RP-91-22

MIDAS: A Travel Demand Forecasting Tool Based on a Dynamic Model System of Household Demographics and Mobility

Final Report

Ryuichi Kitamura
and
Konstadinos G. Goulias

Institute of Transportation Studies
and
Department of Civil Engineering
University of California, Davis
Davis, CA, 95616
USA

June 1991

Preface

The genesis of this effort to develop a dynamic simulation model system for household demographics and mobility dates back to the summer of 1987 when crude simulation runs were made on SPSSx at Bureau Goudappel Coffeng (BGC), almost as an afterthought of a project for DVK. The intent of the exercise was to illustrate how dynamic models can be used in forecasting and scenario analysis. The model system was comprised of an ordered-response probit model of household car ownership, a linear trip generation model, and a binary-logit modal split model. The model components were all "dynamic" in the sense that they contained lagged dependent variables, representing the state of the behavioral process at the last observational period. The representation of household demographics and socio-economics, on the other hand, was extremely simple.

At that time, research had commenced at BGC with a probe into dynamic characteristics of travel behavior and addressing the validity of cross-sectional models, especially their applicability to forecasting. The new stream of research was in part motivated by the observation that the economic recession of 1978 - 1983 resulted in much smaller reductions in mobility than cross-sectional models had indicated. The need to clearly distinguish long-term and short-term responses (or elasticities) was pointed out (Goodwin, 1987); the roles of behavioral inertia and response lags in short-term behavioral responses recognized; and the treatments of long-term and short-term responses in existing forecasting models studied.

This project was conceived when the still ongoing evolution of dynamic travel behavior analysis was in its infancy. Critical in this evolution has been the availability of the Dutch National Mobility Panel data set (see Baanders and Slootman, 1989; Golob, et al, 1986; van der Loop, 1988; van Wissen & Meurs, 1989). As more waves of data became available, more extensive analyses and more elaborate model developments were made possible. It was in this context that the proposal was made to Projectbureau that a simulation model system be developed integrating a micro-analytic demographic accounting system and a dynamic household car ownership and mobility model system. It was intended that the Dutch Panel data set be the principal data source for the model development effort.

This report summarizes the cumulative results of the MIDAS development effort. It is based on the 1989 report, however many sections have been extensively modified and new sections have been added. Section 3 was modified to incorporate new results of household type transition analyses. Section 5.5 was introduced to describe the new income models used in MIDAS. In section 6 the trip generation models have been entirely redeveloped, and a new modal split model and travel distance models have been introduced. Section 7 is a new section which presents the results of a validation analysis of the mobility model components. Sections 8 and 9, also new sections, contain the forecasting results.

The project is an innovative approach based on entirely new concepts and methods. MIDAS itself has been revised numerous times with a number of parameters introduced and updated to better replicate household evolution and represent more extended ranges of scenarios.

The primary objective of the project is to demonstrate that travel demand forecasting using micro-simulation with dynamic models and parameters estimated by the Dutch Panel data, is practical and meaningful. This objective has been fulfilled. Moreover, MIDAS is now able to entertain endless "what if" questions. With its forthcoming PC version, MIDAS could be utilized extensively as a versatile decision support tool for Dutch transportation planning and policy analysis. However, the software should be tested rigorously prior to its distribution to other users.

1. Introduction

The Dutch National Mobility Panel (Golob, et al., 1986; van der Loop, 1988; van Wissen and Meurs, 1989) has contributed tremendously to the development of dynamic travel behavior research by offering a unique and rich data set that has made possible the examination of many aspects of travel behavior that could not have been studied with cross-sectional data. This panel data set has been applied to the more traditional subject areas of mode use, car ownership, trip generation, and trip chaining, and to more novel subject areas such as habitual behavior, response lags, and adaptation behavior[1].

A new approach to travel demand forecasting is proposed in this study in which the following two concepts are integrated to form a simulation model of household travel. The first is the dynamic model of travel behavior. Panel data enable us observe changes, thus making possible the development of models that relate behavioral changes to changes in contributing factors. Using such dynamic models, future behavior can be predicted by extending longitudinally observed changes. It is in this respect that the use of dynamic models in forecasting is critically different from the use of cross-sectional models which, unfortunately, involves the untested assumption that future behavior can be depicted by longitudinally extrapolating cross-sectional variations (see Kitamura, 1990).

The development of dynamic models in this study is a continuation of earlier work by one of the authors (Kitamura, 1987a, 1988a). That study, which used the Dutch National Mobility Panel data and was funded by Dienst Verkeerskunde (DVK), formulated a dynamic model system of car ownership and mobility, and discussed its application to forecasting. This present study adopts the model system, refines it, and uses it as a component of the forecasting model system developed in the study.

The second concept is the notion of demographic accounting system[2]. The research results accumulated in the past decade or so have offered ample evidence that household structures and lifecycle stages have profound impact on household travel (Kitamura, 1988c). Despite this finding, transportation planners have considered demographic and socioeconomic forecasting to lie outside their regime. At the same time, available demographic forecasts unfortunately have not provided data that are adequate for models typically used in transportation planning, e.g., individual- or household-based disaggregate choice models. In particular, the majority of travel demand models are developed through multivariate statistical analysis that fully takes advantage of the information available in the data. Demographic and socioeconomic forecasts, on the other hand, do not in general offer a multivariate distribution. They therefore do not necessarily capture the correlation that exists among variables typically used in travel demand models. The result is a dubious basis for travel demand forecasting.

In this study, a demographic accounting system is developed using data from Waves 1, 3, 5, 7 and 9 of the Dutch National Mobility Panel data set, covering the four-year period of April 1984 through April 1988. This accounting system is integrated with an econometric model system of household car ownership and mobility, to form a model system that replicates the evolution of households through lifecycle stages and, at the same time, determines their transportation system use.

An effort by the Transport Studies Unit group at Oxford University to develop a demographic accounting system for transportation planning led to a simulation model called MIDCAT (Goodwin, Dix and Layzell, 1987). This earlier effort forms a basis for the present study. MIDCAT itself, however, is not adopted in this study due to the availability of the Dutch Panel data which offer a more extended range of modeling possibilities for the study. It is also due to the system of car ownership and mobility models which calls for more detailed accounting of household and person characteristics.

The resulting model system, MIDAS, is a simulation model that can be used to evaluate a variety of scenarios for transport policy development. Employment, income, driver's license holding, household lifecycle stage, and education levels are among the variables that are internally generated in the simulation, then used to predict household car ownership and mobility. Thus many explanatory variables that are exogenous for other forecasting models are endogenous in MIDAS.

MIDAS provides parameters which can be adjusted by the user to manipulate these internally generated demographic and socio-economic variables to represent future scenarios of interest. Changes in highway and transit levels of service can be represented by modifying accessibility measures that are used in the car ownership model. In these scenario analyses, MIDAS does not forecast future behavior through longitudinal extrapolation of cross-sectional variations; its forecast is firmly based on longitudinally observed patterns of changes.

MIDAS embodies the causal structure underlying socio-demographic evolution of a household. This leads to several advantages. It simulates household changes and, therefore, is capable of accounting for changes at the behavioral decision unit in a manner compatible with disaggregate models of travel behavior. Consisting of many interlinked components, each dealing with a pertinent socio-demographic element (e.g., employment and driver's license holding), MIDAS can be readily modified to adapt to a specific scenario. Also because of its structure, MIDAS performs forecasting while realistically replicating and extrapolating the internal correlation among variables that contribute to household car ownership and mobility. In these respects, MIDAS is believed to be more useful as a planning tool than aggregate time series models which treat behavioral processes as a black box.

Many components of MIDAS have dynamic model structures, and forecasts are made along the time dimension with an increment of one year. MIDAS can therefore be applied to examine how a particular growth path will influence car ownership and mobility levels. For example, it can be used to evaluate the impacts of a gradual increase in gasoline prices, or to evaluate their sudden hike. MIDAS can also be used to measure short-term and long-term changes and evaluate the difference between short-term elasticities and long-term elasticities of mobility levels.

These advantages, however, must be critically appraised in future effort as MIDAS is still in its developmental stage. Also it must be kept in mind that there are certain disadvantages due to the fact that MIDAS is a dynamic model system. Most importantly, the model system is complex and requires more extensive and detailed data. The appropriateness of extrapolating observations obtained in four years (April 1984 through April 1988) for long range forecasting needs to be thoroughly examined. It has not been examined how the fact that the data were collected during a period of economic expansion has affected the model estimation. Incorporating supply-demand interaction, the impact of regional economy, threshold effects, and non-linear responses, is beyond the scope of the current effort. Despite these limitations, the study results indicate that the approach taken has been more than worthwhile, and the outcome is a useful and versatile policy tool.

The specific tasks of the project are:

- Improve the car ownership and mobility model system developed earlier for DVK (Kitamura, 1987a, 1988a) by integrating it with a demographic simulator, including trip distances as endogenous variables of the model system, and, if possible, travel costs as explanatory variables,

- Perform simulation experiments to develop long term forecasts for the year 2010 based on the data used in the preparation of SVV II,

- Compare the forecasts with other comparable forecasts (Gunn, van der Hoorn and Daly, 1987; van den Broecke, 1988),

- Derive short-term and long-term elasticities to changes in income, employment, and driver's license holding, and compare the results with other forecasts (van den Broecke, 1988), and

- Evaluate the results and develop recommendations for future effort.

The emphasis in this report is to introduce the concept of dynamic models of household demographics as well as car ownership and mobility, to offer a detailed account of the structure of the components of the simulation system together with discussions on the underlying assumptions and model development processes, and to present the results of the simulation and compare them with other forecasts.

This report is organized as follows. In the next section, the overall modeling approach and the structure of MIDAS are presented. In particular, the use of household types as a major modeling element is discussed, and the scheme used to define household types is described. Following this, the transition among household types is discussed in Section 3 together with a set of logit models used in MIDAS to simulate the evolution of household types. Section 4 presents the results of a causal analysis of employment, driver's license holding, and personal income (Although the resulting causal models were not considered robust enough to be used as forecasting models, they nonetheless offered a basis for the many modeling decisions that had to be made during the development of MIDAS). The demographic component of MIDAS and its program elements are described in detail in Section 5. Section 6 describes the mobility component of MIDAS, which consists of a car ownership model, motorized-trip generation models, a modal split model, and trip distance models by mode. Section 7 reports the results of a validation exercise where predictive accuracy of the models in the mobility component is examined using Wave-10 data that were not used to estimate them. Section 8 contains descriptions of MIDAS input parameters and discusses the weighting procedure applied to the sample households used in MIDAS simulation runs. In Section 9, MIDAS forecasting results are presented and compared with other existing forecasts, car ownership and mobility growth under different income growth scenarios are evaluated, and short-term and long-term elasticities are examined. Section 10 contains a summary and recommendations.

---

[1] A sample of travel behavior studies using the Dutch National Mobility Panel data set can be found in Goulias and Kitamura (1989, 1991), Golob (1986, 1989, 1990), Golob and Meurs (1986, 1987),Golob and van Wissen (1988), Golob, van Wissen and Meurs (1986), Goodwin (1987), Kitamura (1987a, 1987b, 1987c, 1988a, 1988b, 1989a, 1989b), Kitamura and Bovy (1987), Kitamura and Bunch (1990), Kitamura and van der Hoorn (1987), Meurs, van de Mede, Visser and van Wissen (1987), Meurs, Gloerich, van de Mede, Visser and Klok (1987), Recker, Golob, McNally and Leonard (1987), and van der Hoorn and Kitamura (1987).

[2] See Bachman, O'Malley and Johnston (1978), Davidson (1972), Juster and Land (1981), Land and Rogers (1982), Land and Spileman (1975), Orcutt, Caldwell and Wertheimer (1976), Spengler and Duncan (1956), Rossi and Gilmartin (1980), Schoen (1988), Willis (1971).

2. Structure of MIDAS

This section offers an overview of the modeling approach taken to develop MIDAS, and presents a broad picture of how it replicates household evolution over time. Following this, the household type classification scheme used in the study is described. The transition among household types is a governing relationship in MIDAS; changes in many household and personal attributes are conditioned on household type transition. The discussions in the second half of this section and Section 3 are directed to this subject.

2.1. Framework

The mechanisms underlying changes in household attributes are difficult to identify. This is in part due to the fact that many attributes are so intricately interwound that identifying the primordial factors that trigger changes is almost impossible. For example, consider the labor force participation by married women and the presence of pre-school children in the household. Does a woman choose to stay home because of the presence of children, or does she (and her household) choose to have children, and therefore leave the labor force?

Furthermore, it is likely that many causal structures exist that apply to a given change. For example, a recent analysis of trip chaining behavior (Kitamura, Nishii and Goulias, 1990) report that several alternative causal structures explain observed behavior equally well, suggesting that there are many causal relationships that underlie observed behavior. In some instances it may be the presence of children that prevents a woman from participating in the labor force, and in other occasions it may be a woman's conscientious choice to leave the labor force in order to raise children.

Developing a simulation model of household evolution is not a trivial task because it requires a model of causal relationship underlying changes in household attributes. As the above example indicates, even the well studied event of birth becomes a difficult subject to model. Yet, erroneous forecasts may result if the inter-relationships among variables are not properly accounted for. Obviously a model would require a set of simplifying assumptions to be operational. Modeling household evolution is further complicated because of interactive changes at two levels: the household and the individual. The attributes of a household will change as a new member enters it or its existing member exits from it. Its attributes may also change when its members' attributes change. Changes at these two levels must be consistently reproduced for a simulation model to properly function. In MIDAS, the evolution of household characteristics and mobility over time is replicated in a recursive manner by simulating transitions in the household type, changes in the attributes of household members, changes in car ownership, and levels of mobility. The overall dynamic structure of MIDAS is presented in the block diagram of Figure 2.1.

The transition between household types is viewed as the most fundamental element of household evolution in MIDAS. Ample evidence exists in the literature demonstrating the importance of household lifecycle in travel behavior analysis and demand modeling. The household types used in MIDAS are closely related to the concept of lifecycle stage. MIDAS treats the progression of a household through lifecycle stages as the basic building block of its dynamic structure. Changes in person attributes and mobility are modeled around a model of household type transition.

Given a transition in the household type, new household members are generated, or existing household members are eliminated, and member characteristics are altered in MIDAS. The transition in household types thus serves in MIDAS as a control that constrains the number and characteristics of household members.

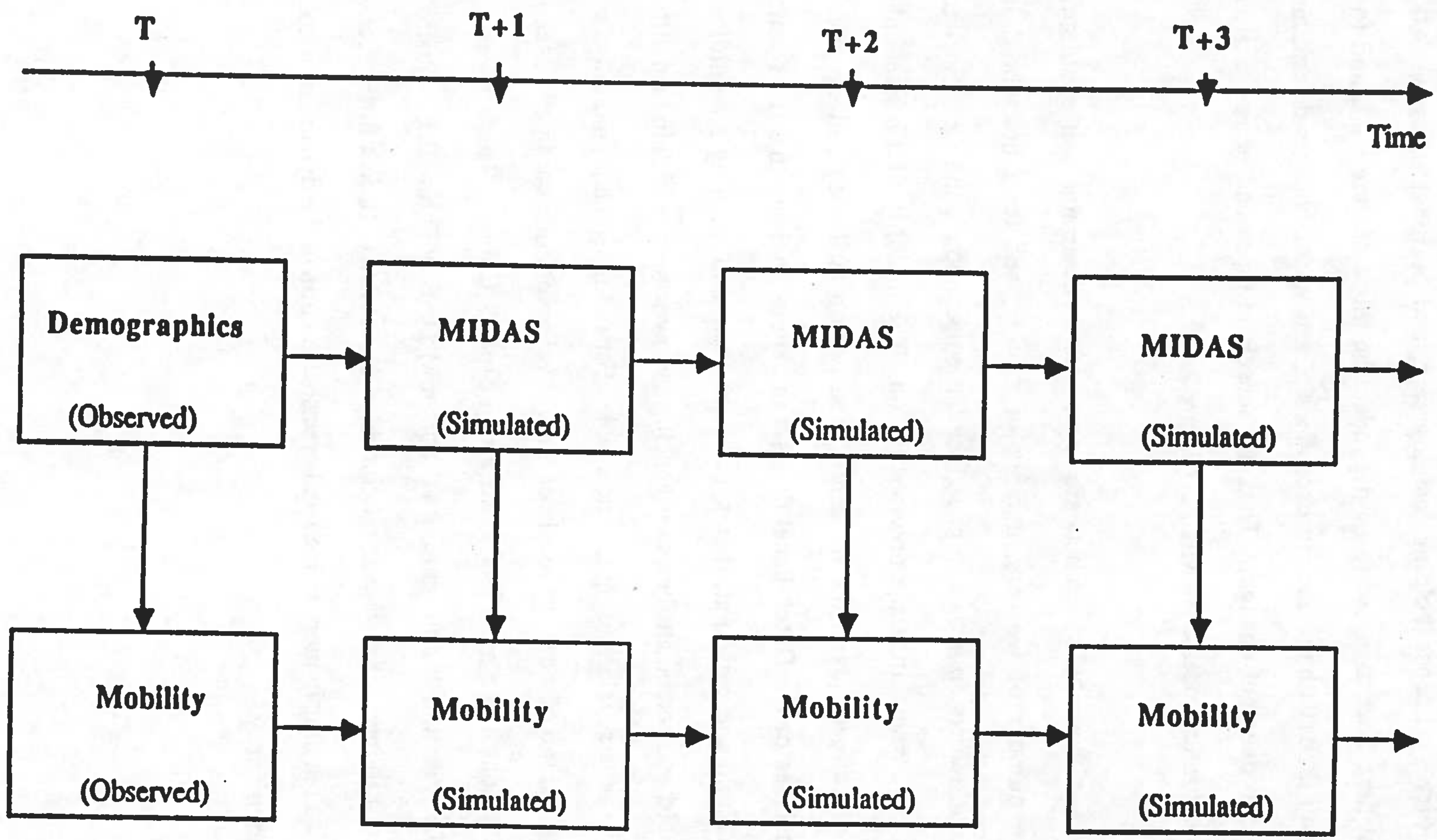


Figure 2.1

Midas and Mobility Component

All pertinent person attributes are endogenously determined in MIDAS, including: education, driver's license holding, employment, and personal income. All person attributes at time t are assumed to be dependent on those of time t - 1, and to change randomly over time (with the obvious exceptions of age, which changes deterministically, and sex, which does not change). Thus the household types and person attributes are viewed as stochastic processes in MIDAS (Figure 2.2).

The attributes of household members are aggregated to determine household attributes, such as the number of workers, number of drivers, and total household income. Household car ownership is determined given these household and person attributes and the level of car ownership in the previous period. The mobility of a household is then randomly determined, given the household and person attributes and car ownership. Previous analyses of the Dutch Panel data set in general indicate that the relationships among household and person attributes, car ownership, and mobility are unidirectional, with household and person attributes influencing car ownership, which in turn influencing mobility. The reverse relationship is conceivable. For example, a high mobility level may cause the acquisition of a car at some later time. Availability of a car in a household may encourage its non-driver members to acquire a driver's license. These reverse causal relations are not assumed in the current version of MIDAS as Figure 2.1 indicates. This, however, is not to imply that reverse relationships do not exist. In fact further research is needed in this area to gain more refined and precise depiction of the dynamics in household evolution and mobility.

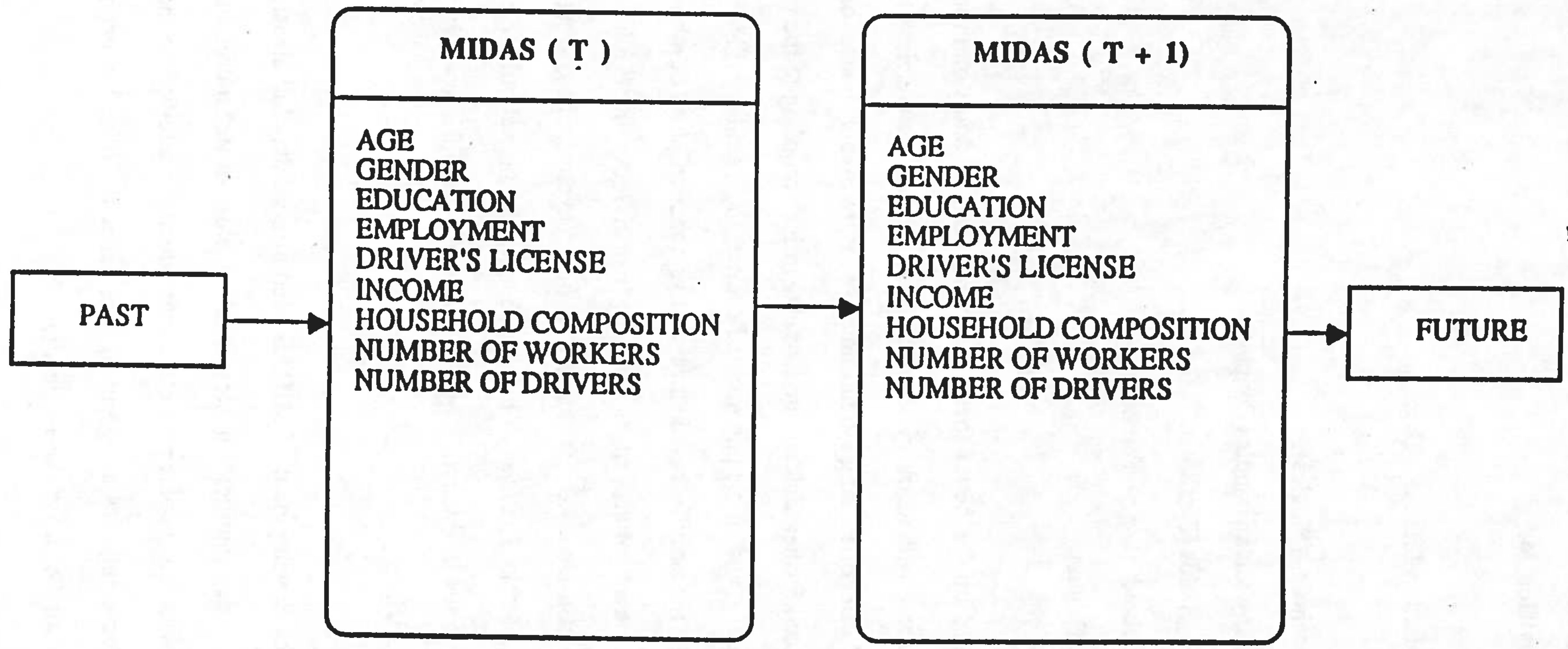


Figure 2.2

A WINDOW ON MIDAS EVOLUTION

2.2. Household Classification

Five household types are used in MIDAS. These are

Single-person households (Singles),
Households of a man-woman couple (Couples),
Nuclear family households (Families),
Single-parent households (Single Parents), and
Other households (Others).

This classification is based on the belief that the composition of adult members of a household is closely associated with its travel behavior. Thus households are classified into the following three broad categories: single-adult households, two-adult households, and others. The first two are each further divided into two categories according to the presence or absence of children: singles and single parents; and couples and families. This is based on the major conclusion of the activity-based travel analysis that children of a household importantly influence the travel patterns of its adult members (Goodwin, 1983; Jones, et al.,1983). The classification also reflects the notion of lifecycle (Jones, et al, 1983; Kostyniuk and Kitamura, 1982). The ages of the members of the household are not used to define household categories in this study because they are variables used in various model components of MIDAS.

The analysis preceding the development of MIDAS used a more detailed classification scheme that included "extended families" (nuclear families plus "other" individuals) and "extended couples" (man-woman couples plus other individuals) in addition to the above five categories. These two categories are grouped together with "other" households in MIDAS due to the limited sample size of these categories.

A household is classified into one of the five household types on the basis of the number of adult men, number of adult women, each member's position in household as recorded in the Dutch Panel data file, and number of children. The classification process is summarized in Figure 2.3.

The frequency distribution of household types thus defined is presented in Figure 2.4 for Waves 1, 3, 5, and 7 of the Dutch Panel survey. The distribution varies noticeably across waves due to attrition and sample refreshment. The majority (85 to 90%) of the Dutch Panel households fall in the first three household types; 15 to 20% of the households are singles, approximately 25% are couples, and 45 to 50% are families.

The distribution of household types is presented in Table 2.1 by municipality class. A clear tendency emerges from this table. Single persons tend to be in larger metropolitan areas (BOV's) while families tend to be in commuter communities and smaller communities in rural settings. The fraction of single-person households exceeds 30% in larger urban areas (BOV's), but is below 10% in smaller communities that are not served by train. Nuclear family households, on the other hand, account for less than 30% in the largest urban areas (BOV-Large), while their fractions exceed 50% in smaller communities. The correlation between the household type and the size of urban area is evident and reflects preferences in residential location choice. (This correlation is not reflected in the current version of MIDAS which does not include a residential location component. Consideration of it is obviously essential when the scope of MIDAS is extended to include residential choice.)

1. Household Type is Single Person when:

Household Size=1

2. Household Type is Couple when:

Number of Men =1, and
Number of Women = 1, and
Number of Heads = 1, and
Number of Wives = 1, and
Number of Children = 0, and
Number of Other Members = 0

3. Household Type is Family when:

Number of Men = 1, and
Number of Women = 1, and
Number of Heads = 1, and
Number of Wives =1, and
Number of Children > 0, and
Number of Other Members = 0

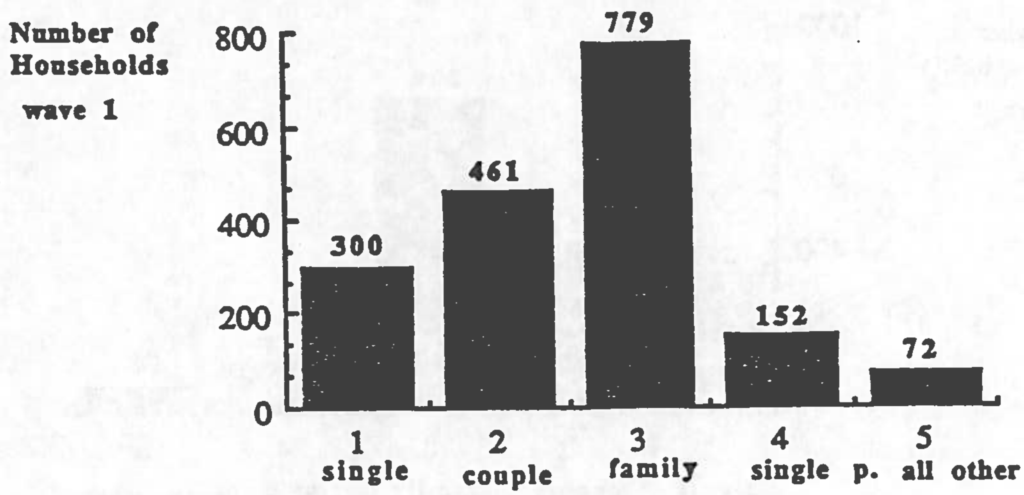
4. Household Type is Single Parent when:

Number of Men + Number of Women = 1, and
Number of Children > 0, and
Number of Other Members = 0

5. Household Type is Other when:

The Household cannot be classified in any of the above.

