). 113 | Discussion Papers | April 1994

Dag G. Wetterwald

Car Ownership and Private Car Use

A Microeconometric Analysis Based on Norwegian Data

Discussion Papers no. 113 • Statistics Norway, April 1994

Dag G. Wetterwald

Car Ownership and Private Car Use
A Microeconometric Analysis Based on Norwegian Data

Abstract:
In this paper we analyze household's car ownership and private car use decisions in a model proposed by de Jong (1990). The model, which incorporates variable and fixed costs of car use, can be used to predict the effects of changes in policy measures on the car stock and aggregate use. The model is estimated on Norwegian household data for 1985.

Keywords: Car ownership and car use, Tobit model, Heckman's two-stage procedure, misspecification.

JEL classification: D12, R41

Acknowledgement: I am greatly indebted to John Kr. Dagsvik who both initiated this project and was my supervisor. Also, I thank Rolf Aaberge, Margaret Simpson and Pam and Frode Johansen for reading earlier versions and giving valuable comments, and Rune Johansen for computational and mathematical assistance.

Correspondence: Dag Wetterwald, Statistics Norway, Research Department,
P.O.Box 8131 Dep., 0033 Oslo. **E-mail:** daw@ssb.no

1 Introduction

Models of the car ownership decisions are numerous in the literature; for example, see Manski and Sherman (1980) and Berkovec and Rust (1985). By contrast, models of the joint decision of car ownership and private car use are, to our knowledge, only studied by Train (1986), de Jong (1990), Institute of Transport Economics (ITE) (1990) and Crawford (1993).

Train (1986) was probably the first to introduce a model in which car ownership and use are treated in a coherent framework. His model explains households' car ownership and use decisions in terms of operating cost, socioeconomic factors and availability of public transport. Whereas Train (1986) allows for multiple cars in each household, de Jong (1990) restricts the car ownership decision to the choice between no car and one car. However, de Jong (1990), unlike Train (1986), realizes that car use is associated with both variable and fixed costs and incorporates both types of costs as explanatory factors. ITE (1990) exploits de Jong's model in developing a model of multiple car ownership and use in households. Crawford (1993) study a model of multiple car ownership and use in households, however he does not incorporate fixed costs.

The point of departure for this paper is the model introduced by de Jong (1990). The incorporation of variable and fixed costs of car use makes it possible to discriminate between the effects of changes in policy measures directed towards fixed and variable costs on aggregate car use. For example, we can study the relative effectiveness of an increase in the fuel tax with an increase in the vehicle excise duty. Also, the effects of policy changes on aggregate car use can be considered both in a short- and a long-term perspective[1]. Therefore, this model may prove useful to policy makers as a means to evaluate how different policy scenarios affect aggregate car use.

de Jong's model has, however, some serious shortcomings. First, it fails to account for the intertemporal nature of car ownership decisions; within the model, the car ownership decision is independent of previous ownership decisions. Suppose we have two households, one that possessed a car last period and another that did not. According to de Jong's model, the probability of currently owning a car will be equal for these households, *ceteris paribus.* Second, fixed and variable costs of use are assumed constant across households. Since these costs depend on characteristics of the car owned, such as make, vintage, size, fuel economy,

[1]In our concern, long-term means the effect on aggregate car use due to changes in the car stock.

it is unlikely that households face identical fixed and variable costs.

de Jong (1990), using Dutch household data, finds that all but one explanatory variable have significant effect on households' car ownership and use decisions, and that the estimated parameters have the predicted signs. However, as de Jong points out, the estimated standard deviations of the error terms "are rather large revealing a substantial unexplained variation." Despite this fact, de Jong uses the estimated model in microsimulation studies, without recognizing the possibility that it will have very low explanatory power.

The purpose of this paper is to study how well de Jong's model explains household's car ownership and private car use decisions. Recalling the above mentioned shortcomings, there are reasons to believe that this model fails to explain the car ownership and use decisions very well. Moreover, the empirical model of de Jong fails to include some essential explanatory variables, such as availability of public transport and dependency on the car. Thus, there might be misspecification due to missing variables.

We pursue the following procedure to study the explanatory power of de Jong's model. Since we do not possess the dataset employed by de Jong, we commence by reestimating de Jong's model on Norwegian household data and compare our results with those of de Jong. Unless the estimation results are similar, we can not determine whether our conclusions are due to the dataset employed or the econometric model. We compute a goodness of fit measure for the car use equation and, based on this and the estimation results, we offer some conclusions about the model. The next stage is to relax some of the theoretical restrictions imposed on the parameters in de Jong's model. The substantial difference between our generalized model and de Jong's model is that the generalized model treats car ownership and car use as separate decisions, whereas de Jong's model treats them as simultaneous decisions. Separating the decisions allows us to draw separate inferences at each decision stage. Note that we cannot justify the generalized model theoretically within the theoretical framework developed by de Jong.

The structure of the paper is as follows. Sections 2 and 3 introduce the theoretic and econometric models, respectively. Section 4 presents the data and section 5, the estimation methods and results. A summary and discussion of the results are given in Section 6.

2 Theoretical model

The econometric model discussed in this paper is derived from the neoclassical theory of consumer behavior. Thus, we assume that the preferences over consumption bundles, consisting of annual driving distance and consumption of all other goods and services, can be represented by a strictly increasing and quasi-concave utility function. Specifically, let A be a household's private car use (measured in 100 kilometers per year) and X be the household's annual consumption of all other goods and services. The prices for one unit of A and X are v and 1, respectively. v is interpreted as the variable cost of driving 100 kilometers. In addition, each household faces fixed costs, C, of owning a car. Variable costs comprise fuel and oil costs, and some part of maintenance, repair, depreciation and insurance costs. The other part of maintenance, repair, depreciation and insurance costs as well as items such as road tax, registration costs, vehicle excise duty comprise fixed costs. Thus, the household's budget can be expressed as:

$$B = \begin{cases} ((A, X); Y \geq C + X + vA, & \text{if } A > 0), \\ ((A, X); Y \geq X, & \text{if } A = 0) \end{cases} \tag{1}$$

where Y is the household's annual disposable income. Note that the budget set is non-convex due to the fixed cost of car ownership. We should, at this point, emphasize some of the restrictions imposed on the model by this specification. First, the setting above cannot account for differences in car costs, since car ownership and use are treated as homogeneous commodities. Thus, while the prices v and C associated with realized choices are assumed constant across households in this model, households can in reality choose among a variety of types of cars with different costs, implying that the realized costs vary across households. Second, there is no gain in utility from pure car ownership. Ownership is merely a prerequisite of usage. Third, the model considers one period only implying that current decisions are independent of the past.

