



Munich Personal RePEc Archive

**The use of a parametric and a  
semi-parametric estimation method for  
the binary choice model: Probit  
Maximum Likelihood versus Maximum  
Score**

Peeters, H.M.M.

University of Tilburg

May 1989

Online at <https://mpra.ub.uni-muenchen.de/28104/>

MPRA Paper No. 28104, posted 13 Jan 2011 08:32 UTC

**Het gebruik van een parametrische en een  
semi-parametrische schattingsmethode voor  
het binaire keuzemodel :**

**Probit Maximum Likelihood versus Maximum Score**

**H.M.M. Peeters**

**April, 1989**

**Tilburg, K.U.B.**

## Voorwoord.

Met deze scriptie sluit ik mijn studie econometrie aan de Katholieke Universiteit Brabant af.

Bertrand Melenberg en Theo Nijman hebben mij het idee voor de scriptie aan de hand gedaan en de begeleiding tijdens de uitwerking van het idee verzorgd. De data die gebruikt zijn bij een toepassing (in paragraaf 6) zijn door Arthur van Soest ter beschikking gesteld. Rob Aalbers heeft mij tijdens het schrijven van de scriptie enkele nuttige adviezen gegeven.

Al deze personen wil ik hartelijk bedanken voor hun bijdrage aan de totstandkoming van deze scriptie.

Marga Peeters

## Inhoudsopgave.

Paragraaf 1	Inleiding / 1
Paragraaf 2	Probit Maximum Likelihood en een toets op normaliteit / 4
Paragraaf 3	Maximum Score van Manski / 11
Paragraaf 4	"Bootstrap"-methoden / 18
Paragraaf 5	Enkele Monte-Carlo-experimenten / 28
Paragraaf 6	Toepassingen / 45
Paragraaf 7	Conclusies / 59
	Referenties / 63



## § 1 Inleiding.

Bij het modelleren van relaties tussen economische grootheden is het gebruikelijk om veronderstellingen te maken met betrekking tot de functionele vorm van het model en niet waargenomen variabelen. Vaak wordt er voor een model gekozen met een eindig aantal onbekende parameters. Deze onbekende parameters kunnen dan worden geschat met een schattingsmethode die gebaseerd is op onder andere de gemaakte veronderstellingen.

De meest bekende econometrische modellen verklaren de endogene variabele(n) rechtstreeks of via een achterliggend latent model, lineair uit de exogene variabelen. Het niet verklaarde gedeelte van een endogene wordt gerepresenteerd door een storingsterm. Ten aanzien van de niet waar te nemen storingen wordt vaak verondersteld dat ze onafhankelijk en identiek verdeeld zijn. In heel veel modellen wordt zelfs het 'identiek' verdeeld zijn, vervangen door 'identiek normaal' verdeeld zijn. Door dergelijke veronderstellingen te maken wordt het zoeken naar een consistente en efficiënte schattingsmethode, aanzienlijk vereenvoudigd.

Het standaard lineaire regressiemodel met de sterke condities (zie bijv. Van der Genugten, 1988) is een voorbeeld van een model waarin de storingen onafhankelijk en identiek normaal verdeeld zijn. De schattingsmethode voor dit model, de kleinste kwadraten of OLS-methode, levert consistente en efficiënte schatters. Indien in werkelijkheid de verdeling van de storingen geen normale verdeling is, zal de OLS-schatter aan efficiëntie verliezen, maar vaak nog wel consistent zijn.

In bepaalde andere modellen, waaronder de keuzemodellen, is het maken van een verkeerde verdelingsveronderstelling veel minder onschuldig; het model schatten met de methode die uitgaat van de verkeerde verdeling hoeft dan zelfs geen consistente schatters meer te geven.

Semi-parametrische schattingsmethoden zijn vaak wel consistent en baseren zich niet op een functionele vorm van de storingsterm, wat onder andere impliceert dat ze geen bepaalde verdeling veronderstellen.

In deze scriptie behandelen we één semi-parametrische schattingsmethode: de Maximum-Score-methode van Manski (1988). Deze methode stelt ten aanzien van

de storingen een hele zwakke eis; ze baseert zich alleen op de mediaan van de storingen die, gegeven de exogenen, de waarde nul aanneemt.

Vanwege het feit dat de Maximum-Score-methode een schattingsmethode voor het binaire keuzemodel is, behandelen we dit model uitvoerig.

Het doel dat wij hierbij hebben gesteld, is het vergelijken van de semi-parametrische methode met de Maximum-Likelihood-methode gebaseerd op normaliteit, een methode die wél expliciet gebruik maakt van de verdelingsveronderstelling. Bekend is dat Maximum Likelihood een hele goede schattingsmethode is, die meestal hele prettige eigenschappen heeft (zoals consistentie, asymptotische efficiëntie en asymptotische normaliteit van de Maximum-Likelihood-schatters), mits de methode gebaseerd is op de juiste verdeling. Wij trachten te achterhalen wat de gevolgen kunnen zijn van het verkeerd gebruik van de Maximum-Likelihood-methode, dat wil zeggen Maximum Likelihood baseren op een verkeerde verdelingsveronderstelling en proberen een antwoord te vinden op de vraag hoe 'slecht' Maximum Likelihood en hoe 'goed' Maximum Score in dergelijke gevallen zal zijn.

Om beide schattingsmethoden met elkaar te kunnen vergelijken, bekijken we eerst de theoretische achtergrond. In paragraaf 2 behandelen we de Maximum-Likelihood-methode voor het probitmodel, dat is het binaire keuzemodel met normaal verdeelde storingen. Bovendien bespreken we in deze paragraaf kort een mogelijkheid om te toetsen op de hypothese dat de storingen normaal zijn verdeeld. De wijze waarop het binaire keuzemodel wordt geschat met de Maximum-Score-methode, behandelen we in paragraaf 3. In deze paragraaf zullen we het begrip 'semi-parametrisch' ook nader toelichten. Omdat het voor Maximum-Score-schatters, in tegenstelling tot de Maximum-Likelihood-schatters, niet eenvoudig is om op basis van de asymptotische verdeling standaardfouten, toetsen en betrouwbaarheidsintervallen af te leiden, bespreken we in paragraaf 4 enkele "bootstrap"-methoden. Deze methoden bieden de mogelijkheid om op een vrij eenvoudige manier voor willekeurige schatters, dus ook voor Maximum-Score-schatters, maatstaven van nauwkeurigheid te berekenen.

Om in kleine steekproeven het gebruik van beide methoden te vergelijken, presenteren we in paragraaf 5 de resultaten van enkele Monte-Carlo-experimenten. In paragraaf 6 bespreken we dan de Maximum-Likelihood- en Maximum-Score-resultaten van twee toepassingen. Bij beide toepassingen

passen we de in paragraaf 2 besproken toets op normaliteit toe en geven de resultaten van enkele in paragraaf 4 besproken "bootstrap"-methoden.

In paragraaf 7, de conclusies, zullen we tot slot terugkomen op de vraag welke schattingsmethode voor het binaire keuzemodel het best gebruikt kan worden. We proberen deze vraag te beantwoorden met behulp van de theorie van beide methoden en de resultaten uit paragrafen 5 en 6.



## § 2 Probit Maximum Likelihood en een toets op normaliteit.

Econometrische modellen waarin de afhankelijke variabele slechts een eindig aantal waarden kan aannemen, worden keuzemodellen genoemd. Eén van de meest eenvoudige voorbeelden hiervan is het binaire keuzemodel. In deze paragraaf zullen we nagaan hoe een speciaal geval van de binaire keuzemodellen, nl. het probitmodel, geschat kan worden met de Maximum-Likelihood-methode (ML-methode). Tevens zullen we een eenvoudig uitvoerbare toets op normaliteit kort bespreken.

Het binaire keuzemodel wordt vaak als volgt weergegeven (zie bijv. Maddala (1983), blz. 22):

$$\left. \begin{aligned} y_t^* &= x_t' \beta + \varepsilon_t & \beta \in B \subset \mathbb{R}^k \\ y_t &= 1 \text{ als } y_t^* > 0 \\ y_t &= 0 \text{ als } y_t^* \leq 0 \\ x_t &\sim F_{x_t} \\ \varepsilon_t &\sim F_{\varepsilon_t} \\ (x_t', \varepsilon_t)' &\text{ onafhankelijk van } (x_s', \varepsilon_s)' \quad s \neq t \\ x_t &\text{ onafhankelijk van } \varepsilon_t \end{aligned} \right\} \quad s, t = 1, 2, \dots, T \quad (2.1)$$

De waargenomen afhankelijke variabele  $y_t$  neemt slechts twee waarden aan:  $y_t = 1$  als een bepaalde gebeurtenis plaatsvindt en  $y_t = 0$  als de gebeurtenis niet plaatsvindt. De variabele wordt verklaard door middel van de latente variabele  $y_t^*$  uit  $k$  exogenen  $x_{t1}, x_{t2}, \dots, x_{tk}$  en storingsterm  $\varepsilon_t$ . Ten aanzien van de vector van exogenen  $x_t$  en storingsterm  $\varepsilon_t$  wordt verondersteld dat ze onderling en van elkaar onafhankelijk zijn. De onbekende te schatten parametervector  $\beta$  legt het lineaire verband tussen de latente variabele en de exogenen. Wij zullen uitgaan van het geval dat voor de verdeling  $F_{\varepsilon_t}$  van  $\varepsilon_t$  geldt:  $F_{\varepsilon_t} = N(0, \sigma^2)$ . Dit model wordt het probit- of ook wel het normitmodel genoemd. In deze paragraaf zullen we het probitmodel met deterministische exogenen bespreken. Dit impliceert dat de veronderstellingen met betrekking tot  $x_t$  in (2.1) vervallen. Alle afleidingen die wij maken voor het probitmodel met deterministische exogenen

gelden ook voor het probitmodel met stochastische exogenen als de stochastische grootheden opgevat worden m.b.t. de conditionele verdeling onder  $x_t$ .

De latente variabele  $y_t^*$  overschrijdt de drempelwaarde 0 als  $y_t = 1$ . Noteren we  $\Phi$  als de verdelingsfunctie van de standaardnormale verdeling, dan is de kans hierop:

$$P\{y_t = 1\} = P\{y_t^* > 0\} = P\{\epsilon_t > -x_t'\beta\} = \Phi\left(\frac{x_t'\beta}{\sigma}\right).$$

De kans dat de gebeurtenis niet waargenomen wordt, is dan:

$$P\{y_t = 0\} = 1 - \Phi\left(\frac{x_t'\beta}{\sigma}\right).$$

Definiëren we vervolgens  $I_1 := \{t : y_t = 1\}$  en  $I_0 := \{t : y_t = 0\}$  dan ziet de likelihoodfunctie  $L$  er als volgt uit:

$$\begin{aligned} L &= \prod_{t \in I_1} P\{y_t = 1\} \prod_{t \in I_0} P\{y_t = 0\} \\ &= \prod_{t=1}^T P\{y_t = 1\}^{y_t} P\{y_t = 0\}^{1-y_t} \\ &= \prod_{t=1}^T \Phi\left(\frac{x_t'\beta}{\sigma}\right)^{y_t} \left(1 - \Phi\left(\frac{x_t'\beta}{\sigma}\right)\right)^{1-y_t} \end{aligned} \quad (2.2)$$

De onbekende parameters, de parametervector  $\beta$  en de parameter  $\sigma$ , komen in deze likelihoodfunctie alleen als het quotiënt  $(1/\sigma)\beta$  voor. De likelihoodfunctie kan dan ook slechts met betrekking tot dit quotiënt geoptimaliseerd worden. Maximaliseren van  $L$  of  $\mathcal{L} := \log L$  levert Maximum-Likelihood-schatters (ML-schatters) voor  $\vartheta := (\beta/\sigma)$ . Aangezien  $\beta$  en  $\sigma$  dus niet beide geïdentificeerd zijn, wordt vaak de normalisatie  $\sigma^2 = 1$  genomen. De afgeleide van  $\mathcal{L}$  naar  $\vartheta$  (de zgn. score) is

$$S(\vartheta) := \frac{\partial \mathcal{L}}{\partial \vartheta} = \frac{\partial}{\partial \vartheta} \left[ \sum_{t=1}^T y_t \log \Phi(x_t'\vartheta) + \sum_{t=1}^T (1 - y_t) \log(1 - \Phi(x_t'\vartheta)) \right]$$

$$\begin{aligned}
 &= \sum_{t=1}^T \left[ y_t \frac{\varphi(x_t' \theta)}{\Phi(x_t' \theta)} x_t - (1 - y_t) \frac{\varphi(x_t' \theta)}{1 - \Phi(x_t' \theta)} x_t \right] \\
 &= \sum_{t=1}^T \frac{y_t - \Phi(x_t' \theta)}{\Phi(x_t' \theta)(1 - \Phi(x_t' \theta))} \varphi(x_t' \theta) x_t
 \end{aligned} \tag{2.3}$$

Hierin is  $\varphi$  de standaardnormale dichtheidsfunctie. Omdat de ML-schatter voor  $\theta$  analytisch niet is te berekenen, wordt deze numeriek bepaald. Een numerieke optimaliseringsprocedure is de Newton-Raphson-methode. Deze bepaalt op basis van een schatting  $\hat{\theta}_i$  voor  $\theta$  de schatting

$$\hat{\theta}_{i+1} = \hat{\theta}_i + H(\hat{\theta}_i)^{-1} S(\hat{\theta}_i) \tag{2.4}$$

met  $H(\theta) := -\frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'}$ , en bereikt een minimum voor  $-\mathcal{L}$  (dus een maximum voor  $\mathcal{L}$ ) als convergentie optreedt.

Omdat de log-likelihoodfunctie bij het probitmodel concaaf is (Amemiya (1985), blz. 273-274), is het optimum dat op deze manier wordt gevonden het globale maximum. Dit is een niet onbelangrijk verschil met de schattingsmethode die in de volgende paragraaf zal worden behandeld.

Onder zwakke regulariteitsvoorwaarden is in het binaire keuzemodel de ML-schatter  $\hat{\theta}_{ML}$  een consistente schatter voor  $\theta$ . Bovendien is bekend dat ML-schatters (eveneens onder bepaalde regulariteitsvoorwaarden) asymptotisch normaal zijn en de Cramer-Rao ondergrens bereiken:

$$\sqrt{T}(\hat{\theta}_{ML} - \theta) \overset{a}{\rightsquigarrow} N(0, \lim_{T \rightarrow \infty} [E\{-\frac{1}{T} \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'}\}]^{-1}) \tag{2.5}$$

Als consistente schatter voor de asymptotisch covariantiematrix van  $\hat{\theta}_{ML}$  kan de inverse van de informatiematrix, gedefinieerd als

$$[I(\theta)]^{-1} := [E\{-\frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'}\}]^{-1}, \text{ geëvalueerd in } \hat{\theta}_{ML} \text{ worden genomen. Noteren we}$$

$$\mathcal{L} = \sum_{t=1}^T \mathcal{L}_t \text{ met } \mathcal{L}_t \text{ de loglikelihoodbijdrage van waarneming } t, \text{ dan geldt}$$

voor het probitmodel indien  $T \rightarrow \infty$ :

$$I(\theta) := E\{-\frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'}\} = E\{\sum_{t=1}^T \frac{\partial \mathcal{L}}{\partial \theta} x_t \frac{\partial \mathcal{L}}{\partial \theta'} x_t'\}$$



$$\begin{aligned}
 &= E\left\{ \sum_{t=1}^T \frac{(y_t - \Phi(x_t'\theta))^2}{[\Phi(x_t'\theta)(1-\Phi(x_t'\theta))]^2} \varphi^2(x_t'\theta) x_t x_t' \right\} \\
 &= \sum_{t=1}^T \frac{\varphi^2(x_t'\theta)}{\Phi(x_t'\theta)(1-\Phi(x_t'\theta))} x_t x_t' \quad (2.6)
 \end{aligned}$$

De laatste gelijkheid volgt uit:

$$\begin{aligned}
 E\{(y_t - \Phi(x_t'\theta))^2\} &= (1 - \Phi(x_t'\theta))^2 \Phi(x_t'\theta) + \Phi(x_t'\theta)^2 (1 - \Phi(x_t'\theta)) \\
 &= \Phi(x_t'\theta) (1 - \Phi(x_t'\theta))
 \end{aligned}$$

De matrix  $[I(\hat{\theta}_{ML})]^{-1}$  als consistente schatter voor de asymptotische covariantie van  $\theta_{ML}$ , kan ook eenvoudig worden gebruikt om toetsen en betrouwbaarheidsintervallen af te leiden.

Voor de consistentie van de ML-schatters in het binaire keuzemodel (in tegenstelling tot vele andere modellen, bijv. het lineaire regressiemodel) is het in het algemeen van belang dat de veronderstelling dat de storingen normaal zijn verdeeld, juist is. Het ligt dan ook voor de hand om te toetsen op normaliteit telkens als het probitmodel geschat wordt.

Een van de meest eenvoudig uitvoerbare toetsen op normaliteit is de Lagrange-Multiplier toets (LM-toets) op normaliteit. Wij zullen deze kort bespreken. Voor een uitvoerige beschrijving van deze toets verwijzen we naar Bera, Jarque en Lee (1984).

De LM-toets toetst de  $H_0$ -hypothese ' $\epsilon_t \sim N(0, \sigma^2)$ ' tegen het alternatief  $H_1$  ' $\epsilon_t \sim P$ ' met  $P$  een verzameling van verdelingen uit de Pearson-familie, behalve de normale verdeling. Een algemene dichtheid  $g(\epsilon_t)$  van  $\epsilon_t$  die afhankelijk is van de variabelen  $c_0, c_1$  en  $c_2$  representeert alle verdelingen uit de Pearson-familie. De likelihood behorend bij het binaire keuzemodel uitgaande van dichtheid  $g(\epsilon_t)$ , kan dan worden opgesteld. De variantie  $\sigma^2$  die door  $c_0$  wordt gerepresenteerd, wordt gelijk aan 1 genomen (normalisatie). De likelihood is dan afhankelijk van de onbekende parametervector

$\eta' := (\beta', c_1, c_2) \in \mathbb{R}^k$ . Onder de  $H_0$ -hypothese is  $c_1 = c_2 = 0$  en is de likelihood gelijk aan (2.2).

De toetsingsgrootheid in de LM-toets,  $\xi_{LM}$ , wordt gebaseerd op de score  $S(\eta)$  en de informatiematrix  $I(\eta)$ , beide geëvalueerd in de ML-schatter onder de  $H_0$  - hypothese ( $\hat{\eta}_{MLO}$ ) voor  $\eta$ . Ze wordt gedefinieerd als:

$$\xi_{LM} = S(\hat{\eta}_{MLO})' [I(\hat{\eta}_{MLO})]^{-1} S(\hat{\eta}_{MLO}) \quad (2.7)$$

Te bewijzen is dat

$$\xi_{LM} \stackrel{a}{\sim} \chi^2_{\ell} \quad (2.8)$$

met  $\ell$  het aantal gerestringeerde parameters onder  $H_0$ , dus  $\ell = 2$ .  
Voor de score geldt dat

$$S(\hat{\eta}_{MLO})' = \left. \frac{\partial \mathcal{L}}{\partial \eta'} \right|_{\hat{\eta}_{MLO}} = \left[ \left. \frac{\partial \mathcal{L}}{\partial \beta'} \right|_{\hat{\eta}_{MLO}}, \left. \frac{\partial \mathcal{L}}{\partial c_1} \right|_{\hat{\eta}_{MLO}}, \left. \frac{\partial \mathcal{L}}{\partial c_2} \right|_{\hat{\eta}_{MLO}} \right]$$

met  $\hat{\eta}_{MLO} = (\beta'_{ML}, 0, 0)$ . (2.9)

Omdat onder  $H_0$  de ML-schatter  $\hat{\beta}_{ML}$  voor  $\beta$  wordt bepaald door  $\left. \frac{\partial \mathcal{L}}{\partial \beta} \right|_{\hat{\eta}_{MLO}} = 0$  en (2.7) dus ook te schrijven als:

$$\left[ \left. \frac{\partial \mathcal{L}}{\partial c_1} \right|_{\hat{\eta}_{MLO}}, \left. \frac{\partial \mathcal{L}}{\partial c_2} \right|_{\hat{\eta}_{MLO}} \right] J' [I(\hat{\eta}_{MLO})]^{-1} J \left[ \left. \frac{\partial \mathcal{L}}{\partial c_1} \right|_{\hat{\eta}_{MLO}}, \left. \frac{\partial \mathcal{L}}{\partial c_2} \right|_{\hat{\eta}_{MLO}} \right]' \quad (2.10)$$

Hierin is  $J$  de matrix bestaande uit de laatste twee kolommen van de identiteitsmatrix met dimensie  $k + 2$ .

