



A Dynamic Random effects Multinomial Logit Model of Household Car Ownership

Søren Leth-Petersen and
Thomas Bue Bjørner

May 4:2005

akf working paper contains provisional results of studies or preliminary work of reports or articles. Therefore, the reader should be aware of the fact that results and interpretations in the finished report or article may differ from the working paper. akf working paper is not covered by the procedures about quality assurance and editing applying to finished akf reports. akf working paper is only available on www.akf.dk and not in a printed version.

A Dynamic Random Effects Multinomial Logit Model of Household Car Ownership

Søren Leth-Petersen [†]

Thomas Bue Bjørner [‡]

Version 200405

Abstract

Using a unique household panel we estimate demand for car ownership by means of a dynamic multinomial model with correlated random effects. Results suggest that the persistence in car ownership observed in the data should be attributed to both true state dependence and to unobserved heterogeneity (random effects). It also appears that random effects related to single and multiple car ownership are correlated, suggesting that the IIA assumption employed in simple multinomial models of car ownership is invalid. Relatively small elasticities with respect to income and car costs are estimated. It should, however, be noted that quantitative importance of state dependence is considerably larger for households with single car ownership as compared with multiple car ownership. This suggests that the holding of a second car will be more affected by changes in the socioeconomic conditions of the household and by economic policy shocks.

Jel codes: C23, C25 and R41

Key words: car ownership, panel data, dynamic multinomial model

Acknowledgements: We thank Kenneth Train for providing the mixed logit estimation software developed by Kenneth Train, David Revelt, and Paul Ruud. The research was supported by the Danish Environmental Research Programme.

[†] akf, institute of local government studies – Denmark, Nyropsgade 37 DK-1602, Copenhagen V, Email: slp@akf.dk

[‡] The Secretariat of the Danish Economic Council, Amaliegade 44, DK-1256 Copenhagen K, Email: tbb@dors.dk

1. Introduction

The stock of cars held by individual households is observed to be persistent when the household is observed repeatedly. If a given household has one car in one year then it is likely to be observed holding one car also in the following year. This pattern will have important implications for assessing the impact of policy shocks on the probability of changing the car stock in the short and long run. In this paper we present an empirical analysis of the dynamics of car holdings that allow us to characterize the persistence in the car stock at the household level.

Persistence in the car stock at the household level can arise for different reasons. One possibility is that persistence is caused by unobserved household specific preferences for car holdings that are constant across time. This is known as “spurious” state dependence, Heckman (1981). Alternatively, it can be due to, for example, the presence of transaction costs associated with adjusting the size of the car stock or with habit formation. Transaction costs are unobserved but not fixed across time, and will show up as “true” state dependence. These two sources of persistence nevertheless have very different implications in terms of policy analysis. For example, consider a policy introducing a new tax on car ownership. If persistence is caused by unobserved fixed differences in preferences for cars then this policy will have an immediate effect that is identical in both the short and the long run. On the other hand, if persistence is generated by the presence of transaction costs associated with adjusting the car stock then the policy will have an effect that is different in the short and the long run.

An econometric model should be capable of handling both sources of persistence. To distinguish between these two types of persistence we estimate a dynamic discrete choice model with unobserved time-invariant heterogeneity employing the approach suggested by Wooldridge (2002a). Estimating this model will permit us to evaluate the importance of time invariant unobserved heterogeneity versus state dependence. The analysis is based on an extraordinary data

set allowing us to follow the car stock of a large number of individual households over a period of ten years. The dataset is based on merged public administrative register data that give us information on the car stock, income, family composition, age etc. for a large number of households for the period 1992 to 2001. Most previous studies are based on either aggregate data or micro cross-section data. Some studies based on repeated cross-section data employ pseudo panel methods; see for example Dargay and Vythoulkas (1999), and Dargay (2001). Both types of data lack the idiosyncratic aspect, and such methods cannot be used to distinguish between persistence due to unobserved heterogeneity and persistence due to state dependence.

The few previous studies, which have employed micro panel data, have either made arbitrary assumptions regarding the nature of the persistence in data or used less general specification of the empirical model than the one applied here. As an example, Meurs (1993) assumed that the persistence in car ownership should be attributed to unobserved heterogeneity. Kitamura and Bunch (1990) estimated models which included both state dependence and unobserved heterogeneity. However, they applied an ordered probit model, which is restrictive in the sense that the same index function is used to determine both single and multiple car ownership (with the exception of a threshold parameter). That is, changes in socioeconomic variables have the same effect (on the index) for single and multiple car ownership. The multinomial logit model is more flexible as it allows for different parameters for single and multiple car ownership. This is likely to be important if a first car does not serve the same purpose as a second car. In a previous paper based on the same data source as applied here Bjørner and Leth-Petersen assessed the dynamic properties of single adults and households consisting of two adults (“couples”). For couples the decision to hold two cars relative to one was modelled as a separate decision from the decision to hold one car relative to no car. This assumption simplifies estimations, but the assumption is questionable, because households that for some unobserved reason prefer to hold one car more than no car are also likely

to have unobserved preferences for holding two cars. Such unobserved correlated heterogeneity in demand for car ownership can arise, for example because households have different access to public transport that can be used as a substitute for car transportation. In this paper a more complete analysis is presented for the decision for couples to hold 0, 1 or 2 cars. Here the choice of car stock is modelled as a dynamic multinomial choice model that allows for unobserved correlated heterogeneity across car stock categories.

