FORECASTING

THE DEMAND FOR

CHILD CARE ARRANGEMENTS:

A NEW STATISTICAL TECHNIQUE

GORDON CLEVELAND

Funding from the Child Care Initiatives Fund of the Department of National Health and Welfare is gratefully acknowledged. Additional copies of the Handbook may be obtained from CHILD CARE 2000, 297 St. George St., Toronto M5R 2P8 or from the Department of National Health and Welfare in Ottawa. Cette publication est aussi disponible en francais. Addressez-vous au CHILD CARE 2000, 297 St. George St., Toronto M5R 2P8 ou au Ministère de la Santé et du Bien-Etre à Ottawa.

The views expressed herein are solely those of the author and do not represent the views or policy of the Department of National Health and Welfare or of DIRECTION 2000.

PROJECT SPONSORED BY
DIRECTION 2000: DEVELOPING COMMUNITY DAY CARE
TORONTO

APRIL, 1992

© Gordon Cleveland

TABLE OF CONTENTS

LIST OF TABLES

LIST OF CHARTS

ACKNOWLEDGEMENTS

The Child Care Initiatives Fund of The Department of National Health and Welfare has supported the research, writing and production of this Handbook from the beginning. Sharon Gribbon, Project Officer, and Carol Levesque, Financial Officer, have been unfailingly helpful in shepherding this project through the bureaucratic maze, and deserve my warm thanks.

Doug Hyatt wrote Appendix A on Computer Software and Hardware, made important contributions to Appendix B, and has provided insightful comment on the entire penultimate draft. Morley Gunderson, Gordon Anderson, and Michael Denny have helped improve my understanding of logit models and their interpretation. They should not be held responsible for my use or abuse of this knowledge.

A Project Committee from various parts of the country has met twice, has reviewed and commented on drafts of the material. Their contributions have been collectively and individually invaluable and have made this Handbook a different, and more readable, product than it would have been. Lyse Frenette from Montreal, Brian Powell from Ottawa, Jane Bertrand and Petr Varmuza from Toronto, Mary Humphrey from Winnipeg, and Peter Ashmore from Vancouver deserve considerable credit for their unremunerated but not unappreciated labours. My thanks, too, to Sonia Ostrowska and Penney Kirby -- who helped organize and staff the Project Committee meetings and who laboured secretarially. Sue Colley provided advice, criticism, and much secretarial labour, too.

Gilbert Bélisle has spent more time than perhaps he would have wished learning the vocabulary of statistics and economics in order to transform a document written originally in English into a very natural French translation. He has been unfailingly efficient and co-operative in this endeavour.

This project, formally called The Child Care Demand Forecasting Project, has been sponsored by DIRECTION 2000, a non-profit organization which develops new day care services and provides research and other services which contribute to the development of day care. Lorraine Duff, as the Executive Member in charge of the project, has kept me on the

straight and narrow way, monitoring the progress of the Handbook, signing cheques as needed, and ensuring the provision of reports to the Board from time to time. Evelyn McKee has handled the financial side of my relationship with DIRECTION 2000 with infinite patience and good humour and considerable hours of labour. Sharron Richards and the rest of the Board of DIRECTION 2000 deserve credit for their willingness to champion a project which must have seemed occasionally to be only distantly related to the concrete task of building a better system of day care for our children. I hope the result will justify their trust.

EXECUTIVE SUMMARY

Many people agree that Canada's licensed child care system should be expanded dramatically to meet needs which are now inadequately served. To plan this expansion, we need to know what kinds of families want to use which types of care, and how this will change over the years ahead. In addition, we want to know how government policies to change the affordability, availability and quality of care will affect the use of different types of child care.

None of the methods presently used to study and forecast the demand for different types of child care is able to measure how parents' child care choices will change when prices, incomes, availability, quality, family size, and other factors change. This handbook is devoted to describing and explaining one method of doing just that.

This handbook presents and explains a new technique for measuring the demand for child care of different types (e.g., day care centre, neighbourhood sitter, care by a relative). This technique (logit regression analysis) is complex because it tries to do a good job measuring something which is complex -- the demand for child care. This handbook is designed to be a non-mathematical introduction to this technique for the non-specialist.

It is useful to examine the differences between "demand" and "preference" or "need". Some studies try to determine what kind of child care families *want* to use or *wish* they could use. In the child care literature, this is called studying families' child care "preferences". But these preferences may or may not ever be realized. If prices are too high, if services are inconvenient, or if the available quality is poor, a family may never actually decide to use the care it would prefer. The key question to address is "Under what circumstances will a family's wishes actually turn into a decision to use a particular form of child care?" A study of "demand" will give us an answer to that question.

Some studies try to determine the total *need* for child care. There may well be a large number of families in Canada who "need" licensed day care, because this would really be the best solution to their child care problems. Measuring the number of families who "need" child care requires a social definition of need. Maybe all families with working

parents "need" day care; maybe only families having working single parents "need" day care; whatever our definition of "need", such studies do not inform us about how many families will actually decide to use day care, under given circumstances.

SIX STEPS

There are six steps involved in this new method of measuring the demand for different types of child care; in brief, the steps are:

1. Work out a model (a hypothesis, if you prefer) of how families make child care decisions;
2. Collect data on a large group of families that will allow you to test this model;
3. Figure out what statistical technique will best allow you to test your model;
4. Estimate your model using the appropriate computer software and hardware;
5. Use various statistical tests to figure out whether your results are acceptable, and in what ways your model might change to get better results;
6. Use your final estimates to forecast future demand for different types of child care and/or analyze the impact of alternative government policies on families' child care choices.

These steps are explained in detail in the handbook. Further, a concrete example, analyzing the child care choices of families with employed mothers in Metropolitan Toronto in 1976, is presented. A number of tables are used to show how the estimates from our mixed logit estimation can be used to forecast the effect of changes in different variables or to analyze the effects of various government policies. A chapter is devoted to examples and discussion of statistical tests which can be used to judge whether the results of our study of child care choice are statistically valid or not.

Forecasting the demand for child care also requires a forecast of the number of mothers with young children in the paid labour force. We review current methods of producing such forecasts and possible directions for future research.

1

INTRODUCTION

Almost everyone agrees that Canada's licensed child care system needs to be expanded dramatically to meet needs which are now inadequately served. To plan this expansion, we need to know what kinds of families want to use which types of care, and how this will change over the years ahead. In addition, we want to know how government policies to change the affordability, availability and quality of care will affect the use of different types of child care.

Our existing techniques for measuring the demand for child care do not allow us to do forecasting and policy analysis of this kind. The one described in this handbook does.[1]

This handbook presents and explains a new technique for measuring the demand for child care of different types (e.g., day care centre, neighbourhood sitter, care by a relative). This technique is complex because it tries to do a good job measuring something which is complex -- the demand for child care. However, this handbook is designed to be a non-mathematical introduction to this technique for the non-specialist. Mathematical and technical details (except for one rather innocuous equation) are confined to the footnotes and appendices.

---

[1] Economists are usually careful to use the terms "projections" and "projecting" instead of "forecasts" and "forecasting" to describe what their models will produce. This terminology is intended to remind the user of two things:

(1) a projection assumes that the behaviour patterns of consumers will remain relatively stable. If this assumption is incorrect, the projection will be inaccurate;

(2) the model (in this case, our model of child care demand) only refers to a small segment of the economy, and makes no predictions about what will happen in the rest of the economy.

We have used the term "forecasts" because it is more understandable for a non-specialist audience, but the qualifications mentioned above still apply.

WHO WILL WANT TO USE THIS HANDBOOK?

This study describes a method of measuring the impact of high prices, poor availability and inadequate quality on the child care choices families make.

This handbook will be very useful to anyone in federal, provincial or municipal governments or in a university who plans to do a study of the amount of child care needed and wanted in a particular geographic region, or by a particular group in the population. This handbook reviews a number of techniques for studying the demand for child care before recommending and discussing in detail a particular one. It also reviews a wide range of academic and other literature on the demand for child care and reviews computer programs available to estimate demand. If you are planning to do a study of the demand for child care, this handbook is for you.

This handbook will also be helpful to policy makers, child care advocates, parents or child care workers who are called on from time to time to read, interpret or criticize what others have written (or assumed) about the demand for child care in this country and how we should go about satisfying it. This study describes a method of measuring the impact of high prices, poor availability and inadequate quality on the child care choices families make. It describes the difference between measuring "demand" and measuring "need". And it gives a concrete example of how to interpret the results of a study of child care choice.

DISCRETE CHOICES

These techniques are not designed only to measure child care choices; they can be adapted to apply to almost any choice families make between discrete alternatives. "Discrete" means clearly distinct or separate, like the choice between owning a house or not owning a house; or the decision to drive a car, take a bus or walk to work. Alternatives such as buying ten litres of milk a month rather than thirteen, seventeen, or thirty litres are better viewed as "continuous" choices rather than "discrete" ones. For continuous choices, you ask "how much?"; for discrete choices, you ask "which one?" Continuous choices require different statistical techniques.

To illustrate the application of the techniques described in this handbook, we examine the choice made by mothers in the paid work force between a day care centre, a sitter or care by a relative.[2] The techniques can be adapted to analyze a different or broader range of choices; similarly, they could be adapted to analyze the two-stage decision of women to work or not to work and then the decision to choose a particular type of child care arrangement. Chapter 6 discusses some of these possibilities; until then we will assume that the choice between a day care centre, a sitter and care by a relative is the one that interests us.

CONTENTS OF THE HANDBOOK

Chapter 2 reviews other ways of measuring preferences, needs and demand for child care. In particular, it looks at what federal, provincial and municipal governments do when they want to measure the demand for day care. Chapter 3 gives a general description of a more reliable statistical technique for measuring child care choice. Chapter 4 provides a concrete example, based on data from Metropolitan Toronto, of an estimate of the influence of various factors on child care choice. Chapter 5 describes the use of various statistical tests to judge and to improve estimates of child care choice. Chapter 6 describes and assesses different techniques for forecasting the number of mothers in the labour force who need child care. This may be separate from, or integrated with, our forecast of child care choice. Chapter 7 briefly concludes and summarizes the handbook. Four appendices contain information which is too specialized for the general reader, but valuable for those more interested. Appendix A evaluates packaged computer programs for logit regression and related software and hardware requirements. Appendix B describes how the 1988 National Child Care Survey data might be used to do a study of the demand for child care as described in this Handbook. Appendix C summarizes economic literature on the choice between different types of child care service. Appendix D provides some details about the statistical model used for estimating child care choice. A bibliography of relevant source materials and a brief glossary conclude the Handbook.

[2]Licensed family home day care is an important alternative type of child care. However, in the Project Child Care data set which we use in Chapter 4, licensed family home care is not distinguished as a separate alternative.

2

NEED, PREFERENCE, AND DEMAND: CURRENT METHODS OF MEASURING THE DESIRE FOR ADDITIONAL CHILD CARE FACILITIES

Child care studies have used a bewildering array of different concepts to describe and measure the amount of child care that should be available. Some measure the "need" for child care; others talk about the "needs and preferences" of parents. Some others purport to study the "demand" for child care; a final group measures the number of children currently using each type of child care. Many people use the terms as if they were almost interchangeable; some studies try to measure various concepts and discuss the relationships between them.

Getting these terms straight is the first step in figuring out what child care parents want, need and will actually use. This chapter will discuss how these different terms have been used and the techniques that have been used to measure them. Then it will explain why "demand" is the right concept if we are interested in forecasting the future use of different types of child care and thinking about how government policies can affect parents' choice of child care arrangements.

CURRENT USE

A survey of the types of child care currently being used by parents (such as Statistics Canada, 1981) tells us how many families have, under existing circumstances of price, quality and availability of care, ended up choosing which types of child care, for how many hours a week, and so on. It gives us a snapshot of child care choices at one point in time. By itself, this type of study does not give us any information about parents' needs, preferences or demand for child care. It does not tell us what choices parents will make in the future, or how parents' choices of child care arrangements would change if conditions were different. However, the data collected on current use might be the raw material for a study of demand such as described in this handbook. Combined with other

information or assumptions about need, preference or demand, data on parents' current child care choices can be very valuable.

NEED

In common usage, there are two meanings of the term "need". Parents who really *should have* access to child care services can be described as needing child care. These parents, because of their employment status, income, marital status, or other reasons are "needy"; they *should have* (a particular type of) child care. Another use of the term "need" has nothing to do with whether parents are socially deserving or not. Parents who *would like* to have access to (a particular type of) child care but cannot find a space for their child (perhaps there are no facilities at all, perhaps they are unaffordable or of too low quality) could be described as "needing" child care and not being able to find it. We use the term "need" to refer to the first meaning above, and "preference" to refer to the second (both those who find the type of child care they want and those that do not).

A study of child care needs will define the types of families who have a legitimate social need for child care and then estimate how many such families exist. A concrete example can help to make the point. Bali Ram, from the Demography Division of Statistics Canada, has done a small study of child care need in Canada up until the year 2001. He makes his definition of "need" admirably clear:

> *The need is measured by the number of children below age 15, whose mothers do not have sufficient time or resources to provide them with adequate care; they either choose or are forced to depend on extra-parental care....*
>
> *(i) children in lone-parent families, and*
>
> *(ii) children in husband-wife families, where the mother is in the labour force.* (Ram, *1987*, pp.2-3)

Having defined need, Ram uses Statistics Canada projections about fertility and lone parent families, and assumptions about labour force participation of women, to produce a couple of different scenarios about how many children will need child care in the future. The work is well done and interesting.

However, such studies are frequently misinterpreted. It would be wrong to interpret Ram's figures as a measure of the total number of children whose parents want to use child care (otherwise it would include some children of student parents and of parents not in the paid labour force). It would be equally erroneous to interpret his figures as a measure of the number of children who are likely to actually use purchased child care services (many of the children with employed mothers will be cared for by fathers or other relatives; many lone parent mothers care for their own children at home; many children over ten years of age care for themselves). Rather, Ram's study, and other similar studies of child care need, calculate a ballpark estimate of the number of children who are in family situations or types of families that have (in the past) been likely to use some form of non-maternal child care. It answers the question "how much will the pool of *potential child care users* (those who we believe should have access to child care services) grow or shrink over the next number of years?"

Perhaps the most widely quoted measures of "need" for licensed day care are those published annually in *"The Status of Day Care in Canada"* from the Department of National Health and Welfare. Each year, the Department calculates the number of full day licensed day care spaces available across Canada in various age groups (which is equivalent to current "use" if all spaces are filled with children for the full day). The number of day care spaces is compared to the number of children who might be considered to need day care. Four alternative definitions of "need" are used: children with mothers in the labour force; children with full-time working parents; children with parents who are working full-time or are students; children with parents who are working more than 20 hours a week or are students.[3]

This yields alternative calculations of the percentage of children in each age group who can be served by the existing supply of day care spaces. One hundred percent minus the percentage of children served in each age group can be viewed as a measure of "the degree of unmet need."

[3]Lero et al. (1985) presents another, and broader, view of the legitimate categories of "need" for child care services.

Status of Day Care in Canada: 1988 carefully points out that:

> *It should not be construed to mean that these [i.e., the four alternative definitions of need] are the numbers of children whose parents would choose a formalized day care setting if it were made available. (1988; p.4)*

In other words, this is not a measure of the "demand" for day care; it tells us nothing about how many children would use day care if conditions were to change.

PREFERENCES

Status of Day Care in Canada does venture an opinion (based on survey information) about parental preferences:

> *The existing evidence of parental preference is sketchy at best, but it does seem to suggest that about 50% of those not using formalized care would do so if provided the opportunity. However, there are a number of variables that need to be researched before definitive statements can be made.* (1988; p.4)

Survey information about parental preferences can be very interesting; it is one way of asking "what if" questions that move beyond the confines of the present child care situation. The trouble is that no one knows whether (or how much) to trust the information collected; stated preferences are not necessarily a good guide to what parents will actually choose.

For instance, if 50% of parents say they would use licensed day care if it were available, what question are these parents really answering? Did "available" mean free of charge, or at regular market prices, or at the full price if fair wages were paid to staff? What quality level was assumed when parents answered the question? Just how convenient does "available" mean? Every parent answering such a survey will make different assumptions. Of course, you can ask more detailed questions about preference and that will help to sketch a pattern of attitudes and preferences, but the more hypothetical the questions are, the more difficult it will be for parents to provide accurate answers.

Laura Johnson's (1977) study *Who Cares? A Report of the Project Child Care Survey of Parents and Their Child Care Arrangements in Metropolitan Toronto* illustrates the complexities of interpreting "preference" information as an index of what parents would actually choose. Asked "What would your ideal daycare arrangement be?", nearly 50% of parents in that survey chose a day care centre, about 30% chose some kind of sitter, about 20% chose care by a relative.

The same parents were also asked to rank the child care alternatives they would consider using if they were engaged in choosing a new child care arrangement and they had the choice between a day care centre, a sitter and a relative at different distances from their home. It turned out that distance from home was the key factor in parents' preferences: any type of care two miles from the child's home was preferred to any type of care five miles from the child's home, and so on. Further, at any particular distance, care by a relative was ranked highest, day care centres second and sitters third. So, what should we say this group of parents really prefers? Is it day care centres (their ideal arrangement according to the earlier question), care by a relative (the most likely to be chosen at any particular distance), or whatever type of care is closest to the child's home? The survey of parental preferences would give some support to each answer.

Of course, it is possible to rationalize the variation in these answers to different questions and Johnson does so quite capably. This example shows, however, that "preferences" are affected by the conditions under which the alternatives are available. "Preference" information may be an unreliable guide to what parents will actually choose when faced with a particular price, level of convenience, set of quality attributes, in a particular family situation, and so on.

"Preference" information may be an unreliable guide to what parents will actually choose when faced with a particular price, level of convenience, set of quality attributes, in a particular family situation, and so on.

The point is that the correct answer to the question "How many children will use a day care centre (or a sitter, or care by a relative)?" is "It depends." Understanding this point is the key to understanding why it is important to measure "demand" rather than "preference" or "need".

DEMAND

The "demand" for a particular type of child care is the number of families who are "willing and able" to use a particular form of child care, and how their willingness and ability changes when conditions change.

We need a measure of parents' decisions to use child care that takes into account a number of the key factors upon which these choices depend. We need, in other words, to measure the "demand" for child care. The "demand" for a particular type of child care is the number of families who are "willing and able" to use a particular form of child care, and how their willingness and ability changes when conditions change.

Both the willingness and the ability are important. As family incomes change, as the prices of day care, sitters and care by a relative change, as the convenience and quality of different types of care change, we want to know what happens to families' actual decisions to use or not to use a particular form of child care.

Therefore, knowing the demand for day care means knowing how much effect each of a series of factors has on the number of families who will use a day care centre. Some factors will increase the demand for day care, others will reduce it. We need to know, on average, the amount of effect each factor has and whether this effect is positive or negative. When we know this, we can forecast how use of day care will change when any one or several of these factors change.

