UNIVERSITEIT
BRABANT
POSTBOX 90153
5000 LE TILBURG
THE NETHERLANDS

[redacted]

[redacted]

DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

[redacted]

THE PROBLEM OF NOT OBSERVING SMALL EXPENDITURES IN A CONSUMER EXPENDITURE SURVEY

Rob Alessie*)
Raymond Gradus*)
Bertrand Melenberg*)

FEW 293

March 1987
revised October 1987

*) Tilburg University. The authors would like to express their thanks to Arie Kapteyn, Theo Nijman, Jean Marc Robin, and Arthur van Soest for their valuable comments, and to Jan-Berno van Lochem for his excellent research assistence. Rob Alessie and Raymond Gradus also wish to thank the Netherlands Organisation for the Advancement of Pure Research (ZWO) and Rob Alessie also to the NMB bank for financial support.

Abstract

In consumer expenditure surveys one often faces the problem that, in practice, full information on small consumption expenditures is hardly available. In this paper we propose a method to correct for the underestimation of consumption resulting from not recording the small expenditures. We apply this method to a Dutch panel which is divided into the subsamples A and B. For subsample B all expenditures are registered, but from subsample A only expenditures above 10 Dutch guilders. The method consists of constructing a model, which explains for each consumption category the sum of expenditures below 10 guilders made during a month, by explicitly taking into account that each expenditure should be below 10 guilders, and estimating this model using data of subsample B. The model can then be used to calculate the expected values of the small expenditures made by households in subsample A.

§ 1. Introduction

In consumer expenditure surveys one often faces the problem that information on all consumption expenditures is not complete. Especially collecting complete information on small expenditures is a difficult task.

In the 1980-1981 Consumer Expenditure Survey of the Netherlands, conducted by the Netherlands Central Bureau of Statistics, for example, information on small expenditures is only gathered during a so-called registration month, once a year. Afterwards, the values of annual expenditures are obtained by inflating the monthly figures. In another Dutch panel, the so-called Expenditure Index, conducted by a private marketing research agency (INTOMART), the sample is divided into two subsamples. The respondents in the first subsample, A, say, is only asked to give information on large expenditures, defined as expenditures in excess of 10 Dutch guilders, whereas the other subsample, B, say, is asked to give also information on small expenditures (expenditures less than or equal to 10 Dutch guilders). The consequence of this procedure is that we do not know the values of the small expenditures of the households in subsample A. Table 1.1 clearly shows that using only expenditures in excess of 10 Dutch guilders to determine the total values of the expenditures would lead to considerable underestimation for several expenditure categories in case of subsample B. It is likely that such underestimation would also occur for subsample A.

Table 1.1: Mean level of the sum of expenditures below 10 Dutch guilders and of expenditures in excess of 10 Dutch guilders by the households of subsample B in April 1984.

Consumption category	Expenditures in excess of Dfl. 10	Expenditures below Dfl. 10
1) Food	470	47
2) Clothing and Footwear	217	4
3) Housing, including rents and interest payments on and redemptions of mortgage payments	630	0
4) Domestic decoration including furniture, expenditures on do-it-yourself articles and on gardening	404	10
5) Recreation, Entertainment including holiday expenditures	183	11
6) Vehicles, including purchases of cars, bicycles etc.	48	1
7) Transportation, including expenditures on fuel and public transportation	96	2
8) Insurance	140	0
9) Appliances, including electric appliances, such as hifi-equipments and washing machines and other personal expenditures	162	7
10) Other expenditures, including medical expenditures, gifts and donations	140	11

In this paper we describe an approach to correct for this underestimation. Our approach is to construct an econometric model that explains for each consumption category, for each period (one month), and for each household the level, Y, of expenditures below 10 Dutch guilders. Once the model has been estimated by using data from subsample B we can use this model to calculate the expected values of the small expenditures, EY, made by the households in subsample A.

The model consists of two parts. The first part which we call the Count Model explains on a household level the number of expenditures, N, below 10 Dutch guilders by using a probability distribution defined on the nonnegative integers, whereas the second part which we call the Amount Model explains the amount of each expenditure that is below Dfl. 10 by using a (conditional) probability distribution defined on the [0;10]-interval. Household characteristics are included by parameterization of the parameters of the probability distribution. This model bears some resemblance with that of Robin (1987).

It is clear that there are alternative ways of modelling the monthly sum of the small expenditures Y. For instance, one can formulate a TOBIT-model, where the latent variable corresponding to the sum of the small expeditures, Y^*, depends on some household characteristics. But this way of modelling neglects the information, that each single expenditure should be below Dfl 10,-. Moreover, there is a second argument for our way of modelling. If we had for all households information about the number of small purchases at our disposal, it would be easy to incorporate this information in our model. In that case, we only have to estimate the Amount Model and we would be able to predict for the households in subsample A the expected values of the small expenditures given the number of these purchases by $EY|N$. We presume that this would lead to a better prediction of the sum of small penditures, than if one uses EY, the unconditional expectation. Therefore, it would be desirable to collect in the future some information about N.

Notice that this way of modelling can be used in other applications as well. For example, instead of small expenditures we can explain gasoline consumption by car owners by first modelling the number of times an individual visits a gas station and then modelling the quantity of petrol the person buys anytime he or she visits a gas station. In the same way other examples can be conceived of.

The structure of the paper is as follows. In the second section we present the model and discuss some estimation strategies. Also, we derive an explicit formula for the expected values of expenditures below 10 Dutch guilders. In Section 3 we present the empirical application of our model to the data of the Expenditure Index. Section 4 concludes.

§ 2. The model

In this section we present the model. Each consumption category will be considered separately. The following diagram, which represents the small expenditures made by some arbitrarily chosen household during a given period, is the starting point of our modelling

Y_i: Amount of i-th purchase $\longrightarrow$

N: Number of purchases ↓		1	2	...	n	n+1	
	N=0	$Y_1=0$	$Y_2=0$	...	$Y_n=0$	$Y_{n+1}=0$	...
	N=1	$Y_1 \in [0;10]$	$Y_2=0$	...	$Y_n=0$	$Y_{n+1}=0$	...
	N=2	$Y_1 \in [0;10]$	$Y_2 \in [0;10]$	...	$Y_n=0$	$Y_{n+1}=0$	...
	⋮	⋮	⋮		⋮	⋮	
	N=n	$Y_1 \in [0;10]$	$Y_2 \in [0;10]$	...	$Y_n \in [0;10]$	$Y_{n+1}=0$	...
	⋮	⋮	⋮		⋮	⋮	

N denotes the number of purchases, $N \in \{0,1,2,...\}$. Y_i denotes the amount of the i-th purchase; if $i \leq N$, then $Y_i \in [0;10]$, otherwise $Y_i \equiv 0$, $i \in \{1,2,...\}$.