Results show that both unobserved time invariant heterogeneity and state dependence are important factors to include in the analysis of household car ownership. Households are shown to respond very little in the short run to changing income and user cost levels. This has important implications for understanding the effects of policy measures in the short run.

In the next section we present the dataset. In section 3 the econometric framework is lined up and results are presented in section 4. Section 5 assesses the economic importance of different policy shocks. Section 6 concludes.

2. Data

The analysis is based on a household level panel data set with information about 10,565 households that we are able to follow in all years in the period 1992-2001. The dataset is constructed by merging different public administrative registers at the individual level. This is possible because each individual in Denmark has a unique civil registration number that is linked to the information in the different registers. The civil registration number allows us, together with the address, to construct household units. In this way we are able to characterize the complete household in terms of car holdings, income, age, family composition, location of residence, labour market participation status. It has been widely recognized that the combination of public administrative registers and the unique civil registration number yields longitudinal data, which are quite remarkable by

international standards, see e.g. Frank (2000). Over the last years these data have been used in fields like medicine and labour market research. By combining the data with information on car ownership obtained from the Danish Central Register for Motor Vehicles we are able to link information on car ownership. The information from the Central Register for Motor Vehicles is used to collect annual ownership taxes and is therefore considered very accurate. Based on this information we calculate the degree of car ownership during the year and subsequently define a discrete car ownership variable (0 if the degree of car ownership during the year was less than 0.5, 1 if the degree was between 0.5 and 1.5 etc.).

Company cars available to private households, but owned by a company cannot be linked with households based on the information from the Danish Central Register for Motor Vehicles. However, information about the presence of a company car in a household was obtained from a tax register (as individuals with a company car in Denmark are to pay income tax on the benefits of having a car at their disposal).

Socioeconomic variables related to the household were extracted from the tax register and other sources. We have information on income (before and after tax), social transfers, demographic information, labour market status and location at municipal level. The municipality of the workplace was also obtained and used to calculate a measure of commuting distance. An index for the cost of car ownership was calculated from aggregate information about fuel prices, ownership tax, repair costs, insurance costs, price of new cars and net rate of return (alternative cost). There is only variation in the car cost index across time, but not between households in a given year.

For the analysis we consider a selected sample of households. First of all we consider only households consisting of couples, and we consider the choice between holding 0, 1 or 2 cars.¹ For single households the choice of car stock is in practice binary. This analysis is presented in Bjørner

¹ Only 1-2% of the singles have multiple private car ownership, while less than 1% of the couples own 3 cars (calculated for households without self-employed and without company cars).

and Leth-Petersen (2005). Moreover, households where one person is self-employed are deselected. This is because self-employed individuals have highly unstable incomes when measured by the tax assessed income (which may not reflect their real consumption possibilities). We consider only households where the oldest person is aged 18 years or more. Finally, we do not consider households that have a company car. We prefer to focus on privately owned cars, because it seems likely that the decision-making process to own a private car is different from the process of obtaining a company car.

Descriptive statistics

We have postulated that car ownership status is persistent across time. This claim is backed by the evidence in table 1. The table shows the number of changes in car ownership status across households in the period 1992-2001. The table shows that more than half of the households in the sample never change ownership status in the observation period, and that including up to two shifts in ownership status accounts for roughly 95% of the sample.

Table 1. Number of changes in car ownership status across households in the period 1992-2001

No. households=10,565	0	1	2	3	4+
Share of households %	56	25	13	4	2

Among those never changing ownership status in the observation period car ownership status is distributed as shown in table 2. The overwhelming majority of households with stable ownership status have held one car.

Table 2. Distribution of car ownership status for households with stable ownership status

No. households=5,939	0	1	2
Share of households %	14	82	4

The distribution for those households changing status one time in the observation period is given in table 3. Most shifts are from either no car to one car or from one car to 2 cars, i.e. there is indication of a general tendency for accumulating cars within the households in the sample.

Table 3. Distribution of changing pattern for households changing ownership status one time in the observation period

No. households =2,652	0 →1	1 →0	1 →2	2 →1
Share of households %	35	15	28	22

Selection of explanatory variables and choice of transformation of those were based on previous studies and preliminary estimations using pooled multinomial logit models. As income measure we use log of household income *after* tax measured in 1997 price level (*linc*). This measure includes wage, pensions, net capital income as well as the most important non-taxed public transfers like child support (given in Denmark independent of income), subsidies for housing rents and social benefits. Age is included both in linear and squared forms. A number of dummies indicate labour market status for males and females distinguishing between status as employed (*work*), unemployed (*unemp*). The reference is individuals outside the labour market (as described, households with self-employed are excluded). For respondents employed we calculate a measure of commuting distance based on the mean distance between municipality of living and working. For individuals living and working in the same municipality the expected commuting distance was calculated based on the size of the municipality. The square root of commuting distance was included in the models

(denoted *dism_sr* and *distf_sr* for males and females, respectively).² Dummy variables are included to indicate the presence of children under 18 years of age (*dchild_m*) and adult children living with their parents (*dchild_a*). The variable (*lusc*) is the log of car user costs, giving the development in car cost (purchase, ownership and use) relative to consumer prices (normalized to 1 in 1997). A *trend* variable normalized at 0 in 1993 is included to account for time effects (annual dummies cannot be included along with *lusc*). Finally, two dummy variables are included to indicate degree of urbanization. One dummy (*cph*) indicates if the household resides in Copenhagen and another dummy (*town*) indicates if the household lives in an urban area outside Copenhagen. The reference is households living in rural areas.