Perhaps it is useful to reiterate the differences between "demand" and "preference" or "need". "Demand" does not mean "want". It is interesting to study what kind of child care families *want* to use or *wish* they could use. In the child care literature, this is called studying families' child care "preferences". But these preferences may or may not ever be realized. If prices are too high, services are inconvenient, or the available quality is poor, a family may never actually decide to use the care it would prefer. The most interesting question is "Under what circumstances will these wishes actually turn into a decision to use a particular form of child care?" A study of "demand" will give us an answer to that question.

Neither is demand the same thing as "need". There may well be a large number of families in Canada who "need" licensed day care, because this would really be the best solution to their child care problems. Measuring the number of families who "need" child care requires a social definition of need. Maybe all families with working mothers "need" day care; maybe only families having working single parents "need" day care; whatever our definition of "need", it tells us nothing about how many families will actually decide to use day care.

GOVERNMENT MODELLING OF CHILD CARE NEEDS, PREFERENCES AND DEMAND

Provincial Governments

In the last several years, l'Office des Services de Garde à l'Enfance (the government agency responsible for child care in Quebec) has worked on developing a model to forecast future child care needs, using some information on parental preference as well. There are four major elements in the model: the number of children in each age group; the labour force participation rate of mothers; the average portion of a full time day care space used by each child, and parental preferences for different types of care. The number of children in each age group is derived from Census data[4] for each region and sub-region in Quebec. The labour force participation rates are based on current averages for mothers of young children. The space usage information is calculated from l'Office data. The parental preference information was originally based on a 1982 survey conducted across Quebec by the University of Montreal. More recently, l'Office has funded two surveys, in 1986 and 1989, to collect its own information on parental preference and current use patterns. The latter survey gives statistically valid data on parents' stated child care preferences for the various regions and sub-regions of Quebec. Preferences vary according to the age of the child; these preferences by age are applied to the rest of the data to get forecasts of the number of child care spaces the Quebec government should make available over the coming years.

The model is a useful planning tool, and more sophisticated than anything used in other provinces. It incorporates one major factor upon

[4] Family Allowance data will be used in future projections.

which parents' choices depend -- the age of the child. Yet, as a forecast of parents' actual choices, it is still subject to important limitations, because it does not include other factors (such as price, availability, quality, and various family characteristics) that also appear to affect families' child care choices. In effect, it assumes that these other factors have negligible impact on the demand for each type of care.[5]

Other provinces have not developed demand, use or need projections in other than a back-of-the-envelope way. Alberta and Ontario both have fairly detailed, child care information collection systems. Alberta has done some work attempting to forecast the number of families who will apply for and receive subsidy in any month. Ontario has not yet developed any child care demand or use projections.

The Federal Government

The federal government does not model child care demand and only predicts future needs or preferences based on estimates provided by provinces and territories. The Women's Bureau of the Department of Labour (in 1968 and 1975) and Statistics Canada (in 1981) have conducted statistical studies of current child care use. Statistics Canada, the Department of National Health and Welfare, and some provincial governments have recently co-operated with a team of university professors to conduct a major new survey of child care use and parental preferences. The results of the 1988 National Child Care Survey are not yet available. Appendix B discusses how this data might be used to measure the demand for different types of child care.

Municipal Governments

There are two situations in which some municipalities have been called on to study and forecast the demand for child care. Where municipalities are mandated (as in Ontario), or wish (as in Alberta) to become involved in providing, financing, and regulating the provision of child care services, they may need to project future demand for child care. For instance, Metropolitan Toronto's Day Care Planning Task Force published in 1986 a calculation of future child care use, perhaps based on Johnson (1977). *Blueprint For Child Care Services: Realistic*

---

[5]Because the Quebec surveys collect information on current type of child care used by each family, as well as its preferences, they permit useful comparison of stated preferences with actual use patterns.

Responses to the Need assumed that half of all children with working parents would use licensed services: "The limited research available suggests that 50% of parents prefer some kind of formal, licensed child care arrangement."

Municipalities may also need to assess the demand for child care if they adopt a linkage ordinance which compels residential or nonresidential developments to offset the additional child care demands which they create. San Francisco initiated the use of such ordinances to create additional day care supply in the mid-1980's. Since then, Milwaukee, New York, Chicago, Boston, Contra Costa County, San Ramon, Danville, and Concord (California) have passed some type of linkage ordinance and Fremont, Sacramento, and the State of Massachusetts are considering them. Both Vancouver and the City of Toronto are studying these American developments. Most of these ordinances use simple rule-of-thumb methods of forecasting the impact of any development on the demand for child care services.[6]

NEEDS ASSESSMENT STUDIES

The fundamental job of needs assessment studies is usually to discover the number of potential users of child care.

Needs assessment studies are a popular tool for groups seeking to assess whether a new child care facility or additional public funding of child care is needed. Questions asked on such surveys vary widely, as do methods of presenting and analyzing the results[7]. Typically, they ask some mixture of needs and preference-type questions. The fundamental job of needs assessment studies is usually to discover the number of potential users of child care. Accordingly, the major

[6]These are reviewed in Cleveland (1990b) and Eichman, Reisman, and Euben (1990).

[7]Recently, a manual and draft questionnaires designed specifically for studying child care needs have been published in Ontario. See Friendly (1989).

questions asked in needs assessment surveys seek to identify how many children, of what ages, for what hours, currently use or will use in the near future some form of non-parental child care. Sometimes, the concern is with a particular form of care like evening and overnight care or rural care.[8] Then the major concern is to identify potential users of this type of care.

Needs assessment studies usually ask additional questions designed to gather information about parents' preferences and satisfaction with different types of child care. For instance, parents may be asked about problems with the cost, convenience, and quality of the type of child care services they are currently using and the form of child care they would prefer to use. Needs assessment studies can be useful[9] but they are distinct from studies of the demand for child care. To adapt a lesson from the Rolling Stones' song, parents do not always get what they want when it comes to child care services. Nor, even if they "try real hard", do they always get what they need. So even a very careful study of parents' preferences and needs does not necessarily tell us very much about what child care services parents will actually use. Nor does a needs assessment study answer the question which is really at the heart of a study of the demand for child care: "When parents can't get what they really want, exactly what effect on their final child care choice will each constraint (cost, lack of availability, poor quality, family income) have?"

None of the methods presently used to study and forecast the demand for different types of child care is able to measure how parents' child care choices will change when prices, incomes, availability, quality, family size, and other factors change. This handbook is devoted to describing and explaining one method of doing just that.

[8]See George Brown College's (1990) needs assessment study for a good example of the use of survey and other information to determine which type of extended hour child care facility should be provided, during what hours of the day, and how many children of health care workers would be likely to use it.

[9]Needs assessment studies are particularly useful in assessing the likely use of a particular type of proposed child care service in a well-defined neighbourhood (or in a well-defined user group). In such cases, projections based on statistical averages may be less reliable than the results of a well-designed needs assessment study.

In forecasting the demand for child care services of different types and in analyzing the effect of policy changes on child care demand, we are not interested in what kind of child care parents wish they were using; rather, we are interested in what kind of child care parents will actually decide to use "under the circumstances". We are also interested in discovering what will happen to the actual choices parents make when these circumstances change. (The information about what parents wish they were doing may, of course, be relevant for policy; it is not really relevant for forecasting, however).

None of the methods presently used to study and forecast the demand for different types of child care is able to measure how parents' child care choices will change when prices, incomes, availability, quality, family size, and other factors change. This handbook is devoted to describing and explaining one method of doing just that.

3

A NEW METHOD OF MEASURING AND FORECASTING PARENTS' CHOICE OF CHILD CARE ARRANGEMENTS: A GENERAL OVERVIEW

There are six steps involved in this new method of measuring the demand for different types of child care; in brief, the steps are:

1. Work out a model (a hypothesis, if you prefer) of how families make child care decisions
2. Collect data on a large group of families that will allow you to test this model
3. Figure out what statistical technique will best allow you to test your model
4. Estimate your model using the appropriate computer software and hardware
5. Use various statistical tests to figure out whether your results are acceptable, and in what ways your model might change to get better results
6. Use your final estimates to do forecasting and policy analysis

The rest of this chapter will be occupied with explaining these six steps more fully. First, we will go through them in just a bit more detail, then again quite thoroughly. The following chapter provides a concrete illustration of each of these steps using the Project Child Care data set from Metropolitan Toronto.

IN BRIEF

The place to start when trying to measure child care demand (Step 1) is to answer the who, what, how and why questions. Who chooses child care? What are the main distinct types of child care from which each family chooses? How does a family make the decision? Why do families make the choices they do (i.e., what factors influence their decisions)? The answers to these questions will make up our model of how the choice of child care arrangements is made.

This model, when put in mathematical form (still Step 1), can then be tested against the data on child care choice collected from a wide variety of families (data collection is Step 2). The data will best represent the entire population if it is chosen randomly. Of course, we must collect data on all the factors that our model tells us are important in influencing child care choices.

The statistical technique we use allows us to calculate the unknown parts of our model (called "estimating the values of unknown coefficients" in the model). This handbook argues in favour of using "logit regression" (Step 3 is choosing the most appropriate statistical technique).

Step 4 is actually performing the estimation of the unknown coefficients in the model. To do this we need to use a fairly powerful computer with the appropriate computer software; Appendix A makes recommendations on what you should use. Our "estimate" consists of a group of numbers, which will help us to predict the demand for child care.

Getting good "estimates" is part science and part art.

There are a number of statistical tests (Step 5) which can be used to help judge whether our "estimated model" (i.e., when the estimated coefficients are slotted into the mathematical form of the model) provides a good explanation of the observed data or not. We can try various amendments of our original model to see whether the model can be improved.

Finally (Step 6), our amended "estimated model" can be used to analyze the effect of certain changes on the child care choices families will make. We assume that the weight of different factors in child care choice remains the same (i.e., our estimates are still valid), but as the amount of family incomes or day care prices or other factors change, it affects the choices made. By slotting different predicted values of, for instance, family income into our data set, we can calculate new predicted values of demand for each type of child care.

STEP 1: THE MODEL

Who Chooses?

Individual families make their own choices about child care arrangements, based on their own needs, desires and financial capabilities. Mothers are likely to have a larger role in the decision than fathers in many families -- this will be reflected in our review of the factors which affect child care choice -- but the decision is influenced by the overall economic and social characteristics of the family, not just of the mother.

Most studies of child care choice concentrate only on families having employed mothers, rather than mothers both inside and outside the paid work force. It seems likely that the factors affecting the child care choices of families with employed mothers are influenced by different factors than the child care decisions of families with a stay-at-home parent. By implication, if we want to look at the child care choices of these two groups of families, we should really develop and test two models of child care choice rather than just one.

Data from the 1981 *Statistics Canada* "Survey of Child Care Arrangements" seem to confirm this. Families with mothers who are not engaged in paid work do not typically use child care for any substantial number of hours per week; if child care is used it is likely to be unpaid. Sixty-five percent of families with mothers outside the paid labour force do not use any form of non-parental child care for their preschoolers. Of the remaining 35%, more than half use free public kindergarten, and a majority of the remainder have their children cared for in their own homes, in most cases by a live-in relative. The average number of hours of care received by preschool children of mothers not in the paid labour

force (excluding non-users) was only 13 hours a week. In contrast, the average number of hours of care received per week by preschoolers whose mothers were employed either part-time or full-time was slightly over 30 hours. This evidence suggests modelling child care choice as the decision of families having employed mothers, or the use of different models for different types of families.

What is Chosen?

In the concrete example of the next chapter, and in a number of other child care studies, the child care alternatives are defined as day care centre, sitter, and care by a relative (where "relative" includes the father of the children).

An alternate version of our model could separate care by fathers from care by other relatives. Studies by Yaeger (1978) and Hofferth and Wissoker (1990) have done this successfully, finding that some factors have different influences on these two types of child care.

The decision about what types of child care should be included in the model is affected by what we, as the researchers, want to find out from our work. If a major focus of policy interest is, for instance, the choice between profit and non-profit day care centres (see Lero, 1981), then our model (and the design of our questionnaire) must reflect this. Otherwise, we should restrict our model to a small number of fairly clear and distinct alternatives.

How Is The Choice Made?

How does a family decide what type of child care to use? Essentially, a family will choose the child care alternative which seems most beneficial, given the family's current situation and child care circumstances. Economists describe this as choosing the alternative which gives the greatest amount of utility to the family; this terminology is convenient, so we will use it. In this framework, what we are doing is modelling the utility of each type of child care to the family, and then comparing the total utility from each.

Why Choose This Type of Child Care?

Our model must identify the main factors which will affect the utility of each type of child care to a family. Existing research can help us with this. We know, for instance, a lot about why some parents use day care and others decide to use other arrangements. The Katie Cooke *Task Force on Child Care* summarized our current knowledge in this way:

> ***...selecting a child care arrangement is a complex decision-making process involving numerous variables. Parents choosing child care consider the age, number and health of their children; the availability and cost of different types of care in their neighbourhood or work area; the suitability and availability of various work schedules (full-time, part-time); and the level of family income. (Cooke, et al. 1986, pp.91)***

A review of empirical studies of child care choice (see Appendix C) confirms that four groups of variables are important to the child care choices families make: attributes of the different types of child care service, such as price, convenience and quality; child variables, such as the number and ages of children in the family; mother's employment variables, such as the number of hours the mother is engaged in the paid work force and her wage; and various socioeconomic variables which reflect the specific needs, preferences or financial ability of the family.

four groups of variables are important to the child care choices families make: attributes of the different types of child care service, such as price, convenience and quality; child variables, such as the number and ages of children in the family; mother's employment variables, such as the number of hours the mother is engaged in the paid work force and her wage; and various socioeconomic variables which reflect the specific needs, preferences or financial ability of the family.

The Mathematical Form Of The Model

As noted in the last chapter, the demand for a particular type of child care is the number of families who are willing and able to use that type of care, and how their choice of child care arrangement changes when conditions change. Demand is therefore not *one number;* it is not really meaningful to simply say that the demand for day care is 250,000 spaces. Instead the demand for day care is a relationship between numbers; the demand for day care is 250,000 spaces when family incomes are such-and-such, the price of day care is such-and-such and the work hours of mothers are such-and-such.

The easiest way to capture this kind of relationship is with a mathematical equation. (Many people have developed a severe allergy to anything mathematical. This is the only equation in the main text of this handbook; spending some time on it is well worth while.)

Imagine writing a mathematical equation for the utility a family gets from using a particular type of child care (our model will require one for each type of child care available):

$$y = bx + az$$

where each letter in the equation stands for a number. In this case, the "y" (the dependent variable) represents the thing we are trying to explain -- e.g., the utility given by use of a day care centre -- and the "x" and "z" (the explanatory variables) represent the factors that explain why a family chooses this arrangement. In the equation, "b" is multiplied by "x"; "a" is multiplied by "z"; the sum of "bx" and "az" equals "y". "x" might be the number of hours worked by the mother in this particular family; then "b" (called the coefficient of "x") represents the effect of each extra hour of work on the utility gained from day care. Similarly, "z" might be the weekly price of a day care centre. Then "a" (the coefficient of "z") measures the effect that each additional dollar cost of day care has on the utility gained from using day care.

This mathematical equation expresses in numbers what we have been, up till now, expressing in words when we discuss our model of child care choice; the choice of a type of child care (the relative value of "y") depends on certain factors (the "x's" and "z's"). For each family in our collected data set, we know whether the family chose a particular type

of child care or not and we know the value of the explanatory variables (mother's weekly hours of employment and day care price, in our example). We do not know the value of "a" and "b", which tell us how much influence each factor has on child care choice and whether the influence is positive or negative (i.e., increases or decreases the likelihood that this particular type of child care will be chosen). Our statistical techniques are designed to provide us with the best possible (i.e., statistically most likely) estimate of the "a's" and "b's" (the coefficients).

STEP 2: THE DATA

To estimate the values of all the different coefficients in the model (usually one for each explanatory variable), we need information on child care decisions collected from hundreds of families. We need to know which child care arrangement each family has chosen (and, if there are several arrangements, which is the main one); we need to know the attributes (price, availability, quality) of each type of child care from which they are choosing; we need to have information on a variety of other variables that affect the child care choices families make. All of this data must be included in the data set.

Why do we need information on hundreds of families? Our statistical technique gets its results by looking at how families behave under the different circumstances they face; we need to have a wide range of examples of different families (i.e., with different values of the explanatory variables) in order to calculate how each explanatory variable affects the choices that families make.

The data collected must, of course, be chosen in a way that makes it representative of the larger population from which it is chosen. Usually this means taking a random sample of all families, or of families having employed mothers. The sample could be a stratified random sample, with larger amounts of data collected from certain groups; the final results could then be weighted to correct for this oversampling. Alternatively, as in some studies of choice of different types of transportation, the sample could be choice-based; in this case, users of each type of care are sampled separately (perhaps from lists provided by caregivers) and the results are weighted to correct for this non-random sample design.

Often, estimates of the influence of different factors on child care choice are prepared using data already collected for a slightly different purpose by another group or agency. For instance, the estimates presented in Chapter 4 come from the Project Child Care data set. The data was collected to study child care choice (amongst other things) but the questionnaire was not designed ideally for our purposes.

Using existing data saves enormous amounts of time and trouble, but it has its own costs. The model in Chapter 4 does not measure the effects of quality on child care choice, and it imputes prices of the types of care not chosen by each family. Both of these modelling decisions were required by the lack of appropriate data in what is otherwise a very good data set, collected for different purposes.

STEP 3: THE STATISTICAL TECHNIQUE

The new statistical technique described in this handbook is called logit (pronounced LO-JIT) regression.
Sometimes the terms "mixed" or "multinomial" or "conditional" will appear before the word "logit"; these represent slight variations on the general logit model.

Regression analysis is the statistician's way of figuring out how much effect each of several explanatory variables has on a particular outcome, when they have all changed together. There are various kinds of regressions. The most popular are known as OLS regressions, or Ordinary Least Squares Regressions. The type we concentrate on here are called logit regressions, which are more appropriate for analyzing discrete choices (i.e.,"Which one?" rather than "How much?" choices). Whatever kind of regression you are doing, its purpose is to statistically calculate the influence of each of the explanatory variables on the dependent variable.

The new statistical technique described in this handbook is called logit (pronounced LO-JIT) regression.

A logit regression will give us a series of estimated coefficients (the "a's" and "b's" of our previous example). Each coefficient is simply a

number (e.g., 1.35 or 0.047). These coefficients are a measure of the average influence of the different explanatory variables on the relative probability that a particular type of child care will be chosen.

Technical details about logit regression analysis appear in Appendix D. A fuller account can be found in a number of places, including Ben-Akiva and Lerman (1985), Domencich and McFadden (1975), Maddala (1983), and Amemiya (1985). Details on the interpretation of estimated logit coefficients appear in Chapter 4 and in Appendix D.

There are two other issues that deserve discussion here. The first is the distinction between multinomial and conditional logit variables. The second is the importance of the Independence from Irrelevant Alternatives Assumption (I.I.A.) in the logit regression model.

