can lead to extreme values.

valuation data analysis typically has assumed that agents have a positive (although possibly small) willingness to pay [...] For the class of parametric distributions considered, to allow a nontrivial fraction of the distribution to fall close to 0, if most agents have a substantial willingness to pay for the good, the right tail of the distribution is forced to increase, leading to large and implausible estimates of the mean. Often 25% or more of agents surveyed indicate that they are not willing to pay even a small amount, however, indicating that these models are misspecified” (page 479).

Regarding the wealth and income effect, and returning to the Kahneman and Knetsch (1992) experiments, the WTP values were small relative to reported incomes; therefore, the embedding effect could not be explained by constraints of wealth or income. For the first sample, the median stated willingness to pay was \$37.50 for families with an income under \$20,000 (23%), \$50 for an income between \$20,000 and \$40,000 (39%), and \$100 for families with an income over \$40,000 (38%). The corresponding means were \$97, \$131, and \$230.

From the strictly economic theory point of view, the embedding effect is usually thought to arise from the non-existence of individual preferences. Quoting an experiment conducted by Schulze et al. (1993) where they asked for WTP for partial and complete clean up of contamination of the Clark Fork National Priorities List Sites in Montana, Diamond and Hausman (1994) discussed the non-existence of economic preferences in willingness to pay studies. Whether or not clean up of contamination has some utility value to the respondent, and he/she is expressing his/her true preferences, it is expected that a complete cleanup would be worth more than a partial one. Nonetheless, the willingness to pay amount was almost the same for both alternatives (\$72.46 for the first, and \$72.02 for the second; outliers were removed). In addition, respondents were asked whether their responses were just for this clean up (in Montana), or partly to clean up other sites, or a contribution for all environmental, or other causes. The majority of respondents (73%) recognised an embedding effect in their answer: just 17% answered “just for this clean up”. To those 73%, it

was asked what percentage of their previous values was just for this clean up, and after adjustments, the mean WTP for a complete clean up was \$40, and for a partial clean up was \$37.15. In conclusion, Diamond and Hausman (1994) argue that the contingent valuation method fails in assessing economic preference due to internal consistency problems. Consequently, this is an inappropriate tool for government decision making.

2.2.2. Anchoring effect

This effect occurs when in multiple questions, WTP answers change when the sequence of questions asked is changed. The point of critique here is that, if a survey question reveals a true valuation, it should not matter whether the question is asked by itself or with other questions (embedding effect), nor on the order of questioning (anchoring effect) (Kahneman and Ritov, 1994).

An example of the anchoring problem was shown by Samples and Hollyer (1990). In a survey, they asked for values of preserving seals and whales. A group of respondents were asked for their WTP to preserve seals first, followed by a question about whales. Another group was asked for willingness to pay to preserve whales first, and after that, to preserve seals. While whale values were not affected by the sequence of questions, seal values tended to be lower when asked after whale values. To the authors, respondents behave more benevolently towards whales than seals, because whales are more popular. Therefore, in the seal-whale evaluation whale values were inflated. However, this behavioural anchoring effect did not exist in the whale-seals version, where whales were valued first.

In another experiment, Kahheman et al., (1999) conducted two different studies to evaluate whether when the objects belong to different categories, comparisons and isolated valuations can yield discrepant results. The same pair of issues was used in both studies: damage to coral reefs caused by cyanide fishing in Asia, and increased incidence of multiple myeloma among the elderly.

They assert that the latter issue would be perceived as a fairly minor public health problems, whereas a threat to coral reefs would appear significant in an ecological context. Also, those public health problems would be assigned a higher general priority than ecological problems, but that this priority would only become relevant in a direct comparison.

The participants were first asked to evaluate one problem; they were then shown the other problem and were asked to respond to it, with an explicit instruction to consider both problems before responding. The independent variable was the order of presentation of the two problems. The participants in Study 1 were 100 visitors at the San Francisco Exploratorium. They were asked to rate the importance of each problem and the satisfaction they would expect to get from making a contribution to its solution. The data for Study 2 are drawn from a larger study, in which the participants were jury-eligible residents of Austin, Texas. Some participants (N=130) provided ratings of satisfaction as in Study 1. Others (N=261) indicated their WTP to contribute to a solution. When they encountered the second problem they were instructed to treat it as the only cause to which they would be asked to contribute.

The authors' hypothesis about the role of context in judgments predicts a statistical interaction effect in each of the panels: the difference between the valuations of the myeloma and coral reefs problems is expected to be larger when these items appear in the second position than in the first. The rationale for this prediction is that the difference between the categories of ecological and human problems is only salient when the issues are directly compared, not when they are valued in isolation. The predicted interaction was highly significant ($p < 0.001$).

The anchoring effect was also consistent even in referendum contingent valuation, where the dichotomous-choice (yes or no) method is used. Jacowitz and Kahneman (1995 *cited in* Green et al., 1998, page 93) conducted an experiment in two phases. Respondents in the first phase provided open-ended estimates of fifteen uncertain quantities. The 15th and the 85th percentiles of the

responses to each question were selected as anchors for the second phase. Subjects in two new groups, recruited from the same population as before, answered two questions for each quantity: they first evaluated whether a number (one of the anchors) was higher or lower than the quantity, then they estimated the quantity. The novel finding of the study was that the percentage of positive responses ("the quantity is higher than X") for subjects shown a high anchor was closer to 30%, instead of the value of 15% which is expected if the open-ended and the referendum questions probe the same underlying belief. This robust effect was limited to high anchors, probably because the uncertainty about the quantity is asymmetric when zero provides a firm lower bound. It appears that the consideration of a proposition (e.g., that the tallest redwood is more than 4000 ft tall) tends to increase the plausibility of that proposition. This suggestion effect is apparently quite automatic, although it can be justified if the subject's reason that a value mentioned in a question is unlikely to be absurd.

All the experimental results until now have concluded that the contingent valuation method does not elicit consistent statements of economic preferences. This methodology is strongly affected by embedding and anchoring problems, and its results are systematically biased, and so is a non-reliable method to economic analysis. Those effects have proved to be remarkably consistent across WTP studies, even though some strategies to reduce its effect were adopted.

However, some authors defend that arguments against contingent valuation methodology is based mainly on studies of public goods, and studies of private or mixed goods or services (e.g. water, sewage) show more consistency in responses (Whittington et al., 1993; Griffin et al., 1995).

Griffin et al. (1995) compared the findings of a survey conducted with families in Kerala State in India to ascertain their willingness to pay for household connections to a piped water supply system. The original survey was conducted in 1988. Families were surveyed again in 1991 to compare if they behaved as they were predicted to behave. The accuracy of their findings was high.

In spite of the accuracy of their forecast, these results did not provide evidence that contingent valuation is a reliable method. In fact, accuracy can happen by chance, and meteorology scientists are aware of this possibility. The main question posed on contingent valuation method is if the approach is psychometrically robust, in that results cannot be altered substantially by changes in survey format (to test embedding and anchoring effects), and instructions that should be inconsequential when behaviour is driven by maximisation of rational preferences (McFadden, 1994). The authors recognised that regardless of their accurate results, contingent valuation still presented a biased model in assessing preferences.

2.2.3. Willingness to pay as an attitude expression

In the middle of the debate about contingent valuation misspecification, a different proposal has been put forward to explain what in fact people have expressed through their statements of WTP in contingent valuation surveys. The main hypothesis comes from studies conducted by psychologists who have suggested that statements of WTP are, in fact, individual's attitude expressions, which capture affective value (positive and negative, good and bad, etc.), rather than economic preferences (Kahneman et al., 1999).

Kahneman and Knetsch (1992) were the first authors to develop the hypothesis that respondents were expressing a willingness to acquire a sense of moral satisfaction by a voluntary contribution to the provision of a public good. This hypothesis was also supported by economists who have developed critical studies on the willingness to pay approach (Diamond and Hausman, 1994). Kahneman and Knetsch (1992) stated that *"public goods differ in the degree of moral satisfaction that they provide to the individual making a contribution. Saving the panda may well be more satisfying for most people than saving an endangered insect, and cancer research may be a better cause than research on gum disease. The quality of causes as sources of moral satisfaction will*

reflect individual tastes and community values. Our first hypothesis is that differences in willingness to pay for various causes can be predicted from independent assessments of the moral satisfaction associated with these causes” (page 64).

Recent development of this hypothesis assumes a more general assumption, involving concepts borrowed from social psychology, and confronts ideas with economic theory, mainly the definition of economic preferences in behavioural approaches. The central idea is that in contingent valuation surveys, each individual's moral satisfaction is an expression of his/her attitude to an environmental problem. A dollar (or any other currency used for rating WTP) is just a scale of an *affective value* (that can range from extremely positive to extremely negative) similar to that used in semantic-differential scales. Thus, respondents are expressing their *attitudes* towards some environmental good rather than their *preferences* (economic preferences) related to it (Kahneman and Ritov, 1994; Kahneman et al., 1999).

Results from different experiments have demonstrated strong consistency of statements of WTP as an expression of attitude (Payne et al., 1999). Comparing different affective scales, Kahneman and Ritov (1994) studied the valuation of 37 topics including environmental problems and other public issues. Participants for the research were visitors at the San Francisco Exploratorium. The total sample size was 1441, and the number of respondents to any particular version of a problem was 50-115. An example of an issue was “fertility loss due to pollution threatens several species of reptiles on the Mexican coast”. Proposed interventions varied depending on the three versions of the reptile problem presented to respondents, but in general could be represented by the follow intervention: “support program to increase fertility by hormonal treatment”. Different scales were used: stated willingness to pay (SWTP), degree of political support for the intervention (a 0-4 rating scale), personal satisfaction expected from making a voluntary contribution (also a 0-4 scale), and

rating the importance of the problem as a public issue (0-6 rating scale). The results are presented in Table 2.2.

The correlation between the mean evaluations for the 37 issues was high and consistent even using different scales. The authors asked what prediction of moral satisfaction, statements of political support and an indication of willingness to pay have in common, and concluded that *“these expressions share a common affective core, which is so prominent that it allows the public attitude order over objects to be measured almost interchangeably by ostensibly diverse responses”* (Kahneman et al., 1999).

Table 2.2. Rank correlations between mean evaluation of 37 issues

	SWTP	Support	Importance	Satisfaction
SWTP	(.87)			
Support	.84	(.85)		
Importance	.76	.84	(.88)	
Satisfaction	.84	.87	.85	(.90)

Source: Kahneman and Ritov, 1994, page 15

Many economists agree with Kahneman and colleagues that SWTP does not measure the economic preference that it attempts to measure, and accept in some sense that what respondents are evaluating is an opinion poll on concern about the environment in general rather than preferences (Diamond and Hausman, 1994; Varian, 1999). However, Kahneman and colleagues also concluded that the concept of economic preferences is not itself valid (Kahneman et al., 1999). This statement has generated some debate.

Varian (1999) shared the same opinion with Kahneman and colleagues regarding the capacity of willingness to pay study to measure preferences, but disagreed about their conclusion that economic preference is not a valid concept. He defended that the concept of preferences is useful to

define social choices, mainly because society must make choices about consume, and this experimental design (economic preferences) provides strong evidence in assessing social choice, recognising however, that this theory is an idealisation, and so it is not expected to hold exactly.

In spite of the divergence on the existence or not of preferences to determine consumer behaviour, the conclusions from diverse authors to contingent valuation studies are the same: this method is misspecified, biased, and so, unreliable for governmental/intervention decision making. Or in Varian's words (1999, page 241), "*talk is cheap and economics is concerned with scarcity. Perhaps this is why economists have little faith in concepts like 'stated willingness to pay'*".

As an alternative to the misspecified, biased and unreliable contingent valuation method, some researchers have adopted more robust methodologies to assess demand for sanitation. These methods, based on the utility maximisation assumption, provide a more reliable assessment for demand for sanitation, and are further discussed in the following sections.

2.3. Probabilistic discrete choice models applied to sanitation studies

Discrete choice model is a reliable method of evaluating the behavioural process of choice. Originally developed by Warner (1962), they were improved and acquired their most modern form through research developed by McFadden (1974) and applied to a choice of transport modes. Assuming different variety and applications, in probabilistic discrete choice models consumers are presumed to have made a choice from a discrete set. These choices are ranked by utility measures and, often, based on revealed or stated preferences. These models are flexible enough to account for different characteristics of alternatives that compose a choice set. The most popular models are the multinomial logit model (MNL), nested logit (NL), conditional logit (CL), and mixed logit (ML) (Domencich and McFadden, 1975; Hensher and Greene, 2003). Chapter 4 of this thesis presents the properties and advantages in using these methods, in particular, the mixed logit model.

There are few studies using probabilistic discrete choice models applied to sanitation. Apart from the Jenkins study (1999) which will be the focus of a detailed analysis in the following paragraphs, there is one study involving one of these methods and the demand for water (Alaba and Alaba, 2002), and three studies involving water and sanitation (toilet facilities) – Persson, 2001 and 2002; and Abou-Ali, 2003.

Persson (2001) estimated the demand for water and sanitation in Bangladesh. The objective of the study was to analyse household choice of toilet facilities and drinking water sources for a sample of households in that country. Using a conditional indirect utility function (conditional since it depends on the price of the chosen alternative and not all prices), the results revealed that the educational level of the household was an important determinant of household choice of toilet facilities.

A further study in Egypt showed that the educational level and region of residence (rural or urban) affected household sanitation choice. People who lived in urban areas and had a medium or

higher education level tended to choose services for sanitation that had a better quality - modern facilities with flush pit latrine (Abou-Ali, 2003). Despite the sophistication of the models used to estimate demand in these two studies, the number (and quality) of variables reduced their explanatory power.

In Persson's study, the dependent variables to estimate demand for toilet facilities were "no toilet facility at all", "traditional type of toilet", "slab toilet provided by UNICEF", and "modern type of toilet". The independent variables were defined only as "landholding size", "number of people in the household", and "number of years of schooling of the household head". Psychometric variables were not included. Abou-Ali's study also presented a few variables to explain demand.

In consumer theory, the decision process of choice behaviour is believed to incorporate a series of psychometric variables which are important for the explanation of variations in choices across individuals. Many economists have agreed that consumer decisions are a complex process and cannot be explained only with the traditional socio-economic and demographic variables. Apart from socio-economic/demographic variations, additional variations in demand are not seen as a simple *ad hoc* disturbance, but as an essential part of the decision process that can be explained by different beliefs, perceptions, motives, attitudes and contextual and historic variables (Simon, 1959; Lancaster, 1966a,b; McFadden, 1994).

This thesis also assumes that consumer choice behaviour is a complex process, and that sanitation is a type of service for which the traditional economic theory of demand is not able to give reasonable explanations for consumer choice. In view of the enormous complexity of the decision tasks, utility maximisation clearly transcends human cognitive abilities, and it becomes obvious that consumer choices are probably inefficient most of the time (Lancaster, 1966a,b).

2.4. Attitudinal-Behaviour Studies

Experimental evidence has demonstrated that traditional economic theory based on rationality-maximisation assumption has failed to explain consumer behaviour, specially in non-market contexts (McFadden, 1998). This theory assumes that the consumer is not only rational in making choices, but also a maximiser who will settle for nothing less than the best. In this context, only the rational-maximiser survives, and the volatile consumer who makes inefficient choices is punished (with high prices, non-optimal choice) in competitive markets (Simon, 1959). Some studies from economics, social psychology and psychology have demonstrated that, in view of the complexity of the decision tasks, it is expected that consumer choices are probably inefficient most of the time, and therefore, an explanation for the process that generates this volatility should be elaborated (Simon, 1959; Lancaster 1966a,b; Kahneman *et al.*, 1999).

Behavioural decision studies have investigated and demonstrated that the so-called inefficiency seen in the rational-maximiser model is related to several elements of cognitive process that involve perceptions and beliefs based on available information, and a process of learning and adaptation which is strongly dependent on history and context, and influenced by attitudes, motives, and, from the point of view of economists, by preferences (Lancaster 1966a,b; Tversky and Kahneman, 1981; McFadden, 1998)¹⁰. From this, a variety of approaches to evaluate the decision-making process based on behaviour decision theory have been developed, for different fields of study. Those approaches have been broadly characterised as attitudinal-behaviour studies, in view of their aim to identify the cognitive elements influencing the behaviour decision process (Ajzen and Fishbein, 1980; Jenkins, 1999).

¹⁰ In chapter 3 a more detailed discussion involving traditional economic theory and behavioural theories based on cognitive process is presented and a framework of analysis is proposed.

Studies in sanitation focusing on the consumer decision-process have been scarce. Most research has been on the contingent valuation method and water access/demand (Whittington, 1998). Supply side investigations dominated studies on choices of excreta disposal facilities in developing countries over several years. However, sanitation experts have seen the necessity for research into the demand side of excreta disposal, and some studies have been carried out (Jenkins, 1999).

A pioneer study based on attitudinal-behaviour and consumer choice theories was conducted by Jenkins (1999) which aimed to examine the decision of private households to install a pit latrine in rural Benin. She investigated the role of several elements of the cognitive process, including cultural, historic, and context based information, perception formation, motives, drives and barriers to the use of latrines in a rural area where no sanitation programme had previously been implemented, and latrine implementation was a private household decision.

Her research led to important findings with implications for marketing promotion. Traditionally, marketing promotion for latrines was based on preventing faecal-oral transmission of infectious disease and hygiene behaviour promotion, and technical language that was difficult for the population to understand and absorb. In fact, most of their motivation to install latrine had nothing to do with those elements. Jenkins demonstrated that, in rural Benin, individuals were motivated by other factors when installing a latrine. Most important factors, in order of frequency of mentioning, were prestige, comfort, physical safety and personal protection, privacy, restricted mobility and health. Health and safety concerns consisted of personal protection from supernatural dangers (some citizens in rural Benin were practicing Voodoo), or family protection from mundane injuries and accidents. Thus, one of the factors that could explain the weak demand for sanitation in developing countries was the “insensibility” of latrine promotion programmes, their failure to incorporate individual’s real motives for adopting latrines, by focusing only on the infectious

diseases issue. Other factors included a lack of financial resources and the high costs of latrine installation.

Jenkins used an anthropological approach for qualitative in-depth interviews with 40 respondents (25 were adopters). A questionnaire based on these interviews was developed and applied to a sample of 320 households to study attitudes, cultural, social, and historic aspects, drives and constraints acting on consumer behaviour. Household attitudes and perception, drives and constraints were measured by a semantic differential scale (Osgood et al., 1975, *op cit* by Jenkins, 1999). In this scale, qualities, and non-qualities, advantages, and disadvantages are stratified in scores that varies from 1 to 7, for instance, where 1 can mean very dirty, 7 very clean, and 4 neutral. She also used some open questions where the interviewee answered spontaneously on some specific topics, as well as on intention to adopt (to non-adopters).

Results from her research have influenced marketing promotion in developing countries to improve demand for sanitation, and changed the focus from traditional socio-economic/demographic characteristics *versus* cost variables (traditional analysis of demand) to analysis of consumer behaviour focusing also on cultural and contextual dependent aspects. However, her attempt to construct a *causal* behaviour model to forecast demand (preference, intention and choice) was not successful. Her model included a series of restrictions that prevented it being a true behavioural model. Causal relationships among variables was not tested or demonstrated. The relationships among variables should be understandable and demonstrate a causal mechanism between the household decisions and the attributes of the latrine, not only replicate the results of conditions existing at the time of the survey. It is important to understand why individual's decisions vary as conditions change, and this is only possible if causal relationships among decision variables are understood (Domencich and McFadden, 1975).

To understand why Jenkins' model did not establish a causal relationship between the attributes of the latrine and the decisions of the individuals to determine a behavioural model, a summary of her thesis will be presented in the section 2.4.1. The critique is focussed on ideas and techniques argued by Jenkins which will be useful to construct my framework of analysis for sanitation, as discussed in Chapter 4.

2.4.1. Jenkins' study of latrine adoption in Benin rural: general critique

In a world where resources are limited and two billion people do not have excreta disposal services, programmes to promote latrine adoption should be efficient in evaluating what people want, and how they are conditioned by their environment and influenced by intervention programmes. Unless these results are found, the planner cannot realistically evaluate the effects of service improvements or the introduction of alternative technologies on demand.

Jenkins deduces that demand for latrines in rural Benin was a function of socio-economic household characteristics, demographic characteristics, lifestyle, cultural factors, village environmental factors, past latrine experience and exposure, beliefs and attitudes toward latrine, drives and constraints acting on latrine choice.

Her conceptual framework to explain sanitation choice was based on social psychology and consumer choice theories. Jenkins adopted three models of behaviour and consumer decision processes. According to her "*these were chosen because they complement each other in contributing to a general understanding of discrete choice behaviour and have been operationalised and tested in part if not in full*" (page 11).

Two models, Fishbein and Ajzen's theory (Fishbein, 1967, Fishbein and Ajzen, 1975; Ajzen and Fishbein, 1980) and the Engel, Kollat, and Blackwell model, or EKB model, (Engel et al., 1968, 1978), were used to describe the motivational forces affecting the individual's behaviour.

The second model is an expansion of the first one, and incorporates concepts from psychological cognitive theory, describing the relationship amongst five main stages of the decision process: 1) problem recognition, 2) search, 3) alternative evaluation, 4) choice, and 5) outcome (as a post-decision evaluation). Cultural aspects, reference group (and family), circumstances (anticipated and unanticipated), motives, lifestyle, normative compliance, evaluative criteria, beliefs, attitudes, intention, information and experience, search, and satisfaction are aspects that interfere in these stages of the decision process.

In addition, a third model, the Mokhtarian and Salomon (1994, 1996) model, developed to analyse the individual decision to telecommute, and based on a previous model conceptualised by McFadden (1986) for market research, was incorporated. In this model, each alternative is evaluated in terms of how effectively it satisfies the drive, and the individual's attitudes toward it (Mokhtarian and Salomon, 1996). Hence, alternative evaluation is oriented not for the characteristics of the good *per se*, but for the satisfaction to the drive that it carries. Thus, in evaluating individual decisions for telecommuting, which is a multidimensional alternative, the choice set may or may not contain telecommuting, and, according to the authors, probably contains other alternatives having nothing to do with telecommuting (page 1860). Jenkins used this last model to construct her framework of analysis because, in her opinion, latrine adoption is the result of choice among functionally non-comparable alternatives which differ for each individual, and this last model permits handling of this particularity. Implications for these assumptions are also discussed below in section 2.4.1.1.

From a qualitative investigation using an anthropological approach, Jenkins (1999) interviewed 40 households heads, their motives and the barriers they faced in adopting latrine use in rural Benin. Her model asserts that *"the key conditions for a latrine to be chosen in rural Benin are the presence of at least one active drive or dissatisfaction from among the 11 found to motivate*

adoption, and the absence of factors acting as constraints on adoption among 13 related to awareness, physical implementation, and psycho-social issues”(page 63).

Drives motivating latrine adoption were grouped into three categories: prestige-related drives, well-being drives, and situational drives. Associated beliefs and attitudes related to these drives were also summarised. Furthermore, 13 factors acting as constraints or facilitators, factors influencing the presence of drives for latrine adoption and past latrine experience and exposure were detailed.

Complementing her model, she described (without testing or demonstrating a causal relationship) the cultural dimension of latrine adoption, how lifestyle factors (formal education and wealth) emerge from well-being related perceptions, and how gender, occupation, education, social linkages, mobility and travel patterns, physical and social characteristics, and village environment act on motivation for latrine adoption.

A questionnaire was developed based on those 11 drives and 13 constraints, and included identified attitudes toward latrines (perceived from 40 interviews), and direct intention to adopt. This questionnaire was applied for a sample (not designed to be statistically significant) of 320 households.

Results showed large gaps between those who stated they preferred latrines to open defecation (97%), those who intend to adopt (25%) and those that have actually adopted (5%). Jenkins attributed these findings to the presence of constraints related to misunderstanding about how latrines function and their safety, lack of awareness about the advantages and benefits of installing a latrine, psycho-social factors, difficulties in saving money to build latrines, and the high cost. Along with demographic and socio-economic variables, she categorised attitudes, drives and constraints rating those variables through a semantic differential method, for adopters and non-adopters of latrines. For attitude variables, scales ranged from 1 point (most negative rating) to 7 points (most positive rating). Drives ranged from 1 to 4 (most positive). All frequencies were presented.

A factor analysis was applied to reduce the large set of interrelated variables into a smaller conceptual set of constructed factors to define drives and constraints underlying latrine adoption. Due to the exploratory character of her research, and inconsistencies between the census data and survey information regarding adopter status and occupations, unweighted data was preferred for analysis. Thus, together with unweighted socio-demographic variables, she defined unweighted drives and constraints to adopters and non-adopters to be analysed in her final logit regression.

Lastly, Jenkins' final model to forecast preference, intention and choice was presented. Independent variables were defined as "attitudinal variables", "drives", and "constraints".

Dependent variables were modelled through the sum of qualities (and non-qualities) stated by households, and the outcome defined according to this statement. Outcomes were defined as representing preference, intention, and choice or as a combination among them.

In spite of the interesting results of the psychometric assessment, Jenkins was not able to disentangle many interrelated variables and demonstrate the causal mechanism among them. The relationship between socio-economic, demographic, and village characteristics were not linked with latrine attributes perceived by the households, and so, the causal link was obscure and the individuals were treated as a "black box": the behaviour process that underlines their decision was impossible to understand. Without demonstrating the *causal* mechanism between the household socio-economic and demographic characteristics, village characteristics, and the attributes of a latrine, on the one hand, and choice-made on the other, Jenkins' results only replicated the conditions existing at the time of the survey (Domencich and McFadden, 1975).

From her model, some fundamental policy questions facing planners could not be answered. For instance: what behaviour underlies latrine demand, and how does it respond to changes in the attributes of latrine? How can one obtain, from the knowledge of behaviour, the implications of a sanitation intervention programme or changes in sanitation policy?

In fact, her results presented an effective inventory designed to gather information about beliefs (perceptions), attitudes (feelings), cultural experience and barriers related with household decision to install a latrine, rather than a causal model to forecast demand. In the following subsections, I will discuss the main restrictions of her model. Basically, the critiques were focused on: (1) the choice of an infinite choice set; (2) the use of pure factor scores; and (3) the non-causal relationship model to estimate preferences, intentions and choice.

2.4.1.1. Choice set in decision analysis

In defining the choice set to evaluate latrine adoption in rural Benin, Jenkins assumed that latrine choice “*is the result of choice among functionally non-comparable alternatives which differ for each individual (household)*” (page 11), she also stated that “*on first examination, latrine adoption looks like merely a choice among alternative defecation sites. However, one needs to separate the choice to use a latrine, repeated at each defecation occasion, from the choice to adopt or install one, a one-time, long term, high consequence consumption decision [...]. Latrine adoption typically results from choice involving alternatives having nothing to do with defecation (construction of a rain-water cistern or cement block house, purchase of plush salon furniture, painting living room walls with colourful paint, cementing the interior house floor, and so on). Furthermore, these alternatives vary between individuals*” (page, 69).

Correctly, she identified that, from the perspective of the behaviour decision, the use of latrine (or non-use) is the main determinant of the choice set. This is because the objective of the consumer is not to acquire the good or service for itself, but in acquiring this good or service he/she is trying to satisfy his/her drive, or as pointed out by Lancaster (1966a,b), the commodities or needs, not the goods themselves, are what the consumer really cares about.

Nevertheless, in assuming that the choice set was composed of alternatives having nothing to do with defecation (as, for instance, cementing the interior house floor), she assumed that the choice set was formed not of alternatives to satisfy the necessity for defecation (which could be a private latrine, open defecation, etc.), but for *all* possible consumer alternatives faced by the consumer – an infinite choice set. This assumption has important implications for latrine adoption analysis.

Before discussing the implications of a data set misspecification, it is necessary to discuss the origin of Jenkins' ideas in defining the choice set in rural Benin.

Jenkins' ideas were based on Mokhtarian and Salomon's model for telecommuting (1994, 1996). In this model, telecommuting was described as the "*use of telecommunications technology to work at home, or at a location close to home, during regular work hours, instead of commuting to a conventional work place at the conventional time*" (Mokhtarian and Salomon, 1994, page 749). As was stated by the authors, telecommuting is a multidimensional alternative, and choices for telecommuting (as for many goods/services) are usually based *not* on the service (good) itself, *but* as part of a decision process to satisfy a specific drive. Hence, due to the multidimensional character of telecommuting it is reasonable (and also expected) that a series of drives are linked to this service. Thus, alternatives presented in the choice set to satisfy a drive have nothing to do (directly) with telecommuting. Although, the choice set was defined as a *function* of the telecommuting service. In other words, the universal choice set defined by the authors was *directly or indirectly* related with telecommuting, as we can see in the Table 2.10.

A choice set should be defined in accordance with the relevant alternatives related to the good or service being analysed. The researcher must distinguish the individuals' inventory of relevant substitutes from all other goods (Haab and Hicks, 1997). Mokhtarian and Salomon in defining the "accommodate travel costs" as a alternative, for instance, as part of the choice set for satisfying drives that are "travels", they did not exclude telecommuting from the alternatives. To the contrary,

telecommuting is an indirect part of this choice set: an individual will telecommute from someplace to have time to organise his/her finance related with “travel” (Table 2.10). Note that telecommuting was an indirect alternative to a direct one (accommodate travel costs). The statement that alternatives become more and more non-comparable implies that reasons for telecommuting are largely variable and linked with indirect alternatives, and so, almost impossible to analyse in a model to forecast demand. A model with such a large number of alternatives would require a highly complex analysis with high costs (to investigate attributes for each alternative for a sufficient, and also, for a huge number of people), and possibly produce results difficult to interpret. Also, for the policy-maker’s point of view, an impractical tool for evaluation. Therefore, the choice of a dichotomous dependent variable (telecommuting and non-telecommuting) for analysing this market makes the model more useful.

In sanitation, in particular focusing on latrine adoption, the alternatives usually are easily identified. In Appendix A of her thesis, Jenkins described the types of latrines found during her interviews with households. Pit latrines with one or two cabins (sometimes three) were the most frequent latrines found. Variations in latrine hole style, sitting and access were also observed, but the basic definition of a latrine was quite invariant. In Jenkins’ words: *“small variations in cabin, hole, ventilation, size, and construction style exist, but the basic concept of a dry pit latrine is the same. Some institutional and public latrines are ventilated improved pit (VIP) or composing types. Public flush toilets are extremely rare even in urban centres”* (page 215).

Complements and substitutes also are clearly identified. In rural Benin, number of cabins, combined shower, and latrine hole style are examples of complementary goods for sanitation. The substitute is open defecation. In other contexts, complements can include flush toilet, different types of seats, different types of connections (septic tanks, local programme of sanitation, etc.), etc. In addition to open defecation, individuals can deposit their faeces in a bag and throw it away.

Table 2.3. The universal choice set for telecommuting (original title: the 'universal' choice set adaptive responses to life-style-related-drives)

Choice	Drives				
	Travel	Work	Family	Leisure	Environmental ideology
Accommodate travel costs	•				
Reduce travel costs	•			•	•
Socialise at work				•	
Change work trip departure time (unofficially)	•	•	•		•
Work unpaid overtime		•			
Take work home		•	•		
Change route	•				•
Buy time		•	•	•	
Adopt flex-time	•		•		•
Adopt compressed work week	•		•	•	•
Change mode	•			•	•
Invest in productivity-enhancing technology at home	2 nd order	•	•		•
Telecommute from home (part or full time)	•	•	•	•	•
Telecommute from a local work centre (part or full time)	•	•	•	•	•
Change to a new job in a new location	•	•	•	•	•
Change to a new job in the same location		•	•	•	•
Relocate home	•	•	•	•	•
Change from part-time to full-time work		•	•		
Change from full-time to part-time work	•		•	•	
Start a home-based business (or put more effort into a existing one)	•	•	•		
Quit work	•		•	•	
Divorce family		•			

• = potential responses

Source: Mokhtarian and Salomon, 1994, page 761

What are the implications of a choice set misspecification? The literature was focussed on the discussion of two main points: 1) misspecification can make the analysis obscure, and cause bias from the omission of the relevant substitute goods, and 2) in omitting the substitute goods, analysis cannot include the true individual preference, and so, cannot predict real demand for the product (Haab and Hicks, 1997; McFadden, 1980).

To identify the choice set correctly is fundamental to the analysis. First of all, because the individual can know exactly what he/she is evaluating. Second, preferences can be stated and alternative attributes can be matched with individual's characteristics, establishing a behavioural link.

It was understood that when Jenkins assumed a dichotomous dependent variable "adopter" and "non-adopter", it was related to latrine use (adopter) and open defecation (non-adopter), in view of the fact that in rural Benin the majority of the population use the open-air for defecation. Also, the household head was asked if he/she understood what a latrine was. However, this approach limits the analysis and may not capture the individual's likely choice because not all possible substitutes were identified. As pointed out by DeShazo et al. (2004), *"a possible discrepancy between the consideration set relevant to the individual and that assumed by the researcher may arise when the researcher misinterprets individual's choice of the "no-purchase" option (in Jenkins's case, "non-adopter – my note) [...]. When the individual selects the no-purchase option, the researcher typically assumes that the individual is opting out of the entire category of goods that are the focus of this choice occasion, allocating their entire income instead to the set of numeraire goods. However, especially in the stated preference case, the individual may know of actual goods that are close substitutes to those in the researcher-defined choice set that they intend to purchase in the near future. When this is the case, the research fails to acknowledge the individual's full relevant choice set by failing to include these omitted close substitute goods"* (page 4). It is especially true in other contexts where sanitation choice or latrine choice is more complex.