Summary statistics for all the variables included in the analysis is given in the appendix.

3. Econometric model

The purpose of the paper is to estimate probability models of car ownership status that can fall into one of three categories: 0, 1 or 2 cars. We estimate three versions of the multinomial logit model with an increasing degree of sophistication. The reference model is the (static) pooled multinomial logit model.

In the second model we expand the pooled multinomial logit model by introducing unobserved heterogeneity. It is assumed that the unobserved heterogeneity of each category is uncorrelated with the explanatory variables in the model, i.e. a random effect. However, we allow the random effects of the categories to be correlated to take into account that households that for

² Municipality of workplace was not recorded in 2% to 5% of the cases (for persons working). To accommodate this in our econometric models we include dummy variables taking the value one if this information is missing.

some unobserved reason prefer to own one car instead of no car may also be likely to prefer two cars.

In the third model the random effects model is extended by also allowing for state dependence, i.e. inclusion of lagged car ownership status. By introducing state dependence we will be able to assess the quantitative importance of the two sources of persistence that appears in the raw data series.

In terms of estimation, the model with random effects and state dependence is the most complicated. The econometric setup is therefore outlined in terms of this model. The less complicated models are special cases of this model. In the next section we present how to estimate a multinomial logit model with random effects for panel data, and finally we describe how the initial conditions problem associated with dynamic panel data models is handled. Here we follow the approach proposed by Wooldridge (2002a).

3.1 Estimation of the dynamic multinomial logit model with random effects

Consider the discrete choice model, where a given individual choose the stock j , ($j = 0,1,2$ cars), that gives the highest indirect utility at time t

$$y_{jit} = \begin{cases} 1 & \text{if } V_{jit} > V_{kit} \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j = 0,1,2 ; n = 1,...,N ; t = 1,...,T \quad (3.1)$$

so that $y_{it} = (y_{0it}, ..., y_{Jit})$,

Indirect utility, V_{jit} , is given by

$$V_{jit} = \gamma_j y_{it-1} + \lambda_j y_{i0} + \beta_j x_{it} + \varepsilon_{jit} \quad (3.2)$$

where y_{it-1} is past observed holdings, x_{it} is a vector of K observed exogenous variables and y_{i0} is the initial stock. The inclusion of the initial stock (y_{i0}) in the model is related to the initial condition problem and will be motivated further in section 3.2. Finally, ε_{jit} is an unobserved error term consisting of two parts:

$$\varepsilon_{jit} = \mu_{ji} + \nu_{jit} \quad (3.3)$$

μ_{ji} is an unobserved household effect specific to the car stock so that a given household is allowed to have an idiosyncratic time invariant preference for a particular stock of cars. ν_{jit} is an iid error term. To make this setup operational for estimation we assume that μ_{ji} follows a J -dimensional multivariate normal distribution, and that ν_{jit} is independent extreme value distributed. Moreover rewrite (3.3)

$$\varepsilon_{it} = C\xi_i + \nu_{it} \quad (3.4)$$

where ε_{it} is a $J \times 1$ vector of unobserved components, $\mu_i = C\xi_i$ where μ_i is a $J \times 1$ dimensional vector of multivariate normal distributed (conditional on (y_{i0}, x_{it})) idiosyncratic effects, and ν_{it} is a $J \times 1$ vector of unobserved independent extreme value distributed residuals. The fact that μ_i is allowed to be multivariate implies that we do not impose the IIA assumption. ξ_i is a $J \times 1$ vector of independent normally distributed variables, and CC' is the $J \times J$ covariance matrix of μ_i and C is the lower triangular Cholesky factorization of it, containing the unknown parameters of the multivariate normal distribution of time constant idiosyncratic effects. C is given by

$$C = \begin{bmatrix} c_{00} & & \\ c_{10} & c_{11} & \\ c_{12} & c_{21} & c_{22} \end{bmatrix} \quad (3.5)$$

Substituting (3.4) and (3.3) into (3.2) and writing it compactly gives

$$V_{it} = \gamma y_{it-1} + \lambda y_{i0} + \beta x_{it} + C \zeta_i + v_{it} \quad i = 1, \dots, N ; t = 1, \dots, T \quad (3.6)$$

where V_{it} is a $J \times 1$ vector of utilities for individual i at time t , γ and λ are $J \times 1$ vectors of parameters to be estimated, β is $J \times K$ vector of parameters to be estimated, and C contains the parameters of the covariance structure that are also to be estimated.