De toetsingsgrootheid  $\xi_{LM}$  is echter ook heel eenvoudig te berekenen als voor  $I(\hat{\eta}_{MLO})$  de schatting  $\sum_{t=1}^T \left. \frac{\partial \mathcal{L}}{\partial \eta} \right|_{\hat{\eta}_{MLO}}^t \left. \frac{\partial \mathcal{L}}{\partial \eta'} \right|_{\hat{\eta}_{MLO}}^t$  (equivalent aan (2.6) geëvalueerd in  $\hat{\beta}_{ML}$ ) wordt genomen. De toetsingsgrootheid (zie (2.7)) is dan gelijk aan

$$t' Q (Q' Q)^{-1} Q' t \quad (2.11)$$

met  $t = [1, 1 \dots 1]' \in \mathbb{R}^T$  en als  $t^e$  rij van matrix  $Q \in \mathbb{R}^{T \times (k+2)}$ :

$$\left[ \left. \frac{\partial \mathcal{L}}{\partial \beta'} \right|_{\hat{\eta}_{MLO}}^t, \left. \frac{\partial \mathcal{L}}{\partial c_1} \right|_{\hat{\eta}_{MLO}}^t, \left. \frac{\partial \mathcal{L}}{\partial c_2} \right|_{\hat{\eta}_{MLO}}^t \right].$$

De toets komt dan neer op een vermenigvuldiging van T met de multi-pele correlatiecoëfficiënt van een lineaire regressie van  $t$  op de kolommen van Q. Voor het probitmodel geldt:

$$\begin{aligned}\frac{\partial \mathcal{L}_t}{\partial \beta} \Big|_{\hat{\eta}_{\text{MLO}}} &= \frac{\varphi(x'_t \beta)}{\Phi(x'_t \beta)(1-\Phi(x'_t \beta))} [y_t - \Phi(x'_t \beta)] x_t \\ \frac{\partial \mathcal{L}_t}{\partial c_1} \Big|_{\hat{\eta}_{\text{MLO}}} &= \frac{\varphi(x'_t \beta)}{\Phi(x'_t \beta)(1-\Phi(x'_t \beta))} [y_t - \Phi(x'_t \beta)] [(x'_t \beta)^2 - 1] \\ \frac{\partial \mathcal{L}_t}{\partial c_2} \Big|_{\hat{\eta}_{\text{MLO}}} &= \frac{\varphi(x'_t \beta)}{\Phi(x'_t \beta)(1-\Phi(x'_t \beta))} [y_t - \Phi(x'_t \beta)] [3(x'_t \beta) + (x'_t \beta)^3] \quad (2.12)\end{aligned}$$

Deze toets is equivalent aan de toets op normaliteit voor het probitmodel volgens Chesher en Irish (1986).

/De  $H_0$  - hypothese wordt met een onbetrouwbaarheid van 5% geaccepteerd als

$$\xi_{\text{LM}} = T * R^2_{t|Q} < \chi^2_{2(0.05)} = 5.99 \quad (2.13)$$

met  $R^2_{t|Q}$  de  $R^2$  van een lineaire regressie van  $t$  op de kolommen van Q.

Als de  $H_0$ -hypothese niet geldt, hebben we in het algemeen geen garantie dat ML - schatters consistent zijn. In bepaalde gevallen is die garantie er wel; zo heeft bijv. Ruud (1983) aangetoond dat ML-schatters consistent kunnen zijn ondanks een verkeerd gespecificeerde verdeling van de storingsterm. Wij komen hier in paragraaf 5 op terug.

Wij zijn hier uitgegaan van storingen die een normale verdeling volgen. Op dezelfde manier kunnen natuurlijk voor willekeurige verdelingen likelihoodfuncties en toetsen worden afgeleid. Als echter met toetsen blijkt dat de verdelingsveronderstelling wordt verworpen, rijst natuurlijk de vraag op welke manier het model dan kan worden geschat. Bovendien is het mogelijk dat bij het uitvoeren van de LM-toets op normaliteit de  $H_0$ -hypothese wordt geaccepteerd, terwijl de ware verdeling van de storingen toch geen normale verdeling is; de  $H_0$ -hypothese en de alternatieve hypothese bevatten immers alleen verdelingen van de Pearson-familie.

Semi-parametrische schattingsmethoden stellen veel zwakkere eisen ten aanzien van de storingsterm en leiden vaak wel tot consistente schattingen. We zullen hier in de volgende paragraaf nader op ingaan.



### §3 Maximum Score van Manski.

Het binaire keuzemodel kan worden geschat met de semi-parametrische schattingsmethode Maximum Score van Manski (1985). In deze paragraaf zullen we deze methode bespreken. Eerst lichten we echter aan de hand van het multipele lineaire regressiemodel toe wat onder het begrip semi-parametrisch verstaan wordt.

Veel econometrische modellen bevatten een eindig aantal onbekende parameters. We nemen als voorbeeld het multipele lineaire regressiemodel met de sterke condities (zie bijv. van der Genugten (1988), blz. 120 e.v.):

$$\left. \begin{array}{l} y_t = x_t' \beta + \varepsilon_t \quad \beta \in B \subset \mathbb{R}^k \\ x_t \sim F_{x_t} \\ \varepsilon_t \sim N(0, \sigma^2) \\ (x_t', \varepsilon_t)' \text{ onafhankelijk van } (x_s', \varepsilon_s)' \quad s \neq t \\ x_t \text{ onafhankelijk van } \varepsilon_t \end{array} \right\} \quad t, s = 1, 2, \dots, T \quad (3.1)$$

In dit model zijn eindig veel (nl.  $k + 1$ ) onbekende parameters. Modellen met een eindig aantal onbekende parameters worden parametrisch genoemd. De verdeling van  $\varepsilon_t$  is bekend op één parameter ( $\sigma^2$ ) na. Als echter omtrent  $\varepsilon_t$  alleen bekend zou zijn dat deze een verdeling  $F_{\varepsilon_t}$  heeft, zonder verder iets te weten over de functionele vorm van  $F_{\varepsilon_t}$ , dan bevat het model niet langer een eindig aantal parameters. Het heeft dan een 'parametrisch', dat wil zeggen een eindig dimensionaal gedeelte ( $\beta$ ) en een 'niet-parametrisch', dat wil zeggen een niet eindig dimensionaal gedeelte (de verdeling van  $\varepsilon_t$ ). Dergelijke modellen worden semi-parametrisch genoemd. Een model wordt niet-parametrisch genoemd als het niet parametrisch is. Een semi-parametrisch model is dus een bijzonder geval van een niet-parametrisch model.

In parametrische modellen kunnen meestal op eenvoudige wijze consistente en (asymptotisch) efficiënte schatters (bijv. ML-schatters) voor de onbekende parameters worden gevonden. Als  $T \rightarrow \infty$  is bovendien (als de storingen onafhankelijk en identiek verdeeld zijn) de centrale limietstelling van toepassing en kunnen toetsen uitgevoerd en betrouwbaarheidsintervallen voor de onbekende parameters worden opgesteld. Omdat hierbij nagenoeg alles is

gebaseerd op de verdelingsveronderstelling van de storingsterm, is het van belang dat deze veronderstelling juist is. Als in het parametrische lineaire regressiemodel (3.1) storingsterm  $\varepsilon_t$  in werkelijkheid geen normale verdeling heeft, zal de (OLS-)schatter voor  $\beta$  in het algemeen aan efficiëntie verliezen maar vaak wel consistent zijn. Als  $E\{x_t x_t'\}$  en  $E\{\varepsilon_t^2\}$  bestaan en bovendien  $E\{x_t x_t'\}$  niet singulier is, vereist consistentie van de OLS-schatter voor  $\beta$  in het multipele lineaire regressiemodel namelijk alleen dat  $E\{x_t \varepsilon_t\} = 0$ . In het geval dat  $x_t$  en  $\varepsilon_t$  onafhankelijk zijn, is  $E\{x_t \varepsilon_t\} = E\{x_t\}E\{\varepsilon_t\}$  en is de OLS-schatter voor  $\beta$  consistent mits slechts geldt dat  $E\{\varepsilon_t\} = 0$ .

In het binaire keuzemodel (zie 2.1) heeft het uitgaan van een verkeerde verdelingsveronderstelling veel ernstigere gevolgen, alleen al met betrekking tot de consistentie van de schatters. Vandaar dat we een semi-parametrische schattingsmethode voor dit model nader zullen bespreken. De keuze is hierbij, mede bepaald door de aanwezige programmatuur, gevallen op de Maximum Score methode (MSCORE) van Manski (1985). Deze stelt aan de storingsterm slechts een vrij zwakke eis en levert consistente schatters. Voor een algemeen overzicht van semi-parametrische schattingsmethoden voor zowel lineaire regressiemodellen als keuzemodellen verwijzen we naar Robinson (1988).

Manski gaat uit van het volgende model:

$$\left. \begin{aligned} y_t^* &= x_t' \beta + \varepsilon_t \quad \beta \in B \subset \mathbb{R}^k \\ y_t &= 1 \text{ als } y_t^* \geq 0 \\ y_t &= -1 \text{ als } y_t^* < 0 \\ x_t &\sim F_{x_t} \\ \text{Med}(\varepsilon_t | x_t) &= 0 \\ (x_t', \varepsilon_t)' &\text{ onafhankelijk van } (x_s', \varepsilon_s)' \quad s \neq t \\ x_t &\text{ onafhankelijk van } \varepsilon_t \end{aligned} \right\} \quad s, t = 1, 2 \dots T \quad (3.2)$$

Een verschil in notatie tussen dit model en het binaire keuzemodel (2.1) is dat Manski i.p.v. de waarde 0 de waarde -1 toekent aan  $y_t$  als  $y_t^* < 0$ . Bovendien veronderstelt hij  $y_t = 1$  als  $y_t^* = 0$ . Het wezenlijke verschil zit echter in de veronderstelling met betrekking tot de storingsterm. De notatie



$\text{Med}(\epsilon_t | x_t) = 0$  heeft de betekenis dat de mediaan van  $\epsilon_t$  gegeven  $x_t$  ( $\epsilon_t | x_t$ ) de waarde 0 heeft en uniek is. Dus als  $\epsilon_t | x_t$  de (onbekende) verdeling  $F_{\epsilon_t}$  heeft, moet gelden dat  $F_{\epsilon_t}(0) = 1/2$ . Omdat deze mediaan uniek is moet bovendien gelden:  $F_{\epsilon_t}^{-1}(\{1/2\}) = \{0\}$ . Deze veronderstelling laat toe dat  $\epsilon_t$  heteroskedastisch is. Ze is dus veel zwakker dan bij het probitmodel. Voor de latente variabelen in het model volgt dat:

$$\text{Med}(y_t^* | x_t) = x_t' \beta \quad (3.3)$$

Op basis van de relatie tussen de waargenomen variabelen  $y_t$  en  $x_t$  moet  $\beta$  geschat kunnen worden. Manski maakt daarom gebruik van het feit dat:

$$\begin{aligned} E\{y_t | x_t\} &= 1 \cdot P\{y_t^* \geq 0 | x_t\} + (-1) \cdot P\{y_t^* < 0 | x_t\} \\ &= 2 \cdot P\{y_t^* \geq 0 | x_t\} - 1 \end{aligned} \quad (3.4)$$

Gegeven  $x_t' \beta > 0$  volgt dat  $E\{y_t | x_t\} > 0$ , immers:

$$x_t' \beta > 0 \Rightarrow \text{Med}\{y_t^* | x_t\} > 0 \Rightarrow P\{y_t^* \geq 0 | x_t\} > 1/2 \Rightarrow E\{y_t | x_t\} > 0 \quad (3.5)$$

De omgekeerde implicaties gelden ook. Analoge equivalenties gelden als  $x_t' \beta = 0$  en  $x_t' \beta < 0$ . Dus:

$$\begin{aligned} x_t' \beta > 0 &\Leftrightarrow E\{y_t | x_t\} > 0 \\ x_t' \beta = 0 &\Leftrightarrow E\{y_t | x_t\} = 0 \\ x_t' \beta < 0 &\Leftrightarrow E\{y_t | x_t\} < 0 \end{aligned} \quad (3.6)$$

Manski zwakt dit zelfs iets af en gaat uit van de volgende equivalenties:

$$\begin{aligned} x_t' \beta \geq 0 &\Leftrightarrow E\{y_t | x_t\} \geq 0 \\ x_t' \beta < 0 &\Leftrightarrow E\{y_t | x_t\} < 0 \end{aligned} \quad (3.7)$$

Hij noteert deze als

$$\text{sgn}(x_t' \beta) = \text{sgn}(E\{y_t | x_t\}) \quad (3.8)$$

waarbij  $\text{sgn}(z) = 1$  als  $z \geq 0$  en  $\text{sgn}(z) = -1$  als  $z < 0$ .

Indien de veronderstelling  $\text{Med}(\epsilon_t | x_t) = 0$  vervangen zou worden door  $E\{\epsilon_t | x_t\} = 0$  is een dergelijk verband tussen  $y_t$  en  $x_t$  niet te leggen. In plaats van de mediaan kan ook een ander kwantiel van de verdeling van  $\epsilon_t | x_t$  worden genomen. Manski leidt ook voor deze gevallen relaties tussen  $x_t$  en  $y_t$  af. Wij beperken ons tot de mediaan van  $\epsilon_t | x_t$ .

Vergelijking (3.8) kan slechts voldoende informatie over de ware  $\beta$  bevatten als deze  $\beta$  in ieder geval op grond van (3.8) te onderscheiden is van elke willekeurige  $b \in \mathbb{R}^k$  ( $b \neq \beta$ ). Daarvoor moeten er eisen aan de verklarende variabelen  $x_{t1}, x_{t2} \dots x_{tk}$  gesteld worden. Manski definieert:

$$X_b := \{ x_t \in \mathbb{R}^k ; \text{sgn}(x_t' \beta) \neq \text{sgn}(x_t' b) \}$$

en

$$R(b) := P \{ x_t \in X_b \} \quad (3.9)$$

Als voor een  $b$  ( $b \neq \beta$ ) geldt dat  $R(b) = 0$ , is  $\text{sgn}(x_t' b) = \text{sgn}(x_t' \beta)$  voor bijna alle  $x_t \in \mathbb{R}^k$ . Het is dan niet mogelijk om op basis van (3.8),  $\beta$  te onderscheiden van  $b$ . Manski toont dan ook aan dat voor de identificatie van  $\beta$  ten opzichte van  $b$  het in ieder geval noodzakelijk is dat  $R(b) > 0$ . Bovendien is het van belang dat minstens één van de verklarende variabelen (gegeven de overige verklarende variabelen) continu is. Zo kan bijv. voor een model met alleen twee dummy-variabelen worden aangetoond dat de scorefunctie op een vrij groot interval hetzelfde maximum bereikt, zodat de ware parameter niet te onderscheiden is van andere punten op dit interval. Eenvoudig is in te zien dat  $\beta$  slechts geïdentificeerd is op een schaalfactor na. Immers met  $\beta$  voldoet ook  $a\beta$  met  $a > 0$  aan (3.8). Voor een beschrijving van de voldoende condities van de identificatie van  $\beta$  verwijzen we naar Manski (1985).

Om een consistente schatter voor  $a\beta$  te vinden, bekijken we als criteriumfunctie de populatie-scorefunctie gedefinieerd door:

$$S(b) := E \{ y_t \text{sgn}(x_t' b) \}$$

$$\begin{aligned}
 &= P \{ y_t^* \geq 0, x_t' b \geq 0 \} + P \{ y_t^* < 0, x_t' b < 0 \} \\
 &- P \{ y_t \geq 0, x_t' b < 0 \} - P \{ y_t < 0, x_t' b \geq 0 \}
 \end{aligned} \tag{3.10}$$

Maximalisatie van deze scorefunctie levert de ware parameter  $\beta$  op een schaalfactor  $a$  na. Vanwege de schaalfactor  $a$  maakt Manski de keuze om  $S(b)$  te maximaliseren op de verzameling van alle  $b \in \mathbb{R}^k$  waarvoor geldt dat  $\|b\| = 1$ :

$$\text{Max}_{b \in B} S(b) \quad \text{met} \quad B := \{ b \in \mathbb{R}^k : \|b\| = 1 \} \tag{3.11}$$

Er wordt dan een schatter gevonden voor  $\beta^* := \frac{\beta}{\|\beta\|}$ . Als de populatie-scorefunctie wordt vervangen door de empirische scorefunctie, dat wil zeggen  $P$  in (3.10) vervangen door de empirische kans met betrekking tot de aselechte steekproef  $(y_t, x_t')$  ter grootte  $T$ , dan kan de steekproef-scorefunctie voor willekeurige  $b \in B$  berekend worden:

$$\begin{aligned}
 S_T(b) &:= \frac{1}{T} \sum_{t=1}^T y_t \text{sgn}(x_t' b) \\
 &= P_T \{ y_t \geq 0, x_t' b \geq 0 \} + P_T \{ y_t^* < 0, x_t' b < 0 \} \\
 &- P_T \{ y_t \geq 0, x_t' b < 0 \} - P_T \{ y_t < 0, x_t' b \geq 0 \}
 \end{aligned} \tag{3.12}$$

Manski toont aan dat maximalisatie van  $S_T(b)$  op  $B$  een schatter levert (we noemen deze de MSCORE- of MS-schatter) die met kans 1 convergeert naar de ware  $\beta^*$ , mits aan bepaalde voorwaarden is voldaan.

Om de scorefunctie (3.12) te maximaliseren is er een numeriek optimaliseringsalgoritme ontwikkeld, het zogenaamde MS-algoritme of MSCORE. Dit wijkt op een aantal punten af van andere optimaliseringsmethoden vanwege de bijzondere vorm van de criteriumfunctie. We zullen dit toelichten door eerst iets te vertellen over veel gebruikte numerieke optimaliseringsprocedures zoals bijv. het Newton-Raphson-algoritme (zie 2.4).

Uitgaande van een schatting  $\vartheta_0$  voor de onbekende parameter  $\vartheta$  wordt in deze algoritmen eerst de richting gekozen waarin een betere parameterschatting te vinden is. Daarna wordt de stapgrootte in deze richting bepaald. Omdat deze optimaliseringsmethoden uitgaan van een continue en twee maal differentieerbare criteriumfunctie ( $\mathcal{L}$ ), wordt de richting met behulp van



de eerste afgeleiden (gradiënt) gevonden; de gradiënt wordt voorvermenigvuldigd met een positief definitie matrix. Het Newton-Raphson-algoritme neemt als positief definitie matrix  $[-\frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'}]^{-1}$ . In deze richting wordt gezocht naar de stapgrootte die de criteriumfunctie maximaliseert.