Define $Z \equiv (Y_1, Y_2, \ldots)$. Then we construct the probability distribution of (N,Z) as follows [1),2)].

$$Pr\{N=n, Z\in\zeta\} = Pr\{N=n\} * Pr\{Z\in\zeta | N=n\} \qquad (2.1)$$

Pr{N=n} is the Count Model, whereas Pr{Z∈ζ|N=n) is the Amount Model.

With regard to Pr{Z∈ζ|N=n} we assume

$$Pr\{Z\in\zeta | N=n\} = \Pi_{i=\{1,2,\ldots,n\}} \int_{\zeta_i} f(v)dv * \Pi_{j\in\{n+1,n+2,\ldots\}} II_{\{0\}}(\zeta_j) \qquad (2.2)$$

where $\zeta = \zeta_1 \times \zeta_2 \times \ldots$ $(= \Pi_{i\in\mathbb{N}}\zeta_i)$ and

$$II_{\{0\}}(\zeta_i) = 1 \quad \text{if } 0 \in \zeta_i$$
$$= 0 \quad \text{otherwise}$$

f(v) stands for a probability density function.
Notice that we assume for $k,\ell \le N$ the amount of the k-th purchase to be stochastically independent of the ℓ-th purchase, conditional upon N = n. This means that we do not take into account possible correlation between Y_k and Y_ℓ with $k \ne \ell$ and $k, \ell < N$.
The second part of the right-hand-side of (2.2) says that the conditional probability is positive only if $Y_j \equiv 0$ for $j > N = n$.

---

1) An alternative approach is to take $Y_1, Y_2, \ldots$ to be i.i.d. random variables, distributed on the interval [0;10], and to assume that $Z = (Y_1, Y_2, \ldots)$ and N are independent. The difference between this approach and the one presented in the main text lies in the definition of the total expenditures. Following the main text we have as total expenditures $\Sigma_{j\in\mathbb{N}} Y_j$, whereas the alternative approach results in $\Sigma_{j=1}^{N} Y_j$.

2) To be precise: The probability space with respect to which (N,Z) is defined is $((\mathbb{N}\times\mathbb{R}^{\mathbb{N}}), \mathcal{B}(\mathbb{N}\times\mathbb{R}^{\mathbb{N}}), P)$, with $Z \equiv \Pi_{i\in\mathbb{N}} Z_i \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$.

The next step consists of choosing families of distribution functions from which one will be chosen by way of estimation. For Pr{N=n}, the Count Model, we first consider the Poisson (λ) model, i.e.,

$$Pr(N=n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

The parameter λ is specified as follows

$$\lambda = \exp(X'\beta) \tag{2.3}$$

where $X :=$ vector of exogenous variables (household characteristics)
$\beta :=$ vector of parameters.

The Poisson distribution is restrictive in several ways (see Cameron and Trivedi (1986), Gourieroux, Montfort and Trognon (1984b)). For instance, the assumption, that the conditional mean and variance of N given X are equal, may be too strong. One way to relax this restriction is to allow for unobserved heterogeneity in λ by replacing (2.3) by the following equation

$$\lambda = \exp(X'\beta) \exp(\varepsilon) \tag{2.4}$$

where $\varepsilon :=$ error term with $E(\exp \varepsilon | X) = 1$.

Equation (2.4) implies, that N given X and ε is Poisson (λ) distributed. Since ε is an unobservable random variable, we must integrate it out to obtain the conditional distribution of N given X. Cameron and Trivedi (1986), among others, show that if $\lambda \sim$ Gamma (φ,ν), with $\varphi = \exp(x'\beta)$, then $N|X=x \sim$ Negative Binomial (φ,ν), i.e.,

$$Pr(N=n|X=x) = \frac{\Gamma(n+\nu)}{\Gamma(n+1)\ \Gamma(\nu)} \left[\frac{\nu}{\nu+\varphi}\right]^{\nu} \left[\frac{\varphi}{\nu+\varphi}\right]^{n} \tag{2.5}$$

Notice that $E(N|X=x) = \varphi = \exp(x'\beta)$ and $\mathrm{var}(N|X=x) = \varphi + \varphi^2/\nu$.

The Negative Binomial distribution will be abbreviated as NEGBIN distribution. Concerning the parametrization of ν, we consider three possibilities:

a) $$\nu = \alpha^{-1}\varphi = \alpha^{-1} \exp(x'\beta) \qquad \varphi = \exp(x'\beta) \tag{2.6}$$

b) $$\nu = \alpha^{-1}\varphi^2 = \alpha^{-1} (\exp(x'\beta))^2 \qquad \varphi = \exp(x'\beta) \tag{2.7}$$

c) $$\nu = \exp(\gamma_0 + x'\gamma) \qquad \varphi = \exp(x'\beta) \tag{2.8}$$

With regard to (2.6) and (2.7), the parameterization of the NEGBIN distribution coincides with the NEGBINI- and NEGBINII-parameterization of Cameron and Trivedi (1986), respectively. The NEGBINI and NEGBINII distribution reduce to the Poisson-distribution if $\alpha \downarrow 0$. In the sequel, the NEGBIN distribution with ν parameterized as in (2.8) will be called the NEGBINA-distribution. Notice that the NEGBINI and NEGBINII distributions are special cases of the NEGBINA distribution.[3)]

The densities f(v) of formula (2.2.) are modelled in two ways:
The uniform distribution:

$$f(v) = \frac{1}{10} v \, I_{[0;10]}(v) \tag{2.9}$$

The Beta-distribution:

$$f(v) = \frac{1}{10} \frac{1}{B(p,q)} \left(\frac{v}{10}\right)^{p-1} \left(1-\frac{v}{10}\right)^{q-1} I_{[0,10]}(v) \tag{2.10}$$

$I_{[0,10]}(v)$ is an indicator function defined by

$$I_{[0;10]}(v) = 1 \text{ if } v \in [0;10]$$
$$= 0 \text{ otherwise.}$$

---

3) To get from the NEGBINA distribution the NEGBINI distribution, just define $\alpha^{-1} = \exp(\gamma_0)$ and set $\gamma = \beta$, to get the NEGBINII distribution also define $\alpha^{-1} = \exp(\gamma_0)$ and set $\gamma = 2\beta$.