There are two main ways in which explanatory variables can influence the utility that families get from different types of child care. Some variables, usually characteristics of the family doing the choosing, are the same no matter which type of care is chosen. So, for instance, in our model where families with employed mothers are doing the choosing, the ages of children in the family, the mother's wage and hours of work, the family income, etc. will not be different for the choice of day care, sitter or care by a relative. Other variables, particularly the attributes of each alternative, are different according to which choice is being considered. So, for instance, the price of care will be different for each of the alternative types of child care being considered. Likewise, the quality and convenience of each alternative will be different for the same family. If a different value of the explanatory variable is included for each alternative, it is a conditional logit variable.

The estimates given in the next chapter provide a good illustration of the differences between the two types of variable. The coefficient of the price variable in Chapter 4 (the "a" in our equation) is the same for every alternative type of child care. In that case, for conditional logit variables, the value of the explanatory variable changes while its estimated coefficient stays the same across different types of care. In contrast, for multinomial logit variables (all the other variables in the estimate in Chapter 4), the value of the explanatory variable stays the same while the estimated coefficients are different across different types of care. We will review this specific example and explain the difference in interpretation in Chapter 4. Hoffman and Duncan (1988) have a good

description, with examples, of the differences between these different types of variables.

The chief potential problem with the use of logit regression techniques is something called the Independence from Irrelevant Alternatives Assumption. Briefly stated, it is assumed in the logit model that, for any particular family, the ratio of the probability of any two alternatives that might be chosen is unaffected by a change in the probability of choosing any other alternative. Often this assumption, described more thoroughly in Appendix D, is a reasonable approximation to reality, but occasionally it is not. In these cases, the normal logit model is inappropriate and should not be used. A probit or nested logit model would be required.

The Independence from Irrelevant Alternatives assumption has two practical implications for our work estimating the demand for child care. First, we should take care to define the types of child care from which families choose so they are not too similar, one to another. When alternatives are quite similar, the I.I.A. assumption is more likely to be violated. Second, we should use a statistical test (see Chapter 5 for details) to judge whether the I.I.A. assumption is a reasonable approximation in estimating our particular model of child care choice.

STEP 4: PERFORMING THE ESTIMATION

Inputting the data so questionnaire results are available in a machine-readable data matrix will, of course, take time and effort. So too will the process of creating and defining both dependent and explanatory variables in the form wanted for the final estimation process. If variables, such as the price variables described in the next chapter, have to be imputed,[10] separate regressions or other steps may be required.

Actually estimating the value of different coefficients in our model is not usually difficult, once the data and variables are prepared. Packaged computer programs are available to estimate models with multinomial logit or conditional logit variables or a combination of the two (mixed logit). These programs, whose strengths and weaknesses are reviewed

[10] Imputation occurs when we do not have data for a particular family or a particular variable and we have to develop approximate measures of the missing data based on other evidence in the data set.

in Appendix A, automatically calculate certain statistics which help us to judge how good a particular set of estimates is.

STEP 5: USING STATISTICAL TESTS

A variety of statistical tests are available to judge whether the estimates of the coefficients in our model are acceptable or not. Chapter 7 of Ben-Akiva and Lerman (1985) is an excellent source of information and advice. Experience in statistical analysis and an intuitive knowledge of the behaviour being modelled (i.e., child care choice) are desirable, so that the information from statistical tests can be interpreted carefully and used to amend and develop a more refined model of child care decision making.

The most basic statistical test is the asymptotic t-test (or, simply, t-test) which is typically used to tell us whether each and every estimated coefficient is significantly different from zero (e.g., if the value of each t-statistic is close to or above 2.00, we can have a high degree of statistical confidence that the variables tested are important factors in families' child care choices)

There are similar tests to tell us whether the coefficient values of any particular group of variables are significantly different from zero (or from some other value). There are tests to judge whether different groups of consumers make their child care decisions in approximately the same way, and can therefore both be included in the same estimated model. There are tests in which any particular group of child care users can be dropped from the estimation process; the test tells us whether dropping the group of observations has any significant effect on the overall results.

Further, there are tests to determine whether a variable is best entered as a conditional logit or multinomial logit variable in the model, and, if entered as conditional logit, whether it is reasonable to constrain the estimated coefficients across types of child care to be equal. As mentioned above, there are a number of ways of testing whether the Independence from Irrelevant Alternatives assumption is a reasonable one for our model. If the I.I.A. test rejects our model, there are alternative estimation procedures which can be used.

The use of many of the available tests is illustrated in the concrete estimation example given in the next chapter.

STEP 6: FORECASTING AND POLICY ANALYSIS

The purpose of estimating the coefficients of a model of child care choice is to be able to forecast future demand for different types of child care and to predict how changes in child care policy will affect child care demand.

The simplest way to do this is to use our data set in reverse. Up till now, we have used data on each family's child care choices and data on the explanatory variables for each family to find the unknown coefficients of each explanatory variable. Now we use the "known" (i.e., estimated) coefficients together with new values of the explanatory variables to predict which families will choose which types of child care. The explanatory variables will have new values because we will change them to reflect some new situation (e.g, rise in family incomes, fall in day care prices, change in the number and ages of children in families, changes in mothers' work patterns, etc.) that we want to simulate. Each simulation will require a new pattern of changes to explanatory variables of each family in the data set, and will produce a different prediction of the number and characteristics of families choosing each type of child care. A simulation may be designed to reproduce expected changes in the society and economy over the next few years; then, our child care predictions are a forecast of the future. Alternatively, a simulation may change only one or two explanatory variables to mimic the effects of some change in child care policy (e.g, subsidization of day care prices, or an increase in tax benefits for families with young children); in this case, our child care predictions are a form of policy analysis.

4

HOW TO MEASURE AND FORECAST DEMAND FOR CHILD CARE ARRANGEMENTS: A CONCRETE EXAMPLE

In the mid-1970's, Project Child Care conducted a major questionnaire survey of child care use in Metropolitan Toronto. Sponsored by the Metro Toronto Social Planning Council, its results were analyzed in a series of reports, of which *Who Cares?* and *Taking Care* are the best known.

We have used the Project Child Care data set, to which the Metro Toronto Social Planning Council has kindly given us access, to illustrate the use of the statistical and analytical techniques described in this Handbook, and to demonstrate how the results may be used for forecasting and policy analysis.

Let's go through the six steps described in the last chapter, using the Project Child Care data set.

THE MODEL

Who, what, how, and why; these are the questions our model of child care choice must answer.

Who?

Our main focus of interest is the child care choices of mothers currently engaged in the paid work force, because we believe these families are affected by different factors in their child care choices than families with

mothers in school or doing unpaid work at home during the day. Because the Project Child Care data set only includes in its sample families with at least one preschool child, we model the child care choice behaviour of families with employed mothers and preschool children. These parts of our model define which families shall be included and excluded from the data set for our estimation.

What?

Each family is viewed as choosing only one of three broad types of child care arrangement: a licensed group day care centre; some form of paid sitter unrelated to the family; or care by a relative (including husband or sibling) which may be paid or unpaid. If a family, in fact, simultaneously uses several forms of care (either more than one type per child or several children using different types), the type of care used for the largest number of hours during the week was taken to represent the choice of the family. This part of our model defines the dependent variables.

How?

Each family chooses the type of care that is most satisfactory to them; to put it in economist's language -- the type of care that gives them the greatest amount of utility.

Why?

The amount of utility given by each type of child care will depend on a series of factors. We choose as explanatory variables those factors which we believe have a major impact on the choices families make. In analyzing the Project Child Care data, our selection of explanatory variables was based on the kind of thinking reflected in the quote from Katie Cooke mentioned earlier. There are four kinds of variables that would appear to be important. The first are the attributes of the different kinds of child care services being considered -- in particular the price, convenience and quality of each service. The choice of a child care arrangement is a complex process, of course, but each family tries to do the best they can with what resources they have. Each family looks for an arrangement that provides safe, stimulating, and conveniently located care at a price it can afford. Naturally then, arrangements that are of higher quality, greater convenience, and lower cost tend to be preferred. We do not have enough information in our

data set to know the quality of each arrangement from which a particular family chooses, so we cannot include some measure of quality level (e.g., staff-child ratio, caregiver qualifications) as an explanatory variable. But we can find some measures of price and convenience.

A family's choice of main child care arrangement will be influenced by the ages and number of children in the family, so this is the second category of explanatory variables we should consider. Part of the reason for the effect of these "child" variables is that parents' views of the ideal type of care for an infant, preschooler, and schoolage child tend to differ; the accessibility of different types of care for different ages of children may also be important. A third reason is that the price of child care will vary with the ages and number of children in the family. Ideally, we would like to devise explanatory variables which would account separately for these separate influences on the utility of each type of care; we are able to do so partially by modelling price as the total child care price for the family, so that price varies according to the age and number of children.

Whether by choice or compulsion, mothers have typically been responsible for most of the fundamental child rearing decisions in Canadian families, and have been expected to provide child care if they are not occupied in the paid work force. This household division of labour may be in the process of changing; so far, this change has not shown up in estimates of the factors which affect child care demand. The type of child care chosen is, accordingly, likely to be affected by the mother's work situation rather than the father's. It seems likely that the number of hours worked by the mother may have some bearing on the type of care chosen, with more structured and dependable alternatives being preferred the greater the number of hours the mother works in the paid labour force. Mothers with higher wage levels have both a greater ability to pay and suffer a greater financial burden if normal arrangements break down. This may affect their choices. The other major employment-related factor which may influence child care choice is the shift on which a mother is employed. Employment on the night shift will generally require the use of care by a relative, both because sitters and day care are rarely available in nightime hours, and care by a relative may well be preferred for preschool children during non-daytime hours.

Finally, there are a group of socioeconomic variables, some of which may reflect differences in parental tastes and preferences, some of which will reflect constraints on what a particular family can afford. These might vary somewhat from community to community. For our analysis we have included variables which reflect the immigrant status of the family, the mother's education level, the number of parents living with the child, and the family's income.

We expect that many families who have recently immigrated to Toronto, particularly if they do not speak English and cannot readily integrate into society's institutions, will want their children brought up in a culturally and linguistically familiar environment. This is likely to encourage them to seek child care by relatives -- members of the extended family.[11]

Mothers with higher levels of education may be more aware of the long term value to children of care provided by skilled caregivers and developmentally appropriate programmming. In this case, mothers with more education would prefer care in a day care centre to other types. Single parent families are thought by many to have child care needs which are fairly distinct from those of two parent families. Some researchers have excluded single parent families from child care choice analysis, or analyzed them separately, for this reason. We include both groups of families in our sample, but record "lone parent family" as an explanatory variable to pick up differences in the choice process for these families.

Family income may influence the child care choices families make; it is commonly remarked that only the rich (because of their affluence) and the poor (because of eligibility for subsidies) can afford to use day care. In our estimation, subsidy eligibility is separately accounted for in the price variable, so we would expect higher income to be a significant predictor of the choice of a day care centre. It is unclear what effect income might have on the choice of other forms of care.

THE DATA

If we were collecting our own data for the estimation of child care choice, there would be a host of specific issues to discuss here, dealing with how best to elicit accurate information on our dependent and

[11] Patterns may vary in other parts of the country or for immigrants from different countries.

explanatory variables from a representative sample of the relevant population. Some of these issues are raised in Appendix B, in discussion of the strengths and weaknesses of the 1988 National Child Care Survey. In our case, we use an existing data set. This constrains both the population which our sample can be said to represent and the variables we can include in the model.

The Project Child Care data set has information on the dependent variables and explanatory variables as discussed in the section above. In fact, of course, data availability influenced the development of the model. Quality variation from one type of child care to another is not considered as an explanatory variable precisely because there was no information available on the quality level of the types of care not chosen by each family. There were problems with the price data as well. The Project Child Care questionnaire asked families about the price of the particular type of child care chosen. It did not ask about the price of those other types of care available to the family but rejected. Yet, the demand for each type of child care is affected by both sets of prices, not just one.

Making certain assumptions, detailed in Cleveland (1990a), we have imputed price information where it did not exist, in order to solve this problem. This solution is not ideal; the availability of information collected directly from families would clearly be preferred (see Appendix B for a discussion of alternative ways of solving this data problem). In essence, we assume that most families would have to pay the average price for care for each child in a particular age group. The total price of a particular type of child care for the family would be the sum of the prices for each child in the family. So, for instance, for each child less than 2 years of age, families would face the average price of day care for infants, of sitter care for infants, and of paid relative care for infants. These averages were calculated from the information in the data set provided by users of those types of care. In two situations, families were assumed to have a lower price than these averages for each child. If a family was deemed to be eligible for day care subsidy, the day care price was presumed to be $2.50 a week (the Metro Toronto minimum payment at the time). If a family had a live-in relative (other than the mother's husband), or the husband was not currently employed, then the family was considered to have access to free child care by a relative.

We recite these details both to illustrate the difficulties of using existing data, and the creative ways in which its limitations can be (partially) overcome. Because the cost, in time and money, of data collection can be high, use of existing data sources may be preferred, especially for preliminary forays into use of these new techniques. When new data is being collected, it is important to consider collection of data which, for the purposes of maintaining a policy analysis and forecasting model, could be updated in future without having to complete an entire new survey. For instance, information on prices of day care centres and sitters could be updated using information collected by resource and referral centres, municipalities, or others. Information on family incomes, numbers of lone parent families, number and ages of children could be updated using census data or municipal assessment information.

THE STATISTICAL FORM OF THE MODEL

For all the reasons described elsewhere in this handbook, we use a logit model for the estimation of unknown coefficients. In the next chapter, we present results of a test of the Independence from Irrelevant Alternatives assumption, which confirms that the logit model is a reasonable one to use. It remains to decide whether the explanatory variables should be entered as conditional logit variables or multinomial logit ones. For all those variables which do not change in amount when the family chooses one alternative rather than another, these must be considered as multinomial logit variables. This applies to all child variables, employment variables, and socio-economic variables.

For those variables which are attributes of the alternatives themselves, it is possible to enter them as multinomial or conditional logit. Hoffman and Duncan (1988) present a reasonably convincing set of arguments for preferring to consider attributes of alternatives as conditional logit. However, software or data limitations may make this impossible. In this case, use of a multinomial logit specification will provide an acceptable and interpretable set of estimates (see Cleveland, 1990a, for an example). In this estimation, we include price as a conditional logit variable, having the same value for each alternative. Data on access to child care within the child's neighbourhood is only available for day care and sitters, so two separate multinomial logit explanatory variables are defined.

INTERPRETING THE RESULTS OF THE ESTIMATION

Table 1 displays the basic results of our estimation of the factors affecting the choice of child care arrangements in families with preschool children and working mothers. At first sight, the table is somewhat complicated; let us go through the different columns and symbols, first, and then look at an overall description of what the results can tell us.

There are eight columns in the table. Column 1 lists the explanatory variables in their four categories (attribute, child, mother's employment and socioeconomic variables). Column 2 lists the units in which these variables are measured (so, for instance, the prices are in dollars per week; day care in the neighbourhood is a "dummy" variable, which either has the value zero, if day care is not available in the neighbourhood for a particular family, or the value one, if day care is available). The next six columns present two types of information. Three of the columns are headed by a "b"; the other three are headed by a "t". The columns labelled with a "b" present the estimated coefficients of our explanatory variables as calculated by the computer. The columns labelled with a "t" present what are called "t-statistics" for each variable. The t-statistics, as mentioned in Chapter 3, tell us whether each particular estimated coefficient is significantly different from zero (statistically speaking). The asterisks which appear in Table 1 are an extra aid to reading the table quickly. If the t-statistic for a particular variable shows that the coefficient is highly statistically significant, then a double asterisk (**) will appear beside the coefficient. If the t-statistic shows the variable to be statistically significant, but at a slightly lower level, then a single asterisk (*) will appear.

TABLE 1

MIXED LOGIT ESTIMATES:
CHOICE OF DAY CARE CENTRE, SITTER, OR CARE BY RELATIVE

Explanatory Variables	Units	Day Care vs Relative		Sitter vs Relative		Day Care vs Sitter	
		b	t	b	t	b	t
Constant		-3.49**	2.42	-1.15	1.06	-2.34	1.59
Attribute Variables							
Price of type of care (1)	$/wk	-0.04**	5.77	-0.04**	5.77	-0.04**	5.77
Day Care in Neighbourhood	dummy	1.02**	2.71	0.05	0.19	0.97**	2.58
Sitter in Neighbourhood	dummy	0.13	0.37	0.08**	3.33	-0.67**	2.06
Child Variables							
Child Under Two Years	dummy	-1.99**	2.45	-0.14	0.39	-1.85**	2.32
Child 2-5 Years Old	dummy	0.08	0.14	0.38	1.20	-0.30	0.51
Child 6-9 Years Old	dummy	-0.73	1.45	-0.11	0.40	-0.61	1.24
Child 10-15 Years Old	dummy	-1.48**	2.81	-0.36	1.21	-1.13**	2.12
Mother's Employment Vars.							
Hours of Paid Employment	hrs/wk	0.07**	2.37	0.05**	2.16	0.02	0.68
Employed Nights/Weekends	dummy	-0.37	0.87	-1.22**	3.57	0.84*	1.81
Mother's Wage	$/hr	0.05**	1.68	-0.007	0.17	0.06	1.30
Socio-Economic Variables							
Non English Speaking Immigrants	dummy	-0.74*	1.83	-0.97**	3.41	0.23	0.54
Mother's Education Level	9 levels (1)	0.22**	2.05	0.05	0.62	0.17	1.53
Lone Parent	dummy	0.79	1.48	0.41	0.91	0.38	0.80
Family Income	$000's/yr	-0.02	0.98	-0.03**	2.16	0.02	0.73

Notes:

** Significant at 5% level

* Significant at 10% level

(1) Price of type of care is modelled as a conditional logit variable, with an estimated co-efficient constrained to equality across different types of care. The interpretation of this co-efficient is different from that of multinomial logit co-efficients, as detailed in the text.

Child variables appear to have their effect primarily on a family's use of a day care centre. Having an infant makes a family significantly less likely to use day care compared to either care by a relative or care by a sitter. Presumably this occurs either because families prefer a home-like setting for their infants or because few infant spaces are available in day care centres, or a combination of the two.

You will notice the peculiar headings above columns 3 to 8: day care vs. relative, sitter vs. relative, day care vs. sitter. These refer to the most useful way of interpreting what the estimated coefficients mean. As explained further in Appendix D, the coefficient estimates can be interpreted as the effects of the explanatory variables on the odds (the relative probability) of choosing one type of care rather than another.

Take as an example the "Hours of Paid Employment" variable, about half-way down column 1. Reading across, we find that this variable records the number of hours per week a mother engages in paid work, that the estimated coefficient for the choice of day care rather than care by a relative is 0.07 and that this coefficient is highly statistically significant. In other words, the number of hours a mother is employed is found to have a significant impact on the odds that a family will choose a day care centre rather than care by a relative. Because the coefficient is positive rather than negative (+0.07, rather than -0.07), we know that "hours of mother's work" *increases* the odds that day care will be chosen. Similarly, the number of hours a mother is employed each week has a significant and positive impact on the relative probability that the family will choose a sitter rather than care by a relative. On the other hand, the coefficient on mother' hours of paid employment in Column 7 shows a positive, but insignificant effect on the odds of choosing day care rather than a sitter. Statistically speaking, we cannot really be sure that an increased number of employment hours will make choosing day care more likely than choosing a sitter.