Omdat de scorefunctie niet continu is, is het in een bepaald punt veel moeilijker om te bepalen in welke richting de functie toeneemt. Gegeven een bepaalde richting is het echter vrij eenvoudig om het maximum van de scorefunctie in die richting te bepalen. Manski toont aan dat er slechts enkele, snel uitvoerbare berekeningen nodig zijn om de scorefunctie (gegeven de richting) te optimaliseren.

Het MS-algoritme begint daarom niet te zoeken naar een richting waarin de scorefunctie geoptimaliseerd kan worden, maar selecteert een aantal richtingen. Langs al deze richtingen wordt gezocht naar het maximum van de scorefunctie.

Zo kiest het algoritme op basis van een parameterschatting  $\beta_0 \in \mathbb{R}^k$   $k - 1$  vectoren  $c^1, c^2, \dots, c^{k-1}$  die tesamen met  $\beta_0$  een orthonormale basis vormen voor  $\mathbb{R}^k$ . Langs de grote eenheidscirkel (de "great unit circle") die  $\beta_0$  met  $c^1$  verbindt, wordt dan de scorefunctie gemaximaliseerd:

$$\text{Max}_{b \in A} S_T(b)$$

$$A := \{b \in \mathbb{R}^k; \|b\| = 1\} \cap \{b \in \mathbb{R}^k; b = \lambda \beta_0 + \mu c^1 \text{ met } \lambda, \mu \in \mathbb{R}\}$$

Het punt of één van de punten (indien er meerdere maxima zijn) dat de scorefunctie maximaliseert, wordt als nieuwe schatting voor  $\beta$  genomen mits deze schatting beter is dan  $\beta_0$ . Het wordt via de grote eenheidscirkel met  $c^2$  verbonden en langs deze cirkel wordt weer gezocht naar een betere schatting voor  $\beta$ . Dit gaat zo door tot alle  $k - 1$  richtingen onderzocht zijn. In een volgende iteratie wordt een nieuwe basis (inclusief de beste parameterschatting tot dan toe) voor  $\mathbb{R}^k$  gekozen. Als er in een iteratie geen betere schatting voor  $\beta$  is gevonden, stopt de procedure.

Indien er maar twee te schatten parameters zijn ( $k=2$ ), hoeven slechts de punten langs de ene eenheidscirkel onderzocht te worden. Het globale maximum van de scorefunctie wordt dan altijd in één iteratie gevonden. Als er echter

meer te schatten parameters zijn ( $k > 2$ ), kan het voorkomen dat het algoritme een lokaal maximum vindt. Immers lang niet alle eenheidscirkels worden onderzocht. Het MS-algoritme biedt de mogelijkheid om extra iteraties (zogenaamde "endgame iterations" of "endgames") toe te voegen, die telkens aselect een andere orthonormale basis voor  $\mathbb{R}^k$  nemen. De kans dat het globale maximum van de scorefunctie wordt gevonden, kan op deze manier worden vergroot en convergeert naar 1 als het aantal extra iteraties naar oneindig gaat.

Tot op heden zijn er over MS-schatters, naast convergentie met kans 1, weinig 'bruikbare' eigenschappen bekend. Met betrekking tot de asymptotische verdeling van de MS-schatter  $\beta_{MS}$  voor  $\beta$  is bekend dat  $T^{1/3}(\beta_{MS} - \beta)$  naar een niet-normale verdeling convergeert (Kim and Pollard, 1989). Deze niet-normale verdeling is weliswaar bekend, maar dermate gecompliceerd dat ze niet goed bruikbaar is.

Als de exacte verdeling niet bekend is, worden in parametrische modellen standaardfouten van de schatters consistent geschat met behulp van de asymptotische covariantiematrix: voor de schatter  $\hat{\theta}$  voor  $\theta$  geldt immers vaak dat  $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$  (met  $V$  de asymptotische covariantiematrix). Toetsen en betrouwbaarheidsintervallen worden dan ook gebaseerd op de asymptotische verdeling. Omdat deze verdeling van de semi-parametrische MS-schatter hiervoor niet geschikt is, zullen er andere methoden moeten worden gebruikt. Een mogelijkheid om standaardafwijkingen van schatters met 'moeilijke' verdelingen te bepalen, is de "bootstrap"-methode. Met deze methode kunnen bovendien betrouwbaarheidsintervallen voor de onbekende parameters en toetsen afgeleid worden. We zullen hier meer aandacht aan besteden.

#### § 4 "Bootstrap"-methoden.

Als voor een willekeurige onbekende parameter een schatter gebruikt wordt, wil men vanzelfsprekend meer weten over bepaalde eigenschappen van deze schatter. De "bootstrap" - methode of "bootstrapping" (Efron (1986)) maakt het mogelijk om zowel met als zonder parametrische veronderstellingen of benaderingen, uitspraken te doen over de nauwkeurigheid van een schatter. De methode is dus ook geschikt voor de schatter die behandeld is in paragraaf 3, de MS-schatter, die een zgn. 'moeilijke' verdeling heeft. Bij de bruikbaarheid van de bootstrap-methoden moet echter een kleine kanttekening worden geplaatst; vanwege het feit dat de bootstrap-methoden vrij nieuw zijn, is nog niet precies bekend welke waarde aan de methoden kan worden gehecht.

In deze paragraaf zullen we eerst bespreken hoe bootstrap-schattingen van standaardfouten voor een schatter verkregen kunnen worden. Daarna laten we een toepassing van de bootstrap-methode in het binaire keuzemodel zien en lichten we een schematische weergave van de bootstrap-methode toe aan de hand van deze toepassing. Tenslotte worden enkele mogelijkheden aangegeven om betrouwbaarheidsintervallen met behulp van bootstrapping te benaderen. In de volgende paragrafen maken we gebruik van deze bootstrap-methoden om iets meer te kunnen zeggen over de precisie van de MS-schatters.

Vanwege het feit dat bootstrap-methoden algemeen toepasbare methoden zijn en dus niet speciaal betrekking hebben op MS-schatters, kan deze paragraaf min of meer onafhankelijk worden beschouwd van de overige paragrafen.

We gaan uit van de stochastische variabele  $x$  met een onbekende verdeling  $F$ :  $x \sim F$ . Op basis van een aselechte steekproef ter grootte  $T$  uit  $F$ , wordt een schatter voor een onbekende parameter  $\vartheta \in \mathbb{R}^k$  bepaald. We noteren deze schatter als  $\hat{\vartheta}(y)$  met  $y := (x_1, x_2, \dots, x_T)'$  en  $x_t$  de  $t^e$  trekking uit  $F$  ( $t = 1, 2, \dots, T$ ). We zijn nu geïnteresseerd in de standaardafwijking van  $\hat{\vartheta}(y)$ . Deze is afhankelijk van de ware verdeling  $F$ , de steekproefomvang  $T$  en de schatter  $\hat{\vartheta}(y)$  voor  $\vartheta$ . Omdat  $T$  en  $\hat{\vartheta}(y)$  bekend zijn, noteren we  $\sigma(F)$  als de standaardafwijking van  $\hat{\vartheta}(y)$

$$\sigma(F) = (V\{\hat{\vartheta}(y)\})^{1/2}$$



$$= [\int \int \dots \int (\vartheta(x_1, x_2 \dots x_T) - E\{\vartheta(x_1, x_2 \dots x_T)\})^2 dF(x_1) dF(x_2) \dots dF(x_T)]^{1/2} \quad (4.1)$$

met  $E\{\vartheta(x_1, x_2 \dots x_T)\} = \int \int \dots \int \vartheta(x_1, x_2 \dots x_T) dF(x_1) dF(x_2) \dots dF(x_T)$ .

Omdat de ware verdeling  $F$  onbekend is, is deze standaardafwijking niet te berekenen. Daarom wordt in plaats van  $F$  een empirische verdeling gebaseerd op de steekproef  $x_1, x_2 \dots x_T$  genomen. Deze empirische verdeling wordt gedefinieerd als:

$$\hat{F} : \text{puntmassa } 1/T \text{ in } x_t \quad t = 1, 2 \dots T \quad (4.2)$$

Als schatter van  $\sigma(F)$  wordt dan genomen

$$\hat{\sigma} := \sigma(\hat{F})$$

$$= [\int \int \dots \int (\vartheta(x_1, x_2 \dots x_T) - \hat{E}\{\vartheta(x_1, x_2 \dots x_T)\})^2 d\hat{F}(x_1) d\hat{F}(x_2) \dots d\hat{F}(x_T)]^{1/2} \quad (4.3)$$

met  $\hat{E}\{\vartheta(x_1, x_2 \dots x_T)\} = \int \int \dots \int \vartheta(x_1, x_2 \dots x_T) d\hat{F}(x_1) d\hat{F}(x_2) \dots d\hat{F}(x_T)$ .

Vanwege het feit dat de empirische verdeling discreet is, staan in (4.3) de integraalteken voor sommatietekens.

De schatter  $\hat{\sigma}$  is weliswaar te berekenen, maar zal vanwege de sommatie over  $T$  indices in het algemeen veel rekenwerk kosten. Daarom wordt  $\sigma(\hat{F})$  met behulp van Monte-Carlo technieken benaderd. We geven hiervan een beschrijving.

Uit de empirische verdeling  $\hat{F}$  worden  $T$  onafhankelijke trekkingen met teruglegging gedaan:  $x_1^*, x_2^* \dots x_T^*$ . Noteren we  $y^* := (x_1^*, x_2^* \dots x_T^*)$  dan wordt  $y^*$  een bootstrap-steekproef genoemd. Op basis van zo'n steekproef kan een schatter  $\hat{\vartheta}(y^*)$  voor  $\vartheta$  op dezelfde manier als  $\hat{\vartheta}(y)$  worden berekend. Wordt dit proces een aantal malen herhaald, dat wil zeggen we trekken  $B$  steekproeven  $y^*(1), y^*(2) \dots y^*(B)$  uit  $\hat{F}$  en bepalen  $\hat{\vartheta}(y^*(1)), \hat{\vartheta}(y^*(2)) \dots \hat{\vartheta}(y^*(B))$ , dan kan de gecorrigeerde steekproef-standaardafwijking worden berekend:

$$\hat{\sigma}_B := \left( \frac{\sum_{b=1}^B (\hat{\vartheta}(y^*(b)) - \bar{\hat{\vartheta}})^2}{B - 1} \right)^{1/2} \text{ met}$$

$$\hat{\sigma}_B := \frac{\sum_{b=1}^B \hat{\vartheta}(y^*(b))}{B} \quad (4.4)$$

Als het aantal bootstrap-steekproeven toeneemt, zal  $\hat{\sigma}_B$  convergeren naar de standaardafwijking van  $\hat{\vartheta}(y)$  met  $F$  vervangen door de empirische verdeling  $\hat{F}$ :

$$\hat{\sigma}_B \rightarrow \hat{\sigma} := \sigma(\hat{F}) \quad B \rightarrow \infty$$

Voor deze convergentie is het van belang dat de bootstrap-steekproefomvang net zo groot is als de steekproef uit  $F$ . Volgens Efron (1986) wordt een goede benadering van  $\sigma(F)$  door  $\hat{\sigma}_B$  verkregen door  $B$  ongeveer tussen 50 en 200 te nemen.

Andere bootstrap-methoden maken, net als de bootstrap-methode voor standaardfouten, gebruik van een empirische verdeling  $\hat{F}$  als vervanging voor de ware verdeling  $F$ . Bovendien is elke bootstrap-methode vaak te beschrijven als een Monte-Carlo-algoritme (MC-algoritme) dat uit drie stappen bestaat. Voor de beschreven bootstrap-methode voor standaardfouten ziet het algoritme er als volgt uit:

- (1) Trek eerst  $B$  bootstrap-steekproeven  $y^*(b)$  ( $b = 1, 2 \dots B$ ) van  $T$  aselecte trekkingen met teruglegging uit  $\hat{F}$ ;
- (2) Bereken dan voor elke bootstrap-steekproef de schatting  $\hat{\vartheta}(y^*(b))$  ( $b = 1, 2 \dots B$ );
- (3) Bereken de gecorrigeerde steekproef-standaardafwijking van de  $B$  schattingen uit stap (2) (zie 4.4).

Alleen stap (3) van dit algoritme dient bij de andere bootstrap-methoden vervangen te worden.

In bepaalde gevallen is het heel eenvoudig om de bootstrap-standaardafwijking (4.4) te berekenen. We lichten dit toe aan de hand van een voorbeeld.

Veronderstel dat  $\vartheta$  de verwachting is van  $x$ , met  $x \sim F$ :

$$\vartheta = E_F\{x\} \quad (4.5)$$

en noteer de variantie van  $x$  als:

$$\mu_2(F) = E_F\{(x - E_F\{x\})^2\} \quad (4.6)$$

Hier betekent  $E_F\{.\}$  dat de verwachting met betrekking tot de verdeling  $F$  genomen wordt. Op basis van de steekproef  $x_1, x_2, \dots, x_T$  ligt het voor de hand om  $\vartheta$  te schatten door het steekproefgemiddelde:

$$\hat{\vartheta} = \bar{x} := \frac{1}{T} \sum_{t=1}^T x_t \quad (4.7)$$

De standaardafwijking van  $\hat{\vartheta}$  is dan:

$$\sigma(F) = \left( \frac{\mu_2(F)}{T} \right)^{1/2} = \left[ \frac{1}{T} \int (x - E_F\{x\})^2 dF(x) \right]^{1/2} \quad (4.8)$$

De bootstrap-methode bestaat wederom uit het vervangen van  $F$  door  $\hat{F}$ . De bootstrap-schatting voor  $\sigma(F)$  is dus:

$$\begin{aligned} \sigma(\hat{F}) &= \left( \frac{\mu_2(\hat{F})}{T} \right)^{1/2} = \left[ \frac{1}{T} \int (x - E_{\hat{F}}\{x\})^2 d\hat{F}(x) \right]^{1/2} \\ &= \left[ \frac{1}{T^2} \sum (x_t - \bar{x})^2 \right]^{1/2} \end{aligned} \quad (4.9)$$

Deze bootstrap-schatting is eenvoudig te berekenen. Vandaar dat het hier niet nodig is om  $\sigma_B(\hat{F})$  te benaderen met behulp van  $B$  bootstrap-steekproeven uit  $\hat{F}$ .

De bootstrap-schatting  $\sigma(\hat{F})$  benadert vrij goed de schatting

$$\bar{\sigma} := \left( \frac{\bar{\mu}_2}{T} \right)^{1/2} \quad \text{met} \quad \bar{\mu}_2 := \frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^2 \quad (4.10)$$

Dit is een voor de hand liggende schatter voor de standaardafwijking van het steekproefgemiddelde  $\bar{x}$ , immers  $\bar{\mu}_2$  is een zuivere schatter voor  $\mu_2(F)$ .

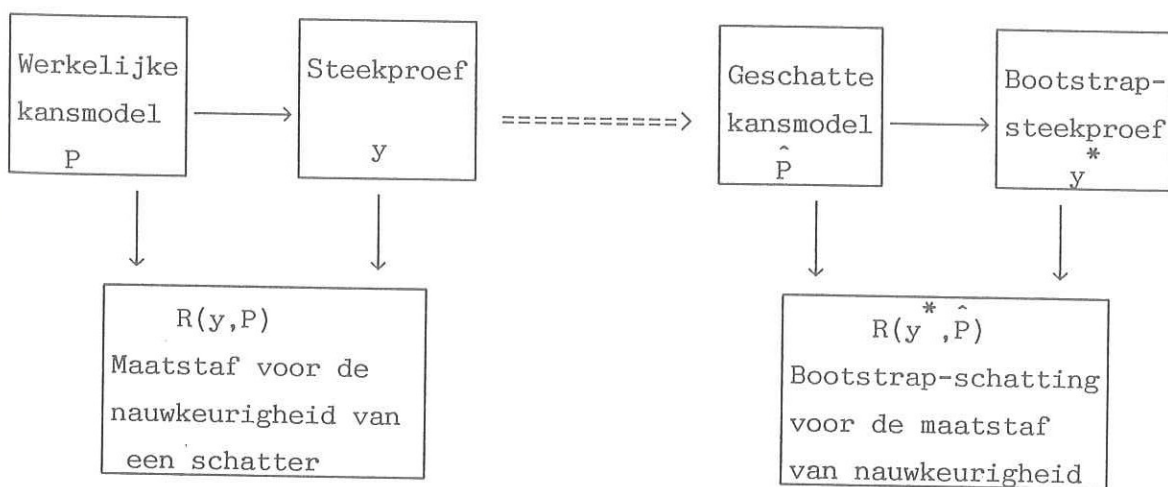
Essentieel bij de bootstrap-methode is de empirische verdeling. Op deze verdeling zijn immers de bootstrap-schatters gebaseerd. Het definiëren van de empirische verdeling kan op twee manieren gebeuren, namelijk uitgaande van parametrische en niet-parametrische veronderstellingen. Het MC-algoritme zoals wij dat hebben besproken, was op niet-parametrische veronderstellingen gebaseerd; het maakt geen gebruik van kennis omtrent de ware verdeling van de stochastische variabele.

De empirische verdeling kan echter wel gebruik maken van parametrische veronderstellingen. Als bijv. verondersteld wordt dat de ware verdeling van een schatter normaal is, maar men kent de parameters niet, kan als empirische verdeling ook een normale verdeling genomen worden met schattingen voor de variantie en verwachting. Bootstrap-steekproeven worden dan uit deze normale verdeling getrokken. Het MC-algoritme blijft verder hetzelfde. Een voordeel van het gebruik maken van parametrische veronderstellingen zou kunnen zijn dat efficiëntere bootstrap-schattingen worden verkregen.

Tot dusver hebben wij ons beperkt tot stochastische variabelen die dezelfde verdeling  $F$  hebben. De bootstrap-methode kan ook eenvoudig toegepast worden met stochastische variabelen die verschillende verdelingen volgen ( $x_t \sim F_t$  in plaats van  $x_t \sim F$ ).

Efron geeft de gehele bootstrap-methode op basis van een steekproef  $y := (y_1, y_2 \dots y_T)$  schematisch als volgt weer:

Figuur 4.1





De dubbele pijl geeft de essentiële stap in de methode aan, namelijk de stap van de steekproef naar de empirische verdeling. Wij zullen dit schema toelichten aan de hand van het binaire keuzemodel:

$$\left. \begin{array}{l} y_t^* = x_t' \beta + \epsilon_t \quad x_t, \beta \in \mathbb{R}^k \\ y_t = 1 \quad \text{als } y_t^* > 0 \\ y_t = 0 \quad \text{als } y_t^* \leq 0 \\ x_t \sim F_{x_t} \\ \epsilon_t \sim F_{\epsilon_t} \\ (x_t', \epsilon_t)' \text{ onafhankelijk van } (x_s', \epsilon_s)' \quad s \neq t \\ x_t \text{ onafhankelijk van } \epsilon_t \end{array} \right\} \quad t, s = 1, 2 \dots T \quad (4.11)$$

Op basis van  $T$  trekkingen  $x_1, x_2, \dots, x_T$  en  $y_1, y_2, \dots, y_T$  kan  $\beta$  worden geschat met het MS-algoritme (zie § 3). Dit levert een MS-schatting  $\hat{\beta}_{MS}$  voor  $\beta$ . Een empirische verdeling  $\hat{F}$  kan dan worden gedefinieerd door

$$\hat{F} : \text{puntmassa } 1/T \text{ in } z_t := (y_t, x_t) \quad t = 1, 2 \dots T \quad (4.12)$$

met  $(y_t, x_t)$  de  $t^e$  steekproefrealisatie. We definiëren  $B$  bootstrap-steekproeven als:

$$\begin{aligned} z^*(b) &:= (z_1^*(b), z_2^*(b) \dots z_T^*(b))' \text{ met} \\ z_t^*(b) &:= (y_t^*(b), x_t^*(b)) \quad b = 1, 2 \dots B \\ &\quad t = 1, 2 \dots T \end{aligned} \quad (4.13)$$

Voor elke bootstrap-steekproef kan een MS-schatting voor  $\beta$  worden berekend. Zo vindt men dus  $B$  bootstrap-schattingen voor  $\beta$ .

