Total mortality analysis of the Rotterdam sample of the Kaunas-Rotterdam Intervention Study (KRIS)

Mulder, P.G.H.; Hempenius, A.L.

Publication date:
1984

FEW

subfaculteit der econometrie

RESEARCH MEMORANDUM

[redacted]

TILBURG UNIVERSITY
DEPARTMENT OF ECONOMICS
Postbus 90153 - 5000 LE Tilburg
Netherlands

[redacted]

FEW
152

TOTAL MORTALITY ANALYSIS OF THE ROTTERDAM SAMPLE OF THE KAUNAS-ROTTERDAM INTERVENTION STUDY (KRIS)

by

P.G.H. Mulder
A.L. Hempenius

August 1984

1. Introduction

In 1972 a representative sample of approximately 4000 men was drawn from the Rotterdam male population aged 45-59. These men were invited to participate in the so-called Kaunas Rotterdam Intervention Study (KRIS), a small-scale study aimed at testing the feasibility of carrying out a double-blind randomized drug intervention trial in the area of cardiovascular diseases in an open male population, with special interest in operational and behavioral aspects. This study was initiated by the World Health Organization and consisted of two parallel studies simultaneously carried out in two different health care systems: in Kaunas (Lithuanian SSR, USSR) and in Rotterdam (The Netherlands); see Glasunow et al. (1982).

From September 1972 till the end of 1973, 3,365 Rotterdam men actually participated in an initial screening examination which was the first of a number of various selection stages for participants in order to become eventually eligible for the intervention trial. Hence this initial screening examination is the only extensive examination common for all 3,365 participants. At the time of this examiniation a follow-up registration started by which each subject was followed through the population registers of Rotterdam and other Dutch municipalities, in order to register one of the following endpoints of follow-up: emigration or death. In the latter case, the mortality causes were collected from the Dutch Ministry of Health. A third possible endpoint of the follow-up is study termination: March 1, 1982. This endpoint applies for a subject if he was still alive at that time in the Netherlands.

The purpose of this paper is to analyze the relationship between a number of covariables measured at the initial screening examination and the subsequent incidence of death during the approximately 9 years of follow-up of the Rotterdam sample. The data is analyzed using two different models: the loglinear survival model and the (less appropriate but often applied) logistic model. For the loglinear survival model two estimation methods are used: a full likelihood method which leans heavily on David and Moeschberger (1978) and the well-known partial likelihood method of Cox (1972 and 1975).

In this paper only mortality from all causes is considered an endpoint. Subsequent papers will deal with several mortality causes, with special emphasis on cardiovascular mortality and with using the results in a cost-effectiveness model developed in Mulder and Hempenius (1982).

2. Survival analysis

2.1. Introduction

The purpose of survival analysis is to analyze the dependency of a subject's survival time probability on a number of explanatory variables. Survival time is defined as the time elapsed since the start of follow-up (here the date of the initial screening examination) till either the occurrence of a number of mutually exclusive, independent and competing true "failure" causes (here death and emigration) or the occurrence of independent termination of a subject's follow-up, so-called "censoring" (here the predetermined and fixed censoring date of March 1, 1982, if a subject is still alive in the Netherlands at that time). One may, however, always dichotomize the causes of the occurence of an endpoint by defining one "true" failure cause (the one of interest), and grouping the other failure causes under the censored group. This will indeed be done: mortality from all causes is the failure cause of interest and emigration is treated as independent censoring.

2.2. The full likelihood function

The log likelihood function for the follow-up times of all subjects is derived in the appendix:

$$(2.1) \qquad \ell n\, L_1 = \sum_{i \in M} \ell n[\lambda_i(t_i)] - \sum_{i \in H} \int_0^{t_i} \lambda_i(x)dx,$$

where

t_i : the follow-up time of subject i in years;

$\lambda_i(t)$: the mortality rate of subject i as a function of time t;

M : the set of subjects who died;

H : the set of all subjects.

First only the simplifying assumption is made that the mortality rate $\lambda_i(t)$ is a constant in time: λ_i. Assuming λ_i to be a log-linear function of age a_i and a vector of covariables z_i as measured in all subjects i at the initial screening examination:

$$(2.2) \qquad \lambda_i = \exp(\beta_0 + \beta_a a_i + \beta' z_i),$$

the log likelihood function (2.1) boils down to

$$(2.3) \qquad \ell n\, L_l = m\beta_0 + \beta_a \sum_{i\in M} a_i + \beta' \sum_{i\in M} z_i - \exp(\beta_0) \sum_{i\in H} t_i \exp(\beta_a a_i + \beta' z_i),$$

where m equals the number of subjects died. This log likelihood has to be maximized with respect to the coefficients β_0, β_a and β', with β' a row vector of coefficients belonging to the covariables in z.

The above assumption for the mortality rate λ_i of constancy in time only covers the cross-sectional information contained in the data and disregards the longitudinal information which to a certain extent is also contained in the data[1]. In order to extract also this longitudinal information one may go about as follows. Partition the total follow-up time interval of a subject into a number of subject-years and define a time variable t = 0,1,2,... for, respectively, the 1st, 2nd, 3rd,... follow-up year of a subject[2]. Next model the possible behavior of the z_i-variables in time. This is done as follows. The time dependency of $\lambda_i(t)$ is defined as:

$$(2.4) \qquad \lambda_i(t) = \exp\{\gamma_0 + \gamma_a(a_i + t) + \gamma' z_i(t) + \gamma_t t\}$$

with $\gamma_a(a_i + t)$ the effect of age and $\gamma_t t$ an autonomous trend term. The model for $z_i(t)$ is assumed to be of the form:

$$(2.5) \qquad z_i(t) = z_i + bt + Cz_i t$$

where again the z_i are the initial measurements $[z_i = z_i(0)]$, b is a column vector and C a diagonal matrix. Substitute $z_i(t)$ into $\lambda_i(t)$ and get:

$$(2.6) \qquad \lambda_i(t) = \exp\{\gamma_0 + \gamma_a a_i + \bar{\gamma} t + \gamma' z_i + \gamma_*' z_i t\}$$

with $\bar{\gamma} = \gamma_a + \gamma_t + \gamma' b$ and $\gamma_*' = \gamma' C$.