The most important use of the results presented in Table 1 is to make an overall judgement about whether our views of child care choice, as

expressed in our model, are supported by the data. In particular, we can judge whether the pattern of effects of various variables on child care choice that we expected to see is confirmed. For this we essentially want to look at the signs of the estimated coefficients (positive or negative) and whether at least the major explanatory variables are statistically significant.

Viewed from this perspective, the estimates in Table 1 strongly confirm our basic views about the fundamental factors which determine whether a particular family will decide to use day care or some other arrangement. The attributes of different types of care (i.e., the Attribute variables), without exception, affect child care choices in the direction and with the statistical significance that we would expect. The price of each type of care has a significant negative impact on the odds that this type of care will be chosen relative to other types of child care (i.e., higher prices discourage use). When day care is available in the child's neighbourhood, it significantly increases the likelihood that day care will be chosen rather than its alternatives. As expected, availability of day care will not affect the relative probability of choosing sitter versus a relative. The same pattern of significance is found for the variable "sitter in the child's neighbourhood"; when a sitter is available, the odds of using a sitter are significantly and positively affected, but other odds ratios are unaffected.

The Child variables tell an interesting story, consistent with our model, but also providing us additional information. Only the presence of infants in the family and the presence of children above normal child care age have statistically significant effects on the family's choice between different types of care. (Another way of putting the same information -- knowing that a family has a child 2-5 years of age or 6-9 years of age provides no basis for predicting what kind of child care the family will decide to use.) Child variables appear to have their effect primarily on a family's use of a day care centre.

Having an infant makes a family significantly less likely to use day care compared to either care by a relative or care by a sitter. Presumably this occurs either because families prefer a home-like setting for their infants or because few infant spaces are available in day care centres, or a combination of the two. Having a ten to fifteen year old child similarly reduces the odds that day care will be used rather than either of its alternatives. Presumably, children in this age range provide some

amount of care for their younger siblings after school, and this combines more easily with sitter or relative as the major form of care than it does with day care.

The fact that mothers rather than fathers bear major responsibility for child care in the household division of labour is evident in the clear effect of Mother's Employment variables on child care choice. As anticipated, increased hours of work by the mother significantly increases the odds that children will be cared for by either a sitter or a day care centre rather than a relative. However, if a mother works evenings or weekends the odds of using a sitter rather a relative are significantly decreased. Even the odds of using a day care centre over a sitter are significantly (at 10%) increased when mothers work evenings, nights or weekends. Since few day care centres provide any non-day care, this must reflect increased odds of using day care rather than a sitter during the day. As the mother's wage increases, she is significantly more likely to have her children cared for in a day care centre rather than by a relative.

The fact that mothers rather than fathers bear major responsibility for child care in the household division of labour is evident in the clear effect of Mother's Employment variables on child care choice.

There are some Socioeconomic variables that influence the choice of child care arrangements, but others that, contrary to popular opinion, apparently do not. Families in which both parents are immigrants and where English is not the main language spoken in the household are significantly more likely to use child care by relatives than either of its alternatives. This is true even though our definition of the price of relative care already accounts for the greater likelihood that such families have live-in relatives available to provide care. Such first generation, non English speaking, immigrant families are likely to feel considerably more comfortable using culturally and linguistically similar caregivers from their extended families to provide child care. The mother's education level has a significant positive effect on the odds of choosing day care rather than

a relative, probably reflecting a preference for more educational, developmentally-structured care.

Lone parent families are no more or less likely to choose any particular form of care than two parent families, reflected in the lack of significance of lone parent status for all types of care. Family income does not really have the anticipated influence on the choice of child care. We expect that higher family income will encourage the use of day care (the most expensive of the types of care). The only statistically significant effect of family income is to reduce the odds of using a sitter rather than care by a relative. We should recognize (see Appendix C) that many other studies of child care choice have concluded that family income is not a significant predictor of the type of care a family will use.

Families in which both parents are immigrants and where English is not the main language spoken in the household are significantly more likely to use child care by relatives than either of its alternatives. This is true even though our definition of the price of relative care already accounts for the greater likelihood that such families have live-in relatives available to provide care. Such first generation, non English speaking, immigrant families are likely to feel considerably more comfortable using culturally and linguistically similar caregivers from their extended families to provide child care.

It is important to note, in thinking about the, perhaps surprising, results for lone parent families and for family income that many factors related to these variables are already separately accounted for in our regression. For instance, lone parent families are much more likely than two parent families to be eligible for a day care subsidy and generally less likely to have a live-in relative available to provide care; they are also likely to have fewer children under ten years of age. These factors are already accounted for in the lower day care prices, higher prices for care by a relative, and smaller number of children under ten years of age of many lone parent families in the data set. What the estimated coefficients for "lone parent" imply is that, independent of the influence of all these other explanatory variables, lone parent status has no separate effect. The same interpretation applies to the effect of income on the choice of

child care arrangements. In other words, apart from the greater likelihood that low income families will get a day care subsidy, or that high income families will have a mother who works long hours or has a higher level of education, income by itself has no effect on the choice between a day care centre and its alternatives.

TESTING AND REVISING THE ESTIMATES

Statistical tests, apart from the brief discussion of t-statistics above, are taken up in the next chapter. These tests can be used to help find the best form of our model of child care choice. In fact, the estimates presented in the section above have already been subjected to a battery of statistical tests and represent our preferred final specification of a child care choice model.

FORECASTING AND POLICY ANALYSIS

Table 1 presented the results of our statistical estimation in their raw form. Those same estimates can be manipulated to give us a great deal more information. Ultimately, combined with assumptions about what will happen to our explanatory variables in the future, we can use the estimates for forecasting and for policy analysis.

You will remember our mathematical equation for the utility given by a particular type of child care. In its simplest form we wrote it as:

$$y = bx + az$$

where y is utility from a particular form of care, x and z are particular explanatory variables and b and a are unknown coefficients. Our final model turned out to have many more than two explanatory variables, each with its own coefficient; apart from that our initial equation is still valid.

Our data set contains all the information on the values of each of the explanatory variables for each family (481 families in total for the Project Child Care estimates). But the "unknown coefficients" are no longer unknown; we now have an estimated value for each one of them (see Table 1). We can substitute these estimates for the a's and b's in our

equation and therefore find out the value of y -- the amount of utility -- for each type of care.

This may not seem to get us much farther ahead -- utility is a very abstract concept, after all -- but it does. You will remember that families are assumed to choose the type of child care that gives them the greatest amount of utility. So, our calculations now give us a prediction, for each family in the data set, of which type of child care the family will choose -- the one that gives the greatest estimated total of utility.

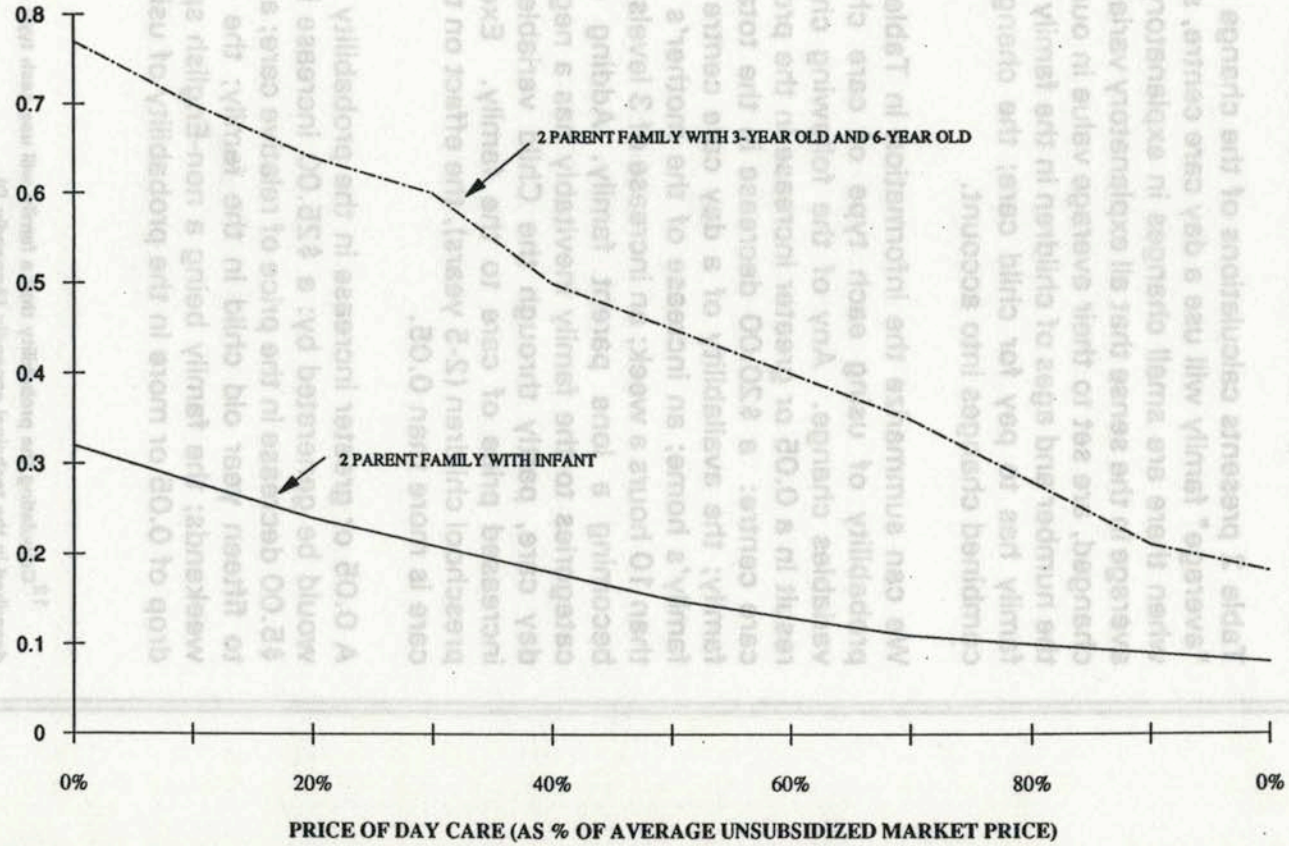
Since our data set is a representative sample of a particular population, we can now predict how many families will decide to use each type of child care. In particular, if any of the values of the explanatory variables were to change, we can also make a new prediction of the number of families who will choose each type of child care under the new circumstances.

We can look at examples of how the predicted use of child care changes when values of the explanatory variables change. For instance, Chart 1 presents a graph of the effects of changes in some of the explanatory variables on the probability that a particular type of family will use a day care centre. This chart depicts the reactions of a particular family -- a working two-parent family with pre-tax income of $50,000, in which the mother, having completed high school, works 40 hours a week on the day shift, earning ten dollars an hour. The family is Canadian-born, and all child care arrangements are available to it at market prices. This seems like a fairly typical family type which might be of interest; somewhat different graphs could be drawn for different family types.

Chart 1 shows the effect on the likelihood that this family will use a day care centre of changes in the price of day care and of the ages and number of children in the family. When day care is fully subsidized (i.e., price = 0% of normal market price), our "typical" family with a 3-year old and a 6-year old will have a greater than 75% probability of using a day care centre. The same family, but with an infant, would have only about 30% probability of choosing a day care centre, even if it were fully subsidized. As the price of day care rises from 0% to 100% of the normal market price, the probability that our "typical" family will use a day care centre falls. At full market price, the probability that the two child family will use day care is less than 20% and for the family with an infant, the probability is less than 10%.

CHART 1

PREDICTED PROBABILITY OF DAY CARE CENTRE CHOICE BY TWO PARENT FAMILY, BY AGE OF CHILD AND DAY CARE PRICE



2 PARENT FAMILY IS ASSUMED TO HAVE PRETAX INCOME OF $50,000, MOTHER EMPLOYED ON DAY SHIFT, 40HRS/WK, EARNS $10/HR, HAS COMPLETED HIGH SCHOOL, NOT RECENT IMMIGRANTS, OTHER CHILD CARE ARRANGEMENTS ARE AVAILABLE IN NEIGHBOURHOOD AT AVERAGE MARKET PRICES

These demand curves, drawn for the individual family, allow some basic results of a logit estimation to be presented in a clear graphic manner.

Table 2 presents calculations of the change in the probability[12] that an "average" family will use a day care centre, sitter and care by a relative when there are small changes in explanatory variables. This family is average in the sense that all explanatory variables, besides the one being changed, are set to their average value in our data sample. Changes in the number and ages of children in the family also changes the price the family has to pay for child care; the change in probability takes the combined changes into account.

We can summarize the information in Table 2 by looking at how the probability of using each type of care changes when explanatory variables change. Any of the following changes apparently[13] would result in a 0.05 or greater increase in the probability of choosing a day care centre: a $20.00 decrease in the total cost of day care to the family; the availability of a day care centre in close proximity to the family's home; an increase of the mother's work week by a bit more than 10 hours a week; an increase of 3 levels in the mother's education; becoming a lone parent family. Adding children in different age categories to the family inevitably has a negative effect on the use of day care, partly through the Child variables and partly through the increased price of care to the family. Except in the case of older preschool children (2-5 years), the effect on the probability of using day care is more than 0.05.

A 0.05 or greater increase in the probability of using care by a relative would be generated by: a $25.00 increase in the price of day care; a $5.00 decrease in the price of relative care; an infant, schoolaged or ten to fifteen year old child in the family; the mother working nights or weekends; the family being a non-English speaking immigrant one. A drop of 0.05 or more in the probability of using care by a relative could

[12] Calculating the probability that a family will use each type of care requires one extra step which is described in the technical appendix (Appendix D).

[13] For the purposes of this description, we assume that doubling or tripling the size of the change in the explanatory variable, as shown in the second column of Table 2, will double or triple the size of the change in probability. So, for instance, we assume that a $20.00 increase in the price of day care would reduce the family's probability of choosing a day care centre by .06. This assumption is convenient for providing a general sense of the magnitude of the results.

TABLE 2

CHANGES IN THE PROBABILITY OF CHOOSING A DAY CARE CENTRE, SITTER AND AND RELATIVE DUE TO CHANGES IN EXPLANATORY VARIABLES AT SAMPLE MEANS

Explanatory Variables	Change in Explanatory Variable	Change in Probability of Choosing Day Care	Change in Probability of Choosing Sitter	Change in Probability of Choosing Relative
Probabilities	Base Case	0.08	0.36	0.55
Attribute Variables				
Price of Day Care	+$10/week	- 0.03	+0.01	+0.02
Price of Sitter	+$10/week	+0.01	- 0.08	+0.07
Price of Relative	+$10/week	+0.02	+0.08	- 0.10
Day Care in Neighbourhood	+1.00	+0.08	- 0.02	- 0.06
Sitter in Neighbourhood	+1.00	- 0.02	+0.18	- 0.17
Child Variables				
Child Under Two Years	+1.00	- 0.11	- 0.01	+0.12
Child 2-5 Years Old	+1.00	- 0.30	+0.05	- 0.02
Child 6-9 Years Old	+1.00	- 0.09	- 0.004	+0.08
Child 10-15 Years Old	+1.00	- 0.08	- 0.05	+0.13
Mother's Employment Variables				
Number of Hours Paid Employment	+$10/week	+0.04	+0.08	- 0.12
Mother Employed Nights/Wknds	+1.00	+0.001	- 0.23	+0.23
Mother's Wage	+$1.00/hour	+0.005	- 0.003	- 0.001
Socio-Economic Variables				
Non English Speaking Immigrants	+1.00	- 0.03	- 0.20	+0.22
Mother's Education Level	+1 (of 9)	+0.02	+0.005	- 0.02
Lone Parent	+1.00	+0.06	+0.06	- 0.12
Family Income	+$1,000/year	0.00	- 0.01	+0.01

Notes:

(1) Changes in the number and ages of children in the family will change the price of each type of child care the family faces. Figures for the first three categories of child variable reflect this.

be generated by a day care centre or sitter becoming available in the neighbourhood; an increase of 5 hours in the mother's work week (and therefore of the number of hours of child care needed); or the mother becoming a lone parent.

An increase of .05 or greater in the probability of using care by a sitter could be brought about by a rise of $10.00 a week in the price of care by a relative; the convenient availability of sitter care close by; having a child 2-5 years of age; an increase of $10.00 a week in the number of hours mother is employed; or becoming a lone parent family. On the other hand, a comparable reduction in the probability of sitter care would result from a rise of $10.00 a week in the price of a sitter; having a child 10-15 years of age; having to work evenings, nights or weekends; or being a non English speaking immigrant family.

Table 2 refers only to a single, though average, family. More relevant from the point of view of policy analysis or forecasting is the effect of changes in explanatory variables on the population of all families who might use child care. Table 3 presents the same changes as Table 2, but for all families rather than just one.

Broadly, Table 3 shows a pattern of change similar to that depicted in Table 2. Table 3 usefully demonstrates the sensitivity of child care demand to changes in the age composition of families, immigration patterns, the prevalence of evening, night and weekend employment, and the price and convenient availability of different types of care.

Table 4 extends this exercise a step further; however, the assumed changes in explanatory variables are frequently larger, and sometimes more complicated. We continue to assume that these changes in explanatory variables do not affect the estimated coefficients of child care choice. Table 4 presents the predicted effect of alternative child care policy changes on the percentage of families choosing different forms of care. Five types of policy are simulated: an increase in the supply of day care centres conveniently available to families (perhaps through a program of capital grants); an operating grant which lowers the price of licensed day care by $10.00 per child per week; an extension of subsidy eligibility to include both low and middle income families; a child care expense deduction which allows the deduction of the cost of any form of child care from income tax payable; and a family income supplement of a flat amount per year, irrespective of the child care choice made.