Figure 2.3
Household Classification Scheme



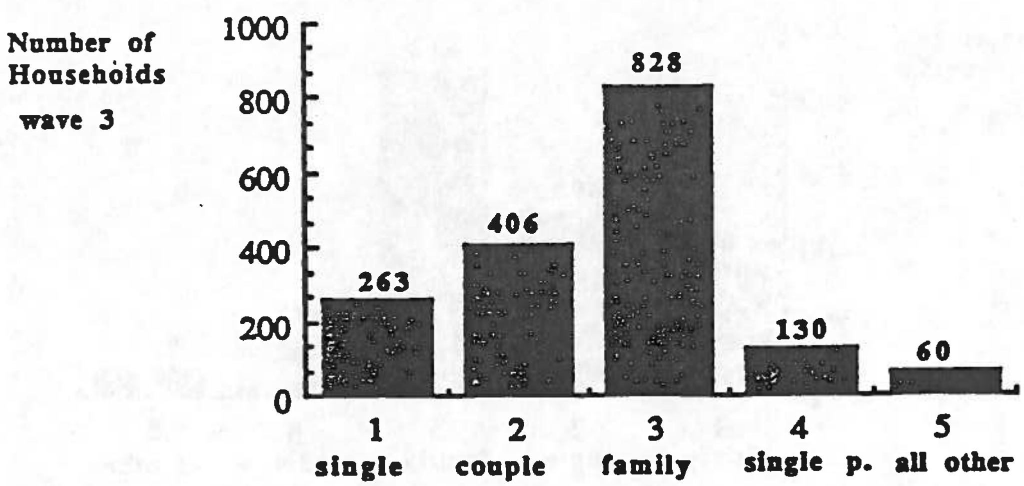
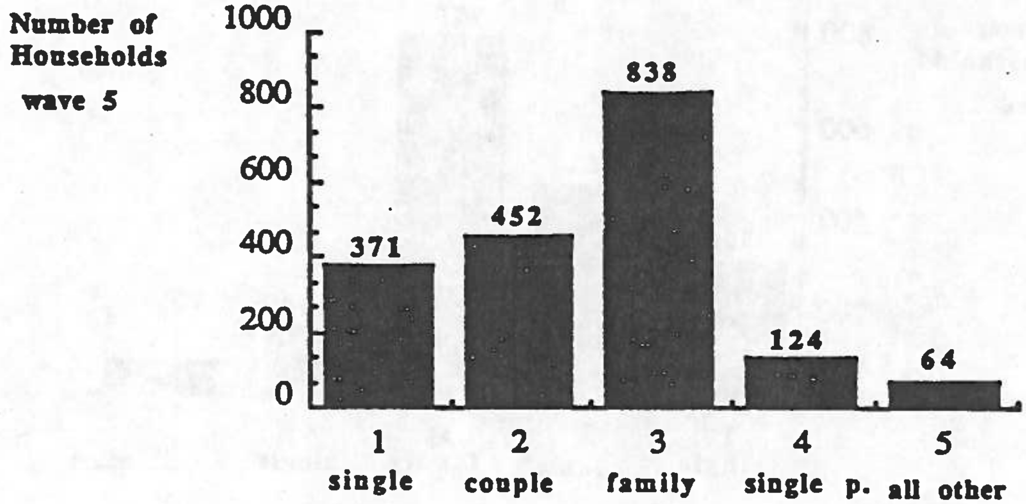


Figure 2.4

Household Type Frequencies in Wave 1 and in Wave 3



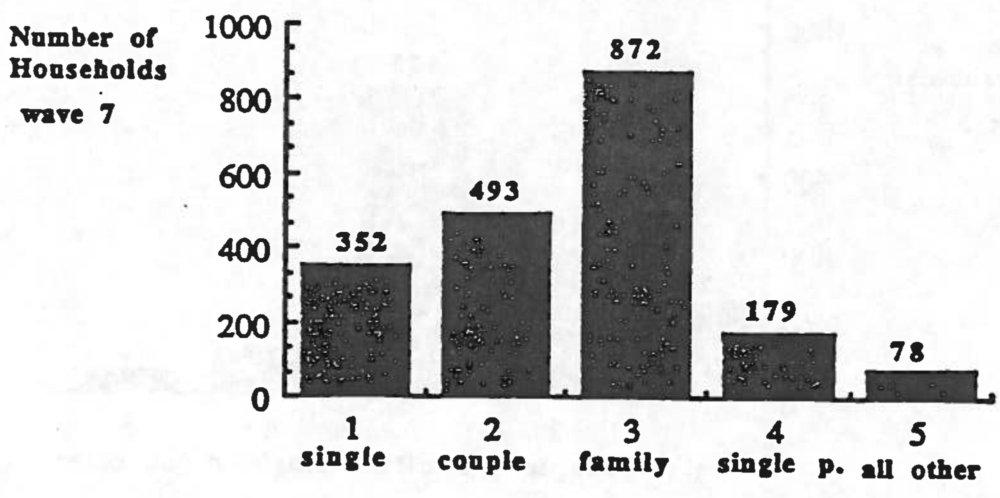


Figure 2.4 (continued)

Household Type Frequencies in Wave 5 and in Wave 7

Table 2.1.a
Distribution of Household Types by Urban Area Class in Wave 1

	Household Type					
Area Class	Single	Couple	Family	Single P.	Other	Total
BOV-Large	56	50	52	16	8	182
	(30.8)	(27.5)	(28.6)	(8.8)	(4.4)	(100)
BOV-Small	92	71	86	19	12	280
	(32.9)	(25.4)	(30.7)	(6.8)	(4.3)	(100)
BOS	58	97	147	39	17	358
	(16.2)	(27.1)	(41.1)	(10.9)	(4.7)	(100)
Community with Train	38	75	112	28	7	260
	(14.6)	(28.8)	(43.1)	(10.8)	(2.7)	(100)
Community without Train	7	34	69	7	4	121
	(5.8)	(28.1)	(57.0)	(5.8)	(3.3)	(100)
Others	49	134	313	43	24	563
	(8.7)	(23.8)	(55.6)	(7.6)	(4.3)	(100)
Total	300	461	779	152	72	1764
	(17.0)	(26.1)	(44.2)	(8.6)	(4.1)	(100)

( ): Row Percentage

Table 2.1.b
Distribution of Household Types by Urban Area Class in Wave 3

Area Class	Household Type: Single	Couple	Family	Single P.	Other	Total
BOV-Large	66	75	75	23	12	251
	(26.3)	(29.9)	(29.9)	(9.2)	(4.8)	(100)
BOV-Small	67	57	78	15	14	231
	(29.0)	(24.7)	(33.8)	(6.5)	(6.1)	(100)
BOS	39	61	136	28	16	280
	(13.9)	(21.8)	(48.6)	(10.0)	(5.7)	(100)
Community with Train	26	55	108	13	3	205
	(12.7)	(26.8)	(52.7)	(6.3)	(1.5)	(100)
Community without Train	17	37	92	12	5	163
	(10.4)	(22.7)	(56.4)	(7.4)	(3.1)	(100)
Others	48	121	339	39	10	557
	(8.6)	(21.7)	(60.9)	(7.0)	(1.8)	(100)
Total	263	406	828	130	60	1687
	(15.6)	(24.1)	(49.1)	(7.7)	(3.6)	(100)

( ): Row Percentage

Table 2.1.c
Distribution of Household Types by Urban Area Class in Wave 5

Area Class	Household Type: Single	Couple	Family	Single P.	Other	Total
BOV-Large	64	59	51	19	7	200
	(32.0)	(29.5)	(25.5)	(9.5)	(3.5)	(100)
BOV-Small	119	68	84	13	14	298
	(39.9)	(22.8)	(28.2)	(4.4)	(4.7)	(100)
BOS	78	97	168	38	15	396
	(19.7)	(24.5)	(42.4)	(9.6)	(3.8)	(100)
Community with Train	41	76	129	17	7	270
	(15.2)	(28.1)	(47.8)	(6.3)	(2.6)	(100)
Community without Train	15	27	80	7	3	132
	(11.4)	(20.5)	(60.6)	(5.3)	(2.3)	(100)
Others	54	125	326	30	18	553
	(9.8)	(22.6)	(59.0)	(5.4)	(3.3)	(100)
Total	371	452	838	124	64	1849
	(20.1)	(24.4)	(45.3)	(6.7)	(3.5)	(100)

( ): Row Percentage

Table 2.1.d
Distribution of Household Types by Urban Area Class in Wave 7

Area Class	Household Type: Single	Couple	Family	Single P.	Other	Total
BOV-Large	63	68	72	18	11	232
	(27.2)	(29.3)	(31.0)	(7.8)	(4.7)	(100)
BOV-Small	109	63	89	13	19	293
	(37.2)	(21.5)	(30.4)	(4.4)	(6.5)	(100)
BOS	76	108	153	32	17	386
	(19.7)	(28.0)	(39.6)	(8.3)	(4.4)	(100)
Community with Train	30	79	117	20	9	255
	(11.8)	(31.0)	(45.9)	(7.8)	(3.5)	(100)
Community without Train	15	26	82	8	4	135
	(11.1)	(19.3)	(60.7)	(5.9)	(3.0)	(100)
Others	59	149	359	38	18	623
	(9.8)	(23.9)	(57.6)	(6.1)	(2.9)	(100)
Total	352	493	872	129	78	1924
	(18.3)	(25.6)	(45.3)	(6.7)	(4.1)	(100)

( ): Row Percentage

2.3 Summary

In MIDAS, household evolution over time is modelled at two levels: the household and the individual. The building block of the household evolution is the household type transition. Around this transition, household members are made to change education, driver's license holding, employment, and personal income. The classification of household types consists of single person, couple, family, single parent, and all other households. Evolution in MIDAS is achieved by first determining the household type transition, then updating each household member's characteristics (attributes), and finally simulating the household mobility measures. All transitions are determined probabilistically. This procedure is repeated continuously until the final year of simulation. An investigation on the possible limitations of MIDAS due to the lack of a component of relocation is outlined. This will be further elaborated upon in Section 10.

3. Household Type Transition

Household type transition probabilities are an essential element of the MIDAS sociodemographic component. An analysis of the information in the 5 odd-numbered panel waves (Waves 1, 3, 5, 7, and 9) has led to the adoption of tabulation schemes and weighting procedures that are different from those in the earlier effort (Kitamura and Goulias, 1989).

3.1 Selectivity in Panel Participation

Theoretical considerations and empirical evidence in the literature (e.g., Goodwin, 1987; Kitamura and Bovy, 1987) suggest that panel survey participants tend to present low frequencies of household type change. This conjecture stems from the consideration that participation decision by panel respondents may be influenced by changes in the household. For example, the members of a household may be less willing to continue to participate in a panel survey after a divorce, death in the family, or other events that lead to a change in the household type. Also it is likely that changes in household type are sometimes concurrent with changes in residential location, reducing the likelihood of continued participation in the survey. If such tendencies are in fact present but are not taken into account in the analysis, then biased estimates of household type transition probabilities will result. In this study a probabilistic model of attrition is developed and applied in a weighting procedure that is designed to eliminate possible attrition bias.

In light of the recognition that the sample of households in the Dutch Panel data may not represent the transition of household types in the population, effort was made to include external demographic information to verify results obtained from the Panel data. Data from a number of sources have been collected and reviewed for possible inclusion as external information in the analysis of household type transition. Unfortunately, available statistics are not comparable with the analytical framework of this project. Both the PRIMOS and LIPRO transition probabilities are person-based. Usefulness of the information from these sources in the present study is limited because it can be used as an external check only for single-person households. Another incompatibility is the household type classification; MIDAS has adopted a classification scheme that is radically different from the one in PRIMOS or LIPRO. Furthermore, our inspection revealed certain inconsistencies in the person-based transition probabilities furnished to us. Due to these difficulties, it was decided not to devise an elaborate procedure to match the person-based transition probabilities and the household-based transition probabilities in MIDAS (External information is used to develop a set of weights that are applied to Panel households for forecasting, see Section 8.)

Another possibility is the use of weights developed by Bureau Goudappel Coffeng. These weights were developed by comparing the distribution of households by income, lifecycle stage, and municipality of residence, between the Dutch Panel sample and the OVG sample. The BGC weights are not used in the analysis of this section because these three stratification variables are included in the household type transition models developed in this section; thus their effects are already accounted for in the estimation of MIDAS household type transition models.

Due to these limitations in the available information, external validation of household type transition probabilities obtained from the Dutch Panel sample is difficult to perform. Consequently eliminating possible biases in statistics obtained from the Dutch Panel sample becomes crucial. The analysis of attrition is extremely important because of the possible correlation between attrition and household type transition discussed above (non-response bias upon the initial attempt of contact, on the other hand, is not considered to play a major role here because no systematic relation is anticipated between initial response and household type transition afterwards).

3.2 Multi-Wave Attrition

The participation of respondents in the Dutch Panel has been extensively discussed in Wissen and Meurs (1989). Their analysis, however, is primarily person-based while MIDAS is constructed using the household as its base unit. This calls for a household-based analysis of panel participation, which is the subject of this section. Figure 3.1 schematically presents the participation of households across the five survey waves. Note the large number of households leaving the panel between waves (they shall be hereafter called "leavers"). Intermittent participation (as opposed to continuous participation in every wave) is very infrequent; once a household leaves the panel, the chance of it returning is negligibly small.

As reported in Wissen and Meurs (1989), the new participants in later waves are substantially different from the continuing participants (or "stayers") from the initial wave. In addition, the households that left the panel are also substantially different from the stayers (Kitamura and Bovy, 1987). If this self-selective attrition is systematically related to household type transition, then the sample will lead to biased inferences. The approach taken in this study to account for systematic attrition is discussed below.

The availability of measures of household attributes in earlier panel waves permits the construction of probabilistic models of attrition. These models can be used to develop weights to correct for attrition biases (Hensher, 1987). Kitamura and Bovy (1987) have shown that such weights can be constructed for trip generation analysis, based on a system

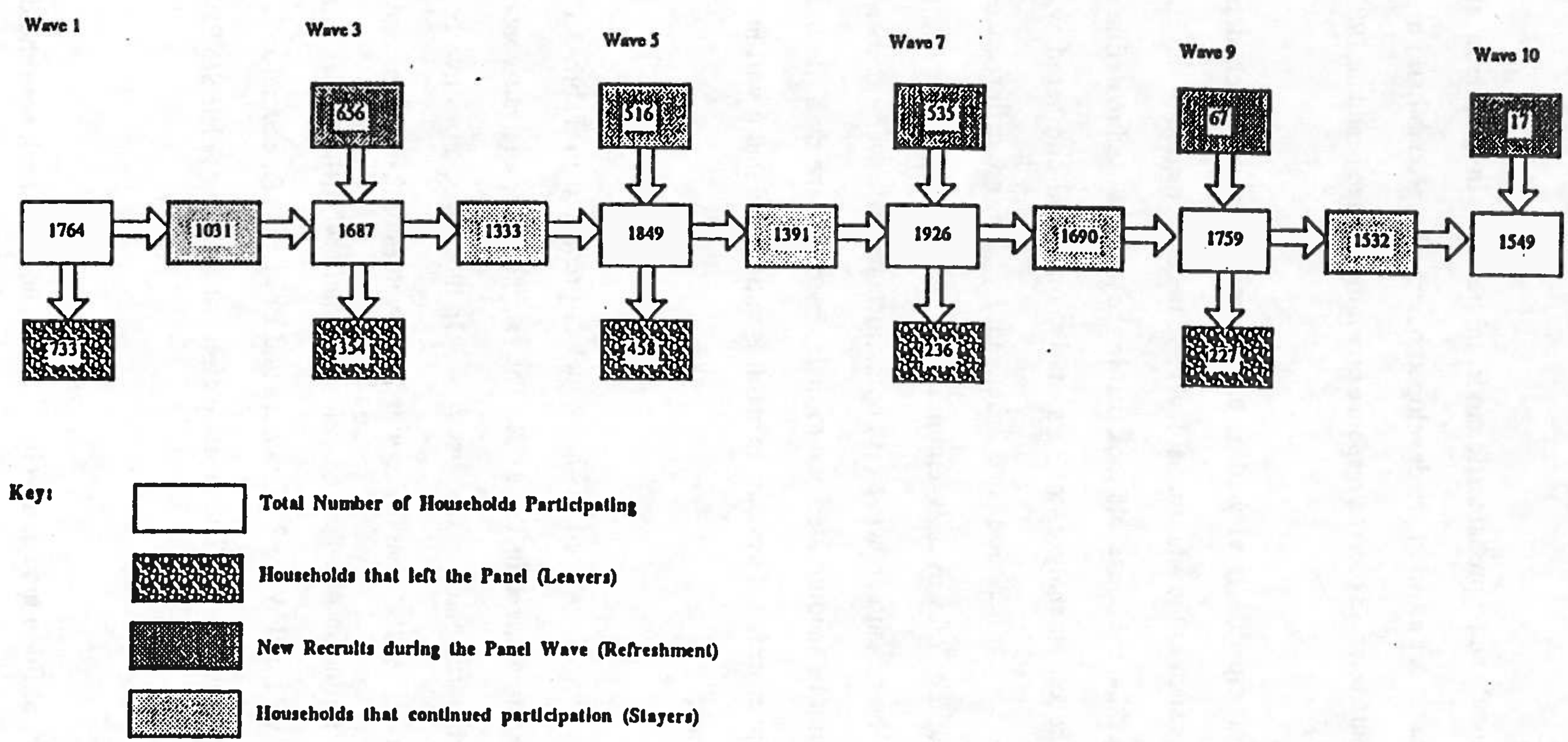


Figure 3.1

Attrition and Sample Refreshment in Dutch National Mobility Panel: Waves 1,3,5,7,9,and 10

of trip generation models and a probabilistic model of attrition. In the present study, the weights used to account for attrition are developed based on a probabilistic multi-wave attrition model, formulated as an ordered response probit model (see Maddala,1983).

Household panel participation is viewed as an ordered response which depends on background variables such as the educational level and the composition of the household. The number of waves that the household took part in the panel is used to define response categories (as in the other analyses of this study, only odd-numbered waves are considered). Intermittent participation is so infrequent that it is not differentiated in the analysis, i.e, the number of waves is enumerated the same way whether they are consecutive or not. The measure, however, is approximate because even-numbered waves (2, 4, 6, and 8) can not be incorporated due to their absence in the data file used in the project. The resulting error is not systematic and it is anticipated that it will not bias the outcome of the analysis.

The definition of the variables used in the model is presented in Table 3.1 and the estimation results are summarized in Table 3.2. The sample of the analysis consists of all the households which participated in the first wave. In the Table, a positive coefficient estimate of a variable implies that households with larger values of that variable have higher probabilities of staying in the panel longer. In agreement with the earlier analysis of attrition between Wave 1 and Wave 2 (Kitamura and Bovy, 1987), education is a major determinant of panel participation. The positive effect of mobility (TRIPS/NPRCRD) is also evident in the results.

Using this multi-wave attrition model, a weight is formulated for each household and for each wave as the reciprocal of the probability that the household will participate in that wave. For example, consider a household which participated in the wave-1 survey. If the

Table 3.1
Definition of Variables and Participation Levels in the Ordered Response Probit Model of Multi-Wave Attrition

Variable	Definition
NPRCRD	Number of respondents in the household
CHLD11⁻	Number of children less than 11 years old.
CHLD11⁺	Number of children more than 11 years old.
SINGLE	1 if a single person household; 0 otherwise.
COUPLE	1 if the household is a couple; 0 otherwise.
FAMILY	1 if the household is a family; 0 otherwise.
SGLPRN	1 if a single person household; 0 otherwise.
ELEM	1 if the highest level of education in the household is elementary school; 0 otherwise.
PROF	1 if the highest level of education in the household is professional school; 0 otherwise.
HSCH	1 if the highest level of education in the household is high school; 0 otherwise..
MIDPRO	1 if the highest level of education is; 0 otherwise.
WORKERS	Number of employed persons in the household.
LOINC	1 if the total annual personal income in the household is less than dfl 17000 ; 0 otherwise.
HDNOJOB	1 if the head of the household is unemployed; 0 otherwise.
DRIVERS	Number of drivers in the household.
ONECAR	1 if the household owns one car; 0 otherwise.
TWOCAR	1 if the household owns two or more cars; 0 otherwise.
TRIPS/NPRCRD	Number of trips made divided the number of respondents.
BOVLARGE	1 if the household resides in large metropolitan area with a transit district; 0 otherwise.

Dependent Variable Number of Waves Participated	Frequency	Cumulative Frequency
1	679	679
2	173	852
3	179	1031
4	38	1069
5	610	1679

Table 3.2
Ordered Response Probit Model of Multi-Wave Attrition

Variable	$\hat{\beta}$	t
NPRCRD	-0.066	-0.50
SINGLE	0.487	2.18
COUPLE	0.533	2.82
FAMILY	0.501	2.49
SGLPRN	0.456	1.91
ELEM	-0.628	-4.88
PROF	-0.248	-2.92
HSCH	-0.257	-2.98
MIDPRO	-0.229	-2.67
ONECAR	-0.082	-0.92
TWOCAR	-0.292	-2.15
BOVLARGE	-0.419	-4.26
WORKERS	-0.078	-1.28
DRIVERS	0.063	1.14
LOINC	-0.275	-3.03
CHLD11-	0.115	2.14
CHLD11+	-0.005	-0.04
TRIPS/NPRCRD	0.016	6.71
HDNOJOB	-0.031	-0.32
α_1	0.345	1.10
α_2	0.631	2.01
α_3	0.926	2.95
α_4	0.990	3.15
N	1679	
L(0)	-2237	
L(c)	-2170	
$L(\hat{\beta})$	-2059	
$-2(L(0)-L(\hat{\beta}))$	357	
$-2(L(c)-L(\hat{\beta}))$	222	

* Chi-square distributed with degrees of freedom 1656
** Chi-square distributed with degrees of freedom 1660

probability that this household participates in the wave-3 survey is estimated by the model as 0.8, then the wave-3 record of this household will be weighted by 1.25 (= 1/0.8). Weights thus computed are used in some of the tabulations of household type transition presented below.

3.3 Household Type Transition Frequencies Based on Alternative Pooling Schemes of Multi-Wave Observations

A set of logit models is used in MIDAS to describe the transition between household types. One logit model is formulated for each "origin" household type. (Suppose a household belongs to type i in period t - 1, and type j in period t. Then we may say that the household made a transition from the origin state, i, to the destination state, j, between periods t -1 and t.) This increases sample size requirements because a sufficiently large number of observations is needed for each household type. Accordingly records from the five odd-numbered waves are combined, or "pooled," to form a large data set while treating them as if they came from two time points, say, t - 1 and t. Note that a transition from a household type to itself implies that the household did not change its type.

The frequency of transitions among household types is summarized in Table 3.3 as a matrix. Transitions are observed over a period of one year using the five household types described in the previous section. A pooled data set of wave pairs is used to prepare the transition frequency matrix. In this pooled data set, paired observations from Waves 1 and 3, Waves 3 and 5, and Waves 5 and 7 are combined together. This data set shall be called the "repeat" data set.

It is clear from the table that the transition in household types is far from being volatile. The large frequencies of diagonal cells indicate that the type of a given household tends to

Table 3.3
Household Type Transitions of the Repeat Data Set

Household Type at t	Household Type at t+1					
	Single	Couple	Family	Single Par.	Other	Total
Single	857 (94.8)	7 (0.8)	7 (0.8)	7 (0.80)	26 (2.9)	904 (100)
Couple	19 (1.4)	1200 (91.3)	78 (5.9)	0 (0.0)	18 (1.4)	1315 (100)
Family	2 (0.1)	57 (2.1)	2545 (95.5)	46 (1.7)	16 (0.6)	2666 (100)
Single Par.	10 (2.6)	3 (0.8)	38 (9.9)	298 (77.4)	36 (9.4)	385 (100)
Other	13 (7.4)	34 (19.4)	9 (5.1)	9 (5.1)	110 (62.9)	175 (100)
Total	901 (16.5)	1301 (23.9)	2677 (49.2)	360 (6.6)	206 (3.8)	5445 (100)

( ): Row percentage

be the same when observed one year apart. The result, however, may be due to panel attrition, if households tend to drop out of the panel when their compositions change. The generality of this finding, therefore, must be carefully examined.

Salient transitions (arbitrarily defined as those with 25 or more observed transitions) are: couples to families, families to couples, families to single parents, single parents to families, and others to couples. Most salient transitions involve couples, families, and single parents. Singles, on the other hand, are relatively detached from the rest. This transition matrix obtained using the Panel sample households thus indicates that singles tend to remain as singles over time.

Frequencies of household type transitions are tabulated using another data set that is obtained using a different pooling scheme. The first repeat, data set included all observed transitions. If a household participated in all five odd-numbered waves, then it offers four observations of transition. The results are presented in Table 3. In the second data set, on the other hand, only the first transition observed for each household is included. This data set shall be named "no-repeat" data set because no household has repeated observations in it. The results are summarized in Table 3.4.

The household transition matrix from the no-repeat data set presents higher frequencies of households changing household types than that from the repeat data, supporting the conjecture that panel participation and household type transition are correlated. The difference between the two, however, is rather small. Some of the characteristics of these transition frequency matrices are summarized below.

Table 3.4
Household Type Transitions of the No-Repeat Data Set

Household Type at t-1	Household Type at t+1: Single	Couple	Family	Single Par.	Other	Total
Single	390 (93.8)	3 (0.7)	4 (1.0)	5 (1.2)	14 (3.4)	416 (100)
Couple	8 (1.4)	531 (89.7)	42 (7.1)	0 (0.0)	11 (1.9)	592 (100)
Family	1 (0.1)	27 (2.6)	983 (94.2)	26 (2.5)	6 (0.6)	1043 (100)
Single Par.	5 (3.2)	0 (0.0)	22 (14.3)	118 (76.6)	9 (5.8)	154 (100)
Other	6 (6.5)	20 (21.7)	2 (2.2)	4 (4.3)	60 (65.2)	92 (100)
Total	410 (17.8)	581 (25.3)	1053 (45.8)	153 (6.7)	100 (4.4)	2297 (100)

( ): Row Percentage

Transition from Single: The repeat data show that 94.8% of singles remained single one year later, while 5.2% changed status. The no-repeat data show that a slightly smaller 93.8% of singles remained single and 6.2% changed status. The probabilities of no change as reported by the LIPRO study are always below 0.9 for individuals of at least 14 years old. The no-repeat data set offers transition probabilities that are closer to the LIPRO results.

Transition from Couple: The repeat data set yields a 91.3% probability of no change in household type, while the no-repeat data set offers a corresponding probability of 89.7%. Again the no-repeat data set indicates more frequent changes in household type. Transition to Family is most frequent from Couple.

Transition from Family: Family is the household type that is most stable, presenting the largest probability of no change; the repeat data set indicates 95.5% and no-repeat data set shows 94.2%. Transitions to Couple and Single Parent are most frequent. The estimated transition probability from Family to Single Parent is noticeably larger in the no-repeat data set than in the repeat data set (2.5% vs. 1.7%).

Transition from Single Parent: Single Parent is a household type that is more volatile than the above three, with a much lower frequency of remaining within its own category (77.4% in the repeat data set and 76.6% in the no-repeat data set). In the no-repeat data, the estimated transition probability to Family (14.3%) is noticeably larger than that indicated by the repeat data (9.9%), while the transition probability to Other is smaller (5.8% vs. 9.4%).

Transition from Other: This is the most unstable household type with the probability of transition to itself of 62.9% in the repeat data set and 65.2% in the no-repeat data. Due to the heterogeneous nature of this household type, no logit model is developed to explain transitions from this state (see Section 3.6).

In conclusion, the no-repeat data set which is least influenced by selective attrition exhibits, in general, less stability in household type transition. It can be expected that estimating logit models using the no-repeat data will yield better results. However, additional analyses are needed before selecting a data base.

3.4 Weighted Transition Frequency Tables

Weights developed using the attrition model presented in Table 3.2 are used in tabulating the household type transition frequency tables shown in Tables 3.5 and 3.6. Table 3.5 is based on the repeat data set, while Table 3.6 on the no-repeat data set. It is notable that the application of the weights have increased the stability in transition. Contrary to our expectation, households that changed household types did not necessarily have a higher estimated probability of leaving the panel. The use of attrition weight is nonetheless justifiable and preferable in light of the high attrition rate in the panel data.

The differences between the two weighted transition tables are examined using approximate statistical measures (they are approximate because the two tables are not independent and also because the frequencies are inflated due to weighting). However, while the two are different, their difference is not substantial. In fact, the application of the weights has decreased the differences between the two tables based on the two data pooling schemes. Thus the pooling scheme is not expected to substantially influence the result of logit model estimation.