In the empirical application of our model we do not further parameterize the parameters p and q.

Given the above modelling we can derive for each consumption category a likelihood function of the observations. We assume independence across households and across time. So, for a particular consumption category, the likelihood function is

$$L = \Pi_{t\in T}\, \Pi_{h\in H}\, Pr_{ht}\{N_{ht} = n_{ht}\} \qquad \text{if } n_{ht} = 0$$

$$\qquad\qquad (2.11)$$

$$= \Pi_{t\in T}\, \Pi_{h\in H}\, Pr_{ht}\{N_{ht} = n_{ht}\}\, \Pi_{i\in\{1,2,\ldots,n_{ht}\}}\, f_t(y_{iht}) \text{ if } n_{ht}>0$$

H stands for the set of households, T is the set of periods, n_{ht} is the number of purchases by household h in period t, and y_{iht} is the (positive) amount of the i-th purchase by household h in period t, $i = 1,\ldots,n_{ht}$. We add the subscripts ht and t because in the Count Model we assume the exogenous variables to be dependent on household characteristics and the parameters to be time dependent whereas in the Amount Model in case of the Beta distribution we assume the parameters p and q to be dependent on time. Let us now consider the log of the likelihood:

$$\log L = \Sigma_{t\in T}\, \Sigma_{h\in H} \log Pr_{ht}\{N_{ht} = n_{ht}\} +$$

$$+ \Sigma_{t\in T}\, \Sigma_{h\in H}\, \Sigma_{i\in\{1,\ldots,n_{ht}\}} \log f_t(y_{iht}) \qquad (2.12)$$

where the second part of the right-hand side does not appear if $n_{ht} = 0$. As long as the sets of parameters of the first and of the second part of the right-hand-side are disjoint, maximizing the log L as function of the parameters can be done by maximizing each part separately. In our case we assume the parameter sets of the Count and the Amount Model to be disjoint. But even if the parameters of the Amount Model depend on some household characteristics or on the number of purchases, N, we would still be able to estimate both models independently of one another. E.g., in formula (2.10) we could specify $p = p(n,x)$, $q = q(n,x)$. As a consequence, in formula (2.2) we should replace $f(v)$ by $f_n(v)$.

Concerning the Count Model two estimation methods are considered:

1. Maximum Likelihood (ML)
2. Pseudo Maximum Likelihood (PML)
   (See Gourieroux, Montfort and Trognon ((1984a),(1984b))

The PML estimators are obtained by maximizing a likelihood function associated with some family of probability distributions, which does not necessarily contain the true distribution. Gourieroux, Montfort and Trognon (1984a) show that under certain assumptions the PML-method will in the case of a linear exponential family give consistent estimators of the parameters appearing in the first order moment of the true distribution (cf. the parameters β in the equations (2.4) and (2.5))
Since, in that case, the PML-method only assumes a correctly specified mean, the ML-method will give more efficient estimates than the PML-method, if the distribution of $N|X$ is correctly specified. However, the PML-method is more robust. In our empirical application, we shall consider the PML-estimators associated with the Poisson family. A consistent estimate of the variance-covariance matrix of the PML-estimates is calculated by using the results of Gourieroux, Montfort, and Trognon (1984a). See also Cameron and Trivedi (1986, page 37). Note, that PML estimates can also be derived from the NEGBINA-distribution with the parameter-vector γ being given (see Gourieroux Montfort and Trognon (1984b)).

Once the model has been estimated by using data of the subsample B, we can use the model to predict the values of the small expenditures made by the households in subsample A. This prediction can be made by taking the expectation of the total value of the small expenditures Y, where $Y \equiv \Sigma_{j \in N} Y_j$. We have

$$E(Y) = E(E(Y|N)) = E(E(\Sigma_{j \in N} Y_j|N)) = E(\Sigma_{j \in N} E(Y_j|N)) =$$

$$= E(\Sigma_{j \in N} E(V)\ I_{\{j,j+1,\ldots\}}(N)) \qquad (2.13)$$

where $I_{\{j,j+1,...\}}(N) = 1$ if $N \in \{j,j+1,...\}$
$= 0$ otherwise

and where E(V) denotes the expectation of a stochastic variable V with density function f(v).

E(V) does not depend on j. So

$$E(Y) = E(V)\ E(\Sigma_{j \in \mathbb{N}}\ I_{\{j,j+1,...\}}(N)) = E(V)\ E(N) \tag{2.14}$$

§ 3. Data, empirical model and results

Data

The data stem from a panel survey conducted in the Netherlands between april 1984 and december 1986. In this study we only use data of the period april 1984-december 1984. Information about expenditures on different categories is collected on a monthly basis, while data on background variables such as net household income and family size are gathered once a year.

The data set consists of about 800 households per month, of which approximately 10 per cent is in the B-subsample. This group reports all expenditures distinguished on the basis of a detailed goods classification. The other group (the A-subsample), however, records only purchases in excess of 10 guilders in a broader goods classification. In this study only data of the B-subsample are used to estimate the model formulated in Section 2.

From a first inspection of the data of subsample B it appears, that purchases below 10 guilders occur most frequently in the following consumption categories:

1) Food
2) Clothing and Footwear
3) Domestic Decoration
4) Recreation, Entertainment
5) Appliances
6) Other Expenditures

In the case of the other consumption categories (Housing, Vehicles, Transportation, and Insurance), such purchases are rarely found in the B-subsample. Therefore, we do not estimate our model for these consumption categories.

In the Count Model (see the equations (2.4) and (2.6)) the following regressors appear for each consumption category

1. C := Constant term
2. URB := Degree of Urbanization
   1 = rural municipality
   2 = commuter towns and small towns
   3 = medium sized cities (30,000 -100,000 inhabitants)
   4 = large cities (over 100,000 inhabitants)
3. CHILD = 1 if household contains children younger than five years of age
   = 0 otherwise
4. FS := Family size
5. WORKH = 1 if head of household doesn't have a paid job
   = 0 otherwise
6. AGEP := Age of the housewife (age of the head of the household if there is no housewife.)
7. SOC := Social group
   1. upper class
   2. upper middle class
   3. middle class
   4. lower middle class
   5. lower class
8. SINGLE = 1 if head of household lives alone
   = 0 otherwise

The NEGBIN models (see the formulae (2.4) and (2.6)) are chosen as a starting point of our analysis. To begin with, we assume that the parameter vectors of the NEGBIN-models do not change over time (the sample period). The resulting ML and PML estimates of the NEGBINI, NEGBINII and NEGBINA-models are presented in table 3.1.