Jenkins correctly identified that *"one needs to separate the choice to use a latrine, repeated at each defecation occasion, from the choice to adopt or install one, a one-time, long term, high consequence consumption decision"* (page 69). In other words, the use (or non-use) of latrine is the starting point for investigation of latrine adoption. From this, it is possible to identify all possible

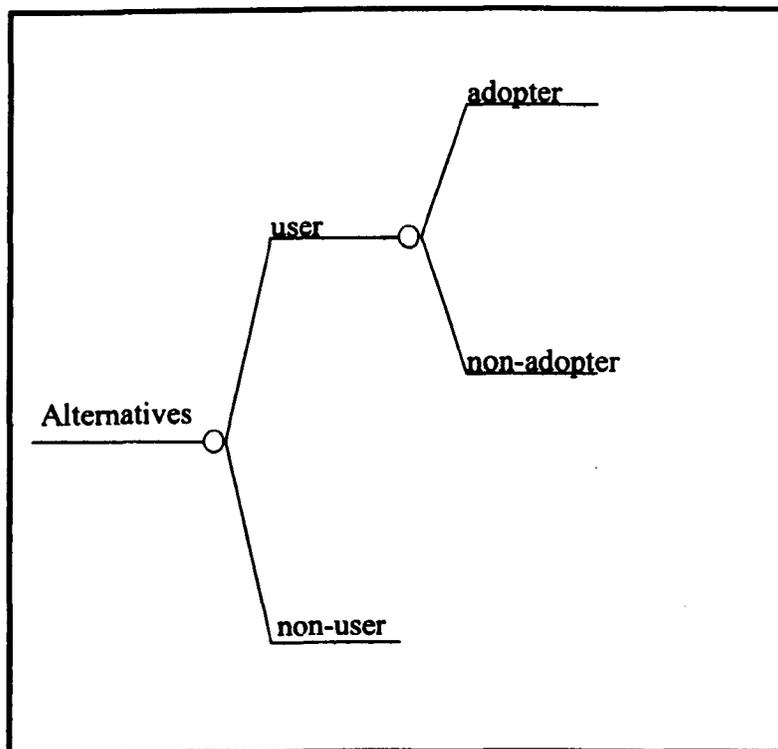
alternatives facing individuals and relate attributes of these alternatives with individuals' characteristics and choice. Analysis would be something like that one demonstrated in the Figure 2.1, where alternatives identified in rural Benin could be demonstrated and evaluated.

For instance, Jenkins estimated that latrines are used by about 15% of the rural population and owned by only about 5% of households. This information has some implications for analysis: how frequently do households share a latrine? In what proportion do they share latrines and use open-defecation?

From the point of view of public health, if an individual is isolating his/her faeces, it is considered a good thing, and owning a latrine is not important. Furthermore, if an individual shares a latrine and uses open defecation from time to time, the probability of this individual adopting a latrine in the future is higher than that for one who uses only open defecation. Therefore, identifying all of the alternatives for defecation is really important to avoid misinterpretation and bias in the results.

In analysing non-market demand, it is important to identify the competing alternatives which satisfy the drives, and not all alternatives that in some sense compete among them. In general, everything competes with everything in making a choice, because people have a budget constraint which limits their choices. In analysing demand for travel, for instance, food, clothes, wall painting, etc. should not be included in the choice set for analysis, competing alternatives that *in fact* are important to analyse travel demand because they are part of the analysis, e.g., car ownership, bus, walking. Also the service levels of competing alternatives should be analysed. If access to materials to build a latrine is difficult (or very difficult), it is possible that this cost is more important than the material price *per se*. Access to building materials may make the cost of construction extremely high (due to transport etc), or even prohibitive; therefore, it is an important factor to be evaluated.

Figure 2.1. Use of latrine in rural Benin: the choice set



In other words, the model must attempt to describe the causal relationship between socioeconomic and latrine (sanitation) characteristics, on the one hand, and latrine choice-making on the other. It is necessary for the model to explain why latrine decisions vary as conditions change. Unless this is done it is not possible to anticipate how the household will behave if its circumstances change or the terms upon which the competing alternatives are offered change. In short, only by explaining the causal relationships can the model be used to forecast the effects of future changes in the performance of the sanitation system. Otherwise, the model will simply replicate the effects of the sanitation system that existed when the model was originally calibrated (Domencich and McFadden, 1975).

2.4.1.2. Factor analysis, correlation, ordinal variables, and causal relationship

This section discusses briefly the limitations of factor analysis as a technique used to reduce a large sample of correlated variables into a smaller uncorrelated one. In addition, the implications of using this method with ordinal variables, and the question of reliability of attitude measures as predictors of behaviour, particularly when the model assumes a causal relationship between attributes and behaviour. These limitations and implications are applied to Jenkins' study. Despite critiques of the use of factor analysis for uncorrelated variables and for reducing the dataset, a fundamental limitation of Jenkins's study was the absence of any demonstration of statistical associations among those complex variables. Nevertheless, this association was only described and assumed to be true. As the mechanism of action between them was obscure, the model was not predictive, hence, not policy oriented, because policy makers can not deal with those variables.

Use of factor analysis applied to different studies has become popular (with special attention to the social sociology and psychology fields) because the availability of electronic resources which facilitate the use of this technique. Parallel to the increase in popularity, the misuse of this method has increasingly been criticised (Eysenck, 1952; Glass and Maguire, 1966).

Factor analysis is a variant of the method of principal component analysis and it attempts to reduce a number of correlated variables to a small number of hypothetical causes or factors called "factor scores" (Eysenck, 1952). These factor scores are obtained "*by weighting scores on variables by the loadings of the variables on the rotated factor, or that factors are obtained by summing the scores on the variables which have "high" (in absolute value) loadings on the factor*" (Glass and Maguire, 1966, page 297). Here is the main point of criticism: both procedures are incorrect and produce "*a set of 'factor scores' having few of the properties of the legitimate factor scores desired, and [...] place the unwary researcher in the position of treating the observed correlated scores as though they were uncorrelated*" (Glass and Maguire, 1966, page 298).

The idea of selection of factor scores has its roots in high values of a matrix (varimax procedure) and was originally developed to be applied to principal component analysis (PCA) by, and his procedure presented well described properties associated to PCA, and has been reasonably accepted by many experts in this method (Jackson, 1991). However, this approach was then imported from principal component analysis to factor analysis without any theoretical rigour, creating factor scores that, in fact, may have high intercorrelations and correlate only moderately with the correct computed factor score (Glass and Maguire, 1966).

In principal component analysis, the factor scores are combined with the identity matrix (it is uncorrelated and has variances of unity) and the matrix of intercorrelations of the variables, and the principal component solution is derived from this combination. The problem occurs when some researchers assume the *pure* factor score, either by summing the standard score on the individual variables, each score being weighted by the varimax factor loading of the variable on the factor, or a variable is said to load only on that factor for which its loading is largest (in absolute value – as in Jenkins' study). These procedures lack good statistical justification, and may obtain varimax factors scores with high intercorrelations (Cooley and Lohnes, 1962; Glass and Maguire, 1966).

Other limitation of Jenkins' study is the use of factor analysis as a method to select psychometric factors represented by ordinal variables.

Use of ordinal measurement to formulate theoretical ideas has been controversial. As ordinal variables have no unit of measurement or a standard which defines them, classical measurement theory disqualifies their use (Kampen and Swyngedouw, 2000). Due to their subjective character (as much in their definition as in their assessment), some authors believe that models based on ordinal variables generate weak conclusions since causal structure itself is uncertain with a poor fit between data and the theoretical model (Wilson, 1971). However, some progress has been made, and the use of some models has been suggested such as hybrid conjoint models, latent variable models, latent

variable combination with multinomial logit models (McFadden, 1986). These models, through rigorous statistical techniques, assume that perceptions and attitudes can be scaled metrically and define the vectors of indicators (McFadden, 1986). Hence, important psychometric data on perceptions and tastes can be included in the analysis, establishing a causal relationship between attributes of the good or service and individual's characteristics.

Conclusions from Jenkins' study are weak due to the poverty of the formal technique by which ordinal variables could be manipulated. Elementary errors in the use of factor analysis to evaluate psychometric data (as briefly discussed above) imply that the results achieved by Jenkins may differ from those achieved by another analysis of the same table of intercorrelations since the ordinal variables assume a certain subjectivity of choice and factor analysis is a complex and difficult method which should be used with great caution, not just assuming the high loading scores (Eysenck, 1952).

Nonetheless, the absence of any attempt to demonstrate the causal relationship among the main variables of her model, even estimating a modest power of her method putting the conclusions in perspective, these remain a weakness in Jenkins's model. Psychometric data are without doubt an important part of the explanation of consumer behaviour, and essential for forecasting demand. However, understanding how these variables can be used objectively for policy design is essential in constructing statistical models for prediction.

2.4.1.3. Final model to forecast demand for latrine in rural Benin: problems with definition of preferences, intentions, and choices

There is a sharp distinction between the meanings of preferences when used by social psychologists and economists. In social psychology and psychology, preferences are defined as a context-dependent process influenced by attitudes which are expressed by evaluating a particular entity with some degree of favour or disfavour (Kahneman et al., 1999). To economists, preferences are only comparative judgements between entities. These also imply a cognitive process that is simplified to define *preference maximisation* which has well defined properties. The main one is the assumption that the consumer is rational and a maximiser. In addition, preferences can be represented on a numerical scale defined as utility (McFadden, 1998). Thus, when someone refers to utility maximisation it is assumed that he/she is referring to economic preferences.

The objective of this section is not to discuss the different points of view between different fields of research. The purpose here is to argue the necessity of clear definitions to the conduct of an analysis. If terms are not well defined, the consumer may not know what he/she is evaluating and the researcher may obtain inconsistent results.

Different approaches to defining preferences were assumed in Jenkins' thesis. At the same time she assumed preferences are a context-based process based on individual feelings toward alternatives (attitudes), but also that these preferences can be ranked on a utility (maximisation) scale.

In chapter 7, Jenkins presented her mathematical model of preference, intention, and choice to adopt a latrine by household heads in rural Benin. Dependent variables were preferences (ordinal ratio – sum of qualities ratings), intention and choice (ordinal and dichotomous). Independent variables were drives, attitudes, and constraints which were defined before.

Now, if a logit model was performed assuming utility maximisation, that means that whether “A is preferred to B”, utility of A is higher than utility of B, including conditions of completeness ($a \geq b$ or $b \geq a$, $\forall a, b \in$ choice set) and transitivity ($a \geq b$ and $b \geq c \Rightarrow a \geq c$, $\forall a, b, c \in$ choice set). Then, the consumer will maximise his/her utility (which are ordered: X_1 is preferred to X_2 which is preferred to X_3 and so on) subject to a budget constraint, i.e.:

$$\text{Maximize } U = U(X_1, X_2, \dots, X_n) \quad (2.1)$$

$$\text{Subject to } \sum_{i=1}^n P_i X_i = Y \quad (2.2)$$

Where X_i is the quantity of the i th commodity, P_i is the price of the i th commodity and Y is total income.

Whether economic preferences are assumed or not, preferences are on a domain of attitudes which are themselves a unitary valuation and these are not measured in a preference order (Kahneman et al., 1999). This last case looks like being Jenkins’ case. If so, why did she expect variability in preferences? In pages 228 and 229, she recognised that “*the hypothetical frequency of choosing to use a household latrine to defecate failed to adequately capture variability in preference: 96.9% of household heads (adjusted N=310 out of 320) said they would ‘always or almost always’ choose to defecate in a latrine at their house rather than in open. Among the remaining ten, nine chose ‘often’ and one (a male farmer) chose ‘very little’ [...]. In retrospect, this question suffered from several weaknesses: first, using and adopting latrines are distinctly separate choices, making it difficult to infer preference for one from preference for the other [...]. Only slightly more variability was measured about the hypothetical frequency of choosing a household latrine for children to defecate (90.3% or adjusted N=291 out of 317 indicated a frequency of ‘always or almost always’. No difference in response by class was apparent*”

One hypothesis can arise from these findings: what in fact household heads were measuring was an attitude towards a latrine through an affective value to a dichotomous choice between

“latrine” and “open defecation”. Hence, the consumer was evaluating an entity (a latrine) with some positive or negative sign, and so, expressing his/her attitude towards a latrine (or to open defecation). Variability of preferences will be observed whether stated preferences had been questioned through presentation of a well defined choice set, or related to it. This choice set could be defined from the revealed preferences for the *use* of local to defecation, as well was observed by Jenkins. The poorly defined choice set compromised the observation of variability among household head preferences, and we could conclude that Jenkins only observed attitudes toward a latrine, not preferences between alternatives. The same conclusion could be made about her model of “choice”. Whether a pure binary choice “adopter” and “non-adopter” was assumed without differentiation among alternatives, and also without the causal relationship among variables, we are only observing the choice that was made when the survey was applied. That is not very useful for policy makers.

It is also difficult to define intention. However, I agree in some sense she could identify intention for latrine adoption in her model. Whether we can assume definition of intention in a broad sense to include beliefs, which are probability of judgments, and attitudes, which are stable psychological tendencies to evaluate particular entities with favour or disfavour, we also can assume that she could evaluate intention towards a latrine (McFadden, 1998). Nevertheless, intention is only a part of the cognitive process, that together perceptions, beliefs, attitudes, preferences, and motives will perform the task of producing choice in a behavioural causal model to latrine adoption. Unfortunately, it was not demonstrated in Jenkins’ thesis.

2.5. Summary of discussion

From the discussion conducted in this chapter, we have an overview of the literature applied to sanitation and what are, in my opinion, the gaps and failures that they presented in explaining the demand for sanitation. The contingent valuation method fails to demonstrate the preferences that it assumes to demonstrate, and presents several biases to be a method useful for policy.

Attitudinal-behavioural approach, on the other hand, provides a tool to evaluate the perception and attitude towards sanitation formation, and understanding of household behaviour, historical and cultural aspects influencing household choice. However, the relationship among these variables that reflect household's perception, attitude, beliefs, historical, and cultural aspects, and that subjectively and objectively influence behaviour, should be demonstrated, for policy purposes. In general, for a model to be able to forecast demand and be useful for policy, it is necessary (at least) to:

1. Test and demonstrate the causal interaction between individual characteristics, alternative attributes, and choice made;
2. Incorporate variables which can be manipulated by policy makers;
3. Be generalised from city-to-city, not just replicate the results of existing conditions of a determined site;

From Jenkins's model, I could identify some limitations that restricted the association of behavioural/attitudinal variables and observable variables, evaluated in a discrete choice model to demonstrate a causal relationship among them. The filling of this gap, between behavioural theory and discrete choice model applied to sanitation, is the objective of this thesis. In so doing, to establish whether this model can be developed, so that it can be used for estimation of demand for sanitation.

Chapter 3: Economic Choice Behaviour Literature and Contributions to the Development of a Framework of Analysis for Sanitation Demand Studies

3.1. Introduction

Several critiques have been undertaken on classical economic consumer theory. Despite critics, an economic approach to assessing behavioural decision processes has been developed, especially with the introduction of discrete choice models based on Random Utility Maximisation hypothesis (RUM). These models have demonstrated enough flexibility to encompass psychometric variables that are important in explaining the cognitive process of choice.

This chapter presents a discussion of traditional economic consumer theory and its contribution for the development of the indirect utility maximisation model (Lancaster model) and discrete choice. The discussion carried out in this chapter will provide the basis of the understanding of the new Hybrid Choice Model, which will be assumed as a framework of analysis to explain demand for sanitation, and will be presented in the next chapter. The Hybrid model permits the introduction of latent variables (non-observed variables) into the analysis, to give a more robust answer for planning proposes. There is strong evidence that the Hybrid approach could outperform conventional methods of evaluating demand, specifically in a market highly influenced by perceptions, beliefs, historical and cultural aspects, such as sanitation.

3.2. Classical economic theory of consumer behaviour: contributions and limitations in understanding the consumer's decision process of choice

A long debate on determinants of consumer's decision process of choice, involving mainly economists and psychologists, has been carried out, at least for the last five decades. Whereas psychologists focus on ideas that behaviour is context dependent, complex and mutable, exploring how decision elements are established and modified by experience, focus in traditional economics is on the mapping of objective (and observable) information inputs to choice. The decision process is usually viewed as a "black box" where the consumer behaves as if the information was processed to form perceptions and beliefs following normative axioms. These perceptions and beliefs are generally defined as preferences, or values, and are defined as primitive, consistent, and immutable. Thus, the cognitive process is simply preference maximisation that can be represented by a utility scale, faced with a budget constraint. This whole process, assumed as "consumer rational behaviour", has a direct relationship with the concept of maximisation (Tversky and Kahneman, 1981; McFadden, 1998).

The classical theory of consumer demand, where the "rational economic man" assumes the centre of analysis, had its origin in the work of Hicks (1939) and Samuelson (1947). Hicks (1939) developed the concept of complementary and substitutability among entities, and some years later, in his "theory of value", he explained the pattern of relative prices in an economy and the resulting allocation of resources. Samuelson (1947) provided an integration of concepts of economic theory, demonstrating in a mathematical language the relationship between economic agents. Based on Bergson (1938), who developed the social welfare function¹¹, Samuelson defined the necessary and sufficient conditions for the allocation of resources where the movement from one allocation to

¹¹ Social welfare function linked the produced and consumed commodities and the resources used to produce them (Bergson, 1938)

another will make the individual better off without disadvantages for other individual (Pareto optimisation), given a set of alternative allocations and individuals. In his model, alternatives could be ordered (transitive ranking – better, worse, indifferent) in utility set (utility function).

Nevertheless, a more dynamic and fashionable descriptive axiom for behavioural decision theory was developed by von Neumann and Morgenstern (1944). This gave more impulse and power to economic consumer theory. It was fashionable because elements from psychology were introduced in the analysis, aggregated with economic variables. The authors developed a formal and axiomatic structure to demonstrate how beliefs could be represented (in terms of agents' preferences) in a decision choice process under uncertainty and game theory. Here, the rational maximiser man is rational in the sense that he/she makes choices which maximise his/her perceived utility (preferences). Note that elements from psychology are constructed within a belief axiomatic structure, represented by preferences which are previously determined, as whether they are genetically coded "taste template". Originally, the experiments were conceived to evaluate choice among lotteries, but these ideas were quickly incorporated in economic decision making analysis (McFadden, 2001)¹².

Later, ideas from von Neumann and Morgenstern were aggregated to a Bayesian belief structure¹³ by Savage (1954), who also developed the paradigm of decision under uncertainty. Independently, Luce et al., (1971) and Krantz et al., (1971) brought out a similar model, however, without the restrictions and unrealistic combinations of events proposed by Savage. It was the "independence condition" where there is no strict separation of act and states, and events can be

¹² It is important to point out that, in this model, the elements from psychology are modelled in a *preference* structure which, presumably, underly consumer perceptions (beliefs) and attitudes.

¹³ If one believe an uncertain event has possible outcomes (say, x_i), each with utility $u(x_i)$ and probability $P(x_i)$, then the subjective expected utility will be $\sum u(x_i) P(x_i)$. If one compare this outcome with another one, say y_j , with probability $P(y_j)$, the subjective expected utility of this second one will be $\sum u(y_j) P(y_j)$. Then, decision process will depend on the order of different beliefs (different utility functions) – putting values on those outcomes: better, worse, indifferent (Savage, 1954).

repeated independently. Luce's model permitted a new dynamic for economic experiments and still influencing several researchers.

These collections of ideas form what is, currently, the basis of economic consumer behaviour theory found in economic textbooks.

According to economic consumer theory (see Varian, 1992; and Mas-Colell et al., 1995), each consumer is able to compare alternatives in a choice set (C). Using a preference-indifference operator (\geq , prefer/indifferent to). Thus, if $a \geq b$, the consumer prefers a to b , or is indifferent ($a > b \leftrightarrow a \geq b$ and not $b \geq a$, strict preference; or, $a \sim b \leftrightarrow a \geq b$ and $b \geq a$, indifference). The preference-indifference operator is supposed to follow some rationality assumptions of¹⁴:

- a) Completeness: $a \geq b$ or $b \geq a$, $\forall a, b \in C$ (people have preferences over all goods, and can rank all of them, i.e., any two bundles can be compared);
- b) Transitivity: $a \geq b$ and $b \geq c \leftrightarrow a \geq c$, $\forall a, b, c \in C$ (if a is preferred (or indifferent) to b , and b is preferred (or indifferent) to c , then a is preferred (or indifferent) to c);

In addition, if the preference relation is "well-behaved"¹⁵, it also assumes these two properties:

- c) Monotonicity: if $b \gg a$ then $b > a$, $\forall a, b \in C$ ¹⁶ (it refers to a strict preference over outcomes, without considering the case where the decision maker is indifferent between outcomes: a gamble (b) which assigns a higher probability to a preferred outcome will be preferred to one which assigns a lower probability to a preferred outcome);
- d) Convexity: for all a in C , the "at least as good as set", $\geq a$, is convex, that is, if $b \geq a$ and $c \geq a$ then $\alpha b + (1 - \alpha)c \geq a$ for any α in $(0,1)$, $\forall a, b, c \in C$ ¹⁷ (a mixture of bundles are

¹⁴ A preference relation that satisfies completeness and transitivity is often referred to as a rational preference or a preference ordering (Mas-Colell et al., 1995).

¹⁵ See Mas-Colell et al. (1995).

¹⁶ Also should be included the assumptions of strong monotonicity and local non-satiation (more of either product is better), but, for simplification, I decide just included the "monotonicity" assumption (for more details see Mas-Colell, 1995).

¹⁷ Let a be in choice set C . Then, I can have the following options (according with Mas-Colell, 1995):

1. $\geq a \equiv \{b \in C \mid b \geq a\}$ "at least as good as set"

preferred to the bundle itself, e.g., mixture of b and c (or b or c) is at least as preferred as a).

The utility function (U), that summarizes a preference ordering by consumers, is said to have the expected utility property whether, for a gamble a with outcome $\{a_1, a_2, \dots, a_n\}$ with probabilities $\{p_1, p_2, \dots, p_n\}$, respectively:

$$U(a) = p_1u(a_1) + p_2u(a_2) + \dots + p_nu(a_n) \quad (3.1)$$

A utility function is a way of assigning a real number to every possible consumption bundle such that the more preferred bundles get assigned larger numbers than the less preferred bundles. That is, a bundle (a_1, a_2) is preferred to a bundle (b_1, b_2) if, *and only if*, the utility of (a_1, a_2) is larger than the utility of (b_1, b_2) .

The choices of the consumer depend on the consumer's income and the prices of the goods. The problem faced by the consumer is how to maximize utility subject to a budget constraint, i.e., the bundle that represents the optimal choices (consumers choose the most preferred bundle from their budget sets). Then, the consumer selects that combination of goods affording him/her highest utility. The problem facing to consumer is to

$$\text{Maximize } U = U(X_1, X_2, \dots, X_n) \quad (3.2)$$

$$\text{Subject to } \sum_{i=1}^n P_i X_i = Y \quad (3.3)$$

Where X_i is the quantity of the i th commodity, P_i is the price of the i th commodity and Y is total income.

The necessary conditions for an optimum are:

2. $\leq a \equiv \{b \in C \mid a \geq b\}$ "no better than set"
3. $> a \equiv \{b \in C \mid b > a\}$ "better than set"
4. $< a \equiv \{b \in C \mid a < b\}$ "worse than set"
5. $- a \equiv \{b \in C \mid a \sim b\}$ "indifference set"

- 1) That the consumer equates the marginal rate of substitution between any two commodities with the ratio of their prices;
- 2) That the budget constraint be satisfied. The optimal bundle will be characterised by the condition that the slope of the indifference curve (the marginal rate of substitution) will equal the slope of the budget line.

Therefore, the demand function for n commodities may be derived from the first order condition to express the optimal quantity of commodity as a function of all prices and the income - the functional relationship being dependent on the consumer's preferences. If the demand function is substituted back into the utility equation, we can derive the indirect utility function which is the basis of discrete choice model, which will be discussed later.

Despite the sophistication and popularity of economic consumer theory, experimental evidence has been accumulated demonstrating that the model presents failures, mainly in non-market analysis, in evaluating the consumer decision process. Showing that the consumer does not behave like a rational-maximiser most of the time, and that there are other important latent factors influencing behaviour - which cannot be captured by the preference structure -, these experiments pointed to the necessity to explore the unobserved factors within and across individuals, and not treat them as "anomalies", aggregated into a disturbance term attributed to data measurement error. Economists and psychologists have demonstrated through their experiments that these "anomalies" are, in fact, part of decision process (Allais, 1953; Kahneman and Tversky, 1979; Tversky and Kahneman, 1981).

Tversky and Kahneman (1981), for instance, have demonstrated failures in the normative process rules with focus on preference-utility properties of transitivity, cancellation, variance and dominance. Also, they questioned the validity of assumption of the consumer rationality-maximisation in the decision making process, particularly observing the anchoring cognitive

phenomenon - previously discussed in Chapter 2 -, (Kahneman and Tversky, 1979; Kahneman, 2003). They also stated that *“the logic of choice does not provide an adequate foundation for a descriptive theory of decision making. We argue that the deviations of actual behaviour from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the normative system”* (Tversky and Kahneman, 1981, page S252).

Recognising the limitations of the economic traditional model to predict consumer behaviour in the decision choice process, many researchers have developed sophisticated approaches and experiments trying to aggregate inputs from cognitive psychology and social-psychology in the decision choice process analysis. The main objective is to understand how consumer behaviour is affected not only by objective variables (socio-economic, demographic, alternatives attributes), but also by non-observed variables (latent variables: attitudes, perceptions, tastes, etc.) within a dynamic process. Research contributions come from different disciplines, such as transportation, economics, psychology, public policy, geography, etc., and some key articles are: Ben-Akiva et al.(1999), Johansson (2005), McFadden (1986, 1998, 2001). These new approaches will be discussed in the following paragraphs.

What I would like to discuss now is why, in spite of limitations and restrictions for analysis, the structure of preference-rationality-maximisation of the traditional model of economic consumer behaviour persist dominating policy evaluations, especially in economics and marketing research. A general answer could be that, as a planner's objective is to predict consumer decisions that can be used to (then) develop strategies and plan interventions, a model where the causal relationship among objective variables (and non-observed variables, when it is possible) is explicit, is useful and necessary.

There is no doubt of the necessity to explore how latent variables act on consumer decision process and on the necessity to rethink the traditional model. And there is also no doubt that many economists (and researchers from other sciences) are truly involved in this process of restructuring in some sense this standard model. Nonetheless, the structure of preference-utility-maximisation remains the core of analysis.

Psychologists such as Kahneman (2003) have recognised that efforts have been made for the past decades to incorporate psychometric data into decision process analysis. Nevertheless, he pointed out that *"[...] conventional economic analysis is now being done with assumptions that are often much more psychologically plausible than was true in the past. However, the analytical methodology of economics is stable, and it will inevitably constrain the rapprochement between the disciplines. Whether or not psychologists find them odd and overly simple, the standard assumptions about the economic agent are in economic theory for a reason: they allow for treatable analysis. The constraint of tractability can be satisfied with somewhat more complex models, but the number of parameters that can be added is small. One consequence is that the models of behavioural economics cannot stray too far from the original set of assumptions. Another consequence is that theoretical innovations in behavioural economics may be destined to be noncumulative: when a new model is developed to account for an anomaly of the basic theory, the parameters that were modified in earlier models will often be restored to their original settings. Thus, it now appears likely that the gap between the views in the two disciplines has been permanently narrowed, but there are no immediate prospects of economics and psychology sharing a common theory of human behaviour"* (pages 165 and 166).

One of reasons for the difficulty in the construction of a common theory of human behaviour could be the different approach to analysis in economics and psychology. Psychologists work on a process of deconstruction, revealing the irregularities and idiosyncrasies features of behaviour

In this model, heavy arrows trace the direct relationship among variables, and coincide with the economic standard model. Dashed arrows correspond to psychological (and latent) factors in the decision-making process.

History matters. Past historical experience has an influence on actual consumer decision. Experience related to the good or service can influence the consumer in making a choice. Some authors include in the variable "history" socio-economic and demographic characteristics of individuals to explain experience. History operates on memory which influences perceptions/beliefs and preferences, and motivation/affect. Perceptions are cognitive of sensation, while beliefs are mental models of the world (probability judgments). Nonetheless, they are broadly used together due to their similarity and they are difficult to observe them separately. Perceptions and beliefs are formed by experience and information (market information, attributes of goods) and form stated perceptions. Motivation and affect refers to the emotional state of the decision-maker. They have direct influence on attitudes that affect valuation of a particular entity.

Attitude, the most important concept in psychology (borrowed from socio-psychology), is defined as a stable psychological tendency to evaluate a particular good or service with some degree of favour or disfavour, through an affective value (from extremely positive to extremely negative). Attitudes are multidimensional, with no requirement of consistency across them, and can be interpreted from attitude scales which are manageable in a statistical analysis (Ajzen and Fishbein, 1980; Kahneman et al., 1999, 2003; McFadden, 1986, 1998, 2001).

Preferences have different meaning for economists and psychologists. Whereas psychologists' preferences are functions of consumer experience, perceptions, and personal characteristics, and mainly, are adaptive and imitative, for economists, preferences are stable and predetermined, and can be represented by utility scales in a maximisation structure. And this is one of the main points of divergence between psychologists and economists.

Several criticisms can be made of the definition, formation and analysis of preferences in the decision process. However, the structure of preferences is still a useful tool in economic analysis. Preferences are a vital concept in economics. Observing the choices made by consumers (revealed preferences), or the choice that he/she considers to make (stated preferences) using a utility scale within a maximisation structure, seems useful, tractable, and a practical method to evaluate intervention. Also, results from studies based on economic preferences strongly suggest consistency in predicting demand (Morikawa et al., 1996; Persson, 2001, 2002). These could be an additional reason why the analytical methodology of economics (utility maximisation) remains stable.

In economics, whether a conditional distribution can be constructed, and whether this distribution depends on individual's history of economic status and choices, his/her experience could be expressed through preferences. Stated preferences can be determined through hypothetical choice situations or asking the individual for their preferences in experiments. Revealed preferences are choices that the individual made, the observed choice. Both have considerable psychological and economic appeal, in the sense that they can incorporate notions of perceptions and attitudes concomitantly with directly observed attributes of the choice (stated or revealed). Also, they have appropriated properties which can be operationalised into models that are flexible enough to take into account the complexity of the decision process (Ben-Akiva and Morikawa, 1990).

In addition, the advent of new models, which permit elucidation and measurement of empirical experiments, including attitude scales have made it possible to work with preferences in a more realistic and efficient way to evaluate economic policies. These models, discrete choice analysis based on the Random Utility Model (RUM) hypothesis, have the potential to incorporate (and explore) the role of perceptions and attitudes, in a psychological sense, in connection with economic decision structure, including economic consumer preferences (McFadden, 2001). Or, as stated by McFadden (2001, page 374), "[...] I believe that the basic RUM theory of decision-making, with a

much larger role for experience and information in the formation of perceptions and expression of preferences, and allowance for the use of rules as agents for preferences, can describe most economic choice behaviour in markets, surveys, and the laboratory. If so, then this framework can continue for the foreseeable future to form a basis for microeconomic analysis of consumer behaviour and the consequences of economic policy”.

Therefore, if preferences can be explored together with cognitive psychological elements to define the decision process, these are powerful tools in determining the causal mechanism of the choice process. It appears to be the case of these new approaches based on RUM hypothesis where preferences (stated and revealed) are elicited and evaluated. In a complex world of scarce resources, society must make choices, and planning is essential. Preferences seem to be a useful structure to make such decision, specially into the framework of discrete choice RUM model, which will be discussed in the following paragraphs.

3.3. The Lancaster model and RUM hypothesis: developments in economic process analysis

The process of questioning and re-thinking the theory of consumer behaviour in a more realist way is an effort that many economists (and other researchers) have been involved in for many years. The main influence on these authors comes from psychological theory (Simon, 1959).

According to the psychological theory of consumer behaviour, a consumer purchases a commodity because he or she expects to satisfy one or more basic “needs” or “drives”. Some consumer needs are *biogenic* in origin and some are *psychogenic*. Biogenic needs arise when tension is created in the psychological system by the body entering disequilibrium state – for example, hunger, thirst, sleep and pain. Psychogenic needs are precipitated by tension that develops as a result of an individual’s association into the need for affiliation, the need for achievement and the need for power. A product or service that satisfied a biogenic or psychogenic need becomes a reinforcing agent or what is often called a goal object. Whenever reinforcement takes place, the likelihood is increased of the person seeking the same goal object whenever he or she experiences the same drive (Zaltman, 1965; Currie et al., 1972).

Lancaster (1966 a, b) assumed similar ideas from psychological theory to develop his approach to consumer demand. For him, goods and services are expected to satisfy needs or drives. Consumers do not choose goods, but the qualities or characteristics inherent in goods. Whereas in traditional economic consumer theory good qualities or characteristics are treated in an implicit way, and where explicitly what consumers choose are goods (not characteristics), in the Lancaster model this relationship is inverted.

In his model, the market for goods and services are merely inputs into the consumption process. Together with time and human capital reflecting the skills and experiences of the

individual, these market inputs are transformed into basic commodities in line with a personal production function. The commodities or needs, not the goods themselves, are what the consumer really cares about. For instance, consumers do not demand food itself, but rather the nutritional ingredients and flavours in the food. Consequently, the utility function for a household is maximised and related to these commodities. Utility concept in his model has also a different meaning from that in the traditional approach to demand: utility is derived from the proprieties and characteristics of the goods.

Considering that consumption is a complex production process and that utility is related to basic commodities, the satisfaction of a specific need does not depend on a single market good. On the contrary, there may be a great many combinations of market inputs, time and skills, which give rise to the same commodities, and same characteristics, may be shared for different goods.

In view of the complexity of the decision tasks, utility maximisation clearly transcends human cognitive abilities. It becomes obvious that consumers' choices are probably inefficient most of the time: *"in consumption, as in production, the prime reasons for inefficient use of the existing technology are ignorance and lack of managerial skill. The consumer may not be aware that a certain good possesses certain characteristics or that certain goods may be used in a particular combination to give a specified bundle of characteristics. Unlike commercial production, the market mechanism does not eliminate inefficient consumers; as a result, inefficiencies in consumption will persist to a certain degree, even in a highly competitive market system"* (Lancaster, 1966a, p.18 and 19).

In Lancaster approach (1966b, pages 135 and 136), individual goods or a collection of goods are associated to a scalar (level of activity), and the relationship between the level of activity and the goods consumed are assumed to be linear and objective¹⁸.

¹⁸ Assuming that the relationships are objective mean that the equations are invariant, i.e., are assumed to hold for all individuals.

If k and y_k are the level of activity, and x_j is the j th commodity, we have:

$$x_j = \sum_k a_{jk} y_k \quad (3.4)$$

Where a_{jk} is determined by the intrinsic properties of the goods themselves and, possibly, the context of technological knowledge in the society.

The vector of total goods required for a given activity vector is given by:

$$x = Ay \quad (3.5)$$

Where A is the consumption technology, assumed be fixed.

Also, Lancaster assumed that each consumption activity produces a fixed vector of characteristics and that the relationship is again linear, and coefficients b_{ik} are objectively determined for some arbitrary choice of the units of z_i . Then z_i is the amount of the i th characteristic:

$$z_i = \sum_k b_{ik} y_k \quad (3.6)$$

or,

$$z = By \quad (3.7)$$

Where B is also the consumption technology.