TABLE 3

CHANGES IN THE PREDICTED PERCENTAGE OF FAMILIES CHOOSING A DAY CARE CENTRE, SITTER AND RELATIVE DUE TO CHANGES IN EXPLANATORY VARIABLES

Explanatory Variables	Change in Explanatory Variable	Change in Percentage Choosing Day Care	Change in Percentage Choosing Sitter	Change in Percentage Choosing Relative
Probabilities	Base Case	12.0	29.3	58.7
Attribute Variables				
Price of Day Care	+$10/week	- 6.1	+4.6	+1.4
Price of Sitter	+$10/week	+4.0	+11.0	+7.0
Price of Relative	+$10/week	+1.5	+8.9	- 10.3
Day Care in Neighbourhood	+1.0	+10.9	- 7.5	- 3.4
Sitter in Neighbourhood	+1.0	- 7.3	+30.4	- 23.0
Child Variables				
Child Under Two Years	+1.0	- 15.4	+2.2	+13.2
Child 2-5 Years Old	+1.0	- 6.8	+4.8	+2.0
Child 6-9 Years Old	+1.0	- 14.6	+3.2	+11.4
Child 10-15 Years Old	+1.0	- 10.7	- 0.6	+11.3
Mother's Employment Variables				
Number of Hours Paid Employment	+$10/week	+4.4	+7.9	- 12.3
Mother Employed Nights/Wknds	+1.0	+4.5	- 32.3	+27.8
Mother's Wage	+$1.00/hour	+1.5	- 1.1	- 0.3
Socio-Economic Variables				
Non English Speaking Immigrants	+1.0	- 3.9	- 31.1	+35.1
Mother's Education Level	+1 (of 9)	+2.7	- 1.4	- 1.3
Lone Parent	+1.0	+7.3	+5.5	- 12.8
Family Income	+$1,000/year	0.0	0.0	+0.4

Notes:

(1) Changes in the number and ages of children in the family will change the price of each type of child care the family faces. Figures for the first three categories of child variable reflect this.

TABLE 4

POLICY SIMULATIONS ON FULL DATA SET

POLICY SIMULATED	CHANGE IN PERCENTAGE PREDICTED TO USE DAY CARE	CHANGE IN PERCENTAGE PREDICTED TO USE SITTER	CHANGE IN PERCENTAGE PREDICTED TO USE RELATIVE
BASE CASE PREDICTED PERCENTAGES	12.0	29.3	58.7
DAY CARE CENTRE AVAILABLE IN EACH NEIGHBOURHOOD	+3.2	- 1.8	- 1.4
OPERATING GRANT WHICH LOWERS DAY CARE PRICE BY $10.00 IN EACH AGE CATEGORY	+6.6	- 4.4	- 2.2
ALL LONE PARENT FAMILIES EARNING LESS THAN $15,000/YEAR ELIGIBLE FOR DAY CARE SUBSIDY	+1.3	- 1.3	0.0
ALL FAMILIES EARNING LESS THAN $15,000/ YEAR ELIGIBLE FOR DAY CARE SUBSIDY	+10.1	- 4.6	- 5.5
ALL FAMILIES EARNING LESS THAN $25,000/ YEAR ELIGIBLE FOR DAY CARE SUBSIDY	+24.9	- 13.6	- 11.3
CHILD CARE EXPENSE DEDUCTION ALLOWS REDUCTION OF 40% IN TOTAL CHILD CARE COSTS:			
- VIEWED AS REDUCED PRICES FOR ALL FORMS OF CHILD CARE	+2.2	+2.1	- 4.3
- VIEWED AS INCREASE IN FAMILY INCOME	0.0	- 0.2	+0.2
FAMILY BENEFIT GRANT OF $1,000/YEAR FOR EACH FAMILY	0.0	- 0.2	+0.2
FAMILY BENEFIT GRANT OF $5,000/YEAR FOR EACH FAMILY	0.0	- 1.3	+1.3

An increase in the supply of day care centre services available has only a modest impact on the choice of care type, largely because the majority of families in the data set already have a day care centre available to them. The marginal impact of increased accessibility to day care is an increase of about 3.2 percentage points, or 25% in the consumption of day care, with approximately equal percentage point reductions in the use of sitters and care by relatives.

A reduction of about 1/3rd in the price of a day care centre, brought about by an operating grant, results in a rise of over 50% (6.6% points) in the use of day care. About 2/3rds of the new day care consumers are former sitter users, 1/3rd former users of care by a relative.

Table 4 records the effect of three alternative new definitions of eligibility for day care subsidy, reducing the price of these services to $2.50 per week. We presume that all families eligible for subsidy can actually receive it, without waiting lists or other restrictions, an assumption that is, historically speaking, unrealistic. Small changes in subsidy eligibility have minor effects on consumption, whereas major changes have large effects. Extending day care subsidy to all lone parent families earning less than $15,000 a year increases day care use by just over 1%. Extending eligibility to two parent families earning the same amount, however, nearly doubles the consumption of day care; if the income limit is raised to $25,000, the use of day care rises to triple its original level (about 1/4 of all child care users would use a day care centre).

The latter two types of child care policy are designed to provide assistance no matter which type of child care is chosen. They are designed to be, and are, more neutral in their effects on the choice of child care type.

The Child Care Expense Deduction relieves a family from the tax it would otherwise pay on a portion of its income --that portion used to pay for child care expenses. Although this tax benefit is therefore received in a lump sum, after the end of the calendar year in which the expenses took place, it can be viewed as a reduction in the effective price of child care paid by a family. We first model the child care deduction in this way.[14] The deduction applies to care provided by a day care centre, sitter or care by a relative (as long as the relative is not part of the immediate

[14]The Child Care Expense Deduction already existed, in less generous form than at present, at the time the Project Child Care survey was completed. It therefore affects, in unknown ways, the behaviour of families in the sample. For simplicity, we model the Deduction as if it was a new program.

family and under 21 years of age). We ignore complications of modelling the change in the tax liability of each family and simply presume a marginal tax rate, and therefore tax reduction, of 40% of the cost of child care for each family. Despite the relatively large price change, the effects on child care choice are relatively small (i.e, more neutral) because the price reduction occurs simultaneously for all forms of care. The use of both day care centres and sitters are increased by just over two percentage points, with a corresponding reduction in care by relatives.

Parents may view this tax benefit as a supplement to family income, rather than a reduction in the price of child care. If so, the changes in child care choice are even smaller, though not in the same direction. Increases in family income reduce the use of care by sitters slightly and increase the use of care by relatives. There is essentially no change in the percentage of families who decide to use a day care centre. This is true whether the supplement to family income comes in the form of an expense deduction or in the form of a family benefit grant designed to reduce the financial burden of families with children irrespective of their child care choice.

If our policy objective is to increase the use of day care centres, either because of their benefits for children in low-income families, or their benefits to children of all employed mothers, there is evidence that changes in the price of day care centre services will be most effective. Increases in the supply of services at the same price will have some impact; increases in family income will have none. Price changes can be brought about by direct grants to services or subsidization of target families. If our policy objective is to reduce the tax burden on employed mothers caused by child care expenses or to supplement family incomes, we can be reasonably assured that these policies will have only minor impacts on the choice of child care by those families which have a mother currently in the paid labour force.

Table 5 and Chart 2 present yet another way of using our estimated coefficients of child care choice. They answer the question, "What would it take to get virtually every family to use a day care centre for its children?", or to put it slightly differently, "Why do many families not currently use day care?" Table 5, in numbers, and Chart 2, in the form of a bar chart, indicate the number of families who are predicted, using our estimates, to use each type of child care, as successive changes are made in explanatory variables. Scenario #1 is the base case, where all families have values of explanatory variables as given in the Project Child

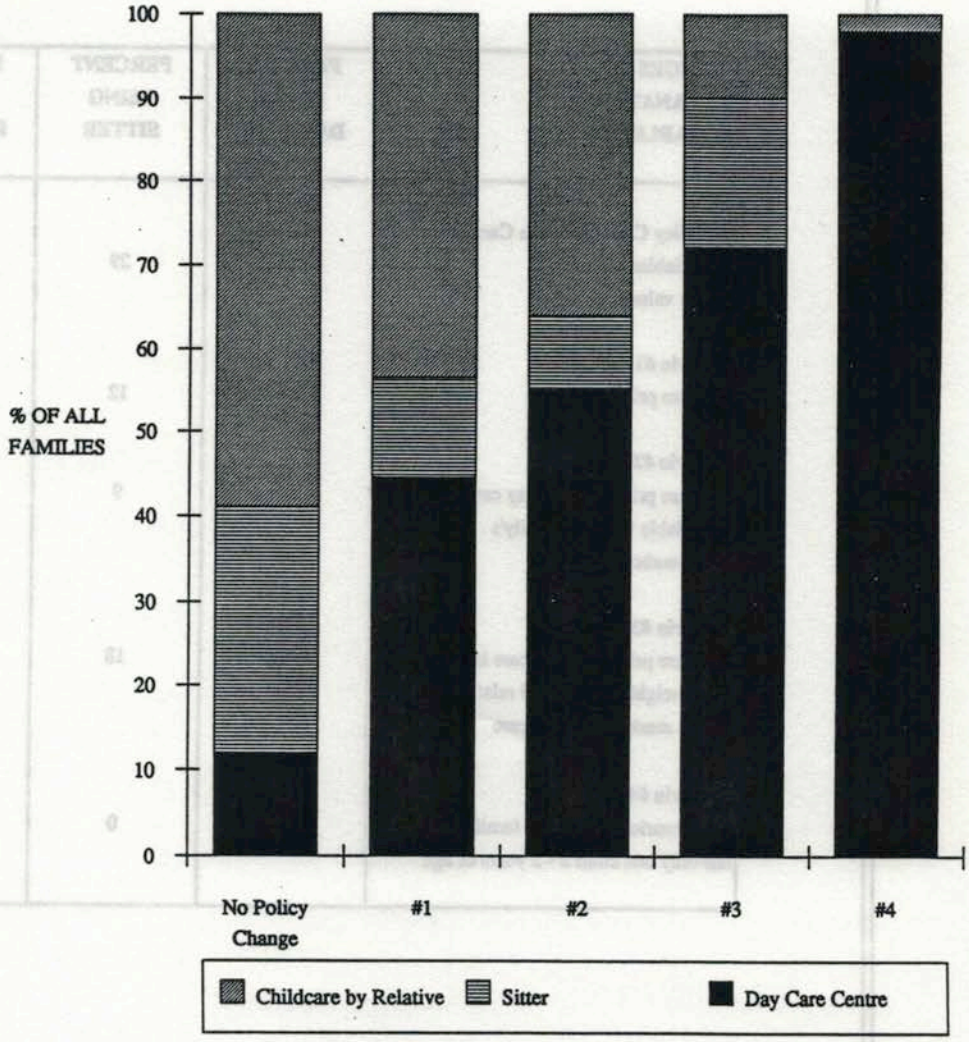
TABLE 5

PERCENT OF FAMILIES CHOOSING A DAY CARE CENTRE, SITTER AND RELATIVE AS A RESULT OF CHANGES DESIGNED TO INCREASE THE DEMAND FOR DAY CARE

CHANGES IN EXPLANATORY VARIABLES	PERCENT USING DAY CARE	PERCENT USING SITTER	PERCENT USING RELATIVE
No Policy Change: Base Case All variables set at sample values	12	29	59
Scenario #1 Day care price = 0	44	12	43
Scenario #2 Day care price = 0 and day care is available in every family's neighbourhood	55	9	36
Scenario #3 Day care price = 0; day care in every neighbourhood; all relatives charge market price for care	72	18	10
Scenario #4 As Scenario #3 but each family has only one child 2 - 5 years of age	98	0	2

CHART 2

PERCENT OF FAMILIES CHOOSING A DAY CARE CENTRE, SITTER AND RELATIVE AS A RESULT OF CHANGES DESIGNED TO INCREASE THE DEMAND FOR DAY CARE



Care Survey. In this case, as we have seen in Table 4, 12% of families choose a day care centre, 29% choose a sitter, and 59% choose care by a relative.

Scenario #2 sets the price of day care equal to zero for all families, with no other changes in explanatory variables. The impact is dramatic: the majority of families previously using a sitter switch to day care, and a significant proportion of those previously using a relative switch to a day care centre as well. In total, more than three times as many families decide to use day care, compared to the base case.

Scenario #3 keeps the price of day care at zero, but ensures that a day care centre is also conveniently available in each family's neighbourhood. As a result, an additional 11% of families decide to use a day care centre. Still, 36% of families continue to use care by a relative, either because of pure preference for this type of care (e.g., for an infant), or because care by a relative may be just as cheap as and even more convenient than a day care centre.

Scenario #4 provides some indication of how families are affected by the fact that care by a relative is typically very low cost or free. This scenario changes the price of care by a relative, for all families, to be equal to the full unsubsidized market price for day care (this could occur, perhaps, when labour was in very high demand and any relative would have to give up another job to care for your child). As a result, five of every seven families previously (i.e., in Scenario #4) using child care provided by a relative change their child care arrangement. Not all of these families decide to use a day care centre, but most do. Day care use rises to 72%.

Our final simulation, Scenario #5, changes the composition (and therefore perceived child care needs) of all families in the data set (while the earlier changes, described above, remain in force). All families now have only one child, and this child is between two and five years of age -- an age at which day care centres are popular. Now, 98% of families use a day care centre, and only 2% continue to prefer care by a relative.

This Table suggests that a very considerable portion of the decision to use one type of child care rather than another depends upon the price and convenience attributes of the care, rather than on pure preference. Changes in the price and availability of day care centres and care by relatives can potentially increase use of day care six-fold, as indicated in Scenarios 1 through 4. On the other hand, some portion of the population

of families not currently using a day care centre is unresponsive to changes in the attributes of this care; the age and number of children in their families makes them prefer either sitter care or care by a relative.

5

GETTING THE MEASUREMENT RIGHT: USING STATISTICAL TESTS TO IMPROVE ESTIMATES OF CHILD CARE CHOICE

There are many statistical tests which can help us to judge the results we have obtained from our logit estimation. We list the more useful ones below:

- t-test - As we have seen in the previous chapter, each of the estimated coefficients (our "a's" and "b's") can be tested to see how much confidence we should place in the numerical estimates we have made.

- Likelihood ratio test - Actually, this is a group of related statistical tests. In its simplest, and most common version, this is a test of whether the whole body of estimated coefficients together do much of a job explaining the choices families make. In other forms, this test can be used to tell us whether any particular group of explanatory variables (for instance, the child variables) are significant in explaining the choices that families make. Or it can be used to tell us whether one segment of our sample population has different tastes than another segment.

- rho-squared and rho-bar-squared tests - Not a different statistical test, but a more convenient way of expressing the results of a likelihood ratio test, these calculations can be used to test which of two fairly similar sets of explanatory variables explains families' choices best.

- Anderson's amended likelihood ratio test - This can be used to test whether dropping a certain class of observations affects the results significantly or not. We can use this test, for instance, to judge whether dropping all lone parent families and reestimating new results produces any significant changes. If it does, we can say that lone parent families choose child care in a way that is significantly different from two parent families.

- Prediction test - The coefficients you have estimated can be used to calculate which families in the data set would be predicted to choose which type of child care. These predictions can then be compared to the actual choices made to calculate the percentage of correct predictions.

- Test of the Independence from Irrelevant Alternatives - The Independence from Irrelevant Alternatives (I.I.A.) assumption has been explained briefly in Chapter 3. Briefly, the logit model presumes that, for any particular family, the ratio between the probabilities of choosing any two alternatives is unaffected by what happens to a third alternative. There are various tests of the I.I.A. assumption available. All of them test whether this assumption, implicit in logit estimation, is a reasonable one for our model and our data set.

Explaining the theoretical rationale and exact method of measurement of various statistical tests goes beyond the scope of this handbook (except for a brief discussion in the next paragraph). A relatively accessible account of most tests is provided in Ben-Akiva and Lerman (1985); Anderson's test is explained in Anderson (1987). Other sources of information are listed in the Bibliography. Let us give some examples, however, of the use of statistical tests in judging estimates from the Project Child Care data set.

SPECIFICATION TESTS

Most of the statistical tests we discuss are based on comparison of "likelihood values" or "log likelihood values". An intuitive grasp of what these things are will help give you some feeling for what the statistical tests do. Estimates such as those presented in Table 1 are made using

what are called maximum likelihood methods (i.e., by maximizing the likelihood function). In essence, what this procedure does is to choose values of the unknown coefficients (i.e., the a's and b's) which make our mathematical model predict that what did happen would happen (i.e., that the predicted choices of families are as close as possible to what they actually chose). An important byproduct of "maximizing the likelihood function" is a number: the value of the likelihood function when it reaches its highest possible total. If we change our model (e.g., add another explanatory variable, change the definition of a variable, omit all socioeconomic explanatory variables from the model) the value of the likelihood function when it reaches its maximum will be different. Most of the statistics discussed below compare likelihood values (or log likelihood values, which give equivalent results) for different versions of our child care choice model, in order to judge which version of the model is preferable.

Table 6 below presents some of the basic statistics associated with the estimates provided in Table 1. There are 481 observations (data on 481 families) in the data set. The log likelihood at convergence (the value of the logarithmic version of the likelihood function when it reaches its maximum) is given, as is the log likelihood when all estimated coefficients are constrained to equal zero. A comparison of the log likelihood at convergence with the log likelihood at zero gives us a measure of how much our model is able to explain. This is what the "all variables" likelihood ratio statistic does, and in a transformed way, what the rho-squared and rho-bar-squared statistics do. The likelihood ratio statistics which refer to attribute variables, child variables, mother's employment variables, and socio-economic variables each compare the main log likelihood at convergence number to the log likelihood when each particular group of variables is omitted from the logit regression. The resulting statistic is a measure of how much explanatory value the particular group of variables has. As the table shows, each group of variables proves to be highly statistically significant. Adding another explanatory variable cannot possibly make the log likelihood at convergence a lower number, at worst it will have no effect on increasing the maximum.

Rho-bar-squared statistics are a useful tool for deciding between two different versions of a model, one of which has more explanatory variables than another, because rho-bar-squared discounts for the

TABLE 6

Number of Observations		481
Log Likelihood at Convergence -- L(b)		-368.34
Log Likelihood at Zero -- L(0)		-528.17
<u>Likelihood Ratio Statistics</u> (all significant at .01)		
All variables		319.66 at 29 d.f.
Attribute variables		59.28 at 5 d.f.
Child variables		25.61 at 8 d.f.
Mother's Employment variables		28.50 at 6 d.f.
Socio-economic variables		32.56 at 8 d.f.
$\rho^2 = 1-\{L(b)/L(0)\}$	(similar to R^2)	.30
$\bar{\rho}^2 = 1-\{[L(b)-k]/L(0)\}$	(similar to $\bar{R}^2$)	.25

addition of extra variables. One reason for preferring the final version of our model with the price variable specified as a conditional logit variable was its better score on the rho-bar-squared criterion.

Lone Parents

An amended likelihood ratio test (Anderson, 1987) can be used to judge whether dropping a group of observations significantly affects the results for the observations which remain. This test can be used to determine whether the estimation of child care choice for a sample excluding the 65 lone parents produces results which are significantly different from the combined sample. We use:

$$-2((\{[T/Tr] \times Lr\} - \{[T \times n]/2 \times \ln[Tr/T]\}) - L)$$

where T = the number of observations in the complete data set
Tr = the number of observations in the restricted data set
Lr = the likelihood value for the restricted data set
L = the likelihood value in the complete data set
n = the number of options

The resulting statistic, distributed as chi-square with n(T-Tr) degrees of freedom,[15] is insignificant. We may conclude that lone parent families behave much the same as other families in their choice of child care arrangements, after accounting for differences in education, income, eligibility for day care subsidy, likelihood of availability of low cost relative care, and so on.