The first hypothesis one might want to test is $\gamma_* = 0$ (i.e., C = 0, assuming $\gamma \neq 0$), meaning no interaction effects with time in $z_i(t)$. Another interesting

---

1) As mentioned in Section 1, only $z_i(0)$ is available for <u>all</u> subjects. Therefore $z_i(t)$ can be modelled in one way only.

2) See also Holford (1976).

hypothesis is $\bar{\gamma} = \gamma_a$ (i.e., $\gamma'b + \gamma_t = 0$), meaning that the combined pure time effects of the covariables in $z_i(t)$ plus the effect of an autonomous time trend are zero. If one accepts both hypotheses one may write:

$$(2.7) \qquad \lambda_i(t) = \exp\{\gamma_0 + \gamma_a(a_i + t) + \gamma' z_i\}.$$

Possibly, the (estimated) model (2.2) approximates this (estimated) model very well.[3] which leads to an interesting point: the maximum likelihood estimation procedure with model (2.7) requires an evaluation of the second summation term of the log likelihood (2.1) over the set of subject-years[4], which set is approximately 9 times as large as the set H of subjects needed for the estimation procedure with model (2.2). This maximization has been done by means of the Newton-Raphson method using a selfwritten FORTRAN program.

2.3. Cox's partial likelihood function

The partial likelihood is derived by Cox (1972 and 1975) as:

$$(2.8) \qquad L_2 = \prod_{i \in M} \{\lambda_i(t_i) / \sum_{r \in H_i} \lambda_r(t_i)\}$$

with H_i the set of subjects r still surviving just before the mortality time t_i of subject $i \in M$. The essence of this partial likelihood is the following specification of the mortality rate as a function of age a, covariable vector z and follow-up time t for subject i:

$$(2.9) \qquad \lambda_i(t) = \lambda_0(t)\exp(\delta_a a_i + \delta' z_i),$$

implying that the partial likelihood (2.8) is only a function of the coefficient δ_a and the row vector of coefficients δ', leaving the purely time-dependent part $\lambda_0(t)$ as arbitrary and irrelevant for the estimation of δ_a and δ'. It has to be noted that the values of the variables a_i and z_i in the model (2.9) are those of one measurement point only (viz., the initial screening examination) and that the longitudinal characteristics of the mortality rate

---

3) Note that this is purely an empirical matter.

4) See the appendix for the extension of (2.1) to subject-years.

can only be investigated by the introduction of an interaction term $z_i t$ in (2.9).

The log partial likelihood without interaction terms follows from expressions (2.8) and (2.9):

$$(2.10) \qquad \ell n\, L_2 = \delta_a \sum_{i \in M} a_i + \delta' \sum_{i \in M} z_i - \sum_{i \in M} \ell n \Big\{ \sum_{r \in H_i} \exp(\delta_a a_r + \delta' z_r) \Big\},$$

which has to be maximized with respect to δ_a and δ'. For this purpose program 2L of the BMDP-package (1981) has been used.

3. Logistic analysis

With logistic analysis the probability of death within some specified time period of follow-up of a subject, who participated in the initial screening examination at the start of this period, is specified as a logistic function of a number of covariables measured at this examination. Logistic analysis, however, does not utilize all the information contained in the data of a prospective (incidence) study; see Green and Symons (1983). More specifically, logistic analysis does not take the actual follow-up time into account, despite its variation among subjects, and thus cannot properly deal with competing failure causes, although it is still often recommended and applied for prospective studies in epidemiology.[5] It is appropriate for cross-sectional (prevalence) studies, where no follow-up time is present.

The probability that a subject i with age a_i and covariable vector z_i dies within some specified time period, is specified as a logistic function of a_i and z_i:

$$(3.1) \qquad \Pr[\text{death} \mid a_i, z_i] = \{1 + \exp(-\zeta_0 - \zeta_a a_i - \zeta' z_i)\}^{-1}$$

where ζ_0, ζ_a and row vector ζ' are to be estimated. From the logistic function (3.1) the following log likelihood can be derived:

$$(3.2) \qquad \ell n\, L_3 = m\zeta_0 + \zeta_a \sum_{i \in M} a_i + \zeta' \sum_{i \in M} z_i - \sum_{i \in H} \ell n\{1 + \exp(\zeta_0 + \zeta_a a_i + \zeta' z_i)\}.$$

This log likelihood has to be maximized with respect to ζ_0, ζ_a and ζ'. For this purpose program LR of the BMDP-package (1981) has been used.

---

5) This disadvantage of logistic analysis becomes of less importance as the follow-up time interval becomes shorter and thus the number of mortality cases becomes relatively smaller.

4. The data

The covariables measured at the initial screening examination and chosen for the mortality analyses are listed and defined in Table 1.[6] Of the 3,365 participants, a number of 3,284 subjects had non-missing observations on all of these covariables. Of these 3,284 subjects, 342 died and 44 emigrated during the follow-up time interval between the initial screening examination and the censoring date of 1 March 1982. Table 1 also gives the range, mean and standard deviation for each covariable in the group of 3,284 subjects. The correlation matrix of these covariables is given in Table 2, again for the group of 3,284 subjects.

Besides the above covariables, the following two variables are also necessary for survival analysis:

(i) the follow-up time, which is defined in years and also given in Tables 1 and 2;

(ii) a failure cause indicator, which is defined here as 0 for non-failure (2,898 subjects), 1 for failure due to death (342 subjects) and 2 for failure due to emigration (44 subjects).

As mentioned in Section 3, follow-up time is not used in logistic analysis. As mentioned in Section 2.1, failure cause 1 (death) is the failure cause of interest and failure cause 2 (emigration) is treated as censoring along with "failure cause" 0 (non-failure).

---

6) Tables and figures may be found after the appendix.

5. Results

5.1. Introduction

The major emphasis is on an application of the full likelihood approach for a log-linear dependency model of the mortality rate, the results of which are compared with those from Cox's partial likelihood approach and a logistic model of the mortality probability. As a criterium helpful for these comparisons, but mainly used for covariable selection, the large-sample chi-square likelihood ratio statistic is used. This statistic is defined as

$$(5.1) \qquad \chi^2_{LR}(\nu) = 2(\ell n\, L_f - \ell n\, L_r)$$

and is based on the ratio of the likelihood L_f of the relevant full model to the likelihood L_r of a restricted model. The number ν of degrees of freedom is the difference in the number of covariables between both models.