Table 3.5
Weighted Household Type Transitions of the Repeat Data Set

Household Type at t	Household Type at t+1: Single	Couple	Family	Single Par.	Other	Total
Single	1806 (95.2)	15 (0.8)	12 (0.6)	16 (0.8)	49 (2.6)	1898 (100)
Couple	28 (1.2)	2205 (92.2)	126 (5.3)	0 (0.0)	32 (1.3)	2391 (100)
Family	2 (0.1)	98 (2.6)	3641 (94.8)	70 (1.8)	29 (0.8)	3840 (100)
Single Par.	16 (2.5)	4 (0.6)	21 (3.3)	521 (82.8)	67 (10.7)	629 (100)
Other	35 (8.6)	68 (16.6)	13 (3.2)	10 (2.4)	283 (69.2)	409 (100)
Total	1887 (20.6)	2390 (26.1)	3813 (41.6)	617 (6.7)	460 (5.0)	9167 (100)

( ): Row Percentage

Table 3.6
Weighted Household Type Transitions of the No-Repeat Data Set

Household Type at t	Household Type at t+1: Single	Couple	Family	Single Par.	Other	Total
Single	865 (94.3)	6 (0.7)	8 (0.9)	12 (1.3)	26 (2.8)	917 (100)
Couple	10 (0.9)	966 (90.8)	67 (6.3)	0 (0.0)	21 (2.0)	1064 (100)
Family	0 (0.0)	46 (3.1)	1395 (93.6)	37 (2.5)	12 (0.8)	1490 (100)
Single Par.	9 (3.8)	0 (0.0)	11 (4.6)	199 (83.6)	19 (8.0)	238 (100)
Other	11 (5.2)	45 (21.3)	2 (0.9)	2 (0.9)	151 (71.6)	211 (100)
Total	895 (22.8)	1063 (27.1)	1483 (37.8)	250 (6.4)	229 (5.8)	3920 (100)

( ): Row Percentage

3.5 Stationarity and Higher-Order History Dependence

The use of pooled data is based on the assumption that the transition matrix remains stable over time. This stationarity assumption was statistically examined in the study. The results offered empirical evidence that the transition matrix is indeed stable over time.

Another assumption examined in the course of this study is that of history independence. The presentation of household type transitions in the matrices of Tables 3.3 though 3.6 assumes that the household type at time t is dependent on that at time t - 1, but the history prior to time t - 1 does not influence the household type at t. In other words, household type transition is assumed to be conditionally independent of the past given the household type of the previous time period;

$$Pr[H(t) \mid H(t-1), H(t-2), H(t-3)...] = Pr[H(t) \mid H(t-1)]$$

where H(t) denotes the household type at time t, and Pr[A | B] represents the conditional probability of event A given event B.

A statistical examination of this history independence assumption offered an indication that the assumption is not valid. An attempt to construct a model with higher-order history dependence, however, ran into the problem of insufficient sample size. For example, consider the following second-order history dependent model:

$$Pr[H(t) \mid H(t-1), H(t-2), H(t-3)...] = Pr[H(t) \mid H(t-1), H(t-2)]$$

The household type at t, H(t), is now influenced by the household type at t - 2, H(t-2), as well as the household type at t - 1, H(t-1). Estimation of this model, however, requires data from three time points (t, t - 1, and t - 2), significantly reducing the sample size. Furthermore, it is extremely rare that a sufficient number of observations are available to estimate the probability of transitions that involve infrequent household types such as single parents and others. For these reasons, such elaborate representation of history dependence is not adopted in the current form of MIDAS.

3.6 Logit Models of Household Type Transition

An extensive set of variables is examined for inclusion in the models of household type transition. These variables include often-used person attributes, such as age, sex, employment, license holding, and income. In addition, a wide range of variables are defined to represent many additional household characteristics, e.g., employment of the household head, employment of the spouse of the head and the number of children by age group. The definition of the variables used in the analysis is presented in Table 3.7. The resulting models are described below.

Transition from Single: Due to the limited frequency of transitions from Single to the other household types, the model is formulated as a binary response model involving the transition from Single to Single (no change) and from Single to other (change in household type). The age, gender, education, income and employment status of the person are the explanatory variables of this model (Table 3.8). The model coefficients indicate that individuals of 18 to 25 years old tend to change their status. The variable, HD18-35, is suppressed in the model to avoid complete linear dependency among the age variables. On the other hand individuals older than 35 years are less likely to change. Non-employed single individuals tend to remain as

Table 3.7
The Variables Used in Logit Model Formulations

Variable	Definition
HD18-35	1 if the age of the head of the household is between 18 and 35 years.
HD25-35	1 if the age of the head of the household is between 25 and 35 years.
HD35-65	1 if the age of the head of the household is between 35 and 65 years.
HD65+	1 if the age of the head of the household is above 65 years
WF18-35	1 if the age of the spouse of the household is between 18 and 35 years.
CHLD06	Number of children less than 6 years old.
CHLD11	Number of children between 6 and 11 years old
CHLD17	Number of children between 12 and 17 years old
CHLD18	Number of children at least 18 years old.
MALE	1 if the head is male.
HDNOJOB	1 if the head is unemployed
WFNOJOB	1 if the spouse is unemployed
HDHIEDUC	1 if the head of the household has at least a University degree
WFHIEDUC	1 if the spouse of the household has at least a University degree
SQRTINC	Square root of the total personal income (dfl) divided by 1000.
Definition of Household Types	
SINGLE	1 if the household is a single person
COUPLE	1 if the household is composed of two adults of different gender
FAMILY	1 if the household is composed of two adults of different gender and there is at least one child
SGLPRN	1 if the household is composed of an adult and at least one child
OTHER	1 if the household is not part of any of the above categories

Table 3.8
Logit Model of Transition from SINGLE

Variable	$\hat{\beta}$	t	Weighted $\hat{\beta}$	t^*
Intercept	1.1662	1.07	1.4250	1.78
HD25-35	0.5665	1.07	0.6380	1.65
HD35+	1.4189	2.27	1.5396	3.43
MALE	-0.3365	-0.78	-0.3405	-1.08
HDNOJOB	1.6037	2.71	1.6850	3.95
HDHIEDUC	0.0929	0.21	-0.0083	-0.00
SQRTINC	0.0018	0.24	-0.0001	-0.03
$L(0)$	-281.4			
$L(c)$	-96.3			
$L(\hat{\beta})$	-86.3			
$-2(L(0)-L(\hat{\beta}))$	390.1			
$-2(L(c)-L(\hat{\beta}))$	20.0			
Observed Frequencies				
SINGLE-SINGLE	380		376	
SINGLE-OTHER	26		26	
Total	406		402	

* t-statistics are not consistent due to weighting.

singles. The other variables are not significant in either weighted or unweighted logit models.

Transition from Couple: The age and education of the head of the household and the employment status of the spouse are the explanatory variables of the model of transition from Couples (Table 3.9). The model is specified for three destination states: Couple, Family, and Other and Single Parent lumped into one category. A couple tends to remain to be a couple when the head of the household is over 65 years old and is highly educated. When the spouse is 18 to 35 years old, the couple is less likely to remain as a couple. The spouse's age is obviously related to the probability of having a child, thus making the transition from a couple to a family. A transition to another household type is more likely when the spouse is not employed.

Transition from Family: The model is specified for three destination states: Family, Couple, and Other (Table 3.10). Variables representing the number of children by age group comprise the explanatory variables of the model. The presence of pre-school children (0 to 6 years old) or children in the primary-school age (7 to 11 years old) leads to a higher probability of remaining as a family, while the presence of children of at least 18 years old increases the chance of changing to another household type. A variable representing the age of the head of the household has been examined, but found to be insignificant.

Transition from Single Parent: This model is specified for two destination states: Single Parent and Other (including Single, Couple, Family, and Other). The probability of remaining as a single parent is expressed as a function of the gender

Table 3.9
Logit Model of Transition from COUPLE

Variable	$\hat{\beta}$	t	Weighted $\hat{\beta}$	Weighted t+
Intercept(1)*	4.7456	8.89	4.6927	11.69
Intercept(2)**	0.8128	2.71	0.6979	3.00
HDNOJOB	0.6527	1.65	0.8508	2.75
HD35-65(2)	-0.0880	-0.17	0.0168	0.04
HD65+	0.3352	0.46	0.0265	0.05
WFNOJOB	-1.0475	-3.22	-1.1470	-4.42
WF18-35	-1.7987	-3.64	-1.7785	-4.81
L(0)	-650.4			
L(c)	-234.2			
$L(\hat{\beta})$	-209.0			
$-2(L(0)-L(\hat{\beta}))$	882.7			
$-2(L(c)-L(\hat{\beta}))$	50.3			
Observed Transitions				
COUPLE-COUPLE	531		518	
COUPLE-FAMILY	42		41	
COUPLE-OTHER	19		19	
Total	592		578	

* Intercept for Couple to Couple
** Intercept for Couple to Family
+ t-statistics are not consistent due to weighting.

Note: Since the model is trinomial, variables can appear in two out of the three alternatives. The number in the parenthesis indicates where the variable is included

Table 3.10
Logit Model of Transition from FAMILY

Variable	$\hat{\beta}$	**t**	Weighted $\hat{\beta}$	t^{+}
Intercept(1)*	2.9616	8.09	2.8937	10.05
Intercept(2)**	-0.2007	0.77	-0.1418	0.70
CHLD06	1.3605	2.62	1.5454	3.47
CHLD17	1.5442	4.39	1.4802	5.46
CHLD18	-0.9389	2.57	-0.8429	2.97
HD18-35	0.1036	0.20	0.1470	0.34
L(0)	-1145.9			
L(c)	-270.9			
$L(\hat{\beta})$	-241.1			
$-2(L(0)-L(\hat{\beta}))$	1809.5			
$-2(L(c)-L(\hat{\beta}))$	59.5			
Observed Transitions				
FAMILY-FAMILY	983		979	
FAMILY-COUPLE	27		27	
FAMILY-OTHER	33		32	
Total	1043		1038	

* Intercept for Couple to Couple
** Intercept for Couple to Family
+ t-statistics are not consistent due to weighting.

Table 3.11
Logit Model of Transition from SINGLE PARENT

Variable	$\hat{\beta}$	t	Weighted $\hat{\beta}$	t^*
Intercept	2.2266	4.49	2.3473	6.27
MALE	-1.8860	-3.47	-2.1748	-5.29
HD35-65	-0.0739	-0.13	-0.1732	-0.41
$L(0)=$	-89.4			
$L(c)=$	-55.6			
$L(\hat{\beta})=$	-49.9			
$-2(L(0)-L(\hat{\beta}))$	79.1			
$-2(L(c)-L(\hat{\beta}))$	11.5			
Observed Transitions				
1. SGLPRN-SGLPRN	109		106	
2. SGLPRN-OTHER	20		20	
Total	129		126	

* t-statistics are not consistent due to weighting.

of the head. The estimation result indicates that single fathers are more likely to change their single-parent status than are single mothers (Table 3.11). The age of the head of the household does not seem to influence the household type transition for single parent households.

These logit models are used in MIDAS to determine the probabilities that a given household will make a transition to the respective destination states. These probabilities are in turn used to simulate transition of household types. Due to limitations in the sample size, it was not possible to formulate logit models such that a transition probability is determined for every pair of household types. For those pairs for which logit models are not formulated, observed transition frequencies are used as transition probabilities.

3.7 Summary

In this section the adoption of tabulation schemes and weighting procedures for the household type transition are explained. The topics presented are selective panel attrition and household type transitions, alternative data pooling schemes and household type transitions, and models depicting household type transitions. The use of weights based on a multi-wave ordered response probit model showed that panel attrition does not bias the household type transitions substantially. The no-repeat pooling data scheme (in which households appear in the data set only once) is preferred to the repeat pooling scheme (in which households appear in the data set for as many times as they responded to the panel survey) because it exhibits less stability in household type transition and is least influenced by selective panel attrition. Four logit models of household type transition are estimated using the no-repeat data set. Two versions for each model are presented–one for the attrition-weighted data set and one for the unweighted data set. In MIDAS the probability that a given household will make a transition is determined based on the logit models estimated from the no-repeat data set.

4. Preliminary Analysis of Person Attributes

The development of MIDAS is preceded by a series of causal analyses of demographic and socioeconomic characteristics of household members. The methodology and the results of the causal analyses are summarized in this section. The analysis is a preliminary investigation aiming at the identification of the causal structures that determine household members' employment, personal income, and driver's license holding. The resulting causal models are not used in MIDAS because the analysis did not lead to parsimonious and robust models. Nevertheless the results of this analysis have offered useful insights that are the basis of the model development discussed in the next section. The log-linear model of cross-classification table analysis is used here because of the ease it offers in evaluating many alternative causal structures (the VMS version of the BMDP Statistical Package is used in the log-linear analysis of this section). For a comparison of the log-linear model and the structural equations modelling approach with binary response variables, see van Wissen and Golob (1990).

4.1. Method

The log-linear model is appealing when the variables involved in an analysis are categorical with relatively few categories. It determines an expected number of observations in each cell of a cross-classification table. The expected cell frequency is modeled using the "main effects" and "interaction effects" of the variables used to formulate the table (these effects are quite analogous to those in the analysis of variance, ANOVA). The effects are combined in a multiplicative form such that the logarithm of an expected cell frequency is expressed as a linear combination of the effects (hence the term, log-linear model).

For example, consider a three-way classification table defined by three variables, X, Y, and Z. Let i, j, and k be categories of X, Y, and Z, respectively. A saturated log-linear model, which exhausts the degrees of freedom and completely replicates the observation, can be presented as

$$\ln[m(i,j,k)] = \mu + U(i) + V(j) + W(k)$$
$$+ UV(i,j) + VW(j,k) + UW(i,k) + UVW(i,j,k)$$

where

m(i,j,k)	=	expected frequency of cell (i,j,k),
μ	=	grand mean,
U(i), V(j), W(k)	=	main effects of X (= i), Y (= j), and Z (= k), respectively,
UV(i,j)	=	interaction effect of X (= i) and Y (= j), etc., and
UVW(i,j,k)	=	interaction effect of X (= i), Y (= j) and Z (= k).

An important difference exists between ANOVA and log-linear models. In ANOVA there is the distinction between a response (dependent) variable and explanatory (independent, or, grouping) variables. Log-linear models can be applied without such distinction. ANOVA attempts to account for the variation in the response variable using the main and interaction effects of the explanatory, or classification, variables. In the log-linear model approach, it is attempted to assess and "describe the structural relationship among the variables corresponding to the dimensions of the table" (Fienberg, 1973).

A set of nested log-linear models can be used to evaluate whether a certain causal relationship fits the observations represented as a cross-classification table (Goodman, 1973). In this analysis, many alternative causal structures are tested and most plausible ones are chosen to describe selected attributes of household members.

The definition of a few new terms introduced in this section is due at this point. The relation between factors observed at the same time point is designated as a "synchronous relation" or "synchronous link." A synchronous relation can be assumed between the employment status of a household member at time t and his/her personal income also at time t. The association of a factor with itself across time points is called an "inertial link." A good example is the relation between a household member's personal income at time t-1 and that at time t. The association between different factors across different time points is called a "cross-lagged link," e.g., employment at time t-1 and income at time t. An ordered combination of factors and links is called a "causal chain" and represents how factors influence each other.

Causal chains are developed for the following critical attributes of the household members as the response variables:

Employment at time t (Employed, Not Employed),
Personal income at time t (Low, Medium, High), and
Driver's license holding at time t (Yes, No),

where the categories used are presented in the parentheses. The following personal characteristics are considered as input (explanatory) variables:

Age at t,
Gender,
Education at t,
Employment at t-1,
Personal income at t-1, and

Driver's license holding at t-1.

The distinction between explanatory and response variables is derived by considering the nature of the variables used. For example, age cannot be influenced by personal income but personal income can be influenced by age. Therefore, age may be used as an explanatory variable and personal income as a response variable.

4.2. Causal Chain for Employment

The causal analysis performed for employment involves five variables:

AGE(t)	[A]
EDUCATION(t)	[E]
GENDER	[S]
EMPLOYMENT(t-1)	[O]
EMPLOYMENT(t)	[U].

EMPLOYMENT(t) is the response variable. The series of models considered as possible causal chains are presented in Table 4.1. The models are described in the table using the symbols shown above in brackets.

Each model is specified as a set of effects, e.g., [AES,EO,AO,SO,OU]. AES represents the three-way interaction term involving A (AGE(t)), E (EDUCATION(t)), and S (GENDER), and EO is the two-way interaction of E (EDUCATION(t)) and O (EMPLOYMENT(t)). All models are hierarchical, and the inclusion of an interaction effect implies that the lower-order interaction effects and main effects nested in it are also included in the model. For example, the inclusion of effect AES implies that two-way interaction effects, AE, ES, and AS, and main effects, A, E, and S, are also included in the model. Thus the model, [AES,EO,AO,SO,OU], comprises AES, AE, ES, AS, EO, AO, SO, OU, A, E, S, O, and U.

Figure 4.1 presents the causal chain developed for employment. The chain is based on model [AES,AO,SO,OU], which was chosen because it effectively explains the variation in the observation with its relatively simple structure. This choice, however, is subjectively made balancing the model's fit and simplicity. Also note that none of the models presented in Table 4.1 fits the observation well as indicated by the large χ^2 values, suggesting that no parsimonious causal structure will be able to well explain employment status.

The significant links are the inertial link between EMPLOYMENT(t-1) and EMPLOYMENT(t), and the synchronous link between EMPLOYMENT(t) and AGE(t). The former link indicates the presence of strong continuity and inertia in employment status. The latter reflects such immediately recognizable associations as young and employed, or old and retired.

Other important linkages include the synchronous link between GENDER and EMPLOYMENT(t), which reveals the presence of gender differences in labor force participation. Also, important is the synchronous link between EDUCATION(t) and EMPLOYMENT(t), revealing the anticipated relation that education influences the probability of employment. Finally, the three-factor interaction involving AGE(t), GENDER, and EDUCATION(t) is significant, suggesting cohort effect in female participation in labor force.

Table 4.1

Selected Log-Linear Models of Employment

Model	d.f.	χ^2	α
1] AES,EO,AO,SO,OU	40	237.29	0.0000
2] AES,EO,AO,SO,AU,OU	30	176.13	0.0000
3] AES,EO,AO,SO,SU,OU	39	215.14	0.0000
4] AES,EO,AO,SO,EU,OU	39	232.56	0.0000
5] AES,EO,AO,OU	41	607.77	0.0000
6] AES,EO,SO,OU	43	794.23	0.0000
7] AES,AO,SO,OU	41	240.67	0.0000

A = AGE(t)
E = EDUCATION(t)
S = GENDER
O = EMPLOYMENT(t)
U = EMPLOYMENT(t-1)

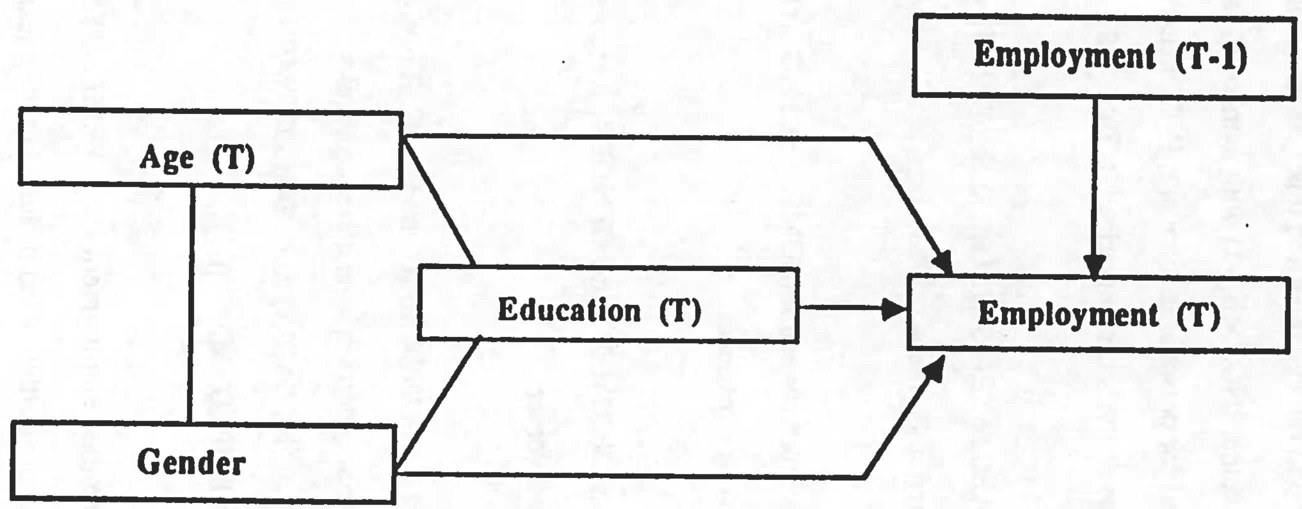


Figure 4.1

The Log-linear structure for Employment

4.3. Causal Chain for Personal Income

In the causal analysis of personal income, the employment status is viewed as its primary determinant. In addition to the variables and linkages involved in the employment causal chain, past personal income, INCOME(t-1) with symbol P, is introduced into the causal chain. The symbol used for personal income at t, INCOME(t), is I. A selected set of the models considered in the analysis is summarized in Table 4.2.

The model chosen is [AES,AO,EO,SO,OI,EI,AI,SI,OU,IU,IP]. The causal chain implied by this model is presented in Figure 4.2. The significant linkages in this causal chain are:

- the synchronous link between INCOME(t) and AGE(t), presumably reflecting the effect of seniority or experience,

- the link between INCOME(t) and GENDER, which reveals income differences between men and women,

- the association between INCOME(t) and EDUCATION(t) exhibiting the anticipated correlation between education level and wage, and
- the link between INCOME(t) and EMPLOYMENT(t) and the link between INCOME(t) and EMPLOYMENT(t-1).

The implication that the past employment status (EMPLOYMENT(t-1)) influences the current income is quite noteworthy. Also included in the causal chain are links present in the employment causal chain, including the inertial link between EMPLOYMENT(t-1) and EMPLOYMENT(t) and the links between EMPLOYMENT(t) and age, gender, and education.

Table 4.2
Selected Log-Linear Models of Personal Income

Model	d.f.	χ^2	α
1] AES,EO,AO,SO,OI,EI,AI,SI,OU, EU,IU,EP,OP,IP	247	501.81	0.0000
2] AES,EO,AO,SO,OI,EI,AI,SI,OU, IU,EP,OP,IP	248	504.15	0.0000
3] AES,EO,AO,SO,OI,EI,AI,SI,OU, EU,EP,OP,IP	249	575.20	0.0000
4] AES,EO,AO,SO,OI,EI,AI,SI,OU, EU,IU,OP,IP	249	514.58	0.0000
5] AES,EO,AO,SO,OI,EI,AI,SI,OU, EU,IU,EP,IP	249	516.09	0.0000
6] AES,EO,AO,SO,OI,EI,AI,SI,OU, EU,IU,EP,OP	298	2112.41	0.0000
7] AES,EO,AO,SO,OI,EI,AI,SI,OU, IU,OP,IP	250	516.93	0.0000
8] AES,EO,AO,SO,OI,EI,AI,SI,OU, EU,OP,IP	251	587.97	0.0000
9] AES,EO,AO,SO,OI,EI,AI,SI,OU, IU,IP	252	596.69	0.0000

A = AGE(t)
E = EDUCATION(t)
S = GENDER
O = EMPLOYMENT(t)
U = EMPLOYMENT(t-1)
I = INCOME(t)
P = INCOME(t-1)

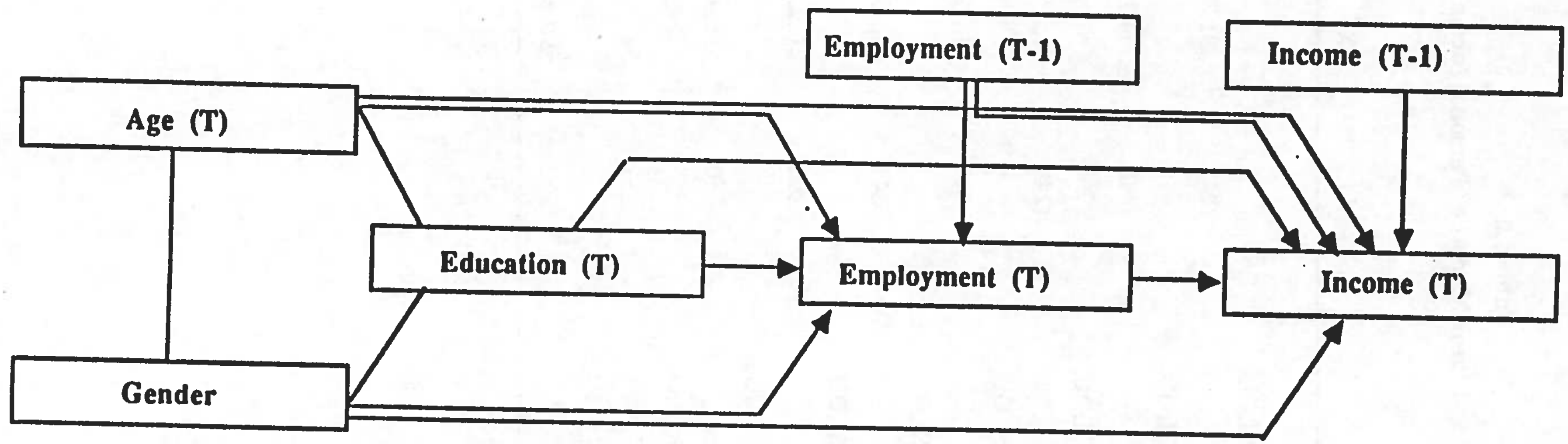


Figure 4.2

The Log-linear structure for Personal Income

4.4. Causal Chain for Driver's License Holding

A causal chain for driver's license holding, LICENSE(t), is developed following similar steps. The factors present in this causal chain are: AGE(t) [A], GENDER [S], EDUCATION(t) [E], EMPLOYMENT(t) [O], INCOME(t) [I], and past driver's license holding (LICENSE(t-1)) [D]. As before the symbols used are indicated in the brackets. The symbol for LICENSE(t) is L. Selected log-linear models are summarized in Table 4.3, and the final causal chain, which is obtained from Model 1 of Table 4.3, is shown in Figure 4.3.

The analysis shows that links exist between age and license holding, gender and license holding, education and license holding, employment and license holding, and income and license holding. The most important is the inertial link between driver's license holding in the past (t-1) and that of the present (t). The significant inertial linkages found commonly for employment, and license holding form the basis for the use of a Markovian transition mechanism to simulate these personal attributes in MIDAS.

4.5 Summary

In this section a preliminary investigation on the possible causalities among person attributes is examined. Through the formulation of causal chains, factors affecting critical attributes of household members are selected. The log-linear model structures analyzed are used as modelling guidelines for MIDAS' socio-demographic component. The most salient finding is that age and gender together with a lagged structure or inertial link (e.g. past employment) are more likely to depict the evolution of person attributes more closely. Most importanlty, the significant inertial links for employment and driver's license holding suggest that Markovian transition models should be used for employment and driver's license holding. For the personal income a more complex structure may be needed to represent most of the significant effects.

Table 4.3
Selected Log-Linear Models of Driver's License Holding

Model	d.f.	χ^2	α
1] AES,EO,AO,SO,OI,EI,AI,SI, IL,AL,SL,EL,OL,DL	249	578.40	0.0000
2] AES,EO,AO,SO,OI,EI,AI,SI, IL,SL,EL,OL,DL	251	609.95	0.0000
3] AES,EO,AO,SO,OI,EI,AI,SI, IL,AL,EL,OL,DL	250	588.12	0.0000
4] AES,EO,AO,SO,OI,EI,AI,SI, IL,AL,SL,OL,DL	250	582.91	0.0000
5] AES,EO,AO,SO,OI,EI,AI,SI, IL,AL,SL,EL,DL	250	647.60	0.0000
6] AES,EO,AO,SO,OI,EI,AI,SI, AL,SL,EL,OL,DL	251	688.66	0.0000
7] AES,EO,AO,SO,OI,EI,AI,SI, IL,AL,SL,EL,OL	251	2902.42	0.0000

A = AGE(t)
E = EDUCATION(t)
S = GENDER
O = EMPLOYMENT(t)
I = INCOME(t)
L = LICENSE(t)
D = LICENSE(t-1)

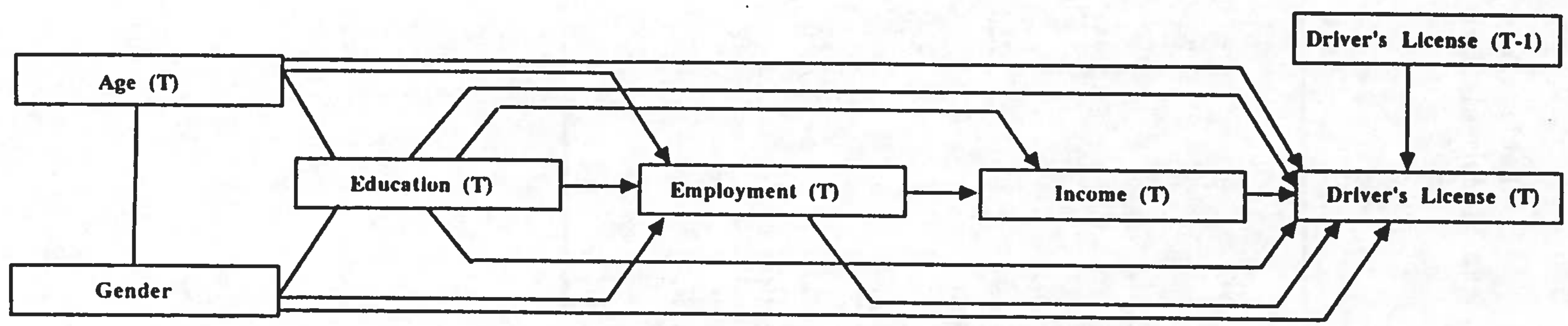


Figure 4.3
The Log-linear Structure for Driver's License Holding

5 Socioeconomic and Demographic Component of MIDAS

MIDAS is a simulator of household characteristics and mobility. Its aim is to realistically recreate the progression of a household through life cycle stages, and simulate changes in the household members' socioeconomic attributes and demographic attributes, such as employment status and driver's license holding, then to use these endogenously generated socioeconomic attributes to forecast household car ownership and mobility. The understanding gained through the causal analysis reported in the previous section is fully utilized in the development of MIDAS. The socio-demographic components are integrated with mobility components (car ownership, trip generation, modal split , and travel distance by mode) to form a comprehensive simulation system.