Conditional on ζ_i , the probability for a particular household choosing car stock j at time t is then

$$\text{Prob}(y_{it} = j | \zeta_i) = \frac{e^{\gamma_j y_{it-1} + \lambda_j y_{i0} + \beta_j x_{it} + C_j \zeta_i}}{\sum_{k=1}^J e^{\gamma_k y_{it-1} + \lambda_k y_{i0} + \beta_k x_{it} + C_k \zeta_i}} \quad (3.7)$$

The probability that household i is observed with a sequence of stocks y_{jit} for $j = 0, 1, 2$ and $t = 1, \dots, T$ is

$$\text{Prob}(y_i | \zeta_i) = \prod_t \prod_j \text{Prob}(y_{it} = j | \zeta_i)^{y_{jit}} \quad (3.8)$$

where C_j is the j^{th} row of C . The unconditional choice probability is

$$\text{Prob}(y_i) = \int_{\xi_i} \text{Prob}(y_i | \xi_i) f(\xi_i) d\xi_i \quad (3.9)$$

where $f(\xi_i)$ is the multivariate distribution for ξ_i . The log likelihood function is

$$\log L = \sum_{i=1}^N \text{Prob}(y_i) \quad (3.10)$$

We evaluate the integral in (3.8) by drawing ξ_i^d from the distribution of ξ_i , calculating $\text{Prob}(y_i | \xi_i^d)$, and repeating this $D=100$ times, to obtain an average hereof.

$$\text{Prôb}(y_i) = \frac{1}{D} \sum_{d=1}^D \text{Prob}(y_i | \xi_i^d) \quad (3.11)$$

Instead of using pseudo random draws we use Halton draws. For details we refer to Train (2003).

Setting the location

The model is estimable in difference form where we evaluate utility of one alternative relative to a reference alternative. For example, if the reference stock is zero cars then we specify the estimable model in terms of $V_{jit} - V_{0it}$ for $j=1,2$. This is because the probabilities for choosing either of the stocks 0, 1, and 2 must sum to unity. Therefore a reference stock must be chosen for which the probability is given by 1 minus the sum of the probabilities of choosing the other stocks. In the example where stock $j=0$ is the reference and the probabilities for the three stocks are denoted (p_0, p_1, p_2) then $p_0 = 1 - p_1 - p_2$. Thus, estimates of the parameter sets (γ_1, β_1, C_1) and

(γ_2, β_2, C_2) are always relative to the base category. The same applies to C in (3.6). Therefore we assume C to take the form

$$C = \begin{bmatrix} 0 & & \\ 0 & c_{11} & \\ 0 & c_{21} & c_{22} \end{bmatrix} \quad (3.12)$$

So that the covariance matrix of the random effects term in difference form becomes

$$CC' = \begin{bmatrix} c_{11}^2 & \\ c_{11}c_{21} & c_{21}^2 + c_{22}^2 \end{bmatrix} \quad (3.13)$$

The term $(c_{11}c_{21})$ allows for correlation across alternatives thereby avoiding imposing the assumption of independence across alternative. For example, consider a household with a high unobserved preference for holding two cars and a positive value of the term $(c_{11}c_{21})$. This household is then likely also to have a high preference for holding one car as opposed to not holding any cars. Allowing for this sort of unobserved preference correlation is potentially important, because without such correlation the preference of a given household for holding two cars is completely independent of its preference for holding one car. The sign and size of $(c_{11}c_{21})$ are to be estimated and are of course not restricted a priori.

3.2 The initial conditions problem

Estimation of dynamic panel data models with unobserved effects is a nontrivial problem. The difficulty arises because it is implausible to assume that the initial observation, y_{i0} , is independent

of the unobserved effect, μ_i . In our application this amounts to assuming that the initial stock (i.e. the one firstly observed) is independent of the unobserved preference for holding cars. This is clearly untenable. Here we follow the approach of Wooldridge (2002a) to handling the initial conditions problem by modelling the distribution of the unobserved effect, assumed to be normal, conditional on the initial value of the dependent variable, y_{i0} .

Apart from being fully parametric the main assumption underlying Wooldridge's approach is that we specify a parametric model for the density of μ_i conditional on the initial observation of the dependent variable y_{i0} . In practice this amounts to including y_{i0} as an additional regressor³. This modelling approach has previously been applied by Erdem and Sun (2001), albeit without allowing for correlation of the unobserved effects across alternatives.

4. Results

Estimation results are presented in table 4. The table shows results from estimating three models: The pooled multinomial logit, the random effects multinomial logit, and the random effects multinomial logit model with state dependence. The order of the presentation of the models represents the increasing level sophistication. The pooled model gives estimates from the standard multinomial model. The random effects model conditions on unobserved time invariant heterogeneity that can be correlated across alternatives, but is assumed to be uncorrelated with the explanatory variables. The final model conditions on the lagged level of the dependent variable and on unobserved time invariant heterogeneity. The lagged dependent variable will capture the

³ We use a simplified version of the approach suggested by Wooldridge (2002a). He also conditioned on the observed history of the exogenous explanatory variables. However, in our case inclusion of variables for the history of some explanatory variables in preliminary regressions indicated that these generally were insignificant and could be excluded without affecting the remaining parameters.

persistence in car ownership status. As the number of conditioning factors increase the models become shorter run in nature. We therefore think of the obtained estimates as having relevance for understanding short run responses.

Considering first the importance of introducing more conditioning elements into the model it is seen that the loglikelihood value becomes numerically smaller as the model becomes richer. This indicates that both unobserved heterogeneity and state dependence are relevant aspects to include in the model from a statistical point of view. It is also evident that the parameters describing the covariance structure become much smaller when state dependence is introduced. This is an indication that (true) state dependence absorbs most of the persistence in the data. It is noticeable that in both models where the covariance structure is estimated there is evidence of correlation across alternatives. This is evidence that the independence of irrelevant alternatives (IIA) invoked in the pooled multinomial logit model is restrictive. The positive parameter on the covariance term indicates that individuals having a preference for holding one car relative to no car also have a preference for holding two cars.