Als maatstaf voor nauwkeurigheid van de MS-schatter  $\hat{\beta}_{MS}$  nemen we de MSE-matrix (mean square error matrix):

$$\Sigma = E_F \{ (\hat{\beta}_{MS} - \beta)(\hat{\beta}_{MS} - \beta)' \} \quad (4.14)$$

Deze kan geschat worden met behulp van de  $B$  bootstrap-schattingen  $\hat{\beta}_{MS}^*(1), \hat{\beta}_{MS}^*(2) \dots \hat{\beta}_{MS}^*(B)$ :

$$\Sigma_B = \frac{1}{B} \sum_{b=1}^B (\beta_{MS}^*(b) - \hat{\beta}_{MS}) (\beta_{MS}^*(b) - \hat{\beta}_{MS})' \quad (4.15)$$

Merk op dat deze matrix niet is te vergelijken met de gebruikelijke steekproef-covariantiematrix; de afwijkingen van de bootstrap-schattingen worden ten opzichte van  $\hat{\beta}_{MS}$  genomen terwijl de gebruikelijke steekproef-covariantiematrix wordt berekend door afwijkingen ten opzichte van het gemiddelde van de B bootstrap-schattingen te nemen.

De empirische verdeling  $\hat{F}$  maakt geen gebruik van veronderstellingen ten aanzien van F, dus deze methode is niet-parametrisch.

Komen we terug op het schema (fig. 4.1), dan zien wij dat in het binaire keuzemodel, het werkelijke kansmodel P de parameter  $\beta$  en de verdeling F omvat:  $P = (\beta, F)$ .

Hieruit komt via  $\varepsilon := (\varepsilon_1, \varepsilon_2 \dots \varepsilon_T)'$  de steekproef  $y := (y_1, y_2 \dots y_T)'$  voort. Als maatstaf voor nauwkeurigheid is de MSE-matrix genomen. Dus:

$$R(y, P) = E \{ (\hat{\beta}_{MS} - \beta) (\hat{\beta}_{MS} - \beta)' \}$$

Het geschatte kansmodel  $\hat{P}$  omvat de MSCORE-schatting  $\hat{\beta}_{MS}$  en de empirische verdeling  $\hat{F}$ , dus  $\hat{P} = (\hat{\beta}_{MS}, \hat{F})$ . Tot slot kan op basis van  $\hat{F}$ ,  $\hat{\beta}_{MS}$  en  $z(1), z(2) \dots z(B)$  uit (4.13), de MSE-matrix (4.15) worden geschat.

Het berekenen van betrouwbaarheidsintervallen voor de onbekende parameter  $\vartheta$ , wordt gebaseerd op de verdelingsveronderstelling van een schatter  $\hat{\vartheta}$  voor  $\vartheta$ . Welke bootstrap-methode de beste benadering voor het ware betrouwbaarheidsinterval is, hangt uiteraard af van de ware verdeling van  $\hat{\vartheta}$ . Wij beperken ons hier tot  $\vartheta \in \mathbb{R}$  en de gevallen dat  $\hat{\vartheta}$  een verdeling volgt met een eindig aantal onbekende parameters. We zullen drie bootstrap-betrouwbaarheidsintervallen bespreken die allen weer gebruik maken van het MC-algoritme. De besproken bootstrap-methoden nemen in algemeenheid toe, dat wil zeggen dat de eerste methode minder algemeen is dan de tweede en de tweede minder algemeen dan de derde. Bij bootstrap-methoden voor betrouwbaarheidsintervallen geeft Efron als richtlijn aan dat minimaal 1000 bootstrap-steekproeven noodzakelijk zijn ( $B = 1000$ ).

Het eenvoudigste bootstrap-interval wordt het 'standaard-interval' genoemd:

$$\vartheta \in \hat{\vartheta} \pm \hat{\sigma} u_{\alpha} \quad (4.16)$$



Hierin is  $\hat{\sigma}$  de bootstrap-standaardafwijking van  $\hat{\vartheta}$  en  $u_{\alpha}$  het  $(1-\alpha)^e$  kwantiel van de standaardnormale verdeling, dat wil zeggen:

$$P\{x \leq u_{\alpha}\} = 1 - \alpha \quad \text{als} \quad x \sim N(0,1) \quad (4.17)$$

Intuïtief is duidelijk dat het standaard-interval geen goede benadering voor het  $(1-2\alpha)$ -betrouwbaarheidsinterval voor  $\vartheta$  is, als niet bij benadering geldt:

$$\hat{\vartheta} \sim N(\vartheta, \sigma^2) \quad (4.18)$$

De tweede bootstrapmethode voor betrouwbaarheidsintervallen schat de eindpunten van het betrouwbaarheidsinterval op basis van de parametrische verdelingsfunctie  $\hat{G}(s)$  van  $\vartheta^*$ , die wordt gedefinieerd door:

$$\hat{G}(s) = P\{\vartheta^* < s\} \quad (4.19)$$

Hierin is  $\vartheta^*$  een schatter voor  $\vartheta$  uit stap (2) van het MC-algoritme. Op basis van de bootstrap-verdeling van de B schattingen  $\vartheta^*(1), \vartheta^*(2) \dots \vartheta^*(B)$  kunnen de grootste  $s_1$  en de kleinste  $s_2$  bepaald worden waarvoor geldt:

$$\begin{aligned} \hat{G}(s_1) &\leq \alpha \\ \hat{G}(s_2) &\geq 1 - \alpha \end{aligned} \quad (4.20)$$

Het betrouwbaarheidsinterval met eindpunten  $s_1 = \hat{G}^{-1}(\alpha)$  en  $s_2 = \hat{G}^{-1}(1-\alpha)$  wordt het 'percentiel-interval' genoemd. Voor het geval dat voor de bootstrap-verdeling van  $\vartheta^*$  geldt

$$\vartheta^* \sim N(\hat{\vartheta}, \hat{\sigma}^2) \quad (4.21)$$

is het percentiel-interval bij benadering gelijk aan het standaard-interval. Een voordeel van de percentiële methode is dat deze invariant is ten aanzien van monotone transformaties; noemen we  $v_{\alpha}$  het  $(1-\alpha)^e$  kwantiel van de verdeling van een monotone transformatie  $\hat{\varphi} = \tilde{\varphi}(\hat{\vartheta})$  met  $\varphi = \tilde{\varphi}(\vartheta)$  en  $w_{\alpha}$  het  $(1-\alpha)^e$  kwantiel van de verdeling van  $\hat{\vartheta}$ , dan geldt:

$$v_{\alpha} = g(w_{\alpha}) \quad (4.22)$$

Efron laat zien dat het percentiel-interval voor  $\vartheta$  het correcte interval is als de monotone transformatie  $\hat{\varphi}$  normaal is verdeeld:

$$\hat{\varphi} \sim N(\varphi, \tau^2) \quad (4.23)$$

(Hierin is  $\tau^2$  een constante). De percentiel-methode kan dus algemener genoemd worden dan de eerste methode, de standaard-methode.

Efron bespreekt een derde bootstrap-methode, die gebaseerd wordt op een monotone transformatie  $\hat{\psi} = \tilde{\psi}(\hat{\vartheta})$  met  $\psi = \tilde{\psi}(\vartheta)$  die normaal is verdeeld en een correctie maakt voor een mogelijke afwijking van de ware parameter  $\psi$ :

$$\hat{\psi} \sim N(\psi - z_0 \tau, \tau^2) \quad (4.24)$$

Hierin is  $z_0 = \Phi^{-1}(\hat{G}(\hat{\vartheta}))$  en  $\Phi$  de standaardnormale verdelingsfunctie. Het betrouwbaarheidsinterval voor  $\vartheta$  wordt dan ook 'het voor de "bias" gecorrigeerde interval' (BC-interval) genoemd. Noemen we  $u_{\alpha}$  het  $(1-\alpha)^e$  kwantiel van de normale verdeling, dan heeft het BC-interval als eindpunten (Efron, 1986):

$$G^{-1}(\Phi(2 z_0 + u_{1-\alpha})) \quad \text{en} \quad G^{-1}(\Phi(2 z_0 + u_{\alpha})) \quad (4.25)$$

Als  $z_0 = 0$  is het BC-interval gelijk aan het percentiel-interval. Het is ook mogelijk om betrouwbaarheidsintervallen met behulp van bootstrapping te berekenen, als de monotone transformatie  $\hat{\xi} = \tilde{\xi}(\hat{\vartheta})$  met de  $\xi = \tilde{\xi}(\vartheta)$  normale verdeling  $N(\xi - z_0 \tau_{\xi}, \tau_{\xi}^2)$  met  $\tau_{\xi}^2 = 1 + a \xi$ , benadert. De parameter  $a$  wordt onder andere bepaald door een maatstaf voor de scheefheid van de verdeling. Betrouwbaarheidsintervallen behorende bij deze verdeling worden 'BC<sub>a</sub>-intervallen' genoemd. We verwijzen hiervoor, evenals voor niet-parametrische betrouwbaarheidsintervallen en betrouwbaarheidsintervallen voor meerdere parameters, naar Efron (1985).

Enkele van de bootstrap-methoden die hier zijn besproken, zullen we in paragraaf 6 toepassen op de MS-schatters. In deze paragraaf is slechts de

werkwijze van de methoden uitgelegd en in sommige gevallen is iets gezegd over de bruikbaarheid van de bootstrap-resultaten. Helaas is, zoals gezegd in de inleiding, de precieze bruikbaarheid van de resultaten (nog) niet bekend. Wij hebben laten zien dat bootstrap-methoden geen eenduidige, maar wel algemeen toepasbare methoden zijn; zonder gebruik te maken van veel kennis omtrent de schatters kan met behulp van bootstrap-methoden toch iets gezegd worden over de nauwkeurigheid van deze schatters.

In de volgende paragraaf komen we terug op het binaire keuzemodel. We vergelijken dan de schattingsmethoden ML (onder normaliteit) en MSCORE aan de hand van resultaten die zijn verkregen met Monte-Carlo-experimenten.



## § 5 Enkele Monte-Carlo experimenten.

Uitgaande van het binaire keuzemodel (zie (2.1) of (3.2)) blijft, na voorgaande paragrafen, de vraag open welke schattingsmethode het best gebruikt kan worden voor dit model: ML of MSCORE ?

Het antwoord hangt af van enige kennis of vermoedens omtrent de verdeling van de storingsterm. Indien er een vermoeden bestaat over de verdeling, ligt ML voor de hand. De in paragraaf 2 behandelde Probit-Maximum-Likelihood-methode (uitgaan van de normale verdeling met constante variantie) is hier een voorbeeld van. Schatten met MSCORE en vergelijken van de MSCORE- met de ML-resultaten en/of toepassen van de LM-toets kan dan altijd gebeuren om te controleren of het vermoeden juist is geweest.

Is de verdeling van de storingsterm onbekend en de steekproef op basis waarvan het model geschat moet worden groot, dan zou MSCORE geadviseerd kunnen worden; MSCORE levert immers (onder bepaalde voorwaarden) consistente schattingen. De vraag blijft dan welke schattingsmethode in kleine steekproeven het beste is. Om een antwoord te kunnen geven op deze vraag hebben we enkele Monte-Carlo experimenten (MC-experimenten) uitgevoerd; zowel met homoskedastische als met heteroskedastische storingen. In bepaalde experimenten is eenvoudig te berekenen wat volgens de theorie de schattingsresultaten zouden moeten zijn. In deze gevallen zullen we dan ook controleren of de resultaten met de theorie overeenkomen. Bovendien zullen we in alle experimenten onze vermoedens omtrent de beste schattingsmethode, ML of MSCORE, uitspreken en vergelijken met de resultaten. We geven een vrij uitvoerige beschrijving van de experimenten, maar zullen eerst een toelichting geven op de verkregen schattingsresultaten en een overzicht van de verschillende gesimuleerde modellen.

Van het PC-pakket LIMDEP hebben we de Probit-ML-methode gebruikt om de modellen te schatten. (In het hierna volgende zullen we Probit-ML aanduiden met MLN, dat wil zeggen Maximum Likelihood onder normaliteit). Het pakket maakt gebruik van de Newton-Raphson-methode (zie (2.4)). Als startwaarden worden de resultaten van een lineaire regressie van de binaire variabele op de exogenen genomen (OLS).

De MSCORE-resultaten zijn verkregen met de PC-versie van het MS-algoritme (MSCORE) van Manski en Thompson (1987). Als er geen startwaarden worden



opgegeven neemt het algoritme de basis  $e_1, e_2 \dots e_k \in \mathbb{R}^k$  met  $e_i$  de vector met als  $i^e$  element 1 en de andere elementen gelijk aan 0 ( $i = 1, 2 \dots k$ ). De variabele  $k$  hierin geeft het aantal te schatten parameters aan. Het algoritme neemt dan als  $i^e$  element van startpunt  $\vartheta_0$  de waarde van de steekproef-scorefunctie (zie (3.12)) geëvalueerd in  $e_i$ . Startpunt  $\vartheta_0$  wordt wel genormeerd zodanig dat  $\|\vartheta_0\| = 1$ .

Zowel bij de uitvoeringen van het ML- als het MS-algoritme zijn de startwaarden door het algoritme zelf bepaald.

Bij alle experimenten hebben we het binaire keuzemodel met twee verklarende variabelen als uitgangspunt genomen:

$$\left. \begin{aligned} y_t^* &= \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \\ y_t &= a \quad \text{als} \quad y_t^* > 0 \\ y_t &= 1 \quad \text{als} \quad y_t^* \leq 0 \\ x_{1t} &\sim F_{x_{1t}} \\ x_{2t} &\sim F_{x_{2t}} \\ \varepsilon_t &\sim F_{\varepsilon_t} \\ (x_{1t}, x_{2t}, \varepsilon_t)' &\text{ onafhankelijk van } (x_{1s}, x_{2s}, \varepsilon_s)' \quad s \neq t \\ x_{1t} \text{ en } x_{2t} &\text{ onafhankelijk van } \varepsilon_t \end{aligned} \right\} s, t = 1, 2 \dots T \quad (5.1)$$

Bij het schatten van het model met MLN is  $a$  gelijk aan 0 en bij het schatten van MSCORE gelijk aan -1. Dit is slechts een kwestie van notatie. Het feit dat bij MSCORE  $y_t = -1$  als  $y_t^* = 0$  (zie (3.2)), is te verwaarlozen omdat wij enkel uitgaan van continue verdelingen van  $\varepsilon_t$ . In alle onderzochte modellen zijn 100 aselechte trekkingen gedaan uit een gekozen verdeling  $F_{\varepsilon_t}$  (dus  $T = 100$ ) en met door ons bepaalde  $\beta' := (\beta_1, \beta_2)$ ,  $x_{1t}$  en  $x_{2t}$  is de latente variabele  $y_t^*$  gegenereerd. Op deze manier kan  $y_t$  worden bepaald. Met de gegevens  $y_t, x_{1t}$  en  $x_{2t}$  ( $t = 1, 2 \dots 100$ ) is  $\beta$  vervolgens met MLN en MSCORE geschat. Dit is voor elk model, dus voor alle verschillende keuzes voor  $F_{\varepsilon_t}$ ,  $\beta$ ,  $x_{1t}$  en  $x_{2t}$  100 keer gedaan. Elke keer dat een model wordt geschat, noemen we een replicatie of "run". Wij voeren dus voor elk model 100 (onafhankelijke) replicaties uit en vinden dan 100 schattingen voor de

'onbekende' parameter  $\beta$ . Deze schattingen vergelijken we met de ware parameter  $\beta$ .

We geven nu een overzicht van de gesimuleerde modellen.

A. Twee modellen met een constante, een stochastische regressor en homoskedastische storingen.

$$\beta' = (1,1)$$

$$\left. \begin{array}{l} x_{1t} = 1 \\ x_{2t} \sim U(-3,3) \end{array} \right\} t = 1,2 \dots 100$$

$$\text{Variant 1) } F_{\varepsilon_t} = N(0, 4/3) \quad t = 1,2 \dots 100$$

$$\text{Variant 2) } F_{\varepsilon_t} = U(-2,2) \quad t = 1,2 \dots 100$$

B. Twee modellen met twee stochastische regressoren en homoskedastische storingen.

$$\beta' = (1,1)$$

$$\left. \begin{array}{l} x_{1t} = U(-1,1) \\ x_{2t} \sim U(-3,3) \end{array} \right\} t = 1,2 \dots 100$$

$$\text{Variant 1) } F_{\varepsilon_t} = N(0, 4/3) \quad t = 1,2 \dots 100$$

$$\text{Variant 2) } F_{\varepsilon_t} = U(-2,2) \quad t = 1,2 \dots 100$$

C. Twee modellen met een constante, een stochastische regressor en heteroskedastische storingen.

$$\beta' = (1,1)$$

$$\left. \begin{array}{l} x_{1t} = 1 \\ x_{2t} \sim U(-3,3) \end{array} \right\} t = 1,2 \dots 100$$

$$\left. \begin{array}{l} \text{Variant 1) } F_{\varepsilon_t} = N(0, \sigma_t^2) \\ \sigma_t^2 := s^2 h^2(x_{2t}) \quad \text{met } s^2 = 4/3 \\ h^2(x_{2t}) = \exp\{x_{2t} - 6/5\} \end{array} \right\} t = 1,2 \dots 100$$

$$\left. \begin{array}{l} \text{Variant 2) } F_{\varepsilon_t} = N(0, \sigma_t^2) \\ \sigma_t^2 := s^2 h^2(x_{2t}) \quad \text{met } s^2 = 4/3 \\ h(x_{2t}) = |1/2 x_{2t} + 1/2| \end{array} \right\} t = 1,2 \dots 100$$

De trekkingen uit de uniforme en normale verdelingen zijn met behulp van resp. de congruentie- en de rejectiemethode verkregen (Kleijnen (1988), blz.

24 en blz. 59). De verdelingen onder A en B zijn zodanig gekozen dat de varianties overeenkomen. Net als alle replicaties zijn bij onderdeel A en B de storingen en de regressoren onderling onafhankelijk. Bij de modellen onder C moet opgemerkt worden dat eerst  $x_{2t}$  getrokken is uit  $U(-3,3)$  en vervolgens storingen getrokken zijn uit de gegeven verdelingen. De verdelingen van CV1 en CV2 (resp. variant 1 en variant 2 van onderdeel C) zijn dus conditionele verdelingen. Bovendien geldt onder CV1 dat  $E\{h^2(x_{2t})\} = \pm 1$  en onder CV2 dat  $E\{h^2(x_{2t})\} = 1$ . Voor het feit dat wij juist voor deze zes modellen hebben gekozen, is er verder geen speciale reden.