Define the conditional indirect utility:

$$V_1 = V_1(v, C, Y) = \max_{(A,X) \in B, A > 0} U(A, X) \tag{2}$$

and let:

$$V_0 = U(0, Y). \tag{3}$$

V_1 and V_0 are the utilities of owning and not owning a car, which means that the household will chose to own a car if $V_1 > V_0$. If this is the case the intensity of use is determined by

Roy's identity:

$$-\frac{\partial V_1/\partial v}{\partial V_1/\partial Y} = A. \tag{4}$$

We should emphasize that, although many households own more than one car, this model specification excludes that possibility. Although ITE (1990) has extended the model to a situation where each household may own two cars, throughout this paper we shall only consider the case where each household owns at most one car.

Consider next the problem of specifying a tractable functional form of the utility function. One possibility is to specify the conditional indirect utility V_1 and apply Roy's identity to derive the demand function for usage. Rather than starting from the indirect or direct utility function, we follow de Jong (1990) in specifying the household's demand function for kilometers and applying Roy's identity to find the indirect utility function[2].

Thus, following de Jong, we assume that the demand for kilometers in a household that owns a car is given by:

$$\ln A = \alpha \ln(Y - C) - \beta v + \delta, \tag{5}$$

where α and β are unknown parameters assumed to be non-negative and δ contains socioeconomic and demographic characteristics of the household. de Jong's argument for using this functional form is that it performed well in an empirical analysis by de Jong and Cramer (1987).

From Roy's identity and (5) we obtain the following partial differential equation:

$$-\frac{\partial V_1/\partial v}{\partial V_1/\partial Y} = A = (Y - C)^{\alpha} e^{-\beta v + \delta}. \tag{6}$$

The solution of (6) has the form

$$V_1(v, C, Y) = \psi(v, Y - C) \equiv \frac{1}{\beta} e^{\delta - \beta v} + \frac{1}{1-\alpha}(Y - C)^{1-\alpha} \tag{7}$$

By letting the variable cost, v, in the expression for V_1 approach infinity, an expression for V_0 can be obtained. In general, this limit does not exist. However, in our case, with the assumptions that the marginal utilities of both goods are positive and β is non-negative, the limit exists and is equal to the utility of not owning a car. Formally we can express this as:

$$U(0, Y) = \lim_{v \to \infty} V_1(v, C, Y) = \frac{1}{1-\alpha} Y^{1-\alpha}. \tag{8}$$

[2]Whether we specify the indirect utility function or the demand function is of course a matter of convenience because there is a one-to-one correspondance between the two.

The last equality follows from (7). We can see from (8) that the household's utility in the case where it does not own a car depends only on α and Y. Note that (8) does not contain C. This is due to the fact that only households that own a car incur the fixed cost.

3 Econometric specification

To this point we have considered a single, unspecified household. In order to distinguish between households we add a subscript i to identify the ith household. Further, we have not addressed the uncertainty with which we, as econometricans, are faced. There are various types of unobservables in this model, such as measurement errors and taste differences between households. We shall now explicitly take into consideration these unobservables by incorporating two random variables in the model. The first one, which accounts for unobservables in the households utility and taste differences between the households, is incorporated in the model via δ in (5). We assume that

$$\delta_i = \gamma S_i + \varepsilon_i \tag{9}$$

where γ is a coefficient vector, S_i is a vector containing observations on socioeconomic and demographic variables for the ith household and ε_i is a random variable that we assume is IID$(0, \sigma_\varepsilon)$. The second random variable, defined as the difference between intended and observed car use, accounts for measurement errors and factors that affect the driving decision after the car ownership decision is made, such as unanticipated driving and factors on the supply side of car use, for example availability of road infrastructure. Formally, we state the relationship between intended and observed car use as

$$K_i = \ln A_i + \omega_i, \tag{10}$$

where K_i is the logarithm of observed annual car use and ω_i is a random variable that we assume is IID$(0, \sigma_\omega)$ and independent of ε_i. Note that only ε_i enters the indirect utility function, while both ε_i and ω_i enter the expression for observed annual car use.

4 Data

We have estimated the model using the Norwegian Survey of Consumer Expenditure (Statistics Norway) data set of 1985. The data set contains figures on disposable income, number of cars owned, annual car use and several household characteristics. An explanation of the socioeconomic and demographic variables used is given below. The data set contains observations on 1559 Norwegian households. From this data set we have excluded 263 households that own two or more cars, 34 households with one car but for which data on car use is missing and 4 households for which data on income is missing. The resulting sample consists of 1258 households. The sample contains 365 observations on households without a car and 893 households with only one car.

The following definitions are used for the socioeconomic and demographic variables in the model:

DF: Dummy for female head of household.

DS: Dummy for rural area.

DT: Dummy for urban area, except Oslo, Bergen and Trondheim.

AGEH: Age of head of household measured in five-year intervals (where class 1: ≤ 20 year; class 2: $21-25$ year; ...; class 11: 65 and more).

LNUMB: Logarithm of $10 \times$ household size.

The S_i vector contains a constant and observations on these five variables.

In table 1 we have given summary statistics of the variables used.