Availability of Day Care Centre Services

The Project Child Care Survey asked respondents who did not choose day care whether they had looked for an arrangement and been unable to find one. This question allows us to exclude the 41 families whose choices were constrained because of the absence of adequate day care supply. The remaining observations presumably reflect the

[15]The likelihood ratio statistic at 220.30 is less than the relevant chi-square value of 228.297 for 195 degrees of freedom.

unconstrained demand relationships.[16] A likelihood ratio test (Anderson, 1987) indicates that we cannot reject (at .05) the hypothesis that both constrained and unconstrained observations share the same choice parameters.[17]

Test of Taste Variations

A market segmentation procedure (Ben-Akiva and Lerman, 1987) can be used to test for systematic variation of parameters among population subgroups. We bisect the data set at $17,000 of family income, remove the income variables, and run the original specification on income segments. The resulting likelihood ratio test statistic is 31.08, insignificant at 25 degrees of freedom.

A similar segmentation by values of the variable "non-English speaking immigrants" also produces an insignificant likelihood ratio test value (18.66). These results indicate that the effect of attributes, child variables, mother's employment variables and the remaining socioeconomic variables on the choice of child care arrangements is not significantly different for high and low income families or non-English speaking immigrant families and other families.

Multiple Mode Users

We model child care choice so that each family chooses one and only one type of child care. Where more than one mode is actually used, the type of care attended for the largest number of hours in the week is deemed to represent the family's choice. Most families in the Project Child Care data set use a single type of child care; only 66 of 481, or 14%, use more than one type over the course of a week. Further, 39 of these multiple mode users have only one child (which, arguably, makes the attribution to them of a single main choice more plausible), leaving only 27 of 481 families as true multiple mode users.

Anderson's amended likelihood ratio test (Anderson, 1987) can be used to test whether omitting all 66 multiple mode users, or omitting the 27

[16]This will not necessarily be true of infant day care, which is scarce enough to create a "discouraged parent" effect.

[17]The likelihood ratio statistic, at 138.25, is not significant.

multiple mode users with two or more children, significantly affects the mixed logit results. Although close, neither test rejects the pooling of multiple mode and single mode users in our child care choice model.[18]

Test of the Independence from Irrelevant Alternatives

The Independence from Irrelevant Alternatives (I.I.A.) assumption , a maintained hypothesis of the multinomial logit model, is also tested. The test (Small and Hsiao, 1982) requires splitting the data set into two sections. One half is used to provide logit estimates of the three-way choice: day care centre, sitter, and care by a relative. The other half is used to provide estimates of a two-way choice, after one alternative is dropped. The test, comparing log likelihood values, judges whether the dropping of one of the choices affects our estimates of the remaining two choices dramatically or not; the Independence from Irrelevant Alternatives assumption presumes that dropping one alternative should not significantly alter our results on the alternatives which remain. In our case, the test showed no significant change.

[18]For deletion of the 66 observations, Anderson's test yields 233.84, insignificant at .05 for chi-square with 201 d.f. For deletion of 27 observations, the test yields 98.08, insignificant at .05 for chi-square with 81 d.f.

6

FORECASTING THE NUMBER OF MOTHERS IN THE LABOUR FORCE NEEDING SOME TYPE OF CHILD CARE ARRANGEMENT

Measuring the demand for different types of child care by families having mothers in the paid labour force really requires two steps. One, which we have analyzed in this handbook, is figuring out what percentage of the families which have working mothers and young children will choose each type of child care (and why). The other step is figuring out how many families there are having working mothers and young children.

The second step may seem to be a trivial exercise. After all, each December's Labour Force Survey produced by Statistics Canada tells us how many mothers having children in various age ranges there are in the labour force. However, there are two problems with this simple solution to the problem. First, if we are preparing a forecast of child care demand several years in the future, how do we predict how many families will have young children and mothers in the paid labour force at that time? Second, for either policy analysis or forecasting, how do we predict the effect that things like the price, availability and quality of child care will have on the decisions of mothers to participate in the paid labour force?

This chapter does not provide a complete and definitive answer to these questions. This is an area requiring future work. Nonetheless, this chapter does suggest the best solution currently available and the directions for future research.

THE CURRENT SOLUTION

At the present time, methods of predicting how many mothers of young children will be in the labour force in the future are rough and ready. Essentially, we must combine various population, fertility and labour force estimates to come up with a "best guess" about what the future will bring.[19]

Bali Ram's (1987) study provides an excellent illustration of this method for two parent families. The total number of children in two parent families where the mother was in the paid labour force in 1981 was found from Census data. Two alternative scenarios of future birth rates were taken from projections by Statistics Canada (1985). Recent trends in the growth of labour force participation of mothers were then assumed to continue through to the year 2001. Together, these numbers are used to provide alternative guesstimates (a high scenario and a low scenario) of the number of children who will probably use some type of child care in the year 2001.

The advantage of this technique is that it provides us with a readily understandable forecast of the total number of families or children who will use some type of child care. Our estimates can then be applied to this base to predict how many will use each type of care. The disadvantage of this technique is that the prediction of the number of families using child care is completely insensitive to all child care variables. In other words, if, for instance, our forecast predicts that the price of day care centre care will triple over the forecast period, this does not have any effect on our prediction of the number of mothers who will decide to participate in the paid labour force. We may or may not regard this as a reasonable assumption; if the effect of child care variables on mothers' decisions to work is small, then it may be a reasonable approximation. If the effect of child care variables on mothers' work decisions is large, the assumption may (partially) invalidate our forecasts.

The same statement may be even more true for policy analysis. Table 4 in Chapter 4 provided predictions of the effects of certain changes in child care policy on the percentage of families who would decide to use

[19]Existing macroeconomic models operated by some universities and consulting firms may also be able to provide future population and labour force estimates as a byproduct of regular economic projections.

each type of child care. In each case this is a prediction of the effect *on families who currently have mothers in the paid labour force.* Yet each policy change could also affect the number of mothers in the labour force. If the Child Care Expense Deduction rises, we expect that some mothers currently outside the labour force will decide to enter, because their potential take-home pay from a job has now risen. If the Child Care Expense Deduction is eliminated, and not replaced with anything else, we expect that some mothers will decide that it is no longer sufficiently economically rewarding to continue paid employment. However, our analysis of the effects of child care policy in Table 5 does not account for this.

RESEARCH DIRECTIONS

Several authors have developed models which analyze how mothers' labour force participation decisions are related to child care choice. Heckman's work (1974) is probably the best known amongst economists, but Blau and Robins (1988) and Connelly (1988; 1989a; 1989b) have made more recent and more accessible contributions. Yaeger (1978) also discusses modelling child care choice and labour force participation, although her empirical work deals with child care choice alone. These contributions are reviewed in Cleveland (1990a).

From this work, future research may develop a more complete estimable model of the linked child care choice and labour force participation decisions, which will allow for the prediction of both choices from a consistent set of estimates of the key coefficients. One promising avenue would be to assume that child care choices are made in a nested fashion. Labour force participation decisions would, in this case, be affected by the anticipated values of child care variables, then child care decisions would be taken conditional on a fixed labour force participation decision. Alternatively, the two decisions could be modelled in a fully simultaneous framework.

7

CONCLUDING OBSERVATIONS

We have seen that current techniques for analyzing parents' choice of different types of child care focus on the measurement of needs and preferences. The need for child care is usually determined by adding up the number of families who have certain economic and social characteristics. Parental preferences for different types of child care are generally determined by responses to survey questions about the type of care families would prefer to use. These techniques generally find a considerable gap between either needs or preferences and current use of licensed or regulated forms of child care. This gap is attributed to prices which are too high, inadequate availability of services and substandard quality: barriers which prevent parents from realizing their true preferences.

There are two main problems with these techniques. First, they do not tell us how many families will actually use any particular type of child care under prevailing conditions of price, convenience and quality. Self-reported preferences for a particular type of child care do not generally specify what price, level of convenience, or level of quality should be assumed by the respondent when the preference is stated. Different respondents have different assumptions in mind, no doubt. Second, these techniques give us no way of figuring out how the effect of different child care policies, changes in family situations, or changes in work patterns will affect the demand for different types of care.

Given the right data, the statistical technique suggested and described in this Handbook overcomes both of these problems. The price, convenience and quality situation in which choice occurs can be directly recorded and incorporated in the estimation procedure; this technique

estimates a very adaptable mathematical equation which can be used to project the average reponse of families to changes in any of the underlying variables. Further, the estimated equation can be used to calculate which families will be most affected by any particular policy change so that government child care policy can be designed for maximum impact on families in greatest need.

Our suggested statistical technique is not a magic elixir, of course. It is only valid to the extent that its underlying assumptions are reasonable. One of the more potent of these assumptions is that families do consciously choose a certain type of child care. Our statistical technique uses data on families' child care use patterns to calculate which factors affect child care choices by families. If families do not really choose child care, but are forced into using something they really do not want with no alternative, then our statistical technique will not produce sensible results.

However, it does seem reasonable to suggest that families choose, quite consciously, the type of child care their children use. They do not necessarily choose the child care of their dreams -- their ideal type of care. But families generally seem to do the best they can, under the circumstances. These circumstances include the price, availability and quality of the care they find around them, as well as their incomes, work situation and family needs. So the choice process is certainly not unconstrained, but it is truly a choice process. Our statistical technique seeks to measure the effect of each of the different types of constraint on the child care choices families make.

Current techniques of measuring parental preferences for child care may provide a rough approximation of future demand. Our statistical technique will give more reliable answers to questions about how demand will change as conditions change, as well as providing a quantitative analysis of the effects on demand of alternative kinds of child care policy. Some investment in development of statistical skills is necessary to use the technique; the payoff would seem to make the investment well worth while.

APPENDIX A

COMPUTER SOFTWARE AND HARDWARE FOR ESTIMATING LOGIT MODELS

The purpose of this Appendix is to describe and evaluate some of the more popular software packages for estimating what are generally called discrete choice or qualitative choice models. Multinomial logit is the qualitative choice model of greatest interest for us.

Because of the virtual explosion of statistical software that has entered the market over the last few years, this Appendix cannot hope to be exhaustive. Rather it will focus on a few of the more popular packages, point out strengths and weaknesses and indicate factors to consider in the evaluation of any alternative statistical software.

It may surprise you to hear that qualitative choice models can now be quite comfortably estimated on personal computers. In fact, much of the PC-based software is superior to many of the popular mainframe statistical programs. In addition, econometric software is not prohibitively expensive. Prices are similar to those one might expect to pay for a typical spreadsheet package.

This Appendix will review some of the more widely known and used mainframe statistical software. It will be noted that some of these packages are not well suited for estimating the models of child care demand outlined in this monograph. A number of PC-based packages will then be examined. While each of these packages has its own features, the ability to handle large data sets remains a stumbling block for all but one of the packages. Some of the discussion will be devoted to a consideration of the hardware requirements which are absolutely necessary, and to other configurations which make life much easier. It will be noted that math co-processors, expanded memory and large hard disks will significantly expedite the statistical work. Summary comments will conclude this Appendix.

MAINFRAME SOFTWARE

For many of those who have done any statistical work in university courses or during their professional career, the mainframe computer has been the traditional workhorse. Among the most widely used mainframe statistical packages are SAS and SPSS-X. TSP has been popular for analyzing time-series data.

Both SAS and SPSS-X are excellent for data manipulation, report writing, file handling and numerous statistical routines. However, neither is outstanding for estimating qualitative choice models, in general, and the multinomial logit model,in particular.

SAS does have some routines such as PROC MLOGIT and PROC LOGIST for estimating logit models (only the former estimates multinomial logit models), but these routines are part of what SAS calls its "supplemental library". The routines in this supplemental library are typically written by SAS users. They are provided to all the installations that license the base SAS software, but not all installations make them available. These routines are not supported by SAS. It is possible to purchase a Supplemental Library User's Guide but if this manual does not answer all of your questions, you are on your own when troubles arise. Your only hope may be to call the author of the software and ask him/her for help.

Nonetheless, if you only have access to SAS PROC MLOGIT, estimating multinomial logit models with it is certainly possible. The Project Child Care example discussed in Chapter 4 was estimated using this SAS routine. The MLOGIT procedure has a number of idiosyncracies, many of which are specific to the computer environment in which SAS is operating. A good example of such a peculiarity is that if the procedure is not written entirely in capital letters, it will not run. Another significant problem is the absence of diagnostic information provided by SAS if the PROC MLOGIT routine fails to execute.

SPSS-X was designed as a statistical package for the social sciences. It has some nice data and file handling capabilities and performs a number of statistical functions. Since it was not designed as a heavy duty econometrics package, however, its capabilities for logit analysis are modest. It will perform simple binary logits but the number of explanatory variables is restricted to 20. It has no capability to estimate multinomial choices.

TSP and SHAZAM and LIMDEP are other popular mainframe packages that have qualitative choice modelling capabilities of various dimensions. Since these packages are also available in PC versions, we will discuss their features in the next section. In brief, both TSP and LIMDEP are superior to SAS and SPSS-X for estimating qualitative choice models. However, neither TSP nor LIMDEP has the data handling and manipulation features of SAS or SPSS-X. Ultimately, the user might find the most effective strategy is to get the data into shape using SAS or SPSS-X, then write the subset of relevant variables to a separate file to do multinomial logit analysis with one of the speciality packages.

PC-BASED SOFTWARE

While there are a number of statistical software packages for personal computers, this section will concentrate on three: TSP, SHAZAM, and LIMDEP. These broadly represent the bulk of what is available and most other PC packages bear close similarities, especially to TSP and SHAZAM. Spreadsheet packages, such as LOTUS 1-2-3 and SYMPHONY, while including some basic statistical functions, do not have the capability to perform multinomial logit estimation without a considerable amount of programming. Spreadsheets can, however, be convenient for manipulating data, particularly when the data sets are not extremely large.

PC TSP is a fully interactive statistical package which performs a number of econometric routines. It is capable of writing and reading LOTUS 1-2-3.WK* which can be handy if data input comes via a spreadsheet. It will also read regular ASCII data files.

PC TSP and Mainframe TSP are both capable of estimating multinomial logit models. The PC version does, however, have a serious drawback. The working space available for estimation is restricted to about 8000 data points (that is, the number of observations times the number of variables). Put differently, a multinomial logit model with 10 explanatory variables could not be estimated for any data set exceeding about 800 observations (assuming the machine being used has a memory of 640K). Note that the Project Child Care example described in Chapter 4 used 36 explanatory variables.

A version of TSP for 286 and 386 machines that have at least 2Mb of extended memory (in addition to the base 640K), and a math co-

processor, makes available a working space of 50,000 data points. In this case, a multinomial logit model with 10 explanatory variables could be estimated for a data set of about 5000 observations. It should be noted that TSP is capable of many other statistical procedures. These include ordinary least squares, probit and tobit regression.

To summarize, TSP is capable of multinomial logit estimation in addition to more standard econometric techniques. It also has the advantage of being familiar to many individuals who have taken econometrics at university. On the downside, it has important limitations on the size of the model it can estimate, unless considerable PC hardware resources are available.

SHAZAM is another popular PC econometrics package. While it can do a number of useful things very well, it cannot directly perform multinomial logit analysis. If one was particularly careful and resourceful, it would be possible to estimate the multinomial logit using other routines in SHAZAM. Since PC packages are available with "canned" multinomial logit routines, it is generally advisable to opt for these more convenient packages. The lesson here is that not all PC statistical packages, including those more widely used, are readily suitable for multinomial logit estimation. Thus, it is important to shop with caution.

LIMDEP is a statistical package that is specifically designed for estimating qualitative dependent variable models, such as the multinomial logit model, and limited dependent variable models such as Tobit (hence the name LIMDEP: LIMited DEPendent variables). Like the other packages, LIMDEP can also handle more traditional statistical analyses. It is capable of reading data in spreadsheet and DIF formats as well as direct keyboard data entry and standard file types like ASCII. For most routines, including the multinomial logit model, there is no limit on the number of observations that LIMDEP will handle. This is a key feature that makes this package the preferred choice for multinomial logit analysis.

LIMDEP will run on IBM and compatible machines with at least 640K of memory and two diskette drives. The speed at which statistical procedures can be performed improves greatly if you have access to a 386 machine with a hard disk and a math co-processor. Extended memory that permits creation of a sizeable RAM disk can also help speed

things up. The benefits of a faster machine should not be underestimated. Multinomial logit estimation is very memory and input-output intensive. Gains in speed can be dramatic, reducing the time required to estimate complicated models on large data sets from hours to less than one hour.

LIMDEP output for multinomial logit estimation includes maximum likelihood estimates of the model parameters, standard errors, t-statistics, means and standard deviations for each of the independent variables. In addition, it produces log-likelihood values for the estimated model and the restricted model (where the slope coefficients are constrained to zero), the chi-square statistic and the significance level for the likelihood ratio test. LIMDEP does not provide other "goodness of fit" tests such as the Maddala, Cragg-Uhler, McFadden R^2's as part of its routine output. These can be quite easily calculated with the output which is generated, however.

Those particularly comfortable with statistics and computer programming might want to consider another mathematical package called GAUSS. This software permits the user to program statistical routines relatively easily. Therefore, the only restrictions lie in the time, resources and ingenuity of the user. Programming one's own routines does require a strong knowledge of math and generous availability of time. The company that manufactures GAUSS has developed a number of preprogrammed statistical routines to reduce the onerous programming burden somewhat.

RECOMMENDATIONS

In summary, the optimal hardware and software configuration for estimating multinomial logit models of child care demand would include:

1. an 80-286 (AT) or 80-386 based IBM or compatible microcomputer with at least a forty megabyte hard disk and 640K of memory. An 80-387 math co-processor and a clock speed in excess of 12 Mhz, and some extended memory would greatly improve performance.

2. The LIMDEP software is preferable to other packages due to its wide array of qualitative choice modelling capabilities. A good

data base package, such as dBase or FOCUS, will prove useful in managing large data bases and performing the manipulations necessary to construct the variables which will ultimately be used for multinomial logit estimation. While LIMDEP, like the other statistical packages, is capable of most types of data manipulation, maximum efficiency is obtained when the data set to be analyzed is in final form before being read into the statistical program.

3. Finally, it is useful to have a good, full screen editor such as KEDIT or BRIEF. While LIMDEP can be used interactively, it is often easier to write programs and submit them to LIMDEP in batch mode. Experienced users will know that the EDLIN line editor which is supplied with DOS is not particularly convenient. Editors are generally inexpensive. BRIEF, for example, costs approximately $100.

4. If mainframe resources are available, LIMDEP remains the statistical package of preference. Access to SAS or SPSS-X would serve the useful purpose of getting the data "into shape".