In the simulation, a household member will age, form an independent household, gain employment, obtain a driver's license, marry, give birth, and so on. The size and composition of the household will change accordingly. A household member may be added to a household through a marriage, or a household may be split into two through a divorce. A child will leave his parents and form a new household. Such changes are probabilistically generated in the simulation. The model parameters that determine the probability of these events are obtained from the Dutch Panel data set.

Many household and person characteristics are correlated with each other. For example, the employment status of a woman is related to the number of small children. There is strong association between the age of the head and the age of the spouse of a household. The ages and number of children in a family are strongly associated with the age of the mother. It is critically important that these internal correlations are accurately reflected in the simulation. MIDAS achieves this by specifying the probability of a change as a function of pertinent household and person characteristics.

5.1 Household Type Transition

For each household in the simulation, its characteristics are first read from an input file comprising records of sample households from the Dutch Mobility Panel data set. Following this, the transition between household types is simulated for each time period (one year is used as the time interval of the simulation). This process is based on the set of logit models described in Section 3 that determine transition probabilities for each household considering its attributes, e.g., the adult household members' age, education, employment, and presence of children by age group.

Given a transition from a household type to a new type, the household attributes are modified to conform to the new household type. This sequential process is based on the following identity:

$$\Pr[H(t), X(t) \mid H(t-1), X(t-1)]$$
$$= \Pr[X(t) \mid H(t), H(t-1), X(t-1)]\Pr[H(t) \mid H(t-1), X(t-1)]$$

where H(t) is the household type at time t, and X(t) denotes a vector of household attributes.

A set of subroutines has been developed to probabilistically change the attributes of household members, generate new members, or to remove individuals from the household. Table 5.1 summarizes the subset of subroutines that are called in the simulation program

Table 5.1
MIDAS Subroutines Used in Connection with Household Type Transition

Household Type Transition

From:	To: Single	Couple	Family	Sgl Prnt	Others
Single	VANISH	MARRIG	MARRIG	BIRTH	MOVEIN
Couple	DIVORC SURVIV	(a)	BIRTH	(b)	MOVEIN
Family	DIVORC SURVIV	MOVEOT	FERTIL MOVEOT	DIVORC SURVIV	MOVEIN
Sgl Prnt	MOVEOT	(b)	MARRIG MOVEOT	FERTIL MOVEOT	MOVEIN
Others	DELETE	DELETE	DELETE	DELETE	(a)

(a) No change assumed

(b) The transition assumed to be impossible

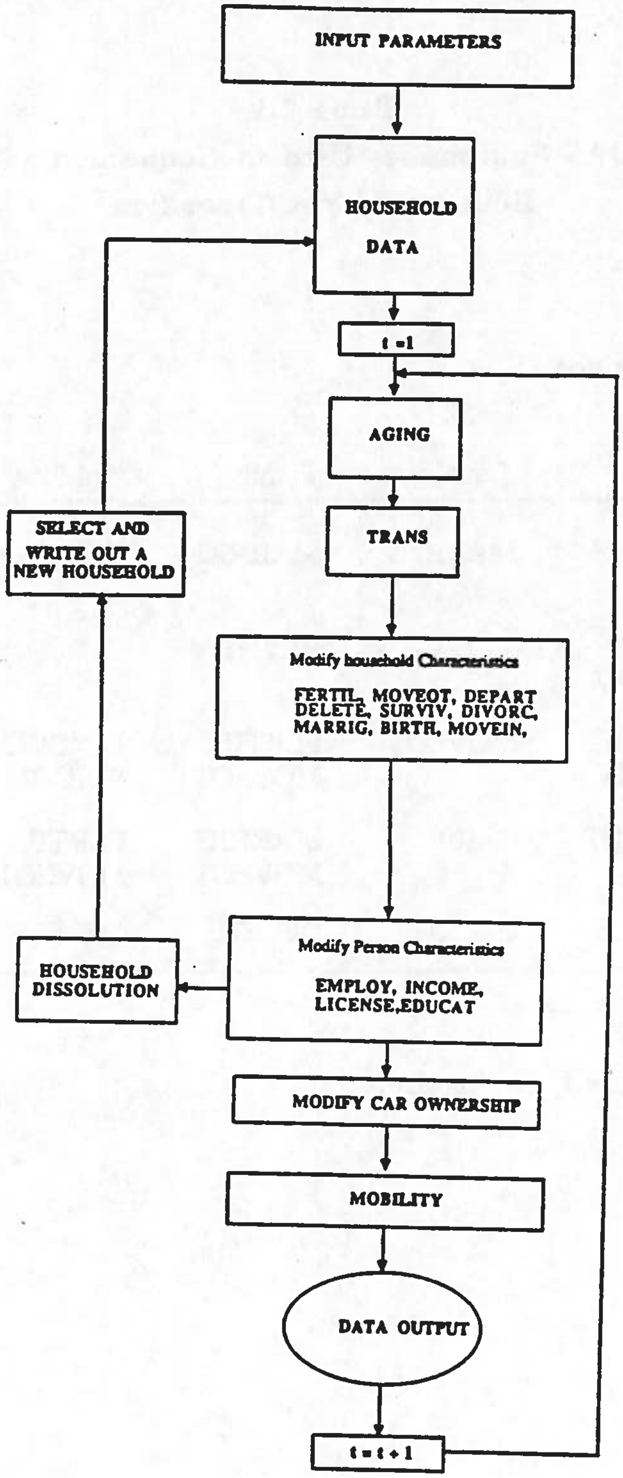


Figure 5.1

Simulation Process Flowchart

for respective pairs of household types while Figure 5.1 summarizes the algorithm followed in moving households and persons from one year to the next.

The composition of a household may change even when the household type remains unchanged. For example, the birth of a second child obviously increases the household size, but the household type will remain the same. Such changes without changes in household type are accounted for in MIDAS by examining the possibility that a new member will be introduced or an existing member will leave the household. For example, subroutines FERTIL and MOVEOT are called in connection with the transition from family to family, or from single parent to single parent, when the number of children is two or more.

The routine VANISH, called in connection with the transition from single to single, accounts for the possibility that the member of a single-person household passes away, thus the household vanishes. The possibility that a household dissolves is also examined in subroutine AGING, which is called before the transition of household types are simulated. If subroutine AGING indicates that one (or more) of the household members will pass away (or leave the household), then no further change in household composition is assumed. Otherwise, subroutine TRANS is called to examine the possibility of change in the household type.

5.2 Birth and Death

The probability that a woman in a household will give birth to a child in a given year is expressed as a function of the age and employment status of the woman, and the number of children that already exist in the household. Observed frequencies obtained from the Dutch Panel data set are used to determine the probability.

A birth may be implied by a change in the household type (e.g., a couple to a family). In such cases, the logit models of household type transitions depict the probability of a birth. For example, the probability of a transition from couple to family is expressed as a function of the man's age and education, and the woman's employment status. The event of birth is randomly generated in the simulation using these probabilities.

The probability of death is considered as a function of the age and sex of the individual. The death rate is determined from the decrease in the size of each age group in the population, provided by van den Broecke. The following relation is used:

$$A'(t+5)/A(t) = (1 - r)^5$$

where $A(t)$ is the size of an age group that contains individuals of age i to $i + 5$, measured in year t; $A'(t+5)$ is the size of the same cohort measured at time $t + 5$ (therefore contains individuals of age $i + 5$ to $i + 10$); r is the death rate for the age group.

A single-person household is removed when a death takes place in the simulation. The possibility of death is also considered in connection with the transition from couple (or family) to single (or single parent). If a death does not take place in the simulation, then the transition is regarded as a result of a divorce, and the household is split into two households.

5.3 Independent Households Formed by Children

The event of "leaving the nest," i.e., a child moving out and forming an independent household, is modeled as a function of the age, sex, and employment status of the child.

Similar to the case of birth, this event is implied by household type transition from family to couple, or from single parent to single. The probabilities of these transitions are represented by the logit models as function of the number of children by age.

When the event of nest leaving takes place in the simulation, a new household is added to the data file with a certain probability (this represents the probability that the new household will remain in the same municipality). The evolution of this new household is simulated through the rest of the simulation period.

5.4 Employment

The employment status of a person is determined using transition matrices developed by sex and age group (Table 5.2). Each matrix contains the probability of change in employment from one status to another. For example, the two-by-two matrix for men in the 18 to 24 age bracket indicates that a person who is employed at time t will also be employed at time t + 1 with probability 0.929, and will not be employed with probability 0.071. Similarly, a person who is not employed at time t will gain employment with probability 0.160, and remain unemployed with probability 0.840, at time t + 1.

The use of transition matrices reflects the assumption that the probability of employment in period t + 1 depends on the employment status of period t. It is evident from the table that men in the age brackets of 35 to 44 and 45 to 55 have extremely high probability of continuous employment. A child is assumed to be not employed until he reaches the age of 18 years old.

Table 5.2
Employment Status Transition Matrix
By Age and Gender

Age		Men: Employed	Men: Not Employed	Women: Employed	Women: Not Employed
18-24	Employed	.929	.071	.882	.118
	Not Employed	.160	.840	.129	.871
25-34	Employed	.967	.033	.865	.135
	Not Employed	.354	.646	.094	.906
35-44	Employed	.982	.018	.909	.091
	Not Employed	.435	.565	.072	.928
45-54	Employed	.992	.008	.854	.146
	Not Employed	.130	.870	.097	.903
55-64	Employed	.878	.122	1.000	.000
	Not Employed	.024	.976	.019	.981
65	Employed	.750	.250	.000	1.000
	Not Employed	.014	.947	.000	1.000
Total	Employed	.966	.034	.881	.119
	Not Employed	.167	.833	.075	.925

5.5 Income Models

Given the employment status, the personal income is determined using a set of models. Each model is formulated with a lagged dependent variable and a serially correlated error term. Thus the personal income at time t is assumed to be determined in part by the personal income at time t - 1. It is also assumed that the unexplained effect of time t - 1 and that of time t are correlated with each other.

Models are developed for the four possible combinations of the employment status at time t - 1 and time t: (not employed, not employed), (employed, not employed), (not employed, employed), and (employed, employed). Two models are estimated for each status pair using data from Waves 1, 3, 5, 7 and 9. The square-root of annual income is the dependent variable of one of the models, while the income itself is used as the dependent variable of the other. The results are summarized in Tables 5.3 and 5.4. Note that the income of an unemployed person is not automatically assumed to be 0 (possible association between the error term of the income model and the change in employment status is ignored in this analysis).

Both types of models offer similar behavioral indications for each employment status pair. Age, sex, and household size are the most significant variables in the models for status pair (not employed to not employed), while sex and education are most dominant and age has relatively smaller effects in the model for (employed to employed). Sex and age are important contributing factors in the models for the remaining pairs with changes in employment status.

Striking is the mostly insignificant coefficients of the lagged dependent variables. Only the models for (employed to employed) have significant coefficient estimates, but the

Table 5.3
Personal Income in $\sqrt{\text{Annual D.Fl / 1000}}$, by Employment History

Variable	Unemployed in t and t-1 $\hat{\beta}$	t	Employed in t-1 and Unemployed in t $\hat{\beta}$	t	Unemployed in t-1 and t Employed in t $\hat{\beta}$	t	Employed in t and t-1 $\hat{\beta}$	t
Intercept	1.68	12.7	1.99	3.5	3.21	10.5	3.25	34.6
AGE2544	1.21	14.4	0.76	1.7	0.25	1.2	0.51	6.8
AGE4564	1.35	15.2	1.65	4.0	0.39	1.3	0.71	9.4
AG65+	2.76	26.4	1.96	2.7	0.71	1.2	0.82	2.0
MALE	1.45	25.1	2.45	9.6	0.98	6.7	1.45	38.6
HIEDUC	0.18	2.9	0.33	1.2	0.33	2.2	0.48	12.2
CHLD11	0.30	4.2	0.76	2.9	-0.01	-0.1	0.09	2.4
CHLD17	0.18	3.2	0.65	2.5	-0.21	-1.5	0.00	0.0
CHLD18	0.48	8.1	0.42	1.7	0.41	2.9	-0.02	-0.6
HHSIZE	-0.68	-16.1	-0.73	-4.5	-0.28	-2.8	-0.10	-4.3
INCOME(t-1)	0.01	0.8	0.07	1.0	0.05	1.5	0.03	3.8
$\hat{\rho}$	0.39	20.4	-0.02	-0.2	0.07	1.0	0.37	18.6
N	2926		166		250		2388	
R^2	0.866		0.746		0.622		0.843	
Replication								
Predicted	1.555		2.482		3.078		4.717	
Observed	1.578		2.500		3.239		4.723	
MAE	0.760		1.145		0.802		0.429	
MSE	1.026		2.047		1.070		0.359	
Prediction								
N	1327		95		96		1206	
Predicted	1.828		3.104		3.100		4.817	
Observed	1.916		3.120		3.366		4.948	
MAE	0.770		1.068		1.081		0.485	
MSE	1.061		1.954		1.947		0.429	
R^2	0.883		0.737		0.289		0.861	

$\hat{\rho}$ = coefficient of serial correlation
MAE = Mean absolute error, average of the absolute difference between observed and estimated value
MSE = Mean square error, average of the squared difference between observed and estimated value

Definition of Variables used in the Income Models

AGE2544	1 if the age of the head of the household is between 25 and 44 years.
AGE4564	1 if the age of the head of the household is between 45 and 64 years.
AG65+	1 if the age of the head of the household is above 65 years
MALE	1 if the head is male.
HIEDUC	1 if the head of the household has at least a University degree
CHLD11	Number of children less than 11 yeras old.
CHLD17	Number of children between 11 and 17 yeras old
CHLD18	Number of children at least 18 years old.
HHSIZE	Household size

Table 5.4
Personal Income in Annual D. Fl / 1000, by Employment History

	Unemployed in t and t-1		Employed in t-1 and Unemployed in t		Unemployed in t-1 and Employed in t		Employed in t and t-1	
Variable	$\hat{\beta}$	t	$\hat{\beta}$	t	$\hat{\beta}$	t	$\hat{\beta}$	t
Intercept	4.09	6.3	3.89	1.4	10.53	5.6	9.39	12.8
AGE2544	5.76	14.0	4.74	2.1	3.01	2.4	4.89	8.3
AGE4564	7.62	17.6	11.07	5.3	4.20	2.3	7.31	12.1
AG65+	13.84	27.0	15.07	4.0	6.32	1.8	7.79	2.4
MALE	7.42	26.4	12.88	9.8	6.77	7.4	11.56	38.9
HIEDUC	1.05	3.3	2.94	2.1	2.71	2.9	4.50	14.4
CHLD11	1.06	3.1	4.00	3.0	0.23	0.2	0.86	2.7
CHLD17	0.76	2.7	3.47	2.6	-0.89	-1.0	0.19	0.7
CHLD18	1.59	5.5	1.72	1.3	2.47	2.8	-0.27	-0.9
HHSIZE	-2.41	-11.6	-3.08	-3.7	-1.78	-2.9	-0.57	-2.9
INCOME(t-1)	0.01	1.2	0.11	1.9	0.05	1.4	0.06	5.7
$\hat{\rho}$	0.41	22.2	-0.08	-1.0	0.07	1.1	0.38	19.1
N	2926		166		250		2388	
Correlation	0.869		0.773		0.629		0.857	
Replication								
Predicted	6.531		10.356		11.479		23.506	
Observed	6.539		10.865		12.196		23.509	
MAE	3.381		5.721		5.028		3.588	
MSE	22.555		55.655		40.748		21.761	
Prediction								
N	1327		95		96		1206	
Predicted	7.804		14.080		11.427		24.692	
Observed	8.316		14.000		13.266		25.967	
MAE	3.780		6.346		6.640		4.261	
MSE	30.539		68.555		77.687		30.499	
Correlation	0.882		0.748		0.312		0.884	

$\hat{\rho}$ = coefficient of serial correlation
MAE = Mean absolute error, average of the absolute difference between observed and estimated value
MSE = Mean square error, average of the squared difference between observed and estimated value

highly significant with t-statistics of around 20 for both (employed, employed) and (not employed, not employed).

The results indicate that annual income is not as stable as one might have thought; the income of year t is not significantly influenced by the income of year t - 1, unless the person was employed in both periods. Unobserved factors exert similar influences over time when the employment status does not change, as indicated by the highly significant coefficients of serial correlation. Presumably there are individual-specific income effects that are longitudinally stable. The estimation results suggest that those who were earning more (or less) than the amount expected for an otherwise identical individual would do so over time unless the employment status changes, but the exact earning tends to vary over time.

The models estimated here all offer extremely good fit to the data as the high R^2 values indicate. In order to evaluate further the models' usefulness in forecasting, they were "validated" using the Wave-10 data that were not used in model estimation. Wave-10 income was predicted using the observed Wave-10 explanatory variables and the coefficient estimates obtained using the data from Waves 1 through 9. The results are shown also in Tables 5.3 and 5.4. It is evident from the tables that these models not only "replicate" the data used for estimation well, but "predict" incomes in the data not used for their estimation. It is worthy to note that the "prediction" R^2's are almost as good as, and in one occasion better than, the "replication" R^2's. Only exceptions are the models for the (not employed, employed) pair. Overall, the usefulness of these models in forecasting is evident through this validation analysis.

Since there are no discernible differences between two classes of models, the ones specified for annual income itself are used in MIDAS. The personal incomes of household members are added up in the simulation to obtain total household income. It is important to note that the employment transition matrices and the parameters of the income models are estimated using data obtained in a period of economic expansion (1984 through 1988). These parameters must be appropriately adjusted if the model is to be applied for a period of stable economy or economic recession. This adjustment requires examination of the impact of regional and national economy on the parameters of these model components, which is obviously outside the scope of this study.

5.6 Driver's License and Education

The driver license holding is determined using transition matrices similar to those for employment status (Table 5.2). Compared with the transition matrices for employment status, the matrices of Table 5.5 in general have larger diagonal elements, which correspond to the transition from licensed to licensed, or from non-licensed to non-licensed. This implies that license holding status is less variable than employment status. Also notable is the stability in the transition probabilities across the age groups.

The level of education is an important variable as the causal analysis has indicated. Because education is among the explanatory variables used in the MIDAS mobility component, it is necessary to determine education levels for those household members that are internally generated in the simulation process. This determination is not based on detailed modeling of education levels as it is clearly beyond the scope of this study.

For children that are generated in the simulation, their education levels are determined randomly using the distribution of education levels by sex, obtained for individuals of 18 through 28 years old in the panel data. Education levels of new members that enter a household through a marriage are determined using the correlation between the education

Table 5.5
Driver's License Holding Status Transition Matrix
By Age and Gender

Age		Men: Licensed	Men: Not Licensed	Women: Licensed	Women: Not Licensed
18-24	Licensed	.988	.012	1.000	.000
	Not Licensed	.232	.768	.176	.824
25-34	Licensed	.974	.026	.975	.025
	Not Licensed	.103	.897	.000	1.000
35-44	Licensed	.987	.013	.983	.017
	Not Licensed	.059	.941	.077	.923
45-54	Licensed	.978	.022	.989	.011
	Not Licensed	.111	.889	.000	1.000
55-64	Licensed	.986	.014	.957	.043
	Not Licensed	.063	.938	.015	.985
65	Licensed	.962	.038	1.000	.000
	Not Licensed	.000	1.000	.061	.939
Total	Licensed	.980	.020	.981	.019
	Not Licensed	.128	.872	.056	.944

through 28 years old in the panel data. Education levels of new members that enter a household through a marriage are determined using the correlation between the education levels of married men and women. For example, the probability that a man (or, a groom) has a given education level is determined by the education level of the woman (the bride) who has been a member of the household in the simulation.

The education levels of new "other" household members (not the head, spouse, or their children) are determined using the distribution of education levels of "other" individuals by age and sex, obtained from the Dutch Panel data. Further discussions on the generation of new individuals can be found in the following section.

5.7 Attributes of New Household Members

A set of personal attributes needs to be generated whenever a new household member is introduced in the simulation. As discussed earlier, in case where a new person enters a household through a marriage, his/her age and education level is determined based on the existing member's age and sex. The new member's employment and income are then determined given his/her age and sex.

For a newborn member of a household, only sex is determined at the time of birth; the rest of person attributes are determined when he/she reaches the age of 18, using the probabilities of employment, license holding, and income as described above.

The person attributes of "other" household members are determined as follows. First, the age and sex of the "other" individual are randomly generated based on the age of the head of the household. Given age and sex, employment, license holding, education, and income, are randomly determined base on the observed distribution of the attributes of "other" persons, by age and sex.

5.8 Household Dissolution

A household is split into two, or eliminated from the simulation, after a divorce or other events that cause its dissolution. If children are present in the household, they are randomly assigned to the respective parents. The current version of MIDAS assumes that the mother will have the custody of a child with a probability of 75%. This, however, is an arbitrary assumption that should be improved in the future with appropriate data.

MIDAS assumes that only a fraction of newly formed households (formed through divorces or by children gaining independence) remain in the simulation. In preliminary simulation runs of this study, 15% of new households are retained. This may be viewed to represent the case where 15% of divorced household members or children leaving the parents remain in the same area. The particular value, 15%, is chosen because, with this rate, new households roughly replace households that disappear due to death, and keep the total number of households in the simulation stable over simulation years. Thus this value represents a demographically stable region. The parameter can be increased or decreased to reflect demographic growth or decline.

5.9 Summary

The socioeconomic and demographic component of MIDAS is made of a set of subroutines that probabilistically change the attributes of existing household members, generate new members and their attributes, and remove individuals from existing households during the simulation. These subroutines use models of household type transition, models of birth and death, transition matrices of employment, models of personal income, transition matrices of driver's license holding and education, and probabilistic assignment methods of values to new household member's attributes.

The household type transition models are the logit models of Section 4 for the most frequent transitions, while for the less frequent transitions observed panel data transitions are used. The probability of a woman in a household having a child is given as a function of age and employment status of the woman, and of the number of children in the household. The probability of death is a cohort-based death rate. The employment status of a person is determined based on observed transition matrices by age group[1]. Similarly, transition matrices have been used for driver's license holding and education. The models of personal income are a set of four dynamic models by employment history (i.e., for a person that is employed at t and t-1, unemployed at t and t-1, employed at t and unemployed at t-1, or unemployed at t and employed at t-1). The models include lagged dependent variables and serial correlation.

The attributes of new household members are computed in a similar fashion to the attributes of old household members. The possible household dissolution is determined based on the possible occurrence of a divorce or a death. Children are allowed to form new households based on age, gender and employment status of the child. When a new person enters a household through a marriage, his/her age and education level is determined based on the partner's age and sex. The new member's employment and income are then determined given his/her age and sex based on the same models and parameters used for the other household members.

In this way the progression of a household through life cycle stages is recreated and changes in the household members' socioeconomic and demographic attributes are simulated. These endogenously generated socioeconomic and demographic attributes are then used to forecast household car ownership and mobility using the models in the mobility component.

---

[1] In MIDAS two kinds of transition matrices have been used. The first was obtained from the Dutch National Mobility Panel data set and the second from published CPB transitions.

6. Mobility Component

The MIDAS mobility component consists of a car ownership model, household motorized-trip generation models, a modal split model, car-trip distance models, and transit-trip distance models. All models are formulated for weekly totals. Note that these mobility measures are obtained from the Dutch Panel Survey results in which only those household members of at least 12 years old were requested to report trips, and trips made by individuals below 12 years of age are not reflected in the measures. Consequently the MIDAS mobility component does not reflect trips made by individuals below 12 years old.

6.1. Approach

The effort of this study sought to improve the modal split model developed in the earlier effort which had been found to over-estimate transit trips (Kitamura, 1987,1988). Use of the accessibility measures now available to the project and other attempts, however, did not lead to appreciable improvement of model performance.

The primary reason for this failure is believed to be the infrequent transit trips made by the Panel households. For example, among the 977 households that participated in Waves 3, 5, and 7, and were included in the data base of this model development, the average 1986 (Wave-5) number of train trips is only 0.71 trip per week per household, and the number of BTM (bus, train, and metro) trips is 1.94. On the other hand, the total number of person trips is 55.42, and the number of car trips is 16.84 per week per household. Transit trips thus represent only a small fraction of total person trips, with many households generating no transit trips at all.

Another difficulty is the lack of accurate level-of-service (LOS) measures. Elaborate accessibility measures developed by the Hague Consulting Group (Geinzer and Daly, 1981)are incorporated in the car ownership model of this study. However, accessibility measures are by definition not developed for specific pairs of origin and destination. Only trip-based LOS data that are available are developed by BGC. However, these LOS measures have been developed only for the work trips (for the mode used and an alternative mode) that were made by those individuals who changed job locations during the Panel survey period. Therefore the attributes of the competing modes that serve the origin and destination of a trip in the Panel data file, are in most cases unavailable.

These considerations have led to the decision to develop

1. a trip-end (pre-distribution) modal split model that applies to the total number of motorized trips generated by the household over a one-week period, and

2. a new improved estimation procedure that accounts for the presence of a large number of households whose trips are exclusively by either car or public transit.

The results are presented in Section 6.5. Further details on the estimation method used to obtain the MIDAS modal split model and its performances relative to other estimators can be found in Goulias and Kitamura (1991).

6.2. Car Ownership Model

An ordered-response probit car ownership model is used to determine household car ownership in MIDAS. As in the model developed previously for the Dutch Ministry of

Transport [Kitamura, 1987a,1987b,1988), the model determines the probability that a given household will have no car, one car, or two or more cars.

6.2.1. Model Structure and Estimation Method

The ordered-response probit model probabilistically describes the choice of an alternative from among a set of ordered discrete alternatives. A household's choice of the number of cars to own, falls in this class of choice. The model assumes the presence of a latent variable which cannot be directly measured, but is related to the observed choice--the number of cars owned in this case. Corresponding to a level of car ownership is a range of the latent variable value which is defined by unknown threshold values.

Mathematically, the model can be described as

$$A(i,t) = \alpha' X(i,t) + \theta_1 D_1(i,t-1) + \theta_2 D_2(i,t-1) + \varepsilon(i,t)$$

$$Y(i,t) = \begin{cases} 0, & \text{if } A(i,t) \le q \\ 1, & \text{if } q < A(i,t) \le r \\ 2, & \text{if } r < A(i,t) \end{cases}$$

where i refers to the household and t represents the time (year), and

$A(i,t)$ = latent variable for household i at time t,

$X(i,t)$ = vector of explanatory variables,

$\varepsilon(i,t)$ = random error term,

$Y(i,t)$ = number of cars available,

$D_1(i,t-1)$ = dummy variable that takes on a value 1 if $Y(i,t-1) = 1$;

0 otherwise,

$D_2(i,t-1)$ = dummy variable that takes on a value 1 if $Y(i,t-1) = 2$;

0 otherwise,

q, r = threshold parameters,

α = vector of model coefficients, and

θ_1, θ_2 = scalar coefficients.

The two dummy variables ($D_1(i,t-1)$, $D_2(i,t-1)$) represent the car ownership level in the previous period. The model formulation here thus assumes that the car ownership level at time t is dependent upon the level at time t - 1. The two dummy variables as a set act as a "lagged dependent variable" and characterize the change in car ownership as a history dependent process.