5.2. Full likelihood approach: constant rates

For computational simplicity use is made of model (2.2), i.e., with the mortality rate assumed constant in time, in order to select the eventual set of covariables relevant for total mortality. Subsequently it is checked in Section 5.3 if the results with this selected and relevant set of covariables remain similar when the computationally more complicated models (2.6) and (2.7) are assumed, i.e., with also the longitudinal information included in the mortality rate as described in Section 2.2.

The eventually selected set of covariables, with their estimated coefficients (including estimated standard errors, correlations and t-ratio's) and the likelihood ratio chi-square for including all covariables listed in addition to the constant, are presented in Table 3. An interesting result is that a model with a first as well as a second-order term for systolic blood pressure and cholesterol performes significantly better than a model with either one of these terms for both these covariables, despite the large collinearity between these terms for each covariable. (The correlation coefficient is 0.99 for systolic blood pressure as well as for cholesterol.) Including second order terms in addition to the first order terms for systolic blood pressure and

cholesterol gives a significant[7] improvement of the likelihood: $\chi^2_{LR}(2) = 8.469$; including first order terms in addition to the second order terms also gives a significant improvement of $\chi^2_{LR}(2) = 7.217$. Both these likelihood ratio tests have been computed with a full model containing the 9 covariables listed in Table 3. The first and second order terms for diastolic blood pressure, when used instead of these terms for systolic blood pressure, do not produce significant effects on the mortality rate.

Adding diabetes mellitus to the model containing the 9 covariables listed in Table 3 does not give a significant improvement: $\chi^2_{LR}(1) = 0.969$. Also no significant improvement results from adding education, alcohol, Quetelet Index and their quadratic terms to the model: $\chi^2_{LR}(6) = 3.31$. With respect to smoking, the binary variable " cigarette smoking" appears to be the only variable of interest; additionally including the number of cigarettes smoked does not yield a significant improvement: $\chi^2_{LR}(1) = 0.003$. Lately, it has been found by Garrison et al. (1983) from the Framingham Heart Study data that "lean smokers" demonstrate an excessive mortality risk. This hypothesis has been tested here by adding the Quetelet Index, its quadratic term and the interaction of both these terms with "cigarette smoking" to the model, yielding no significant improvement however: $\chi^2_{LR}(4) = 0.911$. Finally, the interaction of age, as measured at the initial screening, with the other 8 covariables listed in Table 3 has been investigated by adding 8 interaction terms to the model, which also does not yield a significant improvement: $\chi^2_{LR}(8) = 6.258$. All five likelihood ratio tests of this paragraph have been computed with a restricted model containing the 9 covariables listed in Table 3.

In this section the covariables of Table 1 have been used to explain the observed total death rate. Although these covariables have been selected for their relevancy for explaining the death rate from cardiovascular diseases (being approximately forty percent of the total death rate for males in the age category 45-59) the set of covariables of Table 1 may also be regarded as an indicator of the general health condition, with emphasis on the cardiovascular part: ECG anomaly and angina pectoris condition. Not surprisingly this part of the covariables also returns in the explanation of total death. If other variables indicating a frequently occurring special high risk had been present, these variables would also be significant in "explaining" the total

7) Significance levels (in the individual tests) of 5 percent have been used throughout.

death rate. The presence of these variables (here ECG1, ECG2 and AP) evidently allows for less biased estimation of the effects of the "true" more general health indicators.

5.3. Full likelihood approach: variable rates

In models (2.6) and (2.7) follow-up time is included, together with the 9 covariables listed in Table 3, as the discrete variable t = 0, 1, 2, ..., measuring the (t+1)-th follow-up year of a subject. The second term of the log likelihood (2.1), where the summation is over subjects $i \in H$, is therefore evaluated over the set of subject-years, which set is approximately 9 times as large as the set H of subjects. Of course, computing time increases almost correspondingly.

In a first analysis the follow-up year t is simply added: this is model (2.6) without the interaction variables $z_i t$, with z_i the vector of covariables of Table 3, except AGE. This already produces a substantial improvement: $\chi^2_{LR}(1) = 21.762$. In a second analysis also the interaction terms $z_i t$ are added, so as to test $\gamma_* = 0$ in model (2.6). No significant improvement results from adding these 8 interaction terms: $\chi^2_{LR}(8) = 6.603$, implying that the hypothesis $\gamma_* = 0$ in model (2.6) cannot be rejected.

Going back to the first analysis, one may now test the equality of the coefficients γ_a for age a_i and $\bar{\gamma}$ for follow-up year t, assuming $\gamma_* = 0$. The estimated difference equals 0.00475 with a standard error of 0.0248 and thus is not significantly different from zero. This implies that age a_i and follow-up year t can be combined into age in year t as in model (2.7), which is a simple and appealing way to extract the longitudinal information contained in the dataset at hand.

The results of the analysis with model (2.7) are presented in Table 4 in the same way as in Table 3. The estimates in Table 4 are quite similar to those in Table 3, as they should according to the above tests. Table 5 and Figures 1 and 2 give an interpretation of the estimates in Table 4 in terms of the mortality rate ratio. The minimum mortality points for systolic blood pressure and plasma cholesterol content are estimated from Table 4 as 134.3 mm Hg and as 207.5 mg/100 ml, respectively. It is worth noting that these estimated minimum mortality points are near the sample means of 138.1 mm Hg and 201.7 mg/100 ml for systolic blood pressure and plasma cholesterol content,

respectively; see Table 1. The estimatated standard errors[8] for the minimum mortality points are 13.25 and 5.67 for systolic blood pressure and plasma cholesterol content, respectively.

5.4. Comparison with partial likelihood approach and logistic regression

In a first attempt to maximize the log partial likelihood (2.10) with the 9 covariables listed in Table 3, the computational procedure of program 2L (proportional hazards model) of the BMDP-package (1981) did not converge to a solution. However, after replacing the two dummy covariables ECG1 and ECG2 by their sum, program 2L converged to a solution. The combination of ECG1 and ECG2 in their sum is justified by their insignificantly differing estimated separate coefficients in Table 4.