Additional information on the statistical software packages discussed can be obtained from the following sources:

TSP
TSP International
P.O. Box 61015, Station A,
Palo Alto, California, 94306
U.S.A.
Telephone: (415) 326-1927

SHAZAM
SHAZAM,
Dept. of Economics,
University of British Columbia,
Vancouver, B.C. V6T 1W5
Telephone: (604) 228-5062
Fax: (604) 228-5915

LIMDEP
Econometric Software,
Church Street Station,
P.O. Box 3526,
New York, New York,
10008-3526 U.S.A.
Telephone: (516) 822-2389
Fax: (516) 938-2395

GAUSS
APTECH Systems, Inc.,
26250-196th Place SE,
Kent, Washington,
98042
Telephone: (206) 631-6679
Fax: (206) 630-1220

APPENDIX B

HOW TO USE THE 1988 NATIONAL CHILD CARE SURVEY

The methods presented in this handbook could be used with data gathered in different ways from various sources. One useful, recently collected, and very large data sample is the 1988 National Child Care Survey. This survey was conducted by Statistics Canada in co-operation with the National Daycare Research Network and Health and Welfare Canada. The National Daycare Research Network is an interdisciplinary group of researchers from four Canadian universities, headed by Dr. Donna S. Lero from the University of Guelph. Dr. Alan Pence of the University of Victoria is co-director, while Dr. Hillel Goelman of the University of British Columbia and Dr. Lois Brockman of the University of Manitoba are co-investigators.

The National Child Care Survey was administered as a supplementary survey along with the regular monthly Labour Force Survey in September 1988. As a result the population from which the sample is taken, and the design of the sample are similar to that of the Labour Force Survey. This implies, for instance, that the Yukon and Northwest Territories, and persons living on Indian Reserves are not included in the sample. The final sample covers the rest of Canada quite thoroughly, however, with data on the child care choices of over 24,000 families having over 42,000 children under 13 years of age. The questionnaire administered is nearly 100 pages long; the result is a mountain of data on the child care patterns and opinions of Canadian families. Of course, the size of any particular sub-sample of interest may be much smaller than the overall 24,000 family total. For instance, there is data on about 3000 families in Alberta. If our sub-sample comprised Alberta families having an employed mother and at least one preschool child, it would include considerably less than 3000 observations.

The purpose of this Appendix is to outline the type of model which could be estimated using the National Child Care Survey (NCCS) data. We consider the Who, What and Why questions as outlined in Chapters 3 and 4. We discuss some difficulties that arise in defining certain explanatory variables. Finally, we comment briefly on how the results could be used.

THE SAMPLE

The NCCS allows us to choose from among a number of possible samples. We could consider only the choices of families with employed mothers, or those with both parents employed, or both parents employed or going to school full time, or those with parents employed, going to school or not currently in the paid labour force. Of course, we could also look at any particular subsample of interest, such as families headed by a single parent, families with student parents, families with only schoolage children, families with only preschool children, and so on. Our choice of a sample will depend upon the purpose of our estimation; many alternatives are equally valid.

THE DEPENDENT VARIABLE

The NCCS provides information on the method or methods of child care chosen for each child in the family. In addition, one child in the family is randomly selected (the "target child") and more detailed information collected on the child care choice for that child. Thus, the NCCS allows for estimation of models which seek to explain (a) any and all methods of care used by each child in a family, (b) any and all methods of care used by each child while parents are engaged in work-related activity, (c) the primary care arrangement of each child in a family, and (d) the main method of care used by the "target child" in a family.

The model illustrated in this handbook considers the choice between a day care centre, sitter and care by a relative. The NCCS will permit a greater number of alternatives to be defined, if this is useful and gives reasonable results. For instance, licensed family home day care could be considered as a separate child care alternative. Likewise, care by the father, care by the mother while at work, and care in a nursery school or kindergarten could, potentially, be considered as distinct child care arrangements. For schoolaged children, self care or sibling care may be an important alternative.

The menu of possibilities is tempting; however, it is necessary to ensure that the alternatives are dissimilar enough that the information on the independent variables will allow us to distinguish them. That is, we must be able to determine, for example, that there is something measurably different about families which use licensed family home day care and those which use unlicensed sitters. While it is true that a more

detailed model comes closer to reflecting the "real world", it is necessary to ensure that there is sufficient data to uniquely identify, in a statistical sense, each of the types of child care.

THE EXPLANATORY VARIABLES

The NCCS would allow nearly all of the explanatory variables provided in our concrete example in Chapter 4 to be used. The only exceptions are the mother's wage and some measure of the convenience of access to different types of child care. The NCCS would allow some additional explanatory variables to be included. For instance, we can identify whether a particular child has a disability, which could affect child care choice. Our choice of explanatory variables will depend upon our beliefs about what factors are important in families' child care decisions.

DIFFICULT EXPLANATORY VARIABLES

In common with most data sets on child care arrangements (including the Project Child Care Survey), there is one group of variables which provide considerable difficulty. A large amount of the desired data is absent, and approximations for the missing data must be found. These are the data which describe the attributes (generally, price, convenience and quality) of each type of child care from which a family chooses.

It is normal to collect information on the attributes of the type of care, or types of care, that a family ends up choosing. The NCCS provides a considerable amount of very useful data of this kind. It is atypical to find a data set that includes information on the attributes of the types of child care which a family decided not to choose. However, the high prices, inconvenience and poor quality of the types of child care not chosen are certainly important explanatory variables which we want to include.

It does not appear possible to find, in the NCCS, useful approximations for the convenience and quality of the types of care not chosen. However, it should be possible to develop reasonably good approximations for the price of methods of care not chosen. For instance, we could follow Hofferth and Wissoker (1990) in regressing prices of each type of child care on a series of family characteristics.

This would allow us to predict the missing prices on the basis of available data on family characteristics.

The key to a good child care price regression is to find variables which are good predictors of those prices which are substantially below normal market levels. In particular, we need to predict which families will be eligible for day care subsidy and which families will have low priced care available from a relative. The NCCS provides considerable data to allow us to do this.

USING THE RESULTS FROM THE NCCS

The 1988 National Child Care Survey is the most current and extensive data set available for determining the state of child care in Canada. Its strengths include detail, sample size, and national coverage (with the exception of the Yukon and Northwest Territories). It provides nearly all the variables which allowed successful construction and estimation of the child care demand model illustrated in Chapter 4. The menu of child care alternatives, and the treatment of multimodal use are considerably more extensive than previous data sets have provided. The NCCS offers the possibility of building models which can be disaggregated by regions of the country, making the model more representative of the special circumstances of each province.

The estimates produced could be used to forecast the use of different types of child care and analyze the impact of alternative child care policies on demand for care. It would be possible to develop menu-driven software to facilitate the preparation of forecasts and policy analyses by a range of governmental and non-governmental users.

A FOOTNOTE ON FUTURE DATA COLLECTION

Data collected in the future to calibrate models of the type we have described, should consider carefully how best to gather data on the attributes of child care services. Ideally, we want objective, rather than subjective, measures of the price, convenience and quality (perhaps there are additional relevant attributes) of each type of child care available to each family.

Yaeger's (1978) solution to this problem was to ask each family direct questions about the child care alternatives available. For instance:

6.b. How much money do babysitters in your area usually charge for daytime child care (for a child under 6)?

a. No charge
b. About $0.25 per hour
c. About $0.50 per hour
d. About $0.75 per hour
e. About $1.00 per hour

6.c. Is there any (unrelated) babysitter available for regular daytime child care in your home?

Yes.....Y
No.....N

6.d. How long would it probably take you to travel from your home to the home of an available daytime babysitter (unrelated to you)?

a. Less than 10 minutes
b. Between 10 and 19 minutes
c. Between 20 and 29 minutes
d. 30 minutes or more

6.e. Would daytime child care by someone unrelated probably cost "more than", "less than", or "about the same as" child care by someone related?

More.....M
Less.....L
Same.....S

7.a. In a child care centre or school, how much money would full-day child care probably cost you, for a child under 6?

a. Less than $10.00 per week
b. From $10.00 to $20.00 per week
c. From $25.00 to $35.00 per week
d. $40.00 per week or more

7.b. How long would it probably take you to travel from your home to an available child care centre or school for a child under 6?

a. Less than 10 minutes
b. Between 10 and 19 minutes
c. Between 20 and 29 minutes
d. 30 minutes or more

7.c In a child care centre that you know of or have heard about, how many children under 6 are normally cared for, per adult?

a. Up to 3 children per adult
b. 4 to 5 children per adult
c. 6 to 7 children per adult
d. 8 children or more per adult

These questions were designed to collect information on the price, convenience (i.e., travel time), and quality (i.e., adult-child ratio), of all types of child care available to each family in the data set.

In some situations, it may, alternatively, be possible to determine the value of some attributes from outside information. For instance, Hofferth and Wissoker (1990) determine the prevailing child-staff ratio in day care centres by assuming that regulations are followed in each of the 50 states. In transportation studies, such as Domencich and McFadden (1975), the cost of and time required for transportation to work by alternative modes of transport (e.g., car and public transit) are laboriously calculated for each family. This requires collection of information on the location of the residence and work for each family, and the use of road maps and transit schedules, average car costs per mile and for parking, transit fares and the like to compute costs and times.

When neither of these alternatives is available for determining attributes, techniques such as the price regressions described above are necessary.

To obtain the 1988 Canadian National Child Care Survey, write, phone or fax:

Michael Sivyer,
Special Surveys Group,
Statistics Canada,
Jean Talon Building 5A6,
Tunney's Pasture, Ottawa K1A 0T6.
Phone:(613) 951-4598. Fax:(613) 951-0562.

To gain access to the data, you will have to sign a Microdata Licensing Agreement with Statistics Canada, pay $5000 plus G.S.T. and receive after sales service assistance with the data set. Alternatively, you may wish to read and use the published reports which will come from Statistics Canada over the next period of time, or have Statistics Canada prepare custom tabulations for you at a price.

APPENDIX C

A REVIEW OF RESULTS FROM EMPIRICAL STUDIES OF CHILD CARE CHOICE

There have been a considerable number of studies of child care choice behaviour over the last thirty years, since the rising labour force participation of mothers began to raise public policy concerns about the care of their children during the working day. Many of these studies have had an empirical component, but the theories tested and the techniques used have varied widely. A complete analytical review of what we can learn from these studies remains to be done.

In this appendix, we focus on two groups of studies: those analyzing Canadian data and those using regression techniques to examine the child care choice decisions of employed mothers. We therefore do not consider, for instance, studies by Heckman (1974), Blau and Robins (1988), or Connelly (1989a, 1989b).[1]

CANADIAN STUDIES

There are two distinct types of Canadian study of child care use. First, there is the purely statistical survey. There have been three cross-Canada statistical surveys of patterns of child care choice: one by the Women's Bureau of the Department of Labour (1970), and two by Statistics Canada (1975; 1982). A very large and comprehensive new set of child care data, the National Child Care Survey 1988 collected by Statistics Canada, will soon be publicly available. Payette and Vaillancourt (1984) and Henriques and Vaillancourt (1988) provide some further analysis of Statistics Canada (1982); otherwise none of these Canada-wide data sets have been subjected to in-depth, multivariate analysis.

[1] These studies examine mothers' decisions to enter the paid labour force together with their choice of child care type (or amount of expenditure on child care).

Second, there is the use of theory and data analysis techniques to reach conclusions about child care, based on a particular data set. There have been only a small number of such quantitative analyses of child care choice using Canadian data, apart from local needs assessment studies. We will briefly review some central conclusions from six: Johnson (1977), Lero (1981), Stevens (1984), Payette and Vaillancourt (1984), Lero et al. (1985) and Henriques and Vaillancourt (1988).

Johnson

Johnson (1977) studied the child care arrangements of preschool children in Metropolitan Toronto. On the basis of cross-tabulations, Johnson concluded that a number of factors were key in determining whether a family would use a day care centre, a sitter or care by a relative: ethnicity, family size and structure, marital status, family income, the cost of care and the distance of the care arrangement from the family home.

The Project Child Care Survey also probed parental preferences. Johnson found a range of answers according to how questions were posed to parents; when parents were asked about ideal arrangements for their own children, regardless of constraints imposed by family finances or time schedules, day care centres were preferred by nearly half of respondents.

Lero

Lero's (1981) study was based on samples of mothers in Kitchener, Ontario who used full-time child care arrangements for their preschool children in day care centres, licensed family home day care and informal care. Results, based on cross-tabulations, supported Lero's contention that child care choice is a complex process; a central contribution of this study was to explicitly model the various factors and stages in this choice process.

Lero's model separates child care choice into two main stages: the formation of preferences for a particular mode of care, then the search for this mode and the adaptation to constraints in the supply of services available. As a result, actual use may diverge widely from preferences. Lero's results allow comparison of the profiles of the typical users of alternative modes; the construction of the sample does not permit

inferences about the effect individual factors would have on the probability a certain mode would be chosen.

Informal care users were, typically Canadian-born married couples with two children. Cost of child care was not ranked as a major criterion of choice; most genuinely preferred informal arrangements. Only social assistance recipients were permitted to use licensed family home day care in the municipality sampled by Lero. These licensed family home day care users were typically single parents of British background with two children. They had limited education, a family income of $7,000 to $11,000 and received a day care subsidy. Children were likely to be older. Non-profit day care centre users were likely to be married with two children and of British descent. Some received subsidy but most were well-educated middle class parents with a family income of $25,000 or more. Non-profit centre users tended to have permissive child rearing attitudes and practices. Private day care centre users tended to be two parent families with one child, of German or Western European descent, and with a family income over $25,000. Many of these children were under three years of age. Day care centre users rated much more highly than users of other modes those dimensions of the quality of care relating to educational and social stimulation.

Stevens

On the basis of a telephone survey of preschool and schoolage children in the City of Winnipeg, Stevens (1984) analyzed the demand and supply of child care and the quality, accessibility and need for child care arrangements. He found that a large proportion of Winnipeg's preschoolers in care arrangements were using a licensed day care facility. Sitter care was the most prevalent and care by a relative the least prevalent arrangement. Stevens emphasizes the importance of the age of the child, the cost of the arrangement, and the quality and accessibility of the arrangement on the decision to use a particular type of care.

Payette and Vaillancourt

Payette and Vaillancourt (1984) used the Statistics Canada 1981 Survey of Child Care Arrangements to study child care choices of Quebec families with children 0-4 years of age. This study used a series of probits to analyze the mother's decision to work, to use non-parental

child care services, and to choose a day care centre rather than another form of non-parental child care.

The study found that a mother would be more likely to work the older her children, the smaller the number of children in the family, and the higher her remuneration from work (proxied here by age and education). A family would be more likely to use non-parental child care services if the child was age four, the fewer the number of children in the family, the higher the education level of the husband and the greater the work committment of the mother.

Families using non-parental care were more likely to use a day care centre for 3 and 4 year olds, the higher the education level of the mother and by women working part time or not in the labour force rather than by women working full time. This latter result was apparently due to the greater availability of day care subsidy to women of lesser means. Family income typically had no effect on the decision to use a day care centre, except that families with an income level of $20,000 to $25,000 were less likely to use day care than families at the lowest income level.

Lero et al.

Lero et al (1985) is a study of the patterns of child care use, need and preference by 336 Canadian families. The study was supported by the Katie Cooke Task Force on Child Care as an exploration of the range of child care needs experienced by Canadian families. The data gathered is not derived from a random sample survey and results can not therefore be interpreted as representative of all Canadian families. The study argues persuasively that "needs" for child care should be interpreted broadly; child care plays a multiplicity of roles in supporting the functioning of Canadian families, not just the replacement of mothers when they go to work or otherwise are unable to play a primary caregiving role.

Henriques and Vaillancourt

Henriques and Vaillancourt (1988) used the data set collected by Statistics Canada as a supplement to the Labour Force Survey in February 1981. Probits were employed to analyze the decision of families with children 0-4 years of age to use non-parental child care services as opposed to entirely parental care and the decision of these child care users to use a day care centre or not.

Henriques and Vaillancourt concluded that the decision to use child care services was positively influenced by the expected returns to the mother (measured by education and work status) or to the child (presumed to increase with age), and was negatively affected by increased cost (the number of children). Similarly, the decision to use a day care centre was made more likely by increased rewards (age of child and education level of mother) and by increased likelihood of subsidies (reflected by low income of the mother).

MULTIPLE REGRESSION ANALYSIS OF CHILD CARE CHOICE

In this section, we compare our results, as presented in Chapter 4, with those from other studies which analyze the child care choices of employed mothers using multiple regression. In particular, we consider Hofferth and Wissoker (1990), Yaeger (1978), Robins and Spiegelman (1978), Lehrer (1983), Henriques and Vaillancourt[2] (1988), and Leibowitz, Waite and Witsberger (1988). Each of these studies examines the decision by families to use a day care centre rather than some other alternatives. To provide a reasonably standard basis for the comparison of results, we focus on the choice of a day care centre, as analyzed in these empirical studies. This type of comparison is necessarily hazardous because the origin and design of samples, definition of dependent and independent variables and statistical techniques vary markedly. The exercise seems useful, nonetheless, if only approximate.

Comparison of Results

Wherever the price of day care is used as an explanatory variable (Hofferth and Wissoker; Yaeger; Robins and Spiegelman), its effect is negative and significant as found in our own study.[3] Convenience or availability variables (Yaeger; Leibowitz, Waite and Witsberger) inevitably show that increased convenience or availability makes choice of day care

[2]Henriques and Vaillancourt's data include families with mothers outside the paid labour force as well as those inside the paid labour force.

[3]Hofferth and Wissoker's (1990) price results are no longer significant when the generic price specification is relaxed.

more likely. Quality variables (Yaeger; Hofferth and Wissoker) inevitably show that higher quality increases the likelihood of choice. In other words, this group of studies strongly suggests that the price, availability and quality of day care are important factors influencing the decision of families to use this type of child care.

Presence of infants (variously defined) in the family makes the choice of a day care centre less likely, except in Lehrer where the infant category is broadly defined to include children 0-3 years of age.

Mother's hours of work, wage rate, and education are sometimes significant in the choice of day care and sometimes not (Hofferth and Wissoker; Lehrer; Robins and Spiegelman; Yaeger; Henriques and Vaillancourt). When significant, they inevitably bear a positive relationship to the likelihood that day care will be chosen.

Family income or husband's income are typically insignificant in these studies for explaining choice of a day care centre (Hofferth and Wissoker; Yaeger; Robins and Spiegelman; Leibowitz, Waite and Witsberger). Henriques and Vaillancourt's significant finding for family income is apparently due primarily to ineligibility of higher income earners for day care subsidy. Only Lehrer contradicts the basic pattern of insignificance for husband's or family income, finding husband's permanent income (a constructed variable depending on husband's education, age, occupation and some other factors) to be highly significant in explaining the choice of an "organized facility" rather than a relative, though not for the choice of an "organized facility" rather than a sitter.

Overall Conclusions

In general, the pattern of results from these six empirical studies of child care choice, confirms our results as presented in Chapter 4. It seems reasonable to conclude that the attributes of child care services, the ages and number of children in the family, the mother's work situation, and some socio-economic factors (not generally the level of family income on its own), all have important effects on families' child care choices.