For simplification, the linearity form is assumed to be:

$$F_k(z, x) = 0, \quad k = 1, \dots, m \quad (3.8)$$

The relationship between collections of characteristics available to the consumer (vector z) - which are the direct ingredients of his/her preferences and welfare -, and the collections of goods available to him/her (vector x) - representing his/her relationship with the rest of the economy -, is not direct, but indirect, through the activity vector y . A one-to-one (or direct) relationship will just be possible if A and B matrices are squared, and vectors of good characteristics, activities, and goods themselves were equals.

When one considers the relationship which link z and x from (3.5) and (3.7), we will have $z=Bx$.

It was also assumed that the individual possesses an ordinal utility function on characteristics $U(z)$ and that he/she will choose a situation which maximises $U(z)$.

Thus, objective of his model is:

$$\text{Maximise } U(z) \quad (3.9)$$

Subject to $px \leq Y$ (budget constraint)

$$\text{with } z = Bx$$

$$z, x \geq 0$$

Furthermore, the model can be expressed as a transformed utility function:

$$U(z) = V(x), \text{ the utility function of } x \text{ (level of activity characteristics)} \quad (3.10)$$

The demand for a good q can be written as:

$$q_j = f(Y, p_1, \dots, p_m, z_{11}, \dots, z_{1m}) \quad (3.11)$$

Where Y is the income level and p is price.

The focus on characteristics of goods rather than goods themselves in Lancaster model does not mean that these characteristics should be defined in terms of an individual's reaction (e.g., attitudes toward a good), but rather in terms of objective attributes (such as price, for instance), or properties of goods itself. Or, as pointed out by Louviere et al., (2000, page 3): "*Lancaster did not say that there could not be differences between consumers in the way in which they perceive an objective characteristic. However, if such differences exist, they relate to the formation of a preference function for t (or z , in my demonstration – my note) that is outside the domain of his theory*".

The main point of his assumptions is not to know how the function $U(z)$ is formed, but how people will react to changes in alternatives attributes characteristics (functions x and z). However, as his model assumes that many goods can produce one characteristic and many characteristics can be

produced by one good, it implies that goods are infinitely divisible, frequently purchased and of low unit value. This is a direct limitation of Lancaster's approach: many goods are not perfectly divisible, purchased or evaluated. In fact, some goods are indivisible, as sanitation, for instance, which is a discrete good (Louviere et al., 2000).

Rosen (1974) eliminated this limitation in Lancaster's model proposing an approach where alternatives were associated to a continuous range of objective characteristics, defined in terms of price and quantities of the features. According to Louviere et al, (2000, page 4), if a group of goods yields objective peculiarities (t_1, t_2, \dots, t_R) , and if we define all other goods consumed as d , Rosen's model assumes the following properties:

$$\text{maximise } U(t_1, t_2, \dots, t_R) \quad (3.12)$$

$$\text{subject to } p(t_1, t_2, \dots, t_R) + d = M \quad (3.13)$$

Where price of d is arbitrarily set equal to one dollar, M is the consumer's income, and $p(t_1, t_2, \dots, t_R)$ is the price of one good yielding objective characteristics. As Rosen assumes non-divisibility of goods, the budget constraint is non-linear, and characteristics are defined in terms of their absolute levels, not in terms of dollars. Hence, Rosen model continues to link utility directly to the objective characteristics of goods (as in Lancaster's model), and also permits a more appropriated approach to evaluate discrete goods.

Lancaster's model permitted a new dynamic for economic consumer theory. Moreover, data availability from National Census in United States, and the possibility of experiments using these data, during the 50s and 60s, allowed the operationalisation of this new approach and advances in discrete choice modelling (McFadden, 1998).

In 1974, McFadden developed a microeconomic tool to be applied in consumer theory, a statistical model for discrete response called currently the Multinomial logit (MNL, also called conditional logit) model. This model turned the possible link of unobserved preference

heterogeneity to a fully consistent description of the distribution of demands. Domencich and McFadden (1975) analysed the causal relationship amongst the attributes of the transportation system and individual's characteristics, in a model to predict the demand for urban travel.

McFadden's model was based on Marschak's work (McFadden, 1974) who explored the theoretical implications for choice probabilities of maximisation of utilities that contained random elements¹⁹. It was the RUM model, and originally it was a simple binary logit. From RUM approach, McFadden progressed to the MNL (simulated by maximum likelihood estimation), which was consistent with Luce's model and with the axiom of revealed preference (Domencich and McFadden, 1975).

In RUM model, the consumer is assumed to choose among a finite number of mutually exclusive alternatives, faced with conditional indirect utility²⁰. These are very important assumptions of the RUM model, because a short time horizon is assumed, where consumption will take the form of choices between mutually exclusive alternatives (McFadden, 1974; Maddala, 1983).

In this model, the consumer is supposed to have incomplete information and face uncertainty. In this way, utility is modelled as a random variable that reflects this uncertainty. This random variable takes into account all the effects not included in the systematic component of the utility function. The utility of the consumer i associated with alternative a is given by:

$$U_{ia} = V_{ia} + \varepsilon_{ia} \quad (3.14)$$

Where V_{ia} is a deterministic part of the utility, and ε_{ia} is the stochastic (or random) part, capturing the uncertainty. The alternative with the highest utility is supposed to be chosen.

¹⁹ Marschak, in his turn, was influenced by ideas from Thurstone's (1927) research. He explored psychophysical discrimination elements in a binominal probit model. These elements were interpreted as levels of satisfaction (or utility) within a model for economic choice (McFadden, 2001).

²⁰ It is conditional since it depends on the price of the chosen alternative and not all prices (Persson, 2001), and indirect because the utility is subject to various constraints not only budget constraints like in traditional utility theory. For a more complete discussion, see Madalla (1983), McFadden (1974, 1981).

The determinist part of the utility (V_{ia}) can be written as the utility of the individual i associated to alternative a of the a choice set, or

$$V_{ia} = V_i(x_{ia}) \quad (3.15)$$

Where x_{ia} is a vector containing the attributes, both of individual i , and the alternative a .

The function is assumed to be linear in all parameters^{21,22}, and it can be definite as:

$$V_{ia}(x_{ia}) = \beta_1 x_{ia1} + \beta_2 x_{ia2} + \dots + \beta_n x_{ian} \quad (3.14)$$

Where β_1, \dots, β_n are parameters to be estimated.

Therefore, the probability that alternative a is chosen by consumer i within choice set C is:

$$P_c(i) = \exp(V_i) / \sum_{k \in C} \exp(V_k) \quad (3.15)$$

Equation (3.5) is the multinomial logit. Note that:

- 1) Rational-maximization structure (like in traditional economic consumer theory) remain as the core of analysis, *but*
- 2) Utility function is related and derived from individuals' *and* alternatives' characteristics.

McFadden's MNL model, since then, is used in different applications in evaluating discrete choices. It is considered the "workhorse" of discrete choice models because it is the base model for different specifications. However, the MNL model suffers from the restrictive independence of irrelevant alternatives (IIA) property, which states that the odds ratio of two choice probabilities is independent of the other alternatives in the model. In practice, this implies that a change in an attribute of one alternative will have the same proportional impact on the probability of each of the other alternatives being chosen, or that the odds-ratio between two alternatives does not change by

²¹ Linearity in the parameters implicates that β appears with a power of 1 only. In other words, β is not multiplied or divided by any other parameter. For more details about linearity, see Gujarati (2003).

²² The core of specification and determination of logit/probit models is based on the distribution of the error term; it will determine the type of estimation of the model. The linear form of parameters is just to facilitate comprehension.

²³ Difference from McFadden MNL model to Marschak and Luce model was basically that in McFadden model the vector of utility V was the specified function of alternatives attributes while in Marschak and Luce model it was the preferences themselves or $P_c(j) = P_{(i,j)}(i)/P_{(i,j)}(j)$.

the inclusion (or exclusion) of any other alternative (Maddala, 1983; McFadden and Train, 2000). In sanitation analysis it will mean, for instance, that the ratio of the probability of choosing latrine linked to septic tank compared to the probability of choosing latrine linked to other traditional system will remain unchanged when Bahia Azul's programme starts to be offered.

To try to turn this property into a less rigid and to give it more flexibility for analysis, different models and different combinations of these models derived from RUM hypothesis, have been developed²⁴.

A new approach, which is more flexible and does not present the IIA restriction property, has now become quite popular among researchers. It is the mixed logit model (also based on RUM hypothesis) which will be presented in the next section. The flexibility of the mixed logit model allows analysis of large scale problems as well as the inclusion of psychometric variables.

In the context of mixed discrete models, "The Hybrid Choice Model" was developed. The aim of this behavioural framework is to address "*how psychometric data can be used in choice models to improve the definition of attributes and to better capture taste heterogeneity*" (Ben-Akiva et al., 1999, page 1). The model is a new and consistent proposal to deal with idiosyncrasy aspects inherent to consumer behaviour, and aims to fill the gap between behavioural theory and discrete choice models (Walker, 2001).

However, before introduces the Hybrid model, the concept of causality assumed in this thesis will be discussed in the next session. The economic concept of causal analysis will underline the specification and determination of the Hybrid model.

²⁴For example, the generalised extreme value (GEV) models. However, these models still present restrictions in representing all RUM-consistent behaviour.

3.4. The Economic Framework for Causal Analysis

The limitations of classical theory, which assumes the consumer as a “rational-economic man” in a process of utility maximisation, have opened opportunities for fruitful discussions and developments in the economic approach for consumer behaviour analysis. Lancaster model, as discussed in the previous session, introduced the concept of indirect utility where the goods qualities and characteristics not the goods itself is evaluated in a utility-maximisation framework. For both theories, however, the structure of preference-maximisation (or utility maximisation) is the *modus operandi* of analysis for the consumer’s decision process. And this has a special implication for the definition of causality in economics.

In assuming the preference-maximisation axiom as a framework of analysis, we assume the causal inference as a problem of missing data where the choice is revealed from the observed data. It means that in economic analysis based on utility maximisation assumption, only one of the outcomes is observed in the data (the one chosen), the other outcome (the one not chosen – a counterfactual) cannot be observed because the individual cannot do two things at the same time. For instance, suppose that every individual has two possible outcomes, Y_1 and Y_0 , where Y_1 is the outcome if the individual opts for a sanitation programme and Y_0 is the outcome if the individual does not experience it, all else held fixed. Only one of these outcomes can be observed (Y_1 or Y_0), and the causal effect (or effects) is defined by their comparison, e.g., $Y_1 - Y_0$. Therefore, the causal inference becomes a problem of inference with missing data (Rubin, 1974; Mealli and Rubin, 2003; Moffitt, 2003).

This model of causal inference is consistent with the consumer economic theory and econometrics methodology, where the causal effect of the event on Y is estimated on some assumptions or restrictions (ordered preferences, for instance) due to the unobservability of the counterfactual (the not chosen outcome), and those assumptions and restrictions cannot be formally

tested. Thus, they must be justified on the basis of a theory or argument (Heckman, 2000; Moffitt, 2003)²⁵.

In this model, however, only a exogenous variables (say X in a equation $Y = \beta X + \varepsilon$) is a causal variable, in view that the estimated coefficient on X is interpreted as the difference that would occur in Y if a particular individual were exogenously given one more unit of X (Moffitt, 2003). The problem arises when the variable in question whose effect on Y is potentially endogenous. In this case, *“the action-event is also affected by things we cannot measure, for we can never fully explain and measure the determinants of why individuals take an action or experience an event an others do not”* (Moffitt, 2003, page 449). This is the case of latent variables that are used in the model to explain choice for sanitation.

In fact, the causal relationship between the outcome Y and the latent variable cannot be observed directly, but through an “indirect causal path” inside a triangular specification (Hausman, 1983), involving a structural model specification and estimation that, according to Hausman (2003), is the only context where econometric causality can be reliable. These issues will be exemplified in the Chapter 6 where the model for sanitation choice is presented. In the next section, the Hybrid Choice Model is discussed.

²⁵ In this thesis, the theory is the indirect utility maximisation (Lancaster model), operationalised through a discrete choice model (mixed logit). Both approaches are synthesised in the Hybrid Choice Model.

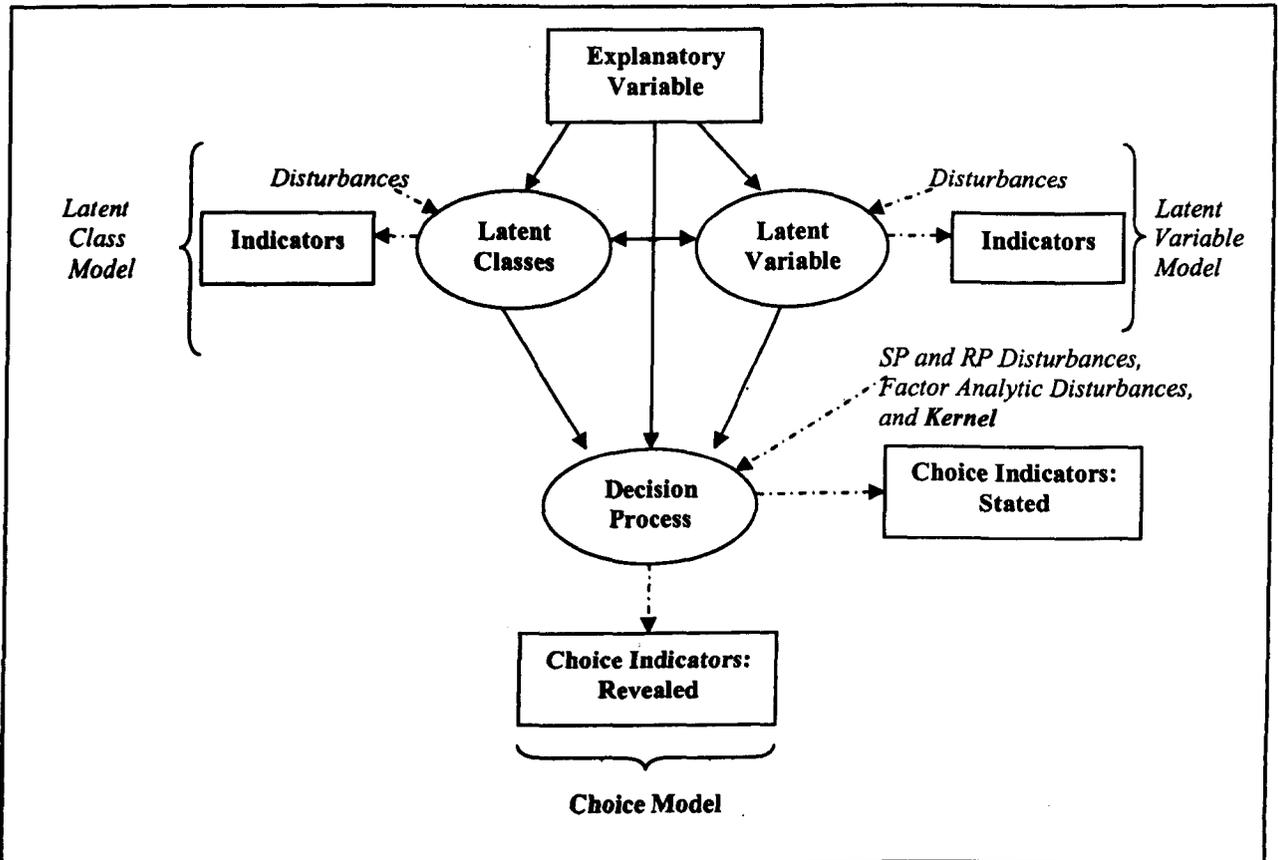
3.5. The Hybrid Choice Model

Recent developments in discrete choice models have permitted incorporation of psychometric data in the analysis of consumer decision process, providing a richer explanation of behaviour through the explicit representation of the formation and effects of latent variables on consumer behaviour. Contributions come from debate among economists, psychologists, marketers, geographers, engineers, and planners. The new approach permits the complicated problem of incorporating idiosyncratic aspects of behaviour, such as perceptions, attitudes, and cognitive processing into the random utility discrete choice models to be addressed. Usually called “The Hybrid Choice Model” or Hybrid model, this new structure of analysis permits a more realistic representation of behaviour in the choice process, with a better predictive power, producing consistent and efficient estimative of the parameters (Walker and Ben-Akiva, 2004; Bolduc et al., 2005).

Essentially, the Hybrid model emphasises that history influences context: the choice process is context dependent and governed by perceptions and attitudes (or other psychometric data) which are constructed throughout the time and vary among individuals (assuming heterogeneity across them). In this model, observable variables are also important to explain behaviour, but they are part of the explanation, not the deterministic behaviour. Therefore, the traditional systematic segmentation of the population based on observable variables to determine tendencies is substituted by a more comprehensive explanation based on individuals’ considerations, criterion, and preferences (Walker, 2006).

The graphical representation of the Hybrid framework is presented in the Figure 3.2. In this figure, the traditional discrete choice analysis can be seen from the vertical relationship among variables (explanatory variables – decision process – choice indicators). The Hybrid model expands the choice process incorporating latent variables and stated preferences in the analysis process.

Figure 3.2. Hybrid Choice Model



Source: Ben-Akiva et al., 1999

The terms in boxes are variables observed directly (or measured by suitable experiments): individual's characteristics and alternative attributes. The terms in ovals are theoretical or latent variables which are not observed directly, but are rather inferred from other variables. Usually, they are inferred by a scale (e.g., a semantic differential scale). The latent variables in this model incorporate elements characterised in socio-psychology and psychology as perceptions and attitudes. The arrows indicate the direction of relationship among the variables: latent variables are influenced by explanatory variable and are linked to the decision process through observed measures that include indicators (which represent manifestations of latent constructs). These arrows indicate the indirect causal path, presented in the previous section. These latent constructs can be evaluated by

an affective value (which varies in sign: positive or negative) and modelled using a structural latent variable model. The dashed arrows are related to the transformed variable (Ben-Akiva et al., 1999).

Explanatory variables are commonly modelled in statistical/econometric analysis as characteristics of individuals (e.g., socio-economic, demographic, experience, etc.) and alternative attributes (e.g., costs), into a random utility model related to an outcome (revealed preference). The challenge here is to incorporate latent variables and understand how they act on the behavioural process.

The integrated model is composed of two parts: a latent variable model (formed by structural equations) and the discrete choice model (measurement equations). Structural equations indicate the cause and effect relationships between the observable causal variable to the latent variable. From the values of the latent variables, measurement equations are constructed to demonstrate the distribution of the indicators (conditioned to the values of latent variables). Measurement equations represent the relationship between observable indicators and the underlying latent variables. This system of equations is then estimated simultaneously or sequentially (Bolduc et al., 2005).

3.5.1. General specification of the model

As described by Ben-Akiva et al., (1999), the general specification of the model – which is formed of two parts - should include the following steps (not necessarily in this order during estimation):

- 1) Define the structural equations for the latent variable model and choice model. As discussed in previous paragraphs, latent variables include different psychometric data (such as attitudes, perceptions) and economic data (preferences). For these latent variables (as

well as for the latent classes²⁶), it is necessary to specify their distribution, given the observed variables, or:

$$X^* = h(X; \gamma) + \eta \text{ and } \eta \sim D(0, \Sigma_\eta) \quad (3.16)$$

Where X^* is the latent variable, X are the observable variables, γ is the set of parameters in the choice model, and η is the error term, which its distribution (D), associated with each of the independent variable, is assumed to be zero and independent (Σ_η). One equation for each latent variable should be specified.

For the choice model, we need to specify the distribution of the utilities, or:

$$U = V(X, X^*; \beta) + \varepsilon \text{ and } \varepsilon \sim D(0, \Sigma_\varepsilon) \quad (3.17)$$

Where U is the distribution of utilities, β is the set of parameters, and ε is the error term.

Note that the systematic utility is a function of both observable and latent variables. It has an important consequence: the individual's utility for each alternative is assumed to be a latent variable, and the observable choices are manifestations of the underlying utility.

2) Define the measurement equations for latent variable model and choice model. As described in the Figure 3.2., values of latent variables are represented by an indicator, then for the latent variable model, we need the distribution of the indicators conditional on the values of the latent variables, or:

$$I = g(X, X^*; \alpha) + v \text{ and } v \sim D(0, \Sigma_v) \quad (3.18)$$

Where I is the latent indicator (that should be constructed to each latent variable), and α is the parameters. Also, note that indicators are function of observable variables and latent variables.

For the choice model, the choice is expressed as a function of the utilities, assuming utility maximisation, or:

²⁶ We represent "latent class" in the model when we are using categorical variables: variables that assume a limited number of cardinal values (nominal or ordinal).

$$y = \begin{cases} 1, & \text{if } U_i = \max \{U_j\} \\ 0, & \text{otherwise} \end{cases} \quad (3.19)$$

Note that equations (3.16), (3.17) and (3.18) determine the latent variables (X^*). Those variables are *indirectly* determined by indicators (I for perception and attitude latent variables) and an utility measure (U), and influenced by exogenous variables (observed directly), X . The indirect causal paths exist because of the correlation of the stochastic disturbance is not diagonal (Hausman, 2003). It means that we cannot observe the principal stratum to which a subject belongs because we cannot directly observe X^* for any subject.

The principal stratum is based on cross-classification (X^* , X) and, by definition, this stratum (or these strata) is not affected by the outcome, therefore, can be used as any other pre-treatment variable to define subgroup causal effects, indirectly (Angrist et al., 1996). Under these assumptions, (3.16), (3.17) and (3.18) estimates are causal effects for the principal stratum of households who are choosing a sanitation programme (Mealli and Rubin, 2003, page 83). A statistical demonstration is presented in the Chapter 6.

3) **Define the likelihood function.** It is determined by integrating equations (3.16) to (3.19) and aims to estimate the unknown parameters. It can be represented as the likelihood function of the integral of the choice model over the distribution of the latent constructs:

$$P(y|X; \beta, \gamma, \Sigma_\sigma, \Sigma_\eta) = \int_{X^*} P(y|X, X^*; \beta, \Sigma_\sigma) f_1(X^*|X; \gamma, \Sigma_\eta) dX^* \quad (3.20)$$

Where $f_1(X^*|X; \gamma, \Sigma_\eta)$ is the distribution function of the latent variable given the observed variable.

Introducing the indicators (I) to improve the accuracy of estimates of the structural parameters, and assuming the error components (η , ϵ , ν) are independent, the joint probability of the observable variables y and I , conditional on the exogenous variables X , is:

$$f_3(y, I|X; \alpha, \beta, \gamma, \Sigma_\epsilon, \Sigma_\nu, \Sigma_\eta) = \int_{X^*} P(y|X, X^*; \beta, \Sigma_\epsilon) f_2(I|X, X^*; \alpha, \Sigma_\nu) f_1(X^*|X; \gamma, \Sigma_\eta) dX^* \quad (3.21)$$

Where $f_2(I|X, X^*; \alpha, \Sigma_\nu)$ is the distribution function of the indicators. The latent variable is only known to its distribution, and so the joint probability of y , I , and X^* is integrated over the vector of latent constructs X^* (Ben-Akiva et al., 1999).

The first term of the integrand corresponds to the choice model, the second term corresponds to the measurement equation from the latent variable model, and the third term corresponds to the structural equation from the latent variable model.

There are two main ways to estimate the equation (3.21). One technique consists of first estimating the latent variable model (3.16) and (3.18), and then these latent variables and their distributions are used to estimate the choice model. The choice model can be estimated by any different forms, such as logit, probit, nested logit, mixed logit, etc. (Ben-Akiva et al., 1999). This method, however, results in consistent, but inefficient estimates of parameters.

Johansson et al., (2005), for instance, estimated a travel model choice using latent variables of attitudes together explanatory variables. Latent variables were modelled and estimated using the Multiple Indicator Multiple Cause model (MIMIC) which was assisted by exploratory and confirmatory factor analysis, and Weighted Least Squares. Attitudinal factors were defined as environmental preferences, safety, comfort, convenience, and flexibility. Their model was estimated sequentially: first all the latent variable models, and then, the discrete choice model, which was a multinomial logit model. They found consistent results for the latent variables to explain the travel model. But, there was no discussion of the efficiency of the estimates.

The second form to estimate the unknown parameters of equation (3.21) uses the maximum likelihood techniques that maximize the logarithm of the sample likelihood function over the unknown parameters (Hensher and Green, 2001; Louviere et al., 2000; Ben-Akiva et al., 1999). The equation is determined as (Ben-Akiva et al., 1999, page 11):

$$\max_{\alpha, \beta, \gamma, \Sigma} \sum_{n=1}^N \ln f_4(y_n, I_n | X_n; \alpha, \beta, \gamma, \Sigma) \quad (3.22)$$

This method results in consistent and efficient estimates of parameters, but “*as the number of latent variables increases, numerical integration methods quickly become infeasible and simulation methods must be employed*” (Ben-Akiva et al., 1999, page 12).

Ben-Akiva et al., (1999) and Walker et al., (2004) provided some examples of simultaneous estimation of equations (latent and choice models). In their examples, they stated that the approach of simultaneous estimation of equations is theoretically and statistically superior to sequential method approach, producing more consistent and efficient estimates of the parameters.

In the way of simultaneous estimation of equations based on Hybrid model, Bolduc et al., (2005) suggested different applicability of this approach applied to large scale models.

3.5.2. Identification of the integrated model

The structure of discrete choice models is well understood. Specifications that define the model were, generally, discussed in the sub-section 3.5.1. However, identification in mixed models is an issue that still needing further investigation. Absence of identification issues can lead to biased estimates and significant loss of fit (Ben-Akiva et al., 1999; Walker et al., 2004).

According to Walker et al., (2004), the model is based on the underlying theory and it is function of unknown parameters. As there is not a unique vector of parameters that solves the equation, infinite solutions can be considered, including infinite amount of data, creating a problem for solution and trustable conclusion from the data and model. Thus, “*the identification problem is*

to determine what conclusion can or cannot be drawn from a model, and under what sets of assumption” (page 4).

The most severe identification problem is about the assumptions, normalisation and restrictions of the model, that can influence the represented behaviour and, therefore, influence its prediction.

In view that identification is related to the assumptions of the model, measurement and distribution of variables, I will discuss it with more details in Chapter 6 related to the specific model for sanitation (the Hybrid model) and the variables collected through a questionnaire.

3.6. Summary

This chapter discussed the evolution and contribution of traditional economic consumer theory in understanding the consumer behaviour in choice analysis. Lancaster's (1966a, b) new approach for economic consumer evaluation, where utility can be expressed as a function of the attributes of the commodities, gave a new vigour for economic analysis in the consumer behaviour field. Incorporating Lancaster ideas to a RUM approach, McFadden (1978) developed the multinomial logit, which still the workhorse of discrete choice models. These approaches are consistent with the causal economic approach based on the counterfactual axiom where the causal inference is a problem of missing data.

Fruitful debate among economists, physiologists, and planners has pointed out for the necessity in developing more real approach for behaviour analysis, incorporating elements from psychology. Historical and cultural elements should be considered when planning and implementing programmes of intervention. Attitudes (and perceptions) are key elements in psychology and socio-psychology, and some of their concepts were imported to the arena of discrete choice models. The Hybrid model was developed in this context, and permits the incorporation of latent variables in the analysis of consumer choice within a discrete choice model structure.

Understanding and predicting the behaviour of decision makers when choosing among discrete can lead to significant changes in product or service design, pricing strategy, distribution channel and communication strategy selection, as well as public welfare analysis.

Benefiting from these developments in behaviour/economic literature, I assumed a Hybrid model to estimate demand for sanitation in Salvador/Brazil which will be operationalised by a mixed logit model, that have been proved flexible enough in incorporating latent structures. Details about these approaches are presented in the next chapter.

Chapter 4: A Framework of Analysis for Sanitation Demand Studies

4.1. Introduction

In the previous chapter, I discussed the contributions of the traditional economic theory to the development of discrete choice models, based on indirect utility maximisation (Lancaster model). The Hybrid model, originated from discrete choice studies and psychological assumptions on attitude and perceptions, aims to fill the gap between both field – economic and psychology.

In this chapter, I will present my model to evaluate demand for sanitation in Salvador/Brazil. It is based on the Hybrid approach.

The anthropological study and Jenkins' (1999) assumptions and findings in rural Benin will form the core for the construction of hypotheses that will be tested in the Chapter 7. Contrary to Jenkins' conceptual model, who assumed many interrelated psychometric measures, I hypothesise that the key psychological based variables to estimate demand for sanitation are perception and attitude towards it. These latent variables are formed through aggregated indicators and influenced by historical and cultural context. The socio-economic and demographic characteristics of households and the attributes of alternatives are also key elements for demand determination.

An overview of the mixed logit method is presented, and its advantages are highlighted. More details about this discrete choice model applied to my Hybrid model for sanitation will be discussed the Chapter 6.

4.2. Choice model for sanitation: framework of analysis

The framework of analysis to evaluate choice for sanitation in Salvador/Brazil is based on the Hybrid model. Details on the operationalisation of variables, including a discussion of the questionnaire, sample, data, variables definition, demonstration of equations, and modelling (using specific software) will be discussed in the Chapter 6. The aim of this section is to discuss the key variables that explain behaviour related to sanitation, which form the analytical framework. Furthermore, the mixed logit model is presented in a general form to explain why it is the best model to fit a Hybrid model for sanitation. Details on mixed logit specification and identification are also discussed in Chapter 6.

4.2.1. Historical and cultural dimensions of sanitation: the case of Salvador/Brazil before the implementation of Bahia Azul's programme

Jenkins (1999) correctly pointed to the importance of the consideration of cultural-based explanations of consumption and defecation-related beliefs have on sanitation evaluation. She stated that *“the cultural dimensions of latrine adoption reveal hidden and complex aspects of sanitation behaviour which are extremely important contributions to: 1) deepening understanding of how and why socio-cultural factors shape consumers' sanitation choices; 2) widening the range of strategies that can be designed to influence those choices; 3) explaining apparent inconsistencies in, and improving design and interpretability of knowledge, attitude, and practice (KAP) studies on sanitation and hygiene behaviour”* (page 49).

The traditional approach to sanitation evaluation and promotion is focused on the idea that sanitation provides health protection from faecal contamination. However, as was demonstrated by Jenkins (1999), in rural Benin, households had different perspectives from the traditional approach: they were looking for prestige associated with social status from affiliation with urban elite,

comfort, physical safety (avoiding open defecation), personal protection (linked to beliefs in Fon culture), privacy, restricted mobility, health benefits, cleanliness, and convenience. Nonetheless, health benefits had no connection with health protection from faecal contamination. It was connected to “reduce flies in my compound” and with health care expenses. The main constraint was the high cost of sanitation (latrine adoption), but this was *just* one part of the explanation of adoption.

Sanitation has an important cultural and social appeal and these characteristics should be incorporated in evaluating interventions. For instance, in Kenya, a latrine must be located outside the extended family compound, reducing its convenience, because using it in the presence of in-laws is a taboo equivalent to undressing before them (Almendon et al., 1994, *op cit.* in Jenkins, 1999). In addition, different latrine designs should be considered before implementation because some societies have different ways of expressing the ways to isolate faeces (Jenkins, 1999).

Localization (whether in urban or rural areas), social and political integration, and socio-economic characteristics also play an important role in sanitation evaluation.

In Salvador/Brazil, an anthropological study to understand the environmental perception and lifestyle of the population was carried out before the implementation of Bahia Azul’s programme of sanitation. The anthropological assessment involved poor populations who had no access to basic sanitation. The methods used were in-depth interviews with fifteen families where six families lived in a more urbanised area (although a poor area) and nine families from non-urbanised area (very poor garbage collection and street coverage, and lack of sanitation facilities). When they were asked for the “main thing they wish to change in their neighbourhood”, the main answers related to sanitation, and especially external connection (Barreto, 2002).

Several poor people in Salvador had a latrine inside their home (linked to septic tanks) before Bahia Azul’s programme of sanitation was implemented, even if it did not have any flush toilet. An

explanation for high adoption of latrine use could be the people's awareness of the direct relationship between lack of sanitation and disease. This information was captured by the anthropological study where people mentioned the words "disease", "diarrhoea", "intestinal infection", "cholera", "dengue", "rat disease (leptospirosis)" as some consequences for the lack of sanitation. Another reason was the social pressure for adoption.

Neighbours who had access to some type of faeces isolation and management put pressure through complaining about bad smells and dirty scenes. Some of them put the non-adopter under threat (in reporting the conflict to Bahia State Public Health Surveillance), considering that the public area should be preserved and family excreta should be treated in a way that did not effect others. Nevertheless, as it is not always possible for families to have access to some type of faeces treatment, the conflicts remained in the poorest areas. To manage this conflict (and be socially accepted by the group, neighbourhood), many families "constructed" some holes near home to put the faeces. When this hole was full, another was dug and the previous one was closed. The other solution was to put faeces in a plastic bag and throw it away, usually in a garbage area. Both solutions, however, caused bad smells and visual problems, and conflicts remained.

The political aspect was also raised during the interviews. Sanitation was viewed as an essential social right, and that should be provided by the State for the whole population. An interesting factor in these interviews was the awareness between public and private rights, in the discussions about sanitation, demonstrated by the interviewees. Interviewees saw external sanitation as a public obligation (in keeping with social welfare), which should be provided by the State, and understood that the internal sanitation (latrine and internal connection) was a private task. However, they also considered that, due to their poverty, the State should provide some support (subsidies or free access) to permit the adoption of latrine use. They also linked their condition to their exclusion from society with a lack of access to basic infrastructure (Barreto, 2002).

Environmental aspects were also important issues of social concerns about the lack of sanitation. It confronted the common assumption in industrialised countries, which assume that people in developing countries do not care about environmental quality, some studies have demonstrated the contrary (Tesh and Paes-Machado, 2004).

Results of a Gallup poll conducted in 1992 revealed that 80% of Brazilians had “a great deal” or “a fair amount” of personal concern. In Mexico, this percentage was 83%, in Uruguay was 82%. The rates of these three countries were similar to United States rate, which was 85% (Tesh and Paes-Machado, 2004).

In a poll with 580 people (243 middle-class people, and 337 poor people) conducted in Salvador/Brazil, results indicated that, when those people were asked about environmental problems, they mentioned lack of urban sanitation (like open-air sewers, polluted beaches, and uncollected garbage) as the most important problem. These aspects were more remarkable for poor people (40% of responses) than for middle-class people (34% of responses). Identified environmental problems included destruction of natural resources, uncontrolled industrialisation, bad policies, and lack of environmental control. Unsanitary conditions remained the most important problem even when compared with general environmental problems (like pollution) in Salvador or local problems (like garbage collection, irregular supply of water) - Tesh and Paes-Machado, 2004.

Results of this research will be discussed further in Chapter 8 and will support the findings presented in Chapter 7.