In all the models income is positively related to the probability of holding both one and two cars. The parameters of the income variable in the category holding two cars are larger than for the category holding one car. It is tempting at this stage to conclude that this indicates that holding two cars relative to one is more income elastic. It is, however, premature to conclude on the quantitative importance based on the parameter estimates, because the model is nonlinear. In section 5 we shall return to the quantitative importance of income. Both parameters on age and squared age are significant in all three models and indicate a concave relationship. In the static models ownership is increasing at all relevant ages while in the model with state dependence ownership is increasing up to ages 50. The peak point of the age profile is similar for both categories within all three models.

In the static models there is evidence that small children increase the probability of having one car. This is reversed in the dynamic model. This could be indicating that child expenditures crowd out car expenditures. In all three models the presence of adult children increases the probability of having two cars. This result may be related to speculation in lower insurance premiums. Young people face very high insurance premiums. Living with their parents they can save money if their car is registered as belonging to one of their parents (given that the parents already have a car and have earned discounts in insurance premiums from collusion free years).

The parameters on the variables indicating degree of urbanization have the expected signs. In more densely populated areas the need for cars conditional on all the other characteristics is smaller. Moreover, the parameter on the user cost variable is in all cases negative. It should be recalled that we only have time variation in the car cost index and the size and significance of the parameter to the car cost index are sensitive to the inclusion/omission of the trend variable, so the impact of changes in car cost should be interpreted cautiously.

Higher commuting distances for females increase ownership probabilities for both one and two cars. Results are equivocal for men, but in the dynamic model it appears that increasing commuting distance increases the probability of having two cars. The parameters of the labour market participation dummies indicate that participating in the labour market is not affecting the probability to hold cars. Recall that this result is conditional on income. Unemployment spells for males are negatively related to the probability of holding both one and two cars, but not for females.

Table 4. Parameter estimates, Pooled multinomial logit, random effects multinomial logit and random effects multinomial logit with state dependence

	Pcar=1		Pcar=2		Pcar=1		Pcar=2		Pcar=1		Pcar=2		Pcar=1		Pcar=2		Pcar=1		Pcar=2		
		s.e		s.e		s.e		s.e		s.e		s.e		s.e		s.e		s.e		s.e	
pcar _{t-1}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
pcar ₀₂	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
linc	1.3946**	0.03865	2.4034**	0.0531	1.5332**	0.0735	2.9611**	0.1026	1.5332**	0.0735	2.9611**	0.1026	1.5332**	0.0735	2.9611**	0.1026	1.5332**	0.0735	2.9611**	0.1026	
age	0.1281**	0.0042	0.1887**	0.0075	0.7763**	0.0131	1.1177**	0.0176	0.7763**	0.0131	1.1177**	0.0176	0.7763**	0.0131	1.1177**	0.0176	0.7763**	0.0131	1.1177**	0.0176	
age_sq/100	-0.1031**	0.0040	-0.1670**	0.0077	-0.7303**	0.0119	-1.0782**	0.0164	-0.7303**	0.0119	-1.0782**	0.0164	-0.7303**	0.0119	-1.0782**	0.0164	-0.7303**	0.0119	-1.0782**	0.0164	
Dehild_m	0.0443**	0.02526	-0.01835**	0.0336	0.1453**	0.0502	-0.3613**	0.0616	0.1453**	0.0502	-0.3613**	0.0616	0.1453**	0.0502	-0.3613**	0.0616	0.1453**	0.0502	-0.3613**	0.0616	
Dehild_a	-0.2355**	0.02984	0.2038**	0.0366	-0.1581**	0.0524	0.3106**	0.0601	-0.1581**	0.0524	0.3106**	0.0601	-0.1581**	0.0524	0.3106**	0.0601	-0.1581**	0.0524	0.3106**	0.0601	
Workmale	-0.0715*	0.0370	-0.1850**	0.0528	-0.3584**	0.0666	-0.0248	0.0862	-0.3584**	0.0666	-0.0248	0.0862	-0.3584**	0.0666	-0.0248	0.0862	-0.3584**	0.0666	-0.0248	0.0862	
Workfem	0.0938**	0.0352	0.1642**	0.0490	0.0452	0.0687	0.3998**	0.0880	0.0452	0.0687	0.3998**	0.0880	0.0452	0.0687	0.3998**	0.0880	0.0452	0.0687	0.3998**	0.0880	
Unempmale	-0.3582**	0.0417	-0.3884**	0.0620	-0.6304**	0.0751	-0.3248**	0.0973	-0.6304**	0.0751	-0.3248**	0.0973	-0.6304**	0.0751	-0.3248**	0.0973	-0.6304**	0.0751	-0.3248**	0.0973	
Unempfem	0.0530	0.0368	0.1894**	0.0528	-0.1035	0.0674	0.1862**	0.0881	-0.1035	0.0674	0.1862**	0.0881	-0.1035	0.0674	0.1862**	0.0881	-0.1035	0.0674	0.1862**	0.0881	
Distm_sr	-0.0177**	0.0047	0.0404**	0.0056	-0.0361**	0.0075	0.0148	0.0098	-0.0361**	0.0075	0.0148	0.0098	-0.0361**	0.0075	0.0148	0.0098	-0.0361**	0.0075	0.0148	0.0098	
Distf_sr	0.0374**	0.0074	0.0840**	0.0088	0.0514**	0.0131	0.0851**	0.0155	0.0514**	0.0131	0.0851**	0.0155	0.0514**	0.0131	0.0851**	0.0155	0.0514**	0.0131	0.0851**	0.0155	
Distm_miss	-0.3086**	0.0631	0.0136	0.0790	-0.3850**	0.1057	-0.1321	0.1311	-0.3850**	0.1057	-0.1321	0.1311	-0.3850**	0.1057	-0.1321	0.1311	-0.3850**	0.1057	-0.1321	0.1311	
Distf_miss	0.0348	0.0473	-0.0197	0.0612	0.1696**	0.0863	0.0131	0.1089	0.1696**	0.0863	0.0131	0.1089	0.1696**	0.0863	0.0131	0.1089	0.1696**	0.0863	0.0131	0.1089	
Lusc	-1.7836**	0.4753	-3.1110**	0.6509	-0.3684	0.8184	-2.4253**	1.0029	-0.3684	0.8184	-2.4253**	1.0029	-0.3684	0.8184	-2.4253**	1.0029	-0.3684	0.8184	-2.4253**	1.0029	
Trend	0.0456**	0.0041	0.0881**	0.0058	0.1388**	0.0070	0.2302**	0.0084	0.1388**	0.0070	0.2302**	0.0084	0.1388**	0.0070	0.2302**	0.0084	0.1388**	0.0070	0.2302**	0.0084	
Cph	-1.3956**	0.0266	-2.0974**	0.0379	-3.5987**	0.1218	-5.0809**	0.1451	-3.5987**	0.1218	-5.0809**	0.1451	-3.5987**	0.1218	-5.0809**	0.1451	-3.5987**	0.1218	-5.0809**	0.1451	
Town	-0.5719**	0.0239	-0.9708**	0.0309	-1.2374**	0.0825	-2.0961**	0.1036	-1.2374**	0.0825	-2.0961**	0.1036	-1.2374**	0.0825	-2.0961**	0.1036	-1.2374**	0.0825	-2.0961**	0.1036	
const	-12.5713**	1.7254	-23.8207**	2.3612	-28.9570**	3.0718	-52.2588**	3.7687	-28.9570**	3.0718	-52.2588**	3.7687	-28.9570**	3.0718	-52.2588**	3.7687	-28.9570**	3.0718	-52.2588**	3.7687	
Covariance structure																					
Covariance structure																					
Covariance structure																					
C ₁₁ ²	-	-	-	-	Parameter	37.4961**	s.e.	1.1600	Parameter	0.2764**	s.e.	0.0920	Parameter	0.2764**	s.e.	0.0920	Parameter	0.2764**	s.e.	0.0920	
C ₂₁ C ₁₁	-	-	-	-		44.3976**		1.3120		44.3976**		1.3120		44.3976**		1.3120		44.3976**		1.3120	
C ₂₁ ² + C ₂₂ ²	-	-	-	-		64.0048**		1.6300		64.0048**		1.6300		64.0048**		1.6300		64.0048**		1.6300	
Loglik	-75.463			-42,105*			-26,150			-26,150			-26,150			-26,150			-26,150		