Given a set of explanatory variables (X(i,t)), the objective of model estimation is to determine α, θ_1, θ_2, q, and r. This can be accomplished using the maximum likelihood method. Because of the inclusion of the lagged dependent dummy variables in the model, however, problems arise under the likely condition that the error terms are serially correlated, namely, $\varepsilon(i,t-1)$ and $\varepsilon(i,t)$ are statistically not independent. If this in fact is the case, the coefficient estimates obtained will be inconsistent (i.e., they are biased and the bias cannot be corrected by increasing the sample size. Note that the estimation problems discussed in this section are generic problems and are not due to the characteristics of the specific data set used).

To account for this problem, a correction term, Q[d(t-1)], developed after Heckman, is introduced into A(i,t) before estimating unknown coefficients (see Kitamura, 1987, and Kitamura and Bovy, 1987, for the definition of the correction term). With this correction term, the maximum likelihood estimation yields consistent estimates with lagged dependent variables and serially correlated errors.

This, however, introduces a new problem. The variance of the error term, $\varepsilon(i,t)$, will not be constant but vary across households (heteroskedasticity), making the estimates of test statistics inconsistent. In the model estimation of this study, weights are used to reduce the effect of the heteroskedasticity. The procedure used here, however, is still incomplete and standard error estimates (therefore estimated t-statistics) are believed to be not consistent. Further developmental effort is needed in the future for consistent estimation of discrete choice models with lagged dependent variables and serial correlation (see Kitamura and Bunch, 1989, for related results).

Another problem is that of initial conditions. The model presented above requires observation of car ownership from the preceding time point (t - 1), but obviously, this is not available for the first time point in a panel survey. Furthermore, the development of the correction term to account for serial correlation requires data from three time points to estimate a model of car ownership at time t, because it requires a lagged dependent variable model of car ownership at time t - 1, which must be estimated using data from t - 1 and t - 2. Using the correction term thus developed, another lagged dependent variable model can be estimated for time t, using data from t and t - 1.

These considerations led to the following five stage process for weighted maximum likelihood estimation of the ordered-response probit model with the correction term. This procedure is used in the estimation of the model presented in the next section:

Stage 1. Initial Condition Model: An ordered-response probit model which uses data from one time point, is developed to predict the level of car ownership at time t = 1 as an initial condition.

Stage 2. Lagged Dependent Variable Model: An ordered-response probit model with a lagged-dependent variable is developed using data from t = 0 and 1, and a correction term obtained from the initial condition model of Stage 1.

Stage 3. Weighted Lagged Dependent Variable Model: The lagged dependent variable model of Stage 2 is re-estimated with weights.

Stage 4. Lagged Dependent Variable Model on 3 Time Points: The model of Stage 3 is applied to observations from t = 0 and 1 to obtain Q[d(t-1)]. The model is estimated using data from t = 1, and 2, together with the correction term thus obtained.

Stage 5. Weighted Lagged Dependent Variable Model on 3 Time Points: The model of Stage 4 is re-estimated with weights.

The model coefficients are estimated using the maximum likelihood method in all stages. It is emphasized that the procedure used is tentative in nature; it is believed to offer better coefficient estimates than a naive estimator which ignores serial correlation, but the efficiency and standard-error estimates are yet to be improved.

6.2.2. Estimation Results

The model is estimated using household records of Waves 1, 3, 5, and 7[1] of the Dutch National Mobility Panel survey. The results of this five stage estimation are summarized in Table 6.1 in terms of the coefficients of the latent variable, A(i,t), threshold values, q and r, their t-statistics, and overall goodness-of-fit statistics.

Note that records are "pooled" to increase the sample size. For example, records of all households in the survey are put together to form a sample of 4,101 households to estimate the Stage 1 model for initial conditions. Households in wave pairs (Waves 1 and 3, Waves 3 and 5, and Waves 5 and 7) are pooled to form a data base for Stage 2 and Stage 3 estimation, and those in wave triples (Waves 1, 3, and 5, and Waves 3, 5, and 7) are pooled for Stages 4 and 5. This results in different sample sizes across stages as shown in Table 6.1.

The explanatory variables of the model are: number of drivers, number of workers, household income (INCOME), number of children of over 18 years old (CHILDREN 18+), work trip accessibility difference (dACCESS(work)), and shopping trip accessibility difference (dACCESS(shop)). In order to represent non-linear effect, the number of drivers and the number of workers are each represented by a set of two dummy variables (ONEDRIVER, MULTIDRIVERS; and ONEWORKER, MULTIWORKERS).

The accessibility measures represent the car and transit service level available for residence zones. The accessibility differences used in the model are based on accessibility indices developed by the Hague Consulting Group using a set of destination choice models (Geinzer and Daly, 1981). The difference, (auto accessibility) - (transit accessibility), is

Table 6.1
Five Stage Estimation Results of Ordered-Probit Car Ownership Model

	Stage1		Stage 2		Stage 3		Stage 4		Stage 5	
			$\hat{\theta}$	t	$\hat{\theta}$	t	$\hat{\theta}$	t	$\hat{\theta}$	t
ONECAR(t-1)			2.398	15.13	2.397	15.13	2.815	18.79	2.817	19.17
MULTICARS(t-1)			4.474	15.13	4.472	15.13	5.271	22.59	5.104	22.70
	$\hat{\alpha}$	t	$\hat{\alpha}$	t	$\hat{\alpha}$	t	$\hat{\alpha}$	t	$\hat{\alpha}$	t
ONEDRIVER	2.201	23.54	1.405	9.00	1.405	9.00	1.271	5.40	1.230	5.28
MULTIDRIVERS	2.949	29.47	1.832	10.27	1.831	10.27	1.683	6.77	1.616	6.60
ONEWORKER	0.101	1.83	0.238	3.09	0.238	3.09	0.243	2.11	0.239	2.14
MULTIWORKERS	0.201	2.79	0.273	2.77	0.272	2.77	0.170	1.19	0.173	1.25
INCOME	0.015	8.89	0.006	2.52	0.006	2.52	0.004	1.28	0.005	1.41
CHILDREN 18+	0.351	10.10	0.283	6.05	0.283	6.05	0.214	3.40	0.212	3.50
dACCESS(work)	0.190	2.07	0.087	0.72	0.087	0.72	0.093	0.54	0.111	0.66
dACCESS(shop)	0.208	2.44	0.107	0.96	0.107	0.96	0.020	0.13	0.011	0.07
Q[d(t-1)]			-0.031	-0.37	-0.032	-0.37	-0.414	-5.83	-0.435	-6.56
q	2.582	21.03	2.880	15.95	2.880	15.95	2.922	10.29	3.037	10.35
r	5.326	38.70	6.571	32.07	6.571	32.07	6.709	21.24	6.846	21.00
L(0)	-4505.4		-3951.7		-3951.7		-1986.3		-1986.3	
L(C)	-3570.1		-3107.0		-3107.0		-1520.0		-1520.0	
$L(\hat{\alpha})$	-2304.5		-1167.9		-1167.9		-559.4		-552.1	
$-2[L(0)-L(\hat{\alpha})]$	4401.8		5567.7		5567.7		2853.8		2868.4	
$-2[L(C)-L(\hat{\alpha})]$	2531.2		3878.3		3878.3		1921.3		1935.9	
N	4101		3597		3597		1808		1808	

taken for work and shopping trips, respectively, and used in the estimation. The same values of the accessibility measures are used in all four waves.

The estimation results indicate that the number of drivers in a household is the most important determinant of the number of cars owned by the household. Also very significant is the number of children who are at least 18 years old, and the presence of a worker. All these variables contribute positively to household car ownership.

The income and accessibility variables are significant in the Stage 1 initial condition model and Stage 3 weighted lagged dependent variable model, but not in the final, Stage 5 weighted lagged dependent variable model with serial correlation. The insignificance of the accessibility variables is presumably due in part to the fact that the values of these variables do not change across waves. Despite their statistical insignificance, these variables are included in the car ownership model of MIDAS because of their importance in policy contexts and because the estimation results offer theoretically supportable signs of the coefficients.

The estimation results indicate that the car ownership level of the previous time period (lagged dependent variable expressed by a set of two dummy variables) is extremely significant, and that the serial correlation of the error term is negative as shown by the coefficient of the correction term, Q[d(t-1)]. The same results were obtained in the previous effort (Kitamura, 1987b, 1988) in which a similar ordered-response probit model was estimated without applying weight. Table 6.1 indicates that the results of unweighted Stage 2 and weighted Stage 3 are virtually identical, and that the difference between Stage 4 and Stage 5 is very slight. This estimation thus offers an empirical indication that the effect of the heteroskedasticity introduced by the use of the correction term is very slight.

Therefore the consistency of the standard-error and t-statistic estimates may not have been impaired.

6.3. Dynamic Motorized-Trip Generation Models

A weekly household motorized-trip generation models, developed using data from Waves 1,3,5,7, and 9, are summarized in Table 6.2. Trip generation models are developed separately for households with cars available and those without a car available.

Demographic and socioeconomic variables are the major variables in the model: number of diary keepers (NDIARIES), number of women (NWOMEN), number of men (NMEN), number of workers (NWORKERS), income categories (INCOME2, INCOME3, INCOME4), multi-car ownership (MULTICARS), number of drivers (NDRIVERS), household type (SINGLE, COUPLE, FAMILY, SGLPARENT), and area type (BOV-Large, BOV-Small, RAIL, NORAIL), and a lagged dependent variable (NTRIPS(t-1)).

In the model for car owners, the number of diary keepers, number of workers, multi-car ownership, number of drivers, and household type indicators are the most significant variables. The results also show that households in the highest income group make more trips than do otherwise identical households. A similar set of variables is significant in the model for non-car owners with the differences that the household type indicators are not significant and the income effect is monotonous with households in a higher income group making progressively more trips. Quite notable is the significant coefficient of BOV-Large, indicating that no-car households residing in a large urban area with well developed transit systems tend to make more trips than comparable households in less transit-oriented urban areas. The result is an interesting empirical indication of the effect of transit development upon trip generation of carless households.

Table 6.2
Household Weekly Motorized-Trip generation Models*

	Car Owners		Non-Car Owners	
Variable	$\hat{\beta}$	t	$\hat{\beta}$	t
Constant	-2.87	-1.0	-1.78	-0.6
NDIARIES	4.54	6.6	5.92	7.1
NWOMEN	0.69	0.7	-2.03	-1.9
NWORKERS	2.08	3.4	2.39	2.5
INCOME2	-1.18	-1.2	1.06	0.9
INCOME3	-0.88	-1.1	1.33	1.2
INCOME4	2.85	3.2	5.95	3.3
MULTICARS	8.90	7.9		
NDRIVERS	4.38	5.7	3.82	4.9
SINGLE	7.77	2.7	-0.17	-0.1
COUPLE	11.49	4.7	1.09	0.4
FAMILY	10.77	4.4	-4.27	-1.6
SGLPARENT	7.35	2.7	-0.60	-0.2
BOV-Large	-0.47	-0.2	6.11	3.9
BOV-Small	-2.79	-1.6	2.03	1.4
RAIL	1.75	1.1	2.54	1.4
NORAIL	-2.92	-1.5	2.38	0.9
NTRIPS(t-1)	0.04	1.4	-0.02	-0.7
$\hat{\rho}$	0.44	13.0	0.29	5.1
N	1630		348	
R^2	0.759		0.718	

* Formulated for the weekly total of car-trips as driver, car-trips as passenger, bus, tram, metro, and train trips

INCOME2	1 if annual household income is between dfl 17,000 and dfl 24,000
INCOME3	1 if annual household income is between dfl 24,000 and dfl 36,000
INCOME4	1 if annual household income is more than dfl 36,000

Each motorized-trip generation model is dynamic with a lagged dependent variable and a serially correlated error. Hatanaka's two-stage method (Hatanaka, 1974) is used in the estimation the trip generation models and the trip distance models described in Section 6.5. The software package used is LIMDEP (Greene, 1990).

The coefficient of serial correlation (ρ) is significant in both models, while that of the lagged dependent variable is insignificant. The models thus indicate the presence of positive serial correlation (or "heterogeneity,") but state dependence is not present. This contrasts with the ordered-response probit car ownership model presented in the previous section, which showed significant positive coefficients of the lagged dependent car ownership dummy variables, which imply positive state dependence. The car ownership model also showed a significant negative coefficient of the correction term, indicating the presence of negative serial correlation.

6.4. Modal Split Model

As noted earlier, LOS data are not available to describe trip characteristics by alternative modes that connect given origin and destination zones. Modal split models that can be developed with this limitation are not trip-interchange (post-distribution) models that focus on modal competition at the disaggregate trip level. Before the discussion of the trip-end modal split model used in MIDAS, the rationale behind it is discussed in further detail.

6.4.1. Rationale behind the Modal Split Model

Analysis of travel survey data quite often encounters the problem of limited supply-side information. While the interview survey data offer detailed measurements of household and person characteristics, measurements of urban land development and transportation system characteristics are available only in terms of aggregate zonal averages, or often not available at all. This is almost inevitably the case when a sample is taken from many geographical areas to represent a nation-wide population. Collecting land use information to cover the entire sample area and network data by travel mode for all trip records in the data set, would be too costly, if possible at all. Consequently forecasting models need to be developed using data with limited information, together with whatever supplementary information available.

This applies to the development of modal choice models using the Dutch National Mobility Panel data set. Land use and transportation network data for the 20 municipalities from which the panel sample was initially drawn, are yet to be compiled (the number of municipalities in which the panel respondents resided is said to be over 100 because of migration). The only measures available on the supply-side are a rough indicator of transit service level by municipality, and accessibility measures by mode based on destination choice models developed in an earlier study (Geinzer & Daly, 1981).

Post-distribution modal choice models that focus on modal competition at the trip level, cannot be developed with the Dutch Panel data. This, although more policy sensitive, is not possible because information on the attributes of alternative modes is not available (except for, as noted earlier, a very limited number of work trips made by respondents who changed their residence locations). However, because the data set contains weekly travel information, it presents many travel mode choices repeated by the same household members. These repeated choices may be collectively explained by accessibility or other macroscopic level-of-service indicators.

Furthermore, mode choice may be made considering not each individual trip but a series of linked trips to be made by the individual as a whole. Then the attributes of trips by alternative modes between a given origin and destination pair may not be as influential as one might think. To the contrary, household car ownership, the number of drivers in the household, overall level of transit development, and other socio-demographic attributes may be the major determinants of weekly household modal split. From this viewpoint, the appropriate measure of mode choice is the relative frequency of trips made by a particular mode rather than the mode chosen for each trip. These considerations motivate the modeling effort reported here.

6.4.2. Binomial-Logistic Model of Weekly Household Mode Choice

In the development of a modal split model for weekly household trips, the assumptions are made that: a) choices of modes made over a one-week period can be viewed as repeated binary choices between the automobile (as either a driver or a passenger) and public transit; and b) there exists a fixed probability that governs these repeated binary choices, that is unique to each household. Then, the number of public transit trips a household makes over a one-week period, given the total number of motorized trips, has a binomial distribution,

$$\Pr[K_i = k \mid T_i = t] = \binom{t}{k} P_i^k (1-P_i)^{t-k},$$

$k=0,1,2,3,\ T_i$

where

K_i = the number of transit trips made by household i,

T_i = the total number of motorized trips made by household i, and

P_i = the probability that a randomly chosen trip made by household i will use public transit,

$\binom{t}{k}$ = t!/k!(t - k)! (binomial coefficient),

t = positive integer representing a possible (or observed) total number of motorized trips, and

k = non-negative integer, no greater than t, representing a possible (or observed) number of public transit trips.

Note that k may take on either 0 (0% transit trips) or t (100% transit trips). The problem that arose during the earlier model development effort, i.e., the presence of many households which made no public transit trips at all, or those which made no car trips at all, is no longer a problem here because the probabilistic formulation of weekly household mode choice behavior here accounts for such extreme cases.

Now, let P_i be represented by the following logistic function:

$$P_i = 1/[1 + \exp(-\beta' X_i)]$$

where

X_i = a vector of explanatory variables for household i, and

β = a coefficients vector.

Combining this logistic probability function with the above binomial distribution function leads to the "binomial-logistic model" of weekly mode choice.

The model can be estimated by the maximum likelihood method. It has been shown (Goulias and Kitamura, 1991) that the likelihood function is concave everywhere, and convergence is guaranteed and quick with the Newton-Raphson algorithm. Goulias and Kitamura (1991) have shown that the binomial-logistic model is superior in data replication than the widely used minimum chi-square approaches (Berkson, 1944, 1953; Gart and Zweifel, 1967; Gart, et al., 1985; Haldane, 1955). For details, see Goulias and Kitamura (1991).

6.4.3. Estimation Results

The results of maximum likelihood estimation using a pooled data set consisting of records from Waves 1, 3, 5, and 7 are presented in Table 6.3. As noted earlier, convergence was rapid with Netwon-Raphson algorithm. The model is highly significant with likelihood-ratio chi-square exceeding 6,800 with 17 degrees of freedom.

The number of diary-keepers in the household, number of cars available, number of drivers and level of public transit availability are the major variables that most significantly influence mode choice (a positive coefficient estimate for an explanatory variable implies that households with larger values of that variable tend to have larger probabilities of transit use). In particular, the results indicate that households without a car available (ZEROCAR) and households in a large urban area with a regional transit district (BOV-Large) tend to have higher fractions of public transit trips.

The model's replication capability is excellent, with the correlation coefficient between the predicted probability (P_i) and observed relative frequency (k_i/t_i) exceeding 0.65. The relative error in the average choice probability is within 1%.

One advantage of the binomial logistic model with a constant term is its ability to replicate observed frequency of choices exactly, in this case the number of transit trips (and therefore the number of car trips; see Goulias and Kitamura, 1991). The correlation coefficient between observed and predicted numbers of transit trips is again high,

Table 6.3
Mode Choice Binomial-Logistic Model on Pooled Data

Variable	$\hat{\beta}$	t
Constant	-3.91	-57.4
NRECORDS	0.46	39.1
NWOMEN	0.15	9.6
NWORKERS	0.09	7.4
INCOME2	-0.09	-2.5
INCOME3	0.16	5.0
INCOME4	0.41	11.4
ZEROCAR	3.02	77.2
ONECAR	0.70	23.1
NDRIVERS	-0.40	-28.6
SINGLE	-0.08	-1.6
COUPLE	-0.60	-12.9
FAMILY	-0.74	-16.5
SGLPARENT	-0.32	-6.0
BOV-Large	1.20	51.5
BOV-Small	0.30	11.6
RAIL	0.41	16.6
NORAIL	-0.45	-10.3
L(C)	-37209	
$L(\hat{\beta})$	-22878	
χ^2	28661	
N	6787	
P		
Observed	0.165	
Predicted	0.164	
% Error	-0.7%	
R^2	0.666	
MAE	0.139	
MSE	0.046	
NT		
Observed	3.05	
Predicted	3.05	
% Error	0.0%	
R^2	0.661	
MAE	2.70	
MSE	19.13	

L(C) = Value of Log-likelihood function with constant only
L(β) = Value of Log-likelihood function at convergence
P = Proportion of transit trips
NT = Number of transit trips
MAE = Mean absolute error, average of the absolute difference between observed and estimated value
MSE = Mean square error, average of the squared difference between observed and estimated value

exceeding 0.65. This modal split is used in MIDAS to assign the total motorized trips generation by a household (estimated by the model of Section 6.3) to the automobile and public transit.

6.5. Trip Length Models

Car- and transit-trip length models are developed to predict average trip length (in km) by mode using household attributes. As before, the models are developed for car-owning households and no-car households separately, and each model is formulated with a lagged dependent variable and serially correlated error (Tables 6.4 and 6.5). Essentially the same set of explanatory variables as in the motorized-trip generation models, is used in these models.

The models fit the data not as well as the trip generation models as indicated by the lower R^2's. In the most significant, car-trip length model for car owners, the lagged dependent variable is not significant with a positive coefficient, while serial correlation is significant. The car-trip length model for no-car owners shows a non significant negative coefficient of serial correlation and a significant and positive lagged dependent variable coefficient. The lagged dependent variable is not significant whereas serial correlation is significant for car owners and insignificant for non-car owners in the transit-trip length models.

6.6 Summary

The MIDAS mobility component consists of a car ownership model, household motorized-trip generation models, a modal split model, car-trip distance models, and public transit-trip distance models. All models are formulated for household weekly totals. These mobility

Table 6.4
Car-Trip Length Models

Variable	Car Owners $\hat{\beta}$	Car Owners t	Non-Car Owners $\hat{\beta}$	Non-Car Owners t
Constant	16.18	5.5	6.48	0.5
NRECORDS	1.42	2.0	-5.36	-1.7
NWOMEN	-1.48	-1.5	4.38	1.1
NWORKERS	1.68	2.7	-6.14	-1.7
INCOME2	0.20	0.2	-7.45	-1.6
INCOME3	0.30	0.4	8.76	2.0
INCOME4	0.34	0.4	11.49	1.6
TWOCAR+	0.81	0.1		
NDRIVERS	-1.75	-2.2	-1.33	1.4
SINGLE	1.34	0.5	10.78	1.0
COUPLE	0.76	0.3	9.90	0.9
FAMILY	-0.13	-0.3	24.48	2.3
SGLPARENT	-4.30	-0.5	16.07	1.5
BOV-Large	-2.50	-1.5	0.58	0.1
BOV-Small	-2.90	-1.0	3.16	0.6
RAIL	-1.41	-1.9	7.08	1.2
NORAIL	-0.39	-1.0	9.45	1.2
Car-Trip Length(-1)	0.36	1.1	0.28	2.9
$\hat{\rho}$	0.29	7.3	-0.13	-0.2
N	1625		265	
R^2	0.417		0.333	

Table 6.5
Transit-Trip Length Models

Variable	Car Owners $\hat{\beta}$	Car Owners t	Non-Car Owners $\hat{\beta}$	Non-Car Owners t
Constant	22.84	1.9	16.58	1.2
NRECORDS	-1.23	-0.4	2.06	0.6
NWOMEN	-4.29	-1.2	-4.75	-1.0
NWORKERS	-0.33	-1.0	-1.90	-0.4
INCOME2	-2.81	-0.6	1.50	0.3
INCOME3	-3.05	-0.8	-4.33	-0.8
INCOME4	6.33	1.5	2.58	0.3
TWOCAR+	10.06	1.9		
NDRIVERS	-1.60	-0.5	7.57	1.9
SINGLE	9.76	0.7	12.70	1.0
COUPLE	17.42	1.6	2.74	0.2
FAMILY	18.99	1.9	9.46	0.8
SGLPARENT	29.38	2.5	6.78	0.5
BOV-Large	-18.57	-2.4	-12.52	-1.9
BOV-Small	9.96	1.6	2.27	0.3
RAIL	-9.78	-1.6	-2.03	-0.2
NORAIL	-11.82	-1.6	-14.26	-0.9
Car-Trip Length(-1)	-0.03	-0.5	-1.19	0.3
$\hat{\rho}$	0.20	2.8	0.14	1.5
N	514		239	
R^2	0.283		0.282	

measures are obtained from the Dutch Panel Survey results in which only household members of at least 12 years old were requested to report trips. Therefore, trips made by individuals below 12 years of age are not reflected in the forecasts. An ordered-response probit car ownership model is used to determine the probability that a given household will have no car, one car, or two or more cars. The model is a dynamic model, i.e., it contains two dummy variables indicating past car ownership levels, with correction terms to account for serially correlated errors. The estimation of this model requires a complex procedure composed of five stages[1]. The two motorized trip generation models, one model for car owners and the other for non-car owners, are dynamic with lagged dependent variables and serially correlated errors[2]. Similarly, car-trip length and transit-trip length models are developed to predict average trip length (in km) by mode using household attributes. The models are developed for car-owning households and no-car households separately, and each model is formulated with a lagged dependent variable and serially correlated error. The relative frequency of trips made by transit is modelled through the use of the binomial logistic formulation. The socio-demographic components (in section 5) are integrated with the mobility component (consisting of the models of car ownership, trip generation, modal split, and travel distance by mode) to form a comprehensive simulation system.

---

[1] The estimation of this model was completed before the wave 9 and 10 data set became available. Since the five stage estimation method is time consuming and the added advantage of including Wave 9 in the data set was judged to be marginal, the model estimated with waves 1,3,5, and 7 is used in MIDAS. The appropriateness of this judgement is conformed in the validation section (Section 7).

[2] For the trip generation models, the trip length models, and the modal split model, data formed by pooling waves 1,3,5,7,and 9 have been used.

7. Validation

The Dutch National Mobility Panel Survey spanned over a period of five years (1984-1989) with a total of 10 waves of which 8 involved weekly trip diary surveys (Wissen & Meurs, 1989). Data derived from five odd-numbered waves (1, 3, 5, 7, and 9) have been used to estimate components of MIDAS. Information from the last wave (Wave 10) is now available, offering the opportunity to use it to validate the model components.

7.1. Procedure and Criteria

In this validation exercise, the models in the MIDAS mobility component are used to predict Wave-10 mobility measures using observed explanatory variable values from the Wave-10 data (plus observed mobility measures of Waves 7 and 9 when the model is dynamic). Predictions thus obtained are then compared against observed measures in the Wave-10 data. The MIDAS mobility components are formulated to predict longitudinal changes. Their predictive accuracy is examined against observed longitudinal changes.

This validation method resembles the test of robustness of regression coefficients where a subset of the sample is set aside for validation and not used in estimation. One important difference is that the validation effort of this study is based on longitudinal data; instead of setting aside a subset of behavioral units for validation, a subset of observational time points (Wave 10, in this case) is set aside for validation.

Another unique feature of this validation effort is that forecasts are produced in simulation by generating random variables for the error terms of the respective models. The intent is to validate the models in an environment that is closest to the one they are applied in, i.e., micro-simulation. If applicable, serially correlated errors are generated using the estimated coefficients of serial correlation.

If the models replicate Wave-10 observations well, it offers evidence that the models are capable of providing adequate short-term forecasting by replicating the sample closely. However, it should be recognized that observed Wave-9 and Wave-10 explanatory variable values are used in the validation. The explanatory variables are thus treated as exogenous in this validation effort. (These variables are endogenous in MIDAS, and their future values are internally generated in the simulation. Examining the accuracy of the forecasts of these explanatory variables is beyond the scope of the effort here.)

The following criteria are used in the validation of continuous mobility measures:

- difference between the observed average and the predicted average,
- percent error of the predicted average,
- mean absolute error,
- mean square error, and
- correlation coefficient between the observations and predictions.

Used for the level of household car ownership, which is a discrete mobility measure, are:

- distribution of observed and predicted numbers of cars, and
- percent of cases correctly predicted.

In the rest of this section, the results of validation effort are presented for the car ownership model, motorized-trip generation models, and modal split model.

7.2. Car Ownership Model

The MIDAS car ownership model (Section 6.2) is dynamic with lagged dependent variables, correlated errors, and Heckman correction terms to account for correlated errors (validation of an earlier, simpler car ownership model developed for DVK using the same data base can be found in Kitamura, 1988). This makes forecasting with this model a complex task. Data from three time points are needed to compute the correction terms and lagged dependent variables that are both needed to predict Wave-10 probabilities for the respective levels of car ownership. A random number is then generated to determine a Wave-10 car ownership level. Table 7.1 compares observed and predicted household car ownership levels for Wave 10.

The first part of Table 7.1 presents the average of five simulation runs, and the second the average of 100 runs. In both cases, car ownership levels are correctly forecast for approximately 90% of the sample households. The average number of cars per household is predicted to be 0.922, while the observed Wave-10 average is 0.945. The error is within 2.5%.