What is important to emphasise from these results is the necessity to explore the historical and cultural of determined intervention, and not treat the evaluation as a simple analysis of observable variables. Observing just direct explanatory variables (such as socio-economic, demographic and attributes of alternatives) is useful in some cases, where it is difficult to conduct surveys to collect

psychometric and anthropological data (due to cost of data collection, for instance). However, this is a necessary effort that brings more consistency to demand forecasting.

4.2.2. Sanitation: is it a private or a public good?

Previous findings indicated that sanitation is a very particular service whose definition depends on the cultural and historical context in which it is inserted. Therefore, its characterisation as a public or private good may fall into different categories when conditions vary, and so it can vary among countries (Riviere-Cinnamond, 2006). Whittington et al. (1993), for instance, classified sanitation as a private good as many of their studies were conducted in African countries, where government sanitation policy is practically non-existent, and households pay for treatment of their excreta disposal.

In Brazil, the government (States and Federal Government) has assumed the role of provider of external sanitation through the construction of treatment stations, garbage collection, and cleanliness of oceans, lakes and rivers, and improvements of street coverage. The government's aim is to offer these services for the whole population, so avoiding the negative externalities due to the lack of sanitation (this point will be discussed in the next section). On the other hand, the population should pay for their sanitation internal facilities (latrine, flush toilet, etc.) and connection to the government sanitation system. Households that already have a sanitation system different from the government one (e.g. septic tank) should change to a new system (e.g. Bahia Azul). Hence, how should this service be classified in Brazil?

Cullis and Jones (1998) defined public goods from two critical properties, it is: 1) a non-rival, and 2) a non-exclusive good/service.

The non-rival in consumption characteristic of a public good implies that the consumption of one individual does not reduce the benefits derived by all other individuals. For instance, with the

investments in the Bahia Azul's program of sanitation, the Bahia Government offered external sanitation for the whole population (external to the houses), the immediate impact of this intervention was on the reduction of risk of diarrhoeal disease: in Salvador, the risk of diarrhoeal disease among children was 98% higher for children living in homes without excreta disposal than for children living in homes with excreta disposal. In addition, the reduction of diarrhoea's incidence for children less than three years was 13% in relation to the period pre-intervention, and the reduction of selected intestinal infectious disease, the impact was higher: 51% of reduction in relation to the same period and age group.

The above example also describes the second property of public good, the property of non-excludability: once that the service was offered to the whole population, the benefits from the programme could be exercised for all individuals.

Therefore, can we assume that sanitation is a public good in Salvador (and Brazil)?

Not really. In Salvador, the sanitation program had two main phases (as discussed in the Chapter 1): 1) the external one, where the government offered sanitation and environmental protection, covering the whole population; 2) the internal one, where households should pay to connect their internal sewer to the programme.

This second phase implies that a household that does not pay for connection remains at risk of acquiring infectious disease preventable by faeces isolation. The reduction of risk of infectious disease associated to sanitation was evident since Bahia Azul was implemented because it reduced the risk of disease acquired from the *external* font. However, the risk of disease from *internal* ones remains. Households without sanitation at home (latrine linked to a connection) have a higher risk of diarrhoeal disease than one who does have access to the service.

Then, if it is not a public good/service, what is it? Cullis and Jones (1998) classified as a mixed good or a quasi-public good those goods or services that are non-rival in consumption, excludable in

principle, but non-excludable from the generated benefits. This is the classification I will assume for sanitation in Salvador (and Brazil).

The benefits of the external sanitation of Bahia Azul were extended for the whole population in Salvador and, as a consequence, the incidence of diarrhoea disease and other intestinal infectious disease was reduced. However, the entire benefits of sanitation could not be deserved for the whole population, because part of them will be excluded of gains. Nevertheless, the more households that are connected to the programme, the more the benefits to the society will be if sewage are not disposed on the street, the risk of food contamination and the acquisition of diseases directly from the excreta disposed in the street will decrease. Hence, in spite of the fact that only part of the population are paying for sanitation, the benefits generated are available to all, creating non-rival and non-excludable benefits. Or, a private-public service *“can be viewed as having private benefits as well as external effects which bear the characteristics of public goods”* - Cullis and Jones (1998, page 51).

4.2.2.1. Externalities

Households' connection to a sanitation programme or any isolation system creates *positive externalities* for the whole population. In other words, it creates benefits that can be extended to all individuals. Nonetheless, the positive externalities of sanitation are not only linked to the reduction of diseases. In Salvador, for instance, the cleanliness of beaches, lakes, rivers, and streets has improved the tourism in that city, increasing the income from this activity.

In 2005, Salvador was a tourist destination for 5.2 million of people (national and international residents). It was estimated that 5% of Bahia GNP was raised from tourism, and in Salvador, it should be higher, as this is the main city destination in the whole Bahia State (data refer 1999 – IBGE, 2004).

The absence of sanitation can affect not only the individuals, but also the Government income and costs. Polluted rivers, lakes, and oceans, and dirty streets affect the image of the city to tourists reducing the Government income raised from it. The lack of sanitation increases the incidence of infection diseases, increasing the governmental costs with treatment (at health facilities and hospitals). In addition, the entire society loses with individuals' absence from work and schools (social costs).

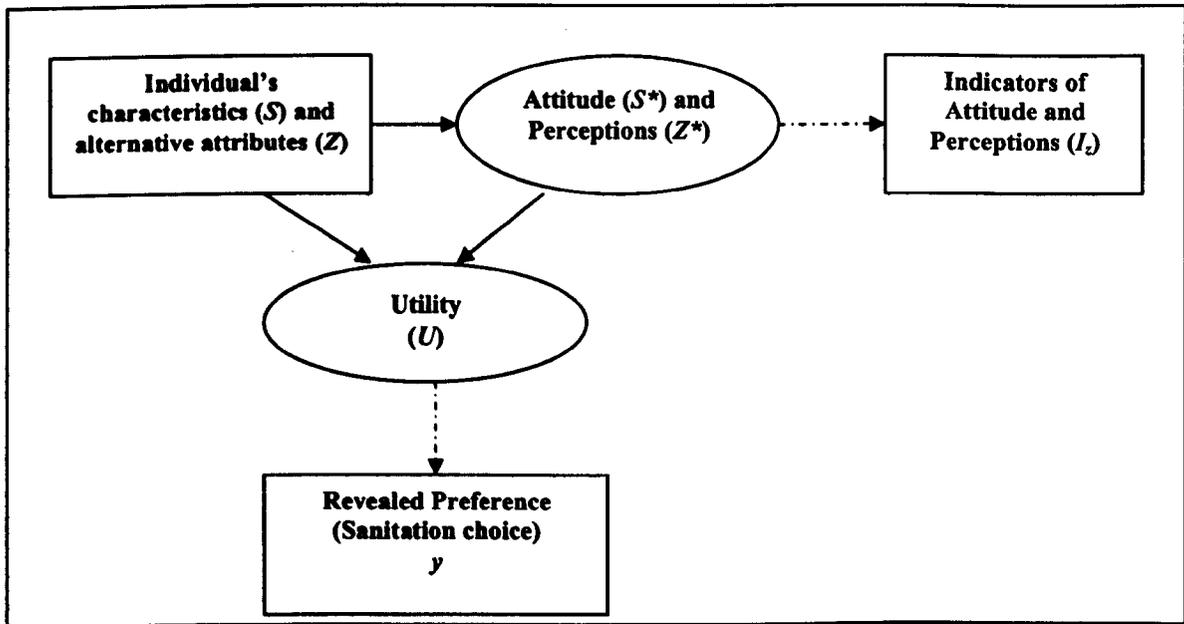
4.2.3. Behavioural conceptual framework for sanitation

I have used the Hybrid model as the framework of analysis for sanitation in Salvador/Brazil. My hypotheses are that choice of sanitation is a function of individual's characteristics (socio-economic, demographic), alternative attributes (e.g., costs), and latent variables of attitudes toward latrine and perceptions of the service. My modelling framework diagram is showed in the Figure 4.1.

As in McFadden (2001), I assumed that history matters. However, as I could not demonstrate the causal relationship between history and the variables in my model (it was not part of my objective), I assumed that cultural and historic aspects underlie those variables. In my model, history is assumed to underlie socio-economic and demographic aspects of the individual (similar to Jenkins, 1999).

Although this model allows the incorporation of different types of psychometric data, only three latent factors were incorporated in the analysis: attitudes, perceptions and preferences. These terms were broadly defined in the section 3.2, in the previous chapter. Nonetheless, some other aspects are discussed below.

Figure 4.1. Behavioural Framework for Choice of Sanitation in Salvador Brazil



Source: adapted from Ben-Akiva et al. (1999).

Attitudes and perceptions of individuals are hypothesised as key drives underlying behaviour, and affect individuals' preferences toward different alternatives and their decision making process. Attitudes reflect individuals' needs, values, tastes, and capabilities, and are affected by experiences (which are formed over time) and socio-economic characteristics. Therefore, attitudes also explain individual heterogeneity. In sanitation, for instance, they can represent the advantages, qualities, or importance of latrine adoption (e.g., importance of cleanliness, health benefits). As well as attitudes, perceptions explain part of the measured error of the choice process captured by the discrete model. Perceptions are mental representations of an idea or image, and form the individuals' beliefs. Examples of perceptions in sanitation are convenience, safety, social status, comfort, health benefits. Preferences represent the desirability of alternative choices, and can be defined through experiments (stated preferences) or observed choice (revealed preferences). The cause-effect behavioural

relationship among these factors is modelled through a causal mapping, using indicators to perform a random component of the utility function (Ben-Akiva et al., 1999).

This approach is different from that adopted by Jenkins (1999) in evaluating sanitation in rural Benin. In fact, the definitions are not uniform, and can encompass each other. Attitudes are a broad term in socio-psychology, and can incorporate perceptions, beliefs, drives, and other affective response. Also, it can mean only a measurement scale, or “elicited verbal attitudes”, which are responses to questionnaires, interviews schedule, or other measurement. Some researchers use only the term “attitude” in a broad sense to explain behaviour (Green, 1954; Schuman and Johnson, 1976). Therefore, definition of the terms depends on the research criteria. I assumed the same definition presented in Ben-Akiva et al., (1999), which link attitude to individuals’ characteristics and perceptions to attributes of alternatives.

I also assumed that individuals will choose sanitation (link to Bahia Azul or other service) if they have a positive evaluation of this service. Otherwise, they will opt for open defecation. Obviously individuals, who evaluated sanitation positively, still cannot adopt it due to their socio-economic conditions or other factors. Therefore, sanitation also has a direct relationship with socio-economic and demographic characteristics of individuals, and alternative attributes, as cost. High cost can be prohibitive for some individuals.

Regarding the socio-economic characteristics, I expected that individuals in better socio-economic conditions would choose to do a connection (latrine connected to Bahia Azul’s programme, septic tank, or municipal system). I expected similar results to those found by Persson (2001), in Bangladesh, and Abou-Ali (2003) in Egypt, in relation to socio-economic characteristics. Their results revealed that household educational level was an important determinant for toilet facilities. People who had a medium to high education level tended to choose services for sanitation that had a better quality - modern facilities with flush pit latrine.

The rational-maximiser assumption was the core of analysis. I believe that consumers maximise their choice in some sense, and compare the utility of alternatives opting for that one which he/she believes that is the most affordable alternative, given his/her budget constraint. Utility here has the same conceptualisation as that in Lancaster (1966a, b) and McFadden (1974): individuals derive utility from the goods attributes (now, including latent aspects as attitude and perception), the indirect utility. Utilities are represented by preferences, and in this thesis, only revealed preferences are analysed.

Details about variable description and analysis are discussed in the method's chapter (Chapter 6). Now, I would like to introduce the mixed logit model, its general form and advantages for analysis. Details about specification and identification for the analysis of sanitation choice will be also discussed in Chapter 6.

4.2.4. *The mixed logit model*

A mixed logit model refers to models that are comprised of a mixture of logit models. From equation (3.14), if we divide the random utility term, ε_{ia} , in a probit-like component and an i.i.d Gumbel (or GEV), we have a mixed logit model²⁷:

$$U_{ia} = X_{ia}\beta + F_{ia}\xi_{ia} + v_{ia} \quad (4.1)$$

Where:

$$\varepsilon_{ia} = F_{ia}\xi_{ia} + v_{ia} \quad (4.2)$$

Note that if $\xi = 0$, we have the multinomial logit.

²⁷ As stated before, this combination can result from different logit and probit models. A probit-like component is just an example; i.i.d Gumbel means: an independent and identically distributed Gumbel, any member of General Extreme Value models (GEV), such as multinomial logit, nested logit and cross-nested logit (see McFadden, 1974, McFadden and Train, 2000).

From equation (4.2), the mixed logit is estimated by the integrated logit probability over all values of ξ weighted by the density of ξ , to obtain the unconditional choice probability for each individual. Or:

$$P(j) = \int_{ia} \left[\frac{e^{\mu(X_{ia}\beta + F_{ia}\xi_{ia})}}{\sum_{j \in C_{ia}} e^{\mu(X_{ia}\beta + F_{ia}\xi_{ia})}} \right] \quad (4.3)$$

McFadden and Train (2000) proved that any well-behaved RUM-consistent behaviour can be represented with any mixed specification, and presented easy implementation specification tests for these models. Although there is not a closed-form solution for this model, integrals in the choice probabilities can be approximated using Monte Carlo techniques, and values of ε is drawn for each individual using Halton sequences, for example²⁸.

The flexibility of mixed logit model permits a series of advantages compared with other models (McFadden and Train, 2000):

- 1) Any number of elements may be included in the random term ε ;
- 2) Mixed logit model does not present the IIA property, which confers a significant flexibility in the analysis;
- 3) It can estimate any number of random coefficients or error components, and ξ is not subject to the identification restrictions of the covariance matrix of the unobserved portions of utility;
- 4) It also can estimate random coefficient in a model that has only two alternatives, a restriction for multinomial probit models, for instance;
- 5) Elements of ε , in mixed logit, can follow any distribution.

These are some of the important features of mixed logit models that make possible the estimation of models where latent variables are included. Whereas parameters in multinomial logit

²⁸ Halton sequence in simulation techniques will be discussed in the Chapter 6.

are estimated as a function of the observable variables, in mixed logit models the parameters can also be function of the latent variables. Thus, mixed logit can be specified to answer many questions about behaviour consumer choice. My Hybrid model for sanitation will be operationalised using a mixed logit model, as will be discussed in the Chapter 6.

Chapter 5 – Descriptive Analysis of a Household Survey of Sanitation Adoption

5.1. Introduction

In Chapter 4, I discussed the proposed behavioural framework of analysis for choice of sanitation in Salvador/Brazil. Results from an anthropological study (Barreto et al., 2002), during the pre-intervention period of Bahia Azul's programme, demonstrated that sewage treatment (or absence of that) had a cultural and social appeal for the population in that city. Poor individuals could associate the lack of sanitation with their poverty condition, the cause of diseases, and environmental problems. In addition, they perceived access to appropriated sewage system as a social right, and therefore, it should be provided by the State. In their vision, the State was responsible for providing external treatment of excreta disposal and households were responsible for the treatment of domestic sewer, which included latrine and external connection (septic tanks, municipal sewer system). However, due to their poverty condition, they understood that the Government should facilitate their access to the external sanitation system through total or partial subsidies.

Historical and cultural dimensions of sanitation are elaborate relationships, and their determination processes are too complex to be represented by a single variable or some of them. For this reason, I assumed that household history and cultural aspects *underlie* the consumer choice process. I also assumed that perception and attitude towards sanitation were expressions of those dimensions and, therefore, influenced by them.

This chapter presents a descriptive analysis of a survey of sanitation adoption. The objective was to understand the defecation practices, sanitation experience, perception, attitude, beliefs related to drives and constraints, and socio-economic and demographic characteristics of households. A subset of those results will be part of demand analysis presented in the Chapters 6 and 7. As discussed in the Chapter 2, a model to forecast demand is only useful if the causal relationship among variables can be demonstrated. As we have several interrelated variables whose causal process is obscure, I have opted to include only the attitude and perception latent variables, as discussed in the previous chapters, understanding that they are influenced in some way by historical and cultural contexts.

Hence, the research goals of this survey were to provide information to construct the latent variables that will be part of demand analysis, and to offer additional explanations for household behaviour when choosing sanitation alternatives. Only descriptive analysis was used at this stage.

Alongside this introduction, the chapter is divided up as follows: in section 5.2, I discuss the methodological issues involved in this study, describing the sampling and data collection procedures. It included details about the study population, questionnaire design and topics, and data analysis procedures. In section 5.3, I present the results of the household survey examining the characteristics of adopters and adopters and differences between them when choosing sanitation. In section 5.4, I present a summary of the results.

5.2. Methodological issues

5.2.1. Study population

My research team visited 1,128 households, which corresponded to those 1,216 children of the cohort of the epidemiological study.

During 1997-1998, to evaluate the prevalence and incidence of diarrhoeal disease, a cohort study with a sample size of the 1,216 children was implemented and final analysis occurred in 2005²⁹ (Barreto et al., 2002, 2005). Inclusion criteria for participation in this cohort study were: 1) children living in areas with good and poor (without) sanitation; and 2) residing in Salvador/Brazil.

Data collection for the demand study occurred between November 2004 and April 2005. The original idea was to compare the information in the cohort study database with the demand estimation, presented in this thesis, however, data about type of connection collected from the first study were confused and duplicated, so I had to choose a non-comparison study.

5.2.2. Estimated sample size for the epidemiological study

For this study, the sample size was estimated to provide a study power of 80% and error type I of 5% to detect a reduction of 25% in the incidence of diarrhoea associated with sanitation between areas with and without a sewage system. The sample size was calculated to be 295 (24.1%) children from sanitary area and 921 (75.9%) children from no sanitary area. This corresponded to 1,128 households, and these were the total number of households visited for the purpose of administering a questionnaire in my demand study.

Sample was designed to be representative of intervention areas, not for the whole population in Salvador.

²⁹ Besides diarrhoeal disease, other selected infectious intestinal diseases were evaluated for different sample sizes and age group.

5.2.3. Effective sample used for demand estimation

All of the 1,128 households that were selected in the cohort study during the pre-intervention period were visited by the research team. From those 1,128 households, we could obtain information from 721; 389 (34%) moved to other neighbourhood or other city; and 18 (2%) were re-visited, but refused to answer the questionnaire. Therefore, my effective sample was composed by 721 (64%) households.

5.2.4. Previous qualitative study and questionnaire design

The questionnaire used in this thesis is very similar to that used in Jenkins (1999). It was a questionnaire based on socio-psychological techniques, and designed to explore variables related to household behaviour towards sanitation. In fact, only a few questions regard preferences (not economic preferences), and characteristics of population, available services of sanitation, costs of latrine, connection, and maintenance, and questions that explored the health component associated with sanitation were included in the original version of Jenkins' questionnaire. Those questions were related to the anthropological study that was carried out in Salvador during the pre-intervention period to evaluate environmental perception and the living conditions of the population.

Some open questions were included in the questionnaire, to evaluate the spontaneous household answers associated with perceptions and attitudes toward sanitation, as well as some related to beliefs and constraints acting on household choice. The questionnaire was originally written in Portuguese. A translated version is in Appendix 1.

5.2.5. Topics included in the survey questionnaire

The survey questionnaire consisted of closed-ended pre-coded questions and some open questions, and incorporated the following topics:

- Characteristics of the house;
- Present defecation practices;
- Perception and attitude of the present defecation site of sanitation's users and non-users; "perception" was characterised as the qualities and non-qualities related to the open defecation (for non-adopters) and latrine (for adopters and non-adopters); it was measured using a 5-point semantic differential scale (1=very dirty, 5= very clean, 4= neutral); "attitude" characterises individuals and their beliefs about sanitation; it was measured as advantages and disadvantages of sanitation (3-point semantic differential: 2=little importance; 3= important; 4= very important) - Osgood et al., 1957. These measuring scales followed the same structure proposed by Jenkins (1999); Association of perception with qualities and non-qualities of sanitation, and attitude with advantages and disadvantages was discussed in the Chapter 3;
- Preference for latrines and non-adopters' intention to adopt;
- Reasons for non-adoption and presence of various constraint factors;
- Importance of advantages (drives) and disadvantages (more constraint factors) of sanitation adoption, besides that ones evaluated in "attitudes" measures;
- Costs of sanitation adoption, connection, and maintenance (for users);
- Socio-demographic characteristics.

5.2.6. Definition of adopter and non-adopter

For this thesis, an adopter was considered the household that had a latrine inside the home connected to any system of external faeces isolation (septic tanks, old municipal sanitation system, Bahia Azul's programme, or other). Non-adopter was one that had no latrine inside home (or outside home), or, even though they had a latrine, it was not connected to any external faeces isolation at all (their sewage was disposed in the streets).

5.2.7. What was evaluated by the questionnaire (questions about attitude and perception)?

Attitude and perception were related to latrine evaluation, not type of connection.

My pilot study with few households to evaluate perceptions and attitudes demonstrated that there was a difficulty in evaluating the connection *per se*. Evaluation was better understood when connected to the latrine. As a latrine is a good with a more direct relationship with faeces isolation, and so, more easily perceived by the household, I opted to evaluate the household perceptions and attitudes (including beliefs) for the latrine instead of the connection, and I assumed that there was an indirect link between these psychometric indicators for latrine and the type of connections.

Households that had a latrine inside home, but had not any type of connection, were grouped into the non-adopter set (as previously defined). Both adopters and non-adopters assessed latrines: qualities (non-qualities), advantages (disadvantages). The households that had no latrine inside home practiced "open-defecation" (deposit faeces in a bag and throw away, in a potty, in a land hole outside home, or other). For this group (non-adopters), we asked them if they had contact with latrine during their life, and how frequent it was. If answer was affirmative, we asked them to evaluate latrine, if not, they did not evaluate it. They were also asked to evaluate their current practice of defecation (open defecation). It permitted us comparisons between latrine and open defecation in terms of perceptions and attitudes.

5.2.8. Possible problems during interviews

A problem that could occur during an interview was that the respondents might try to respond to question to impress or please the sponsor of the survey hoping to receive a free or subsidised latrine and/or connection in the future (Jenkins, 1999). I have two main reasons to think that this problem was minimised. First of all, in Salvador the population often responds to surveys from the university and Government Statistics Agencies. The university has a good reputation among population and, especially the Instituto de Saúde Coletiva (Collective Health Institute) has been working with this population for several years. Second, research assistants presented themselves as “university research assistants”, and explained carefully the purpose and outcome of the survey. All research assistants were dressed with T-shirts identifying the Federal University of Bahia association. Nonetheless, it is possible that for some variables (as household income, for instance), the information could have not been accurate. This point was discussed in the results below.

5.2.9. Researcher assistant training and data collection

There were four research assistants assisting the administration of the questionnaire. All of them had received previous training in applying the questionnaire. They were instructed to read the questions to each household, and repeat the question in case of household misinterpretation. If the respondent did not understand the question, they were instructed to leave the question blank. Supervision was conducted by myself during the first three weeks, when I attended the interviews with them. Each day questionnaires were checked by myself. In case of doubt or incomplete information, the interviewer returned to visit an interviewee to correct, complete, or clarify responses.

5.2.10. Missing data

In spite of our efforts to avoid missing data, I detected some missing or inconsistent information during daily questionnaire reviews. When it happened, research assistants were asked to re-visit the households and collect the necessary information. However, some individuals were not found on the second visit, and some refused to be interviewed again, therefore some missing data remained. From the total sample of 721 households, 12 had missing data:

- Five questionnaires had no information about household income (neither household head income, nor other member of family income's information);
- Attitude toward latrine information was missing for three adopters;
- Two questionnaires had no data about the number of children; and
- Two had no data about household members level of education.

For this descriptive analysis, households that presented missing were excluded from specific analysis. For instance, when evaluating number of children at home, two information were excluded. Same procedure was used when evaluating income, attitude towards latrine and level of education. However, to estimate demand for sanitation, different procedures were adopted. These procedures will be discussed in the Chapter 6.

5.2.11. Ethics

Two ethical committees approved the research: London School of Hygiene and Tropical Medicine, UK, and Federal University of Bahia, in Salvador, Brazil.

All household heads received a letter informing them of the objectives of research. This letter was read to each household by research assistants, and if they agreed to participate, the researcher asked them to sign the consent form (see copy of document in Appendix 2).

5.2.12. Data entry and analysis

Data were entered using EPI-INFO version 6.0 by a computer typist, and checked by myself. For analysis, data were transferred in STATA-Transfer to STATA 9.0.

Survey data were analysed using descriptive statistics: mean, standard deviation, difference and groups and significance tests, and percentages. These statistics provide simple summaries about the sample and the measures, being the basis of quantitative analysis.

5.3. Results

5.3.1. General characteristics of households

In the household sample, 90% of interviewees were female and stated they were the head of the household. Around 62%, however, were housewives with some “sporadic income” from informal work. The sample median age was 32 years and they had low educational level: 8 and half years of study. They had, on average, 2 children in the house (median).

Table 5.1. Socio-economic, Demographic and Characteristics of Sanitation Choice for a Sample of Households in Salvador/Brazil (N=721)

Female: n (%)	649 (90)
Age: median (std.dev.)	31.89 (12.23)
Housewife: n (%)	450 (62.4)
Education: median (std.dev.)*	8.57 (3.83)
Income (monthly average during 2004) : median (std.dev.)**+	462.19 (338.00)
Number of children at home: median (std.dev.)*	2 (1.37)
Type of sanitation: n(%)	
<i>Bahia Azul</i>	499 (69.21)
<i>No sanitation</i>	71 (9.85)
<i>Septic tanks</i>	64 (8.88)
<i>Other</i>	87 (12.07)
Water Closet (WC) toilet: n (%)	514 (71.29)
Cost of sanitation: median (std.dev.)^	83.35 (180.64)

*N=719

** In Brazilian currency Real equivalent to 2004's US\$ 157.94, in nominal prices; + N= 716

^ In Brazilian currency Real equivalent to 2004's US\$ 28.48, in nominal prices;.

Research assistants were told to ask for the person responsible for the household: the person who was the responsible for the main decisions in the house, independent of current occupation (if the person was or not the financial provider). It meant that the person responsible for the household

should respond about children's education, household organisation and decisions about acquisitions to the house (including sanitation choice). Data from the Brazilian Institute of Geography and Statistics (IBGE, 2004) supported my findings: women are assuming the role of the head of household, as much in terms of financial provision as in terms of household expenditure decisions, even if they are not the main resources provider. This is a tendency verified in the whole country, especially for the last decade (from 1994).

Median monthly income was R\$ 462 (US\$ 158). The standard deviation was very high (338). It showed a high variation of income around the mean: the minimum value was R\$ 0 and the maximum R\$ 2,800. Checking the sample, only 28 households stated they had no income at all. If information from these 28 households is excluded from the analysis, the median income is R\$ 481 (US\$ 164), with standard a deviation of R\$ 331.47.

This variability could be explained by two main factors. First, some households could have refused to provide information about their real income fearing to be taxed or excluded from some social benefit. In Salvador, informal job is the most usual income generator for poor and low skilled people. For this kind of work, regulation is difficult and several people do not pay any tax on their income to the State. In addition, the Federal Government maintains important social programmes for specific population (very poor households)³⁰. As most of respondents come from a low socio economic strata, reasons for non-information of correct family income could be fearing to lose benefits, in spite our explanation that we were part of University and not part of the Government.

A second possible explanation was related to the population socio-economic characteristics and areas where the study was conducted. The study was carried out in different areas of intervention, areas with good, poor, or inexistent sanitation. Thus, it is expected that household income differs among those different areas.

³⁰ It also depends on the number of children at home, if they are regularly attending schools and having vaccinations and medical assistance.

Comparing my sample results with other official research conducted about the labour market, such as the Employment and Unemployment research (PED) conducted by the Bahia State Government, Labour Unions, and State foundations (SEI), the monthly median income for the 25% poorest strata of the Economically Active Population (EAP) was similar to my findings: in 2004 it corresponded to R\$ 452 (US\$ 154). Therefore, we can assume that the sample results were a good proxy to the real value (IBGE, 2004).

Regarding the type of sanitation, 69% already had a connection with Bahia Azul's programme of sanitation, and 71% had a water closet toilet (latrine with flush toilet). However, around 10% of investigated population had no sanitation.

Part of explanation for Bahia Azul's high rate of connection is due to the Law number 7.307 from 23rd January 1998, which establish in its clauses the obligation for connection to Bahia Azul programme ("*...household must promote the connection of their internal sewer to the Bahia Azul programme... in a maximum period of 90 days...*" – Bahia, 1998, page 7). Some households stated that they preferred "to sacrifice" some basic consumption, such as clothes, than have problems with the State's inspection.

Another explanation was the neighbours' pressure to connect - which strengthens the assumption of social pressure raised in the Chapter 4. Furthermore, not all households had to pay for the whole sanitation (latrine, flush toilet and connection): most households already had a latrine connected to another system (like municipal system or septic tanks), and they had to pay only for the connection (Bahia Azul's fees). On the other hand, some households had only latrine, but no connection at all, and had to pay the whole cost for connection (building the connection under land), including Bahia Azul's fees. All households who adopted Bahia Azul's programme, however, have to pay a monthly fee for maintenance, which increases their total costs.

About 12% of households had access to another type of sanitation, like the old municipal system. However, there are only a few areas being served by this system, and it is expected that this structure will be integrated to Bahia Azul's programme in the future, when the programme, in fact, covers all the region (All Saints Bay). In this case, therefore, households should pay only the costs for connection and the regular monthly fee for maintenance. The 9% of households that use septic tanks to dispose their sewage were expected to change to Bahia Azul's programme. However, this change will imply in cost of the construction of the linkage between the internal household sewer and the programme, and Bahia Azul's monthly fees for these households, and it is possible they will not be willing to do this, in spite of the Law.

In short, it is expected that the household connection to Bahia Azul's programme will increase in the future years because, principally, of the Law. Nevertheless, because of the costs, it is likely that many domiciles will remain connected to septic tanks.

The median fixed cost of sanitation was R\$ 83.35 (US\$ 28), in 2004. This also had a high standard deviation (R\$ 181; minimum R\$ 1.50, maximum R\$ 1,558). This volatility was expected because the cost included those households who just pay for connection, and those who pay for the whole sanitation implementation (including latrine, flush toilet, and connection). For demand analysis, this variation had no implications, in view that, in discrete choice models, the alternative choice is related to the choice made, and not for all choices – McFadden, 1974.

5.3.2. Perception about sanitation

5.3.2.1. Non-adopters

As discussed in the Chapter 4, *perception* was characterised as a mental representation of an idea or image related to a good or service and therefore associated to it. To capture this idea or image, the variable “perception” was evaluated through the qualities (non-qualities) that were realised by the individual and related to *latrines*.

All 71 non-adopter households had had frequent contact with a latrine. They had used latrines in their job, public toilets, and relatives’ house. Only one household had not a latrine inside home, and practised open defecation. They were asked to evaluate the qualities (non-qualities) they appreciated in their usual domestic practice (open defecation) and latrines.

When asked about the qualities (or non-qualities) they appreciated in “open defecation”, their “qualities” rates were very low. In general, they considered open defecation dirty, dangerous, smelly, not good for health, inconvenient, non-private, and un-suitable. Medium-high rates were given to access and usefulness, 2.57 and 2.92, respectively – rates ranged from 1 to 5 (see Table 5.2). All differences were significant at $p < 0.05$.

Table 5.2. Perception of qualities (or non-qualities) of open defecation and latrines among a sample of non-adopter households in Salvador/Brazil^a

Quality	"Open defecation"		"Latrine"		Difference (between means) ⁺
	Mean	Std. Dev. [^]	Mean	Std. Dev. [^]	
Clean	1.86	1.231	4.76	0.970	2.9**
Safe	1.50	1.092	3.29	1.724	1.79**
Easy access	2.57	1.342	4.29	1.404	1.72**
Not smelly	1.50	1.160	3.06	1.819	1.56**
Useful	2.92	1.847	5.00	0.000	2.08***
Healthy	1.21	0.579	4.76	0.664	3.55**
Convenient	1.29	0.726	4.82	0.529	3.53***
Private	1.57	1.453	4.59	0.939	3.02***
Suitable	1.57	1.284	5.00	0.000	3.43***

^a Scores ranged from 1 (most negative) to 5 (most positive) for each quality; N= 71;

[^] Std. Dev. = standard variation; ⁺ Two-tailed significance of two independent t-test of difference in mean scores: * (p<0.05), ** (p< 0.005), and (p< 0.0005).

Why open defecation is considered as "easy access" needs an explanation. When the questionnaire was described, it was assumed that the population used bush areas for defecation. In reality, it is a rare behaviour in Salvador³¹. Individuals opt for defecation at home or nearby home (at the backyard). Usually they put their faeces in a bag or paper and throw it away, or they have a hole at the backyard, protected from external vision by wood or cardboard pieces. So, when households were evaluating open defecation, they were evaluating these options, not bushes.

And why they did not use bushes for defecation is essentially linked to social pressure for environmental protection: most of the "green areas" were protected areas, in spite of crescent invasion for house construction ("favelas": shantytowns). In addition, the visual dirt is combated by

³¹ Homeless people in Salvador used to do their faeces in the streets and parks/squares. Because of this, many parks and squares are now closed during the night. Also, there are several campaigns to re-socialise these people, and give access to them to public toilets. However, as in many developing countries, these actions are very restricted if poverty remains the central problem.

the whole society in that city, as previously discussed in the Chapter 3. The social pressure for environmental protection was an important (or maybe the most important) reason to explain why people opted for private defecation (open defecation practiced at home or nearby). Furthermore, the possibility of violence (attacks and robberies), and of being attacked by animals (like rats and snakes) were also additional factors that favoured “open defecation” at home or near the home.

It is interesting to observe that the variable “health” received the smallest score in evaluating open defecation (1.21), almost the minimum score in the list. The standard deviation (difference of opinion among households about health) was the smallest among other variables to assess open defecation. It indicated that this is one of the variables of higher consensus among households, and that open defecation was considered bad for the health. As we will see in the following paragraphs, health was an important reason when people have considered adoption of sanitation, contrasting with results found in rural Benin (Jenkins, 1999).

Comparing the household perception of open defecation and latrine, we can see that all nine qualities for latrine were substantially more favourable than those toward open defecation, with highest scores for suitability and usefulness (1st high qualities), convenience (2nd), and cleanliness and health (both 3rd). These results were similar to Jenkins’ study, except for those related to health, which was scored as one of last qualities in rural Benin. Smells coming from latrines was a disadvantage, and had the smallest quality rate (3.06), just after security (3.29) among non-adopters.