** indicates significance at 95% level. * indicates significance at 90% level.

Finally, it should be noted that the size of the parameters in the different models cannot directly be compared because the variance of the error term plus random effects are different, see e.g. Wooldridge (2002b). The impact of changes in income (after tax) and car costs will therefore be described further in the next section. The standard errors of the pooled logit models are generally considerably smaller than in the other models. This is because the standard errors of the pooled logit are calculated subject to the (incorrect) assumption that the errors of each household are uncorrelated. This is clearly not the case, so the standard errors of the pooled logit are strongly downwards biased.

5. Quantitative importance of state dependence, income and user costs

To assess the quantitative importance of state dependence and the economic importance of changes in variables like income and user costs we need to calculate average predicted probabilities. There are no simple estimators for the average probabilities available for the mixed distribution of the logit with normally distributed random effects, see e.g. Wooldridge (2002b).⁴ Therefore, we calculate the average probabilities using a simulation approach, where the probability for each household is calculated many times adding draws from the estimated normal distribution to the index function. Let ε_{ijs} be random draws from the standard normal distribution, where s indexes the draws $(1, \dots, S)$, where we set $S=1000$. The simulated average probability is then:

$$\text{Average Prob}(y_{it} = j | \zeta_i) = \frac{1}{N} \sum_{i=1}^N \frac{1}{S} \sum_{s=1}^S \frac{e^{\hat{\gamma}_j y_{it-1} + \hat{\lambda}_j y_{i0} + \hat{\beta}_j x_{it} + \hat{C}_j \varepsilon_{ijs}}}{\sum_{k=1}^J e^{\hat{\gamma}_k y_{it-1} + \hat{\lambda}_k y_{i0} + \hat{\beta}_k x_{it} + \hat{C}_k \varepsilon_{iks}}}$$

⁴ In contrast, it is easier to calculate average probabilities in the random effects probit, see e.g. Wooldridge (2002a or 2002b). In preliminary estimations we relied on the random effects probit, but it turned out to be difficult to identify the variance parameter of the random effects.