Table 1

Summary statistics

	Mean	St.deviation
Disposable income before interest deductions in kroner	147823	76231
Usage in kilometers (car owning households only)	11168	6014
DF	0.23	0.42
DS	0.29	0.45
DT	0.53	0.50
AGEH	6.90	3.22
LNUMB	2.63	1.36

Data on v and C are not reported here since we shall assume that they are constant across households. The argument behind these assumptions is that every household, in principle, faces the same exogenously given price vector. Due to missing information on variable and fixed costs we have used values on v and C computed by ITE (1990). These computations are based on figures for 1985. The values are:

$$C = 9204 \text{ kroner per year}$$
$$v = 76.37 \text{ kroner pr. } 100 \text{ kilometers.}$$

5 Estimation methods and results

The econometric model derived in section 3 is classified by Amemiya (1989) as a Type II Tobit model. In this section we shall discuss the estimation methods used and present the estimation results. As will be seen later in the section, we will also introduce and estimate a more general version of the model.

Full information maximum likelihood estimation

A Type II Tobit model has two dependent variables, one binary and one censored. In our case the binary variable refers to car ownership or lack of ownership and the censored variable is annual car use measured in 100 kilometers. From Amemiya (1990) it follows that the likelihood function of this model is:

$$\begin{aligned} L = &\prod_{n_0} P\left\{\psi_i(v, Y-C) \leq U_i(0,Y)\right\} \times \\ &\prod_{n_p} P\left\{\psi_i(v, Y-C) > U_i(0,Y)\right\} f\left\{K_i; \psi_i(v, Y-C) > U_i(0,Y)\right\}, \end{aligned} \tag{11}$$

where n_p and n_0 are the number of households with and without private car, respectively. $f\{x;y\}$ stands for the conditional density function of x, given y. Under the assumptions that both ε and ω are normally distributed we have, according to de Jong (1990), that:

$$\begin{aligned} &P\left\{\psi_i(v, Y-C) \leq U_i(0,Y)\right\} = \\ &\Phi\left\{\frac{\ln\left[Y_i^{1-\alpha} - (Y_i - C)^{1-\alpha}\right] - \ln(1-\alpha) + \ln\beta - \gamma S_i + \beta v}{\sigma_\varepsilon}\right\} \equiv \Phi_i \end{aligned} \tag{12}$$

and

$$P\{\psi_i(v,Y-C)>U_i(0,Y)\}\,f\{K_i;\psi_i(v,Y-C)>U_i(0,Y)\} =$$

$$\left\{1-\Phi\left[\frac{N_i-\frac{\sigma_\varepsilon^2}{\sigma_u^2}(K_i-M_i)}{\sigma_\varepsilon\left(1-\frac{\sigma_\varepsilon^2}{\sigma_u^2}\right)^{1/2}}\right]\right\}\left\{\frac{1}{\sigma_u}\cdot\phi\left[\frac{K_i-M_i}{\sigma_u}\right]\right\} \tag{13}$$

where $\Phi(\cdot)$ is the standard Normal probability function, $\phi(\cdot)$ is the standard Normal density function, N_i is the numerator in (12), $u = \varepsilon + \omega$ and $M_i = \alpha\ln(Y_i - C) - \beta v + \gamma S_i$. (12) is an expression for the probability that household i does not own a car and (13) is the car owning households' contribution to the likelihood function. From (11), (12) and (13) we obtain the following log-likelihood function.

$$\ln L = \sum_{n_0}\ln\Phi_i - n_p\ln\sigma_u - \frac{1}{2\sigma_u^2}\sum_{n_p}(K_i-M_i)^2 + \sum_{n_p}\ln(1-\Phi_i^*) - n_p\ln\sqrt{2\pi}, \tag{14}$$

where

$$\Phi_i^* \equiv \Phi\left[\frac{N_i-\frac{\sigma_\varepsilon^2}{\sigma_u^2}(K_i-M_i)}{\sigma_\varepsilon\left(1-\frac{\sigma_\varepsilon^2}{\sigma_u^2}\right)^{1/2}}\right].$$

The estimates were found by maximizing (14) using the Maximum Likelihood application of the statistical package GAUSS, and are given in table 2. In table 2 we have included the estimates obtained by de Jong on Dutch data. Note that de Jong (1990) uses $C = 2536$ guilders per year and $v = 21.09$ guilders per 100 kilometers.

Table 2 shows that our results are quite similar to de Jong's. Moreover, in both studies the estimated standard deviation, σ_u, is rather large compared to mean of the censored variable, K_i (4.62 in our sample). The elasticity of use with respect to variable costs, $-\beta v$, is 0.65 in de Jong (1990) and 0.61 in our study; the reduced income, $Y_i - C$, elasticity of use, α, is 0.33 in de Jong and 0.177 in our study.

Microsimulations revealed that the model predicts car ownership decisions correctly in approximately 70% of the cases. The R^2 statistic, which provides a goodness of fit measure for the demand function for kilometers, equals 0.009. The extremely low value of the R^2 is somewhat surprising even in light of the considerable amount of unexplained variation revealed by the estimated standard deviation. This forces us to question the validity of the model and to conclude that the model explains the car ownership decision fairly well, but does not explain the demand for kilometers. The lack of explanatory power of the demand function for kilometers may indicate that the theoretical and functional form assumptions are too restrictive and that crucial variables are missing in the demand function for kilometers. In the rest of the paper we shall discuss this in more detail.

Table 2

Results from maximization of (14). (t-values in parentheses)

Variable			de Jong	
$\ln(Y_i - C)$, (α)	0.1770	(5.26)	0.33	(7.61)
v, $(-\beta)$	−0.0080	(−28.84)	−0.031	(−26.22)
LNUMB	0.0009	(0.05)	0.016	(0.50)
AGEH	−0.0179	(−4.99)	−0.016	(−4.13)
DF	−0.1140	(−4.73)	−0.094	(−1.94)
DT	0.0673	(2.99)		
DS	0.0517	(2.12)		
DA†			0.048	(−0.97)
CONSTANT	3.0470	(8.01)	1.80	(3.89)
σ_ε	0.1811		0.26	
σ_ω	0.7360		0.50	
σ_u	0.7584		0.57	
Log-likelihood per observation	−0.5896		−0.394	

†DA represents the dummy for farmer head of household.