5.3.2.2. Difference between adopters and non-adopters

Concerning latrines, the most highly rated qualities among adopters were usefulness, suitability, convenience, accessibility, and privacy. Usefulness presented the highest score among qualities of latrines and the smallest standard deviation for the answers, among adopters. Hence, it was the most consistent perceived quality. Health, cleanliness, and smell were in the middle. Safety

was given a small score (2.93), but it was above the medium score of 2.5. It could be related to the low quality of ceramics of some latrines and the bad quality of the seat that could break when someone sat down. These points were made in some spontaneous answers about latrine disadvantages. However, this point needs more investigation for a consistent conclusion.

Adopters perceived latrines as less clean, healthy, private, suitable, safe, convenient, and smelly than non-adopters. When compared with adopters, cleanliness was a quality that presented a higher difference in opinions between the two groups, followed by healthy. The perception of latrine was more favourable among non-adopters than adopters, with significant differences between means for cleanliness, healthy and privacy.

Table 5.3. Perception of qualities (or non-qualities) of latrines among a sample of adopter and non-adopter households in Salvador/Brazil^a

Quality	Adopters		Non-Adopters		Difference (between means) ⁺
	Mean	Std. Dev. [^]	Mean	Std. Dev. [^]	
Clean	3.39	1.434	4.76	0.970	-1.37**
Safe	2.93	1.549	3.29	1.724	-0.36***
Easy access	4.35	1.067	4.29	1.404	-0.06***
Not smelly	3.38	1.354	3.06	1.819	0.32***
Useful	4.74	0.755	5.00	0.000	-0.26***
Healthy	3.94	1.323	4.76	0.664	-0.82**
Convenient	4.48	1.040	4.82	0.529	-0.34***
Private	4.06	1.317	4.59	0.939	-0.53***
Suitable	4.58	0.915	5.00	0.000	-0.42***

^a Scores range from 1 (most negative) to 5 (most positive) for each quality; N= 721;

[^] Std. Dev. = standard variation.

⁺ Two-tailed significance of two independent t-test of difference in mean scores: * (p<0.05), ** (p< 0.005), and (p< 0.0005).

While perceptions of these qualities may seem inconsistent with the expectation that latrines were better ranked among adopters than non-adopters, it would be false to assume that non-adopters perceived latrines more accurately than adopters. In fact, these results could be pointing to the inverse direction. Problems and costs with maintenance were one of the most frequent problems indicated by the households, and these problems could be associated to the lower scores given by adopters to latrines. It is not uncommon that individuals score services or goods that they owned less than those individuals who have not own one because only the experience with these services or goods permits a better assessment.

5.3.3. Attitudes toward sanitation

5.3.3.1. Non-adopters

Evaluating non-adopter's attitude towards latrines, we asked about the advantages of availability of a latrine at home. The main results pointed to increasing market value of the home, avoiding dangers going outside home for defecation (especially during the night), and security (related to violence). Health and other factors also scored well (3.71). Whereas in rural Benin people saw sanitation as a social status, in Salvador social status (prestige and social image in front of neighbours and relatives) had relatively less importance and presented the highest variation in opinions (standard deviation) – Table 5.4.

Table 5.4. Attitudes toward latrines among a sample of non-adopter households in Salvador/Brazil^a

Advantages	Mean	Standard Deviation
Gain prestige	3.53	0.624
Increasing market value of home	3.82	0.393
Easy access	3.71	0.470
Health protection	3.71	0.470
Make my house more modern	3.71	0.470
Make my social image better	3.53	0.514
Avoiding dangers	3.81	0.403
Privacy	3.71	0.470
Security	3.76	0.437
Make my house more comfortable	3.71	0.470

^a Scores range from 2 (most negative) to 4 (most positive) for each advantage; N= 71

Undoubtedly there is a correlation between elements that composed “perception” and “attitude” latent constructs, as for instance, the variables access, health, privacy, and security, are components for both constructs. This correlation was tested in the Chapter 7, when evaluating latent variables.

5.3.3.2. Difference between adopters and non-adopters

Attitudes toward sanitation had less difference between means when comparing adopters and non-adopters. Among adopters, health protection and avoiding dangers were the most well-scored factors followed by easy access, privacy, security, comfort and house monetary valorisation, according to the data presented in the Table 5.5.

Table 5.5. Attitude related to latrine adoption (advantages) among a sample of adopters and non-adopter households in Salvador/Brazil^a

Advantages	Adopter ^{&}		Non-Adopter		Difference (between means) ⁺
	Mean	Std. Dev. [^]	Mean	Std. Dev. [^]	
Gain prestige	3.39	0.639	3.53	0.624	-0.14***
House monetary valorisation	3.55	0.553	3.82	0.393	-0.27***
Easy access	3.64	0.502	3.71	0.470	-0.07***
Health protection	3.70	0.487	3.71	0.470	-0.01***
Make my house more modern	3.48	0.583	3.71	0.470	-0.23***
Make my social image better	3.43	0.642	3.53	0.514	-0.10***
Avoid dangers	3.70	0.485	3.81	0.403	-0.11***
Privacy	3.62	0.516	3.71	0.470	-0.09***
Security	3.62	0.508	3.76	0.437	-0.14***
Make my house more comfortable	3.58	0.531	3.71	0.470	-0.13***

^a Scores range from 2 (most negative) to 4 (most positive) for each advantage; [&]N= 718; Non-adopters = 71.

[^] Std. Dev. = standard variation. ⁺ Two-tailed significance of two independent t-test of difference in mean scores: * (p<0.05), ** (p< 0.005), and (p< 0.0005).

Health was rated better in the evaluation of attitudes toward sanitation than in the perceptions of this service, in spite of having had a good score in the perception assessment. In Chapter 4, I hypothesised that attitudes reflect individuals' needs, values, tastes, and capabilities, and are affected by experience, and socio-economic and demographic characteristics. If attitude is associated with the individuals, and perception with the alternative, it is expected that *health* will be well scored or ranked when it is linked to the individual than to the alternative. The individual (or household) suffers the direct effect of the disease, therefore, the aspect health is better understood when associated to the household, even though this is recognised as an important quality or benefit when choosing or adopting an alternative.

As well as the assessment of perception, the attitude towards sanitation was higher evaluated among non-adopters than among adopters. However, the difference of opinion was much lower and uniform than in the perception appraisal.

5.3.4. Non-adopters: preferences, intention for adoption, drives and constraints for installation

From 71 non-adopters all but one stated they preferred to use a latrine instead of open defecation. Among them, 69 preferred a latrine connected to a sewage system. Similar results were found when they were questioned about their intention to adopt a latrine. Preference and intention for 70 households was for latrine (with flush toilet) connected to Bahia Azul's programme.

According to Jenkins (1999, page 212), *"...preference is formed from attitude jointly shaped by drives to choose latrines and perceptions of factors of a psycho-social nature. Once preference exists, implementation factors typically intervene to affect intention through individual's perceptions of how these factors impede or facilitate their ability to adopt. For example, a person may have a strong desire but little intention to install a latrine because he or she perceives a lack of resources or opportunities needed to act [...]"*.

From this definition, preference anticipates intention that in turn anticipates choice. However, here we are talking not about economic preferences, as discusses in the Chapter 3, but about preferences as general term. The findings in Salvador suggested that preference coincides with intention. 62 households stated they preferred *and* had an intention to install a latrine. They intended to do so between 2004 (year of data collection) and 2006. However, statement of preferences and intention without confrontation with budget and other constraints is a weak indicator of demand. For this reason, an approach establishing a causal relationship among the variables, including constraint variables, is a stronger indicator of demand. This analysis is presented in Chapter 7.

When asked to give a spontaneous answer to a question about the main reason for installing a latrine, 25 households answered "health protection" as a main reason for adoption ("to protect my children against disease", "to protect me against disease", "to protect my family against disease", "to avoid disease", "to avoid diarrhoea"). Health protection corresponded to 35% of the total. Other answers included "comfort", "hygiene", and "usefulness".

For 36 households the preference was for a latrine inside the home. However, 8 considered the possibility of sharing a latrine (collective latrine), but most of them would agree to share with relatives rather than neighbours (unless neighbours were relatives); 28 would not accept even relatives. The family's privacy seemed to be an important aspect among non-adopters.

Regarding social pressure to adopt a latrine, 67 individuals answered that their neighbours insisted that they adopt a latrine, and when they were asked if they thought that if they adopt a latrine their neighbours would be jealous about their better situation (comparing with those ones that did not have a latrine), 52 individuals gave a negative answer. These negative answers reinforce the relative non-importance of "prestige" and "image in front of neighbours" as important factors determining adoption of latrines in Salvador.

Lack of financial resources was the main reason for non-adoption (69 answers related to financial resources). Extreme poverty seemed to be crucial for non-adoption, in spite of recognition of the sanitation benefits. The necessity to survive, acquiring basic goods and services day by day, such as food and transport to go to work, were more important than acquiring a latrine (and a connection). Extreme poverty was clear when some households ironically stated that "they would need to sell their home to buy a latrine" or "I and my family need to eat".

However, it is necessary to differentiate poverty from extreme poverty. Many people in poverty have access to the minimum food for surviving. In Brazil, many people considered poor have access to goods such as television, DVD, and some basic infrastructure (IBGE, 2004).

Extremely poor people most of the time have very few meals during the day, and have to acquire food day by day. Thus, it is understandable that sanitation was not a priority. Nonetheless, it does not mean that they will not opt for this, regardless of their minimum financial resources.

Here, it is important to point out the question of competing alternatives and definition of the universal choice set in evaluating demand. Of course food, rent, transport, and clothes are important basic competing alternatives in analysing consumer demand for any good or service. Nevertheless, these goods or services should not be considered in a universal choice set when evaluating a demand for sanitation (or travel choice, or telecommuting, or other intervention, good, or service). Basic goods and services are what guarantee the survival and reproduction of humans. People will not opt for not eating every day to buy a latrine. It is possible that they will opt to walk to work, and save some money from transport to acquire a latrine, but, if the work is not in walking-distance, the benefit-cost will be considered. Many poor people walk to work, even if it is not walking distance, but because they have no money at all to take transport. Money, most of the time, is allocated to food.

In analysing a demand for some good or service, only the competing alternative regarding that good or service should be considered, otherwise, if everything depends on everything, no analysis will be possible.

Alongside lack of financial resources, the absence of space to install a latrine was one of the most frequent answers (47%).

Interestingly, there was a good understanding about latrine construction and materials used in construction: 65% stated they know how to construct the latrine connection, 70% that some relative who could do the work, and 76% stated that the type of soil was appropriate for installation.

In fact this result is not a surprise, as in urban centres people have access to materials, and access to more information. Also, in Brazil the household and their relatives commonly conduct the civil construction, in the poor strata³².

Questioned about the willingness to pay for sanitation, only 29% (21 individuals) gave a positive answer. The minimum payment was US\$ 0.85 (in 2004 current US\$) and the maximum was US\$ 340. The median payment was US\$ 135.25 (standard deviation: US\$ 270.83). However, it was not a contingent evaluation analysis, and no other exploratory question was included.

5.3.5. Adopters: history of latrine (and connection) installation and disadvantages of adoption

Around 45% of the total adopters stated they had paid someone to install the latrine. For 29% of the total, the latrine already existed when they moved into the house, however, they had to pay for a connection to Bahia Azul's programme. Another 17% made the installation themselves, and for 9% it was installed by a relative or a friend without payment for the labour.

Days of work were lost by 29% of households due to supervision or direct participation in the latrine installation. They lost a median of 2 days of work and US\$ 2.55 per day. It corresponded to 0.13% of household total income in a year³³.

The use of latrines by relatives was confirmed by 17% of households; but only 1% stated that neighbours also used their latrine.

Households had good knowledge about latrine installation: 53% knew about installation and had enough space to install a water closet and most of them also included a shower, 59% knew the correct material to use for installation, and 23% had to contract a specialist.

³² "Construction master" is a popular profession for many unqualified individuals. In spite of low qualification level, in this kind of profession, the years of experience is the most important quality.

³³ It was calculated as: US\$ 2.55 (loss of income) divided by annual total household income (US\$ 157.94 multiplied by 12 months = US\$ 1,895.28).

Households spontaneously cited several advantages. However, the most frequent were usefulness (14%), hygiene (9%), cleanliness (7%), practical, health benefits, and security (6% each one). Among the disadvantages, most frequent were bad smells (5%) and blocked sewage (4%). However, when exploring more the disadvantages of latrine adoption through structured questions the results were scored as (in order of importance): “it was cause of accidents”, “it had bad smells” and “have to pay for Bahia Azul’s connection”, “more work at home” and “have to pay for a septic tank”, “increased the number of visits (neighbours’ inconvenience) to my home”, “increased the number of visits of relatives to my home” – See Table 5.6.

Table 5.6. Disadvantages related to latrine adoption among a sample of adopter households in Salvador/Brazil^a

Disadvantages	Mean	Standard Deviation
More work at home	2.67	0.840
Bad smells	3.00	0.935
Cause accidents	3.39	0.850
Inconvenient’ neighbours	2.63	0.885
Increase the number of visits of relatives	2.50	0.816
Have to pay for a septic tank	2.67	0.767
Have to pay for Bahia Azul’s connection	3.00	0.840

^a Scores range from 2 (most negative) to 4 (most positive) for each advantage; N= 718;

The questionnaire had an assertion to be evaluated by the household that stated the following: “we prefer open defecation”. This statement should be rated from 2 (most negative) to 4 (most positive), similar to Jenkins’ (1999) questionnaire. Nevertheless, the majority of households mentioned to the research assistants that “*this was a stupid question*” and “*who will prefer open*

defecation to latrine?” Some households, however, scored this expression and the mean was 2.33 (0.594 standard deviation). It was one of the lowest score among the disadvantages evaluated.

5.4. Summary of results

In general, non-adopters scored perception and attitude related to sanitation higher than adopters. These results, however, did not indicate a better non-adopters' understanding about latrines/sanitation than adopters. As discussed before, adopters' answers could be influenced by the problems with maintenance of latrines and connections, whereas non-adopters had not had the good and service yet, and so, had not had those kinds of problems.

Health protection was an important quality linked to latrine adoption, and it was in consonance with the health agencies' promotion. The households also appreciated other qualities and advantages. Adopters and non-adopters perceived usefulness, suitability, and convenience as the three most important qualities of a latrine. Advantages were differently perceived between groups: while adopters scored better health protection, avoiding dangers of going out for open defecation, and easy access, non-adopters rated increasing market value of home, avoiding dangers of going out for defecation and security.

Latrine with flush toilet connected to Bahia Azul's programme was the preferred choice of non-adopters, and costs for construction, installation and connection were the main constraint for adoption. In spite of the advantages and qualities of a latrine, adopters indicated several problems related to latrine/connection adoption: bad smells, frequent problems with blocked sewage, accidents, and connection costs.

The results of the household survey pointed out a series of different factors influencing sanitation choice. In the demand analysis, I will use only some of these factors: perception and attitude towards sanitation. However, alongside these two latent variables, socio-economic and demographic characteristics of households and cost of alternatives, are other variables that might be influencing choice.

Chapter 6: Methodology for Estimation of the Demand for Sanitation in Salvador/Brazil

6.1. Introduction

Results from a household survey of sanitation adoption in Salvador were presented in Chapter 5. From those results, we explored perceptions, beliefs and attitudes, as well as the historical and cultural context that underlined household behaviour in their choice of sanitation. Analysis was essentially descriptive, which provided us a quantitative general vision of the data: cleaning (identifying missing data, duplicated, number of valid cases, etc.), comparative analysis between groups, and mean value of variables. It was also a picture of the sanitation conditions of households in the specific areas where the programme Bahia Azul was operating, between 2004 and 2005.

In this Chapter, I present the methodology for estimation of the demand for sanitation in Salvador/Brazil. The focus is on the treatment of variables (in statistics terms), and specification and determination of the integrated model to estimate demand: the Hybrid Choice Model, defined as a latent variable model integrated with a discrete choice approach, the mixed logit. The latent variable model included the specification and determination of perception and attitude latent constructs that were defined theoretically in the Chapter 4. Here, the causal relationship among these variables, the socio-economic and demographic characteristics of individuals, and alternative attributes are statistically tested and demonstrated, as well as their integration to the discrete choice model. Hence, I expect I have contributed to filling the gap between behavioural theory and discrete choice model in sanitation literature. A flow chart is shown in the summary section at the end of this chapter to show a clear visualisation of the steps in this methodology.

6.2. Study Design

6.2.1. Overview

This study was designed as an econometric model to estimate demand for sanitation in Salvador/Brazil. The approach used is the Hybrid Choice Model, which is an integrated framework that permits the incorporation of latent variables in an extended discrete choice model with a flexible error structure. Empirical effort was not to produce an explanation for the construction of latent variable (e.g., to evaluate its socio-psychological determinant, to understand and to explain the causal relationship among latent variables, etc.), although statistical analysis was applied to understand how the psychometric indicators form the latent variables. The objective was to illustrate and test the application of the theory, to understand how the selected latent variables affected the decision-making process, and to produce numerical estimates of model parameters to forecast the demand for sanitation.

The behavioural framework in socio-psychology is complex, with extensive connections between behavioural states and constructions. The aim of socio-psychologists and psychologists is to understand the deconstruction process of decision-making rather than forecast individual's choice, which is the objective of economic models.

The objective of the Hybrid Choice Model is to fill the gap between behavioural theory and discrete choice models. Incorporation of psychological factors leads to a more realistic behavioural representation of the choice process, and gives more explanatory power to choice models (Morikawa et al., 1996; Polydoropoulou, 1997). Nevertheless, mathematical procedures to incorporate and capture the causal process among variables is elaborate, and the inclusion of a large number of variables can make the model non-interpretable

Jenkins (1999) demonstrated that in rural Benin several psychometric indicators were associated with household behaviour towards sanitation. However, no causal relationship amongst

variables was established and, as discussed in the Chapter 2, without the establishment of this causal relationship, the model is only replicating the results of conditions existing at the time of the survey.

In keeping these tasks in mind, I have opted to include only two latent variables in the Hybrid model for sanitation: attitude and perception. Usually, these two latent variables are better explored in psychological studies. Causal relationships between latent variables and socio-economic characteristics of individuals and alternatives attributes were tested.

Several other latent and observed variables can influence behaviour. On the other hand, there is no guarantee that a model with several latent variables would be a more appropriated one for forecasting demand for sanitation. If the causal relationship among these variables cannot be demonstrated, this complexity would generate only a difficult interpretation of the parameters rather than a useful tool for policy planning. However, explorations should be carried out.

6.2.2. Objectives

The main objective of demand estimation was to test and demonstrate the causal relationship among the variables, establishing the behavioural link amongst the attributes of sanitation system, individuals socio-economic and demographic characteristics, and the choice made. An additional objective is to capture through an econometric model the influence of a latent variable on choice. The secondary objectives were the estimation of the price elasticity and the probability of choice for the different sanitation systems in the city of Salvador/Brazil.

6.3. Plan of analysis

I opted for the sequential analysis of the Hybrid model described in Morikawa et al. (1996). As described in section 3.4 (Chapter 3), the Hybrid model is composed of two parts: the latent model and the discrete model. They can be integrated using (1) a sequential estimation, or (2) a simultaneous estimation. Simultaneous estimation is preferred to the sequential one. Simultaneous estimation produces consistent and effective estimated parameters³⁴. Nonetheless, it is an involved process where simultaneous integrals are used. The more variables the researcher uses, the more complex is the identification and estimation in terms of computational modelling.

Before I opted for the sequential estimation, I discussed with experts in discrete choice models the process of estimation to understand how the computational modelling could be constructed - including contact, without success, with some authors who worked with this simultaneous method. My contacts also included statisticians specialised in structural equations and modelling. As the process of estimation was not clear for me, I opted for sequential analysis. Certainly, simultaneous analysis will be a challenge for my future research.

Sequential estimation produces consistent, but inefficient parameters. Inefficient parameters do not mean that the model does not give robust answers for the question posted. The term only means that coefficients could have the *minimum variance* if they were estimated simultaneously. Most important, however, is the consistency of parameters, *“that is, the values of the parameters should be stable. Otherwise, forecasting will be difficult”* (Gujarati, 2003, page 507).

Sequential estimation is the more popular method used in discrete choice studies, due to ease of modelling. I opted for the MIMIC latent variable model integrated with the mixed logit approach, both of which are described below.

³⁴ An efficient estimator is unbiased and has smaller variance (Ben-Akiva and Lerman, 1985; Gujarati, 2003).

6.3.1. Multinomial logit model

It is recommended that before using any advanced discrete choice model, a multinomial logit should be applied as a starting point in the analysis. It helps “*to ensure that the data are clean and that sensible results (e.g. parameter signs and significance) can be obtained from models that are not ‘cluttered’ with complex relationships*” (Hensher and Green, 2001, page 2). Moreover, a specification test based solely on multinomial logit provides a good evaluation for a presence of multicollinearity among the regressors (McFadden and Train, 2000; Gujarati, 2003).

The multinomial logit model was described in the section 3.3 in Chapter 3. The specification and identification of this standard model will not be presented here. For more details, see McFadden (1974), Maddala(1983), Liao (1994), and Louviere et al. (2000).

Due to its characteristics, latent variables were not included in this analysis. The checking of latent variable parameters was done using the MIMIC model. Multinomial logit was only used to check the alternative attributes and household characteristics associated with the revealed choice.

Dependent and independent variables are presented in the Table 6.1. Variable “Bazul” was the variable of reference in the analysis.

The software used to run the model was STATA version 9.0 (STATA Corporation, 2007). It is important to point out that the command in this software refers to the “*clogit*”, which runs the multinomial logit McFadden’s choice model, the conditional logit. The “*mlogit*” command refers to the traditional multinomial logit, which did not includes choice-specific characteristics (STATA reference manual release 8, 2003; McFadden, 1974; Liao, 1994).

Table 6.1. Dependent and independent variables used in the multinomial logit analysis

Dependent Variables	Description
BAZUL (y_1)	Latrine connected to Bahia Azul's programme (reference variable)
NONE (y_2)	No toilet facility and connection at all.
SEPTIC (y_3)	Latrine linked to septic tanks or disposal to drainage under public stairs
OTHER (y_4)	Other type of connection (mainly to the town hall service)
Independent Variables	Description
FEMALE (x_1)	Dummy = 1 for female; 0 for male
AGE (x_2)	In years
INCOME (x_3)	In Brazilian currency "Real", 2004
NCHILDREN (x_4)	Number of children in the house
EDUCATION (x_5)	Number of years that households have studied
COST (x_6)	Cost of sanitation by each household (conditioned to the choice); In Brazilian currency "Real"

Note: symbols in parenthesis will be the same as those used in the equations for determination of the model

6.3.1.1. Goodness of fit

A measure of goodness of fit shows the proportion of total sample variation in the dependent variable which is due to sample variations in all independent variables of the model. Or, in other words, it shows how "well" the sample regression line fits the data. The most well-known measure of goodness of fit is the coefficient of determination R^2 . However, in models with a qualitative dependent variable the use of the coefficient of determination should be avoided, because R^2 has limited value in these models (Gujarati, 2003). In this case, measures similar to R^2 are used. These measures are known as pseudo- R^2 , and there are a variety of them. For multinomial logit analysis I

used the McFadden R^2 (1978). It is denoted as R^2_{McF} , and ranges between 0 (bad fit) and 1 (excellent fit), like R^2 . It is calculated as:

$$R^2_{McF} = 1 - [LL(\eta, \beta) / LL(\eta)] \quad (6.1)$$

Where, LL is the Log Likelihood, and η , β are the coefficients that were estimated.

Nevertheless, as observed by Gujarati (2003, p.606), “*in binary regressand models (and also, in multi-response models – my note), goodness of fit is of secondary importance. What matters is the expected signs of the regression coefficients and their statistical and/or practical significance*”.

Together with McFadden is R^2 test, the likelihood ratio (LR) statistic should be carried out to test the null hypothesis that all the slope coefficients are simultaneously equal to zero. The LR statistic is an equivalent of the F test in the linear regression and, given the null hypothesis, it follows the Chi-square distribution with degrees of freedom equal to the number of explanatory variables. For this reason, some books present the test as a “LR chi2” followed by the P -value ($\text{prob} > \text{chi2}$) – Gujarati, 2003.

Hence, the LR statistic is distributed chi-square with i degrees of freedom, where i is the number of independent variables, and it is calculated as (Huelsenbeck and Crandall, 1997):

$$LR(i) = -2[LL(\eta) - LL(\eta, \beta)] \quad (6.2)$$

Furthermore, the P -value was applied to assess the coefficients.

6.3.1.2. Testing the independence from irrelevant alternatives (IIA) property

As also was discussed in section 3.3 in the Chapter 3, multinomial logit models exhibit the property of independence from irrelevant alternatives (IIA), which states that the ratio of two choice probabilities is independent of the other alternatives in the model, or that the inclusion or exclusion of alternatives does not affect the ratio of the probabilities of the choice set (Green, 2003).

To verify this possibility, Hausman and McFadden (1984) developed a test to examine the IIA property. The basic idea is that, where IIA holds, the two sets of estimates should not be statistically different. Where there is independence, the inclusion of irrelevant alternatives will cause inefficiency, but not inconsistency. In the case of dependence, the omission of some alternatives will lead to inconsistency (Anonymous, 2000b; Louviere et al., 2000; Persson, 2002). The test is defined as (Anonymous, 2000b):

$$\chi^2 = (\beta_a - \beta_b)'(V_a - V_b)^{-1}(\beta_a - \beta_b) \quad (6.3)$$

Where a denotes the restricted model, and b the unrestricted one. β_a are the estimates obtained from the restricted model and β_b , the estimates of the same parameter (from an unrestricted model). V_a and V_b are the estimated asymptotic covariance matrices.

The LR chi2 is also presented in this calculation, and where IIA holds, this statistic has a chi-square distribution with one degree of freedom (Anonymous, 2000b).

6.3.2. The latent variable model

6.3.2.1. Psychometric indicators and latent variables

19 psychometric indicators were used to construct the latent variables “attitude” (10 indicators) and “perception” (9 indicators). This approach followed the same structure as that in Jenkins’ questionnaire. The indicators and their respective latent variable are presented in the Table 6.2.

Table 6.2. Indicator Variables and Scales for “Attitude” and “Perception” Latent Variables*.

Attitudinal Latent Variable (advantages and disadvantages; scale 2 to 4; 2 = less important and 4 = very important)	Perception Latent Variable (qualities and non-qualities; scale 1 to 5; 1 = little and 5 = very)
Gain prestige from visitors (z_1)	It is clean (z_{11})
Increasing market value of home (z_2)	It is dangerous (z_{12})
Make it easy to defecate (z_3)	It is of difficult access (z_{13})
Protect from disease (z_4)	It has bad smelling (z_{14})
Make my house more modern (z_5)	It is useful (z_{15})
Turns my “image” better to my neighbours (z_6)	It is good for health (z_{16})
Avoid dangers (z_7)	It is convenient (z_{17})
It is private (z_8)	It is private (z_8)
It is safe (z_9)	It is adequate (z_{19})
It is comfortable (z_{10})	

*See more details about the variables in Jenkins (1999)

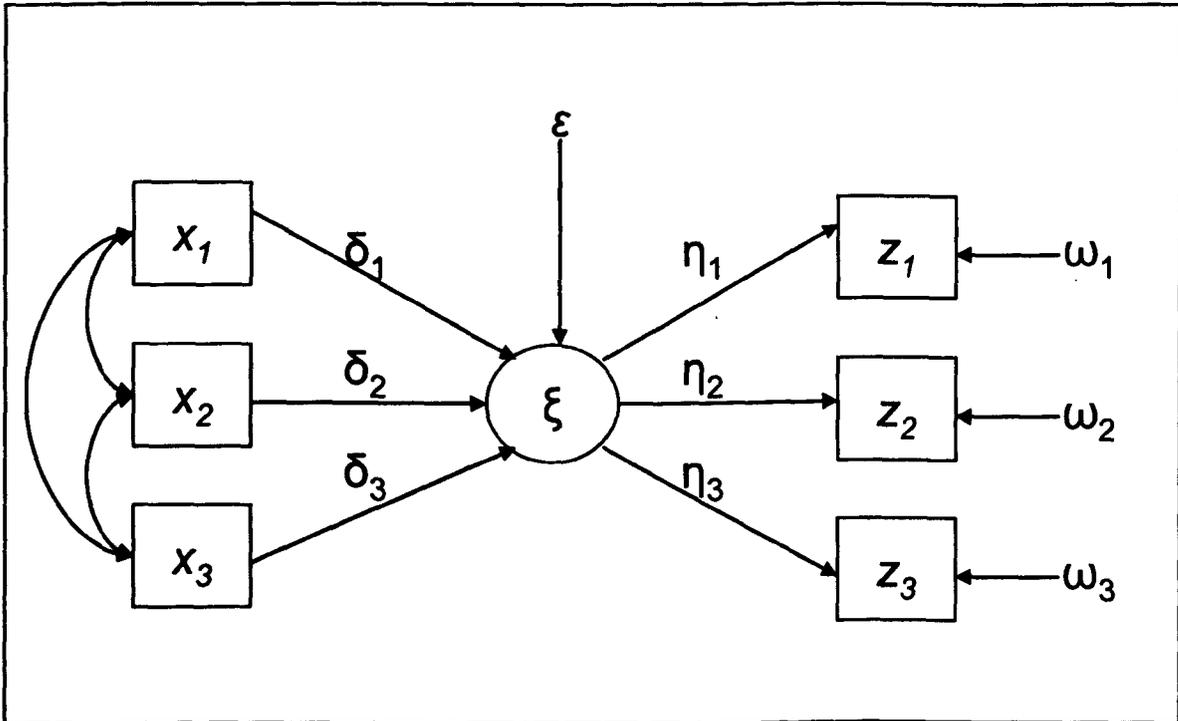
Note: symbols in parenthesis will be the same used in the equations for determination of the model

6.3.2.2. The multiple indicators, multiple causes (MIMIC) model: a general definition³⁵

The MIMIC model for a single latent variable (attitude or perception) involves a single hypothetical variable ξ , which appears as both cause and effect. The model is represented in the Figure 6.1. This model is applied for each latent variable.

³⁵ Part of LISREL methodology (see Joreskog, 1973; Joreskog and Sorbom, 1981)

Figure 6.1. A Model with Multiple Indicators and Multiple Causes for Construction of a Latent Variable



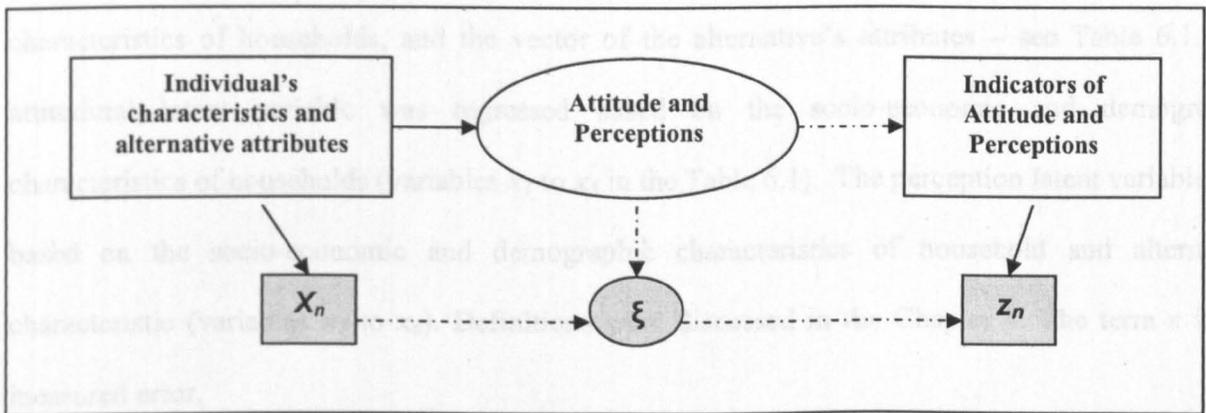
Source: Joreskog and Sorbom (1979)

MIMIC model is a structural analysis of covariance matrices, which is a general method for analysing measurements, in order to detect and assess latent sources of variation. This method will perform the “latent model” part of the Hybrid choice model to evaluate demand for sanitation. The multiple indicator part of the “MI” is executed using exploratory and confirmatory factor analysis. In the Chapter 2, I discussed the implications of using factor analysis for exploratory assessment. But, in general, I used factor analysis because my purpose was to identify the latent variable which was contributing to the common variance in a set of measured variables. In this case of identification, factor analysis would produce a better answer (Joreskog and Sorbom, 1979).

The model was used to produce the variables shown in Figure 3.3 (Chapter 3). Reproducing the Figure 3.3, and linking that with the Figure 6.1, we establish the first causal relationship: among latent variables and socio economic/demographic characteristics of individuals, and alternative

attributes. This, however, did not correspond to the final analysis of demand, which involves the estimation of utilities for each revealed choice of sanitation. For this first part of sequential analysis, we can have the following link between the framework of analysis developed in the Chapter 4 and the MIMIC model, summarised in Figure 6.2.

Figure 6.2. Linking Behavioural Framework for Choice of Sanitation in Salvador Brazil with MIMIC model



6.3.2.3. Specification of the latent model

The measurement equation which will assess the indicators that influence the formation of the latent variable was executed using exploratory and confirmatory factor analysis. The most important part in this analysis is the confirmatory factor analysis. According to Jackson (2003), it is the main reason for the existence of factor analysis.

Confirmatory factor analysis restricts the number of factors extracted to a particular number and specifies particular patterns of relationship between measured variables and common factors. This relationship is shown in the equation (6.4) and Figure (6.1, right side).

$$z = \eta\xi + \omega \quad (6.4)$$

Where z is the vector of 19 observable indicator variables of ξ (10 to “attitude” and 9 to “perception”, see Table 6.2), η is the matrix of unknown parameter to be estimated, ξ is a vector of individual specific latent variables (attitude, perception), and ω is the error. This equation (6.4) corresponds to the part “MI” in the MIMIC model.

The model to assess the causal part of the structural equation is:

$$\xi = \delta x + \varepsilon \quad (6.5)$$

Where δ is the parameter to be estimated, x is the vector of socio-economic and demographic characteristics of households, and the vector of the alternative’s attributes – see Table 6.1. The attitudinal latent variable was regressed based on the socio-economic and demographic characteristics of households (variables x_1 to x_5 in the Table 6.1). The perception latent variable was based on the socio-economic and demographic characteristics of household and alternative characteristic (variables x_1 to x_6). Definitions were discussed in the Chapter 4. The term ε is the measured error.

The complete MIMIC model was estimated simultaneously using *Mplus* software version 3.01 (Muthén and Muthén, 2004).