Table 5 gives the calculated probabilities for holding 0, 1 or 2 cars under different assumptions about lagged car holding. The table clearly shows that probabilities centre on the category of the lagged value. This is most pronounced for the category holding 1 car, and less pronounced for the category holding 2 cars. This indicates that households are quicker to adjust to a policy shock if they hold two cars than if they hold 1 car.

Table 5. The quantitative importance of state dependence: Estimated probability of car ownership in 2001 conditional on different ownership levels in 2000

	Assumed car ownership in 2000		
	0	1	2
0 car	0.723	0.039	0.000
1 car	0.277	0.907	0.469
2 cars	0.000	0.053	0.531

In table 6 partial effects from a one percentage change in income and user costs are presented. The partial effects are calculated as the average probability after the change less the average probability before the change times 100, and they are calculated based on the estimates from the pooled multinomial logit, the random effects multinomial logit (RE) and the random effects multinomial logit with state dependence (RD SD). The bottom row of table 6 gives the elasticities of the total car stock with respect to income and user costs. These elasticities can directly be compared with “macro” elasticities. The numbers in table 6 are most appropriately thought of as characterizing the short run responses⁵.

⁵ In dynamic linear models typically employed when using time series methods, it is straight forward to calculate the long run response using the estimated parameter on the lagged dependent variable, see for example Dargay (2001). This is not possible in a nonlinear model.

Table 6. Effect of changes in income and user costs in 2001

	Income			User costs		
	Pooled	RE	RE SD	Pooled	RE	RE SD
0 car	-0.156	-0.050	-0.017	0.202	0.014	0.112
1 car	0.029	-0.022	-0.028	-0.037	0.086	0.063
2 cars	0.127	0.072	0.045	-0.165	-0.101	-0.175
Car stock elasticity ⁽¹⁾	0.283	0.122	0.062	-0.367	-0.118	-0.293

Note: 'Pooled' is pooled logit without random effects and state dependence. 'RE' is logit with random effects but without state dependence. 'RE SD' is logit with random effects and state dependence. The upper part of the table presents partial effects multiplied by 100.

(1) The elasticity of the total car stock is calculated as the percentage change in the predicted car stock (derived from the changes in probabilities also reported in table 6) following a one percentage change in income/user costs.

The general picture appearing from the income effects in table 6 is that the partial effects get smaller as the level of sophistication of the models increases. The largest effects are found in the pooled multinomial logit model that condition on neither unobserved heterogeneity nor lagged levels of car ownership. The smallest response effects are found in the random effects model with state dependence. The estimates from the static random effects model generally lie between those of the other models, as would be expected. Generally, for the models including random effects, the largest income responses are found for the category holding two cars. This is consistent with the interpretation that the second car has less of a necessity nature (in the jargon of demand analysis). The general conclusion, though, is that income changes have little impact on car ownership in the short run.

The partial effects with respect to user costs do not generally become smaller as the number of conditioning factors increase. The largest responses are found in the pooled model and in the model with random effects and state dependence. The most important effect of including random effects and state dependence is that the partial effect for the category holding no car becomes smaller than

in the pooled model. The pooled model thus exaggerates the extent of downsizing following an increase in user costs.

The short-run income effects in the random effects model with state dependence are considerably smaller than typically found in other studies, e.g. in studies using synthetic panel data, i.e. panel data constructed from repeated cross sections, like Dargay (2001) and Dargay and Vythoulkas (1999). They found short-run income elasticities (based on macro time series methods) ranging from 0.18 to 0.48. In the same studies, long-run income elasticities range from 0.28 to 0.80. This difference is likely to arise because it is not possible to take into account idiosyncratic effects using synthetic panel data. Also other studies based on micro cross-section data, e.g. de Jong (1990) and Ramjerdi and Rand (1992), have found income elasticities at 0.33 and 0.15, respectively. Previous studies based on Danish data (micro cross section) yielded income elasticities at 0.41 (Bjørner, 1999) and from 0.39 to 0.55 (Fosgerau and Nielsen, 2002). As expected these income elasticities are closer to the ones we have found in the pooled logit model.

The car cost (purchase, ownership and variables costs) responses are largest for the category holding two cars, but the general conclusion is that changes in costs have little impact on car ownership in the short run. The estimated responses are in range with what is found in other studies. Dargay (2001) finds a car purchase cost elasticity at -0.13, while Dargay and Vythoulkas (1999) find long-run elasticities with respect to purchase and variable costs at -0.33 and -0.51 (for “middle” levels of income and car ownership). However, it should be recalled that the car cost responses estimated in this study are based on changes in car cost over time for a relatively short period. The estimates could therefore reflect too little variation in the data rather than a genuine behavioural effect, and the estimates should therefore be interpreted with caution.

6. Conclusion

Using a unique panel data set with information on car ownership for 10,565 households observed over the period 1992-2001 we have demonstrated that car ownership status is very persistent. This shows very clearly at the descriptive level, 56% of the households do not change ownership status over the ten-year period where we follow them.

We estimate models of describing ownership status as a function of income, user costs, demographic and geographical characteristics. In the reference model we condition only on the observed characteristics. Next, we condition on unobserved fixed heterogeneity, and finally on both unobserved heterogeneity and lagged ownership status. The two latter models capture idiosyncratic effects, and this can only be done using panel data. The reference model yields estimates that are comparable to those of other studies not based on panel data. The random effects model and the random effects model with state dependence indicate that both unobserved heterogeneity and state dependence are important factors in explaining car ownership in the short run.