In Table 6 the results of four analyses with ECG1 + ECG2 are presented together:

(i) full likelihood approach with model (2.2);

(ii) full likelihood approach with model (2.7);

(iii) partial likelihood approach;

(iv) logistic regression.

For an interpretation of the estimated coefficients in these four analyses it has to be noted that in analyses (i) - (iii) logarithms of rate ratios are estimated and that in analysis (iv) logarithms of probability odds ratios are estimated. The results from analyses (i) - (iii) thus are comparable, while the estimated coefficients from analysis (iv) are in an absolute sense larger than those from analyses (i) - (iii).

It is interesting to compare the chi-square likelihood ratio statistics for including all 8 covariables in addition to the constant for these four analyses. It appears that analysis (ii) yields a substantially higher chi-square than the other analyses, which is caused by the fact that the explanatory variable AGE in analysis (ii) not only represents the cross-sectional age structure at the initial screening examination, but, AGE being the sum of initial age and follow-up year t, it also respresents the effect of (longitudinal) aging. It further appears that analyses (i), (iii) and (iv)

---

8) Here and in Table 5 the estimated standard errors have been calculated using a theorem in Rao (1973, pp. 387-389) on the distribution of a function of statistics. The functions concerned have an asymptotically normal distribution.

yield approximately the same chi-square, where particularly the logistic analysis (iv) merits some attention. As mentioned in Section 3 logistic analysis does not utilize all the information from the data as the stochastic mortality process is simply represented by a random binary outcome, whereas in the other analyses this process is represented by the continuous random variable time of death. Hence, one would expect, as in Green and Symons (1983), the logistic analysis (iv) to yield a markedly lower chi-square than analyses (i) and (iii) considering the long follow-up time interval of approximately 9 years. The reason why apparently not much information is lost by considering the stochastic mortality process only as a simple random binary outcome is the relatively small number of mortality cases despite the long follow-up interval.

6. Discussion

In this paper two mathematical descriptions of the stochastic mortality process are considered: a crude description in terms of probabilities of death during specified time intervals of follow-up conditional upon being alive at the start of these intervals and a more sophisticated description in terms of mortality rates λ, with λdt the probability of death during an infinitely small time interval $(t, t+dt)$, given alive at time t. When measuring the incidence of a relatively rare event, such as mortality observed in a cohort over a small time interval, the distinction between a rate and a probability may seem somewhat academic. However, a description in terms of rates is mathematically much more convenient as a rate is directly related to the probability density function of the continuous random variable time of death. This facilitates the generalization in various directions, such as longer follow-up intervals where death becomes a much less rare event and the presence of competing risks, when the total death rate is the sum of a number of cause specific death rates. When using probabilities of binary outcomes in these more general circumstances one may run into problems, since the information provided by only a binary mortality outcome, gives too crude a description of the mortality process.

Another interesting point to discuss is the dependency specification. In this paper an exponential dependency model is chosen for rates, see expressions (2.2) and (2.9), and a logistic dependency model is chosen for probabilities, see expression (3.1). The reasons for these choices are epidemiological tradition and computational simplicity. As to tradition: coefficients of the exponential dependency model represent logarithms of rate ratios and the coefficients of the logistic dependency model represent logarithms of probability odds ratios, both ratios being well-known measures of association in epidemiology. As to computational simplicity: an exponential function is always positive, as a rate should be, and a logistic function is always in the interval (0,1), as a probability should be, without the need to make any restrictions on the coefficients in both dependency models. It has to be mentioned, however, that no clear biological mechanism leads to these dependency models and that the mortality process is looked at in an isolated way.

Recent developments in this area (Woodbury et al. 1977, 1979 and 1981) consider the total life trajectory of a longitudinally followed cohort in the presence of the simultaneous processes of aging and mortality selection. An

analytically elegant and tractable way derived by Woodbury et al. is to assume that (i) the covariables obey an autoregressive structure in surviving subjects of the cohort (the aging process), and (ii) the mortality rate depends quadratically on the covariables (a convexly shaped mortality selection process). These assumptions then comply with a multivariate normal distribution of the covariables in the surviving cohort, with the attractive property that the time-behaviour of the mean vector and covariance matrix as well as the unconditional survival function can be analytically expressed into the parameters of the aging process and the mortality selection process. This facilitates the possibility of making future projections of the total life trajectory of a cohort. However, the maximum likelihood procedure for estimating the quadratic dependency model for the mortality rate is computationally more complicated as constraints on the parameters of this quadratic dependency model need to be made in order to guarantee the positiveness of the mortality rate. The empirical results in this paper confirm the convexly shaped relationship as postulated by Woodbury et al. between the mortality rate and the continuous explanatory covariables systolic blood pressure and plasma cholesterol, although here as a quadratic function in the exponent, for ease of estimation. Of course, one should be careful when interpreting the increasing mortality rate associated with decreasing systolic blood pressure or plasma cholesterol in terms of cause and effect.

Appendix on the log likelihood (2.1)

Subject i of the total cohort H is followed during the time interval $[0, t_i]$. The only type of censoring present is of Type I, i.e., the total follow-up time t_i of subject i is either predetermined or stochastic and it is stochastic only if one of K failure causes produces time t_i. Correspondingly, two types of events are defined for each subject $i \in H$:

(i) event M_0, occurring if the follow-up ends without failure at predetermined time t_i;

(ii) event M_j ($j = 1,\ldots,K$), occurring if the follow-up ends at stochastic time t_i, produced by failure cause j ($j = 1,\ldots,K$).

The set H is partitioned into the sets $M_0, M_1, \ldots, M_K$. Each subject $i \in H$ belongs to exactly one of these sets. This is also denoted by the zero-one indicators $\delta_{0i}, \delta_{1i}, \ldots, \delta_{Ki}$, which are defined as $\delta_{ji} = 1$ if $i \in M_j$ ($j = 0, 1, \ldots, K$) and $\delta_{ji} = 0$ otherwise.