In case of the NEGBINA-model only the estimates of the parameters appearing in the first order moment of the negative binomial distribution are presented. In order to calculate the expected value of the small purchases (cf. formula 2.14) we only need these parameters. A comparison of the ML and PML-estimates of the parameters, which appear in the first order moment of the NEGBINA-model, shows, that these are similar in sign and magnitude. This observation especially applies to the significant parameters. This is an encouraging result, because both the ML and PML-estimates of these parameters are consistent, if the first order moments are correctly specified. The ML and PML-estimates could, however, differ considerably, if the first order moment of the Negative Binomial distribution is not correctly specified.

Next we turn to the topic of model selection. In section 2 we noted, that the NEGBINI- and NEGBINII-models follow from the NEGBINA model by imposing suitable restrictions. By means of a likelihood ratio test we check whether we may impose such restrictions. This test statistic, which under the null hypothesis, is asymptotically distributed as $\chi^2(7)$, is significant for both the NEGBINI- and NEGBINII-model in case of Food, Recreation and Other Expenditures (c.f. row "LR" in table 3.1). For these consumption categories we select the NEGBINA-model and for the other categories we choose the NEGBINI-model, because from the values of Akaike Information Criterion (c.f. row AIC or Table 3.1) we can conclude that, in terms of this criterion, the NEGBINI model shows a better performance than the NEGBINII model for Clothing, Footwear and a similar performance for Domestic decoration and Appliances.

Now we will discuss the estimation results. For the selected models we have carried out a Likelihood ratio test of the hypothesis, that the parameters corresponding to all explanatory variables except the constant term are equal to zero. The test statistic is asymptotically distributed as $\chi^2(14)$ for Food, Recreation, and Other Expenditures and $\chi^2(7)$ for the other consumption categories. In all cases we must reject the null hypothesis (see row LC in table 3.1).

Table 3.1: ML- and PML-estimates of β for the NEGBIN-models (Asymptotic t- values in parentheses)

	FOOD				CLOTHING. FOOT WEAR			
	ML NEGBINI	ML NEGBINII	ML NEGBINA	PML	ML NEGBINI	ML NEGBINII	ML NEGBINA	PML
Constant	2.234	2.115	2.926	2.810	-0.710	-0.555	-0.712	-0.896
	(9.68)	(13.02)	(6.252)	(10.87)	(-1.68)	(-1.69)	(-1.41)	(-1.91)
URB	-0.145	-0.108	-0.136	-0.104	-0.121	-0.109	-0.093	-0.081
	(-3.86)	(-4.43)	(-2.05)	(-2.96)	(-1.83)	(-2.15)	(-1.13)	(-1.17)
Child	-0.104	-0.055	-0.310	-0.234	-0.075	-0.164	-0.002	0.043
	(-1.05)	(-0.87)	(-1.46)	(-2.15)	(-0.44)	(-1.13)	(-0.007)	(0.22)
FG	0.1145	0.089	0.073	0.082	0.218	0.176	0.239	0.229
	(3.22)	(4.02)	(0.99)	(2.36)	(3.63)	(3.65)	(3.29)	(4.18)
WORKH	0.020	-0.010	0.0032	0.069	-0.086	-0.052	-0.002	-0.096
	(0.24)	(-0.17)	(0.02)	(0.69)	(-0.54)	(-0.425)	(-0.01)	(-0.60)
AGEP	0.0088	0.008	-0.002	0.0003	0.012	0.009	0.007	0.012
	(3.35)	(4.56)	(-0.290)	(0.009)	(2.11)	(2.235)	(0.86)	(2.27)
SOC	-0.075	-0.030	-0.084	-0.115	-0.171	-0.146	-0.139	-0.146
	(-2.40)	(-1.31)	(-1.70)	(-2.55)	(-2.98)	(-3.25)	(-2.07)	(-2.26)
SINGLE	0.182	0.168	0.052	0.042	0.034	0.066	-0.0167	-0.003
	(1.61)	(2.26)	(0.29)	(0.33)	(0.16)	(0.42)	(-0.065)	(-1.51)
α	19.448	0.169	-	-	1.070	0.00076	-	-
	(11.12)	(7.34)	-	-	(6.41)	(5.30)	-	-
Log L.	15446.19	15451.11	15458.70	-	-593.94	-594.77	-589.24	-
LR	25.02	15.18	-	-	9.40	11.06	-	-
LRP	6121.64	6131.48	6146.66	-	184.58	182.92	193.88	-
LC	-	-	71.94	-	49.44	-	-	-
AIC	-30874.38	-30884.22	-30885.4	-	1205.88	1207.54	1210.48	-

Log L = Log Likelihood

LR = Likelihood ratio test of the hypothesis, that the NEGBINA-distribution reduces to the NEGBINI or the NEGBINII distribution (depending on the model selection).

LRP = Likelihood ratio test of the hypothesis, that the number of small purchases is Poisson distributed.

LC = Likelihood ratio test of the hypothesis, that the parameters corresponding to all explanatory variables except the constant term are equal to zero

AIC = Akaike's Information criterion

Table 3.1: continued

	APPLIANCES				OTHER EXPENDITURES			
	ML NEGBINI	ML NEGBINII	ML NEGBINA	PML	ML NEGBINI	ML NEGBINII	ML NEGBINA	PML
Constant	0.236	0.284	0.571	0.489	0.559	0.623	0.467	0.397
	(0.768)	(1.26)	(1.48)	(1.27)	(1.46)	(2.38)	(0.80)	(1.23)
URB	-0.148	-0.118	-0.132	-0.119	-0.499	-0.316	-0.541	-0.559
	(-3.29)	(-3.50)	(-2.52)	(-2.13)	(-9.50)	(-7.38)	(-8.41)	(-8.07)
Child	-0.183	-0.103	-0.342	0.335	-0.104	-0.002	-0.244	-0.266
	(-1.345)	(-1.08)	(-2.08)	(-2.44)	(-0.656)	(-1.98)	(-1.34)	(-1.50)
FG	0.205	0.174	0.176	0.172	0.133	0.083	0.183	0.167
	(4.56)	(5.36)	(2.88)	(3.87)	(2.551)	(2.07)	(2.23)	(3.53)
WORK II	-0.320	-0.010	-0.065	-0.049	0.222	0.112	0.356	0.353
	(-0.29)	(-0.12)	(-0.42)	(-0.43)	(1.699)	(1.16)	(1.98)	(2.78)
AGEP	0.010	0.008	0.007	0.007	0.015	0.011	0.011	0.015
	(2.57)	(2.93)	(1.19)	(1.63)	(3.321)	(0.326)	(1.59)	(3.47)
SOC	(-0.045)	-0.034	-0.072	-0.049	-0.120	-0.117	-0.073	-0.081
	(-1.073)	(-1.14)	(-0.102)	(-1.32)	(-2.69)	(-3.35)	(-1.23)	(-1.65)
SINGLE	0.075	0.151	-0.048	-0.125	0.159	0.153	0.119	0.071
	(0.59)	(1.42)	(0.250)	(-0.79)	(1.019)	(1.29)	(0.55)	(0.42)
α	2.72	0.006	-	-	1.966	0.003	-	-
	(10.69)	(7.81)	-	-	(7.794)	(6.61)	-	-
Log L.	119.49	119.45	125.2	-	-305.80	-319.60	-296.64	-
LR	11.42	11.50	-	-	18.32	45.92	-	-
LRP	848.68	848.60	860.1	-	447.84	420.29	466.16	-
LC	60.92	-	-	-	-	-	146.98	-
AIC	-220.98	-220.90	-218.9	-	629.60	657.20	625.28	-