Two-stage estimation

To determine whether it is likely that the theoretical and the functional form restrictions are too stringent we generalize the model to one in which all the parameters may assume different values, that is there are no restrictions on any of the parameters between the two decision stages. We should note that this generalized model cannot be justified theoretically by the model developed in Sections 2 and 3. However, since our purpose at this stage is merely to detect the weakness of the original model rather than to propose a new model, we proceed as follows. First, we estimate the parameters in the generalized model and compare these with the estimated parameters of the original model. Then, based on this comparison, we discuss the reasons for the lack of explanatory power of the demand function for kilometers in the original model.

At first glance, the most convenient way to estimate the generalized model is to replace (α, β, γ) in (14) by $(\alpha_1, \beta_1, \gamma_1)$ in N_i and by $(\alpha_2, \beta_2, \gamma_2)$ in M_i. However, this will not work since v is constant over households. One solution is to assume that $\beta_1 = \beta_2 = \beta$ and that the intercept terms in N_i and M_i are equal. Even with these restrictions we are not able to estimate the model using the ML procedure because the iteration procedure fails to converge. Rather than pursuing the Maximum Likelihood estimation procedure any further we shall estimate the model using a two-stage estimation procedure.

The two-stage estimation procedure presented here is a version of Heckman's two-stage estimation method. In our context, this method is implemented as follows: We estimate the structural equation for the intensity of use by OLS but, since this equation is defined for car owning households only, we exclude observations on those households without a car. One obvious problem, then, is how to account for the sample selection bias that results. We study this problem for two different distributions of ε, the normal and the logistic c.d.f. To this end, we start by taking the expectation of K_i, conditional on car ownership, in (10). We find that:

$$
\begin{aligned}
\mathrm{E}(K_i \mid \varepsilon_i > N_i) &= \mathrm{E}(\ln A_i + \omega_i \mid \varepsilon_i > N_i) \\
&= \alpha_2 \ln(Y_i - C) - \beta v + \gamma_2 S_i + \mathrm{E}(\varepsilon_i \mid \varepsilon_i > N_i). \qquad (15)
\end{aligned}
$$

Since ε_i is symmetrically distributed around zero, we know that the following must be true

$$
\mathrm{E}(\varepsilon_i \mid \varepsilon_i > N_i) = \frac{1}{1 - F(N_i)} \int_{N_i}^{\infty} \varepsilon f(\varepsilon)\, d\varepsilon > 0, \qquad (16)
$$

where $F(\cdot)$ and $f(\cdot)$ are the distribution function and density function, respectively. Furthermore, we find that

$$
\mathrm{E}(\varepsilon_i \mid \varepsilon_i > N_i) = \begin{cases} \sigma_\varepsilon \left(\dfrac{\phi(\mathcal{N}_i^*)}{1 - \Phi(\mathcal{N}_i^*)} \right), & \text{if } \varepsilon_i \text{ i.i.d. normal} \\[2ex] \sigma_\varepsilon \left(\dfrac{-\mathcal{N}_i^* \Lambda(\mathcal{N}_i^*) - \ln(1 - \Lambda(\mathcal{N}_i^*))}{1 - \Lambda(\mathcal{N}_i^*)} \right), & \text{if } \varepsilon_i \text{ i.i.d logistic,} \end{cases} \qquad (17)
$$

where $\mathcal{N}_i^* = N_i / \sigma_\varepsilon$ and $\Lambda(\cdot)$ is the standard logistic distribution function. If we let $\lambda(\mathcal{N}_i^*) = \mathrm{E}(\varepsilon_i \mid \varepsilon_i > N_i)/\sigma_\varepsilon$, we can rewrite (15) as

$$
\mathrm{E}(K_i \mid \varepsilon_i > N_i) = \alpha_2 \ln(Y_i - C) - \beta v + \gamma_2 S_i + \sigma_\varepsilon \lambda(\mathcal{N}_i^*) \qquad (18)
$$

or

$$
K_i = \alpha_2 \ln(Y_i - C) - \beta v + \gamma_2 S_i + \sigma_\varepsilon \lambda(\mathcal{N}_i^*) + \eta_i, \qquad (19)
$$

where $\eta_i = K_i - \mathrm{E}(\varepsilon_i \mid \varepsilon_i > N_i)$. From (19) we see that if we introduce $\lambda(\mathcal{N}_i^*)$ as an extra regressor we are able to control for sample selection. But since $\mathcal{N}_i^*$ is unknown we must estimate it to use $\lambda(\mathcal{N}_i^*)$ as a regressor. This is done by estimating the parameters in $\mathcal{N}_i^*$ using a probit or logit estimation of the car ownership decision and replacing the parameters by their corresponding estimates. The two-stage estimators are then the OLS estimators of

$$K_i = \alpha_2 \ln(Y_i - C) - \beta v + \gamma_2 S_i + \sigma_\varepsilon \lambda(\widehat{\mathcal{N}_i^*}) + \eta_i, \tag{20}$$

where $\widehat{\mathcal{N}_i^*}$ is the estimate of $\mathcal{N}_i^*$.

Before we can present the results there are two problems to be solved. The first concerns the discrete choice estimation. As we see from (12), the probability that a household does not own a car, the numerator N_i, cannot be transformed to a linear function in α. This prevents us from using a standard Probit/Logit estimation package. However, if we consider the first order Taylor expansion of N_i around $\alpha_1 = 0$ we find that:

$$\begin{aligned} \ln\left(\frac{Y_i^{1-\alpha_1} - (Y_i - C)^{1-\alpha_1}}{1-\alpha_1}\right) &\approx \left(\frac{(Y_i - C)\ln(Y_i - C) - Y_i \ln Y_i}{C} + 1\right)\alpha_1 + \ln C \\ &\equiv \mathcal{F}(Y_i, C)\alpha_1 + \ln C. \qquad (21) \end{aligned}$$

Our simulation experiments indicate that this approximation is extremely good for $\alpha_1 \in [0,1)$. Substitution of this expression in (12) enables us to estimate the fraction $\alpha_1/\sigma_\varepsilon$ by a standard Probit/Logit estimation package.