The likelihood function L_i for subject $i \in H$ can now be written as:

$$(A.1) \qquad L_i = [\bar{F}_i(t_i)]^{\delta_{0i}} \prod_{j=1}^{K} [\lambda_{ji}(t_i)\bar{F}_i(t_i)]^{\delta_{ji}}$$

where $\lambda_{ji}(t_i)$ is the failure rate at time t_i from failure cause j ($j = 1,\ldots,K$) in subject $i \in H$ and $\bar{F}_i(t_i)$ is the survival function at time t_i of subject $i \in H$:

$$(A.2) \qquad \bar{F}_i(t_i) = \exp\Big[-\int_0^{t_i}\Big\{\sum_{j=1}^{K} \lambda_{ji}(x)\Big\}dx\Big].$$

In expression (A.1) $\bar{F}_i(t_i)$ denotes, for subject $i \in M_0$, the survival probability for predetermined time t_i and $\lambda_{ji}(t_i)\bar{F}_i(t_i)dt_i$ denotes the probability of failure from cause j at stochastic time t_i for subject $i \in M_j$ ($j = 1,\ldots,K$). The likelihood function for all subjects in H is, because of independence:

$$(A.3) \qquad L = \prod_{i\in H} L_i = \prod_{i\in M_0} \bar{F}_i(t_i) \prod_{j=1}^{K} \prod_{i\in M_j} \lambda_{ji}(t_i)\bar{F}_i(t_i).$$

This can be expressed into the rates λ_{ji} as follows:

$$(A.4) \qquad L = \prod_{j=1}^{K} \Big\{ \prod_{i\in M_j} \lambda_{ji}(t_i) \prod_{i\in H} \exp\Big[-\int_0^{t_i}\lambda_{ji}(x)dx\Big]\Big\},$$

which follows from the survival function (A.2). This implies that $\ell n\ L$ is additively separable with respect to the causes $j = 1,\ldots,K$:

$$(A.5) \qquad \ell n\ L = \sum_{j=1}^{K} \Big\{ \sum_{i \in M_j} \ell n[\lambda_{ji}(t_i)] - \sum_{i \in H} \int_0^{t_i} \lambda_{ji}(x)dx \Big\}.$$

A separate maximization for each failure cause j maximizes the log likelihood (A.5). Suppressing the index j for a particular failure cause, one has from the log likelihood (A.5):

$$(A.6) \qquad \ell n\ L = \sum_{i \in M} \ell n[\lambda_i(t_i)] - \sum_{i \in H} \int_0^{t_i} \lambda_i(x)dx,$$

which equals the log likelihood (2.1).

The extension to subject-years, assuming constancy of the failure rates λ_{ji} in each year of the follow-up interval, and with i now denoting subject-year i and w_i the length of subject-year i, is as follows:

$$(A.7) \qquad \ell n\ L = \sum_{j=1}^{K} \Big\{ \sum_{i \in M_j} \ell n\ \lambda_{ji} - \sum_{i \in H'} w_i\ \lambda_{ji} \Big\}.$$

with M_j and H' now sets of subject-years.

Table 1. A selected number of variables measured in 3,284 subjects participating in the initial screening examination.

variable	symbol	definition	min.	max.	mean	stand. dev.
systolic blood pressure	SBP	mm Hg	84	246	138.055	20.204
diastolic blood pressure	DBP	mm Hg	26	164	79.971	12.144
cholesterol	CH	mg/100 ml plasma	91	428	201.722	34.036
age	AGE	years	44	60	52.544	4.285
ECG anomaly 1	ECG1	= 1 for Minnesota codes I-1,2 = 0 elsewhere	0	1	0.022	0.145
ECG anomaly 2	ECG2	= 1 for Minnesota codes I-3, IV-1,2,3, V-1,2,3 or VII-1 = 0 elsewhere	0	1	0.081	0.272
angina pectoris	AP	= 1 for a positive Rose questionnaire (Rose and Blackburn 1968) = 0 elsewhere	0	1	0.058	0.234
cigarette smoking	SMOK	= 1 yes = 0 no	0	1	0.635	0.482
number of cigarettes smoked	CIG	= 0 for none = 1 for 1-9 cigarettes/day = 2 for 10-19 cigarettes/day = 3 for $\geq$ 20 cigarettes/day	0	3	1.408	1.221

Table 1 (continued)

variable	symbol	definition	min.	max.	mean	stand. dev.
diabetes mellitus	DM	= 1 yes = 0 no	0	1	0.018	0.132
alcohol use	ALC	= 1 for no use or less than once a month = 2 for regularly or daily a moderate amount = 3 for regularly or daily a large amount	1	3	1.865	0.653
Quetelet index	QI	weight/height^2 (kg/m^2)	16.7	47.9	25.514	3.002
education	EDUC	ordered categories from primary education (1) to university (9)	1	9	3.383	2.621
follow-up time	TIME	years	0.022	9.455	8.464	1.565

Table 2. Correlation matrix of variables listed in Table 1; each correlation coefficient computed from 3,284 pairs.

	SBP	DBP	CH	AGE	ECG1	ECG2	AP	SMOK	CIG	DM	ALC	QI	EDUC	TIME
SBP	1.00													
DBP	0.70	1.00												
CH	0.09	0.08	1.00											
AGE	0.14	0.04	-0.02	1.00										
ECG1	0.02	0.01	0.05	0.07	1.00									
ECG2	0.16	0.12	0.05	0.13	0.22	1.00								
AP	-0.00	-0.00	0.06	0.07	0.13	0.11	1.00							
SMOK	-0.09	-0.13	0.01	-0.05	-0.01	-0.04	-0.01	1.00						
CIG	-0.08	-0.14	0.02	-0.06	-0.02	-0.04	-0.04	0.88	1.00					
DM	0.05	0.05	0.01	0.04	0.03	0.05	0.06	-0.02	-0.02	1.00				
ALC	0.04	0.06	0.05	-0.04	-0.00	-0.00	-0.02	0.01	0.05	-0.02	1.00			
QI	0.19	0.22	0.16	0.01	-0.02	0.07	-0.01	-0.16	-0.11	0.02	0.07	1.00		
EDUC	0.02	0.06	0.03	-0.11	-0.00	0.00	-0.11	-0.24	-0.21	0.00	0.08	0.02	1.00	
TIME	-0.05	-0.04	-0.02	-0.11	-0.12	-0.14	-0.09	-0.05	-0.05	-0.03	-0.00	0.01	0.03	1.00

Table 3. Maximum likelihood estimates (full likelihood approach) with model (2.2) for the mortality rate and with 3,284 subjects.