Log L = Log Likelihood

LR = Likelihood ratio test of the hypothesis, that the NEGBINA-distribution reduces to the NEGBINI or the NEGBINII distribution (depending on the model selection).

LRP = Likelihood ratio test of the hypothesis, that the number of small purchases is Poisson distributed.

LC = Likelihood ratio test of the hypothesis, that the parameters corresponding to all explanatory variables except the constant term are equal to zero

AIC = Akaike's Information criterion

Table 3.1: continued

	DOMESTIC DECORATION				RECREATION, ENTERTAINMENT			
	ML NEGBINI	ML NEGBINII	ML NEGBINA	PML	ML NEGBINI	ML NEGBINII	ML NEGBINA	PML
Constant	0.587	0.514	0.879	0.754	1.363	1.197	1.191	1.091
	(1.94)	(2.201)	(2.11)	(2.50)	(4.25)	(5.49)	(2.98)	(2.36)
URB	-0.199	-0.152	-0.198	-0.206	-0.290	-0.194	-0.294	-0.232
	(-4.30)	(-4.23)	(-3.69)	(-4.52)	(-5.60)	(-5.47)	(-4.52)	(-2.75)
Child	0.131	0.135	0.069	0.060	0.424	-0.258	-0.492	-0.598
	(0.94)	(1.26)	(0.38)	(0.42)	(-3.01)	(-2.62)	(-2.61)	(-3.28)
FG	0.030	0.027	0.033	0.017	0.257	0.158	0.327	0.343
	(0.65)	(0.74)	(0.53)	(0.38)	(5.17)	(4.72)	(5.21)	(6.06)
WORKH	0.042	0.010	0.229	0.094	-0.125	-0.203	0.187	0.268
	(0.37)	(0.11)	(1.59)	(0.80)	(-1.07)	(-2.36)	(1.33)	(1.43)
AGEP	0.015	0.013	0.008	0.012	-0.0025	0.0015	-0.013	-0.012
	(3.80)	(4.37)	(1.26)	(3.73)	(-0.68)	(0.60)	(-2.79)	(-2.65)
SOC	-0.185	-0.152	-0.181	-0.174	-0.188	-0.134	-0.076	-0.111
	(-4.39)	(-4.64)	(-3.27)	(-4.54)	(-4.05)	(-4.29)	(-1.44)	(-1.84)
SINGLE	0.061	0.102	0.022	-0.049	0.922	0.629	-0.928	0.818
	(0.43)	(0.94)	(0.124)	(-0.36)	(5.66)	(5.67)	(4.50)	(4.34)
α	1.543	0.00023	-	-	4.754	0.012	-	-
	(7.56)	(6.06)	-	-	11.85	(8.26)	-	-
Log L.	-458.73	-458.49	-454.57	-	757.45	755.55	776.34	-
LR	8.32	7.84	-	-	37.78	41.58	-	-
LRP	329.22	329.7	337.54	-	1647.60	1643.80	1685.38	-
LC	63.0	-	-	-	-	-	143.42	-
AIC	935.46	934.88	941.14	-	-1496.90	-1493.10	-1520.68	-

Log L = Log Likelihood
LR = Likelihood ratio test of the hypothesis, that the NEGBINA-distribution reduces to the NEGBINI or the NEGBINII distribution (depending on the model selection).
LRP = Likelihood ratio test of the hypothesis, that the number of small purchases is Poisson distributed.
LC = Likelihood ratio test of the hypothesis, that the parameters corresponding to all explanatory variables except the constant term are equal to zero
AIC = Akaike's Information criterion

The age of the housewife, AGEP, has a significant[4] positive influence on the expected number of purchases below Dfl. 10 (see equation (2.6)) except for Recreation, Entertainment and Food. In the Netherlands partners of lower age more frequently have a paid job and consequently have less leisure, than partners of relative higher age. Therefore, our intuition is that most households with partners of low age spend relatively little time on shopping, but buy these goods in large quantities in supermarkets and department stores. The signs of the parameters, corresponding to AGEP, conform to this intuition.

The expected number of expenditures below Dfl. 10 on Recreation, Entertainment decreases significantly if small children are present in the household, and increases, if the household consists of a single person. The last result would conform to a life style of singles, especially those of young age, who spend a fair amount of time away from the home. Social class, SOC, - a variable which is strongly correlated with income - plays a significant role in explaining the number of purchases below Dfl. 10 for Clothing, Footwear and Domestic Decoration. The higher the social class the higher the expected number of small purchases will be.

The degree of urbanisation, URB, has a significant negative, and family size, FS, a significant positive influence on the expected number of purchases below Dfl. 10 for all consumption categories except for Clothing and Footwear in the case of URB and except for Domestic decoration in the case of FS.

Finally, the results suggest that the dummy variable WORKH, which indicates whether the head of household works, has no explanatory power for the number of purchases below Dfl. 10 for any category, except for Other Expenditures.

In the preceding section it was noticed that the Poisson-model is nested in the NEGBIN-model. A likelihood ratio test indicates that the Poisson-model must be rejected for all consumption categories (see row LRP in table 3.1). This is not a surprising result, because the Poisson model is very restrictive in several ways. For instance, the assumption underlying the Poisson model that the conditional mean and variance of the count, N, given the regressors are equal, may be too strong.

---

4) At the 5%-level.