Alvorens we de resultaten bespreken, moet er nog een opmerking met betrekking tot de schattingsresultaten worden gemaakt. Schatten met MLN geeft resultaten voor  $\beta_1/\sigma$  en  $\beta_2/\sigma$  (met  $\sigma^2$  de variantie van verdeling  $F_{\epsilon_t}$ ). Schatten met MSCORE geeft daarentegen resultaten voor  $\beta_1/\sqrt{(\beta_1^2 + \beta_2^2)}$  en  $\beta_2/\sqrt{(\beta_1^2 + \beta_2^2)}$ . Omdat  $\beta$  dus slechts geïdentificeerd is op een schaalfactor na en omdat de MLN-resultaten niet direct te vergelijken zijn met de MSCORE-resultaten, nemen we bij beide methoden als de te onderzoeken parameter:

$$\lambda := \arctan(\beta_2/\beta_1) \quad (5.2)$$

Deze parameter geeft de hoek in radialen tussen  $\beta := (\beta_1, \beta_2)'$  en de  $\beta_1$ -as. Voor alle modellen wordt op basis van de middels 100 replicaties verkregen schattingen  $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_{100}$  voor  $\lambda$  een schatting voor de "bias" berekend, de standaardafwijking van schatter  $\hat{\lambda}_r$  ( $r = 1, 2, \dots, 100$ ) en de "root mean square error" (RMSE). Deze grootheden worden gegeven door resp.:

$$\frac{1}{100} \sum_{r=1}^{100} \hat{\lambda}_r - \lambda \quad (\text{"bias"}) \quad (5.3)$$

$$\sqrt{\left(\frac{1}{100} \sum_{r=1}^{100} (\hat{\lambda}_r - \bar{\lambda})^2\right)} \quad \text{met } \bar{\lambda} := \frac{1}{100} \sum_{r=1}^{100} \hat{\lambda}_r \quad (\text{SDEV}) \quad (5.4)$$

$$\sqrt{\left(\frac{1}{100} \sum_{r=1}^{100} (\hat{\lambda}_r - \lambda)^2\right)} \quad (\text{RMSE}) \quad (5.5)$$

Merk hierbij op dat RMSE bepaald wordt door de andere twee grootheden, immers:

$$(\text{RMSE})^2 = (\text{"bias"})^2 + (\text{SDEV})^2 \quad (5.6)$$

ad A) Twee modellen met een constante, een stochastische regressor en homoskedastische storingen.

Het probitmodel met deterministische exogenen is uitvoerig besproken in paragraaf 2. De eerste variant met  $\epsilon_t \sim N(0, 4/3)$  is een probitmodel met een stochastische exogene. Ondanks deze stochastiek is te verwachten dat MLN goede schattingsresultaten geeft. Voor de likelihoodfunctie  $L(\epsilon_t, x_{2t})$  geldt immers:

$$L(\epsilon_t, x_{2t}) = L(\epsilon_t | x_{2t}) L(x_{2t}) \quad (5.7)$$

Hierin is  $L(\epsilon_t | x_{2t})$  de likelihood van  $\epsilon_t$  gegeven  $x_{2t}$ . Definiëren we  $\vartheta := (\beta_1/\sigma, \beta_2/\sigma)'$ , dan komt maximaliseren van de log-likelihood neer op:

$$\begin{aligned} \text{Max}_{\vartheta} \quad & \log L(\epsilon_t | x_{2t}) + \log L(x_{2t}) = \\ \text{Max}_{\vartheta} \quad & \log L(\epsilon_t | x_{2t}) \end{aligned} \quad (5.8)$$

De functie  $L(x_{2t})$  vervalt in de criteriumfunctie omdat de onbekende parameter  $\vartheta$  hier niet in voorkomt. De te maximaliseren functie blijft dan gelijk aan de log-likelihood besproken in paragraaf 2 met  $x_{2t}$  vervangen door de realisatie van  $x_{2t}$ . Voor de informatiematrix geldt nu (zie (2.6))

$$I(\vartheta) = E \left\{ \sum_{t=1}^T \frac{(y_t - \Phi(\vartheta' x_t))^2}{[\Phi(\vartheta' x_t)(1 - \Phi(\vartheta' x_t))]^2} \varphi^2(\vartheta' x_t) x_t x_t' \right\}$$

met  $x_t' = (1, x_{2t})$ ,  $x_{2t} \sim U(-3, 3)$  en  $\vartheta' = (\frac{1}{2}\sqrt{3}, \frac{1}{2}\sqrt{3})$ .

Deze matrix kan worden geschat door de ware parameter  $\vartheta$  en  $T$  te substitueren en een gemiddelde over alle 100 replicaties te nemen:

$$\frac{1}{100} \sum_{r=1}^{100} \sum_{t=1}^{100} \frac{(y_t - \Phi(\vartheta' x_{rt}))^2}{[\Phi(\vartheta' x_{rt})(1 - \Phi(\vartheta' x_{rt}))]^2} \varphi^2(\vartheta' x_{rt}) x_{rt} x_{rt}' \quad (5.9)$$



Hierin is  $x_{rt} \in \mathbb{R}^2$  de vector met de  $t^e$  waarnemingen in de  $r^e$  run. Uit de resultaten hiervan (en na inverteren) volgt voor de MLN-schatter  $\hat{\theta}' := (\hat{\theta}_1, \hat{\theta}_2)$  voor  $\theta'$  dat:

$$\begin{aligned} \text{Avar}\{\hat{\theta}_1\} &= 0.0486 \\ \text{Avar}\{\hat{\theta}_2\} &= 0.0238 \end{aligned} \quad (5.10)$$

Als de verdeling van de storingsterm geen normale verdeling is, zullen de MLN-schatters voor  $\beta_1$  en  $\beta_2$  in het algemeen inconsistent zijn. Met betrekking tot model AV2 is daarom te verwachten dat MSCORE betere schattingsresultaten zal geven dan MLN; aan de voorwaarde  $\text{MED}(\epsilon_t | x_{2t}) = 0$  is voldaan en MSCORE is derhalve consistent. Hierbij schuilt, vanwege onze keuze van regressoren, echter een addertje onder het gras. Hierop komen we terug na het presenteren van de resultaten.

In tabel 5.1 en tabel 5.2 geven we de schattingsresultaten voor resp. variant 1 en variant 2.

Tabel 5.1 Resultaten model AV1 :  $\epsilon_t \sim N(0, 4/3)$

	MLN	MSCORE
"bias"	-0.0059	0.0107
SDEV	0.0905	0.2931
RMSE	0.0907	0.2933

Tabel 5.2 Resultaten model AV2 :  $\epsilon_t \sim U(-2, 2)$

	MLN	MSCORE
"bias"	-0.0025	0.1020
SDEV	0.1077	0.5042
RMSE	0.1077	0.5144

De ware waarde van  $\lambda$  is  $\frac{1}{4}\pi = 0.7854$ . Vanwege het feit dat de schatting voor  $\lambda$  kan variëren van  $-\frac{1}{2}\pi$  tot  $\frac{1}{2}\pi$  lijkt een "bias" van -0.0059 (model AV1, MLN) daarom niet slecht. Wat meteen opvalt aan deze resultaten is dat MLN met uniforme trekkingen de resultaten van MLN met normale trekkingen sterk benadert. De "bias" is bij uniforme trekkingen zelfs beter, al is het verschil met model AV1 wel minimaal. De RMSE die bepaald wordt door de "bias" en SDEV (zie (5.6)), is bij model AV1 iets beter en dus kan dit model

wel het beste worden genoemd. De modellen schatten met MSCORE, met name model AV2, lijkt wel hele slechte resultaten te geven.

Omdat de schatters voor  $\lambda$  onafhankelijk en identiek verdeeld zijn geldt volgens de centrale limietstelling voor  $\hat{\bar{\lambda}} := \frac{1}{R} \sum_{r=1}^R \hat{\lambda}_r$ , mits

$v := V\{\lambda_r\} > 0$  bestaat:

$$\sqrt{R}(\hat{\bar{\lambda}} - E\{\hat{\lambda}_r\}) \stackrel{a}{\sim} N(0, v) \quad (5.11)$$

Wij hebben 100 replicaties uitgevoerd, dat wil zeggen 100 schattingen voor  $\lambda$ , dus  $R = 100$ . Nemen we voor de verwachting van  $\hat{\lambda}_r$  de ware parameter  $\lambda$  en als schatting voor  $v$  de gecorrigeerde steekproefvariantie ( $GSDEV := \sqrt{\frac{R \cdot SDEV \cdot SDEV}{R-1}}$ ), dan kan er een 'betrouwbaarheidsinterval' voor  $\lambda$  worden opgesteld:

$$\hat{\bar{\lambda}} \pm (GSDEV/10 * u_{\alpha}) \quad (5.12)$$

Hierin is  $u_{\alpha}$  het  $(1-\alpha)^e$  kwantiel van de normale verdeling. Bij MLN met normale trekkingen is  $\hat{\bar{\lambda}} = 0.7795$ . Het 95%-'betrouwbaarheidsinterval' voor  $\lambda$  is dan:

$$(0.7617; 0.7973)$$

De werkelijke waarde van  $\lambda$  is  $\frac{1}{4}\pi = 0.7854$  en valt inderdaad binnen dit interval. De 'betrouwbaarheidsintervallen' bij de andere drie methoden zijn:

$$(0.7384; 0.8539) \quad \text{AV1 met MSCORE}$$

$$(0.7617; 0.8041) \quad \text{AV2 met MLN}$$

$$(0.7881; 0.9867) \quad \text{AV2 met MSCORE}$$

Bij model AV2 geschat met MLN valt de ware  $\lambda$  binnen het 'betrouwbaarheidsinterval' en bij schatten met MSCORE net niet. De verwachting dat de schattingsresultaten van het MS-algoritme beter zouden zijn dan de resultaten van een verkeerde toepassing van het ML-algoritme, wat hier wil zeggen schatten met MLN terwijl de ware verdeling uniform  $U(-2, 2)$  is, is dus niet uitgekomen. Wij komen hierbij terug op het onverwachte aspect van onze keuze van exogenen (het zogenaamde addertje onder het gras). Er blijken situaties te zijn waarin MLN op een schaalfactor na consistent schat, ondanks het feit dat de likelihoodfunctie niet gebaseerd is op de ware

verdeling. Onze keuze van exogenen is één van deze gevallen. Ruud (1983) gaat uit van een model dat een constante bevat en toont aan dat onder bepaalde voorwaarden de MLN-schatters voor alle parameters, behalve de constante, consistent zijn. Naast enkele standaard regulariteitsvoorwaarden, zijn voldoende voorwaarden hiervoor dat:

- (i) trekkingen van exogenen en storingen onafhankelijk zijn;
- (ii) de conditionele verwachting van elke exogene (behalve de constante), gegeven  $z$ , lineair is in  $z$ . Hierin is  $z$  de lineaire combinatie van de ware parameters met de exogenen.

Nemen we  $y_t = \beta_1 + \sum_{k=2}^K \beta_k x_{tk} + \varepsilon_t$  en  $z_t = \beta_1 + \sum_{k=2}^K \beta_k x_{tk}$ , dan komt de tweede voorwaarde neer op:

$$E\{x_{tk} | z_t\} = c + d \cdot z_t \quad \forall k \in \{2, 3, \dots, K\} \quad c, d \in \mathbb{R} \quad (5.13)$$

Het model AV2 voldoet aan de voorwaarden (i) en (ii), immers voor alle  $t$  en  $k \in \{2, 3, \dots, K\}$ :

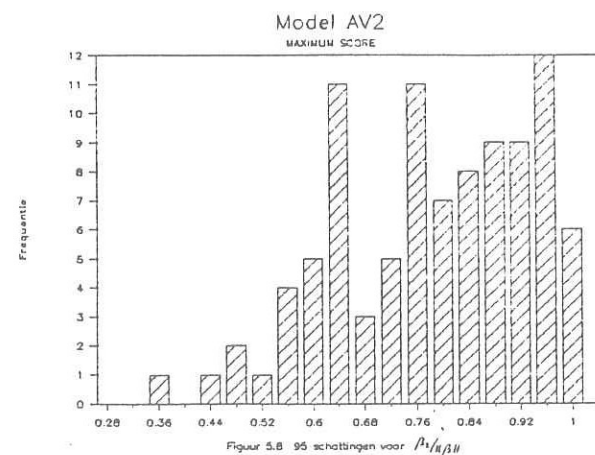
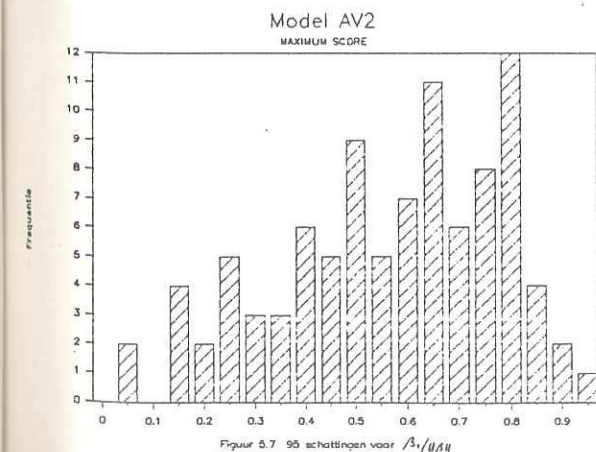
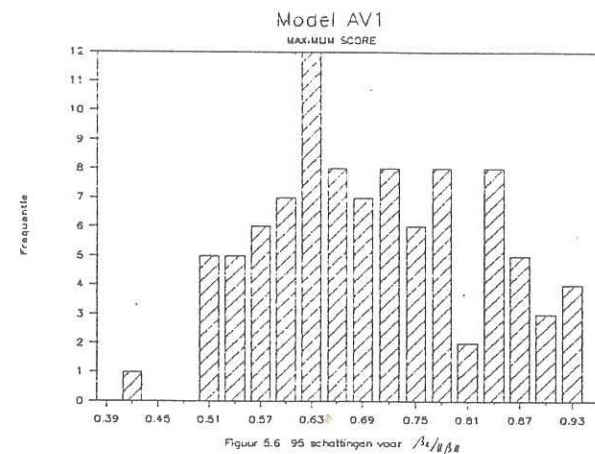
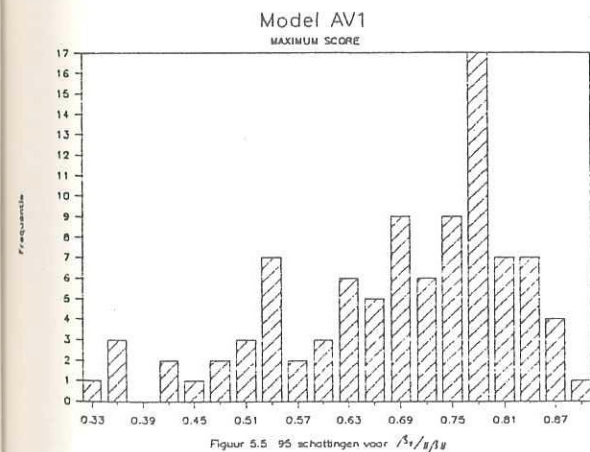
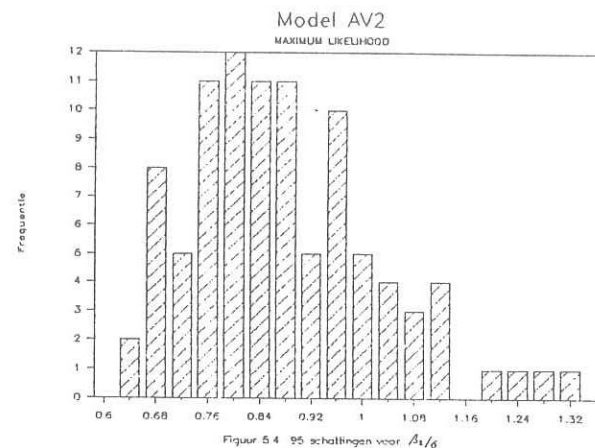
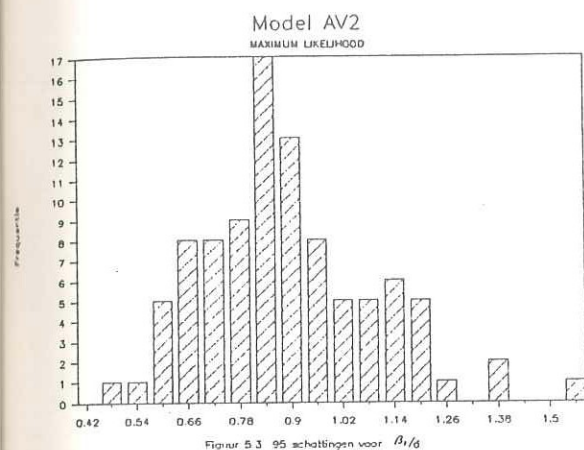
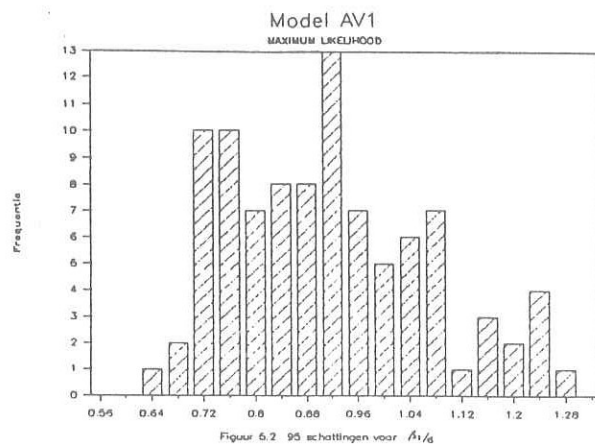
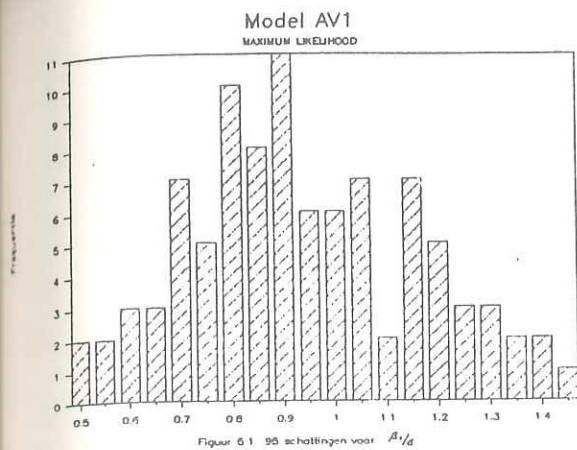
$$z_t = \beta' x_t = x_{1t} + x_{2t} = 1 + x_{2t} \sim U(-2, 4)$$

$$E\{x_{2t} | z\} = x_{1t} + x_{2t} - 1 = -1 + 1 \cdot (x_{1t} + x_{2t})$$

Hiermee is een verklaring gegeven voor de relatief goede schattingsresultaten die werden verkregen voor parameter  $\lambda$  bij schatten van model AV2 met MLN.