	variable	est. coeff.	est. SE	est. coeff./est. SE
0.	constant	-5.97821	1.76629	-3.385
1.	SBP/10	-0.32186	0.18410	-1.748
2.	$SBP^2/1000$	0.11998	0.05983	2.005
3.	CH/10	-0.17506	0.07692	-2.276
4.	$CH^2/1000$	0.04204	0.01683	2.498
5.	AGE	0.09235	0.01334	6.923
6.	SMOK	0.46515	0.12024	3.869
7.	ECG1	0.72672	0.22765	3.192
8.	ECG2	0.81527	0.14702	5.545
9.	AP	0.73095	0.16460	4.441

$\chi^2_{LR}(9) = 173.574$

correlation matrix of estimated coefficients

	0	1	2	3	4	5	6	7	8	9
0	1.00									
1	-0.78	1.00								
2	0.77	-0.99	1.00							
3	-0.44	-0.03	0.02	1.00						
4	0.42	0.02	-0.01	-0.98	1.00					
5	-0.40	0.02	-0.04	-0.03	0.05	1.00				
6	-0.11	0.02	-0.01	0.03	-0.03	0.05	1.00			
7	0.07	-0.04	0.04	-0.06	0.05	-0.01	0.00	1.00		
8	0.03	0.04	-0.07	-0.02	0.01	-0.10	0.05	-0.28	1.00	
9	-0.06	0.03	-0.03	0.10	-0.13	-0.04	0.02	-0.18	-0.09	1.00

Table 4. Maximum likelihood estimates (full likelihood approach) with model (2.7) for the mortality rate and with 29,398 subject-years.

	variable	est. coeff.	est. SE	est. coeff./est. SE
0.	constant	-6.48486	1.74746	-3.711
1.	SBP/10	-0.32722	0.18484	-1.770
2.	$SBP^2/1000$	0.12186	0.06008	2.028
3.	CH/10	-0.17561	0.07629	-2.302
4.	$CH^2/1000$	0.04231	0.01666	2.540
5.	AGE*	0.09492	0.01140	8.326
6.	SMOK	0.47293	0.12028	3.932
7.	ECG1	0.74920	0.22702	3.300
8.	ECG2	0.83022	0.14628	5.676
9.	AP	0.74590	0.16402	4.548

* age in year t

$\chi^2_{LR}(9) = 195.299$

correlation matrix of estimated coefficients

	0	1	2	3	4	5	6	7	8	9
0.	1.00									
1.	-0.79	1.00								
2.	0.78	-0.99	1.00							
3.	-0.45	-0.03	0.02	1.00						
4.	0.43	0.02	-0.01	-0.98	1.00					
5.	-0.38	0.02	-0.03	-0.03	0.04	1.00				
6.	-0.11	0.03	-0.02	0.03	-0.03	0.05	1.00			
7.	0.06	-0.04	0.04	-0.05	0.05	0.00	0.00	1.00		
8.	0.01	0.04	-0.07	-0.03	0.02	-0.07	0.05	-0.28	1.00	
9.	-0.07	0.04	-0.02	0.10	-0.13	-0.02	0.02	-0.17	-0.09	1.00

Table 5. Effects of covariables on the mortality rate in terms of the rate ratio (RR); interpretation of the estimates in Table 4.

variable	est.RR	est.SE*)
AGE (1 year increase)	1.10	0.013
SMOK	1.60	0.193
ECG1	2.12	0.480
ECG2	2.29	0.337
AP	2.11	0.346

For systolic blood pressure and cholesterol see Figures 1 and 2 respectively.

*) See footnote 8.

Figure 1. Interpretation of the estimates in Table 4: the mortality rate ratio (RR) as a function of systolic blood pressure (SBP); the minimum mortality point is 134.3 mm Hg, where RR ≡ 1.