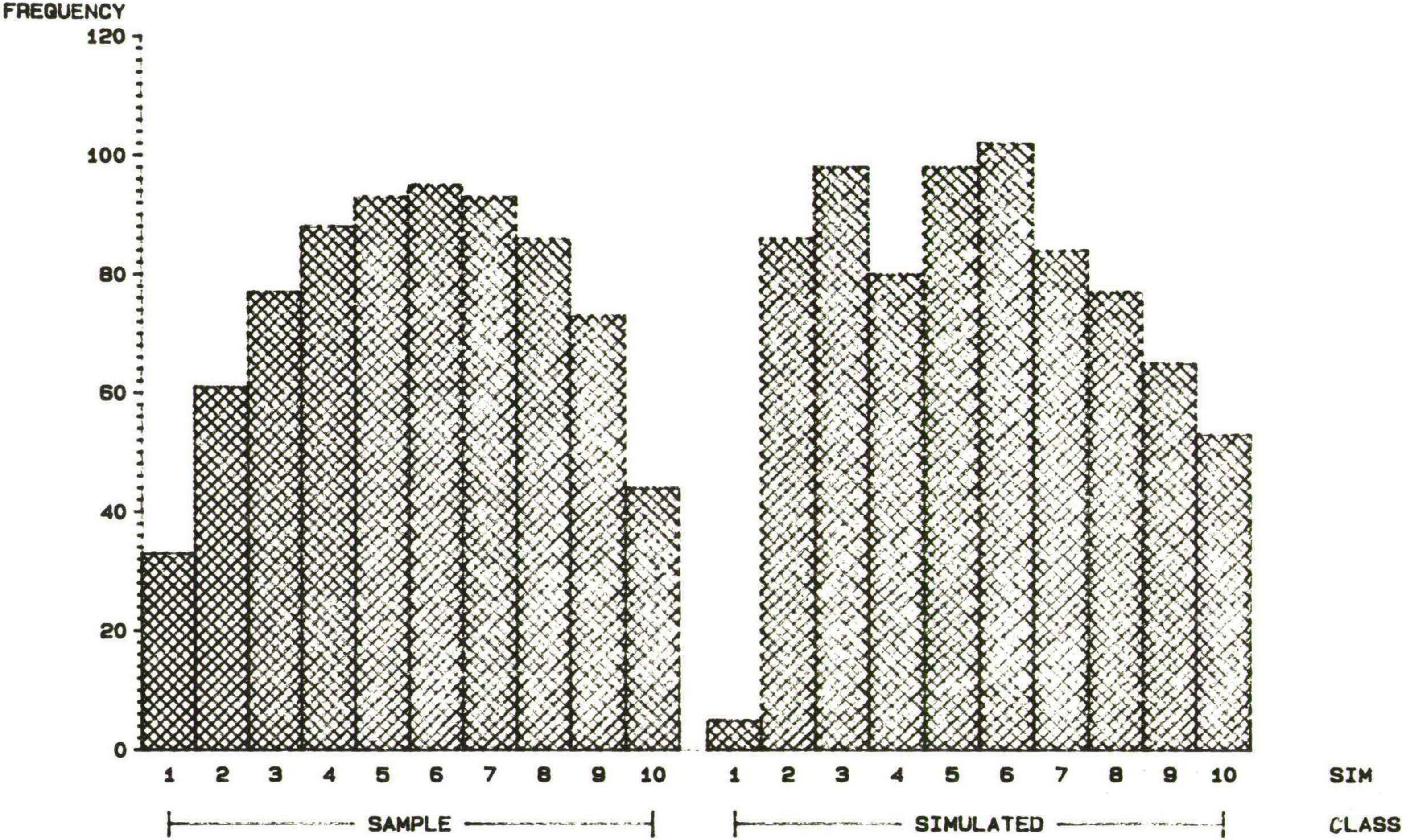
Table 3.2: Comparison of the sample mean and sample variances of the number of purchases below 10 guilders.

	april		may		june	
	mean	variance	mean	variance	mean	variance
1. Food	9.350	79.825	12.145	167.096	10.969	169.312
2. Clothing, footwear	0.713	1.575	0.957	2.101	0.508	1.035
3. Domestic Decoration	1.613	3.734	1.623	4.444	1.631	3.268
4. Recreation, Entertainment	2.463	19.594	2.261	15.225	2.000	18.656
5. Other expenditures	1.550	5.922	1.493	6.342	1.123	3.328
6. Appliances	2.300	10.694	2.043	9.366	2.185	10.684

Table 3.2 suggests, that for all consumption categories the sample variance exceeds the sample mean considerably. By using the negative binomial distribution, one can allow for the inequality of mean and variance.

Next we have tested by means of a likelihood ratio test, whether the parameter vector is varying with time (months). The results, summarized in table 3.3, suggest that the same model may be valid in all months.

FIGURE 1 SIMULATED AND SAMPLE FREQUENCY DISTRIBUTIONS
OF FOOD EXPENDITURES BELOW DFL. 10 IN APRIL 1984



10 CLASSES OF 1 DUTCH GUILDER EACH

Table 3.3: Likelihood ratio test of the hypothesis, that the parameters of the NEGBIN-model do not vary with time (months) (The test statistic is asymptotically χ^2_{72}-distributed in case of the categories Food, Recreation and Other Expenditures and χ^2_{128} for the other categories).

1. Food	76.38
2. Clothing and Footwear	59.98
3. Domestic Decoration	54.24
4. Recreation and Entertainment	118.12
5. Other Expenditures	101.10
6. Appliances	47.26

In Section 2 we have assumed that the amount of the i-th purchase Y_i, follows a Beta ([0,10],p,q) distribution (see equation (2.9)), where the parameters p and q vary over the nine months and across consumption categories. The 108 parameters are estimated by means of maximum likelihood. The following remarks can be made about the results:

- It is well known that, if p = q = 1, the Beta ([0,10],p,q)-distribution reduces to the uniform distribution over [0,10]. A likelihood ratio test indicates that for each month and for each consumption category except Appliances the hypothesis p = q = 1 has to be rejected. From the sample distribution of the small purchases of food in april 1984 (cf. figure 1) it is clear, that one cannot describe this distribution by a uniform distribution. The same remark applies to the other sample distributions.
- By means of a Wald-test we have also tested the hypothesis that the parameters p and q are constant over time (but different across consumption categories). The results are presented in table 3.4. The hypothesis must be rejected for all consumption categories except for Food.

Table 3.4: Test of the null-hypothesis, that the parameters p and q are constant over time. (The Wald test statistic is asymptotically χ^2_{16}-distributed).