In de figuren 5.1 tot en met 5.4 zijn de MLN-schattingen voor  $\beta_1/\sigma$  en  $\beta_2/\sigma$  van resp. model AV1 en AV2 uitgezet. De figuren 5.5 tot en met 5.8 geven MSCORE-schattingen voor  $\beta_1/\sqrt{(\beta_1^2 + \beta_2^2)}$  en  $\beta_2/\sqrt{(\beta_1^2 + \beta_2^2)}$  voor beide modellen. Er is een opdeling gemaakt in 20 intervallen en de waarden op de horizontale as geven de middelpunten van de intervallen aan. We hebben 5% van de schattingen, zogenaamde 'uitschieters', weggelaten omdat deze uitschieters vaak een vertekend beeld geven. Dit is met name het geval bij het schatten met MLN; bij 100 replicaties is er vaak een schatting die extreem slecht is. Een voor de hand liggend criterium voor het elimineren zou gebaseerd kunnen zijn op de schattingen voor  $\lambda$ . Een dergelijk criterium zou echter niet noodzakelijk hele slechte MLN-resultaten (dus hele grote schattingen van de







parameters en varianties) elimineren, terwijl dit wel wenselijk is. Daarom hebben we een criterium genomen dat zulke extreme waarden wel uitsluit. Noemen we de MLN-schattingen voor  $\beta_1/\sigma$  en  $\beta_2/\sigma$  bij de  $r^e$  replicatie resp.  $h_{1,r}$  en  $h_{2,r}$ , dan is het criterium dat gehanteerd is bij de figuren 5.1 tot en met 5.4:

$$(h_{1,r} - \beta_1/\sigma)^2 + (h_{2,r} - \beta_2/\sigma)^2 \quad (5.14)$$

Noemen we  $k_{1,r}$  en  $k_{2,r}$  de MSCORE-schattingen voor resp.  $\beta_1/\sqrt{(\beta_1^2 + \beta_2^2)}$  en  $\beta_2/\sqrt{(\beta_1^2 + \beta_2^2)}$  in de  $r^e$  replicatie, dan is het gehanteerde criterium bij de figuren 5.5 tot en met 5.8:

$$(k_{1,r} - \beta_1/\sqrt{(\beta_1^2 + \beta_2^2)})^2 + (k_{2,r} - \beta_2/\sqrt{(\beta_1^2 + \beta_2^2)})^2 \quad (5.15)$$

De vijf replicaties met de grootste criteriumwaarde zijn geëlimineerd. Bekijken we de figuren 5.1 en 5.2 dan zien we dat de MLN-schattingen geconcentreerd zijn om de intervallen met de ware parameters, resp.  $\vartheta_1 = \frac{1}{2}\sqrt{3}$  en  $\vartheta_2 = \frac{1}{2}\sqrt{3}$ . Duidelijk is dat deze verdelingen de asymptotische normale verdelingen van de MLN-schatters voor  $\vartheta_1$  en  $\vartheta_2$  met verwachting  $\frac{1}{2}\sqrt{3} = \pm 0.87$  zouden kunnen benaderen. De verdelingen van de MLN-schatters met uniform verdeelde storingen blijken ongeveer een zelfde beeld te vertonen. Bekijken we de MSCORE-schattingen, dan zien we hele andere verdelingen. De figuren 5.5 tot en met 5.8 met  $R=95$  vertonen verdelingen die duidelijk niet symmetrisch zijn, zodat niet te verwachten is dat de asymptotische verdeling normaal is. Zoals in paragraaf 3 al is besproken, heeft de MSCORE-schatter een asymptotisch niet-normale verdeling. Een beter beeld van de asymptotische verdeling zou verkregen kunnen worden door een veel groter aantal waarnemingen ( $T \rightarrow \infty$ ) te nemen. De verdeling van de MSCORE-schatter convergeert immers vrij langzaam (langzamer dan de MLN-schatter) naar de asymptotisch verdeling; voor de MLN-schatter  $\hat{\vartheta}_{ML}$  en de MSCORE-schatter  $\hat{\vartheta}_{MS}$  voor  $\vartheta$  geldt resp.:

$$\begin{aligned} \sqrt{T}(\hat{\vartheta}_{ML} - \vartheta) &\xrightarrow{a} N(0, V) \\ T^{1/3}(\hat{\vartheta}_{MS} - \vartheta) &\xrightarrow{a} NN \end{aligned} \quad (5.16)$$

Hierin is  $V$  de asymptotische covariantiematrix van  $\hat{\vartheta}_{ML}$  en  $NN$  de niet-normale asymptotische verdeling van  $\hat{\vartheta}_{MS}$ . Onder bepaalde voorwaarden convergeert de verdeling van het gemiddelde van de schattingen volgens de centrale limietstelling (zie (5.11)) naar een normale verdeling, als het aantal replicaties ( $R$ ) naar oneindig gaat. Helaas is het niet mogelijk om op een eenvoudige manier te verifiëren of deze convergentie voor onze resultaten (met  $R=95$ ) in overeenstemming zou kunnen zijn met de theorie.

Voor het model met de normale trekkingen,  $AV1$ , kan gecontroleerd worden of de varianties van de MLN-schatters overeenkomen met de geschatte asymptotische varianties (zie (5.10)). Omdat wij 100 modellen hebben geschat, en dus 100 varianties voor beide MLN-schatters hebben, nemen we het gemiddelde van deze varianties. Voor deze gemiddelden van  $\hat{\vartheta}_{1ML,1}, \hat{\vartheta}_{1ML,2} \dots \hat{\vartheta}_{1ML,100}$  voor  $\vartheta_1$  en  $\hat{\vartheta}_{2ML,1}, \hat{\vartheta}_{2ML,2} \dots \hat{\vartheta}_{2ML,100}$  voor  $\vartheta_2$  geldt:

$$\begin{aligned} \frac{1}{R} \sum_{r=1}^R V\{\hat{\vartheta}_{1ML,r}\} &\xrightarrow{p} E\{V\{\hat{\vartheta}_{1ML,r}\}\} , \quad R \rightarrow \infty \\ \frac{1}{R} \sum_{r=1}^R V\{\hat{\vartheta}_{2ML,r}\} &\xrightarrow{p} E\{V\{\hat{\vartheta}_{2ML,r}\}\} , \quad R \rightarrow \infty \end{aligned} \quad (5.17)$$

De berekende gemiddelden van de varianties zijn:

$$\begin{aligned} \frac{1}{100} \sum_{r=1}^{100} V\{\hat{\vartheta}_{1ML,r}\} &= 0.0609 \\ \frac{1}{100} \sum_{r=1}^{100} V\{\hat{\vartheta}_{2ML,r}\} &= 0.0317 \end{aligned}$$

Door het aantal replicaties en het aantal waarnemingen te vergroten ( $R \rightarrow \infty$  en  $T \rightarrow \infty$ ) zouden de gemiddelden van de varianties de asymptotische varianties sterk moeten benaderen. De gemiddelden van de geschatte varianties komen dan in de buurt van de verwachting van de varianties (zie (5.17)) en door het grote aantal waarnemingen benaderen deze de asymptotische varianties (te berekenen door de inverse van de informatiematrix (2.6) te nemen). We zien dat bij  $R=100$  en  $T=100$  de

gemiddelde varianties (0.0609 en 0.0317) en de geschatte asymptotische varianties (0.0486 en 0.0238, zie (5.10)) al redelijk in elkaars richting wijzen.

Uit de resultaten van model AV2 bleek MLN een betere schattingsmethode te zijn dan MSCORE, waarschijnlijk vanwege het feit dat MLN met een uniform verdeelde storingsterm in dit model toch consistent is. We bespreken nu een model met een uniform verdeelde storingsterm waarbij MLN geen consistente schattingsmethode is.

ad B) Twee modellen met twee stochastische regressoren en homoskedastische storingen.

Het enige verschil van de modellen BV1 en BV2 met de vorige twee modellen is de stochastiek van de eerste regressor ( $x_{1t} \sim U(-1,1)$ ). Dit verschil zorgt ervoor dat de MLN-schatters bij variant 2 zeer waarschijnlijk niet consistent zijn. Bij model BV2 is dus te verwachten dat MSCORE betere schattingsresultaten geeft dan MLN. De MLN-methode bij model BV1 moet weer duidelijk de beste resultaten opleveren; dit model wordt immers met de juiste methode geschat. MSCORE levert weliswaar ook consistente schatters, maar de convergentie van de MLN-schatters naar de ware parameters treedt eerder op dan de convergentie van de MSCORE-schatters (zie (5.16)).

Berekenen van de asymptotische varianties voor resp.  $\vartheta_1 := \beta_1/\sigma$  en  $\vartheta_2 := \beta_2/\sigma$ , analoog aan (5.9) geeft:

$$\begin{aligned} \text{Avar}\{\hat{\vartheta}_1\} &= 0.1023 \\ \text{Avar}\{\hat{\vartheta}_2\} &= 0.0217 \end{aligned} \tag{5.18}$$

De schattingsresultaten van beide modellen staan in tabel 5.3 en tabel 5.4.

Tabel 5.3 Resultaten model BV1 :  $\epsilon_t \sim N(0,4/3)$

	MLN	MSCORE
"bias"	0.0213	-0.1280
SDEV	0.1677	0.5198
RMSE	0.1691	0.5353



Tabel 5.4 Resultaten model BV2 :  $\epsilon_t \sim U(-2,2)$

	MLN	MSCORE
"bias"	0.1408	-0.1014
SDEV	0.3215	0.8366
RMSE	0.3510	0.8428

Het model BV1 geschat met MLN geeft, zoals verwacht mag worden, inderdaad de beste resultaten. Opvallend aan de resultaten van model BV2 is dat de "bias" bij schatten met MLN iets groter is dan bij schatten met MSCORE, maar dat op basis van de RMSE duidelijk MLN toch de betere methode is. Berekenen van de 'betrouwbaarheidsintervallen' voor  $\lambda = \frac{1}{4}\pi = 0.7854$  (zie (5.12)) geeft:

(0.7737;0.8398)	BV1 met MLN
(0.8629;0.9895)	BV2 met MLN
(0.5550;0.7598)	BV1 met MSCORE
(0.5192;0.8488)	BV2 met MSCORE

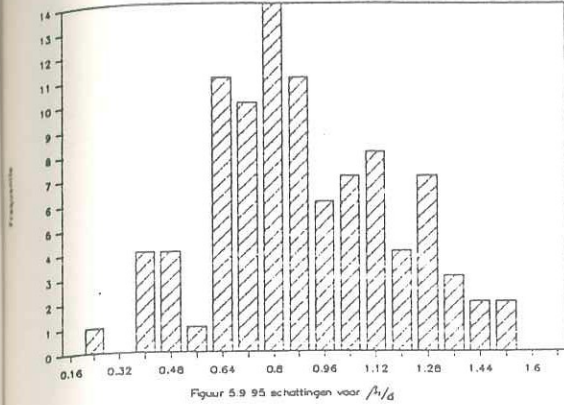
De ware  $\lambda$  valt bij model BV2 geschat met MLN buiten het interval; deze resultaten komen daarom overeen met de theorie dat MLN met uniforme trekkingen in dit model inconsistent is. De MSCORE-schatter bij model BV1 zou op basis van deze 'betrouwbaarheidsintervallen' (net) inconsistent kunnen worden genoemd, iets wat niet strookt met de theorie. We hebben echter slechts 100 replicaties uitgevoerd. Door meer replicaties uit te voeren zal het 'betrouwbaarheidsinterval' de ware parameter waarschijnlijk wel omvatten;  $\lambda$  valt immers maar net buiten het berekende interval.

In de figuren 5.9 tot en met 5.16 zijn voor model BV1 en BV2 de MLN-schattingen voor  $\beta_1/\sigma$ ,  $\beta_2/\sigma$  ( $\pm 0.87$ ) en MSCORE-schattingen voor  $\beta_1/\sqrt{(\beta_1^2 + \beta_2^2)}$  en  $\beta_2/\sqrt{(\beta_1^2 + \beta_2^2)}$  ( $\pm 0.71$ ) uitgezet, na eliminatie van vijf extreme resultaten op basis van de criteria (5.14) en (5.15). Te zien is dat de MLN-schattingen bij model BV1 geconcentreerd zijn om het interval met de ware parameter. De verdeling van de schattingen voor  $\beta_2/\sigma$  bij dit model lijken niet symmetrisch. Door het aantal waarnemingen te vergroten ( $T \rightarrow \infty$ ), zouden de asymptotische normale verdelingen met verwachting  $\frac{1}{2}\sqrt{3} = \pm 0.87$  beter moeten worden benaderd. De verdelingen van de MSCORE-schatters laten duidelijk niet-symmetrische verdelingen zien die meerdere pieken vertonen. Het is dan ook te verwachten dat door een toename van het aantal waarnemingen de verdelingen in elk geval naar niet-normale verdelingen zullen convergeren.



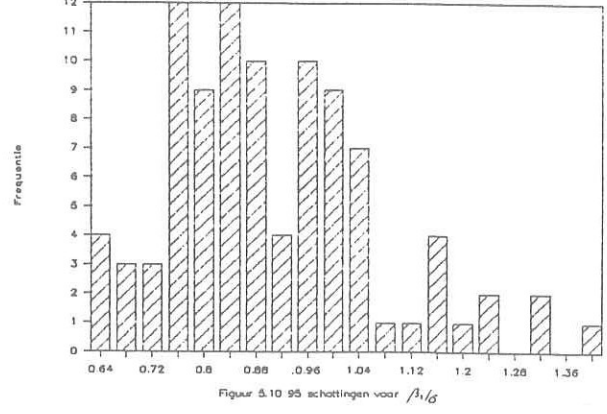
Model BV1

MAXIMUM LIKELIHOOD



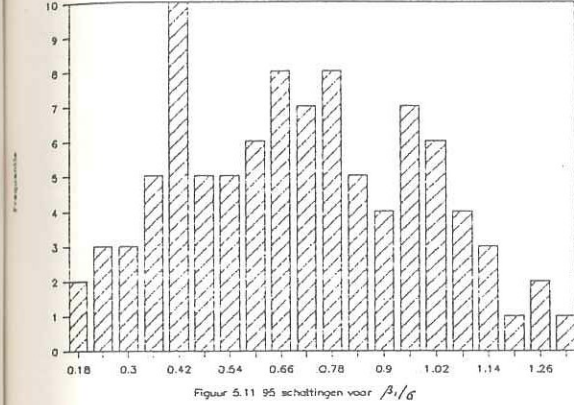
Model BV1

MAXIMUM LIKELIHOOD



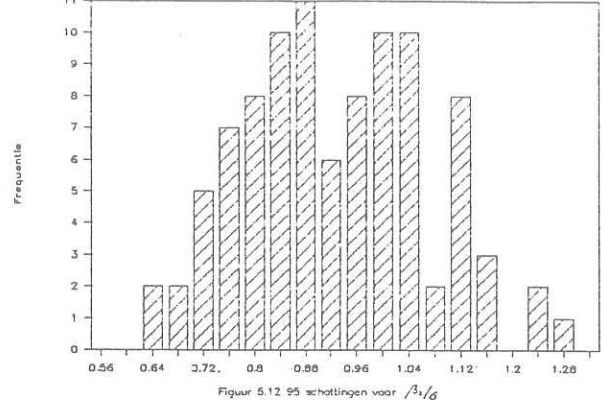
Model BV2

MAXIMUM LIKELIHOOD



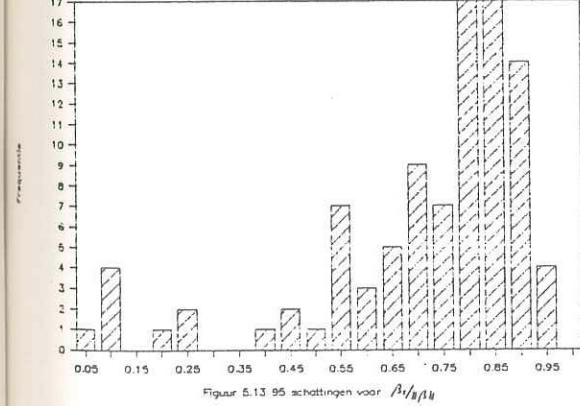
Model BV2

MAXIMUM LIKELIHOOD



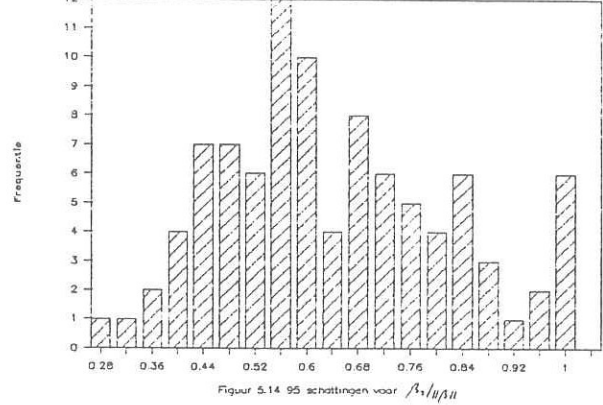
Model BV1

MAXIMUM SCORE



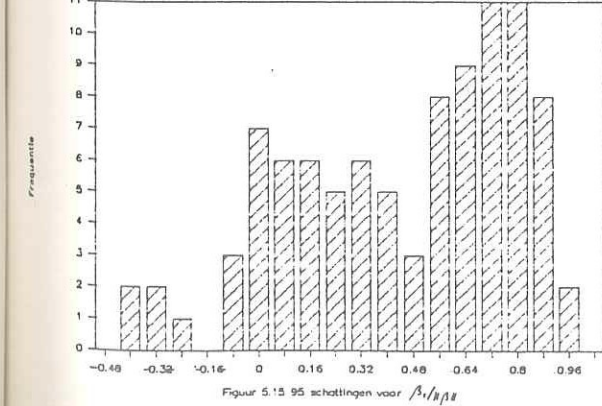
Model BV1

MAXIMUM SCORE



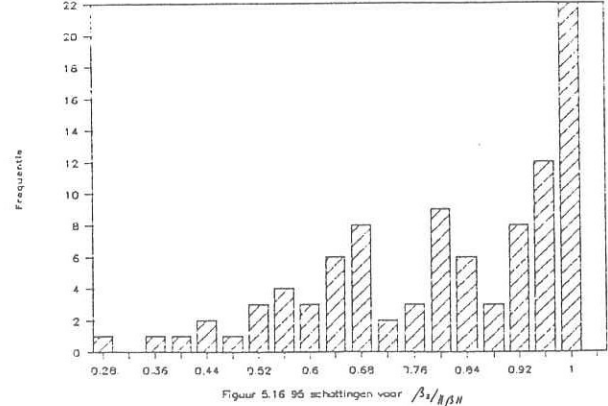
Model BV2

MAXIMUM SCORE



Model BV2

MAXIMUM SCORE



Het berekenen van het gemiddelde van de 100 geschatte varianties van de MLN-schatters voor  $\theta'$  bij model BV1, op basis van (5.17), geeft:

$$\frac{1}{100} \sum_{r=1}^{100} V\{\hat{\theta}_{1ML,r}\} = 0.1175$$

$$\frac{1}{100} \sum_{r=1}^{100} V\{\hat{\theta}_{2ML,r}\} = 0.0274$$

Deze schattingen kunnen de geschatte asymptotische varianties 0.1023 en 0.0217 (zie 5.18) goed benaderen.

Op basis van de geschatte RMSE is MSCORE in deze modellen niet beter dan MLN. Bij de modellen met normale trekkingen was dit te verwachten, bij de modellen met uniforme trekkingen echter niet. De vraag rijst nu natuurlijk of bij een verkeerd gespecificeerde verdeling van de storingsterm in kleine steekproeven, MSCORE wel een beduidend betere schattingsmethode zal zijn dan MLN.

We simuleren twee modellen met heteroskedastische storingen.

ad C) Twee modellen met een stochastische regressor en heteroskedastische storingen.

De modellen CV1 en CV2 hebben een storingsterm die, geconditioneerd op  $x_{2t}$ , normaal is verdeeld. Toch geldt voor deze modellen niet meer dat de likelihoodfunctie uit paragraaf 2 de goede likelihoodfunctie is; in de likelihoodfunctie (2.2) moet de standaardafwijking  $\sigma$  worden vervangen door  $\sigma_t$  waarbij:

$$\sigma_t = \frac{2}{3}\sqrt{3} \exp\left\{\frac{1}{2} x_{2t} - \frac{3}{5}\right\} \quad \text{in CV1} \quad (5.19)$$

$$\sigma_t = \frac{2}{3}\sqrt{3} \left|\frac{1}{2} x_{2t} + \frac{1}{2}\right| \quad \text{in CV2} \quad (5.20)$$

In ieder geval kan derhalve worden verwacht dat de modellen niet consistent worden geschat met MLN. Omdat aan de veronderstelling  $\text{MED}(\epsilon_t | x_{2t}) = 0$  door beide modellen voldaan is, is MSCORE wel een 'juiste' schattingsmethode.