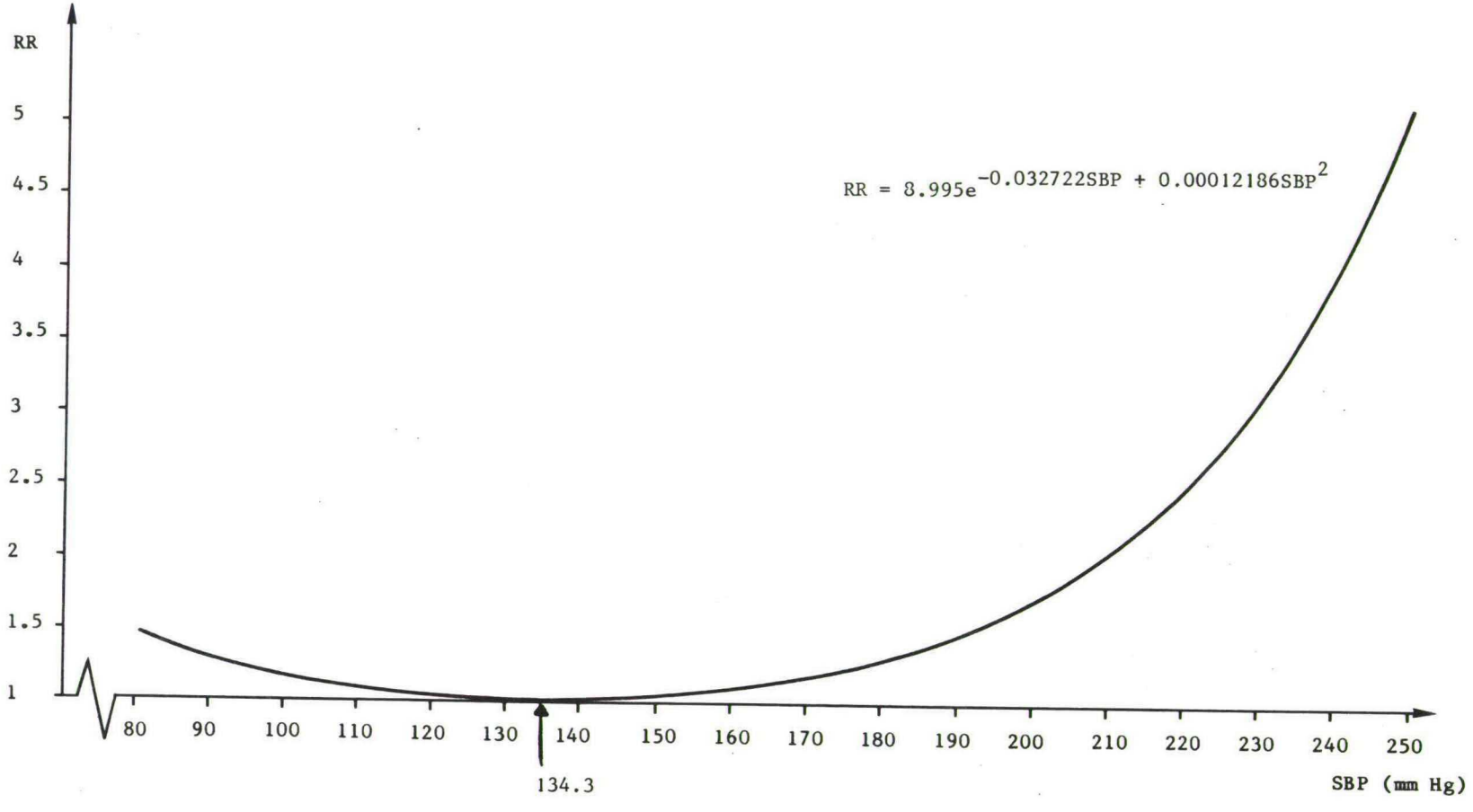

Figure 2. Interpretation of the estimates in Table 4: the mortality rate ratio (RR) as a function of cholesterol (CH); the minimum mortality point is 207.5 mg/100 ml plasma, where RR ≡ 1.

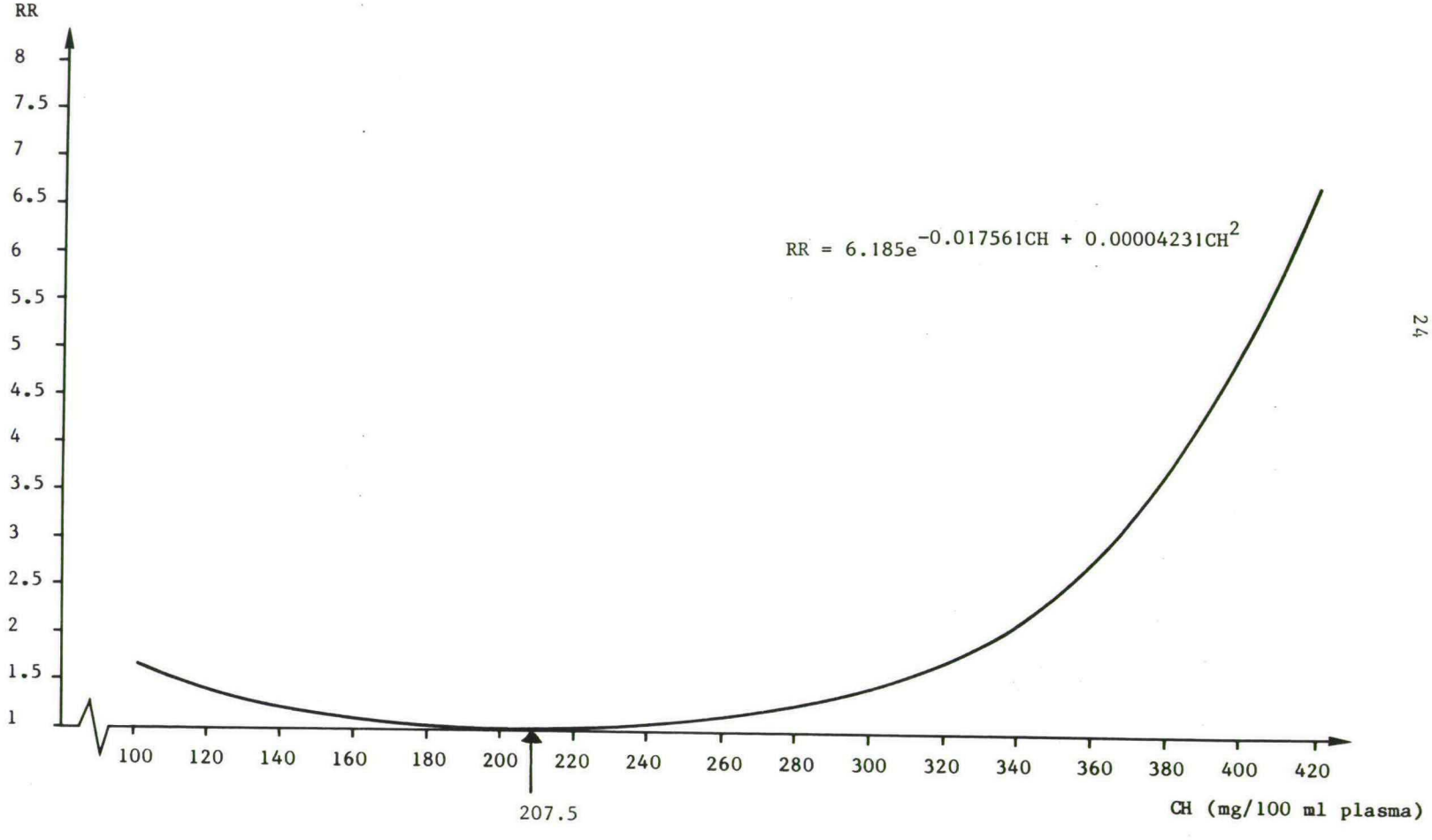

Table 6. Comparison of four analyses.

	analysis (i) full likelihood with model (2.2)	analysis (ii) full likelihood with model (2.7)	analysis (iii) Cox's partial likelihood	analysis (iv) logistic regression
variable	coefficient (SE)	coefficient (SE)	coefficient (SE)	coefficient (SE)
constant	-5.9590 (1.76446)	-6.46784 (1.74575)	-	-2.757 (2.131)
SBP/10	-0.32471 (0.18376)	-0.32978 (0.18450)	-0.3329 (0.1849)	-0.441 (0.221)
$SBP^2/1000$	0.12117 (0.05966)	0.12293 (0.05991)	0.1242 (0.0601)	0.163 (0.073)
CH/10	-0.17578 (0.07685)	-0.17619 (0.07625)	-0.1770 (0.0764)	-0.237 (0.106)
$CH^2/1000$	0.04219 (0.01682)	0.04242 (0.01665)	0.0425 (0.0167)	0.057 (0.024)
AGE*	0.09251 (0.01333)	0.09503 (0.01139)	0.0935 (0.0133)	0.100 (0.015)
SMOK	0.46439 (0.12020)	0.47223 (0.12024)	0.4708 (0.1203)	0.518 (0.131)
ECG1 + ECG2	0.78514 (0.10542)	0.80273 (0.10495)	0.8026 (0.1054)	0.909 (0.130)
AP	0.72681 (0.16408)	0.74228 (0.16354)	0.7430 (0.1638)	0.864 (0.193)
$\chi^2_{LR}(8)$	173.49	195.23	177.55	177.18

* AGE = age in year t in analysis (ii); AGE = age at initial screening (t = 0) in other analyses;

REFERENCES

BMDP Statistical Software manual, University of California Press, Los Angeles (1981).