The second problem we face in this two-stage estimation procedure is that v is constant across the sample, which prevents us from estimating β. However, we can cirumvent this problem by defining:

$$\left(\frac{\kappa}{\sigma_\varepsilon}\right) \equiv \frac{\ln\beta + \beta v - \gamma_0 + \ln C}{\sigma_\varepsilon}, \tag{22}$$

where γ_0, assumed to be equal in the two stages, is the intercept term in the model. Then we can use this definition in (20) to obtain an equation in which all the parameters in stage two can be estimated. The resulting equation is:

$$K_i - \ln C = \alpha_2 \ln(Y_i - C) + \ln\beta + \bar{\gamma}_2 \bar{S}_i + \sigma_\varepsilon \left(\lambda(\widehat{\mathcal{N}_i^*}) - \left(\frac{\kappa}{\sigma_\varepsilon}\right)\right) + \eta_i, \tag{23}$$

where the bar indicates that the intercept term has been subtracted. Thus, $\ln\beta$ is estimated as the intercept term in (23). Finally, we have to substitute $\kappa/\sigma_\varepsilon$ with its estimated value from the first stage.

Table 3

First stage of the two-stage estimation. (t-values in parentheses)

Variable	PROBIT		LOGIT	
$\mathcal{F}(Y_i, C)$, $(\frac{\alpha_1}{\sigma_\varepsilon})$	0.994	(8.82)	1.849	(8.85)
DF	−0.637	(−6.07)	−1.078	(−6.00)
DS	0.272	(2.11)	0.515	(2.29)
DT	0.382	(3.32)	0.706	(3.49)
AGEH	−0.093	(−6.53)	−0.157	(−6.25)
LNUMB	−0.029	(−0.25)	−0.096	(−0.46)
CONSTANT				
$(\ln\beta + \beta v + \ln C - \gamma_0)/\sigma_\varepsilon$	10.368	(8.94)	19.358	(8.95)
Log-likelihood				
per observation	−0.441		−0.438	
Percent correctly				
predicted	80.92		81.24	
McFadden's pseudo-R^2	0.2678		0.2730	
Cragg and Uhler's pseudo-R^2	0.1327		0.1366	

The results from estimation of the Probit/Logit models are given in table 3 and the results from OLS estimation of (23) are given in table 4. The results in table 3 support our earlier results that the model explains the car ownership decision fairly well. In fact, if we divide the estimates from the original model (table 2) by the estimate of σ_ε in table 2 and compare the resulting values with the corresponding estimates in the probit estimation, we find that the values are almost equal, except for LNUMB. Also, LNUMB is the only insignificant variable in the first stage. Table 4 shows that only two variables have significant effect on usage, namely the sample correction variable and the intercept ($\ln\beta$). In addition, the estimate on the sample correction variable, σ_ε, is negative. Based on the results in table 4, we thus conclude that the assumed demand function kilometers does not explain the variation in observed car use well.

Although our purpose here was to check the nature of misspcification, we have in reality only investigated one possible misspecification, namely the imposed restrictions on the parameters in the model of de Jong. However, the analysis has given us some insight to the

Table 4

Second stage: OLS on (23) with Probit or Logit in first stage. (t-values in parentheses)

Variable	Probit		Logit	
$\ln(Y_i - C)$, (α_2)	-0.228	(-1.37)	-0.149	(-1.04)
DF	0.221	(0.163)	0.145	(1.27)
DS	-0.041	(-0.46)	-0.021	(-0.25)
DT	-0.118	(-1.32)	-0.088	(-1.05)
AGEH	0.004	(0.21)	0.007	(0.48)
LNUMB	0.049	(0.74)	0.056	(0.84)
$\lambda(\widehat{\mathcal{N}_i^*})$	-0.923	(-2.40)	-0.396	(-2.27)
$\ln\beta$	-11.276	(-5.01)	-10.360	(-5.22)
σ_η	0.741		0.741	
Number of obs.	893		893	
R^2	0.043		0.042	
$\bar{R}^2$	0.035		0.034	

problems with which we are faced in modelling car ownership and use in households. In the next section we suggest some factors that should be considered.

6 Conclusion

In this paper we have estimated the model of car ownership and private car use proposed by de Jong (1990) on Norwegian household data. Although we obtained significant estimates for all but one parameter. However, the low R^2 value for the demand function for kilometers suggested that the car use decision was poorly explained by the model. We conjectured that the original model was misspecified and, to check the nature of misspecification we estimated a more general model where some of the theoretical restrictions imposed on the parameters by de Jong (1990) were relaxed. The estimation results revealed, however, that the low R^2 value prevailed, suggesting more severe misspecification.

We conclude that de Jong's model does not provide a satisfying description of households' joint decisions on car ownership and use. This may be related to the following shortcommings. First, the model, in its present form, does not account for the possibility of owning more than one car. Second, obvious relevant variables are missing from the model, such as availability of public transportation and dependency on the car. Third, cars are heterogeneous goods; when, as in de Jong, automobiles and usage are treated as homogeneous commodities, we cannot account for the possibility that low income households own cheap cars and high income households own expensive cars. For example, in this approach no household may shift from owning an expensive car to a cheap car if the costs of car ownership and use increase relative to the costs of other goods or to income. Fourth, the dynamic nature of the car ownership decision is not accounted for. And last, no care is taken for the possibility that in some occupations private car use is necessary to do the job, implying endogenous sample selection.

References

Amemiya, T. (1989): *Advanced econometrics*, Basil Blackwell, Oxford.