There is a slight tendency to under-predict car ownership; the marginal total of multi-car households is under-predicted by 20 (160 vs. 140) and that of no-car households over-predicted by 9 (229 vs. 238) in the first table. The tendency is somewhat lessened in the second table. This result may be a reflection of asymmetry in household behavior dynamics (Clarke, et al., 1982; Goodwin, 1977, 1987; Jones, et al., 1990; Kitamura, 1989; Kitamura and van der Hoorn, 1987). In particular, maintaining a car which has already been acquired involves only marginal maintenance costs; thus changes in contributing factors in the direction of fewer cars may not immediately lead to a disposal of the car. Representation of this possibly asymmetry in car ownership behavior

Table 7.1
Model Validation with Wave-10 Observations: Car Ownership Model

Five Simulation Runs

	Predicted			
Observed	Zero Cars	One Car	Two+ Cars	Total
Zero Cars	217	12	0	229
(%)	17.2	0.9	0.0	18.1
One Car	21	816	39	876
(%)	1.7	64.5	3.1	69.2
Two+ Cars	0	59	101	160
(%)	0.0	4.7	8.0	12.6
Total	238	887	140	1265
	18.8	70.1	11.1	100

% of cases correcly classified = 89.7

One Hundred Simulation Runs

	Predicted			
Observed	Zero Cars	One Car	Two+ Cars	Total
Zero Cars	219	10	0	229
(%)	17.3	0.8	0.0	18.1
One Car	20	820	36	876
(%)	1.6	64.8	2.8	69.2
Two+ Cars	0	51	109	160
(%)	0.0	4.0	8.6	12.6
Total	239	881	145	1265
	18.9	69.6	11.5	100

% of cases correctly classified = 90.7

remains as a future task (initial modeling effort can be found in Kitamura, 1989). Overall accuracy of the car ownership model is well demonstrated in this validation study.

7.3. Motorized-Trip Generation Models

Table 7.2 summarizes the validation results of the motorized-trip generation models reported in Section 6.3. Two models have been formulated, separately for car-owning households and car-less households. The models are also dynamic with lagged dependent variables and serially correlated errors.

Predictions are produced with two different methods: (a) using observed Wave-10 car ownership to classify sample households to car-owning households and car-less households and to exogenously determine the value of the multi-car dummy in the model for car-owning households (MULTICARS; see Table 6.2); and (b) using simulated Wave-10 car ownership levels to classify households and to endogenously determine the value of MULTICARS. The latter method, which more closely represents MIDAS simulation forecasting, is subject to additional errors in household classification (leading to the possibility of applying a wrong model) and in the value of MULTICARS.

The results indicate that the models are performing very well, in particular the one for car-owning households. The larger errors observed for the model for car-less households are presumably due to the fact that the model is based on a much smaller sample.

Quite noteworthy is the result that the endogenous prediction method (b) is producing percent errors in predicted averages that are comparable to the exogenous method (a). The result is encouraging because predicted averages are perhaps the single most frequently used measure in forecasting.

Table 7.2
Model Validation with Wave-10 Observations: Weekly Motorized-Trip Generation Models

Five Simulation Runs

	Car Owners		Non Car Owners	
	(a)	(b)	(a)	(b)
N	1036		229	
Trips Observed	32.1		12.1	
Trips Predicted	32.9	31.2	13.0	13.0
%Error	2.65%	-2.64%	7.51%	7.26%
Mean Abs. Error	9.2	10.6	5.7	5.9
Mean Square Error	134.6	182.6	51.1	58.0
R^2	0.725	0.620	0.648	0.597

One Hundred Simulation Runs

	Car Owners		Non Car Owners	
	(a)	(b)	(a)	(b)
N	1036		229	
Trips Observed	32.1		12.1	
Trips Predicted	32.9	32.7	13.0	13.8
%Error	2.65%	1.81%	7.51%	14.21%
Mean Abs. Error	9.2	9.3	5.7	5.8
Mean Square Error	134.6	138.4	51.1	52.9
R^2	0.725	0.714	0.648	0.654

(a) : Observed car ownership levels are used as input.
(b) : Simulated car ownership levels are used as input
Mean Abs. Error= Mean absolute error, average of the absolute difference between observed and estimated value
Mean Square Error= Mean square error, average of the squared difference between observed and estimated value

Note: The measures shown have been computed using the observed car ownership to classify the households into car owners and non-car owners.

Although the endogenous method is inferior in terms of the other measures in the results with five simulation runs, differences between endogenous and exogenous predictions appear to diminish with one hundred simulation runs. The result is extremely important as it suggests that the accuracy of forecasts using endogenously generated explanatory variables, as is the case for MIDAS, can be improved to levels comparable to forecasts using observed explanatory variables by increasing the number of simulation runs.

The prediction errors tend to be positive, over-predicting Wave-10 trip generation. This may have been caused in part by under-reporting of trips in Wave 10 due to panel fatigue. The effect of panel fatigue on the coefficient estimates needs to be examined in a future effort.

7.4. Modal Split Model

The weekly household modal split model of Section 6.4 is validated similarly through simulation. The analysis here used Wave-10 observed explanatory variable values. The model's performance is evaluated in terms of the fraction of transit trips and the number of transit trips. The Wave-10 observed number of motorized trips is used together with a predicted fraction of transit trips to obtain the latter measure. The results are summarized in Table 7.3.

In each simulation run, the number of transit trips is randomly generated for each household according to a binomial distribution. The two parameters of the binomial distribution are determined as follows: The probability of transit choice predicted by the modal split model (= $1/(1 + \exp(-\beta X_i))$) is used as the probability of a "success," and the observed number of motorized trips is used as the total frequency of choices. The ratio of

Table 7.3
Model Validation with Wave-10 Observations:
Weekly Household Modal Split Model

	Five Simulation Runs		One Hundred Simulation Runs
Proportion of Transit Trips			
Observed		0.140	
Pred (1)	0.146		0.146
%Error	4.65%		4.65%
Pred (2)	0.134		0.144
% Error	4.5%		3.2%
Mean Absolute Error	0.121		0.127
Mean Square Error	0.040		0.038
Correlation	0.637		0.654
Number of Transit Trips			
Observed		2.9	
Predicted	2.7		3.0
% Error	8.4%		0.7%
Mean Absolute Error	2.9		3.0
Mean Square Error	22.7		21.2
Correlation	0.519		0.551

Mean Absolute Error= Mean absolute error, average of the absolute difference between observed and estimated value
Mean Square Error= Mean square error, average of the squared difference between observed and estimated value
Pred(1) = Average of $(1/(1+\exp(-\beta'x))$ across observations.
Pred(2) = Obtained by simulation

the simulated number of transit trips to the total number of trips is used as a simulated transit choice probability.

The model is again performing quite well. Its forecasting errors are in most cases within +5%. The results again show that forecasting accuracy can be improved by increasing the number of simulation runs. In particular, the Wave-10 number of transit trips is very well predicted with 100 simulation runs.

The series of validation analyses presented in this section has indicated that the mobility models perform quite well. In fact the correlation coefficients between observed and predicted Wave-10 mobility measures are quite often as good as those obtained during model estimation; the models are not only replicating observed behavior well but also predicting future (i.e., Wave-10) behavior with comparable accuracy. The analysis of this section lends support to the simulation forecasting reported in Section 9.

7.5 Summary

In this section an overview of the short-term forecasting ability of the mobility component is presented. The dynamic car ownership model, the first model in the sequence of mobility models, is shown to perform very well. It correctly predicts the observed car ownership of Wave 10 in 90.7% of the cases. The mechanized trip generation models are formulated separately for car-owning households and no-car households and are dynamic with lagged dependent variables and serially correlated errors. The validation results indicate that the models are performing very well, in particular the one for car-owning households for which errors are within 3%. The weekly household modal split model is also validated through simulation. The model's performance is evaluated in terms of the fraction of transit trips and the number of transit trips. Again the model is performing quite well. Its forecasting errors are in most cases within +5%. The results show that forecasting accuracy can be further improved by increasing the number of simulation runs. In particular, Wave-10 number of transit trips are very well predicted with 100 simulation runs. This exercise has revealed that these models do not offer the same predictive ability. This is in part due to the fact that some measures of travel behavior are intrinsically more difficult to predict than others. Also it is conceivable that some measures are temporally less stable than others. Accuracy requirements for model components of a dynamic simulation system are not well understood yet. These are areas where improvement is needed in future effort. Overall the analysis of this section supports the simulation forecasting reported in Section 9.

8. Application of MIDAS to Forecasting

As noted earlier, forecasting with MIDAS is different from conventional approaches to forecasting. Fundamental differences lie in the facts that

- MIDAS uses dynamic model components that are based on observation of changes over time,

- MIDAS forecasting is based on micro-analytic simulation in which households "march" and evolve along a simulated time axis, and

- As a result many demographic and socio-economic variables that are typically exogenous, are endogenously determined in MIDAS. Mobility forecasts are based on these endogenously generated household and person attributes in MIDAS.

Most model parameters are estimated using subsamples from the Dutch Panel data set. A subsample of Dutch panel households is also used in the simulation. Observed household and person attributes of 1984, 1985 and 1986 are used as initial conditions; demographic and socioeconomic attributes and mobility levels of these (and internally generated new) households are simulated year by year to 2010 in MIDAS.

In this section, the input parameters to MIDAS are briefly reviewed. Following this, the procedure used to weight the Panel households for forecasting by MIDAS is described. Some adjustments made to some of the MIDAS parameters are then discussed. Attempts to develop micro-simulation models can be found in Mackett (1985) and Miller, et al. (1987). The former is based on cross-sectional data, while the primary concern of the latter is residential mobility.

8.1. MIDAS Input Parameters and Modifiers

The parameters in MIDAS can be classified into two categories. The first category contains the coefficients of the dynamic models in the mobility component, and the income models in the demographic component. These coefficients have estimated using subsamples of the Dutch Panel data set using econometric methods, and embedded in the MIDAS programming code. The second category contains parameters of the demographic components, most of which represent probabilities associated with changes.

Most of the parameters in the demographic component are treated as input data. This is to facilitate their individual adjustment for versatile scenario analysis. The MIDAS software package contains 16 sets of parameters representing probabilities of various demographic and socioeconomic changes. The default values stored in 16 separate MIDAS input files, have been estimated using the Dutch Panel data set. These parameters can be modified to represent a particular scenario of interest (e.g., an increase in women's labor force participation). In maintaining MIDAS, effort will be made to update these parameters as more robust statistics become available. These 16 sets of parameters are summarized in Table 8.1.

In addition, the following input parameters can be used for quick modification of MIDAS default settings:

- RINF modifies annual growth rate in personal income (1.0),
- BFCTR modifies birth probabilities (0.0),
- MEFCTR and FEFCTR modify male and female employment transition probabilities, respectively (0.0),

Table 8.1
MIDAS Demographic Input Parameters*

1 The probability that a woman in a household will give birth in a given year, by employment, number of children in the household, and age of the woman.

2 The probability that a child in a household will leave the household in a given year, by age, sex, and employment.

3 Probability distribution of the age category of a male adult in a household given his spouses age category.

4 Probability distribution of the age category of a female adult in a household given her spouses age category.

5 Probability of employment by age and sex (for new household members)

6 Probability of holding a driver's license, by age and sex (for new household members)

7 Probability distribution of the education category of a male adult, given that of the female adult in a household (for new household members)

8 Probability distribution of the education category of a female adult, given that of the male adult in a household (for new household members)

9 Probability distribution of the number of children in a household, by the age of the female adult

10 Probability distribution of the age of the youngest child by the age of the female adult

11 Joint distribution of the age and sex of the head by household type

12 Probability that an "other" household member is employed, by his/her age

13 Probability distribution of the education category of an "other" member

14 Transition probability of employment by age and sex

15 Transition probability of license holdings by age and sex

16 Probability of death in a given year, by age and sex

*Default values have been estimated using Dutch National Mobility Panel samples

- MLFCTR and FLFCTR modify male and female license holding transition probabilities, respectively (0.0),
- SGFCTR modifies single-to-single household type transition probability (0.0),
- CPFCTR modifies couple-to-couple household type transition probability (0.0),
- FMFCTR modifies family-to-family household type transition probability (0.0), and
- SPFCTR modifies single parent-to-single parent household type transition probability (0.0),

where their default values are shown in parentheses. BFCTR through FLFCTR are applied in MIDAS as follows:

$$P_m = 1/[1 + \exp(-(Z + FCTR))]$$

where

$$Z = \ln(P_o/(1 - P_o))$$

and

P_m = modified probability, and

P_o = original probability.

SGFCTR through SPFCTR are used as additive terms in the logit transition probability models. For details, see MIDAS User's Manual (Goulias,1991).

8.2. Household Weighting

MIDAS simulates the evolution of a subset of those Dutch Panel households that participated in Waves-1, 3, and 5 (many models in MIDAS are dynamic, requiring observations from three time points in the simulation). Because of the initial sampling scheme and attrition, this subset of Panel households does not represent the Dutch population. Because only participants of multiple waves are included in the subsample, it is unlikely that existing weights (BGC, 1984) are applicable here (see Table 8.2). Therefore a new set of weights is developed for this particular subsample using available nationwide statistics.

The weights are developed with the principle of making the distribution of household sizes in the MIDAS subsample agree with the nationwide distribution (although the use of household type is more desirable, no comparable nationwide statistics were available for this analysis). Let **c** be a column vector containing the nationwide household size distribution. Using available statistics for 1985 (CBS, 1988),

$$\mathbf{c}' = (c_1, c_2, ..., c_5)$$
$$= (0.279, 0.293, 0.151, 0.190, 0.087)$$

where

c_i = relative frequency of households with i persons, i = 1, 2, .., 4, and
c_5 = relative frequency of households with 5 or more persons.

Now let **A** be a 5 x 5 matrix of a_{ij}, and **w** be a column vector of w_j, where

a_{ij} = frequency of households of type j with size i in the MIDAS subsample, and

w_j = weight for household type j, j = 1 (singles), 2 (couples), 3 (families), 4 (single parents), and 5 (others).

Then the weights, w_j, can be determined by solving the system of linear equations,

$$Aw = cN$$

where

N = total number of households after the weights are applied,

which can be set as desired. In the MIDAS runs presented in this report, N = 1135 is used, with the weights determined as

Singles	4.6
Couples	1.5
Families	0.95
Single Parents	3.1
Others	1.3

8.3. Comparable Statistics

Often representing the probability of changes and being estimated using a longitudinal data set, many of the MIDAS parameters do not find comparable statistics estimated on larger nationwide samples. The attempts to obtain estimates of household type transitions that are

Table 8.2
Distribution of Household Types and Household Size: Dutch Panel 1984-87, 1982 WBO, and 1982 ORIN

	Dutch National Mobility Panel						
	MIDAS Sample	1984 Wave 1	1985 Wave 3	1986 Wave 5	1987 Wave 7	WBO+ 1982	ORIN+ 1982
Singles	9.4	17.0	15.6	20.1	18.3	7.2	7.3
Couples	23.2	26.1	24.1	24.4	25.6	21.1	16.7
Families	58.4	44.2	49.0	45.3	45.3	71.7	76.0
Single Parents	6.9	8.6	7.7	6.7	6.7		
Others	2.1	4.1	3.6	3.5	4.1		
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

+The category for families includes single parents and others.
Note: CBS statistics for 1985 indicate that the fraction of single-person households is 27.7%, substantially larger than the WBO and ORIN figures.

comparable to the household type definition used in MIDAS were unfortunately unsuccessful.

One notable exception is employment status transition probabilities, which have been tabulated using CBS st tistics by age and sex into a comparable format (Table 8.3). Quite surprisingly, the CBS transition probabilities depict much higher stability in employment than do Panel transition probabilities. By comparing Tables 5.2 and 8.3, it is clear that the diagonal elements of CBS transition matrices are always greater, indicating stability. This is quite contrary to the expectation that transition probabilities estimated using a panel sample tend to be biased to over-represent stability because respondents whose job status changed are more likely to leave the panel. Preliminary MIDAS runs were made using both Panel and CBS transition probabilities, but no appreciable differences were found in the simulation results.

8.4. Summary

Forecasts with MIDAS are produced using dynamic model components that are based on observation of changes over time, and they are based on micro-analytic simulation in which households "march" and evolve along a simulated time axis. As a result many demographic and socio-economic variables that are typically exogenous, are endogenously determined in MIDAS. Mobility forecasts are based on these endogenously generated household and person attributes in MIDAS.

The model parameters used in the dynamic models are estimated using subsamples from the Dutch Panel data set. There are 16 sets of parameters that represent various demographic and socioeconomic changes. These parameters can be modified to represent a particular scenario of interest (e.g., an increase in women's labor force participation).

Table 8.3
Employment Status Transition Matrix by Age and Sex Based on CBS Arbeidskrachtentellingen, 1977, 1979, 1981 and 1983

Age		Men		Women	
		Employed	Not Employed	Employed	Not Employed
18 - 24	Employed	.979	.021	.926	.074
	Not Employed	.136	.864	.135	.865
25 - 34	Employed	.994	.006	.971	.029
	Not Employed	.251	.749	.033	.967
35 - 44	Employed	.994	.006	.971	.029
	Not Employed	.251	.749	.033	.967
45 - 54	Employed	.969	.031	.937	.063
	Not Employed	.134	.866	.003	.997
55 - 64	Employed	.943	.057	.903	.097
	Not Employed	.017	.983	.001	.999
>65	Employed	.585	.415	.665	.335
	Not Employed	.000	1.000	.000	1.000
Total	Employed	.977	.023	.913	.087
	Not Employed	.082	.918	.036	.964

Note: Average of transition probabilities observed between 1976-77, 1978-79,1980-81, and 1982-83.

A subsample of Dutch panel households is used in the simulation.This subsample does not represent the Dutch population because of the initial sampling scheme and attrition. Since only participants of multiple waves (stayers) are included in the subsample, it is unlikely that existing weights are applicable in this case. Therefore a new set of weights is developed for this particular subsample using available nationwide statistics (CBS, 1988).

Employment transition probabilities have been tabulated by age and sex using CBS statistics in a format comparable to the format in MIDAS. Contrary to the expectation that transition probabilities estimated using a panel sample tend to be biased to over-represent stability, the CBS transition probabilities depict much higher stability in employment than do Panel transition probabilities[1]. Preliminary MIDAS runs were made using both Panel and CBS transition probabilities, but no appreciable differences were found in the simulation results. The results presented in this report are obtained using the CBS employment transition probabilities.

[1] Note that a similar finding is reported in Section 3 for the household type transition probabilities.

9. Simulation Experiment

The evolution of household demographics and socio-economics, car ownership, and mobility, is simulated with MIDAS using the expanded Panel household sample described in the previous section. A simulation period of 25 years is used that starts in 1986, when the Wave 5 survey was conducted, and ends in 2010. One year is used as the time increment in the simulation. Therefore the characteristics of each sample household is updated 25 times in the simulation. All simulation runs reported here assume a retention rate of 15%, i.e., 15% of households newly generated are kept in the simulation (see section 5.8). The deletion of the remaining households may interpreted to represent out-migration from the study area.

Before proceeding, it is repeated that many variables that are exogenously determined and given as input parameters to other forecasting models, are endogenous to MIDAS and are internally generated during simulation. Because of this, it is not possible to exactly match the socio-economic and demographic factors in MIDAS to those in other forecasting scenarios. It was viewed more meaningful to use MIDAS to observe the socio-economic and demographic evolution in The Netherlands. The mobility forecasts thus obtained can then be compared with existing forecasts with the understanding that the principle of forecasting is fundamentally different in MIDAS.

The tenet of MIDAS has been to extract salient longitudinal relationships in the Dutch National Mobility Panel data, and extend them into the future. One of the fundamental tasks charged to this project is to examine whether such dynamic forecasting is practical and meaningful at all. Because of this, manipulation of the MIDAS parameters that have been estimated using the Dutch Panel data is kept to the minimum; only a parameter to control income growth (RINF) is manipulated in the simulation exercise reported here to represent CPB "referentie," "optimistisch" and "pessimistisch" growth scenarios. As noted earlier, however, many parameters are built in MIDAS to enable a wide range of scenario analysis.

In this section, the results of a "baseline" MIDAS run are first presented. The baseline run assumes an income growth rate similar to the CPB "referentie" scenario. The baseline results are then compared in Section 9.2 to the CPB "referentie" socio-economic and demographic scenario, observed OVG mobility measures, car ownership forecasts by van den Broecke, and mobility forecasts by the national model. MIDAS forecasts with high and low income growth scenarios are presented in Section 9.3. Finally in Section 9.4, short-term and long-term aggregate income elasticities of car ownership and selected mobility measures are presented.

9.1. Baseline MIDAS Forecasts

The MIDAS baseline forecast represents an income growth of 72% by year 2010 (closely approximating the CPB referentie growth rate of 65%). The results are presented in Table 9.1 for year 1986 (base year), 1995, 2000, 2005 and 2010. All MIDAS results presented in this section are averages of five simulation runs repeated for each simulation case using different seeds for random number generation.

Simulation results are given for household size, labor force participation, license holding, automobile ownership, and for five mobility measures: number of motorized trips, number of car trips, number of transit trips, driver vehicle-kilometers, and transit passenger-kilometers. The evolution of these demographic, socioeconomic, and mobility measures through the 25 year simulation period is shown in Figures 9.1. All mobility measures are

Table 9.1
Baseline MIDAS Forecasts: 1986 - 2010

	Base Year 1986	MIDAS Forecasts 1995	2000	2005	2010	Growth
Population (x 10^6)+	14.5				15.1	4.1%
Population, ≥ 12 Years Old (x 10^6)++	12.3				13.0	6.5%
Household Size	2.53	2.31	2.16	2.02	1.92	-24.3%
Labor Force Participation*	50.2%	51.7%	49.3%	46.3%	42.6%	
Average Income per Employed Person	100	138	146	158	172	
Number of Licensed Drivers (x 10^6)**	7.45				10.34	38.8%
Percent of Licensed Drivers	51.4%	59.5%	63.4%	66.5%	68.5%	
Number of Automobiles (x 10^6)**	5.03				7.19	43.0%
Automobiles per Person	0.35	0.40	0.43	0.46	0.48	37.3%
Automobiles per Household	0.88	0.92	0.92	0.93	0.91	4.0%
Automobiles per Driver	0.68	0.67	0.67	0.69	0.70	3.0%
Number of Motorized Trips per Week						
Per Person	9.68	11.76	12.29	12.69	13.00	34.3%
National Total (x 10^6)**	118.6				169.6	43.0%
Number of Car Trips per Week						
Per Person	8.67	10.60	11.05	11.46	11.79	36.0%
National Total (x 10^6)**	106.3				153.8	44.8%
Number of Transit Trips per Week						
Per Person	1.01	1.16	1.24	1.23	1.21	20.1%
National Total (x 10^6)**	12.3				15.7	27.9%
Vehicle-Kilometers Driven per Week						
Per Person	92.4	135.3	134.1	146.0	151.2	63.6%
National Total (x 10^6)**	1132				1972	74.2%
Transit-Passenger Kilometers Trips per Week						
Per Person	24.1	37.7	41.0	41.7	41.2	70.9%
National Total	296				538	81.9%

+CPB "referentie" scenario.

++Van den Broecke Social Research (1987a, Deel I, p.3, Deel IV, Table 1). The 2010 figure adjusted to agree the CPB forecast.

*Among individuals of 15 years old and over (CPB), or 18 years old and over (MIDAS).

**MIDAS forecasts are expended using the national population (of individulas of 12 years old and over for mobility measures).

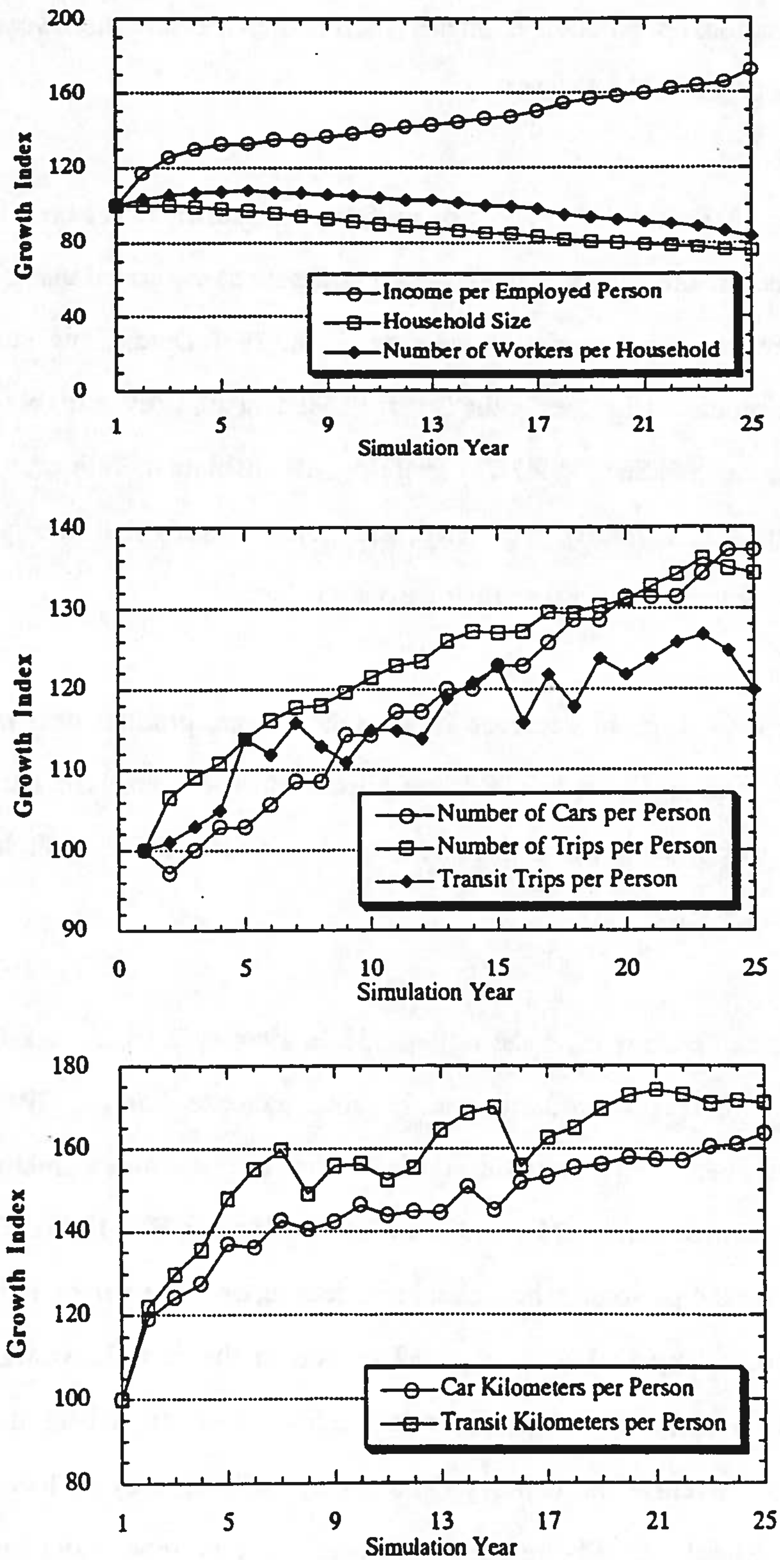


Figure 9.1
MIDAS Referentie Scenario Forecasts

weekly totals including trips made on weekend. Nationwide figures are developed by multiplying national population estimates (used in the CPB referentie scenario) to the per-capita figures generated by MIDAS.

In presenting these absolute values of mobility measures, it is noted that the MIDAS mobility forecasts are based on the mobility component estimated using the Dutch Panel data together with base-year trip rates reported in the 1986 Dutch Panel survey. It has been reported that reported trip rates in the Dutch Panel data are subject to biases due to under-reporting of trips (See Section 9.2.2). Therefore the absolute mobility forecasts reported in this section must be carefully interpreted, and more emphasis should be placed on relative changes in their values rather then their absolute values.

The results show a rapid decrease in household size, gradual decline in labor force participation, and increases in the driver population and household car ownership. All mobility figures show substantial increases, in particular driver vehicle-kilometers and transit passenger-kilometers.

Household size declines from the initial 2.53 in 1986 to 2.16 in year 2000, and 1.92 in year 2010. This represents a much more rapid decrease than the CPB scenario (2.3 in 2010). However, CBS statistics indicate that the average number of persons per household declined from 2.95 in 1975 to 2.54 in 1985 (CBS, 1988). This represents a decline of over 0.4 person per household in a decade, or 0.041 person per year. The above decrease forecast by MIDAS, i.e., 0.37 person in the first 15 years and 0.24 in the following 10 years, may in fact accurately reflect the observed trend. The continuing decline in the twenty-first century depicted by MIDAS may reflect the aging of the population. Should there be reasons to believe that this trend may change in the future, then MIDAS is capable of generating forecasts reflecting such changes.

9.2. Comparison with Other Forecasts

The baseline MIDAS forecasts are compared to available forecasts (or scenarios) and observed mobility measures in this section. They include: CPB referentie socio-economic and demographic scenario, observed OVG mobility measures, driver license holdings and car ownership forecasts by van den Broecke, and mobility forecasts by the national model.