6.3.2.4. Determination of the latent model

The variables FEMALE, AGE, INCOME, NCHILDREN, EDUCATION, and COST were tested for normality using a kurtosis and skew test (Tabachnick and Fidell, 1996). All proved significant kurtosis and skew (P -value for skewness and kurtosis of 0.000). Related to categorical variables, these clearly depart from normality. Hence, the estimation of the structural equation latent variable was performed using a maximum likelihood (ML) estimator.

In general, the ML estimator minimises the difference between the *sample* covariance matrix and the covariance matrix, whose elements are hypothesised to be function of a parameter vector (for more details see Gujarati, 2003, pages 114-118).

Therefore, the determination of the MIMIC model for 19 indicators was:

From equation (6.4), the “MI” was calculated as (η is the matrix of unknown parameters with λ elements):

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \\ z_9 \\ z_{10} \\ z_{11} \\ z_{12} \\ z_{13} \\ z_{14} \\ z_{15} \\ z_{16} \\ z_{17} \\ z_{18} \\ z_{19} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ \lambda_{51} & 0 \\ \lambda_{61} & 0 \\ \lambda_{71} & 0 \\ \lambda_{81} & 0 \\ \lambda_{91} & 0 \\ \lambda_{101} & 0 \\ 0 & 1 \\ 0 & \lambda_{121} \\ 0 & \lambda_{131} \\ 0 & \lambda_{141} \\ 0 & \lambda_{151} \\ 0 & \lambda_{161} \\ 0 & \lambda_{171} \\ 0 & \lambda_{181} \\ 0 & \lambda_{191} \end{bmatrix} \begin{bmatrix} \xi_{attitude} \\ \xi_{perception} \end{bmatrix} + \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \omega_6 \\ \omega_7 \\ \omega_8 \\ \omega_9 \\ \omega_{10} \\ \omega_{11} \\ \omega_{12} \\ \omega_{13} \\ \omega_{14} \\ \omega_{15} \\ \omega_{16} \\ \omega_{17} \\ \omega_{18} \\ \omega_{19} \end{bmatrix}$$

From equation (6.5):

$$\xi = \delta x + \varepsilon$$

The MIMIC model was calculated as (δ is the matrix with τ elements):

$$\begin{bmatrix} \xi_{attitude} \\ \xi_{perception} \end{bmatrix} = \begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} & \tau_{14} & \tau_{15} & \tau_{16} \\ \tau_{21} & \tau_{22} & \tau_{23} & \tau_{24} & \tau_{25} & \tau_{26} \end{bmatrix} \begin{bmatrix} FEMALE \\ AGE \\ INCOME \\ NCHILDREN \\ EDUCATION \\ COST \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

Remember that only variables FEMALE, AGE, INCOME, NCHILDREN, and EDUCATION were associated with the latent “attitude” variable. For the latent “perception” variable, all variables were associated.

The matrices above are the matrices of structural coefficients, representing the complete set of coefficients of endogenous (latent) and predetermined variables (socio-economic, demographic and choice attribute).

6.3.2.5. Goodness-of-fit and reliability coefficients in latent models

The most important test for model fit in the latent model is the Chi-square. This is a test to detect whether departures of the data from the model is dependent on the sample size. However, this test should be used together with other tests due to its limitation in assessing goodness-of-fit. Limitation corresponded to the sample size: for large samples sizes, the test very often will indicate that the model should be rejected, the converse occurs with small sample sizes (Hughes et al., 1986, page 141). To reduce or eliminate this limitation, the ratio of Chi-square to degree of freedom (and *P*-value) is estimated (Wheaton et al., 1977).

Common tests that accompany Chi-square analysis are:

- 1) Cronbach’s alpha test that assesses the reliability in the measurement of an unobserved factor (STATA reference manual release 8, 2003). Reliability is defined as the proportion of the observed variable that is free from error (Lord and Novick, 1968). Nunnaly (1978) has indicated 0.7 (it ranges from 0 to 1) to be an acceptable reliability coefficient, but some

studies in the literature use lower thresholds; also, *t*-statistic is calculated; these tests are related to the factor and their coefficients;

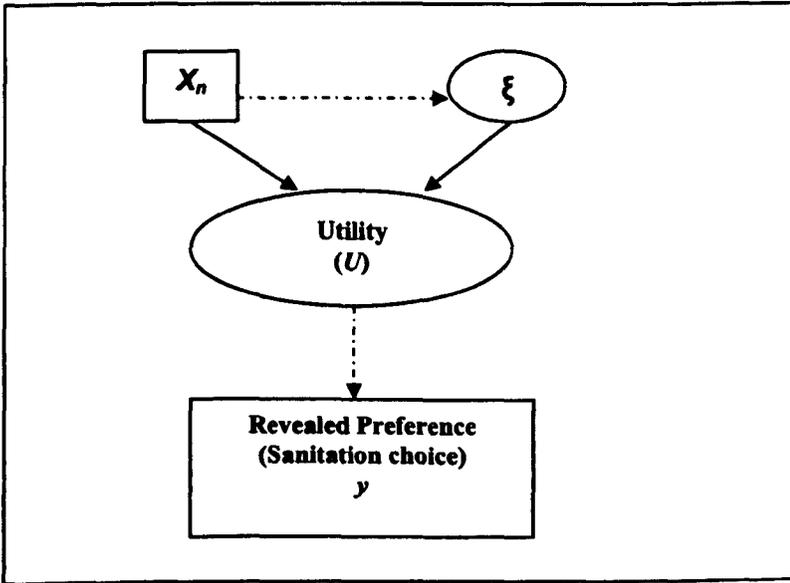
- 2) Root Mean Square Error of Approximation (RMSEA) is a measure of model fit to test the discrepancy per degree of freedom for the model. Good models have a RMSEA of 0.05 or less. Models whose RMSEA is 0.10 or more have poor fit (Browne and Cudeck, 1993).
- 3) Standardized Root Mean Square residual (SRMR). This measure is the standardized difference between the observed covariance and predicted covariance. A value of zero indicates perfect fit. This measure tends to be smaller as sample size increases and as the number of parameters in the model increases. A value less than 0.08 is considered a good fit (Kenny, 2003).

Coefficients were tested using the *z*-test.

6.3.3. The discrete choice model

To estimate the second part of the sequential model, the discrete choice model including the observed and the latent variables. Figure 6.3 outlines this final part of analysis.

Figure 6.3. Discrete choice model with observed and latent variables



Specification of the mixed logit model was presented in Chapter 3 (subsection 3.5.3). However, equations (3.23) and (3.24) are repeated here to facilitate demonstration of the discrete model. We have:

$$U_{ia} = X_{ia}\beta + F_{ia}\zeta_{ia} + v_{ia} \quad (3.23)$$

Where:

$$\varepsilon_{ia} = F_{ia}\zeta_{ia} + v_{ia} \quad (3.24)$$

6.3.3.1. Determination of discrete model

From equation (3.23), a general model including the attributes of alternatives, household characteristics and latent variables could be re-written as:

$$U_j = \mathbf{b}'\mathbf{X}_j + \mathbf{c}'_j\xi + v_j \quad (6.6)$$

$$\xi = \delta\mathbf{x} + \varepsilon \quad (\text{from 6.5})$$

$$y = \begin{cases} j & \text{if } U_j \geq U_k, \forall k \in J \\ 0, & \text{otherwise} \end{cases} \quad (6.7)$$

and,

$$z = \eta\xi + \omega \quad (\text{from 6.4})$$

Where z is a $(q \times 1)$ vector of observable indicators of ξ ; s and t_j are vectors of observable exogenous variables (t_j is mode specific, and x is household specific), ξ is a $(\ell \times 1)$ vector of individual latent variables, x is a $(k \times 1)$ vector of exogenous observable variables that causes ξ (x may or may not be a part of s), a_j , b and c_j are vectors of unknown parameters to be estimated and η and δ are, respectively, $(\ell \times k)$ and $(q \times \ell)$ matrices of unknown parameters to be estimated and $v = (v_1, \dots, v_j)$, ε and ω are measurement errors independent of s , t_j , and ξ .

From (6.6), we have:

$$\begin{bmatrix} y_{\text{BAZUL}} \\ y_{\text{NONE}} \\ y_{\text{SEPTIC}} \\ y_{\text{OTHER}} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} \end{bmatrix} \begin{bmatrix} \text{FEMALE} \\ \text{AGE} \\ \text{INCOME} \\ \text{NCHILDREN} \\ \text{EDUCATION} \\ \text{COST} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{bmatrix} \begin{bmatrix} \xi_{\text{attitude}} \\ \xi_{\text{perception}} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

The final relation in (6.7) is estimated using maximum simulated likelihood estimation with 125 Halton draws. The model was estimated in GAUSS 6.0 (Apetch, undated). Hence, the

parameters to be estimated, β (a, b, and c), and Ω , the parameters that describe the distribution of v , assumed a mixed logit log-likelihood (Train, 1999):

$$\mathcal{G}(\beta, \Omega) = \sum_I \sum_J y_{ij} \log \left[\int_{v_i} \left\{ \frac{e^{s_{ij}\beta_j + \Omega v_i}}{\sum_k e^{s_{ij}\beta_j + \Omega v_i}} \right\} g(v_i | \Omega) \partial v_i \right] \quad (6.8)$$

And the choice probabilities given the parameter vectors β and Ω are approximated by averaging over the values of $\hat{P}_r(j|v)$:

$$\hat{P}(j|\beta, \Omega) = \frac{1}{R} \sum_{r=1}^R \hat{P}_r(j|v_r) \quad (6.9)$$

$\hat{P}_r(j|\beta, \Omega)$ is the estimated choice probability of household choosing alternative j given β and Ω . This simulated choice probability is an unbiased estimator of the actual probability $P(j)$, with a variance that decreases as R increases. The bias is very low when $R = 250$ draws (Brownstone and Train, 1999). Bath (1996) and Train (1999) found that in estimating mixed logits, the simulation error in estimated parameters is lower with 100 Halton draws than with 1000 random draws. In my model I used 150 Halton draws in the estimation.

6.3.3.2. Halton draw

A Halton draw is an “intelligent” draw in a simulation method to reduce the run-times in estimating the integral in some models. Integrals have been approximated through simulation using random draws from the mixing distribution (as mixed logit) - Brownstone and Train, 1999. A large number of draws (random) is usually needed to assure reasonably low simulation error in the estimated parameters. However, a large number of draws is time demanding (Train, 1999).

Bath (1996) tested the Halton sequence for mixed logit and found that with 125 Halton draws, the simulation error was half that with 1000 random draws. Using Halton sequences in place of

random draws allows us to obtain more accurate estimates of model parameters at a fraction of the estimation cost.

6.3.2.3. Goodness-of-fit

The same tests described in the section 6.3.2.1 were used to assess reliability of coefficients and model's goodness of fit. Coefficients were tested using a *t*-test.

6.3.2.4. Marginal price (cost) elasticities

Estimating how household members respond to variations in the price of sanitation plays an important role in the implementation of sanitation programmes. The price elasticity of choice probability was defined as (Gertler and Gaag, 1990):

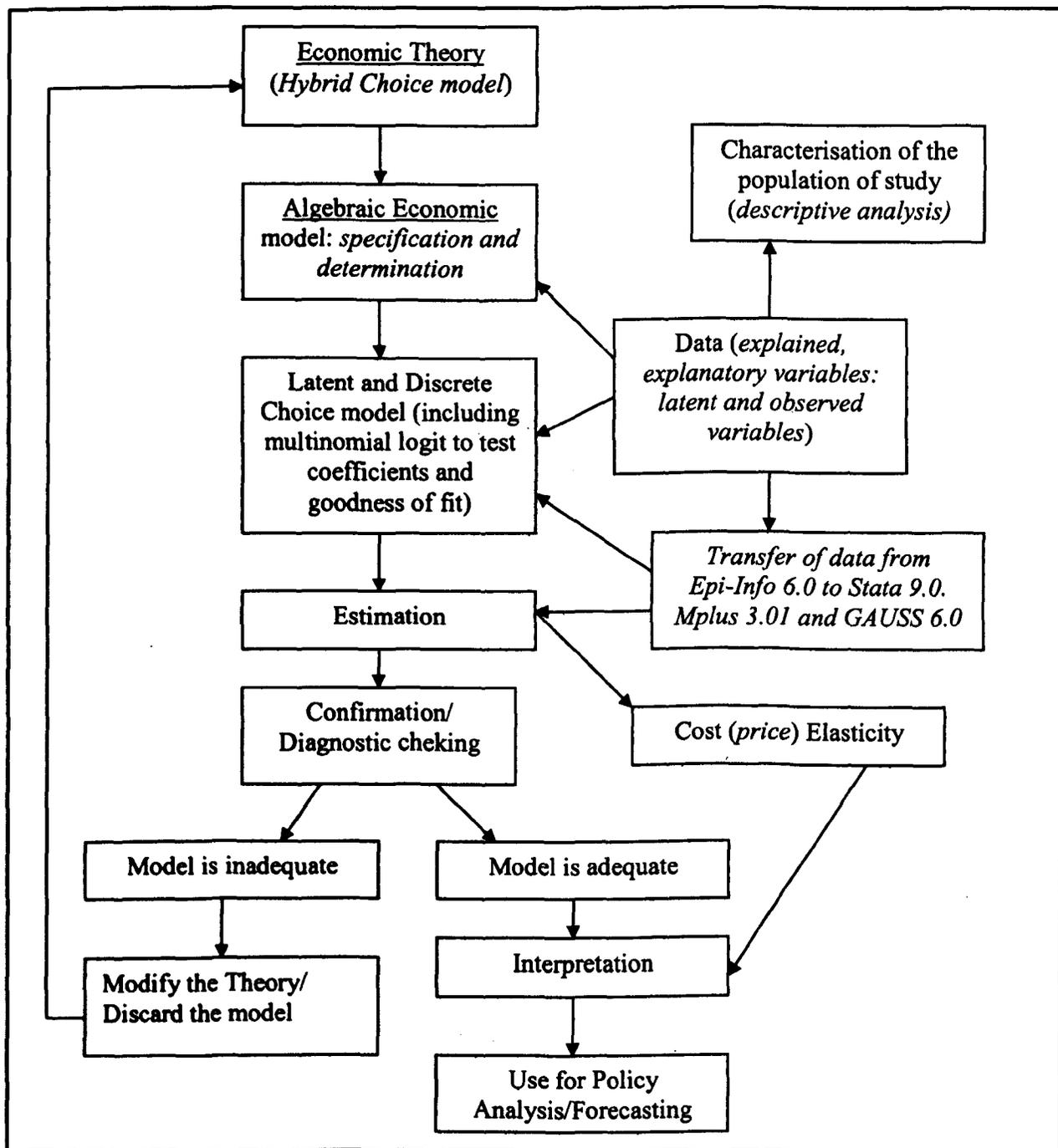
$$\text{Price elasticity}_{\text{alternative } k} = \left(\frac{\frac{Pr_1 - Pr_0}{(Pr_1 + Pr_0)/2}}{\frac{p_1 - p_0}{(p_1 + p_0)/2}} \right) \quad (6.10)$$

Where Pr_0 is the initial probability, Pr_1 is the new probability, p_0 is the initial price and p_1 is the new price. The new prices were assumed as 10% increase in the price of Bahia Azul's programme and septic tank. In view of the tendency of municipal sanitation system be substituted by Bahia Azul's program, it was not included in the analysis.

6.4. Summary of Analysis

This chapter can be summarised in accordance with Figure 6.4.

Figure 6.4. General Modelling (adapted from Seddighi et al., 2000)



Chapter 7 – Results: Estimation of Demand for Sanitation in Salvador/Brazil

7.1. Introduction

This chapter presents the results of estimation of the demand for sanitation in Salvador/Brazil. The approach used was the Hybrid choice model, in which a latent model and a discrete choice model were integrated. Alternatives were evaluated in terms of the signs and magnitudes of estimated parameters. The empirical results were presented in terms of the parameters estimated, measures of the statistical reliability of the parameter estimates (e.g., z and t -statistics), and a measure of elasticity. A discussion in the light of the sanitation literature will be presented in the next chapter.

The presentation of results follows the same structure as that presented in Chapter 6 to facilitate interpretation. Some explanations for specific subjects, not discussed in the previous chapters, were provided, when necessary.

In summary, these results indicated that the modelling approach developed in this study is an extremely promising approach giving an estimation of demand for sanitation studies. Clearly, the inclusion of latent variables as explanatory indicators enriched the model and, together with socio-economic and demographic characteristics of households and alternative attributes, allowed policy makers can obtain robust answers for planning. The causal relationship between those characteristics and household choice could be linked and established.

The Hybrid Choice Model seems to be an advance for demand analysis, especially for goods and services, which do not have a market, defined, as in the case of sanitation.

7.2. Results

7.2.1. Multinomial logit (conditional logit): interpretation and results

The model did not present a multicollinearity problem, when running a multinomial logit analysis.

The results of analysis are shown in Table 7.1. The alternative of reference was Bazul (latrine connected to Bahia Azul's programme). As described in equation (3.15), the multinomial logit is the probability that alternative "i" being chosen given the probability of alternative "k". In other words, this is the difference between the expected value of the probability of Bazul being chosen and the expected values of probabilities of other alternative (NONE, SEPTIC or OTHER).

Before starting the results discussion, it is very important to point out that, in econometrics, the sign (positive or negative) of the coefficient of a variable as well as its magnitude are the centre of the analysis. These indicate how a variable will "behave" when conditions change. For instance, when the variable cost has a negative coefficient sign, it usually means that, *ceteris paribus*³⁶, the demand for a determined good or service will decrease as cost increases.

In terms of the causal inference, the result could be interpreted as: the positive sign turns the reference alternative (*Bazul*) a counterfactual (the not observed choice); for instance, from Table 7.1 we can observe that *None* (no toilet at all) and *Septic* (latrine connected to septic tanks) have a positive sign what means that these both alternatives were preferred to *Bazul*; however, comparing with *Other* (municipal system) *Bazul* turns the preferred (and the observed) alternative, thus *Other* is the counterfactual, the not observed alternative that is indicated with a negative sign. Again, the alternative of reference for all comparisons was the *Bazul*.

From Table 7.1, we can also interpreter the preference utility maximisation axiom presented in the section 3.2. We can verify that alternatives *None* and *Septic* are preferred to *Bazul* therefore,

³⁶ "With other things being the same".

None and *Septic* > *Bazul*, and *Bazul* is preferred to *Other*, thus *Bazul* > *Other*, in terms of utility for households. In short: *None* and *Septic* > *Bazul* > *Other* then, *None* and *Septic* are preferred to *Bazul* and *Other*, in general terms (without testing by socio-economic and demographic characteristic of households). No statement can be done regarding utilities to difference *None* and *Septic*, and so ranking them properly. This will be an attempt for further research when testing multinomial logit model.

Caution is necessary for the interpretation of the results presented in the Table 7.1 as only the estimated coefficient for alternative *None* was statistically significant³⁷. In fact, few variables had coefficients that were significantly different from zero: level of education related to alternative *None* and *cost* (of all alternatives). The level of education associated with the alternative *None* suggested that when the level of education increases, households prefer *Bazul*.

An unexpected result was the sign of the variable *cost*. According to traditional demand theory, when the price of some good or service increases, it is expected that the demand for this good or service decreases, when it is a normal good. We can observe, however, that the contrary happened here, which could suggest that toilet facilities (associated with some of these connections, except *None*) were not a normal good. Persson (2002) found similar results in the Philippines when evaluating demand using a multinomial logit model. Nevertheless, when applying a mixed logit model, her results were different: sanitation “behaved” like a normal good. Similar results were obtained in my analysis (see section 7.2.3). What did this result imply? An answer for this problem is that, possibly, the true responses do not satisfy IIA-assumption, and predictions from a multinomial logit approximation can be very misleading (Green, 2000).

³⁷ In spite of the poor magnitude and significance of the coefficients, the general goodness-of-fit of the model was robust. This point will be discussed in the next chapter, and was not presented here to not break the sequence of presentation of the results

Table 7.1. Estimation Results for the Multinomial Logit Model

Variables	Coefficients	Odds ratio	P-values
None	3.09492	22.08547	0.002
Septic	0.72649	2.067812	0.636
Other	-0.54897	0.577544	0.691
Cost	0.32518	1.384277	0.000
Femaleneone	-0.91661	0.399870	0.164
Femaleseptic	-1.03559	0.355015	0.269
Femaleother	-1.33534	0.263068	0.069
Ageneone	-0.03204	0.968471	0.025
Ageseptic	-0.04089	0.959934	0.090
Ageother	-0.02372	0.997631	0.898
Incomeneone	-0.00003	0.999964	0.995
Incomeseptic	0.00278	1.002788	0.732
Incomeother	0.00287	1.002873	0.696
Nchildrennone	0.03098	1.031463	0.782
Nchildrenseptic	0.03328	1.033846	0.858
Nchildrenother	-0.16268	0.849863	0.426
Educationnone	-0.19247	0.824920	0.000
Educationseptic	-0.05012	0.951118	0.488
Educationother	0.05312	1.054553	0.436

Number of observations: 721

LR chi-square (19 df)*: 1469.88

P-value (model): 0.0000

Pseudo McFadden: 0.7442

* Likelihood Ratio (LR) chi-square with 19 degrees of freedom (df)

In comparing conditional logit, mixed logit and multinomial probit analysis applied to locational choice model, Dahlberg and Eklöf (2003) found that, when relaxing the IIA-assumption (and also the assumption of fixed coefficient inherent to conditional logit model), estimators for the three analyse led to the same conclusions. However, testing for IIA, different results were obtained. They stated that *“as we exclude one of the regressand from the estimated model, we simultaneously increase the variance of the error term in the estimated model. As the estimated coefficients are*

proportional to the inverse of the root of the assumed error variance, we will observe a 'multiplicative' bias in the mean and standard deviation estimates" (page 2).

As the multinomial logit reflects the effects of unobserved portion of utility, it is probable that IIA property is affecting all the coefficients. To test this assumption, one can apply a Hausman-McFadden test to the estimated coefficients (Hausman and McFadden, 1984).

7.2.1.1. A Hausman-McFadden test

The results of the Hausman-McFadden test are presented in Table 7.2.

It was not possible to reject the null hypothesis of independence even at the 30% level, for the *None* and *Septic* choices. However, the result was inconclusive (statistics could not be computed) for the choice *Other*. It was not an uncommon result. According to Hausman and McFadden (1984) and Small and Hsiao (1985), it is a common result in finite sample application, and it happened because the matrix $(V_a - V_b)$ in (6.3) is not positive semidefinite (positive semidefinite is analogous to a positive real number) which can generate inconclusiveness.

Table 7.2. Hausman-McFadden Test Statistic

	Test statistic	P-values
NONE removed	8.58	0.8039
SEPTIC removed	1.57	0.9999
OTHER removed	-48.31	-

Small and Hsiao (1985) also pointed out that the test may encounter computational and inference problems because of lack of nonnegative definiteness of the difference of the estimates of two variance matrices (due to the arbitrary selection of the alternatives by the researcher). However, when the inadequacy of the multinomial logit is revealed, one may then proceed to investigate more

flexible specifications such as mixed logit models. The results of the mixed logit analysis are presented in the subsection 7.2.3.

7.2.2. *The latent variable model*

Cronbach's alpha value for the Multiple Indicator (MI) part of the MIMIC model was high: for grouped *attitude towards sanitation* it was 0.8579, and for grouped *perception* it was 0.7859. The η matrix and its λ elements are shown in Table 7.3.

All indicators showed consistent z-statistic, against the null hypothesis. The only exception was the last indicator of attitude "it is comfortable" (z_{10}). It suggests that this indicator may not be converging to measure a single underlying attitude construct (Huges et al., 1986). However, to confirm this information, further investigations would be needed (Jöreskog and Sörbom, 1979).

In general, the indicators converged very well to explain the latent variables, *attitude* and *perception*, confirming the existence of these latent constructs when their indicators were grouped to represent them.

The model showed a good fit, although RMSEA was higher (0.075). RMSEA is a measure of goodness-of-fit for the entire model, not for the coefficients, and a good RMSEA should have a value of 0.05 or less. However, as discussed in the Chapter 6, other indicators should accompany the analysis to consider that a model has a good or bad fit. The Chi-square (and *P*-value) and SRMR presented a good fit, significant at 5%. Moreover, Cronbach's alpha values for the coefficients were high.

Table 7.3. The η Matrix of Factor Loadings (t -statistic in parentheses)

Indicator	ξ_{attitude}	$\xi_{\text{perception}}$
Z_1	1	
Z_2	1.059 (23.29)	
Z_3	1.085 (23.58)	
Z_4	1.123 (25.96)	
Z_5	1.082 (21.59)	
Z_6	1.079 (21.43)	
Z_7	1.094 (23.42)	
Z_8	1.091 (22.47)	
Z_9	1.109 (24.82)	
Z_{10}	3.993 (1.149)	
Z_{11}		1
Z_{12}		1.053 (11.15)
Z_{13}		1.098 (13.31)
Z_{14}		1.125 (12.50)
Z_{15}		0.922 (12.77)
Z_{16}		1.240 (13.30)
Z_{17}		1.166 (13.90)
Z_{18}		1.054 (11.98)
Z_{19}		1.060 (13.72)
Chi-square (236 df*): 1192.32		
P-value (model): 0.0000		
RMSEA: 0.075		
SRMR: 0.064		

*Degrees of freedom.

The indirect causal relationship among *perception* and *attitude* latent construction and socio-economic and demographic characteristics of households and alternative attributes is tested with the “MIC” model (second part of the MIMIC model), and it is presented in Table 7.4.

Table 7.4. The δ Matrix (z-statistics in parenthesis)

Latent Variable	FEMALE	AGE	INCOME	NCHILDREN	EDUCATION	COST
$\xi_{attitude}$	0.007 (0.755)	0.000 (0.680)	0.000 (1.743)	-0.020 (-1.092)	0.003 (1.596)	-
$\xi_{perception}$	-0.003 (-0.289)	0.001 (0.891)	0.000 (0.580)	-0.045 (-1.941)	0.004 (1.665)	0.000 (2.449)
Correlation “Attitude” and “Perception”: 0.326 (9.714)						

Not all of the variables demonstrated indirect causal relationship with the latent variables. Only variable *cost* was significant, but the magnitude of its coefficient was zero.

Ben-Akiva et al. (1999) noted that sometimes it can be difficult to find good causal relationship for the latent variables, given their subjective characteristics. However, when we tested the latent variables in a mixed logit model, the indirect causal paths of these variables could be captured and they could explain differences among alternatives, contributing to understand the household behavioural in demand studies, related to the alternatives. Their inclusion favours the increase of magnitude and significance of coefficients related to individuals and alternative characteristics. Johansson et al. (2005) observed similar results when testing their model for environmental analysis.

7.2.3. The discrete choice model with latent variables

Before demonstrate the results of the full discrete choice model, with and without the latent variables, I will present the disaggregated model aiming to better understand the results. In the Table 7.5, the results of the alternatives are presented, to determine preferences.

Table 7.5. Estimation Results for the Mixed Logit (MXL) Model: Reference Model (Using 125 Halton Draws)

Variables	Coefficients	t-statistic [^]
None	-1.74323	7.67
Septic	0.56701	4.32
Other	-0.42094	-4.95
Number of observations: 721		
P-value (model): 0.0000		
Pseudo McFadden: 0.6994		LR chi-square: 989.45

The reference alternative, *Bazul*, was preferred to *None* and *Other*, nevertheless, *Septic* was preferred to *Bazul*. Therefore, the preference-utility-maximisation axiom can be defined as: *Septic* > *Bazul* > *None* and *Other*. Again, no investigation comparing *None* and *Other* was conducted thus, we cannot infer utility between them. In addition, this represent the general preferences not stratified by socio-economic and demographic characteristics of households. Goodness-of-fit was high for all parameters: for the model and coefficients.

When the gender variable is included in the model, there are some gains in statistical significance of the model and coefficients. From the Table 7.6, we can observe that females seem to be inclined to opt to *Other* when compared to *Bazul*. It was an expected result because *Other* refers to the old municipal system that is still in operation in some households. As this system performs relatively well in isolating the households sewer, without presenting regular problems with maintenance (as is the case of the septic tank), it is expected that there is no incentive to change

from *Other* to *Bazul*. The household's change of connection from *Other* to *Bazul* possibly would be induced by the Law, not by the households' preferences, as suggested by the results. When comparing with *None* and *Septic* females prefer *Bazul* rather than those alternatives. Although *Femalenone* coefficient was not significant.

Table 7.6. Estimation Results for the Mixed Logit (MXL) Model: Reference Model with Gender Variables (Using 125 Halton Draws)

Variables	Coefficients	t-statistic [^]
None	-2.02758	5.01
Septic	0.64673	3.94
Other	-1.02900	-4.03
Femalenone	-0.57890	-1.27
Femaleseptic	-0.98601	-2.24
Femaleother	1.04326	3.77
Number of observations: 721		
P-value (model): 0.0000		
Pseudo McFadden: 0.6994		LR chi-square: 998.32

The result of the full discrete choice model is shown in Table 7.7. Aiming to compare the contribution of latent variables in explaining the choice model, I estimate two models: a mixed logit without latent variables and a mixed logit with those variables. The results were interesting: the latent variables clearly improved the magnitude and significance of the coefficients of the variables, and the difference between the analysis was extremely significant in terms of the resulting gain in explanatory power and the improved specification of the discrete choice model.

Analysing the other socio-economic and demographic characteristics of household we have that older people prefer *Bazul* rather than other connection systems.

Table 7.7. Estimation Results for the Mixed Logit (MXL) Model: Reference Model (without Latent Variable) and the Latent Variables Enriched (LVE) Model (Using 125 Halton Draws)

Variables	MXL _{Reference}		MXL _{LVE}	
	Coefficients	t-statistic [^]	Coefficients	t-statistic [^]
None	-2.72811	3.37	-3.01781	5.38
Septic	0.84445	5.12	1.34471	7.84
Other	-1.32556	-4.25	-1.58470	-8.02
Cost	-0.22451	-7.67	-0.48713	-5.76
Femalenone	-0.84771	-1.30	-2.00388	-4.12
Femaleseptic	-1.14162	-3.63	-3.56741	-3.32
Femaleother	1.27897	8.11	1.08417	8.18
Agenone	-0.02149	-1.28	-0.97411	-5.12
Ageseptic	-0.02789	-4.60	-0.04871	-7.18
Ageother	-0.00374	-3.82	-0.57410	-5.38
Incomenone	-0.00021	-2.79	-0.01313	-7.77
Incomeseptic	0.00874	1.12	1.00371	5.28
Incomeother	0.00658	2.24	1.00805	5.88
Nchildrenone	0.00871	4.13	0.00323	6.02
Nchildrenseptic	0.03344	4.07	0.87440	5.90
Nchildrenother	-0.99433	-0.20	-2.74106	-7.07
Educationnone	-0.57410	-8.13	-1.87721	-5.16
Educationseptic	-0.02577	-3.68	-1.07895	-2.58
Educationother	-0.05741	3.77	-0.41738	-4.78
Noneattitude			-0.70549	-2.60
Septicattitude			0.54763	2.72
Otherattitude			1.22741	2.77
Noneperception			-0.24004	-1.03
Septicperception			0.60347	2.24
Otherperception			0.16882	3.93
Number of observations: 721				
P-value (model): 0.0000		LR chi-square: 1216.29		1319.63
Pseudo McFadden: 0.778				

[^] t with 95% confidence interval to test for difference, t* = 1.98

Regarding to income, the findings suggest that households with a high level of income preferred *Septic* and *Other* when compared with *Bazul* connection. However, *Bazul* was preferred to *None*, as expected. It was anticipated that *Bazul* would be a preferred alternative when compared with *Septic* because of the maintenance problems related by the households concerning septic tanks. One explanation for this result may be related to the fact that many septic tanks in Salvador are connected to pipes that are linked to the Bahia Azul system, many of them informally. It was an arrangement of households to try to avoid flooding in determined critical areas during the rainy season (March to May) – (SEDUR, 2006). This action may generate no incentives to households change their sewage connection, and so can explain in part the preference for septic tanks. Nonetheless, further investigation is necessary for a final conclusion. The option for *Other* rather than *Bazul* can also be explained by the absence of incentive to change from one well-performed system to another one.

Moreover, the results pointed out that the higher the number of children in the households, the lower the probability that the option *Bazul* will be chosen. But, the alternative *Bazul* overcome alternative *Other* when the variable number of children (*Nchildren*) is analysed: when compared with alternative *Other*, households with more children at home preferred *Bazul*.

When evaluating the variable level of education, however, *Bazul* is the preferred connection system: the higher is the level of education, the higher is the probability of households opting for the Bahia Azul programme.

In estimating a mixed logit model, the cost variable presented the expected result: when the price of alternative *Bazul* increases, the likelihood of *Septic* or *Other* being chosen increase (except for the alternative *None*, which have no link with costs). This reversal from the multinomial logit model may be explained by the capacity of a mixed logit model to better capture the elements in the error term).

The interpretations for all socio-economic and demographic variables are valid to the two models, with and without latent variables. The difference is that, the magnitude and significance of the coefficients of the model with latent variables were more robust statistically, whereas in the model without latent constructs some coefficients were not significantly different from zero (such as “Femalenone”, “Agenone”, “Incomesepitic”, and “Nchildrenother”).

In fact the inclusion of latent variables in the model improved not only the magnitude and significance of the coefficients, but also the goodness-of-fit of the model. The LR chi-square increased when these variables were added, in spite of no changes in the Pseudo-McFadden. Both models were highly significant.

In spite of the MIMIC model has failed to test the causal relationship among latent variables and socio-economic and demographic characteristic of individuals, and alternative attributes, the mixed logit model suggest that these latent constructs contribute to explain demand and difference among households, related to the alternatives. Those actors with positive perception and attitude towards sanitation preferred *Bazul* rather than *None*. However, *Septic* and *Other* were preferred to *Bazul*.

Jenkins (1999) hypothesised that positive perception and attitude towards sanitation favour the adoption. As the assessment of household perception and attitude was based on the connection, as a way to isolate faeces, it was expected that the all sewage systems were preferred to *None*. Further discussion is carried out in Chapter 8.

All but one coefficient were statistically significant different from zero at a 5% significance level. The exception was for variable “Noneperception” (perception related to open defecation). This result can have been influenced by the, relatively, few number of respondents that had no latrine and connection at home, affecting the statistical analysis. However, no further investigation was conducted to check this absence of significance.

7.2.3.1. Price (cost) elasticities

Price (cost) elasticities are presented in Table 7.8.