Berkovec, J. and J. Rust (1985): "A nested logit model of automobile holdings for one vehicle households", *Transportation Research*, Vol. 19B, 275 - 285.

Crawford, I. (1993): "A micro-model of the ownership and use of private cars", *The Institute for Fiscal Studies, Working Paper Series*, No. W93/15.

Dubin, J.A. and D.L. McFadden (1984): "An econometric analysis of residential electric appliance holdings and consumption", *Econometrica*, 52, 345 - 362.

Heckman, J.J. (1979): "Sample selection bias as a specification error", *Econometrica*, 47, 153 - 161.

Institute of Transport Economics (1990): "A model system to predict fuel use and emissions from private travel in Norway from 1985 to 2025", Final Report to the Norwegian Ministry of Transport for the "Climate Project", Hague Consulting Group (Den Haag).

Johnson, N.L. and S. Kotz (1970): *Continuous univariate distributions 1*, Houghton Mifflin, Boston.

Jong, G.C. de (1990): "An indirect utility model of car ownership and private car use", *European Economic Review*, 971 - 985.

Jong, G.C. de and J.S. Cramer (1987): "A censored regression model of private car use", *Institute of Actuarial Science and Econometrics, University of Amsterdam*, Report AE9/87.

Maddala, G.S. (1983): *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge.

Manski, C.F. and L. Sherman (1980): "An empirical analysis of household choice among motor vehicles", *Transportation Research*, 14A, 349-366.

Train, K. (1986): *Qualitative choice analysis. (Theory, econometrics and an application to automobile demand)*, The MIT Press, Cambridge, MA.

Issued in the series Discussion Papers

No. 1 *I. Aslaksen and O. Bjerkholt (1985)*: Certainty Equivalence Procedures in the Macroeconomic Planning of an Oil Economy

No. 3 *E. Biørn (1985)*: On the Prediction of Population Totals from Sample surveys Based on Rotating Panels

No. 4 *P. Frenger (1985)*: A Short Run Dynamic Equilibrium Model of the Norwegian Production Sectors

No. 5 *I. Aslaksen and O. Bjerkholt (1985)*: Certainty Equivalence Procedures in Decision-Making under Uncertainty: An Empirical Application

No. 6 *E. Biørn (1985)*: Depreciation Profiles and the User Cost of Capital

No. 7 *P. Frenger (1985)*: A Directional Shadow Elasticity of Substitution

No. 8 *S. Longva, L. Lorentsen and Ø. Olsen (1985)*: The Multi-Sectoral Model MSG-4, Formal Structure and Empirical Characteristics

No. 9 *J. Fagerberg and G. Sollie (1985)*: The Method of Constant Market Shares Revisited

No. 10 *E. Biørn (1985)*: Specification of Consumer Demand Models with Stochastic Elements in the Utility Function and the first Order Conditions

No. 11 *E. Biørn, E. Holmøy and Ø. Olsen (1985)*: Gross and Net Capital, Productivity and the form of the Survival Function. Some Norwegian Evidence

No. 12 *J.K. Dagsvik (1985)*: Markov Chains Generated by Maximizing Components of Multidimensional Extremal Processes

No. 13 *E. Biørn, M. Jensen and M. Reymert (1985)*: KVARTS - A Quarterly Model of the Norwegian Economy

No. 14 *R. Aaberge (1986)*: On the Problem of Measuring Inequality

No. 15 *A.-M. Jensen and T. Schweder (1986)*: The Engine of Fertility - Influenced by Interbirth Employment

No. 16 *E. Biørn (1986)*: Energy Price Changes, and Induced Scrapping and Revaluation of Capital - A Putty-Clay Model

No. 17 *E. Biørn and P. Frenger (1986)*: Expectations, Substitution, and Scrapping in a Putty-Clay Model

No. 18 *R. Bergan, Å. Cappelen, S. Longva and N.M. Stølen (1986)*: MODAG A - A Medium Term Annual Macroeconomic Model of the Norwegian Economy

No. 19 *E. Biørn and H. Olsen (1986)*: A Generalized Single Equation Error Correction Model and its Application to Quarterly Data

No. 20 *K.H. Alfsen, D.A. Hanson and S. Glomsrød (1986)*: Direct and Indirect Effects of reducing SO_2 Emissions: Experimental Calculations of the MSG-4E Model

No. 21 *J.K. Dagsvik (1987)*: Econometric Analysis of Labor Supply in a Life Cycle Context with Uncertainty

No. 22 *K.A. Brekke, E. Gjelsvik and B.H. Vatne (1987)*: A Dynamic Supply Side Game Applied to the European Gas Market

No. 23 *S. Bartlett, J.K. Dagsvik, Ø. Olsen and S. Strøm (1987)*: Fuel Choice and the Demand for Natural Gas in Western European Households

No. 24 *J.K. Dagsvik and R. Aaberge (1987)*: Stochastic Properties and Functional Forms of Life Cycle Models for Transitions into and out of Employment

No. 25 *T.J. Klette (1987)*: Taxing or Subsidising an Exporting Industry

No. 26 *K.J. Berger, O. Bjerkholt and Ø. Olsen (1987)*: What are the Options for non-OPEC Countries

No. 27 *A. Aaheim (1987)*: Depletion of Large Gas Fields with Thin Oil Layers and Uncertain Stocks

No. 28 *J.K. Dagsvik (1987)*: A Modification of Heckman's Two Stage Estimation Procedure that is Applicable when the Budget Set is Convex

No. 29 *K. Berger, Å. Cappelen and I. Svendsen (1988)*: Investment Booms in an Oil Economy - The Norwegian Case

No. 30 *A. Rygh Swensen (1988)*: Estimating Change in a Proportion by Combining Measurements from a True and a Fallible Classifier

No. 31 *J.K. Dagsvik (1988)*: The Continuous Generalized Extreme Value Model with Special Reference to Static Models of Labor Supply

No. 32 *K. Berger, M. Hoel, S. Holden and Ø. Olsen (1988)*: The Oil Market as an Oligopoly