1. Food	26.12
2. Clothing and Footwear	46.53
3. Domestic Decoration	82.77
4. Recreation	46.19
5. Supplementary Family costs	50.01
6. Appliances	34.56

- In order to check whether Y_i follows a Beta ([0,10],p,q) distribution, we chose to carry out a Pearson's Chi-Square Goodness of Fit test (see Mood, Graybill and Boes (1974), p. 446). For this purpose we have divided the interval [0,10] in ten equal parts $A_1,\ldots,A_{10}$ and have then calculated the predicted probabilities $P(A_i)$, $i = 1,\ldots,10$, to be used in the test. The results indicate, that, given $\alpha = 0.05$, for all months the purchases of Food and Other Expenditures cannot be described by a Beta-distribution. These results are confirmed in figure 1, in which, for the category Food in the month april, the Beta distribution predicts a too high frequency for the expenditures between zero and one Dutch guilder. The same remark applies to the category Domestic Decoration in all months except May, June and October, to Recreation, Entertainment in the months September, November and December, and to Appliances in the months July, August and September. In all other cases the null-hypothesis cannot be rejected. From these results we can conclude that it may be advisable to make the parameters p and q dependent on some family characteristics. This is a topic of future research.

§ 4. Summary and conclusions

In this paper we presented a two-stage model explaining the sum of small expenditures in one subsample (subsample B) and that can be used to predict the sum of small expenditures in the other subsample (subsample A). As noted in the introduction this kind of model can be used in other applications as well.

Concerning the first-stage model, the Count Model, the empirical results are satisfactory for all consumption categories. Especially, the similarity of the PML- and ML-estimates of the selected NEGBIN models is rather striking. Yet, it may be interesting to investigate other propability distributions for the number of small purchases, like the Double Hurdle model and With Zero (WZ) model proposed by Mullahy (1986).

The results for the second-stage model, the Amount Model, show that the Beta distribution does not fit well for all consumption categories. This may be remedied by a further parameterization of the parameters p and q of the Beta distribution. But other distributions could be tried as well.

References

Cameron, A., and P. Trivedi (1986), "Econometric models based on count data", Journal of Applied Econometrics, 1, pp. 29-53.

Gourieroux C., A. Montfort, and A. Trognon (1984a), "Pseudo Maximum Likelihood Methods: theory", Econometrica, 52, pp. 681-700.

Gourieroux C., A. Montfort, and A. Trognon (1984b), "Pseudo maximum likelihood methods: applications to Poisson-models", Econometrica, 52, pp. 701-720.

Hausman, J., B. Hall, and Z. Grilliches (1984), "Econometric models for count data with an application to the patents-R&D relationship", Econometrica, 52, pp. 909-938.

Mood, A.M., F.A. Graybill, and D.C. Boes (1974), "Introduction to the theory of statistics", Third Edition, McGraw-Hill Kogakusha, Tokyo.

Mullahy J. (1986), "Specification and testing of some modified count models", Journal of Econometrics, 33, pp. 341-365.

Robin, J.M. (1987), "A new model to estimate Engel curves from budget expenditure data", Mimeo, E.N.S.A.E., Paris.

IN 1986 REEDS VERSCHENEN

202 J.H.F. Schilderinck
Interregional Structure of the European Community. Part III

203 Antoon van den Elzen and Dolf Talman
A new strategy-adjustment process for computing a Nash equilibrium in a noncooperative more-person game

204 Jan Vingerhoets
Fabrication of copper and copper semis in developing countries. A review of evidence and opportunities

205 R. Heuts, J. van Lieshout, K. Baken
An inventory model: what is the influence of the shape of the lead time demand distribution?

206 A. van Soest, P. Kooreman
A Microeconometric Analysis of Vacation Behavior

207 F. Boekema, A. Nagelkerke
Labour Relations, Networks, Job-creation and Regional Development. A view to the consequences of technological change

208 R. Alessie, A. Kapteyn
Habit Formation and Interdependent Preferences in the Almost Ideal Demand System

209 T. Wansbeek, A. Kapteyn
Estimation of the error components model with incomplete panels

210 A.L. Hempenius
The relation between dividends and profits

211 J. Kriens, J.Th. van Lieshout
A generalisation and some properties of Markowitz' portfolio selection method

212 Jack P.C. Kleijnen and Charles R. Standridge
Experimental design and regression analysis in simulation: an FMS case study

213 T.M. Doup, A.H. van den Elzen and A.J.J. Talman
Simplicial algorithms for solving the non-linear complementarity problem on the simplotope

214 A.J.W. van de Gevel
The theory of wage differentials: a correction

215 J.P.C. Kleijnen, W. van Groenendaal
Regression analysis of factorial designs with sequential replication

216 T.E. Nijman and F.C. Palm
Consistent estimation of rational expectations models

217 P.M. Kort
The firm's investment policy under a concave adjustment cost function

218 J.P.C. Kleijnen
Decision Support Systems (DSS), en de kleren van de keizer ...

219 T.M. Doup and A.J.J. Talman
A continuous deformation algorithm on the product space of unit simplices

220 T.M. Doup and A.J.J. Talman
The 2-ray algorithm for solving equilibrium problems on the unit simplex

221 Th. van de Klundert, P. Peters
Price Inertia in a Macroeconomic Model of Monopolistic Competition

222 Christian Mulder
Testing Korteweg's rational expectations model for a small open economy

223 A.C. Meijdam, J.E.J. Plasmans
Maximum Likelihood Estimation of Econometric Models with Rational Expectations of Current Endogenous Variables

224 Arie Kapteyn, Peter Kooreman, Arthur van Soest
Non-convex budget sets, institutional constraints and imposition of concavity in a flexible household labor supply model

225 R.J. de Groof
Internationale coördinatie van economische politiek in een twee-regio-twee-sectoren model

226 Arthur van Soest, Peter Kooreman
Comment on 'Microeconometric Demand Systems with Binding Non-Negativity Constraints: The Dual Approach'

227 A.J.J. Talman and Y. Yamamoto
A globally convergent simplicial algorithm for stationary point problems on polytopes

228 Jack P.C. Kleijnen, Peter C.A. Karremans, Wim K. Oortwijn, Willem J.H. van Groenendaal
Jackknifing estimated weighted least squares