Table 7.8.

t-statistic in parenthesis

	BAZUL	SEPTIC
10% increase price	-0.34	-0.13
	(-7.60)	(-14.60)

Interpretation of elasticities is quite involved. In general, what is important in the analysis is the sign indicating the direction of choice and the magnitude of the coefficients. In this estimation, we can confirm that sanitation was a normal good because results demonstrated that all own price (cost) elasticities were negative and of great magnitude, in line with the expectations. Implications of these results are discussed in the subsequent chapter.

Chapter 8 – Conclusions and Policy Implications for Sanitation Promotion

8.1. General comments

The primary objective of this thesis was to estimate the demand for sanitation in Salvador/Brazil, using a framework of analysis that allows the causal relationship among the socio-psychological latent constructs, the socio-economic and demographic characteristics of households, and the alternative attributes to be tested and demonstrated. This objective has been met in part.

Despite the inability to verify the causal relationship between the socio-economic and demographic characteristics of households and latent variables, the Hybrid Choice Model applied to sanitation fills the gap between behavioural theory and discrete choice models, demonstrating that those latent variables are manifestations of the underlying revealed preferences. As far as I am aware, this is the first study using observed and latent variables in a mixed logit model to explain demand for sanitation adoption in a developing country.

So far, the models applied to sanitation studies were either misspecified, presenting serious bias and not measuring the preferences/utilities they attempt to measure – as was the case of the contingent valuation approach –, or they were unable to test the causal relationship among variables (latent and observable) – as was the case of Jenkins' model. Other models simply did not include those latent constructs, as the probabilistic choice models. Furthermore, traditional economic theory, based on the direct maximisation of utility and consumer rationality, was inappropriate for this analysis, in view that the core of this assessment is the choice of the goods or services itself, not its qualities or characteristics that satisfy determined needs or drives.

This chapter will demonstrate the extent to which this study's objectives have been met. The conclusions are presented in four parts, alongside this introduction. The second section presents a summary of the main findings in light of the literature, collaborating to the debate on household's behaviour related to sanitation, in developing countries. As behaviour theory applied to sanitation is still an exploratory field, this debate can add and/or confirm hypotheses to this evaluation. In the third section, I discuss the contributions of my model to behavioural theory and discrete choice models applied to sanitation studies. In this section I address to the contribution of the Hybrid model to econometric evaluation and forecasting, with special attention to the incorporation of latent constructs in a discrete choice model, the limitations and possible bias problems in my methodology. The fourth presents a brief discussion on the role of the model accuracy to forecast demand. My position is that it is not the core of the analysis. The core should be the capacity of the model in determining the causal relationship among variables. The fifth section deals with the implications of the study results for sanitation promotion in Salvador (and possibly elsewhere), and the final section suggests areas for future research.

8.2. Summary of the main findings

8.2.1. Perception and attitude affecting sanitation adoption in Salvador/Brazil

As previously discussed, households expected that the service “sanitation” satisfied their needs or drives of faeces isolation. The drives (perception and attitude towards sanitation) of cleanliness, safety, easy access, not smelly, usefulness, healthy, convenience, privacy, suitability, the increasing of the monetary value of the house, security, and comfort, *not* the latrine and connection themselves, were what the households really cared about. And together households socio-economic and demographic characteristics, and attribute of alternatives (including constraints, as costs), the consumption process is determined.

Jenkins (1999) found that without strong drives (perception, attitude and other latent construction) for a latrine, the household would be uninterested in a change in rural Benin. The presence of constraints alone (costs of connection, for instance) was not the main reason for non-adoption. The presence of one or more drives had strong influence on adoption.

My results show that both, adopters and non-adopters, had a positive perception and attitude towards sanitation in Salvador/Brazil. They perceived usefulness, suitability, and convenience as the three most important quality of a latrine. Health protection was also an important quality linked to adoption, and it was in consonance with the agencies of sanitation promotion. This result is contrary to that of Jenkins’ research in rural Benin where households did not associate lack of sanitation to cause of diseases. In addition, Whittington et al. (1993) found that in Kumasi, Ghana, households’ dissatisfaction with public latrines was the inconvenience of using them, not the risks to public health that they posed.

In Salvador, households referred to diarrhoea and other infectious intestinal disease as the result of the absence of sewage system (Barreto, 2005). These results have direct effect on market

research, pointing to the importance of the historical and cultural context when investigating and planning sanitation intervention.

With respect to the construction of latent constructs, the power of the test of the causal relationship between these constructs and socio-economic and demographic characteristics of households was low. However, these results should be taken with caution. Ben-Akiva et al. (1999) stated that, in general, it can be difficult to find solid causal variables for the latent variables, and it is part of the subjective nature of latent construction. As will be discussed in the following sections, possible bias also can be affecting the low power of explanation of the causal relationship among those variables.

Nevertheless, the inclusion of latent constructs into the model improved the specification of the mixed logit approach and the explanatory power of the analysis demonstrating a causal relationship between those variables and alternative attributes. Taken in sum, the results of the Hybrid model appear to be generally internally consistent, and consistent with the existing literature on sanitation demand.

8.2.2. Household's socio-economic and demographic characteristics affecting demand for sanitation

The variables "FEMALE", "AGE", "INCOME", "EDUCATION", and "NCHILDREN" were the household socio-economic and demographic characteristics included in the demand analysis. All of these variables included in the evaluation could explain differences among households.

Women preferred the old municipal system to isolate their household faeces matter. As previously discussed, it was not an unexpected result because the municipal system is still in operation in some households and performs relatively well in isolating households' sewage.

One explanation for what women did not prefer Bahia Azul programme could be linked to the composition of neighbourhood's associations. In Salvador, the formation of these local associations was intensified from 1979, aiming to give more bargaining's power to the local population to negotiate with politicians. These associations are more numerous in poor neighbourhoods. Although household decisions are independent from those organisations, these associations are recognised by politicians as having an important influence on household opinion. For instance, the negotiation for inclusion of Bahia Azul's programme of sanitation, during the early stages of this programme in some neighbourhoods, depended on the power of those associations in negotiating with politicians. The more representative and organised they were, the more power they had for negotiations with politicians (Moraes et al., 2000; Serpa and Brito, 2004).

However, these associations were dominated by men rather than women (Serpa and Brito, 2004). Hence, all technical discussions about the sanitation projects were attended essentially by men, who improved their skills and understanding about the different technologies. For women the central objective of excreta disposal system of treatment is to isolate faeces, and to avoid contamination and disease. It is therefore understandable that the type of technology is not important, but rather, the capacity of this technology to achieve the target of avoiding disease. Therefore, for women the old municipal system could execute the same task of health protection than Bahia Azul programme, without substantial perceived differences among them. For men, however, it is expected that the preference be for Bahia Azul alternative.

In general, older people have a good understanding and stronger drives towards sanitation than young people. Consequently, it was expected that older people would prefer Bahia Azul's programme than other types. My results were in consonance with this expectation.

Concerning the variables "INCOME" and "EDUCATION", it was expected that the wealthiest households would prefer connection to an external system of sanitation rather than open defecation.

Individuals with a high income and high level of education opted for septic tanks and municipal system. A first explanation for this result was that without any incentive to change the type of connection (from municipal system or septic tank to Bahia Azul's programme), a household would prefer to keep their current excreta disposal structure. Nevertheless, a question remains: in spite of household preference in keeping their current choice, why is Bahia Azul's adoption increasing? In fact, it corresponded to 69% of household connections.

A plausible explanation for this high adoption of Bahia Azul's programme was the introduction of a State Law that obliged the population who did not already have access to that service, to link with the system. Most of the poor population opted to pay for adoption, even if it meant the sacrifice of some basic consumer goods. They opted to do that rather than have problems with the Law. An additional explanation, captured by the anthropological study, and confirmed by the descriptive analysis of the household survey, could be the neighbourhood pressure influencing household adoption.

In addition to their social awareness of their rights to access basic infrastructure, and environmental concerns, neighbourhood pressure seemed to play an important role in influencing household adoption in Salvador. Neighbourhood pressure to diminish negative externalities, as sewage disposal in the streets, was mentioned by some interviewees.

"NCHILDREN" indicated that, the higher was the number of children in the household, the lower was the preference for Bahia Azul's programme. It was an expected result.

Hence, all the observed socio-economic and demographic household characteristics that were chosen to compose the model, did contribute to explaining the demand for sanitation in Salvador/Brazil.

In Bangladesh, Persson (2001) found that choice specific characteristics, such as the cost of alternatives, had a significant effect on household choice. No sanitation, traditional toilet and toilet

provided by UNICEF were preferred by households rather than the modern alternative. Nevertheless, households with higher educational level opted for the modern one.

In Egypt, a study showed that educational level and region of residence (rural or urban) affected household sanitation choice. People who lived in urban areas and that had a medium and higher educational level tended to choose services for sanitation that had a better quality (Abou-Ali, 2003).

8.2.3. Attributes of alternatives

The only sanitation attribute evaluated was cost. In a multinomial logit model, analysis suggested that sanitation was not a normal good: if the price increased, the consumption of the service also increased. However, when estimating demand using a mixed logit model, the sign of cost coefficient changed to the expected negative sign, suggestion that sanitation is really a normal good. This result was also supported by the elasticity analysis.

This misspecification of coefficient's sign might be related to the IIA-property, which if not relaxed using some techniques, affects estimation. The IIA-assumption holds when, for instance, we have that: the probability that Bahia Azul's programme (with probability p) is preferred to septic tanks (with probability $(1-p)$) does not change if the municipal system alternative is included in the evaluation. This restriction (IIA-property) was studied by McFadden (1974) in transportation demand analysis, by Arrow (1951) in his studies on voters, and by Tversky and Kahneman (1974, 1981) in their experiments on consumer behaviour. They concluded that human behaviour rarely adheres to this axiom.

8.3. Contributions to theory

8.3.1. The inclusion of latent variables in a model to explain household behaviour to sanitation adoption: the contributions to discrete choice models and behavioural theory applied to sanitation

The incorporation of latent variables in the model to explain demand for sanitation improved all estimated coefficients (with statistical significance), demonstrating that those constructs had a causal relationship with the alternative attributes, and so influence on choices. The effects of perception and attitude towards sanitation were positive for the all alternatives. Exception was alternative “NONE”, as it was expected.

In spite of the poor causal relationship demonstrated between socio-economic and demographic characteristic of households and latent constructs, the Hybrid model with latent variable seems to be consistent in explaining preferences.

The interest in attitude and perception variables from the standpoint of sanitation policy analysis lies in the question of whether planners can influence behaviour by campaigns to modify attitude and perceptions. A demand model with explanatory attitude and perception adds power of explanation for this investigation. However, it is useful if one can incorporate indicators to understand the process of attitude and perception formation within a historical and cultural context – as for instance, the conduction of surveys to understand the formation of these constructs (McFadden, 1974, 1986). In my survey with households in Salvador, it was found that the marketing promotion of sanitation linked to public health coincided with households’ motives to install a latrine and connection. In rural Benin, however, this promotion was not linked with households’ expectations about latrine. They associated latrine with prestige and comfort rather than health benefits (Jenkins, 1999).

Concerning the choice model (mixed logit), the Hybrid Model could explain difference among household's perception and attitude, with improvements in goodness of fit over choices, correct parameter signs, and more satisfying behavioural representation (Ben-Akiva et al., 1999).

The intrinsic subjective properties of sanitation cannot be captured by a direct maximisation utility axiom, where just the direct qualities of the good or service can be derived (as for instance, price and quantity). Critiques from behavioural theorists of economic studies were centred on the aspect that economic models did not consider the individuals' idiosyncrasies in their analysis, and so, the rationality concept inside of the direct utility maximisation structure, the nucleus of economic theory, is unrealistic, treating the individual as a "black box" for which alternative attributes and individual characteristics are inputs where the observed choice corresponds to the output of the choice.

The Hybrid Choice Model opens the black box permitting more flexibility in the evaluation of consumer demand for sanitation. The process of consumer choice is seen inside an *indirect* utility maximisation structure where households maximise their utility (but they are probably inefficient most of the time).

The results of my investigation confirmed the appropriateness of the Hybrid model for demand evaluation: latent variables incorporated to a mixed logit (discrete choice) model improved the explanation of household behaviour towards sanitation.

The findings of this study also support the *indirect* utility maximisation approach where the consumer choice is expected to satisfy his/her needs or drives, and where the choices reflect the skills and experiences of the individuals.

My results not only contributed to the debate on the importance of incorporation of the perception and attitude latent variables, reinforcing previous explorations about this subject, as in

Jenkins' study, but went further, incorporating these variables into a discrete choice model to forecast demand appropriately.

These findings also reinforce Jenkins' statement that sanitation promotion should not be uniform among settings; instead it should focus on the local/regional beliefs, perception and drives to motivate the adoption, bearing the historical and cultural context of different societies (Jenkins, 1999).

This study has contributed to the identification of factors related to context, design and forecast of demand for sanitation. It was the first to look at the association among latent and observable variables in this market, providing insights into how policy makers can deal with these variables and plan their actions. The model was flexible enough to be reproduced in other settings.

8.3.2. Limitation of my Hybrid Choice Model to estimate demand for sanitation in Salvador/Brazil

8.3.2.1. Estimation procedure

Simultaneous estimation is preferred to sequential one because it produces efficient coefficients (Ben-Akiva et al., 1999). Efficiency is related to minimum variance among the coefficients, therefore, an unbiased estimator (Gujarati, 2003). As I used sequential estimation, it was expected that the estimated coefficients were not efficient, however I expected consistency. Consistency is expected when the estimator ($\hat{\theta}$) approaches the true value (θ) as the sample size gets larger and larger (Gujarati, 2003).

In using the latent constructs, I found that those generated a strong result when comparing the mixed logit model without the latent variables: the likelihood ratio test was bigger and significant when comparing the two models, strongly rejecting the null hypothesis of the reference model without latent variables. However, this result should be viewed with caution because the indicators and causes used to construct the latent variable were estimated using the MIMIC model, and not

directly using the discrete choice model (simultaneous analysis), and the causal relationship analysis between household characteristics and latent construction was not demonstrated. It is possible that using a simultaneous estimation, these causal relationships can be verified, but only further investigations may hold up this possibility. In assessing a demand for sanitation using the sequential estimation, however, the researcher should be aware for the identification rules that apply to a traditional latent variable model and discrete choice model, and that were addressed in this study.

According to Ben-Akiva and Lerman (1985), Bollen (1989) and Ben-Akiva et al (1999), for the Hybrid integrated model, a sufficient, but not necessary, condition for identification can be obtained by:

1. Confirming that the measurement equations for latent variables model are identified. In my model, I used a MIMIC approach;
2. Confirm that, giving the latent variables, the structural equations of the latent variable model are identified (also assessed by the MIMIC model, in my analysis);
3. Confirm that, given the distribution of the latent variables, the choice model is identified. I used the mixed logit model, and parameters were reproduced using a maximum likelihood estimator with Halton sequences (see Chapter 6).

Following the rules above, the possibility of the estimation of inconsistent estimates of the parameters is minimised. However, the sequential model still produced inefficient estimators.

8.3.2.2. The possibility of the influence of the interviewer on households answers

The presence of an interviewer could introduce a bias problem related to influence of the interviewer on household's interpretation of the questions, and consequently their answers. However, as many household had low educational levels, the presence of an interviewer was essential as they could read the questions to them. Furthermore, the interviewer was oriented not to

interfere in the household's interpretation and answers, and just read the questions again in case of doubt.

Other important bias to be considered is the family information bias, where the head of household is asked some questions about other members (Schull and Cobbs, 1969; Sackett, 1979). In this study, questions on family income were asked, and there was the possibility that some information was not correct. However, as the median income in my sample was quite similar to the whole population in Salvador (comparing the 25% most poor strata), I believe that real differences were not important.

8.3.2.3. The historical and cultural contexts as factors underlying households perception and attitude towards sanitation: possibility source of bias

In Salvador, previous studies showed that households had a strong environmental trait, and the lack of sanitation was one of the main concerns of population (Tesh and Paes-Machado, 2004; Barreto et al., 2006). Furthermore, social pressure was an important determinant of adoption: neighbours who had access to sanitation put pressure on households that had not access to adopt, which was a generator of conflicts in the neighbourhoods.

The community vision of excreta disposal treatment as a social right also helps to explain the household's perception and attitude towards sanitation. I assumed that all of these aspects underlined the latent constructs of households' perception and attitude. This assumption was based on the complexity of variables' construction, since all these aspects are interrelated, and to determine the deconstruction process of these variables, demonstrating how they are interrelated and explain each other is an involved evaluation, where the causal relationship is unknown (McFadden, 2001). For this reason, I opted to construct only two latent variables to assess demand, considering that the historical and cultural contexts that underlies socio-economic and demographic

characteristics of households, as assumed by Engel et al. (1968, 1978), also underlies perception and attitude variables.

However, without demonstrating the causal relationship among latent variables and historical and cultural context, using some proxy, this assumption of “underlie action” is only hypothetical and not useful for policy purposes, in terms of capacity of the policy-maker capacity of influence behaviour. Nevertheless, understanding the historical and cultural contexts in which the household is inserted (through an anthropological or psychological approach, for instance) is useful to explore, understand, and construct tools for further evaluation.

My study of demand for sanitation was essentially based on an anthropological study, which gave important inputs to the development of my questionnaire that was applied to the households. In this exploratory and relatively under-researched field that is sanitation, a more qualitative approach is necessary. The objective of economic models (that is forecasting) in this area just is possible if latent constructs are introduced. Otherwise, the traditional approach based only on objective variables provides no answer (McFadden, 1980).

Hence, the behavioural framework and the methodology here presented needs to be extended to further bridge the gap between behavioural theory and discrete choice models, including historical, cultural, tastes, awareness, etc. into the analysis. The inclusion of these potential omitted variables may increase the power of explanation of the model and, possibly, establish a causal relationship between them and household socio-economic and demographic characteristics.

8.4. Forecasting of demand for sanitation: is the accuracy of prediction the core of analysis?

Is accuracy a good indicator of the appropriateness of the model? or should this be evaluated by the evidence of its capacity in demonstrating causal relationship among variables (even only established by household's preferences) and capacity to outperform conventional methods?

In my opinion, the model should be evaluated, primary (1) by its capacity in giving answers to policy makers in how to influence behaviour by campaigns to modify attitude towards sanitation; it is possible if causal relationship is demonstrated; (2) by its capacity to overcome misspecified assumptions, as for example, the misspecified assumption of preference of the contingent valuation method, which is an inferior and costly method, and could be substituted by a more robust approach (as behavioural analysis, anthropological and psychological studies, discrete choice analysis, and the Hybrid approach, for instance); (3) that put on context the different settings to be evaluated with simple adaptations of the methodology; in other words, that it is transferable without major cost and collection of data from a setting to another (Domencich and McFadden, 1975; McFadden 2001), and preferably, (4) that be relatively accurate, to predict a *proxy* of the potential demand .

In short, the model should be essentially sensitive to the operational policy decisions facing sanitation planners, *not necessarily* accurate in forecasting demand, as defended by Friedman (1953). The consumer behaviour is much more complex than the relationship between “leaves” and “sunlight” described by Friedman (McPake et al., 2002).

Traditional economic demand models are too simplistic, and unable to give answer in a non-market field. And models with several variables without demonstrated causal relationship among them is unrealistic and non-useful for policy. The Hybrid model is flexible enough to deal with these

complexities of the sanitation market and consumer behaviour, opening the possibility of analysing unexplored (and realistic) aspects of the subject, and improving the quality of sanitation planning.

8.5. Sanitation promotion

8.5.1. Implications of result for sanitation promotion

In Salvador, perception and attitude towards latrines were strongly perceived.

Promotional messages should come, preferably, together with public health education to reinforce not only the importance of sanitation, but also hygiene habits, in order to contribute to the decrease of negative externalities (as infectious intestinal diseases, beaches pollution, etc.), and improve adoption, since that many households are still without connection to treatment systems.

Reinforcing the understanding of households about sanitation construction is also fundamental for the success of such a campaign. As sanitation is not a good that can be found in a specific market, households need to know about the materials necessary for construction and have skills to do that (or contract someone else). Understanding sanitation technology saves money to households as well as to the Government: households save money using a correct technology, avoiding future damage, and Government saves money on the treatment of disease, environmental pollution (including saving resources comes from tourism), and very regular maintenance in the external connections (sewage's street).

8.5.2. Delivery and support to household

Providing better information, as discussed above, improving access to technical support, developing and disseminating new latrine construction, helps to improve the access to the population. New latrine construction, however, should be in consonance with household tastes. VIP latrines are a simple and cheap solution to isolate faeces, but it is uncommon in Salvador, and its promotion should be accompanied with a preceding pilot to test acceptability of this technology for the population. From the questionnaire, household visit, and personal experience, I suspect that it is an unacceptable system in Salvador.

From interviews, the most desirable (and in fact most common) type of latrine in Salvador, is the typical ceramic with plastic seat, linked to an external system of sanitation. However, more investigation is needed.

8.5.3. Financing schemes

Federal Brazilian Government proposed a new plan of intervention to cover the whole population with basic infrastructure: sanitation, access to water supply, transport, etc. This plan will help the States in complementing their efforts in supplying the basic infrastructure to the population (Brazilian Government, 2007). It is expected that the Bahia Azul's programme will expand its operation to the few areas that are still without any type of external treatment of sewer.

Nonetheless, it is also expected that, as in areas where Bahia Azul's programme is already established, the very poor population will not be able to afford the costs of construction and/or connection. Financial incentives and credit should be considered when promoting the programme. In addition, the State should balance the social benefits and costs, and probably design some plan to operate subsidies, user charges etc., aiming to benefit the very poor population.

8.6. Further research and conclusion

Although the findings of this study have addressed the main objective of estimating the demand for sanitation, there are some aspects that were only touched upon in this thesis and deserve further attention.

The importance of historical and cultural contexts was hypothesised as one of the determinants for sanitation, and previous studies using psychometric instruments have suggested that these contexts have some influence on household choice. In this thesis, I assumed that the historical and cultural background *underlies* the latent constructs, and socio-economic and demographic characteristics of households. However, no statistical test or deep theoretical assessment was carried out. Understanding how these aspects influence and determine latent and socio-economic variables could be useful when planning sanitation interventions. Models applied to psychological studies may be useful in understanding the mechanism of action of these complex factors. In addition, recent advances in discrete choice models may permit the inclusion of these aspects in a forecasting analysis, improving the power of explanation of consumer behaviour. Furthermore, future in-depth, case-study research should assess the importance of context, by comparing across a variety of settings.

Other latent constructs can also be explored in the main analysis to explain demand for sanitation. For instance, in Chapter 5 I found that, alongside perception and attitude towards sanitation, households also were influenced by their neighbourhood and relatives to adopt, by their environmental trait, and by the State's Law, and constrained by the costs of the connection system's maintenance. However, before using these variables in a discrete choice model, it is important to test their causal relationship with socio-economic and demographic characteristics of households and alternative attributes. Otherwise, their inclusion could not mean any significant difference in the analysis.

Another important aspect to sanitation policy design is the investigation of the household welfare status when different price schemes (or subsidies) are introduced. It is particularly important when programmes of interventions are being designed. Government must identify, and choose between, interventions that may overcome obstacles to the success of the sanitation programmes. A consumer surplus analysis is a tool that can help in this process of general understanding of the worse-better off condition of households in terms of welfare. Positive externalities could be addressed and/or improved when sanitation is available for the whole population.

Although these aspects have not been evaluated in this thesis, the demand study here support to understanding of the relationship between latent constructs and objective variables, filling the gap between behavioural theory and discrete choice models. This study has provided a powerful tool of analysis of sanitation intervention for policy makers.

In conclusion, the model used to estimate demand for sanitation in Salvador/Brazil, the Hybrid Choice Model, filled the gap between behavioural theory and discrete choice model in the sanitation literature. In general, the model appear to be consistent in demonstrating those latent variables are manifestations of the underlying revealed preferences, and opening the black box of consumer process of choice, contributing to behavioural and discrete choice model debates within the sanitation demand literature.

References

- Abou-Ali, H. (2003). Child mortality, wealth and education: direct versus indirect effects, Goteborg University, Department of Economics. 2004.
 URL: <http://www.handels.gu.se/epc/archive/00003020/01/gunwpe0114.pdf>
- Ajzen, I. and M. Fishbein (1980). Understanding attitudes and predicting social behavior. New Jersey, Prentice-Hall, Englewood-Cliffs.
- Alaba, O. B. and O. A. Alaba (2002). Determinants of demand for infrastructure in rural and urban Nigeria. P. communication. URL: <http://www.wider.unu.edu/conference/conference-2002-4/conference-2002-4-papers/olumuyiwa%20b.%20alaba%20and%20olufunke%20a.%20alba.pdf>
- Almendon, A., T. Mboya, et al. (1994). Siaya Evaluation Study. London, London School of Hygiene and Tropical Medicine.
- Allais, A. (1953). Le comportement de l'homme rationnel devant le risque, critiques des postulats et axiomes de l'école américaine. *Econometrica* 21: 503-546.
- Andrade, M. R. (1997). O saneamento na Bahia: situação atual e perspectivas. *Bahia Análise e Dados* 7: 5-12
- Angrist, J., Imbens, G.W., Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91: 444-472.
- Anonymous (2000a). Mixed logit. Lecture/Discussion: Qualitative Choice Analysis Workshop. Econometrics Laboratory, Berkeley, University of California. 2005.
- Anonymous (2000b). Tests of IIA: Details. Lecture/Discussion. Econometrics Laboratory. Qualitative Choice Analysis Workshop. University of California at Berkeley.
- Aptech. (undated). Gauss (version 6.0).USA, Aptech.
- Arrow, K. J., R. Solow, et al. (1995). Report of the NOAA panel on contingent valuation, Silver Spring, MD: National Oceanic and Atmospheric Administration, Damage Assessment and Restoration Program.
- Bahia. (1998). Governo do Estado da Bahia (Diário Oficial da União). Lei n. 7307: 7-8.
- Barreto, M. L., A. Strina, et al. (2002). Avaliação do impacto epidemiológico do programa de saneamento ambiental da Bahia de Todos os Santos (Bahia Azul). Salvador, Public Health Institute: 351pgs.

- Barreto, M. L. (co-ord.) (2006). Avaliação do impacto epidemiológico do programa de saneamento ambiental da Baía de Todos os Santos (Bahia Azul). Salvador, Instituto de Saude Coletiva, Federal University of Bahia (Report).
- Bath, C. R. (1996). Endogenous mode choice model with an application to intercity travel. *Transportation Science* 31: 34-48.
- Ben-Akiva, M. and S. R. Lerman (1985). *Discrete choice analysis: theory and application to travel demand*. Cambridge, MIT Press.
- Ben-Akiva, M. and T. Morikawa (1990). Estimation of switching models from revealed preferences. *Transp. Res.* 24(6): 485-495.
- Ben-Akiva, M., D. McFadden, et al. (1999). *Extended Framework for Modeling Choice Behavior*. *Marketing Letters* 10: 187-203.
- Bergson, A. (1938). A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics* 52.
- Bolduc, D., M. Ben-Akiva, et al. (2005). *Hybrid Choice Models with Logit Kernel: Applicability to Large Scale Models*. Boston, Elsevier.
- Bollen, K. A. (1989). *Structural Equations with latent variables*. New York, John Wiley and Son.
- Brazilian Government. (2007). *Programa de saneamento*. URL: <http://www.funasa.gov.br/>
- Brookshire, D. S., A. Randall, et al. (1980). Value increments and decrements in natural resources service flows. *American Journal Agricultural Economics* 62: 477-88.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, NJ: Sage.
- Brownstone, D. and K. Train (1999). Forecasting new Product Penetration with Flexible Substitution Patterns. *Journal of Econometrics* 89: 109-129.
- Cairncross, S. (1992). *Sanitation and Water Supply: Practical Lessons from the decade*. Water and Sanitation Discussion Paper Series, No. 9, UNDP - World Bank Water and Sanitation Program, The World Bank. Washington, D.C.
- Checkley, W., R. H. Gilman, et al. (2004). Effect of water and sanitation on childhood health in a poor Peruvian peri-urban community. *Lancet* 363(9403): 112-118.
- Clawson, M. and J. Knetsch (1966). *Economics of outdoors recreation*. Baltimore, John Hopkins University Press.
- Cooley, W. W. & Lohnes, P. R. *Multivariate Procedures in the Behavioral Sciences*, New York: John Wiley, 1962.

- Cullis, J. and P. Jones (1998). *Public finance and public choice*. Oxford, Oxford University Press.
- Curtis, V. and S. Cairncross (2003). "Water, sanitation, and hygiene at Kyoto." *BMJ* 327: 3-4.
- Currie, J., M. Miah, et al. (1972). *The demand for food: an exercise in household budget analysis*. Manchester, Manchester University Press.
- DeShazo, J. R., T. A. Cameron, et al. (2004). A test of choice set misspecification applicable to stated preferences methods. Los Angeles. URL:
http://faculty.spa.ucla.edu/deshazo/pdf/15/DeShazo_Cameron_Saenz_070904.pdf
- Diamond, P. A. and J. A. Hausman (1994). Contingent valuation: is some number better than no number? *The Journal of Economic Perspectives* 8(4): 45-64.
- Diamond, P. (1996). Testing the internal consistency of contingent valuation surveys. *Journal of Environmental Economics and Management* 30: 337-347.
- Domencich, T. and D. L. McFadden (1975). *Urban Travel Demand: A Behavioral Analysis*. New York, North-Holland Publishing Co.
- Dwyer, J. F. and M. D. Bowes (1978). Concepts of value for marine recreational fishing. *American Journal Agricultural Economics* 60: 1008-12.
- Engel, J. F., D. T. Kollat, et al. (1968). *A model of consumer motivation and behavior*. New York, Holt, Rinehart and Winston Inc.
- Engel, J. F., R. D. Blackwell, et al. (1978). *Consumer Behavior*. Hinsdale, Dryden Press.
- Esrey, S., R. Feachem, et al. (1985). Interventions for the control of diarrhoeal diseases among young children: improving water supplies and excreta disposal facilities. *Bulletin of the World Health Organization* 63(4): 757-772.
- Esrey, S., J. Potash, et al. (1991). Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization* 69(5): 609-621.
- Esrey, S. A. and J.-P. Habicht (1986). Epidemiologic evidence for health benefits from improved water and sanitation in developing countries. *Epidemiologic Reviews* 8: 117-128.
- Eysenck, H. J. (1952). Uses and Abuses of Factor Analysis. *Applied Statistics* 1: 45-49.
- Fishbein, M. (1967). *A consideration of beliefs and their role in attitude measurement*. New York, John Wiley.
- Fishbein, M. and I. Ajzen (1975). *Belief, attitude, intention, and behavior: an introduction to theory and research*. Reading, Addison-Wesley.

- Freeman, A. M. (1979). *The benefits of environmental improvement: theory and practice*. Baltimore.
- Friedman, M. (1953). *The methodology of positive economics. Essays in Positive Economics*. Chicago, Chicago University Press.
- Gertler, P. and J. van-der-Gaag (1990). *The willingness to pay for medical care: evidence from two developing countries*. Baltimore, John Hopinks University Press.
- Glass, G. V. and T. O. Maguire (1966). *Abuses of Factor Scores*. *American Educational Research Journal* 3: 297-304.
- Green, D., K. E. Jacowitz, et al. (1998). *Referendum Contingent Valuation, Anchoring and Willingness to Pay for Public Goods*. *Resources and Energy Economics* 20(2): 85-116.
- Green, W. H. (2003). *Econometric Analysis*. New Jersey, Prentice Hall-Pearson Education International.
- Griffin, C. C., J. Briscoe, et al. (1995). *Contingent valuation and actual behaviour: predicting connections to new water systems in the State of Kerala, India*. *The World Bank Economic Review* 9(3): 373-395.
- Gujarati, D. N. (2003). *Basic Econometrics*. New York, McGraw-Hill Companies
- Haab, T. C. and R. L. Hicks (1997). *Accounting for choice set endogeneity in Random Utility Models of recreation demand*. W. P. 9710, East carolina University, Department of Economics.
- Heckman, J. (2000). *Causal parameters and policy analysis in economics: a twentieth century retrospective*. *Quarterly Journal of Economics* 115 (February): 45-97.
- Hausman, J. *Specification and estimation of simultaneous equation models*. In: Griliches, Z., Intriligator, M. (eds.) (1983), *Handbook of Econometrics*, vol. 1. North-Holland, Amsterdam.
- Hausman, J. (2003). *Triangular structural model specification and estimation with application to causality*. *Journal of Econometrics* 112: 107-113.
- Hensher, D. A. and W. H. Green (2003). *The mixed logit model: the state of practice*. *Transportation* 30(2): 133-176.
- Hicks, J. R. (1939). *Value and Capital*. Oxford, Clarendon Press.
- Hoehn, J. and A. Randall (1989). *Too many proposals pass the benefit cost test*. *American Economic Review* 79: 544-51.
- Huelsenbeck, J. P. and K. A. Crandall. (1997). *Phylogeny estimation and hypothesis testing using maximum likelihood*. *Ann. Rev. Ecol. Syst.* 28:437-466

- Hughes, M. A., R. L. Price, et al. (1986). Linking Theory Construction and Theory Testing: Models with Multiple Indicators of Latent Variables. *The Academy of Management Review* 11: 128-144.
- Instituto Brasileiro de Geografia e Estatística (2004). Pesquisa Nacional por Amostra de Domicílios (PNAD). 2004. URL: <http://www.ibge.gov.br>
- Jackson, J. E. (2003). *A user's guide to principal components*. New York, John Wiley & Son.
- Jacowitz, K. and D. Kahneman (1995). Measures of anchoring in estimation tasks, *Personality and Social Psychology Bulletin*.
- Jaswal, I. J. (1991). Two steps forward; three backwards. *Indian J Matern Child Health* 2(2): 36-39
- Jenkins, M. W. (1999). Sanitation promotion in developing countries: why the latrines of Benin are few and far between. *Civil and Environmental Engineering*. Davis, University of California: 416.
- Johansson, M. V., T. Heldt, et al. (2005). Latent Variables in a Travel Mode Choice Model: Attitudinal and Behavioural Indicator Variables. Working Paper. Uppsala University, Department of Economics.
- Joreskog, K. G. and D. Sordom (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA, Abt.
- Abts, K. & Swyngedouw, M. (2000). Anvers: chacun pour soi et le Vlaams Blok contre tous? *Politique: Revue de Débats*, Juillet-Août, (17):27-31
- Kahneman, D. and A. Tversky (1979). Intuitive prediction: biases and corrective procedures. *Studies in Management Science* 12: 313-327.
- Kahneman, D. and J. L. Knetsch (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of Environmental Economics and Management* 22: 57-70.
- Kahneman, D. and I. Ritov (1994). Determinants of stated willingness to pay for public goods: a study in the headline method. *Journal of Risk and Uncertainty* 9: 5-38.
- Kahneman, D., I. Ritov, et al. (1999). Economic preferences or attitude expressions? an analysis of dollar responses to public issues. *Journal of Risk and Uncertainty* 19(1-3): 203-35.
- Kenny, D. A. (2003). Effect of the number of variables on measures of fit in structural equation modelling. *Structural Equation Modelling: a Multidisciplinary Journal* 10(3): 333-351.
- Daniel Kahneman, 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics, *American Economic Review*, American Economic Association, vol. 93(5), pages 1449-1475

- Knetsch, J. L. and J. A. Sinden (1984). Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value." *The Quarterly Journal of Economics* 99(3): 507-521.
- Kothari, G. (1987). Diarrhoea in urban slums: Bombay. *Dialogue Diarrhoea* 31: 4-5.
- Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. (1971), *Foundations of Measurement, Vol. I*, Academic Press, New York.
- Krutilla, J. V. and A. Fisher (1975). *The Economics of Natural Environments*. Baltimore.
- Lancaster, K. (1966a). Change and Innovation in the Technology of Consumption. *The American Economic Review* 56: 14-23.
- Lancaster, K. J. (1966b). A new approach to consumer theory. *The Journal of Political Economy* 74(2): 132-157.
- Liao, T. F. (1994). *Interpreting probability models: logit, probit, and other generalized linear models*. Iowa, Sage Publications.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Louviere, J. J., D. A. Hensher, et al. (2000). *Stated choice methods: analysis and application*. Cambridge, Cambridge University Press.
- Luce, R. D., Krantz, D. et al. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement: Vol. 3. Representation, axiomatization, and invariance*. San Diego, CA: Academic.
- Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, Cambridge University Press Mas-Colell, A., Whinston, M. D., and Green, J. (1995). *Microeconomic Theory*. New York: Oxford University Press.
- Maelli, F., Rubin D.B. (2003). Assumptions allowing the estimation of direct causal effects. *Journal of Econometrics* 112: 79-87.
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics* 3: 303-328.
- McFadden, D. (1986). The choice theory approach to market research. *Marketing Science* 5(4): 275-297.
- McFadden, D. (1994). Contingent valuation and social choice. *American Journal Agricultural Economics* 76(4): 689-708.