9.2.1. CPB Referentie Scenario

A comparison of the MIDAS baseline forecasts and the CPB referentie scenario is given in Table 9.2. Some discrepancies in the base-year figures are presumably due to the uniqueness of the Panel subsample used in MIDAS. As noted in Section 8.2, weighting based on household size was applied to the Panel sample, and led to the MIDAS base-year average household size of 2.53, which closely approximates the CPB average of 2.59.

As discussed earlier, MIDAS results show much faster decline in household size. MIDAS also forecasts faster increase in the driver population in terms of the percentage of licensed drivers among the nationwide population. The discrepancy between the MIDAS and CPB results is much smaller for the percentage in the adult population (individuals of 18 years old and over). The MIDAS results, which use Panel-based transition probabilities of license holding, show a similar probability that an adult individual will be holding a driver's license in 2010. The apparent discrepancies in the number of licensed drivers and their percentage in the entire population, are therefore caused by the difference in the 2010 age distribution in the CPB scenario and that simulated by MIDAS. MIDAS depicts much more rapid aging of the Dutch population.

Table 9.2
Comparison of MIDAS Forecasts and CPB Referentie Scenario

	CPB+		MIDAS	
	1986	2010	1986	2010
Population (x 10^6)	14.5	15.1		
Population, ≥ 12 Years Old (x 10^6)++	12.3	13.0		
Household Size	2.59	2.29	2.53	1.92
Labor Force Participation*	41%	46%	50.2%	42.6%
Average Income per Employed Person	100	165	100	172
Number of Licensed Drivers (x 10^6)**	7.11	9.30	7.45	10.34
Percent of Licensed Drivers in Population	49.0%	61.6%	51.4%	68.5%
Percent among 18 Years Old and Over	66.1%	72.6%	71.2%	73.3%
Number of Automobiles (x 10^6)**	4.54	7.90	5.03	7.19
Automobiles per Person	0.31	0.52	0.35	0.48
Automobiles per Household	0.81	1.20	0.88	0.91

+CPB "referentie" scenario.
++Van den Broecke Social Research (1987a, Deel I, p.3, Deel IV, Table 1).
The 2010 figure adjusted to agree the CPB forecast.
*Among individuals of 15 years old and over (CPB), or 18 years old and over (MIDAS).
**MIDAS forecasts are expanded using the national population.

MIDAS forecasts the number of cars per person in 2010 to be 0.48. This is slightly smaller than the value, 0.52, in the CPB referentie scenario. The number of cars per household forecast by MIDAS is smaller than the CPB value by 24%, largely because of the decline in household size in the MIDAS simulation. The MIDAS forecast of nationwide fleet size is approximately 9% smaller than the CPB value.

In comparing these results, it should be noted that MIDAS utilizes an elaborate dynamic household car ownership model. Although the assumptions underlying the CPB car ownership scenario are not known to the project team (the scenario appears to be after van den Broecke; see Section 9.2.3), it is conceivable that the national 2010 fleet size of 7.90 million vehicles was obtained by extrapolating observed trends, rather than building a model of household car ownership. Car ownership forecasts in MIDAS, on the other hands, reflect changes in the number of drivers, number of workers, number of driving-age children, and household income at the disaggregate household level.

Another important difference is the decline in labor force participation shown by MIDAS. Earlier in the study, two sets of preliminary simulation runs were conducted, one using the employment status transition probabilities estimated using the Panel sample, and the other using those estimates by CBS. These runs did not show any appreciable difference in the results. The decline in labor force participation shown by MIDAS, therefore, cannot be attributed to incorrect parameter values for individual labor force participation. It is conceivable that the CPB forecast of an increase from 41% of the adult population (15 years old and over) in 1986 to 46% in 2010, is based on some assumption that is not shared by MIDAS, e.g., increased labor force participation by women. Again, the philosophy behind the MIDAS results reported here is to use the Panel-based parameter values without modification. Should forecasts based on such assumptions be desired, MIDAS is capable of incorporating them.

9.2.2. Observed 1982 OVG Mobility Measures

Dutch national mobility statistics available in Moning (1983) are used to examine the representativeness of the Panel sample used in MIDAS simulation forecasting. The results are summarized in Table 9.3. Although the survey years are not exactly comparable (the OVG trip rates are from 1981, travel distances from 1982; the base year MIDAS mobility statistics are 1986 observations), the two sets of mobility measures are quite comparable, in particular travel distance measures.

The MIDAS base-year trip rates are consistently below the 1981 OVG trip rates (it is believed that the OVG mobility measures are averages over all days of the week including Saturdays and Sundays). For example, the motorized-trip rate is 17% below the comparable OVG trip rate. This may be due to the well documented trip under-reporting in the Dutch Panel survey (Golob and Meurs, 1986; Meurs, van Wissen and Visser, 1989). As noted earlier, this should be kept in mind in interpreting the mobility forecasts provided by MIDAS.

9.2.3. Van den Broecke's Forecasts

Based on his innovative cohort model, van den Broecke produced driver's license holdings and car ownership forecasts for the Netherlands (van den Broecke, 1987a, 1987b, 1988). His forecasts are compared with MIDAS forecasts in Table 9.4. The driver population and the national car ownership forecasts by van den Broecke are identical to those in the CPB referentie scenario; apparently the former have been incorporated in the latter. The differences between the MIDAS license holding and car ownership forecasts and the

Table 9.3
Comparison of MIDAS Base-Year Mobility Measures with 1982 OVG Observations

	OVG* 1978	OVG* 1982	MIDAS 1986
Population (x 10^6)+			14.5
Population, ≥ 12 Years Old (x 10^6)++			12.3
Number of Motorized Trips per Week			
Per Person		11.69	9.68
National Total (x 10^6)**			118.6
Number of Car Trips per Week			
Per Person		10.57	8.67
National Total (x 10^6)**			106.3
Number of Transit Trips per Week			
Per Person		1.12	1.01
National Total (x 10^6)**			12.3
Vehicle-Kilometers Driven per Week			
Per Person	93.8	93.8	92.4
National Total (x 10^6)**			1132
Transit Passenger-Kilometers Trips per Week			
Per Person	20.3	21.7	24.1
National Total (x 10^6)**			296

*Moning (1983). The OVG trip rates are for 1981.
+CPB "referentie" scenario.
++Van den Broeckė Social Research (1987a, Deel I, p.3, Deel IV, Table 1).
The 2010 figure adjusted to agree the CPB forecast.
**MIDAS forecasts are expanded using the national population of individulas of 12 years old and over.

Table 9.4
Comparison of MIDAS Forecasts with van den Broecke's License and Car Ownership Forecasts

	VDB[1]		MIDAS	
	1985	2010	1986	2010
Population (x 10^6)+			14.5	15.1
Population, ≥ 12 Years Old (x 10^6)++			12.3	13.0
Labor Force Participation*	31%	38%	36.2%	39.8%
Average Income per Employed Person	100	170	100	172
Number of Licensed Drivers (x 10^6)**	6.90	9.30	7.45	10.34
Percent of Licensed Drivers	48.0%	61.0%	51.4%	68.5%
Number of Automobiles (x 10^6)**	4.50	7.90	5.03	7.19
Automobiles per Person	0.31	0.52	0.35	0.48
Automobiles per Household			0.88	0.91

[1]Van den Broecke/Social Research (1987b), "Middenscenario"
+CPB "referentie" scenario.
++Van den Broecke Social Research (1987a, Deel I, p.3, Deel IV, Table 1). The 2010 figure adjusted to agree the CPB forecast.
*Percentage of employed persons in the total population.
**MIDAS forecasts are expanded using the national population.

corresponding CPB values have been discussed in Section 9.2.1.

Quite interestingly, a good agreement exists between van den Broecke and MIDAS in the 2010 labor force participation forecasts (represented here as the percentage of employed persons in the total population). MIDAS assumes practically the same income growth rate as in van den Broecke. Considering the fundamental differences in data and methodology, the compatibility between the van den Broecke forecasts and MIDAS results, including driver's license and car ownership, is quite striking.

9.2.4. National Model

The Dutch National Model provides the only mobility forecasts available to this study (Vrolijk, Gunn and van der Hoorn, 1987; Gunn, van der Hoorn and Daly, 1987). The results are summarized in Table 9.5 along with MIDAS forecasts (the National Model forecasts used here assume changes in demographic factors, driver's license holding and car ownership, but no changes in accessibility and congestion). The differences in household size and labor force participation are similar to those seen earlier.

Quite notably, the 2010 driver's license holding in the National Model forecasts is practically identical to the forecast by MIDAS. Driver's license holdings are forecast in National Model using a set of discrete choice models formulated at the household level. Thus it is not a simple extrapolation of observed trends. MIDAS forecasts are based on transition probabilities of license ownership, while van den Broecke's forecast relies on license ownership probabilities assumed for respective population age cohorts. It is noteworthy that these three, entirely different forecasting methods have produced 2010 driver population forecasts that are within 12% of each other.

Table 9.5
Comparison of MIDAS Forecasts with National Model Forecasts

	National Model# 1986	National Model# 2010	MIDAS 1986	MIDAS 2010	Growth
Population (x 10^6)	14.3	15.1	14.5	15.1	
Population, ≥ 12 Years Old (x 10^6)++			12.3	13.0	
Household Size	2.70	2.29	2.53	1.92	
Total Workforce (x 10^6)	4.6	6.1			
Labor Force Participation*	39.2%	48.5%	50.2%	42.6%	
Number of Licensed Drivers (x 10^6)**	6.6	10.4	7.45	10.34	
Percent of Licensed Drivers	46.2%	68.9%	51.4%	68.5%	
Number of Automobiles (x 10^6)**	4.3	7.9	5.03	7.19	
Automobiles per Person	0.30	0.52	0.35	0.48	
Automobiles per Household	0.81	1.20	0.88	0.91	
Change in Weekday Vehicle-Kilometers[2]	+72%				
Vehicle-Kilometers Driven per Week					
Per Person			92.4	151.2	63.6%
National Total (x 10^6)**			1132	1972	74.2%
Change in Weekday BMT Passenger-Kilometers[2]	-7%				
Change in Weekday Rail Passenger-Kilometers[2]	-2%				
Transit Passenger-Kilometers per Week					
Per Person			24.1	41.2	70.9%
National Total (x 10^6)**			296	538	81.9%

#Vrolijk, Gunn and van der Hoorn (1987), Gunn, van der Hoorn and Daly (1987)
++Van den Broecke Social Research (1987a, Deel I, p.3, Deel IV, Table 1).
The 2010 figure adjusted to agree the CPB forecast.
*Among individuals of 15 years old and over (CPB), or 18 years old and over (MIDAS).
**MIDAS forecasts are expanded using the national population (of individulas of 12 years old and over for mobility measures).
[1]Estimated using the total population and the number of households used in the National Model study.
[2]Read from a graph in Gunn, van der Hoorn and Daly (1987).

Vehicle-kilometrage forecasts are again strikingly similar between MIDAS and the National Model. The National Model's forecasts an increase of 72% by year 2010. The corresponding MIDAS forecast is a virtually identical increase of 74.2%.

The forecasts of public transit use are drastically different between the two, however. The National Model predict a slight decrease in public transit passenger-kilometers by 2010. MIDAS, on the other hand, forecasts an increase of over 80%. No changes in accessibility and levels-of-service are assumed in both forecasts.

This discrepancy in public transit use between MIDAS and the National Model is perhaps the single most important discrepancy. Unfortunately there is no other comparable forecast available to this study to infer which forecast depicts a more likely picture of the future. Both forecasts are based on elaborate model systems formulated at the household level. One important difference is that the National Model is formulated using cross-sectional data and longitudinal changes in population compositions are represented by weighting households. MIDAS, on the other hand, is based on longitudinal data and simulates household evolution over time. Investigation of the relative advantages and disadvantages of these two distinct model systems for long-range forecasting appears to be well warranted.

9.3. Scenario Analysis: Income Growth Effect

As an example to show how MIDAS can be used for scenario analysis, two additional simulation runs were made to produce the income growth in the CPB "optimistisch" scenario and the one in the "pessimistisch" scenario. The results are summarized in Table 9.6 and Figures 9.2a and 9.2b.

Table 9.6
Comparison of Three Income Growth Scenarios

	Base Year 1986	MIDAS Forecasts for 2010 by Income Growth Rate					
		Low		Middle		High	
Population (x 10^6)+	14.5	15.1		15.1		15.1	
Population, ≥ 12 Years Old (x 10^6)++	12.3	13.0		13.0		13.0	
Household Size	2.53	1.89	-25%	1.92	-24%	1.90	-25%
Labor Force Participation*	50.2%	42.3%		42.6%		42.2%	
Average Income per Employed Person	100	141		172		220	
Number of Licensed Drivers (x 10^6)**	7.45	10.31	38%	10.34	39%	10.34	39%
Percent of Licensed Drivers	51.4%	68.3%		68.5%		68.5%	
Number of Automobiles (x 10^6)**	5.03	6.86	36%	7.19	43%	7.57	51%
Automobiles per Person	0.35	0.45		0.48		0.50	
Automobiles per Household	0.88	0.86		0.91		0.95	
Automobiles per Driver	0.68	0.67		0.70		0.73	
Number of Motorized Trips per Week							
Per Person	9.68	12.50	29%	13.00	34%	13.51	40%
National Total (x 10^6)**	118.6	163.0	37%	169.6	43%	176.3	49%
Number of Car Trips per Week							
Per Person	8.67	11.28	30%	11.79	36%	12.26	41%
National Total (x 10^6)**	106.3	147.1	38%	153.8	45%	159.9	50%
Number of Transit Trips per Week							
Per Person	1.01	1.22	21%	1.21	20%	1.25	25%
National Total (x 10^6)**	12.3	15.9	29%	15.7	28%	16.3	33%
Vehicle-Kilometers Driven per Week							
Per Person	92.4	142.3	54%	151.2	64%	156.4	69%
National Total (x 10^6)**	1132	1856	64%	1972	74%	2040	80%
Transit Passenger-Kilometers Trips per Week							
Per Person	24.1	39.2	63%	41.2	71%	42.5	76%
National Total	296	512	73%	538	82%	554	87%

+CPB scenario.

++Van den Broecke Social Research (1987a, Deel I, p.3, Deel IV, Table 1). The 2010 figure adjusted to agree the CPB forecast.

*Among individuals of 15 years old and over (CPB), or 18 years old and over (MIDAS).

**MIDAS forecasts are expanded using the national population (of individulas of 12 years old and over for mobility measures).

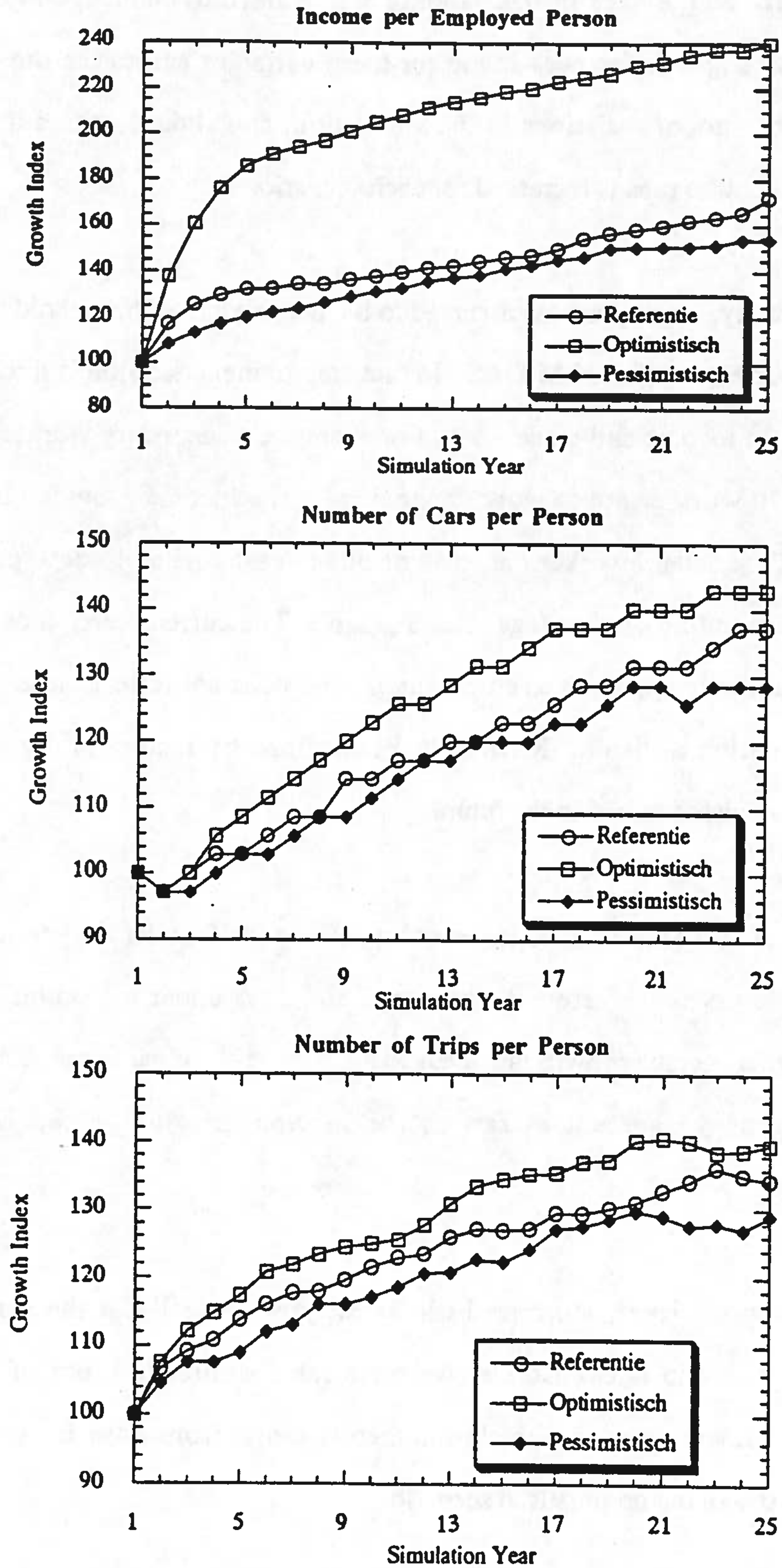


Figure 9.2a
Comparison of Income Growth Scenarios

Household size and driver's license holding are assumed to be independent of income in MIDAS. The slight differences found for these variables across the three scenarios are purely due to random variations in the simulation, and should vanish if the number of repetitive simulation runs is increased for each scenario.

More importantly, employment is assumed to be independent of household income or wage rates in the current version of MIDAS. In fact employment decision is likely to depend on both household income and wage rates. For example, a secondary worker in a household may choose to work or not to work depending on the income provided by the principal worker. The secondary worker may also decide to participate or not to participate in the labor force depending on the wage rates available. The current version of MIDAS uses a simple probabilistic approach to employment, and does not reflect these aspects of labor force participation decision. MIDAS can be refined by incorporating a more realistic employment decision model in the future.

The number of cars is forecast to increase by 36% by 2010 under the pessimistisch growth scenario, 43% under the referentie scenario, and 51% under the optimistisch scenario. Clearly a higher income growth rate leads to a higher level of car ownerships, but these car ownership growths are not as fast as the income growths (41%, 72% and 120%, respectively).

At the nationwide level, motorized-trip rates grow virtually at the same pace as car ownership. Car trip rates also show growth rates similar to those of car ownership. Nationwide growth rates in vehicle-kilometers range from 64% for the pessimistisch scenario to 80% in the optimistisch scenario.

Public transit trip rates, on the other hand, show more complex growth patterns; the trip rate is slightly higher for the pessimistisch scenario (29%) than for the referentie scenario (28%), while it is the highest for the optimistisch scenario (33%). Apparently two forces are at work. When income growth is slow, the lower car ownership level leads to more frequent use of public transit. On the other hand, higher income implies more long distance rail trips, leading to the highest transit trip rate with the optimistisch scenario (the analysis of the Dutch Panel data set in Section 6.4.3 has shown that transit use is positively correlated with income). Transit passenger-kilometrage, on the other hand, grow with income.

The scenario analysis of this section indicates that transit use will increase under any income growth scenario examined here. Income growth and accompanying car ownership growth are not detrimental to transit use. In fact the analysis here shows that both transit trip rates and passenger-kilometrage grow fastest under the fastest income growth scenario.

9.4. Long-Term and Short-Term Elasticities

One of the advantages of simulation forecasting based on dynamic models is its ability to replicate the dynamics inherent in the process under investigation. Changes in car ownership or mobility may not immediately follow a change in a contributing factor, but may involve response time lags. If that is the case, then the impact of a change in a contributing factor cannot be assessed in its entirety until delayed responses take place and their repercussions are complete. With the dynamic models with lagged dependent variables capturing such effects, MIDAS is a dynamic simulation model system ideal for the analysis of behavioral processes.

In this section, the year-to-year changes in mobility measures produced by MIDAS are used to evaluate possible differences in short-term and long-term growth rates. The intent of the analysis is to shed light on the effect, on the value of an elasticity measure, of the time interval with which it is measured. Aggregate elasticities are used as measures of behavioral change. (It must be noted that an aggregate elasticity is not necessarily identical to the elasticity that would be observed for each individual behavioral unit. Positive changes and negative changes cancel each other out as an average measure of change is computed. This information loss in aggregation would lead to aggregate elasticities that tend to under-represent behavioral sensitivity. If study objectives warrant, the MIDAS is capable of producing such disaggregate elasticities.)

Suppose it is desired to evaluate the elasticity of Y with respect to X, i.e., how responsive Y is to a change in X. A measure of the elasticity may be defined as

$$E = (dY/Y)/(dX/X),$$

namely, the ratio of the relative change in Y to the relative change in X. However, if Y does not change instantaneously after X changes, or X changes gradually over time, then the above elasticity measure depends on the time interval, Δ, used to measure dX and dY as well as the time, t, when X and Y are measured. Accordingly E becomes a function of Δ and t, as

$$E(t,\Delta) = \{(Y(t+\Delta) - Y(t))/(X(t+\Delta) - X(t))\} \{X(t)/Y(t)\}$$

Note that dY is replaced by the difference, $Y(t+\Delta) - Y(t)$, and dX by $X(t+\Delta) - X(T)$. If Δ is relatively long (say, a decade), then the resulting $E(t,\Delta)$ may be called a "long-term elasticity," while if Δ is short (a few months or a year), then it may be called a "short-term elasticity." Suppose Y changes instantaneously as X changes, while X changes continuously with time, t. Further let,

$$Y(t) = F(X(t)).$$

Then an elasticity measure using interval Δ is

$$E(t,\Delta) = \{(F(X(t+\Delta)) - F(X(t))/(X(t+\Delta) - X(t))\}\{(X(t)/Y(t)\}.$$

Now, suppose X changes monotonously over time, i.e.,

$$X(t) < X(s) \quad \text{or} \quad X(t) > X(s) \quad \text{for any } t < s.$$

Namely X either increases or decreases over time. Then, for a given t,

$$\{F(X(t+\Delta)) - F(X(t))\}/(X(t+\Delta) - X(t)) \text{ decreases with } \Delta \text{ if } d^2F/dX^2 < 0,$$

$$\{F(X(t+\Delta)) - F(X(t))\}/(X(t+\Delta) - X(t)) \text{ increases with } \Delta \text{ if } d^2F/dX^2 > 0,$$

and

$$\{F(X(t+\Delta)) - F(X(t))\}/(X(t+\Delta) - X(t)) \text{ is constant if } d^2F/dX^2 = 0.$$

Then assuming that $X(t)/Y(t) > 0$ for a given t,

$$E(t,\Delta) \text{ decreases with } \Delta \text{ if } d^2F/dX^2 < 0,$$

$$E(t,\Delta) \text{ increases with } \Delta \text{ if } d^2F/dX^2 > 0, \text{ and}$$

$$E(t,\Delta) \text{ is independent of } \Delta \text{ if } d^2F/dX^2 = 0.$$

Thus the relationship between long-term and short-term elasticities depends on the functional relationship between X and Y. The true relationship between X and Y may not be contemporaneous and X may change in non-monotonous manner. Thus the simplified discussion here may not be immediately applicable to represent observed behavioral dynamics. Nonetheless, the discussion shows that functional relationship between X and Y is one of the factors that influence long-term and short-term elasticities.

Aggregate elasticities of car ownership, trip rate, and vehicle-kilometers with respect to income, are evaluated using $\Delta = 1, 2, 4, 6,$ and 10 years. The results for Year 5 through Year 25 of the simulation with the referentie scenario are used in this computation. The averages and standard deviations of the elasticities are summarized in Table 9.7.

The results for car ownership and vehicle-kilometrage offer clear indications that their aggregate elasticities decline as the time interval for evaluation, Δ, increases. If the above assumption of contemporaneous relations between average car ownership and average income can be accepted, then the result implies that average car ownership is a concave function of average income; the marginal increase in average car ownership gradually declines as average income continues to increase. Similar results can be observed for vehicle kilometrage. The results for trip rate, on the other hand, suggest that average trip rate is linearly related to average income; if the latter continues to increase, the former would also continue to increase at the same rate as the former.

The analysis of this section has shown that MIDAS is capable of generating simulation output that would enable analyses that would otherwise be difficult to perform. Although the analysis here used aggregate elasticities, MIDAS output can be easily generated such that elasticity measures can be evaluated at the household level.

Table 9.7
Aggregate Income Elasticities of Car Ownership, Trip Rate, and Vehicle-Kilometrage by Time Interval for Evaluation (Δ)

Car Ownership

	Δ = 1	Δ = 2	Δ = 4	Δ = 6	Δ = 10
Average Elasticity	1.41	1.25	1.05	1.02	1.00
Standard Deviation	1.85	1.04	.26	.18	.13
Time Points	20	19	17	15	11

Trip Rate

	Δ = 1	Δ = 2	Δ = 4	Δ = 6	Δ = 10
Average Elasticity	.57	.63	.68	.63	.58
Standard Deviation	1.10	.61	.31	.16	.06
Time Points	20	19	17	15	11

Vehicle-Kilometrage

	Δ = 1	Δ = 2	Δ = 4	Δ = 6	Δ = 10
Average Elasticity	1.08	.83	.68	.65	.65
Standard Deviation	2.18	.90	.41	.31	.25
Time Points	20	19	17	15	11

9.5 Summary

The evolution of household demographics and socio-economics, car ownership, and mobility is simulated with MIDAS using the expanded Panel household sample described in Section 8. A simulation period of 25 years is used that starts in 1986, when the Wave 5 survey was conducted, and ends in 2010. The time increment in the simulation is one year.

The baseline run of MIDAS assumes an income growth rate similar to the CPB "referentie" scenario. The results show a rapid decrease in household size, gradual decline in labor force participation, and increases in the driver population and household car ownership. All mobility figures show substantial increases, in particular for vehicle driver-kilometers and transit passenger-kilometers.

Comparing the MIDAS baseline results to the CPB "referentie" socio-economic and demographic scenario, it is found that MIDAS results show much faster decline in household size. MIDAS also forecasts a faster increase in the driver population in terms of the percentage of licensed drivers among the nationwide population. The discrepancy between the MIDAS and CPB results is much smaller for the percentage in the adult population (individuals of 18 years old and over). The MIDAS results show a similar probability that an adult individual will be holding a driver's license in 2010. The apparent discrepancies between MIDAS and CPB are caused by the difference in the 2010 age distribution in the CPB scenario and that simulated by MIDAS. MIDAS depicts much more rapid aging of the Dutch population.

MIDAS forecasts the number of cars per person in 2010 to be 0.48. This is slightly smaller than the value obtained (0.52) in the CPB referentie scenario. The number of cars per household forecast by MIDAS is smaller than the CPB value by 24%, largely because of the decline in household size in the MIDAS simulation. The MIDAS forecast of nationwide fleet size is approximately 9% smaller than the CPB value. Another important difference is the decline in labor force participation shown by MIDAS results. This difference between MIDAS and CPB may be the outcome of an assumption not shared between the two, e.g., increased labor force participation by women.

MIDAS's motorized trip rates are 17% below the comparable OVG trip rates. This may be due to the well documented trip under-reporting in the Dutch Panel survey. This should be kept in mind in interpreting the mobility forecasts provided by MIDAS. A good agreement exists between van den Broecke's models and MIDAS in the 2010 labor force participation forecasts. MIDAS assumes practically the same income growth rate as in van den Broecke. Considering the fundamental differences in data and methodology, the compatibility between the van den Broecke forecasts and MIDAS results, including driver's license and car ownership, is noteworthy.