Cox, D.R., "Regression Models and Life-Tables", Journal of the Royal Statistical Society, Series B, Vol. 34, 187-202 (1972).

Cox, D.R., "Partial Likelihood", Biometrika, Vol. 62, 269-276 (1975).

David, H.A., and M.L. Moeschberger, The Theory of Competing Risks, Charles Griffin & Company Ltd, London (1978).

Garrison, R.J., Feinleib, M., et al., "Cigarette Smoking as a Confounder in the Relationship between Relative Weight and Long Term Mortality", Journal of the American Medical Association, Vol. 249, 2199-2203 (1983).

Glasunov, I.S., Dowd, J.E., et al. (editors), The Kaunas Rotterdam Intervention Study, Elsevier/North-Holland Biomedical Press, Amsterdam (1981).

Green, M.S., Symons, M.J., "A Comparison of the Logistic Risk Function and the Proportional Hazards Model in Prospective Epidemiologic Studies", Journal of Chronic Diseases, Vol. 36, 715-724 (1983).

Holford, T.R., "Life Tables with Concomitant Information", Biometrics, Vol. 32, 587-597 (1976).

Mulder, P.G.H. and Hempenius, A.L., Evaluating Life Time in a Competing Risks Model for a Chronic Disease, Research Memorandum FEW 121, Tilburg University (1982).

Rao, C.R., Linear Statistical Inference and Its Applications, 2nd. ed., John Wiley & Sons, New York (1973).

Rose, G.A. and Blackburn, H., Cardiovascular Survey Methods, WHO, Geneva (1968).

Woodbury, M.A. and Manton, K.G., "A Random-Walk Model of Human Mortality and Aging", Theoretical Population Biology, Vol. 11, 37-48 (1977).

Woodbury, M.A., Manton, K.G., et al., "Longitudinal Analysis of the Dynamics and Risks of Coronary Heart Disease in the Framingham Study", Biometrics, Vol. 35, 575-585 (1979).

Woodbury, M.A., Manton, K.G., et al., "A Dynamic Analysis of Chronic Disease Development: A Study of Sex Specific Changes in Coronary Heart Disease Incidence and Risk Factors in Framingham", International Journal of Epidemiology, Vol. 10, 355-366 (1981).

IN 1983 REEDS VERSCHENEN

126 H.H. Tigelaar
Identification of noisy linear systems with multiple arma inputs.

127 J.P.C. Kleijnen
Statistical Analysis of Steady-State Simulations: Survey of Recent Progress.

128 A.J. de Zeeuw
Two notes on Nash and Information.

129 H.L. Theuns en A.M.L. Passier-Grootjans
Toeristische ontwikkeling - voorwaarden en systematiek; een selectief literatuuroverzicht.

130 J. Plasmans en V. Somers
A Maximum Likelihood Estimation Method of a Three Market Disequilibrium Model.

131 R. van Montfort, R. Schippers, R. Heuts
Johnson S_U-transformations for parameter estimation in arma-models when data are non-gaussian.

132 J. Glombowski en M. Krüger
On the Rôle of Distribution in Different Theories of Cyclical Growth.

133 J.W.A. Vingerhoets en H.J.A. Coppens
Internationale Grondstoffenovereenkomsten.
Effecten, kosten en oligopolisten.

134 W.J. Oomens
The economic interpretation of the advertising effect of Lydia Pinkham.

135 J.P.C. Kleijnen
Regression analysis: assumptions, alternatives, applications.

136 J.P.C. Kleijnen
On the interpretation of variables.

137 G. van der Laan en A.J.J. Talman
Simplicial approximation of solutions to the nonlinear complementarity problem with lower and upper bounds.

IN 1984 REEDS VERSCHENEN

138 G.J. Cuypers, J.P.C. Kleijnen en J.W.M. van Rooyen
Testing the Mean of an Asymetric Population:
Four Procedures Evaluated

139 T. Wansbeek en A. Kapteyn
Estimation in a linear model with serially correlated errors when observations are missing

140 A. Kapteyn, S. van de Geer, H. van de Stadt, T. Wansbeek
Interdependent preferences: an econometric analysis

141 W.J.H. van Groenendaal
Discrete and continuous univariate modelling

142 J.P.C. Kleijnen, P. Cremers, F. van Belle
The power of weighted and ordinary least squares with estimated unequal variances in experimental design

143 J.P.C. Kleijnen
Superefficient estimation of power functions in simulation experiments

144 P.A. Bekker, D.S.G. Pollock
Identification of linear stochastic models with covariance restrictions.

145 Max D. Merbis, Aart J. de Zeeuw
From structural form to state-space form

146 T.M. Doup and A.J.J. Talman
A new variable dimension simplicial algorithm to find equilibria on the product space of unit simplices.

147 G. van der Laan, A.J.J. Talman and L. Van der Heyden
Variable dimension algorithms for unproper labellings.

148 G.J.C.Th. van Schijndel
Dynamic firm behaviour and financial leverage clienteles

149 M. Plattel, J. Peil
The ethico-political and theoretical reconstruction of contemporary economic doctrines

150 F.J.A.M. Hoes, C.W. Vroom
Japanese Business Policy: The Cash Flow Triangle
an exercise in sociological demystification

151 T.M. Doup, G. van der Laan and A.J.J. Talman
The $(2^{n+1}-2)$-ray algorithm: a new simplicial algorithm to compute economic equilibria