- McFadden, D. (1998). Rationality for Economists? *Journal of Risk and Uncertainty* 19(1-3): 73-105.
- McFadden, D., Train, K. (2000). Measurement of Simulation Variance in Parameter Estimation Working Papers _008, University of California at Berkeley, Econometrics Laboratory Software Archive.
- McFadden, D. (2001). Economic Choices. *The American Economic Review* 91: 351-378.
- McPake, B., L. Kumaranayake, et al. (2002). Health economics: an international perspective. London, Routledge.
- Mitchell, R. C. and R. T. Carson (1989). Using surveys to value public goods. Resources for the future. Washington D.C.
- Moffitt, R. (2003). Causal analysis in population research: an economist's perspective. *Population and Development Review* 29 (3): 448-458.
- Mokhtarian, P. L. and I. Salomon (1994). Modeling the choice of telecommuting: setting the context. *Environment and Planning A* 26: 749-766.
- Mokhtarian, P. L. and I. Salomon (1996). Modeling the choice of telecommuting 2: a case of the preferred impossible alternative. *Environment and Planning A* 28: 1859-1876.
- Moraes, L. R., J. A. Caucio, et al. (2000). Impact of drainage and sewage on diarrhoea in poor urban areas in Salvador/Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*.
- Morikawa, T. M., M. Ben-Akiva, et al. (1996). Incorporating Psychometric Data in Econometric Choice Models. Working Paper. Massachusetts, Massachusetts Institute of Technology.
- Muthen, M. (2004). Mplus (version 3.01). Los Angeles, Muthen & Muthen.
- Nunnally, J. (1978). Psychometric theory. New York: McGraw-Hill.
- Osgood, C. E., G. J. Suci, et al. (1957). The measurement of meaning, University of Illinois, Urbana, IL.
- Payne, J. W., J. R. Bettman, et al. (1999). Measuring constructed preferences: towards a building code. *Journal of Risk and Uncertainty* 19(1-3): 243-270.
- Persson, T. H. (2001). Demand for water and sanitation in Bangladesh, Department of Economics, Lund University. 2003. URL: <http://www.nek.lu.se/publications/workpap/workpaps.html>.
- Persson, T. H. (2002). Welfare calculations in models of the demand for sanitation. *Applied Economics* 34: 1590-1518.

- Polydoropoulou, A. (1997). Modelling User Choice Response to Advanced Traveller Information System (ATIS). Department of Civil and Environmental Engineering. Massachusetts, Massachusetts Institute of Technology.
- Regens, J. L. (1991). Measuring environmental benefits with contingent markets. *Public Administration Review* 51(4): 345-352.
- Rijsberman, F. (2004). The water challenge. Copenhagen Consensus Challenge Paper, Copenhagen, Denmark.
- Riviere-Cinnamond, A. (2006). Economics of animal health: implications for public health. *The Economics of Infectious Disease*. J. A. Roberts. New York, Oxford University Press.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy* 82: 34-55.
- Rubin, D.B. (1974). Estimating causal effects of treatments in randomised and nonrandomised studies. *Journal of educational Psychology* 66: 668-701.
- Sackett, D. L. (1979). Bias in Analytic Research. *J Chron Dis* 32: 51-63.
- Samples, K. C. and J. R. Hollyer (1990). Contingent valuation of wildlife resources in the presence of substitutes and complements. Boulder, Westview Press.
- Samuelson, P. (1947). *Foundation of economic analysis*. Cambridge, Mass., Harvard University Press.
- Savage, L.J. (1954). *The Foundations of Statistics*. New York: Wiley
- Schull, W. J. and S. Cobbs (1969). The intrafamilial transmission of rheumatoid arthritis. 3. The lack of support for a genetic hypothesis. *J Chronic Dis* 22(4): 217-22.
- Schulze, W. D., R. D. Rowe, et al. (1995). Contingent Valuation of Natural Resource Damages Due to Injuries to the Upper Clark Fork River Basin, Reported prepared for the State of Montana. Boulder: RCG/Hagler Bailly.
- Schuman, H. and M. P. Johnson (1976). Attitudes and Behavior. *Annual Review of Sociology* 2: 161-207.
- Secretaria de Desenvolvimento Urbano (SEDUR) (2002). Bahia Azul: folheto informativo. Salvador. URL: http://www.sedur.ba.gov.br/programas/ba_programa.asp
- Secretaria de Desenvolvimento Urbano (SEDUR) (2005). Programa Bahia Azul.: folheto informativo. Salvador. URL: http://www.sedur.ba.gov.br/programas/ba_programa.asp
- Seddighi, H. R., K. A. Lawler, et al. (2000). *Econometrics: a practical approach*. London, Routledge.

- Simon, H. (1959). Theories of decision-making in economics and behavioural science. *American Economic Review* 49: 253-283.
- Smith, V. K. (1992). Arbitrary values, good causes, and premature verdicts. *Journal of Environmental Economics and Management* 22: 71-89.
- Soares, L. C., M. O. Griesinger, et al. (2002). Inequities in access to and use of drinking water services in Latin America and the Caribbean. *Rev Panam Salud Publica* 11(5-6): 386-96.
- Stephens, C. (1996). Urbanisation: the implications for health. *Afr Health* 18(2): 14-15.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.
- Teixeira, M. d. G., M. L. Barreto, et al. (2002). Sentinel areas: a monitoring strategy in public health. *Cadernos de Saude Publica* 18(5): 1189-1195.
- Tesh, S. N. and E. Paes-Machado (2004). Sewers, Garbage, and Environmentalism in Brazil. *Journal of Environment & Development* 13: 42-72.
- Thayer, M. (1981). Contingent valuation techniques for assessing environmental impacts: further evidence. *Journal of Environmental Economics and Management* 8: 27-44.
- Train, K. (1999). *Halton sequences for mixed logit*. Berkeley, University of California. 2005.
- Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science* 211.
- United Nations (UN) (2004). *Monitoring the Millenium Development Goals for Water and Sanitation*, United Nations. URL: http://www.euwi.net/file_upload/niki_tmpphpHcPuN9.pdf
- Varian, H. (1992). *Microeconomic Analysis*. 3 ed. New York: W.W. Norton.
- Varian, H. R. (1999). Commentary on Economic preferences or attitude expressions: an analysis of dollar responses to public issues by Kahneman et al. *Journal of Risk and Uncertainty* 19(1-3): 241-242.
- von Neumann, J. and O. Morgenstern (1947). *Theory of games and economics behavior*, Princeton University Press.
- Walker, I., F. Ordonez, et al. (2000). *Pricing, subsidies and the poor*, World Bank working papers. 2003.
- Walker, J. L. (2001). *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. Department of Civil and Environmental Engineering. Massachusetts, Massachusetts Institute of Technology: 208.

- Walker, J., M. Ben-Akiva, et al. (2004). Identification of the Logit Kernel (or Mixed Logit) Model. Massachusetts.
- Walker, J. L. (2006). Opening up the black box: enrichment behavioral models of spatial and travel choices. *Journal of Transport Geography* 14(5): 396-398.
- Warner, S. (1962). Stochastic choice of mode in urban travel. Evanston, Northwestern University Press.
- Werner, M. (1999). Allowing for zeros in dichotomous-choice contingent-valuation models. *Journal of Business & Economic Statistics* 17(4): 479-486.
- Wheaton, B., Muthen, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in penal models, in D. Heise (Ed.), *Sociological Methodology* (pp. 84-136). San Francisco: Jossey-Bass.
- Whittington, D., D. T. Lauria, et al. (1993). Household sanitation in Kumasi, Ghana: a description of current practices, attitudes, and perceptions. *World Development* 21(5): 733-748.
- Whittington, D. (1998). Administering contingent valuation surveys in developing countries. *World Development* 26(1).
- Wichmann, R. (1995). The link between poverty, environment and development. The political challenge of localizing Agenda 21. *Countdown Instanb* 1(5): 3-4.
- Wilson, T. P. (1971). Critiques of Ordinal Variables. *Social Forces* 49(3): 432-444.
- World Health Organisation (WHO) (2000). Global water supply and sanitation assessment. Geneva, World Health Organisation. URL: http://www.who.int/water_sanitation_health/monitoring/jmp2000.pdf
- Water Supply and Sanitation Council (1994). The Problem of Sanitation. Geneva, Switzerland, Revised Draft Paper of the Working Group on Promotion of Sanitation, 15 March 1994, prepared by Peter Kolsky. Water Supply and Sanitation Collaborative Council Working Group on the Promotion of Sanitation (Report).
- Zaltman, G. (1965). *Marketing contributions from the behavioural sciences*. New York, Harcourt, Brace and Wold.

Appendix 1 - Questionnaire

Demand study for sanitation – Bahia Azul's programme

Name of interviewer: _____ Date of interview: ___/___/___

IDENTIFICATION AND CHARACTERISTICS OF THE HOUSEHOLD HEAD

Questionnaire number: _____

1. Name of the interviewee: _____

2. Gender: () Male () Female

3. Type of job: _____

4. Are you the head of household? () Yes () No

5. If no, who is the head? (and what is the type of relationship with the head?)

6. Interviewee age: _____

7. Head of household age: _____

8. head of household gender: () Male () Female

9. Interviewee colour: () White () Black () Mixed () Indigenous () Asian

10. Head of household colour:

() White () Black () Mixed () Indigenous () Asian

11. Marital status of the interviewee: _____

12. Marital status of the household head: _____

13. Region of the house: _____ **Number of the region:** _____

14. Full address: _____

D. EXPERIENCY WITH LATRINES (NON-ADOPTER)

25. Do you understand what "use a latrine" means?

1. I understand 2. I do not understand

26. Cite the places where normally you use latrine (read the options):

1. house of neighbours 5. health centre
 2. house of friends 6. Shopping centre
 3. house of relatives 7. school
 4. work 8. Other (describe: _____)
 99. I never use latrine

27. Which places you have used a latrine for the first time?

1. house of neighbours 5. health centre
 2. house of friends 6. Shopping centre
 3. house of relatives 7. school
 4. work 8. Other (describe: _____)

28. Which sort of latrine you have used for the first time?

1. Latrine com flush 2. Latrine without flush
 3. Other (describe): _____

29. Which sort of latrine do you prefer to use?

1. Latrine com flush 2. Latrine without flush
 3. Other (describe): _____

30. Why do you prefer to use this sort of latrine? _____

31. With which age did you use latrine for the first time?

1. 0 to 19 years 4. 40 to 45 years
 2. 20 to 30 years 5. more than 45 years
 3. 30 to 40 years 99. I have never used latrine

D. EXPERIENCY WITH LATRINES (NON-ADOPTER).....

33. We would like to know the qualities and non-qualities you appreciate in a latrine. In your opinion a latrine is:

- | | | |
|-------------------------------|---------------|------------------------|
| a. DIRTY | MEDIUM | CLEAN |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |
| b. DANGEROUS | MEDIUM | SECURE |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |
| c. DIFFICULT TO ACCESS | MÉDIUM | PRACTIC |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |
| d. SMELLS | MEDIUM | PLESANT |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |
| e. USELESS | MEDIUM | USEFUL |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |
| f. BAD FOR HEALTH | MEDIUM | GOOD FOR HEALTH |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |
| g. INCOVENIENT | MEDIUM | CONVENIENT |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |
| h. INDISCRETO | MÉDIO | DISCRETO |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |
| i. INADEQUATE | MEDIUM | ADEQUATE |
| 1.....2.....3.....4.....5 | | |
| very little | | little very |

E. PREFERENCE AND INTENTION IN ADOPTING A LATRINE

Imagine you have an available latrine in your home...

34. How often you would use the latrine when you needed?

1..... 2..... 3 4 5
 never very few some times very often always

35. And your children?

1..... 2..... 3 4 5
 never very few some times very often always

36. Do you have intention to install a latrine in your home?

() 1. yes () 2. no

37. What type of latrine would you like to install in your home (read the options)?

- () 1. latrine without flush
- () 2. latrine with flush but with no connection with system
- () 3. latrine with flush connected to Bahia Azul
- () 4. latrine with flush connected old system
- () 5. latrine with flush connected to septic tank
- () 9. none of the options above

38. Why have you chosen this kind to latrine and connection? _____

39. if you had decided to install a latrine, would you do within next years?

() 1. yes () 2. no

40. Which year are you planning to install a latrine (year) ? _____

41. Which reasons would make difficult to install a latrine?

Reason 1: _____

Reason 2: _____

Reason 3: _____

Reason 4: _____

E. PREFERENCE AND INTENTION IN ADOPTING A LATRINE

42. Which reasons would make you decide to install a latrine?

Reason 1: _____

Reason 2: _____

Reason 3: _____

Reason 4: _____

Pay attention, question 43 only for those respondents who mentioned "health" as a reason to install a latrine

43. Why do you consider that the installment of a latrine is good for health?

44. Below there are some reasons people normally use to justify the installation or not of a latrine. Please read the options and mark yes or no.

a. I would like a latrine to me or to my family ...

() 1. yes () 2. No

b. I have never thought in installing a latrine to me or to my family ...

() 1. yes () 2. No

c. I don't mind if I and my family had to use a collective latrine (to several people among neighbours, for example)...

() 1. yes () 2. No

d. people in my family want to install a latrine inside home...

() 1. yes () 2. No

e. My friends and neighbours insist that I and my family should install a latrine inside home.....

() 1. yes () 2. No

E. PREFERENCE AND INTENTION IN ADOPTING A LATRINE

f. My friends and neighbours have a latrine and I would also like to have one...

() 1. yes () 2. No

g. I would have problems with my neighbours and friends if I had installed a latrine inside my home, because they have no latrine and this could make them envy me and could cause troubles....

() 1. yes () 2. No

45. Which reason mentioned above is the most important? (put the letter) ()

46. Some people normally mention some problems that make difficult to install a latrine. Below there are reasons, please mark if you agree (yes) or not (no).

a. It is very difficult to build a latrine, no one in my family, including myself, does not know the materials that are needed to build it)...

() 1. yes () 2. No

(if the interviewee answered "no" to the question above, ask the question below)

Which materials do you know that are necessary to build a latrine (cite them)?

b. I and my family, or neighbours and friends, could build a latrine inside my home...

() 1. yes () 2. No

c. There is no space in my home, or in my backyard, where I could build a latrine...

() 1. yes () 2. No

d. The type of soil is not appropriate to build a spetic tank ...

() 1. yes () 2. No

e. I can not pay to make a connection with the Bahia Azul program or other program....

() 1. yes () 2. No

f. it is very expensive to me and my family to buy the materials to build a toilet...

() 1. yes () 2. No

**F. ADVANTAGES AND DISADVANTAGES IN ADOPTING A TOILET WITH
LATRINE**

47. In your opinion, what advantages or positive consequences could happen to you or to your family (read, write the answer, and mark the level of importance)

Advantage1: _____

In your opinion, how is it important?

Little important very important
2.....3.....4

Advantage2: _____

In your opinion, how is it important?

Little important very important
2.....3.....4

Advantage3: _____

In your opinion, how is it important?

Little important very important
2.....3.....4

Advantage4: _____

In your opinion, how is it important?

Little important very important
2.....3.....4

Advantage5: _____

In your opinion, how is it important?

Little important very important
2.....3.....4

48. Which advantage is the most important to you?

- () 1. Advantage 1
- () 2. Advantage 2
- () 3. Advantage 3
- () 4. Advantage 4
- () 5. Advantage 5

**F. ADVANTAGES AND DISADVANTAGES IN ADOPTING A TOILET WITH
LATRINE**

49. I would like that you could say the level importance of some advantages of having a toilet with a latrine inside home

1. I and my family will have prestige with people who visit my home

Little important very important

2.....3.....4

2. This can increase the value of the house

Little important very important

2.....3.....4

3. it is easier to me and my family to defecate

Little important very important

2.....3.....4

4. This protects me and my family against diseases

Little important very important

2.....3.....4

5. My home will be more modern

Little important very important

2.....3.....4

6. This will improve the my image before my friends, neighbours and relatives

Little important very important

2.....3.....4

7. This would prevent the dangers of going to mato (insects, rats, etc)

Little important very important

2.....3.....4

8. it is more discreet (no one is looking)

Little important very important

2.....3.....4

9. I and my family feel safer

Little important very important

2.....3.....4

10. I and my family feel more comfortable

Little important very important

2.....3.....4

**F. ADVANTAGES AND DISADVANTAGES IN ADOPTING A TOILET WITH
LATRINE**

53. I would like that you could say the level importance of some disadvantages of having a toilet with a latrine inside home:

1. I and my family feel stuck at home (we prefer to go to the bush)

Little important very important

2.....3.....4

2. The house work will increase because we need to clean the toilet

Little important very important

2.....3.....4

3. A toilet inside house will cause bad smell

Little important very important

2.....3.....4

4. A toilet inside house can cause accidents

Little important very important

2.....3.....4

5. Some neighbours can become inconvenient and want to use our toilet (latrine)

Little important very important

2.....3.....4

6. The number of people visiting us at home will increase

Little important very important

2.....3.....4

7. Some people will influence us to not use the toilet and latrine

Little important very important

2.....3.....4

8. I and my family would have to build a septic tank

Little important very important

2.....3.....4

9. I and my family would have to pay for the connection with the Bahia Azul program

Little important very important

2.....3.....4

**F. ADVANTAGES AND DISADVANTAGES IN ADOPTING A TOILET
WITH LATRINE**

54. What of the disadvantages cited above are the most important to you (cite only 3 and put the corresponding number of each advantage cited)

- () 1. The first most important advantage
 () 2. The second most important advantage
 () 3. The third most important advantage

G. WILLINGNESS TO PAY FOR SANITATION

55. In your opinion, how much cost a complete SYSTEM (toilet with flush and connected to sewerage system)? In your opinion, what is the approximate cost for each item below?

1. Latrine: R\$ _____
 2. Flush: R\$ _____
 3. Material: R\$ _____
 4. Labour: R\$ _____

56. Would you be willing to pay for the complete system (toilet with flush and connected to sewerage system)?

- () 1. yes () 2. no

57. If the answer to the question 56 above was "no", ask the following question: "why not?"

Answer: _____

58. How much you would be willing to pay for the complete system? R\$

END OF THE APPLICATION

GO TO PAGE 23

H. CHARACTERISTICS AND COSTS FOR THE INSTALLATION OF LATRINE

59. Who have built a bathroom with latrine inside your home?

- () 1. I myself or other person of my family who lives with me
- () 2. Other person of my family without payment
- () 3. A friend (without payment)
- () 4. other person (without payment) – who? _____
- () 5. Other person of my family (with payment)
- () 6. Other person (with payment) – who? _____
- () 7. There had existed bath room with latrine when I arrived in this house

60. If the interviewee has answered questions 1, 2, 3 or 4 above, ask:

- a. For how long does it take to build the work? _____
- b. In each year it was concluded? _____
- c. You and/or other person involved in the work missed the day or days in the job?
- () 1. yes () 2. no
- d. Have you and/or other person involved in the work not received any salary or payment because of the work?
- () 1. yes () 2. no
- e. If yes, how much have you lost? (on average per day) RS _____

61. In the case you have paid for the work, how much have you spent on average with:

1. Latrine: RS _____
2. Material: RS _____
3. Flush: RS _____
4. Labour: RS _____
5. Connection with Bahia Azul: RS _____
6. Connection with other system: RS _____
7. Material for the construction of the septic tank: RS _____
8. Labour for the construction of the septic tank: RS _____
10. Other costs (please describe): _____

**H. CHARACTERISTICS AND COSTS FOR THE INSTALLATION OF
LATRINE**

62. How much do you spend on average per year with maintenance your toilet with:

1. _____

I. CHARACTERISTIC OF THE USE OF THE LATRINE

63. Who are other people who do not live with you but use your latrine (toilet)?

- | | |
|---|---|
| <input type="checkbox"/> 1. Neighbours | <input type="checkbox"/> 4. Friends who live nearby |
| <input type="checkbox"/> 2. Relatives who live nearby | <input type="checkbox"/> 5. Friends who live very far |
| <input type="checkbox"/> 3. Relatives who live very far | <input type="checkbox"/> 6. Others (who?): _____ |

I. CHARACTERISTIC OF THE USE OF THE LATRINE

64. I would like that you could say the level importance of some qualities (non-qualities) of having a toilet with a latrine inside home:

a. DIRTY **MEDIUM** **CLEAN**

1.....2 3..... 4 5

very little little very

b. DANGEROUS **MEDIUM** **SECURE**

1.....2 3..... 4 5

very little little very

c. DIFFICULT TO ACCESS **MÉDIUM** **PRACTIC**

1.....2 3..... 4 5

very little little very

d. SMELLS **MEDIUM** **PLESANT**

1.....2 3..... 4 5

very little little very

e. USELESS **MEDIUM** **USEFUL**

1.....2 3..... 4 5

very little little very

f. BAD FOR HEALTH **MEDIUM** **GOOD FOR HEALTH**

1.....2 3..... 4 5

very little little very

g. INCOVENIENT **MEDIUM** **CONVENIENT**

1.....2 3..... 4 5

very little little very

h. INDISCRETO **MÉDIO** **DISCRETO**

1.....2 3..... 4 5

very little little very

i. INADEQUATE **MEDIUM** **ADEQUATE**

1.....2 3..... 4 5

very little little very

J. PROBLEM TO INSTALL TOILET (LATRINE)

65. When have you decided to install your toilet (latrine), which problems have you found?

Mark as below.

a. I did not know the best way to install a latrine

1. yes 2. no 3. not apply

b. There was no space to install the latrine

1. yes 2. no 3. not apply

c. The type of soil was bad to build a septic tank

1. yes 2. no 3. not apply

d. I had difficulties to know about the best materials to buy

1. yes 2. no 3. not apply

e. I had difficulties to know how to use appropriately the materials

1. yes 2. no 3. not apply

f. I had to contract or consult a specialist (someone else with experience)

1. yes 2. no 3. not apply

g. Other (please describe): _____

L. ADVANTAGES AND DISADVANTAGES TO ADOPT A LATRINE

66. In your opinion, what advantages or positive consequences could happen to you or to your family (read, write the answer, and mark the level of importance):

Advantage 1: _____

In your opinion, how is it important?

Little important very important

2.....3.....4

Advantage 2: _____

In your opinion, how is it important?

Little important very important

2.....3.....4

Advantage 3: _____

In your opinion, how is it important?

Little important very important

2.....3.....4

Advantage 4: _____

In your opinion, how is it important?

Little important very important

2.....3.....4

Advantage 5: _____

In your opinion, how is it important?

Little important very important

2.....3.....4

67. Which advantage is the most important to you?

- () 1. Advantage 1 () 4. Advantage 4
 () 2. Advantage 2 () 5. Advantage 5
 () 3. Advantage 3

INTERVIEWER, PAY ATTENTION: ONLY ASK THE QUESTION 68 IF THE INTERVIEWEE MENTIONED "HEALTH" AS ADVANTAGE TO HAVE A LATRINE

68. Why do you consider that the installment of a latrine is good for health?

L. ADVANTAGES AND DISADVANTAGES TO ADOPT A LATRINE

69. I would like that you could say the level importance of some advantages of having a toilet with a latrine inside home:

1. I and my family will have prestige with people who visit my home

Little important very important

2.....3.....4

2. This can increase the value of the house

Little important very important

2.....3.....4

3. it is easier to me and my family to defecate

Little important very important

2.....3.....4

4. This protects me and my family against diseases

Little important very important

2.....3.....4

5. My home will be more modern

Little important very important

2.....3.....4

6. This will improve the my image before my friends, neighbours and relatives

Little important very important

2.....3.....4

7. This would prevent the dangers of going to mato (insects, rats, etc)

Little important very important

2.....3.....4

8. it is more discreet (no one is looking)

Little important very important

2.....3.....4

9. I and my family feel safer

Little important very important

2.....3.....4

10. I and my family feel more comfortable

Little important very important

2.....3.....4

L. ADVANTAGES AND DISADVANTAGES TO ADOPT A LATRINE

70. What of the advantages cited above are the most important to you (cite only 3 and put the corresponding number of each advantage cited)

- () 1. The first most important advantage
 () 2. The second most important advantage
 () 3. The third most important advantage

71. In your opinion, what disadvantages or negative consequences could occur (to you or to your family)? (read, write and mark the level of importance):

Disadvantage 1: _____

In your opinion, how is it important?

Little important very important
 2.....3.....4

Disadvantage 2: _____

In your opinion, how is it important?

Little important very important
 2.....3.....4

Disadvantage 3: _____

In your opinion, how is it important?

Little important very important
 2.....3.....4

Disadvantage 4: _____

In your opinion, how is it important?

Little important very important
 2.....3.....4

Disadvantage 5: _____

In your opinion, how is it important?

Little important very important
 2.....3.....4

72. Which advantage is the most important to you?

- () 1. Disadvantage 1 () 4. Disadvantage 4
 () 2. Disadvantage 2 () 5. Disadvantage 5
 () 3. Disadvantage 3

L. ADVANTAGES AND DISADVANTAGES TO ADOPT A LATRINE

73. . I would like that you could say the level importance of some disadvantages of having a toilet with a latrine inside home:

1. I and my family feel stuck at home (we prefer to go to mato)

Little important very important

2.....3.....4

2. The house work will increase because we need to clean the toilet

Little important very important

2.....3.....4

3. A toilet inside house will cause bad smell

Little important very important

2.....3.....4

4. A toilet inside house can cause accidents

Little important very important

2.....3.....4

5. Some neighbours can become inconvenient and want to use our toilet (latrine)

Little important very important

2.....3.....4

6. The number of people visiting us at home will increase

Little important very important

2.....3.....4

7. Some people will influence us to not use the toilet and latrine

Little important very important

2.....3.....4

8. I and my family would have to build a septic tank

Little important very important

2.....3.....4

9. I and my family would have to pay for the connection with the Bahia Azul program

Little important very important

2.....3.....4

L ADVANTAGES AND DISADVANTAGES TO ADOPT A LATRINE

74. What of the advantages cited above are the most important to you (cite only 3 and put the corresponding number of each advantage cited)

- () 1. The first most important advantage
- () 2. The second most important advantage
- () 3. The third most important advantage

TO CONTINUE THE APPLICATION ON
PAGE 23

M. DATA ON HOUSEHOLD EXPENSES
(FOR ALL INTERVIEWEES)

75. How many people live in this home? _____

76. How many are children? _____

77. Age, Sex e educational level of each household member.

Member 1: Age? _____ Sex? () 1. F () 2. M Years of schooling: _____

Member 2: Age? _____ Sex? () 1. F () 2. M Years of schooling: _____

Member 3: Age? _____ Sex? () 1. F () 2. M Years of schooling: _____

Member 4: Age? _____ Sex? () 1. F () 2. M Years of schooling: _____

Member 5: Age? _____ Sex? () 1. F () 2. M Years of schooling: _____

Member 6: Age? _____ Sex? () 1. F () 2. M Years of schooling: _____

Member 7: Age? _____ Sex? () 1. F () 2. M Years of schooling: _____

Member 8: Age? _____ Sex? () 1. F () 2. M Years of schooling: _____

78. What is the average income per month for each member?

Member 1: R\$ _____

Member 2: R\$ _____

Member 3: R\$ _____

Member 4: R\$ _____

Member 5: R\$ _____

Member 6: R\$ _____

Member 7: R\$ _____

Member 8: R\$ _____

79. Any family member receives governmental financial aid, from friends, or from any other source (please, describe)?

Member 1: () 1. Yes () 2. No - Which? _____ How much: R\$ _____

Member 2: () 1. Yes () 2. No - Which? _____ How much: R\$ _____

Member 3: () 1. Yes () 2. No - Which? _____ How much: R\$ _____

Member 4: () 1. Yes () 2. No - Which? _____ How much: R\$ _____

Member 5: () 1. Yes () 2. No - Which? _____ How much: R\$ _____

Member 6: () 1. Yes () 2. No - Which? _____ How much: R\$ _____

Member 7: () 1. Yes () 2. No - Which? _____ How much: R\$ _____

Member 8: () 1. Yes () 2. No - Which? _____ How much: R\$ _____

80. Any family member receives non financial aid, from friends, or from any other source (please, describe)? (food, cloth, transport, etc)

Member 1: () 1. Yes () 2. No - Which? _____ How much: RS _____

Member 2: () 1. Yes () 2. No - Which? _____ How much: RS _____

Member 3: () 1. Yes () 2. No - Which? _____ How much: RS _____

Member 4: () 1. Yes () 2. No - Which? _____ How much: RS _____

Member 5: () 1. Yes () 2. No - Which? _____ How much: RS _____

Member 6: () 1. Yes () 2. No - Which? _____ How much: RS _____

Member 7: () 1. Yes () 2. No - Which? _____ How much: RS _____

Member 8: () 1. Yes () 2. No - Which? _____ How much: RS _____

81. How much did you spend in the last 30 days for the items below:

1. Electricity: RS _____

2. Water and sewage system: RS _____

3. other taxes: RS _____

4. Gas: RS _____

5. fuel: RS _____

6. Rent: RS _____

7. Telephone: RS _____

8. Medications: RS _____

9. Clothes: RS _____

10. Education: RS _____

82. How much did you spend in the last week for:

1. Transport: RS _____

2. Food: RS _____

83. How much do you normally spend on average each month for:

1. Transport: RS _____

2. Food: RS _____

<i>ITEMS</i>	<i>QUANTITY</i>	<i>YEAR</i>
STOVE	_____	_____
FREEZER	_____	_____
REFRIGERATOR	_____	_____
WASH MACHINE	_____	_____
DISH MACHINE	_____	_____
TOASTER	_____	_____
VACUM CLEANER	_____	_____
IRON	_____	_____
COLOUR TV	_____	_____
BLACK/WHITE TV	_____	_____
STEREO	_____	_____
RADIO	_____	_____
VIDEO CASSETE	_____	_____
DVD PLAYER	_____	_____
AIR CONDICIONER	_____	_____
FAN	_____	_____
CAR	_____	_____
BIKE	_____	_____
MOTOBIKE	_____	_____
COMPUTER	_____	_____
MICROWAVE OVEN	_____	_____

Appendix 2– Information sheet and consent form



Universidade Federal da Bahia (Federal University of Bahia)
Instituto de Saúde Coletiva (Public Health Institute)
 Rua Pe. Feijó, 29, Canela - 40.110-170 (Address and telephone/fax)
 Salvador-Bahia-Brasil- Tel/Fax: 336-8896

Information Sheet

Dear Sir or Madam,

We are conducting a study about the choice of household for toilet facilities and types of sanitation connection, in the city of Salvador. It is an important evaluation to understand the process of the household choice, and an important tool to local government to define sanitation policy.

We will wish have your support to answer a questionnaire. It is completely voluntary. All information that you give will be treated in strict confidence. Your answer will help us to evaluate ways to understand demand for sanitation.

For any question about this research, please contact: **Andreia C. Santos**, Institute of Collective Health, Federal University of Bahia, **telephone: 336-8896**.

And refers to: "INFORMATION ABOUT DEMAND FOR SANITATION PROJECT"

Interviewer: _____ Date: ____/____/____



Universidade Federal da Bahia (Federal University of Bahia)
Instituto de Saúde Coletiva (Public Health Institute)
 Rua Pe. Feijó, 29, Canela - 40.110-170 (Address and telephone/fax)
 Salvador-Bahia-Brasil- Tel/Fax: 336-8896

Consent Form

(in two copies: one for the interviewee and other for the interviewer)

Project: Demand for sanitation in Salvador/Brazil

Dear Sir or Madam,

We will wish have your support to answer a questionnaire about the choice of household for toilet facilities and types of sanitation connection, in the city of Salvador. It is completely voluntary. All information that you give will be treated in strict confidence. Please, fill the appropriated required information below.

"I (please, write your full name) _____ have read the information sheet concerning this study (or have understood the verbal explanation) and I understand what will be required of me and what will happen to me if I take part in it".

"I understand that at any time I may withdraw from this study without giving a reason and without affecting my normal care and management".

"I agree to take part in this study".

Signature _____ Date: ___ / ___ / ___

For any question about this research, please contact: **Andreia C. Santos**, Institute of Collective Health, Federal University of Bahia, **telephone: 336-8896**.

And refers to: "INFORMATION ABOUT DEMAND FOR SANITATION PROJECT"