Vehicle-kilometrage forecasts are strikingly similar between MIDAS and the National Model. The National Model forecasts an increase of 72% by the year 2010, while the MIDAS increase is 74.2%. The National Model predicts a slight decrease in public transit passenger-kilometers by 2010. MIDAS, on the other hand, forecasts an increase of over 80%. This is the only important discrepancy between MIDAS and the National Model. The salient difference between the two is that the National Model is formulated using cross-sectional data and longitudinal changes in population compositions are represented by weighting households while MIDAS is based on longitudinal data and simulates household evolution over time.

Analysis with two additional income growth scenarios has been conducted to study MIDAS' performance and illustrate the possible use of this forecasting package. The scenarios depict the income growth in the CPB "optimistisch" scenario and the one in the "pessimistisch" scenario. The results indicate that the number of cars is forecast to increase by 36% by 2010 under the pessimistisch growth scenario, 43% under the referentie scenario, and 51% under the optimistisch scenario. Clearly a higher income growth rate leads to a higher level of car ownership, but this car ownership growth is not as fast as the income growth (41%, 72% and 120%, respectively). At the nationwide level, motorized-trip rates grow virtually at the same pace as car ownership. Car trip rates also show growth rates similar to those of car ownership. Nationwide growth rates in vehicle-kilometers range from 64% in the pessimistisch scenario to 80% in the optimistisch scenario.

Public transit trip rates, on the other hand, show more complex growth patterns; the trip rate is slightly higher for the pessimistisch scenario (29%) than for the referentie scenario (28%), while it is the highest for the optimistisch scenario (33%). This is the result of two tendencies. On one hand, when income growth is slow, the lower car ownership level leads to more frequent use of public transit. On the other hand, higher income implies more long distance rail trips, leading to the highest transit trip rate in the optimistisch scenario. Therefore, when income increases rail trips increase and when income decreases the trips made by bus-tram-metro increase. Also, transit passenger-kilometrage increases with income because of the increase in rail trips. In summary, the scenario analysis of this section indicates that transit use will increase under any income growth scenario examined here. In addition, the analysis in this section shows that both transit trip rates and passenger-kilometrage grow fastest under the fastest income growth scenario.

Aggregate elasticities computed from the MIDAS output are used to study the relationship between income growth and mobility measures. The elasticities are computed for different time intervals. The results for car ownership and vehicle-kilometrage offer clear indications that their aggregate elasticities decline as the time interval for evaluation increases. This implies that average car ownership is a concave function of average income; the marginal increase in average car ownership gradually declines as average income continues to increase. Similar results can be observed for vehicle kilometrage. The elasticities for trip rate suggest that average trip rate is linearly related to average income. This indicates that income and trip rates increase at equal rates. Elasticities of this type can be computed to study the relationships between any pair of variables of interest. For example the effect of increases in drivers license holding can be simulated and elasticities constructed to investigate the possible lagged effect this increase may have on car ownership and mobility.

10. Conclusion

The study leading to the development of MIDAS represents an entirely innovative approach to travel demand forecasting. Unlike the conventional approach of taking externally produced demographic and socio-economic forecasts and using them as input to a cross-sectionally estimated model system, MIDAS generates demographic and socio-economic, as well as car ownership and mobility forecasts internally through micro-simulation. A system of dynamic models estimated using the Dutch National Mobility Panel data set is applied in the simulation.

The effort has been motivated by the recognition that no external demographic and socio-economic forecasts are furnished at levels that meet the data requirements of sophisticated discrete choice models currently used in transportation planning. Specifically, no external forecasts are produced to provide a multivariate distribution of the array of explanatory variables typically used in travel choice models, at the levels where these models are formulated, i.e., households or individuals. It has also been motivated by the recognition of the dubious assumption inherent in the use of cross-sectional models in forecasting: behavioral changes over time can be forecast using cross-sectional variations observed at one point in time.

The 2010 forecasts produced by MIDAS clearly indicate that MIDAS is capable of serving as a decision support tool for transportation planning and policy development. MIDAS forecasts are in general similar to other existing forecasts, including those offered by CPB. Important differences are that: 1) MIDAS depicts a much faster decrease in household size, and 2) MIDAS forecasts a decline in labor force participation.

These discrepancies are viewed as a result of the differences in the assumptions underlying the respective forecasts. Although the exact assumptions used in the other forecasts are not known to the project team, it is conceivable that the CPB employment forecasts adopted the assumption that women's labor force participation will continue to increase to 2010. MIDAS, on the other hand, uses the employment transition probabilities observed during a part of the Dutch Panel study (1984 - 1988). The MIDAS household size forecasts are also based on observed changes during the Panel period. The decline shown by MIDAS forecasts is in fact, slightly less than that observed between 1975 and 1985.

These assumptions can be modified in MIDAS using the input parameters which have been furnished precisely for this purpose (see Section 8.1). This was intentionally ignored in this study; as the primary objective of the project is to demonstrate that long-range travel demand forecasting can be practically and meaningfully performed using micro-simulation with a system of dynamic models and parameters estimated using the Dutch National Mobility Panel data.

MIDAS forecasts practically the same levels of increase in the driver population and vehicle-kilometers as the 2010 National Model forecasts. On the other hand, it forecasts somewhat slower growth in the national car park under an income growth rate that is comparable to the one in the CPB referentie scenario. Most importantly, MIDAS forecasts that, by 2010, transit trips will increase by 20% per person and by 28% nation-wide. It also forecasts a 71% increase in the average public transit passenger-kilometers per person, and an 82% increase in the nation-wide passenger-kilometers. These forecasts vary drastically from the National Model forecasts which indicates a slight decline in public transit passenger-kilometers. Probing into the source of this difference is a subject for future research. If such an investigation is made comparing the National Model and

MIDAS, it will offer an ideal setting for detailed and practical evaluation of the relative advantages and disadvantages of cross-sectional vs. dynamic forecasting.

The forecasting exercise reported in Section 9 has offered evidence that MIDAS is a credible forecasting model system. One of its strengths lies in the fact that MIDAS internally generates pertinent demographic and socio-economic factors while maintaining coherent relationships among themselves. This is a significant advantage for MIDAS as a policy tool. Any parameter can be modified by the user to represent the scenario of interest; MIDAS will automatically simulate the repercussions that follow and reflect them in its mobility forecasts. For example, suppose the parameters associated with employment are modified to represent increased labor force participation by women. This will automatically lead to an increase in household income, a decrease in the number of births, and possibly a change in car ownership and mobility.

Despite the relatively small data base used in the estimation of its model components, and the fundamentally different principles it is based on, MIDAS has produced plausible forecasts that are comparable with existing forecasts. Overall, the main objective of the project, to examine whether dynamic models can serve as a decision support tool in transportation planning, has been successfully achieved. MIDAS is able to entertain a wide range of "what if" questions while providing internal consistency that is unmatched by any other transportation demand forecasting model. The large number of model parameters can aid in most closely representing any policy scenario of interest. These are advantages which MIDAS has over conventional cross-sectional forecasting models.

However, it cannot be over-emphasized that the use of dynamic models and micro-simulation in travel demand forecasting and policy analysis, is yet in its infancy. It will require continuous and intensive research effort before dynamic models and simulation become a practical tool for planners and policy analysts. MIDAS, perhaps the first full-fledged dynamic simulation forecasting system in the transportation planning field, is not yet a completed tool. In fact, its current version needs to be improved in a number of ways. In particular, several important dimensions in transportation planning are yet to be incorporated into MIDAS (e.g., residential and employment location). Also, MIDAS, as it stands now, has only limited sensitivity to transportation supply characteristics, e.g., highway congestion, fuel prices, public transit service levels, and automobile purchase prices and maintenance costs. These considerations have led to the development of the following recommendations for enhancing the usefulness of MIDAS as a policy tool.

Recommendations

Outlined below are various directions for improvement spanning from the more "pragmatic" visual input-output enhancement to the more complex modelling of migration and relocation.

1. Identify those parameters that are essential to represent policy scenarios for long-range transportation forecasting, and incorporate them into MIDAS.

   This can be best achieved through meetings involving likely MIDAS users and the research team members. This will enable further development of a tailor-made policy analysis tool which can probe into specific policy questions for the Netherlands.

2. Enhance the MIDAS demographic component through

   - development of a component to represent regional immigration and outmigration,

- further causal analysis of the Dutch National Panel Data using structural equations models to better represent the causal structure leading to changes in household structure and employment,
- analysis of OVG household data, and
- further effort to identify external data sources for model validation, preferably by a Dutch-speaking researcher(s).

The social and travel behavior of newly relocated residents may be substantially different from the behavior of the established residents. The introduction of a migration component, together with a residential location component (see 5 below) will allow MIDAS to reflect regional growth and capture the differences among ex-migrants, new in-migrants and established residents. The use of external data will fine tune MIDAS to more exactly represent the Dutch population and replicate its changes with improved accuracy.

3. Enhance the MIDAS mobility component by

- refining the modal split and distance models by disaggregating the public transit mode into inter-urban (rail) and intra-urban (bus, tram, and metro),
- refining the trip generation, modal split, and trip distance models by formulating them for work trips and non-work trips separately, and weekday trips and weekend trips separately,
- identifying and incorporating better indicators of highway and public transit levels-of-service into the mobility component models,
- analyzing the effect of panel fatigue and trip under-reporting on coefficient estimates of the mobility component models, and

- developing a more efficient consistent estimation procedure of a model system that involves dynamic discrete choice models.

These modifications are recommended to improve the accuracy of the MIDAS mobility component and make its outcome more informative for policy analysis. In particular they address the current weakness that MIDAS is not sensitive to supply variables. The inclusion of better level-of-service indicators will allow the users to study the impact of network improvements on travel behavior in both short and long terms. Estimation techniques for dynamic model systems are rapidly improving, and MIDAS should take advantage of new methods which may lead to more precise estimation and accurate forecasts.

4. Make the PC version of MIDAS user-friendly by

- making MIDAS menu-driven,
- providing graphic displays of simulation results, and
- internally generating dissemination-ready summary tables (similar to Table 9.1 of this report).

The usefulness of a software package largely depends on its user-friendliness. MIDAS' output is made of tables of forecasts for the 25 simulation years. The tables and figures in Section 9 were produced by post-processing the MIDAS output tables. This is a time consuming exercise. A software component which generates dissemination-ready summary tables can be designed to facilitate the inspection of the simulation results and to reduce the time needed for report preparation.

5. Develop a model component for residential location such that household relocation behavior (either within or outside the same urban area) and new or immigrating households' location choice can be represented by MIDAS.

It is well known that employment search, residential location, car ownership, and travel behavior are interconnected processes. The development of a model system that integrates these processes will allow more realistic analysis of travel demand and regional growth. In addition, more accurate prediction of trip length will be possible by introducing residence and employment location into the model system.

6. Conduct a comparative study of the Dutch National Model and MIDAS as long-range forecasting tools and determine the relative advantages and disadvantages of cross-sectional vs. dynamic forecasting.

Further investigation is required to determine the advantages and disadvantages of dynamic forecasting methods relative to cross-sectional methods. The advantages of dynamic methods that motivated this research include:

- Forecasting based on dynamic models does not rest on the untested assumption of cross-sectional forecasting, i.e., cross-sectional variations observed at one point in time, can be extrapolated over time to forecast future behavior;

- Dynamic models with lagged dependent variables in general offer improved predictive accuracy, at least in short-term forecasting;

- Dynamic models are able to represent dynamic aspects of travel behavior such as response lags (e.g., the time lag between the time a household member obtains a driver's license and the time when the household acquires a new car);

- Consequently, dynamic micro-simulation can represent causal relationships more realistically as a "cause" does not lead to an "effect" instantaneously in the real world (e.g., consider changes in land use as a result of residential and employment location);

- Dynamic micro-simulation is able to forecast the evolution of an urban population while maintaining, at disaggregate household or person levels, coherent relationships among socio-demographic factors that contribute to travel demand;

- A dynamic micro-simulation system contains many internal linkages whose parameters can be modified to represent a much wider range of scenarios, while allowing to capture all the repercussions of the changes assumed in a growth scenario;

- The time dimension is explicit in dynamic forecasting, permitting the evaluation of different policy implementation stages; and

- Changes in demand can be forecast along a continuous time axis.

On the last point, it is worthy to note that dynamic micro-simulation is able to forecast the evolution of a national car park – e.g., how conventional cars are replaced by catalytic-converter equipped cars.

The experience gained during the development of MIDAS is in agreement with these expectations. At the same time, it has been recognized that:

- A dynamic micro-simulation model system is complex;

- The development of a dynamic demographic micro-simulation model requires a large amount of data;

- Readily available population statistics may not be useful in developing a dynamic demographic micro-simulation model;

- The estimation of a dynamic mobility model requires more data than does a corresponding cross-sectional model;

- The estimation of dynamic models using panel data requires additional attention due to panel attrition, conditioning and fatigue;

- The accuracy of forecasts produced by dynamic micro-simulation is more difficult to evaluate than the accuracy of forecasts produced by a cross-sectional model; and

- It is difficult to determine the level of accuracy required of each model component in order to achieve a given level of accuracy in demand forecasts.

The research into the use of dynamic models and micro-simulation in travel demand forecasting has just begun. There are numerous questions that need to be answered. A detailed, comparative analysis of the National Model and MIDAS will offer a setting where these questions can be examined. The experience so far gained with MIDAS indicates that this will lead to the development of improved tools for transportation planning and policy analysis.

Acknowledgments

The authors are indebted to the Projectbureau which sponsored this effort into an uncharted research area. In particular, Monique Kockelkoren, Joke Jager, and Max Klok of Projectbureau are gratefully acknowledged for their continuous support, encouragement, and valuable comments throughout this project. Toon van der Hoorn of Rijkswaterstaat is acknowledged for his constructive comments during the early stages of the project. Henk Meurs of BGC (currently with MuConsult) played an important role in the project by steering the research team. Discussions with Phil Goodwin of TSU, Oxford University, were extremely useful in the initial stage of this effort. At BGC, Co Holzapfel and Jackeline Visser helped with the data preparation. Hugh Gunn of the Hague Consulting Group provided the accessibility measures used in the project and Anton van den Broecke provided many pieces of valuable demographic information.

References

Baanders, B. and K. Slootman (1989) A panel for longitudinal research into travel behaviour.

Bachman, J.G., P.M. O'Malley and J. Johnston (1978) Adolescence to Adulthood-Change and Stability in the lives of Young men. In Youth in Transition Volume IV, Institute for Social Research, The University of Michigan, Ann Arbor, MI.

Berkson, J. (1944) Application of the logistic function to bio-assay. Journal of the American Statistical Association, 39, 357-365.

Berkson, J. (1953) A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. Journal of the American Statistical Association, 48, 565-599.

BMDP (1985 & 1988) BMDP Statistical Software Manual. University of California Press, Berkeley, CA.

Bureau Goudappel Coffeng (1984) Weighting of the Panel Sample, Concept Report, Deventer, The Netherlands.

Centraal Bureau voor de Statistiek (1987) Socio-Demographic Accounts, Framework to Measure the Dynamics of Population. Department of Social Accounts, Voorburg, The Netherlands.

Centraal Bureau voor de Statistiek (1988) Statistisch Zakboek 1988. Staatsuitgeverij, CBS-Publikaties, 's-Gravenhage, The Netherlands.

Clarke M. Dix M. and Goodwin P. (1982) Some issues of dynamics in forecasting travel behaviour–A discussion paper. Transportation, 11, 153-72.

Davidson, T.N. (1972) Evolution of a Strategy for Longitudinal Analysis of Survey Panel Data. In Youth in Transition Volume IV, Institute for Social Research, The University of Michigan, Ann Arbor, MI.

Fienberg, S.E. (1980) The Analysis of Cross-classified Categorical Data. 2nd ed., MIT Press, Cambridge, MA.

Gart, J.J. and J.R. Zweifel (1967) On the bias of various estimators of the logit and its variance with application to quantal bio-assay. Biometrika, 54, 181-187.

Gart, J.J., H.M. Pettigrew and D.G. Thomas (1985) The effect of bias, variance estimation, skewness and kurtosis of the empirical logit on weighted least squares analyses. Biometrika, 72,179-190.

Geinzer, J. and A. Daly (1981) Zuidvleugel Study–Report 9, Models of Car Ownership and License Holding. Cambridge Systematics Europe, b.v., The Hague, The Netherlands.

Golob, J.M., L.J.M. Schreurs, and J.G. Smit (1986) The design and policy applications of a panel for studying changes in mobility over time. In Behavioural Research for Transport Policy, VNU Press, Utrecht, The Netherlands, pp. 81-95.

Golob, T.F., (1986) A non-linear canonical correlation analysis of weekly trip chaining behavior. Transportation Research A, 20A, 385-99.

Golob, T.F. (1989) The causal influence of income and car ownership on trip generation by mode. Journal of Transport Economics and Policy, 23, 141-62.

Golob, T.F. (1990) The dynamics of household travel time expenditures and car ownership decisions. Transportation Research, 24A, 443-463.

Golob, T.F., and H. Meurs (1986) Biases in response over time in a seven-day travel diary. Transportation, 13, 163-81.

Golob, T.F., and H. Meurs (1987) A structural model of temporal change in multi-modal travel demand. Transportation Research A, 21A, 391-400.

Golob, T.f., and L. van Wissen (1988) A joint household travel distance generation and car ownership model. UCI-ITS-WP-88-15, Institute of Transportation Studies, University of California, Irvine, CA.

Golob, T.F., L. van Wissen, and H. Meurs (1986) A dynamic analysis of travel demand. Transportation Research, 20A, 5, 401-14.

Goodman, L.A. (1973) Causal analysis of data from panel studies and other kinds of surveys. American Journal of Sociology, 78, 1135-91.

Goodwin P.B. (1977) Habit and hysteresis in modal choice. Urban Studies, 14, 95-98.

Goodwin, P. (1983) Some problems in activity approaches to travel demand. In Recent Advances in Travel Demand Analysis (edited by S. Carpenter and P. Jones), Gower, Aldershot, England, pp. 470-74.

Goodwin, P.B. (1986) Changes in Car Ownership and Bus Use 1981-84: A Panel Analysis. TSU Working Paper 320, University of Oxford, England.

Goodwin P.B. (1987) Family changes and public transport use: A dynamic analysis using panel data. In Analysis of Panel Data, Proceedings of the Round Table Conference on the Longitudinal Travel Study, Project Bureau for Integrated Transport Studies, Netherlands Ministry of Transport and Public Works, The Hague, The Netherlands, pp. 139-172.

Goodwin, P.B., M.C. Dix and A.D. Layzell (1987) The case for heterodoxy in longitudinal analysis. Transportation Research A, 21A, 363-376.

Goulias, K.G. (1991) MIDAS User's Manual.

Goulias, K. and R. Kitamura (1989) A Recursive Model System for Trip Generation and Trip Chaining. Transportation Research Record, 1236, 59-66.

Goulias, K.G. and R. Kitamura. (1991) Analysis of Binary Choice Frequencies with Limit Cases: Comparison of Alternative Estimation Methods and Application to Weekly Household Mode Choice. February.

Greene, W. H. (1981) On the asymptotic bias of the ordinary least squares estimator of the Tobit model. Econometrika, 49, 505-513.

Gunn, H., T. van der Hoorn and A. Daly (1987) Long-range, country-wide travel demand forecasts from models of individual choice. Paper presented at the Fifth International Conference on Travel Behavior Analysis, Aix-en-Provence, France, October.

Haldane, J.B.S. (1955) The Estimation and significance of the logarithm of a ratio of frequencies. Annals of Human Genetics, 20, 309-311.

Jones, P.M., M.C. Dix, M.I. Clarke, and I.G. Heggie (1983) Understanding Travel Behaviour. Gower Publishing, Aldershot, England.

Jones, P.M. F.S. Koppelman and J.P. Orfueil (1988) Activity analysis: State-of-the-art and future directions. Paper presented at the 1988 Oxford Conference on Travel and Transportation: New Developments in Dynamic and Activity-Based Approaches, Oxford University, England, July.

Juster, F.T. and C.K. Land (1981) Social Accounting Systems. Academic Press, New York, NY.

Kasprzyk, D., G. Duncan, G. Kalton and M.P. Singh eds. (1989) Panel Surveys. John Wiley and Sons, New York.

Kitamura, R. (1987a) A Dynamic Model System of Household Car Ownership, Trip Generation, and Modal Split: Model Development and Simulation Experiments. Prepared for Rijkswaterstaat, Dienst Verkeerskunde, the Dutch Ministry of Transport and Public Works.

Kitamura, R. (1987b) A panel analysis of household car ownership and mobility. Proceedings of the Japan Society of Civil Engineers, No. 383/IV-7, 13-27.

Kitamura, R. (1988a) An Analysis of Weekly Activity Patterns and Travel Expenditure. In R.G. Golledge and H.J.P. Timmermans (eds.), Behavioral Modeling Approaches in Geography and Planning, Croom Helm, London, pp. 399-423.

Kitamura, R. (1988b) Formulation of Trip Generation Models Using Panel Data. Transportation Research Record, 1203, 60-68.

Kitamura, R. (1988c) A dynamic model system of household car ownership, trip generation, and modal split: model development and simulation experiments. In Proceedings of the 14th Australian Road Research Board Conference, Part 3, Australian Road Research Board, Vermont South, Victoria, Australia, 1988, pp. 96-111.

Kitamura, R. (1988d) Life-Style and Travel Demand. In A Look Ahead, Year 2020, Special Report 220, Transportation Research Board, Washington, D.C., pp. 149-189.

Kitamura, R. (1989a) A Causal Analysis of Car Ownership and Transit Use. Transportation, 16, 153-173.

Kitamura, R. (1989b) The Asymmetry of the Change in Household Car Ownership and Utilization: A Panel Analysis. In International Association for Travel Behaviour (ed.), Travel Behaviour Research, Gower Publishing, Aldershot, England, pp. 186-198.

Kitamura, R. (1990) Panel Analysis in Transportation Planning: An Overview. Transportation Research, 24A(6), 401-415.

Kitamura, R., and P.H.L. Bovy (1987) Analysis of attrition biases and trip reporting errors for panel data. Transportation Research A, 21A, 287-302.

Kitamura, R. and D.S. Bunch (1990) Heterogeneity and State Dependence in Household Car Ownership: A Panel Analysis Using Ordered-Response Probit Models with Error Components. In M. Koshi (ed.), Transportation and Traffic Theory, Elsevier, pp. 477-496.

Kitamura, R., K. Nishii and K. Goulias (1990) Trip Chaining Behavior by Cen- tral City Commuters: A Causal Analysis of Time-Space Constraints. In P.M. Jones (ed.), New Developments in Dynamic and Activity-Based Approaches to Travel Analysis, Gower Publishing, Aldershot, England, pp. 145-170.

Kitamura, R., and T. van der Hoorn (1987) Regularity and irreversibility of weekly travel behavior. Transportation, 14, 227-51.

Kostyniuk, L.P. and R. Kitamura (1982) Life cycle and household time-space paths: Empirical investigation. Transportation Research Record, 879, 28-37.

Land, C.K. and S. Spilerman (1975) Social Indicator Models. Russell Sage Foundation. New York, NY.

Land C.K and A. Rogers (1982) Multidimensional Mathematical Demography. Academic Press. New York. NY.

Mackett, R.L. (1985) Micro analytic simulation of locational and travel behaviour. In Proceedings of PTRC Summer Annual Meeting, Seminar L, Transportation Planning Methods, PTRC, London, England, pp. 175-188.

Meurs, H., F. Gloerich, P. van de Mede, J. Visser, and M. Klok (1987) The analysis of change in mobility in the Netherlands. Paper to be presented at the International Conference on Travel Behavior, Aix-en- Provence, France, October.

Meurs, H., P. van de Mede, J. Visser, and L. van Wissen (1987) Analysis of Panel Data. Bureau Goudappel Coffeng, Deventer, the Netherlands.

Meurs H. van Wissen L. and Visser J. (1989) Measurement biases in panel data. Transportation, 16(2), 175-94.

Miller, E.J., P.J. Noehammer and D.R. Ross (1987) A micro-simulation model of residential mobility. In Proceedings for the International Symposium on Transport, Communication and Urban Form: 2, Analytical Techniques and Case Studies (edited by W. Young), Clayton, Monash University, Victoria, Australia, pp. 217-234.

Moning, H.J. (1983) The National Travel Survey in The Netherlands. Nota Nummer H.4308-83-E4, Centraal Bureau voor de Statistiek, Voorburg, The Netherlands.

Orcutt G., S. Caldwell and R. Wertheimer, II (1976). Policy Exploration Through Microanalytic Simulation. The Urban Institute, Washington D.C.

Recker, W.W., T.F. Golob, M.G. McNally, and J.D. Leonard (1987) Dynamic tests of a time-space model of complex travel behavior. Paper to be presented at the International Conference on Travel Behavior, Aix-en- Provence, France, October.

Rossi, J.R. and J.K. Gilmartin (1980) The Handbook of Social Indicators. Garland STPM Press. New York, NY.

Schoen, R. (1988) Modeling Multigroup Populations. Plenum Press. New York. NY.

Spengler, J.J. and O.D. Duncan (1956) Demographic Analysis. The Free Press, Glencoe, IL.

van den Broecke, A.A.J. (1988) Long term forecasting of car ownership with the cohort processing model. PTRC 16th Summer Annual Meeting, Proceedings of Seminar D, PTRC Education and Research Services, Ltd., London, England, pp. 183-191.

van den Broecke/Social Research (1987a) Demografische Ontwikkelingen en Verkeers- en Vervoerproblematiek in de Komende 25 Jaar. Amsterdam, The Netherlands.

van den Broecke/Social Research (1987b) De Mogelijke Groei van het Personenautobezit tot 2010. Amsterdam, The Netherlands.

van der Hoorn, T., and R. Kitamura (1987) Evaluation of the predictive accuracy of cross-sectional and dynamic trip generation models using panel data. Presented at the 66th Annual Meeting of the Transportation Research Board, Washington, D.C., January.

van der Hoorn, T. and H. van Hoek (1988) Translation of results from the Dutch National Model into practical information for policy recommendations. PTRC.

van der Loop, H. (1988) Panel analysis of travel behavior for policy applications: methods used and results available. Paper presented at the Oxford Conference on Travel and Transportation: New Developments in Dynamic and Activity-based Approaches, Oxford, England.

van Wissen, L.J.G. and T.F. Golob (1990) Simultaneous-equations system involving binary choice variables. Geographical Analysis, 22(3), 224-243.

van Wissen, L.J.G. and H.J. Meurs (1989) The Dutch Mobility Panel: Experiences and evaluation. Transportation, 16(2), 99-119.

Vrolijk, P., H. Gunn and T. van der Hoorn (1987) Waar komt de groei vandaan? Presented at the 14th Colloquium Vervoersplanologisch Speurwerk, Den Haag, The Netherlands.

Willis K.G. (1971) Models of Population and Income. Research Monograph No 1. Agricultural Adjustment Unit, University of Newcastle upon Tyne, Newcastle, United Kingdom.

Appendix A

Missing Accessibility Data

The accessibility indices in the data files given to the project by the Hague Consultancy Group (HCG) are formulated using a zone system which is not compatible with the system used in the Dutch National Mobility Panel data files. Thus, a conversion procedure was needed in order to use the HCG accessibility indices in the MIDAS mobility component models. The conversion consisted of identifying the correspondence between the zones used in the HCG files and the Panel files. For some of the zones, this conversion was straightforward because a HCG zone was identical to, or completely contained, a Panel zone. In the reverse case where a HCG zone is a part of a Panel zone (therefore the latter contains more than one HCG zones), the conversion had to be performed manually. During the summer of 1989, the HCG data files and Panel Data files were matched for the first time, but only for those zones where conversion can be made automatically. This resulted in a small number of Panel zones for which accessibility indices were missing.

The last two waves of the Dutch Panel survey data were not available in 1989. During 1990, the matching procedure for the accessibility indices was reconvened for the data files from Waves 9 and 10. In summer 1990, the procedure was repeated at Bureau Goudappel Coffeng (BGC) to see if any additional accessibility indices could be added to the panel for the earlier waves as well. This, however, did not result in any substantial gain of information as accessibility indices were identified only for 10 additional households.