The results from these models indicate that responses to changes in income and user costs are much smaller than what was thought based on previous studies. One interesting feature of the dynamic model is that ownership of two cars is more responsive in the short run than ownership of one or no car. This suggests that car holdings of multiple car ownership households respond stronger to changes in incentive. Altogether the general conclusion remains, however, that income and user cost changes have little impact on car ownership in the short run, and that car ownership adjusts very slowly so that policy instruments aiming at reducing car ownership are not likely to be very effective in the short run.

References

- Bjørner, T.B. (1999). Demand for Car Ownership and Car Use in Denmark: a micro econometric model. *International Journal of Transport Economics*, XXVI(3): 377-395.
- Bjørner, T.B. and S. Leth-Petersen (2005). Dynamic Models of Car Ownership at the Household Level. *International Journal of Transport Economics* (forthcoming).
- Dargay, J. (2001). The effect of income on car ownership: evidence of asymmetry. *Transportation Research Part A*, 35: 807-821.
- Dargay, J. and P.C. Vythoulkas (1999). Estimation of a Dynamic Car Ownership Model: a Pseudo-Panel Approach. *Journal of Transport Economics and Policy*, 33(3): 287-302.
- de Jong, G.C. (1990). An Indirect Utility Model of Car Ownership and Use. *European Economic Review*, 34: 971-985.
- Erdem, T. and Sun, B. (2001). Testing for Choice Dynamics in Panel Data. *Journal of Business and Economic Statistics*, 19(2): 142-152.
- Frank, L. (2000). When an Entire Country Is a Cohort. *Science*. 287: 2398-2399.
- Fosgerau, M. and J.E. Nielsen (2002). *Wealth as a Source of Bias in Cross-Sectional Evaluation of Car Ownership*. Discussion Paper (04/07/2002). Danish Transport Research Institute.
- Heckman J.J. (1981). The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process. In: Manski, C.F. and D. McFadden (eds.): *Structural Analysis of Discrete Data with Econometric Applications*, 179-195. MIT Press, Cambridge, MA.
- 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 Koshi, M. (Ed): *Transportation and Traffic Theory*. Elsevier, New York, 477-496.
- Meurs, H. (1993). A Panel Data Switching Regression Model of Mobility and Car Ownership. *Transportation Research A*, 27A: 461-476.
- Ramjerdi, F. and L. Rand (1992). *The National Model System for Private Travel*. Institute of Transport Economics report 150/1992, Oslo.
- Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Wooldridge, J.M. (2002a). *Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity*. The Institute of Fiscal Studies, Department of Economics, UCL, Working paper CWP18/02 (June 2002). [Forthcoming in *Journal of Applied Econometrics* according to homepage of Wooldridge].

Wooldridge, J.M. (2002b). *Econometric Analysis of Cross Section and Panel Data*. The MIT press, Cambridge, MA.

Appendix with summary statistics

	All years					1992					2001				
	Mean	std	min	max		Mean	std	min	max		Mean	std	min	max	
Pear=0	0.1573	0.3641	0	1		0.1993	0.3995	0	1		0.1313	0.3377	0	1	
Pear=1	0.7221	0.4480	0	1		0.6980	0.4592	0	1		0.7357	0.4410	0	1	
Pear=2	0.1206	0.3257	0	1		0.1027	0.3036	0	1		0.1330	0.3396	0	1	
linc	12.467	0.3970	8.6927	15.8293		12.4578	0.3797	8.6927	15.8293		12.4476	0.4256	9.9621	14.8364	
age	50.6183	14.1114	18	95		46.1183	13.8166	18	86		55.1183	13.8166	27	95	
Dchild_m	0.3923	0.4883	0	1		0.4307	0.4952	0	1		0.3370	0.4727	0	1	
Dchild_a	0.1341	0.3408	0	1		0.1357	0.3425	0	1		0.1230	0.3285	0	1	
Workmale	0.6168	0.4862	0	1		0.6536	0.4759	0	1		0.5715	0.4949	0	1	
Workfem	0.5571	0.4967	0	1		0.5715	0.4949	0	1		0.5374	0.4986	0	1	
Unempmale	0.0817	0.2739	0	1		0.1217	0.3270	0	1		0.0526	0.2233	0	1	
Unempfem	0.1209	0.3260	0	1		0.1818	0.3857	0	1		0.0710	0.2568	0	1	
Distm_sr	2.4870	2.6712	0	22.4722		2.6749	2.6277	0	22.2036		2.3343	2.7455	0	22.4722	
Distf_sr	1.7918	2.0084	0	22.5389		1.9137	1.9780	0	21.2132		1.7230	2.0496	0	20.0499	
Distm_miss	0.02526	0.1569	0	1		0.0219	0.1462	0	1		0.0256	0.1578	0	1	
Distf_miss	0.05931	0.2362	0	1		0.0522	0.2223	0	1		0.0544	0.2269	0	1	
Lusc	3.5001	0.0267	3.4663	3.5625		3.4996	0.0109	3.4899	3.5166		3.5446	0.0111	3.5351	3.5625	
Cph	0.2701	0.4440	0	1		0.2754	0.4468	0	1		0.2656	0.4417	0	1	
Town	0.3692	0.4826	0	1		0.3733	0.4837	0	1		0.3685	0.4824	0	1	
Observations		105,650					10,565					10,565			