Tabel 5.5 Resultataten model CV1 :  $\epsilon_t | x_{2t} \sim N(0, \frac{4}{3} \exp\{x_{2t} - \frac{6}{5}\})$

	MLN	MSCORE
"bias"	0.1320	0.0021
SDEV	0.0953	0.1163
RMSE	0.1629	0.1163

Tabel 5.6 Resultataten model CV2 :  $\epsilon_t | x_{2t} \sim N(0, \frac{4}{3} |\frac{1}{2} x_{2t} + \frac{1}{2}|^2)$

	MLN	MSCORE
"bias"	0.0596	0.0036
SDEV	0.0990	0.0295
RMSE	0.1156	0.0297

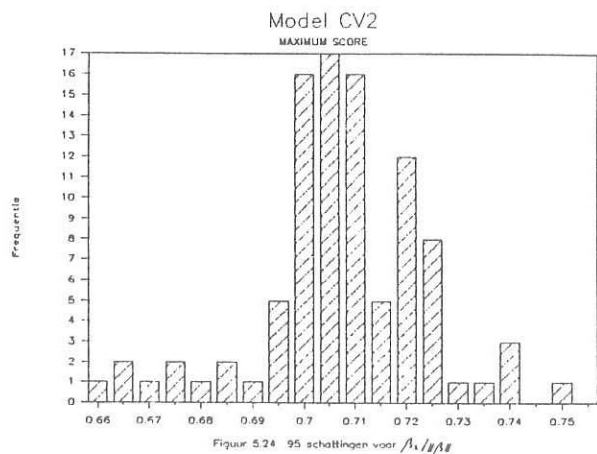
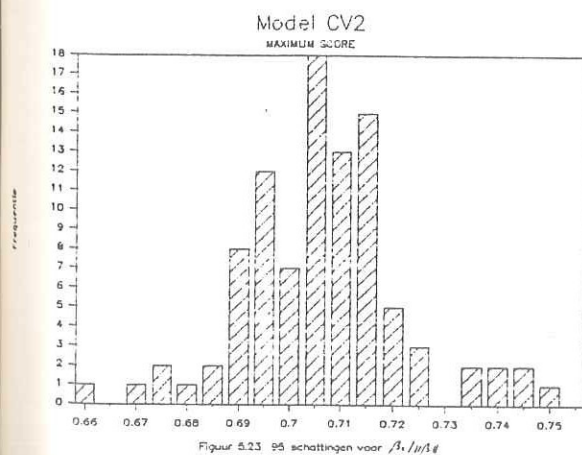
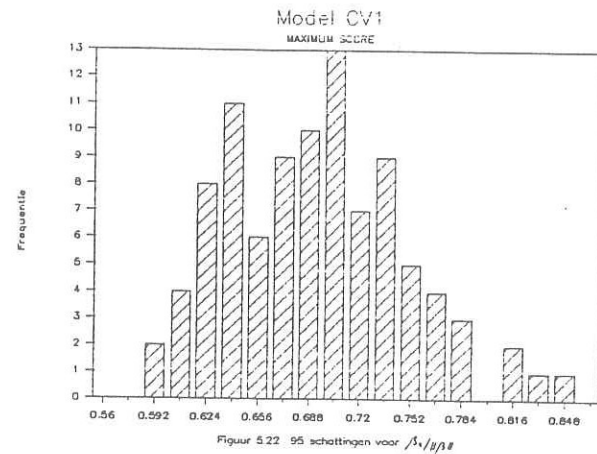
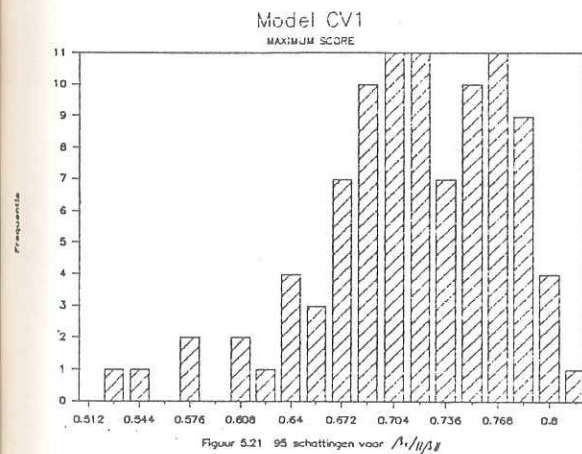
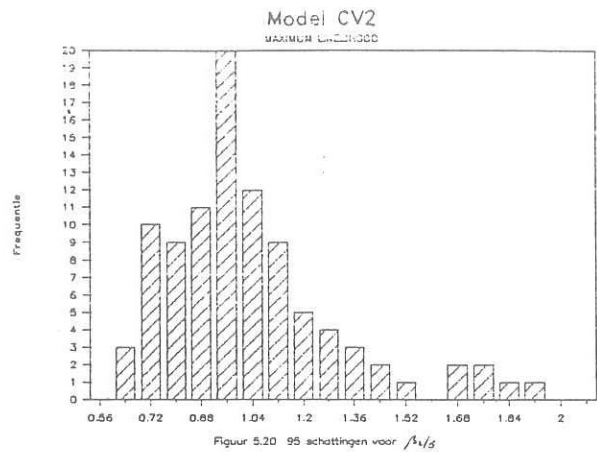
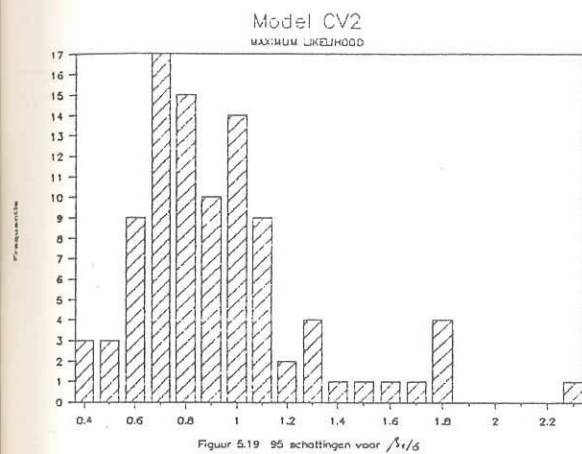
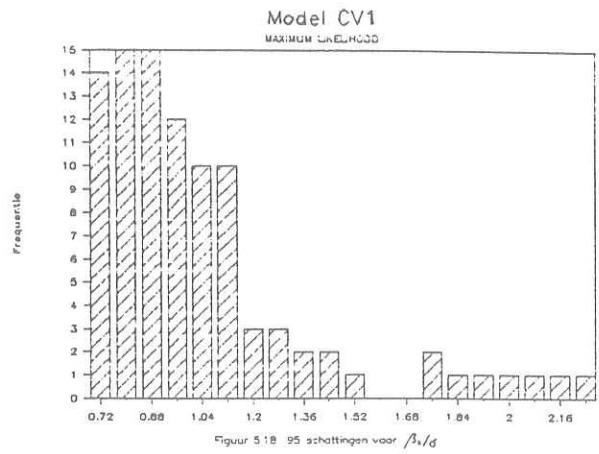
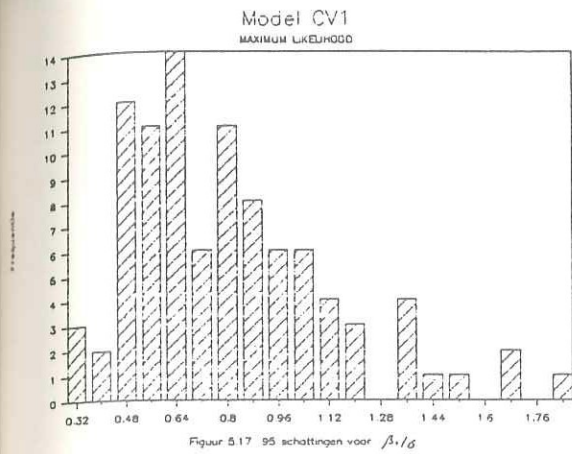
Bekijken we de resultaten in de tabellen 5.5 en 5.6, dan is duidelijk dat op basis van de RMSE MSCORE inderdaad de betere schattingsmethode is bij beide modellen. Berekening van de 'betrouwbaarheidsintervallen' voor de ware parameter  $\lambda = \frac{1}{4} \pi = 0.7854$  geeft:

- (0.8987;0.9362) CV1 met MLN
- (0.8254;0.8645) CV2 met MLN
- (0.7646;0.8104) CV1 met MSCORE
- (0.7832;0.7948) CV2 met MSCORE

Deze 'betrouwbaarheidsintervallen' laten duidelijk zien dat de ware parameter bij schatten met MLN buiten het 'interval' valt. MSCORE is consistent.

De figuren 5.17 tot en met 5.24 geven voor model CV1 de MLN- en MSCORE-schattingen voor  $\beta_1/\sigma_t, \beta_2/\sigma_t, \beta_1/\sqrt{(\beta_1^2 + \beta_2^2)}$  en  $\beta_2/\sqrt{(\beta_1^2 + \beta_2^2)}$ . Bij de schattingen zijn vijf extreme schattingen weggelaten op basis van de criteria (5.14) en (5.15). De MLN-schattingen zijn verdeeld over vrij grote intervallen en hebben naar links scheve verdelingen. De MSCORE-schattingen laten verdelingen zien die geconcentreerd zijn om het interval met de ware parameter  $\theta' = (0.7071, 0.7071)$ . De schattingen hebben als mediaan ongeveer de ware parameter en zijn verdeeld over een vrij klein interval, met name is dit het geval met de MSCORE-schattingen bij model CV2 (figuur (5.23) en (5.24)).





Tot dusver hebben we enkel over Probit-ML (MLN) gesproken, dat wil zeggen dat de likelihood behorende bij het binaire keuzemodel met  $\epsilon_t \sim N(0, \sigma^2)$  als criteriumfunctie werd genomen. De juiste likelihood voor de modellen CV1 en CV2 is echter gebaseerd op  $N(0, \sigma_t^2)$  met  $\sigma_t$  gelijk aan (5.19) resp. (5.20). Te verwachten is dat de varianten CV1 en CV2 geschat op basis van de correct gespecificeerde MLN-functie betere resultaten geven dan MLN op basis van de verkeerd gespecificeerde, Probit-ML-functie en waarschijnlijk ook betere dan de MSCORE-resultaten (tabel 5.5 en 5.6). We hebben deze 'goede' (dat wil zeggen correct gespecificeerde) MLN uitgevoerd. Bij beide varianten zijn daarbij voor elke replicatie de 100 waarnemingen voor  $x_{1t}$  en  $x_{2t}$  ( $t=1,2,\dots,100$ ) gedeeld door de juiste  $h(x_{2t})$ . Vervolgens is de likelihoodfunctie bepaald, analoog aan (2.2). Op deze manier wordt de 'goede' likelihood genomen omdat de vergelijking  $y_t^* = \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t$  uit (5.1) dan wordt gedeeld door  $h(x_{2t})$ . Voor de nieuwe storingsterm  $\eta_t := \frac{\epsilon_t}{h(x_{2t})}$  geldt dan dat:  $\eta_t \sim N(0, 4/3)$ . Er wordt dus wederom een schatting voor  $(\beta_1/\sigma, \beta_2/\sigma) = (\frac{1}{2}\sqrt{3}, \frac{1}{2}\sqrt{3})$  verkregen. De resultaten van deze uitvoering van MLN staan vermeld in tabel 5.7 en tabel 5.8.

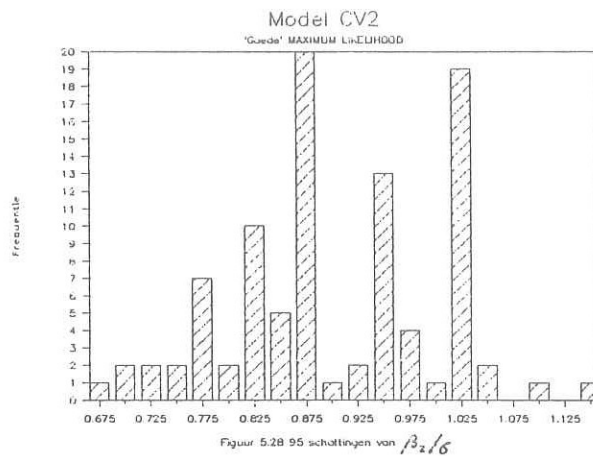
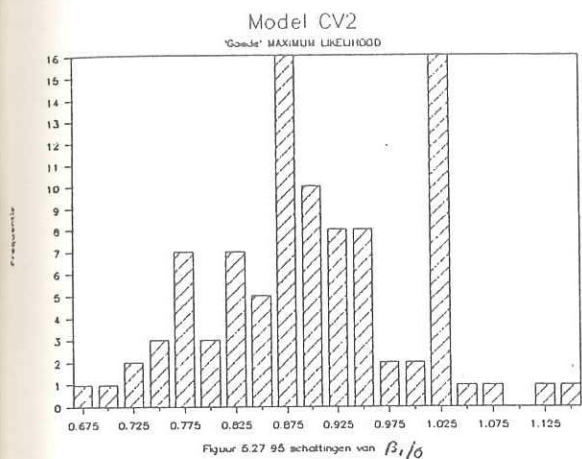
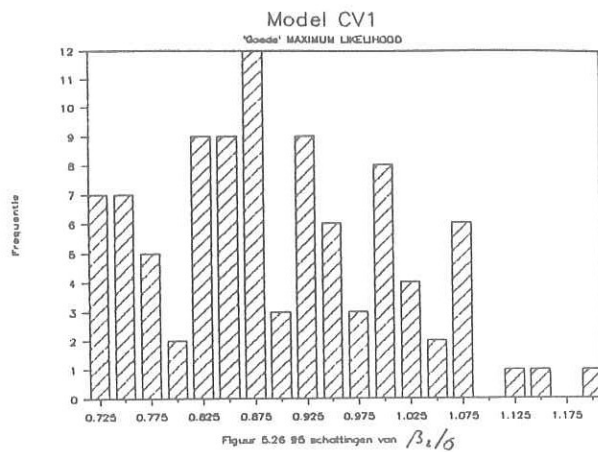
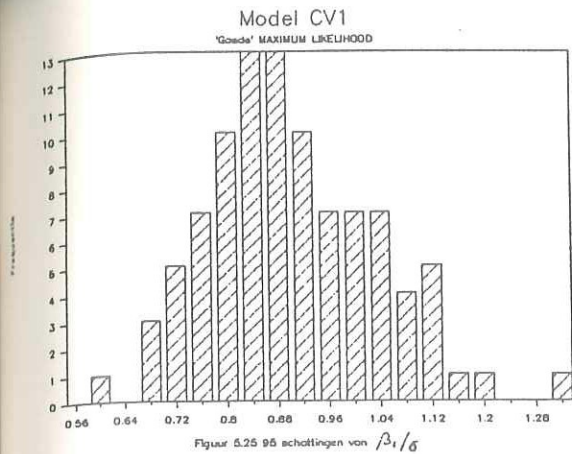
Tabel 5.7 Resultataten model CV1 :  $\epsilon_t | x_{2t} \sim N(0, \frac{4}{3} \exp\{x_{2t} - \frac{6}{5}\})$

	'Goede' MLN
"bias"	-0.0010
SDEV	0.0538
RMSE	0.0539

Tabel 5.8 Resultataten model CV2 :  $\epsilon_t | x_{2t} \sim N(0, \frac{4}{3} |\frac{1}{2} x_{2t} + \frac{1}{2}|^2)$

	'Goede' MLN
"bias"	0.0001
SDEV	0.0195
RMSE	0.0195

Duidelijk is dat deze resultaten inderdaad veel beter zijn dan de MLN-resultaten op basis van de verkeerd gespecificeerde likelihood en de MSCORE-resultaten.





De verdelingen van de schattingen voor de parameters  $\beta_1/\sigma$  en  $\beta_2/\sigma$  zijn uitgezet in de figuren 5.25 tot en met 5.28. Vanwege het feit dat de storingen en exogenen simultaan identiek en onafhankelijk zijn, is de asymptotische verdeling van deze MLN-schatters een normale verdeling. Op de verdeling van de MLN-schattingen voor  $\beta_1/\sigma$  bij model CV1 na, vertonen de verdelingen duidelijk meerdere pieken en geen symmetrie, zodat normaliteit niet waarschijnlijk lijkt. Daarbij dient wel opgemerkt te worden dat de schattingen in vergelijking met de verdelingen onder A en B over een relatief klein interval zijn verdeeld. Door meer waarnemingen te nemen ( $T \rightarrow \infty$ ) zullen de asymptotische normale verdelingen met verwachting  $\frac{1}{2}\sqrt{3} = \pm 0.87$  beter kunnen worden benaderd.

Op basis van deze experimenten kan onder andere worden geconcludeerd dat het maken van een verkeerde veronderstelling ten aanzien van de storingsterm, nl.  $\varepsilon_t \sim U(-2,2)$  in plaats van  $\varepsilon_t \sim N(0,4/3)$ , in kleine steekproeven blijkbaar niet altijd ernstige gevolgen hoeft te hebben; de MLN-schattingsresultaten van model AV2 vertonen geen groot verschil met de MLN-resultaten van model AV1 (zie ook Ruud, 1983). Het schatten van de modellen onder A en B met MSCORE lijkt niet zinvol. Bij model BV2 geschat met MSCORE, is de "bias" weliswaar iets beter dan bij de MLN-schattingen, maar in alle vier modellen zijn de SDEV en de RMSE bij MSCORE minstens twee maal zo groot dan bij MLN. Ondanks het feit dat MLN met uniforme trekkingen inconsistent is in model BV2, blijkt MLN dus toch een betere schattingsmethode te zijn dan MSCORE.

De resultaten onder C tonen aan dat door de heteroskedasticiteit, MSCORE betere resultaten geeft dan MLN. Volgens de theorie en de berekende 'betrouwbaarheidsintervallen' op basis van de centrale limietstelling, zijn de MLN-schattingen inconsistent. De vrij kleine 'betrouwbaarheidsintervallen' bij MSCORE omvatten de ware parameter daarentegen wel; MSCORE lijkt dus duidelijk consistent.

We zijn ons bewust dat het trekken van algemene conclusies op basis van zo weinig experimenten, met grote voorzichtigheid moet gebeuren. We willen ons hier beperken tot deze experimenten en tot de gedane uitspraken, die enkel gebaseerd zijn op de door ons gevonden resultaten. In onze conclusies zullen we terugkomen op de in deze paragraaf gepresenteerde resultaten. Alvorens we

aan deze conclusies toekomen, bespreken we echter nog twee toepassingen van binaire keuzemodellen die wij schatten met behulp van empirische gegevens met MLN en MSCORE.

## § 6 Toepassingen.

In deze paragraaf behandelen we een tweetal toepassingen. Het uitgangspunt is in beide toepassingen een binair keuzemodel dat eerst met probit-ML (genoteerd als MLN) wordt geschat en daarna twee maal met MSCORE: één keer zonder en één keer met opgave van genormeerde MLN-schattingen als startwaarden. Om bootstrap-schattingen voor de standaardafwijking van de MSCORE-schatters te verkrijgen, is de MSE-matrix (4.15) berekend. De wortels uit de diagonaalelementen van deze matrix zijn bootstrap-standaardafwijkingen. Om iets meer te weten te komen over de verdeling van de MSCORE-schatters hebben we een groot aantal bootstrap-steekproeven genomen en de MSCORE-schattingen op basis van deze steekproeven in grafiekjes uitgezet. Met behulp van deze MSCORE-schattingen zijn twee bootstrap-betrouwbaarheidsintervallen, het percentiel interval (4.20) en het BC-interval (4.25), afgeleid. Bovendien wordt bij beide toepassingen de in paragraaf 2 besproken LM-toets op normaliteit uitgevoerd.