No. 33 *I.A.K. Anderson, J.K. Dagsvik, S. Strøm and T. Wennemo (1988)*: Non-Convex Budget Set, Hours Restrictions and Labor Supply in Sweden

No. 34 *E. Holmøy and Ø. Olsen (1988)*: A Note on Myopic Decision Rules in the Neoclassical Theory of Producer Behaviour, 1988

No. 35 *E. Biørn and H. Olsen (1988)*: Production - Demand Adjustment in Norwegian Manufacturing: A Quarterly Error Correction Model, 1988

No. 36 *J.K. Dagsvik and S. Strøm (1988)*: A Labor Supply Model for Married Couples with Non-Convex Budget Sets and Latent Rationing, 1988

No. 37 *T. Skoglund and A. Stokka (1988)*: Problems of Linking Single-Region and Multiregional Economic Models, 1988

No. 38 *T.J. Klette (1988)*: The Norwegian Aluminium Industry, Electricity prices and Welfare, 1988

No. 39 *I. Aslaksen, O. Bjerkholt and K.A. Brekke (1988)*: Optimal Sequencing of Hydroelectric and Thermal Power Generation under Energy Price Uncertainty and Demand Fluctuations, 1988

No. 40 *O. Bjerkholt and K.A. Brekke (1988)*: Optimal Starting and Stopping Rules for Resource Depletion when Price is Exogenous and Stochastic, 1988

No. 41 *J. Aasness, E. Biørn and T. Skjerpen (1988)*: Engel Functions, Panel Data and Latent Variables, 1988

No. 42 *R. Aaberge, Ø. Kravdal and T. Wennemo (1989)*: Unobserved Heterogeneity in Models of Marriage Dissolution, 1989

No. 43 *K.A. Mork, H.T. Mysen and Ø. Olsen (1989)*: Business Cycles and Oil Price Fluctuations: Some evidence for six OECD countries. 1989

No. 44 *B. Bye, T. Bye and L. Lorentsen (1989)*: SIMEN. Studies of Industry, Environment and Energy towards 2000, 1989

No. 45 *O. Bjerkholt, E. Gjelsvik and Ø. Olsen (1989)*: Gas Trade and Demand in Northwest Europe: Regulation, Bargaining and Competition

No. 46 *L.S. Stambøl and K.Ø. Sørensen (1989)*: Migration Analysis and Regional Population Projections, 1989

No. 47 *V. Christiansen (1990)*: A Note on the Short Run Versus Long Run Welfare Gain from a Tax Reform, 1990

No. 48 *S. Glomsrød, H. Vennemo and T. Johnsen (1990)*: Stabilization of Emissions of CO_2: A Computable General Equilibrium Assessment, 1990

No. 49 *J. Aasness (1990)*: Properties of Demand Functions for Linear Consumption Aggregates, 1990

No. 50 *J.G. de Leon (1990)*: Empirical EDA Models to Fit and Project Time Series of Age-Specific Mortality Rates, 1990

No. 51 *J.G. de Leon (1990)*: Recent Developments in Parity Progression Intensities in Norway. An Analysis Based on Population Register Data

No. 52 *R. Aaberge and T. Wennemo (1990)*: Non-Stationary Inflow and Duration of Unemployment

No. 53 *R. Aaberge, J.K. Dagsvik and S. Strøm (1990)*: Labor Supply, Income Distribution and Excess Burden of Personal Income Taxation in Sweden

No. 54 *R. Aaberge, J.K. Dagsvik and S. Strøm (1990)*: Labor Supply, Income Distribution and Excess Burden of Personal Income Taxation in Norway

No. 55 *H. Vennemo (1990)*: Optimal Taxation in Applied General Equilibrium Models Adopting the Armington Assumption

No. 56 *N.M. Stølen (1990)*: Is there a NAIRU in Norway?

No. 57 *Å. Cappelen (1991)*: Macroeconomic Modelling: The Norwegian Experience

No. 58 *J. Dagsvik and R. Aaberge (1991)*: Household Production, Consumption and Time Allocation in Peru

No. 59 *R. Aaberge and J. Dagsvik (1991)*: Inequality in Distribution of Hours of Work and Consumption in Peru

No. 60 *T.J. Klette (1991)*: On the Importance of R&D and Ownership for Productivity Growth. Evidence from Norwegian Micro-Data 1976-85

No. 61 *K.H. Alfsen (1991)*: Use of Macroeconomic Models in Analysis of Environmental Problems in Norway and Consequences for Environmental Statistics

No. 62 *H. Vennemo (1991)*: An Applied General Equilibrium Assessment of the Marginal Cost of Public Funds in Norway

No. 63 *H. Vennemo (1991)*: The Marginal Cost of Public Funds: A Comment on the Literature

No. 64 *A. Brendemoen and H. Vennemo (1991)*: A climate convention and the Norwegian economy: A CGE assessment

No. 65 *K. A. Brekke (1991)*: Net National Product as a Welfare Indicator

No. 66 *E. Bowitz and E. Storm (1991)*: Will Restrictive Demand Policy Improve Public Sector Balance?