APPENDIX D

TECHNICAL DETAILS OF LOGIT STATISTICAL MODELS

It is customary to assume that families seek to maximize utility when they choose a child care arrangement for their children. It is possible, then, to describe a utility function $U_{ij} = U_{ij}(x, z)$ where "i" subscripts the i'th family, "j" subscripts the j'th type of child care arrangement (e.g., day care centre, sitter, care by a relative), and x and z are some group of variables which determine the amount of utility this family gets from choosing this particular type of child care.

Further, it seems appropriate to describe this as an indirect, rather than direct, utility function. The latter is an expression of pure preference, with no accounting for the effects of constraints upon choice. Instead, our utility function measures the maximum attainable utility from the j'th choice, given the price and income constraints facing the family -- an indirect utility function.

We assume that the utility function for each type of child care is common to the group of families in which we are interested, so that the choice behaviour of this whole group of families can be used to provide us information about the unknown coefficients of the function. Naturally, the utility function for different choices is different (i.e., $U_j(X, Z) = U_k(X, Z) = U_l(X, Z)$). It is normal to assume that each utility function is linear, and in order to allow for maximum flexibility of the way in which explanatory variables are specified, linearity in the parameters is normally assumed.

UNOBSERVED UTILITY AND THE LOGIT MODEL

Families are assumed to always choose the type of child care which gives them the greatest utility total. However, it is impossible for us to measure with completeness and accuracy all factors that affect a family's utility from child care. Therefore, even if our assumptions about

the commonality and linearity of the utility function are accurate, the calculation of the values of the unknown coefficients is not simply a mathematical exercise; in addition to the unknown coefficients, there is an unknown amount of error in our measurement process. We account for this unobservable utility -- a different amount for each family -- by adding an error term to the utility function for each choice for each family. So, using "V" to represent the systematic component of utility and "E" to represent the random (more accurately, unobserved and therefore random from the point of view of the observer) component of utility, we have $U_{ij} = V_j(X_i, Z_{ij}) + E_{ij}$, for all j.

If we are considering a choice between only two alternatives, j and k, then alternative j will be chosen if $U_{ij} > U_{ik}$, and alternative k will be chosen if $U_{ik} > U_{ij}$. To put it another way, we can say that the probability that alternative j will be chosen by family i (denoted $P_i(j)$) depends on whether the difference in the size of the unobservable utility components of the two alternatives is less than the difference in the size of the systematic components; in symbols:

$$P_i(j) = P[(E_{ik}-E_{ij}) < (V_{ij}-V_{ik})]$$

The probability that a randomly distributed error term (or the difference of two errors) is less than some determinate amount is given by the cumulative distribution function of that error term; therefore if we know the class of distributions of which the error term is a member, we may be able to express this probability in terms of an explicit functional form, leaving a relatively small set of unknown coefficients to estimate.

If we assume that the error term for each alternative is distributed as Type I extreme value, that the distributions are identical across families and that the error term for one alternative is statistically independent of error terms for other alternatives, then the difference of the error terms $(E_{ik}-E_{ij})$ will be distributed as logistic. In this case the probability of choosing an alternative takes the following explicit form:

$$P_i(j) = e^{V_{ij}}/e^{V_{ij}}+e^{V_{ik}}$$

Although it is less simple to depict, the same argument applies when the choice occurs amongst several, rather than just two, alternatives and, in this case, the probability of choosing the jth alternative is expressed as:

$$P_i(j) = e^{V_{ij}}/\sum_m e^{V_{im}}$$

"e" here is the base of natural logarithms, approximately 2.718. What

the above probability expression does is to express the probability of a particular choice as (some function of) the utility given by that particular choice divided by (some function of) the total utility given by all choices. As the ratio of the utility given by the jth choice to that given by all choices rises, the probability of the jth choice will rise. Obviously, the probability of the jth choice cannot be less than zero, nor can it be greater than one; accordingly it satisfies the requirements of a probability measure.

INDEPENDENCE FROM IRRELEVANT ALTERNATIVES

The chief limitation on the applicability of the logit model to any particular set of choices is something known as the Independence from Irrelevant Alternatives (I.I.A.) assumption. Some implications of the I.I.A. property can be seen by taking the ratio of the probability of choosing any two alternatives in the logit model. $P_{ij}/P_{ik} = e^{V-V}$. In other words, the ratio of the probability that a particular family will choose any two alternatives depends only on the utility of those two alternatives and not on the utility of any other alternative. The choice between these two alternatives is independent of the utility of irrelevant alternatives. The validity of this property in logit depends upon the statistical independence of the error term across different choice alternatives for the family. If the error terms are correlated (i.e., if the unobserved component of utility of different alternatives is correlated), then the logit model, with its assumed independence from irrelevant alternatives, is inappropriate.

It is possible to test for the appropriateness of the I.I.A. assumption (i.e., to test for the independence of the error terms), but it is also desireable to take care in defining the choice alternatives, so they are not too similar. Choice alternatives which are similar are, all else equal, more likely to have uncorrelated components of utility which are not independent of one another.

MAXIMUM LIKELIHOOD

The probabilities, as we have stated them above, are used directly in the calculation of the value of the unknown coefficients of the utility functions for child care choice. We use maximum likelihood techniques

in this calculation; in other words, we choose as the estimated values of the unknown coefficients those values that will make it most likely to observe the set of choices that families have made in our sample. These values will make the joint probability of the sample a maximum, where the probability of each observation is, as given above:

$$P_i(j) = e^{Vij} / \sum_m e^{Vim}$$

The log likelihood function, which our computer calculations maximize the value of, is:

$$\ln L = \sum_i \sum_j Y_{ij} \ln P_{ij}$$

where $Y_{ij} = 1$ if the j'th alternative is chosen by the i'th family, but otherwise equals zero.

MULTINOMIAL AND CONDITIONAL LOGIT VARIABLES

There are two types of variables in common use in logit models: multinomial logit and conditional logit. The former, denoted B_jX_i, take a distinct coefficient for each alternative, while the explanatory variable is constant across alternatives. Coefficients can be interpreted as measuring the effect of a one unit change in the explanatory variable on the relative probability that a certain type of child care will be chosen; more precisely each coefficient is the partial derivative of the log of the odds ratio of two alternatives.

In other words, these coefficients measure the change in the (log of) the ratio of the probability of choosing one alternative rather than another when the explanatory variable changes by a small amount. The ratio of the choice probabilities of two alternatives, the jth and the qth, can be written as

$$P_{ij}/P_{iq} = e^{Xi(Bj-Bq)}$$

for the ith family. Therefore,

$$\delta \ln(P_{ij}/P_{iq})/\delta X_{ik} = B_{jk} - B_{qk}.$$

Because multinomial logit can only estimate the difference in the B's rather than the B's themselves, a normalization has to be imposed to give unique values of the coefficients. Conventionally, the coefficients for one alternative are set to zero. In this case, with $B_{qk} = 0$, $B_{jk} = \delta \ln(P_{ij}/P_{iq})/\delta X_{ik}$.

Conditional logit variables, denoted A_jZ_{ij}, take a different variable value,

and perhaps a different coefficient value for each choice alternative (i.e., coefficient values of conditional logit variables may be constrained to equality across alternatives, if desired). The main difference between conditional and multinomial logit variables is that the former are constrained not to apply to irrelevant alternatives. The other difference of note is that

$$A_{jk} = \delta \ln(P_{ij}/P_{iq})/\delta Z_{ijk}$$

while

$$B_{jk}\text{-}B_{qk} = \delta \ln(P_{ij}/P_{iq})/\delta X_{ik}.$$

Information on the use and interpretation of logit models, on statistical tests, and much more, is now readily available. Several good sources are Ben-Akiva and Lerman (1987), Maddala (1983), and Amemiya (1985).

SELECTED BIBLIOGRAPHY

Amemiya, T. 1985 *Advanced Econometrics* Cambridge, Mass.: Harvard University Press.

Anderson, G.J. 1987. "Prediction Tests in Limited Dependent Variable Models." *Journal of Econometrics* 34:253-61.

Ben-Akiva, Moshe, and Steven R. Lerman. 1987. *Discrete Choice Analysis: Theory and Application to Travel Demand* Cambridge, Mass.: MIT Press.

Blau, David, and Philip K. Robins. 1988. "Child-Care Costs and Family Labor Supply." *The Review of Economics and Statistics* 374-81.

Canada Department of Labour. 1970. *Working Mothers and Their Child Care Arrangements* Ottawa: Queen's Printer.

Cleveland, Gordon. 1990a. "The Choice of Child Care Arrangements by Employed Mothers: Canadian Evidence and Policy Rationale." Ph.D. dissertation, University of Toronto.

_____. 1990b. "Calculating The Impact of New Residential and Nonresidential Developments on Municipal Requirements for Child Care Services: A Report To The City of Vancouver." (Mimeographed.)

Connelly, Rachel. 1988. "Utilizing Market Child Care: An Economic Framework for Considering the Policy Issues." Department of Economics, Bowdoin College, Maine. (Mimeographed.)

_____. 1989a. "The Effect of Child Care Costs on Married Women's Labor Force Participation." Department of Economics, Bowdoin College, Maine. (Mimeographed.)

_____. 1989b. "Determinants of Weekly Child Care Expenditures: A Comparison of Married and Unmarried Mothers." Department of Economics, Bowdoin College, Maine. (Mimeographed.)

_____. 1990. "The Importance of Child Care Costs to Women's Decision Making." Department of Economics, Bowdoin College, Maine. (Mimeographed.)

Cooke, Katie; London, J.; Rose-Lizee, R.; and Edwards, R. 1986. *Report of the Task Force on Child Care* Ottawa: Status of Women Canada.

Domencich, Thomas A., and Daniel McFadden. 1975. *Urban Travel Demand: A Behavioural Analysis* Amsterdam: North-Holland.

Duncan, Greg J., and C. Russell Hill. 1975. "The Child Care Mode Choice of Working Mothers." In *Five Thousand American Families: Patterns of Economic Progress* Vol. V. Edited by Greg J. Duncan and James N. Morgan. Ann Arbor: Institute for Social Research.

Eichman, Caroline; Barbara Reisman; and Donna Euben. 1990. *Building Links: Developer Initiatives For Financing Child Care* Child Care Action Campaign Special Report #2: State Financing Alternatives Project. New York: Child Care Action Campaign.

Frenette, Lyse. 1986. "Document de Reflexion Sur Le Plan Quinquennal de Developpement des Services de Garde À L'Enfance." Working Document. Montreal: Office des Services de Garde À L'Enfance, Gouvernement du Québec.

_____. 1987. "La garde des enfants au Québec in 1986: ce que les gens souhaitent, ce que les parents font." Montreal: Offices des Services de Garde À L'Enfance, Gouvernement du Québec.

_____. 1990. "Estimation Regionale des Besoins en Services Regis." Montreal: Office des Services de Garde À L'Enfance, Gouvernement du Québec.

Friendly, Martha. 1989. *Assessing Community Need For Child Care: Resource material for conducting community needs assessments* Toronto: The Childcare Resource and Research Unit, Centre for Urban and Community Studies, University of Toronto.

George Brown College. 1990. "Feasibility Study: Extended Hour Childcare Project" Toronto: George Brown College.

Gunderson, Morley. 1986. "Implications of Daycare Policies on Female Labour Market Behaviour." Report to the Special Committee on Child Care. Ottawa: House of Commons.

Health and Welfare Canada. 1989. *Status of Day Care in Canada* 1988 Ottawa: Health and Welfare Canada.

Heckman, James J. 1974. "Effects of Child-Care Programs on Women's Work Effort." In *Economics of the Family: Marriage, Children and Human Capital* Edited by Theodore W. Schultz. Chicago: University of Chicago Press.

Henriques, Irene, and Francois Vaillancourt. 1988. "The Demand For Child Care Services in Canada." *Applied Economics* 20(3):385-94.

Hill, C. Russell. 1978. "Private Demand for Child Care: Implications for Public Policy." *Evaluation Quarterly* 2(4):523-46.

Hofferth, Sandra L. 1979. "Day Care in the Next Decade: 1980-1990." *Journal of Marriage and The Family* 649-58.

_____. 1989. "What is the Demand for and Supply of Child Care in the United States?" *Young Children* 28-33.

_____ and Deborah A. Phillips. 1987. "Child Care in the United States, 1970 to 1995." *Journal of Marriage and The Family* 49:559-71.

_____ and Doug Wissoker. 1990. "Quality, Price and Income in Child Care Choice." The Urban Institute, Washington, D.C. (Mimeographed.) Amended version published as "Price, Quality and Income in Child Care Choice." *Journal of Human Resources* 27(1):70-111 (Winter 1992).

Hoffman, Saul D., and Greg J. Duncan. 1988. "Multinomial and Conditional Logit Discrete-Choice Models in Demography." *Demography* 25: 415-27

Johnson, Laura Climenko. 1977. *Who Cares? A Report of the Project Child Care Survey of Parents and Their Child Care Arrangements* Toronto: Social Planning Council of Metropolitan Toronto.

_____. 1978. *Taking Care: A Report of the Project Child Care Survey of Caregivers in Metropolitan Toronto* Toronto: Social Planning Council of Metropolitan Toronto.

Kisker, Ellen E.; R. Maynard; A. Gordon; M. Strain. 1989. *The Child Care Challenge: What Parents Need and What is Available in Three Metropolitan Areas* Princeton, N.J.: Mathematica Policy Research.

Lehrer, Evelyn. 1983. "Determinants of Child Care Mode Choice: An Economic Perspective." *Social Science Research* 12:69-80.

Liebowitz, Arleen; Linda J. Waite; and Christina Witsberger. 1988. "Child Care for Preschoolers: Differences by Child's Age." *Demography* 25(2): 205-220.

Lero, Donna S. 1981. *Factors Influencing Parents' Preferences For And Use of Alternative Child Care Arrangements For Pre-School Children* Ottawa: Health and Welfare Canada.

_____; A. Pence; L. Brockman; and M. Charlesworth. 1985. "Parents' Needs, Preferences, and Concerns About Child Care: Case Studies of 336 Canadian Families." in *Series 5: Child Care Needs of Parents and Families* Background Papers for Report of the Task Force on Child Care. Ottawa: Status of Women Canada.

Lightman, Ernie S., and Laura C. Johnson. 1977. "Child Care Patterns in Metropolitan Toronto." Project Child Care Working Paper #2. Toronto: Social Planning Council of Metropolitan Toronto.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics* Cambridge, U.K.: Cambridge University Press.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behaviour." in *Frontiers in Econometrics* edited by P. Zarembka. New York: Academic Press.

Metropolitan Toronto Day Care Planning Task Force. 1986. *Blueprint for Child Care Services: Realistic Responses to the Need* Toronto: Municipality of Metropolitan Toronto.

Payette, Micheline, and Francois Vaillancourt. 1984. "L'Utilisation des Services de Garde Au Québec." Collation Diffusion, Volume 1. Montreal: Office des Services de Garde À L'Enfance, Gouvernement du Québec.

Presser, Harriet. 1986. "Shift Work Among American Women and Child Care." *Journal of Marriage and the Family* 48: 551-63.

Ram, Bali. (1987). "Family Structure and Extra-Parental Child-Care Need in Canada: Some Projections." Demography Division, Statistics Canada, Ottawa. (Mimeographed.)

Robins, Philip K., and Robert G. Spiegelman. 1978. "An Econometric Model of the Demand for Child Care." *Economic Inquiry* 16(1): 83-94.

Shortlidge, Richard L., Jr. 1975. "Child Care and the Need for Day Care Centers and Homes in 1971." In *Dual Careers: A Longitudinal Analysis of the Labor Market Experience of Women* Vol.3. Columbus: Ohio State University, Center for Human Resource Research. 37-72.

_____. 1977. "Patterns of Child Care Utilization Among Women With Preschool Children." In *Dual Careers: A Longitudinal Analysis of the Labor Market Experience of Women* Vol. 4. Edited by Herbert S. Parnes, et al. Columbus: Ohio State University, Center for Human Resource Research. 119-39.

Small, K., and C. Hsiao. 1982. "Multinomial Logit Specification Tests." Working Paper. Department of Economics, Princeton University, Princeton, N.J.

Special Committee on Child Care. 1987. *Sharing The Responsibility* Ottawa: Queen's Printer.

Statistics Canada. 1975. "Working Mothers and Their Child Care Arrangements in Canada, 1973." in *The Labour Force,* Statistics Canada Bulletin 71-001. Ottawa: Minister of Supply and Services.

Statistics Canada. 1981. *Labour Force Survey Research Paper* Number 31: Initial Results from the 1981 Survey of Child Care Arrangements Ottawa: Minister of Supply and Services Canada.

Stevens, Harvey. 1984. *Child Care Needs and Realities in Winnipeg - 1984* Winnipeg: Social Planning Council of Winnipeg.

Westinghouse Learning Corporation and Westat Research, Incorporated. 1971. *Day Care Survey 1970: Summary Report and Basic Analysis* Bladensburg, Maryland: Westinghouse Learning Corporation.

Yaeger, K. E. 1978. "Modal Choice in the Demand for Child Care by Working Women: A Multinomial Logit Analysis with Quality Adjustment." Ph.D. dissertation, Princeton University.

GLOSSARY

CHILD CARE Any form of care provided for a child while mother is at work or school.

DAY CARE A licensed group day care centre.

SITTER Care for a child by a (baby)sitter who is not related to the child. Care is usually provided outside the child's home, but includes care provided in the child's home. The general term "sitter" includes both licensed and unlicensed family home day care unless specified.

CARE BY A RELATIVE Care provided for a child by any relative while the mother is at work. This relative might be the child's father, a sibling, a grandparent, or other relative whether in the child's home or in the relative's home and whether this care is paid or provided free of charge.

REGRESSION A statistical technique for relating the magnitude of one variable to the magnitudes of other variables. Usually, we believe that the one variable is "caused" by the other variables, so that the results of the regression tell us, on average, what effect each of the other variables has on the one variable. The most common form of regression is called Ordinary Least Squares (or O.L.S.) but in this Handbook, we focus on logit regression.

DEMAND The numerical relationship between the number of users of a particular product or service (e.g., day care) and the factors which affect whether that product or service is chosen. If we are talking about demand by an individual family, rather than by a

group of families, we have to change the above definition to read "the numerical relationship between the probability that an individual family will choose day care and the factors which affect whether day care is chosen."

DEPENDENT VARIABLE In a regression, this is the choice we are trying to explain.

EXPLANATORY VARIABLES In a regression, these are the variables (such as income level, the price of day care, the number of children in the family) that, we believe, explain what is happening to the dependent variable. The explanatory variables are the causes and the dependent variable is the effect. Usually we will use the letter "X" to refer to explanatory variables; for instance, X_1 or X_2.

ESTIMATED COEFFICIENT The demand for a particular type of child care depends in variables. For any particular family, each of the explanatory variables (the "X's") is multiplied by a number (a coefficient) which represents the effect of that explanatory variable on the choice of child care. We use the letter "b" to refer to the coefficients. So, for instance, we would have b_1 and b_2. These coefficients are unknown when we begin the exercise of trying to study the demand for child care. We use statistical techniques to estimate what the value of each coefficient is; the result is the set of "estimated coefficients".