229 A.H. van den Elzen and G. van der Laan
A price adjustment for an economy with a block-diagonal pattern

230 M.H.C. Paardekooper
Jacobi-type algorithms for eigenvalues on vector- and parallel computer

231 J.P.C. Kleijnen
Analyzing simulation experiments with common random numbers

232 A.B.T.M. van Schaik, R.J. Mulder
On Superimposed Recurrent Cycles

233 M.H.C. Paardekooper
Sameh's parallel eigenvalue algorithm revisited

234 Pieter H.M. Ruys and Ton J.A. Storcken
Preferences revealed by the choice of friends

235 C.J.J. Huys en E.N. Kertzman
Effectieve belastingtarieven en kapitaalkosten

236 A.M.H. Gerards
An extension of König's theorem to graphs with no odd-K_4

237 A.M.H. Gerards and A. Schrijver
Signed Graphs - Regular Matroids - Grafts

238 Rob J.M. Alessie and Arie Kapteyn
Consumption, Savings and Demography

239 A.J. van Reeken
Begrippen rondom "kwaliteit"

240 Th.E. Nijman and F.C. Palmer
Efficiency gains due to using missing data. Procedures in regression models

241 S.C.W. Eijffinger
The determinants of the currencies within the European Monetary System

IN 1987 REEDS VERSCHENEN

242 Gerard van den Berg
Nonstationarity in job search theory

243 Annie Cuyt, Brigitte Verdonk
Block-tridiagonal linear systems and branched continued fractions

244 J.C. de Vos, W. Vervaat
Local Times of Bernoulli Walk

245 Arie Kapteyn, Peter Kooreman, Rob Willemse
Some methodological issues in the implementation of subjective poverty definitions

246 J.P.C. Kleijnen, J. Kriens, M.C.H.M. Lafleur, J.H.F. Pardoel
Sampling for Quality Inspection and Correction: AOQL Performance Criteria

247 D.B.J. Schouten
Algemene theorie van de internationale conjuncturele en strukturele afhankelijkheden

248 F.C. Bussemaker, W.H. Haemers, J.J. Seidel, E. Spence
On (v,k,λ) graphs and designs with trivial automorphism group

249 Peter M. Kort
The Influence of a Stochastic Environment on the Firm's Optimal Dynamic Investment Policy

250 R.H.J.M. Gradus
Preliminary version
The reaction of the firm on governmental policy: a game-theoretical approach

251 J.G. de Gooijer, R.M.J. Heuts
Higher order moments of bilinear time series processes with symmetrically distributed errors

252 P.H. Stevers, P.A.M. Versteijne
Evaluatie van marketing-activiteiten

253 H.P.A. Mulders, A.J. van Reeken
DATAAL - een hulpmiddel voor onderhoud van gegevensverzamelingen

254 P. Kooreman, A. Kapteyn
On the identifiability of household production functions with joint products: A comment

255 B. van Riel
Was er een profit-squeeze in de Nederlandse industrie?

256 R.P. Gilles
Economies with coalitional structures and core-like equilibrium concepts

257 P.H.M. Ruys, G. van der Laan
Computation of an industrial equilibrium

258 W.H. Haemers, A.E. Brouwer
Association schemes

259 G.J.M. van den Boom
Some modifications and applications of Rubinstein's perfect equilibrium model of bargaining

260 A.W.A. Boot, A.V. Thakor, G.F. Udell
Competition, Risk Neutrality and Loan Commitments

261 A.W.A. Boot, A.V. Thakor, G.F. Udell
Collateral and Borrower Risk

262 A. Kapteyn, I. Woittiez
Preference Interdependence and Habit Formation in Family Labor Supply

263 B. Bettonvil
A formal description of discrete event dynamic systems including perturbation analysis

264 Sylvester C.W. Eijffinger
A monthly model for the monetary policy in the Netherlands

265 F. van der Ploeg, A.J. de Zeeuw
Conflict over arms accumulation in market and command economies

266 F. van der Ploeg, A.J. de Zeeuw
Perfect equilibrium in a model of competitive arms accumulation

267 Aart de Zeeuw
Inflation and reputation: comment

268 A.J. de Zeeuw, F. van der Ploeg
Difference games and policy evaluation: a conceptual framework

269 Frederick van der Ploeg
Rationing in open economy and dynamic macroeconomics: a survey

270 G. van der Laan and A.J.J. Talman
Computing economic equilibria by variable dimension algorithms: state of the art

271 C.A.J.M. Dirven and A.J.J. Talman
A simplicial algorithm for finding equilibria in economies with linear production technologies

272 Th.E. Nijman and F.C. Palm
Consistent estimation of regression models with incompletely observed exogenous variables

273 Th.E. Nijman and F.C. Palm
Predictive accuracy gain from disaggregate sampling in arima - models

274 Raymond H.J.M. Gradus
The net present value of governmental policy: a possible way to find the Stackelberg solutions

275 Jack P.C. Kleijnen
A DSS for production planning: a case study including simulation and optimization

276 A.M.H. Gerards
A short proof of Tutte's characterization of totally unimodular matrices

277 Th. van de Klundert and F. van der Ploeg
Wage rigidity and capital mobility in an optimizing model of a small open economy

278 Peter M. Kort
The net present value in dynamic models of the firm

279 Th. van de Klundert
A Macroeconomic Two-Country Model with Price-Discriminating Monopolists

280 Arnoud Boot and Anjan V. Thakor
Dynamic equilibrium in a competitive credit market: intertemporal contracting as insurance against rationing

281 Arnoud Boot and Anjan V. Thakor
Appendix: "Dynamic equilibrium in a competitive credit market: intertemporal contracting as insurance against rationing

282 Arnoud Boot, Anjan V. Thakor and Gregory F. Udell
Credible commitments, contract enforcement problems and banks: intermediation as credibility assurance

283 Eduard Ponds
Wage bargaining and business cycles a Goodwin-Nash model

284 Prof.Dr. hab. Stefan Mynarski
The mechanism of restoring equilibrium and stability in polish market

285 P. Meulendijks
An exercise in welfare economics (II)

286 S. Jørgensen, P.M. Kort, G.J.C.Th. van Schijndel
Optimal investment, financing and dividends: a Stackelberg differential game

287 E. Nijssen, W. Reijnders
Privatisering en commercialisering; een oriëntatie ten aanzien van verzelfstandiging

288 C.B. Mulder
Inefficiency of automatically linking unemployment benefits to private sector wage rates

289 M.H.C. Paardekooper
A Quadratically convergent parallel Jacobi process for almost diagonal matrices with distinct eigenvalues

290 Pieter H.M. Ruys
Industries with private and public enterprises

291 J.J.A. Moors & J.C. van Houwelingen
Estimation of linear models with inequality restrictions

292 Arthur van Soest, Peter Kooreman
Vakantiebestemming en -bestedingen