No. 67 *Å. Cappelen (1991)*: MODAG. A Medium Term Macroeconomic Model of the Norwegian Economy

No. 68 *B. Bye (1992)*: Modelling Consumers' Energy Demand

No. 69 *K. H. Alfsen, A. Brendemoen and S. Glomsrød (1992)*: Benefits of Climate Policies: Some Tentative Calculations

No. 70 *R. Aaberge, Xiaojie Chen, Jing Li and Xuezeng Li (1992)*: The Structure of Economic Inequality among Households Living in Urban Sichuan and Liaoning, 1990

No. 71 *K.H. Alfsen, K.A. Brekke, F. Brunvoll, H. Lurås, K. Nyborg and H.W. Sæbø (1992)*: Environmental Indicators

No. 72 *B. Bye and E. Holmøy (1992)*: Dynamic Equilibrium Adjustments to a Terms of Trade Disturbance

No. 73 *O. Aukrust (1992)*: The Scandinavian Contribution to National Accounting

No. 74 *J. Aasness, E, Eide and T. Skjerpen (1992)*: A Criminometric Study Using Panel Data and Latent Variables

No. 75 *R. Aaberge and Xuezeng Li (1992)*: The Trend in Income Inequality in Urban Sichuan and Liaoning, 1986-1990

No. 76 *J.K. Dagsvik and Steinar Strøm (1992)*: Labor Supply with Non-convex Budget Sets, Hours Restriction and Non-pecuniary Job-attributes

No. 77 *J.K. Dagsvik (1992)*: Intertemporal Discrete Choice, Random Tastes and Functional Form

No. 78 *H. Vennemo (1993)*: Tax Reforms when Utility is Composed of Additive Functions

No. 79 *J. K. Dagsvik (1993)*: Discrete and Continuous Choice, Max-stable Processes and Independence from Irrelevant Attributes

No. 80 *J. K. Dagsvik (1993)*: How Large is the Class of Generalized Extreme Value Random Utility Models?

No. 81 *H. Birkelund, E. Gjelsvik, M. Aaserud (1993)*: Carbon/energy Taxes and the Energy Market in Western Europe

No. 82 *E. Bowitz (1993)*: Unemployment and the Growth in the Number of Recipients of Disability Benefits in Norway

No. 83 *L. Andreassen (1993)*: Theoretical and Econometric Modeling of Disequilibrium

No. 84 *K.A. Brekke (1993)*: Do Cost-Benefit Analyses favour Environmentalists?

No. 85 *L. Andreassen (1993)*: Demographic Forecasting with a Dynamic Stochastic Microsimulation Model

No. 86 *G.B. Asheim and K.A. Brekke (1993)*: Sustainability when Resource Management has Stochastic Consequences

No. 87 *O. Bjerkholt and Yu Zhu (1993)*: Living Conditions of Urban Chinese Households around 1990

No. 88 *R. Aaberge (1993)*: Theoretical Foundations of Lorenz Curve Orderings

No. 89 *J. Aasness, E. Biørn and T. Skjerpen (1993)*: Engel Functions, Panel Data, and Latent Variables - with Detailed Results

No. 90 *Ingvild Svendsen (1993):* Testing the Rational Expectations Hypothesis Using Norwegian Microeconomic DataTesting the REH. Using Norwegian Microeconomic Data

No. 91 *Einar Bowitz, Asbjørn Rødseth and Erik Storm (1993):* Fiscal Expansion, the Budget Deficit and the Economy: Norway 1988-91

No. 92 *Rolf Aaberge, Ugo Colombino and Steinar Strøm (1993):* Labor Supply in Italy

No. 93 *Tor Jakob Klette (1993):* Is Price Equal to Marginal Costs? An Integrated Study of Price-Cost Margins and Scale Economies among Norwegian Manufacturing Establishments 1975-90

No. 94 *John K. Dagsvik (1993):* Choice Probabilities and Equilibrium Conditions in a Matching Market with Flexible Contracts

No. 95 *Tom Kornstad (1993):* Empirical Approaches for Analysing Consumption and Labour Supply in a Life Cycle Perspective

No. 96 *Tom Kornstad (1993):* An Empirical Life Cycle Model of Savings, Labour Supply and Consumption without Intertemporal Separability

No. 97 *Snorre Kverndokk (1993):* Coalitions and Side Payments in International CO_2 Treaties

No. 98 *Torbjørn Eika (1993):* Wage Equations in Macro Models. Phillips Curve versus Error Correction Model Determination of Wages in Large-Scale UK Macro Models

No. 99 *Anne Brendemoen and Haakon Vennemo (1993):* The Marginal Cost of Funds in the Presence of External Effects

No. 100 *Kjersti-Gro Lindquist (1993):* Empirical Modelling of Norwegian Exports: A Disaggregated Approach

No. 101 *Anne Sofie Jore, Terje Skjerpen and Anders Rygh Swensen (1993):* Testing for Purchasing Power Parity and Interest Rate Parities on Norwegian Data

No. 102 *Runa Nesbakken and Steinar Strøm (1993):* The Choice of Space Heating System and Energy Consumption in Norwegian Households (Will be issued later)

No. 103 *Asbjørn Aaheim and Karine Nyborg (1993):* "Green National Product": Good Intentions, Poor Device?

No. 104 *Knut H. Alfsen, Hugo Birkelund and Morten Aaserud (1993):* Secondary benefits of the EC Carbon/ Energy Tax

No. 105 *Jørgen Aasness and Bjart Holtsmark (1993):* Consumer Demand in a General Equilibrium Model for Environmental Analysis

No. 106 *Kjersti-Gro Lindquist (1993):* The Existence of Factor Substitution in the Primary Aluminium Industry: A Multivariate Error Correction Approach on Norwegian Panel Data

No. 107 *Snorre Kverndokk (1994):* Depletion of Fossil Fuels and the Impacts of Global Warming

No. 108 *Knut A. Magnussen (1994):* Precautionary Saving and Old-Age Pensions

No. 109 *Frode Johansen (1994):* Investment and Financial Constraints: An Empirical Analysis of Norwegian Firms

No. 110 *Kjell Arne Brekke and Pål Børing (1994):* The Volatility of Oil Wealth under Uncertainty about Parameter Values

No. 111 *Margaret J. Simpson (1994):* Foreign Control and Norwegian Manufacturing Performance

No .112 *Yngve Willassen and Tor Jakob Klette (1994):* Correlated Measurement Errors, Bound on Parameters, and a Model of Producer Behavior

No. 113 *Dag Wetterwald (1994):* Car ownership and private car use: A microeconometric analysis based on Norwegian data

Statistics Norway
Research Department
P.O.B. 8131 Dep.
N-0033 Oslo

Tel.: +47-22 86 45 00
Fax: +47-22 11 12 38

Statistics Norway
Research Department