Het doel bij deze toepassingen is om te controleren of de veronderstelling dat de storingen een normale verdeling volgen, juist zou kunnen zijn. Dit proberen we te doen door de MLN- met de MSCORE-resultaten te vergelijken en de LM-toets op normaliteit toe te passen. Indien blijkt dat normaliteit ten onrechte wordt verondersteld, is het interessant om te bekijken in hoeverre de MLN-schattingsresultaten tot misleidende conclusies leiden.

### Toepassing 1: Wel/niet werkende vrouwen.

Het deelnemen van vrouwen aan het arbeidsproces wordt in toepassingen vaak verklaard met behulp van een binair keuzemodel uit een aantal demografische variabelen, de opleiding van de vrouw en het overig huishoudinkomen. Wij analyseren een dergelijk model waarbij we gebruik maken van een steekproef van 849 vrouwen, variërend in leeftijd van 18 tot en met 61 jaar. Van elke vrouw is bekend of zij werkt ( $y_t = 1$ ) of niet ( $y_t = 0$ ). We veronderstellen dat het deelnemen aan het arbeidsproces is te verklaren middels een niet waarneembare variabele  $y_t^*$ ; als deze de waarde 0 overschrijdt, neemt een vrouw deel aan het arbeidsproces. Van elke vrouw in de steekproef zijn, naast het feit of zij werkt of niet, waargenomen: de leeftijd, gezinsgrootte, het aantal kinderen jonger dan 6 jaar, de opleiding en het



overig inkomen. De latente variabele verklaren we lineair uit deze vijf variabelen (gerepresenteerd door  $x_{2t}$ ,  $x_{3t}$  ...  $x_{6t}$ ), een constante term ( $x_{1t}$ ) en een storingsterm  $\epsilon_t$ . Bovendien veronderstellen we dat storingsterm  $\epsilon_t$  onafhankelijk is van  $\epsilon_s$  ( $s \neq t$ ). Dit impliceert dat de keuze 'gaan werken of niet' die vrouw  $t$  maakte, niet is beïnvloed door de keuze die vrouw  $s$  ( $s \neq t$ ) heeft gemaakt.

We komen dan tot het volgende model (ook geanalyseerd door Arthur van Soest):

$$\left. \begin{aligned} y_t^* &= \sum_{i=1}^6 \beta_i x_{it} + \epsilon_t \\ y_t &= 1 \text{ als } y_t^* \geq 0 \\ y_t &= 0 \text{ als } y_t^* < 0 \\ \epsilon_t &\sim F_t \\ \epsilon_t &\text{ onafhankelijk van } \epsilon_s \text{ } s \neq t \end{aligned} \right\} \quad t, s = 1, 2 \dots T \quad (6.1)$$

Waargenomen variabelen in het model zijn:

- $y_t$  :  $y_t = 1$  respondente  $t$  is een werkende vrouw  
 $y_t = 0$  respondente  $t$  is een niet werkende vrouw;
- $x_{1t}$  : constante;
- $x_{2t}$  : leeftijd vrouw;
- $x_{3t}$  : gezinsgrootte;
- $x_{4t}$  : aantal kinderen jonger dan 6;
- $x_{5t}$  : opleiding;
- $x_{6t}$  : overig inkomen.

De vrouwen in de steekproef hebben elk minimaal twee kinderen. Te verwachten is dat van de gezinsgrootte een negatieve invloed uitgaat op de latente variabele: de bereidheid om te gaan werken neemt af naarmate het aantal kinderen toeneemt. De aanwezigheid van kinderen onder 6 jaar, zal deze bereidheid nog sterker negatief beïnvloeden. De opleiding van een vrouw is waargenomen op een schaal van 1 t/m 5, representerend een vrij lage opleiding t/m de hoogst mogelijke opleiding. Verwacht mag worden dat een vrouw die een hogere opleiding heeft gehad, eerder zal (blijven) werken dan vrouwen met een vrij lage opleiding. De zesde verklarende variabele, het overig inkomen (dat ook het inkomen van de partner bevat), zal in het

algemeen de bereidheid om te gaan werken niet positief beïnvloeden. Alleen over de invloed van de leeftijd van de vrouw hebben wij tot dusver nog geen uitspraak gedaan. Meestal zullen jongere vrouwen een grotere bereidheid tonen om te gaan werken dan oudere vrouwen, al valt hier nog het een en ander op aan te merken.

Tabel 6.1 geeft de MLN-schattingsresultaten uitgaande van het probitmodel (dat wil zeggen  $F_t = N(0, \sigma^2)$ ), de MSCORE-schattingsresultaten indien het MS-algoritme wordt uitgevoerd zonder opgave van startwaarden (MSCORE 1) en de MSCORE-resultaten met opgave van de genormeerde MLN-schattingen als startwaarden (MSCORE 2). De waarden tussen haakjes geven de geschatte standaardafwijkingen van de coëfficiënten (bij de MSCORE-resultaten zijn dit de bootstrap-standaardafwijkingen). Bij de uitvoeringen van MSCORE zijn 10 bootstrap-steekproeven voor de berekening van de standaardafwijkingen genomen en 5 'endgames' (dat wil zeggen extra iteraties, zie §3 blz. 17) uitgevoerd. Voor het uitvoeren van deze extra iteraties bestaat geen richtlijn. Het aantal 5 is dan ook willekeurig. Om een goede schatting voor de standaardafwijkingen van de schatters te krijgen, zouden volgens Efron (zie § 4) minimaal 50 bootstrap-steekproeven genomen moeten worden. De 10 steekproeven die wij nemen zijn dus eigenlijk te weinig, maar wij komen hier later op terug.

Tabel 6.1 Schattingsresultaten model (6.1).

Variabele	MLN	MSCORE 1	MSCORE 2
constante	1.527 (0.267)	-0.788 (0.959)	0.938 (0.0004)
leeftijd	-0.027 (0.005)	-0.003 (0.388)	-0.020 (0.003)
gezinsgrootte	-0.328 (0.051)	0.029 (0.547)	-0.204 (0.002)
aantal kinderen (<6)	-0.354 (0.067)	-0.475 (0.628)	-0.217 (0.002)
opleiding	0.293 (0.050)	0.389 (0.265)	0.179 (0.002)
overig inkomen	-0.0003 (0.0002)	-0.0005 (0.036)	-0.00004 (0.0003)

De MLN-resultaten voor de coëfficiënten zijn, op de laatste coëfficiënt na, significant. Blijkbaar speelt het overig inkomen, bij de keuze tussen al of

niet gaan werken, geen grote rol. Alle tekens zijn precies zoals verwacht. De sterke negatieve invloeden van de gezinsgrootte en van de aanwezigheid van kinderen beneden 6 jaar, komen bovendien duidelijk naar voren. Deze MLN-schattingen lijken dus vrij aannemelijk. Op basis van deze schattingen kan de kans dat een willekeurige vrouw zich aanbiedt op de arbeidsmarkt, gegeven haar leeftijd, gezinsgrootte etc., worden berekend.

Vergelijken van de MLN- met de MSCORE-resultaten is op deze manier moeilijk omdat het ML-algoritme schattingen geeft voor de parameters gedeeld door de standaardafwijking en MSCORE de genormeerde coëfficiënten schat. Nemen we de normalisatie  $\sigma^2 = 1$  en definiëren we  $\beta' = [\beta_1, \beta_2 \dots \beta_6]$ , dan schat MSCORE:

$$g(\beta) = \begin{bmatrix} \beta_1 / (\sqrt{\sum_{i=1}^6 \beta_i^2}) \\ \vdots \\ \beta_6 / (\sqrt{\sum_{i=1}^6 \beta_i^2}) \end{bmatrix}$$

De functie  $g$  evalueren in  $\hat{\beta}_{ML}$ , geeft dan de transformatie van de MLN-schatter in de MSCORE-schatter  $g(\hat{\beta}_{ML})$ .

Met behulp van de Taylor-ontwikkeling kan de covariantiematrix van  $\hat{\beta}_{ML}$  dan ook worden getransformeerd in de covariantiematrix van  $g(\hat{\beta}_{ML})$ :

$$V\{g(\hat{\beta}_{ML})\} \stackrel{a}{=} \left[ \frac{\partial g(\beta)}{\partial \beta} \right]_{\hat{\beta}_{ML}}' V\{\hat{\beta}_{ML}\} \left[ \frac{\partial g(\beta)}{\partial \beta} \right]_{\hat{\beta}_{ML}}$$

Uit deze matrix kunnen standaardafwijkingen worden afgeleid. Met de MLN-schattingsresultaten uit tabel 6.1 geeft dit:

Tabel 6.2 Getransformeerde MLN - schattingsresultaten.

Variabele	$g(\hat{\beta}_{ML})$	$\text{stdev}\{g(\hat{\beta}_{ML})\}$
constante	0.938	0.022
leeftijd	-0.017	0.003
gezinsgrootte	-0.201	0.036
aantal kinderen (<6)	-0.218	0.050
opleiding	0.180	0.050
overig inkomen	-0.00002	0.0001

De waarden van  $g(\hat{\beta}_{ML})$  zijn de startwaarden die gebruikt zijn bij de tweede MSCORE. Het optimum dat daarbij gevonden is, wijkt weinig af van de



startwaarden. Het optimum bij de eerste MSCORE is geen globaal maximum; de scorefunctie in het startpunt van de tweede MSCORE is namelijk al groter dan de score in het eindpunt bij de eerste MSCORE (0.4770 resp. 0.4464). Bovendien zijn de bootstrap-schattingen voor de standaardafwijkingen bij MSCORE 1 veel groter dan bij MSCORE 2 (Hierbij moet natuurlijk wel rekening worden gehouden met het feit dat er eigenlijk te weinig bootstrap-steekproeven genomen zijn).

Om deze redenen laten we MSCORE 1 derhalve buiten beschouwing.

De MSCORE-resultaten lijken weinig af te wijken van de MLN-resultaten. Berekenen van zogenaamde 'betrouwbaarheidsintervallen' voor de parameters door als afwijking van de getransformeerde MLN-schattingen de standaardfouten uit tabel 6.2 te vermenigvuldigen met 2 (ongeveer het 0.95-kwantiel van de normale verdeling), bevestigt dit. Veronderstellen we dat  $\sigma^2 = 1$ , dan zijn de MLN-resultaten schattingen voor  $\beta_1$  t/m  $\beta_6$ . De 'betrouwbaarheidsintervallen' voor  $\beta_1$  t/m  $\beta_6$  worden resp.:

(0.894;0.982)	'BI' voor $\beta_1$
(-0.023;-0.011)	'BI' voor $\beta_2$
(-0.273;-0.129)	'BI' voor $\beta_3$
(-0.318;-0.118)	'BI' voor $\beta_4$
(0.08;0.28)	'BI' voor $\beta_5$
(-0.00022;0.00018)	'BI' voor $\beta_6$

De MSCORE-resultaten (tabel 6.1, MSCORE 2) vallen, op de (met MLN niet significant geschatte) zesde parameter na, binnen de intervallen. Er zou daarom verwacht kunnen worden dat de veronderstelling  $F_t = N(0, \sigma^2)$  bij benadering juist is. Berekenen van de toetsingsgrootte  $\xi_{LM}$  (zie 2.13) van de LM-toets voor standaardnormaliteit levert echter:

$$\xi_{LM} = 849 * 0.0233 = 19.7817$$

Omdat  $\chi^2_{2(0.05)} = 5.99$  wordt de  $H_0$ -hypothese ' $\epsilon_t \sim N(0,1)$ ' dus duidelijk verworpen. Gezien de resultaten is dit heel vreemd; schatten met probit-ML lijkt wel consistent (de genormaliseerde MLN-resultaten benaderen de MSCORE resultaten), terwijl volgens de LM-toets de verdeling van de storingsterm geen normale verdeling is. Een mogelijke verklaring hiervoor zou kunnen zijn dat het model aan bepaalde voorwaarden voldoet, zodat schatten met MLN

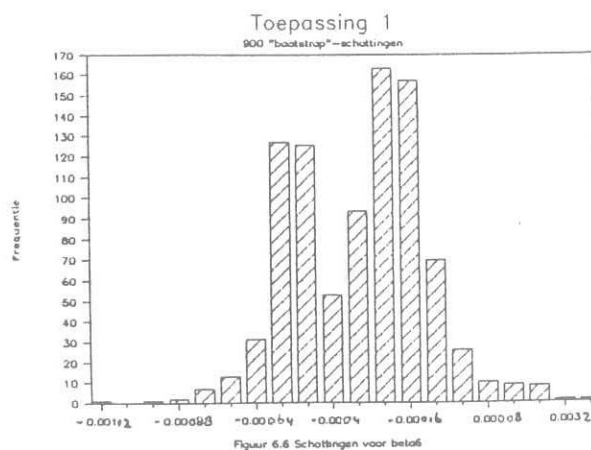
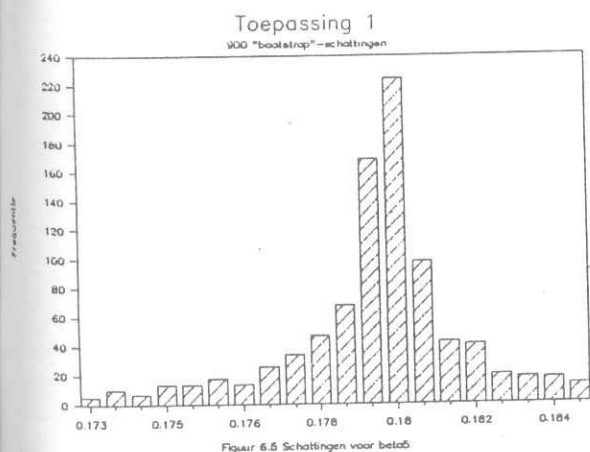
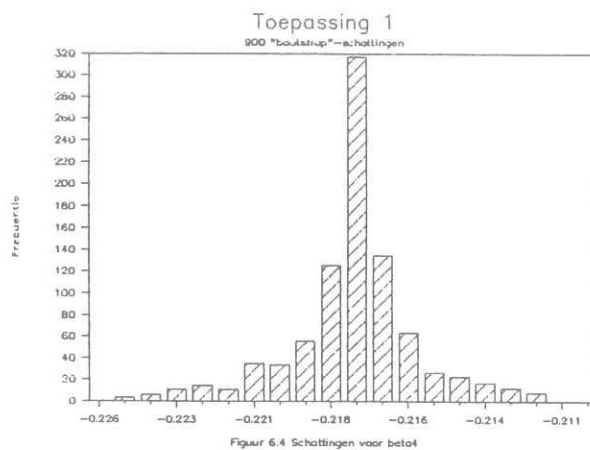
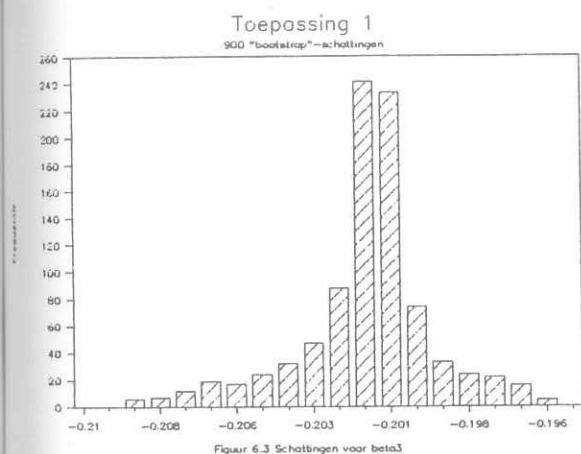
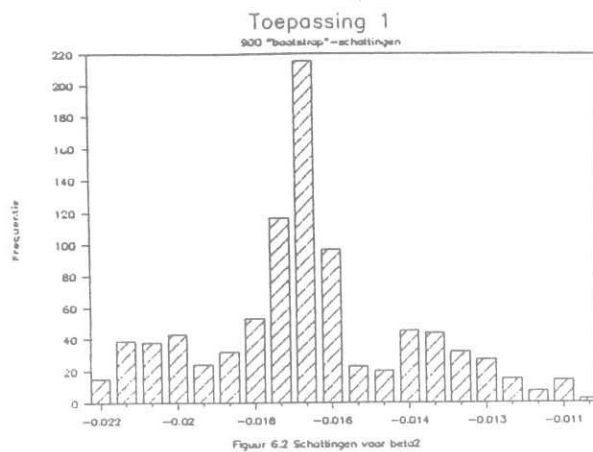
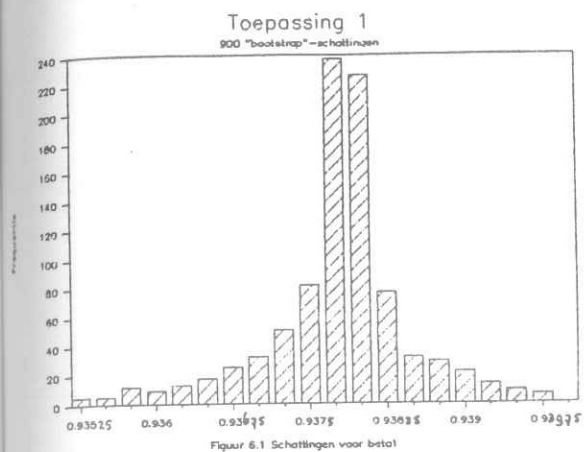
consistent is ondanks misspecificatie van de verdeling van de storingsterm (analoog aan bijv. Ruud, 1983).

Om iets te weten te komen van de verdeling van de MSCORE-schatters hebben we 1000 bootstrap-steekproeven uit de oorspronkelijke steekproef genomen. Voor elke bootstrap-steekproef zijn MSCORE-schattingen op basis van die steekproef berekend. De schattingen voor  $\beta_1$  t/m  $\beta_6$  zijn uitgezet in resp. figuur 6.1 t/m 6.6. Daarbij is een opdeling gemaakt in 20 intervallen. De waarden op de horizontale as geven de middelpunten van deze intervallen aan. We hebben daarbij de 50 extreme schattingen aan beide staarten van de verdeling weggelaten zodat een beter beeld van de verdelingen wordt verkregen. Een verdeling over 20 intervallen van alle 1000 schattingen gaf nl. op de schattingsresultaten van  $\beta_2$  en  $\beta_6$  na, alleen duidelijk een piek bij het interval met de puntschatting uit tabel 6.1. Opvallend bij de verdelingen van de coëfficiënten is dat deze afgekapte verdelingen, op de verdelingen van de schattingen voor de tweede, derde en zesde parameter na, symmetrisch zijn om de puntschattingen (resp. 0.938, -0.217 en 0.179 voor  $\beta_1$ ,  $\beta_4$  en  $\beta_5$ ).

Omdat volgens Efron een goede benadering van de standaardafwijkingen van de schatters pas verkregen kan worden op basis van minimaal 50 bootstrap-steekproeven, geven we in tabel 6.3 voor MSCORE 2 bootstrap-standaardafwijkingen behorend bij de MSCORE-schattingen voor  $\beta_1$  t/m  $\beta_6$  op basis van de zojuist genoemde 1000 bootstrap-steekproeven.

Tabel 6.3. Bootstrap-standaardafwijkingen van de MSCORE-schatters voor  $\beta_1$  t/m  $\beta_6$  o.b.v. 1000 bootstrap-steekproeven.

Parameter	MSCORE-schatting	Bootstrap-standaardafwijking
$\beta_1$	0.938	0.004
$\beta_2$	-0.020	0.005
$\beta_3$	-0.204	0.007
$\beta_4$	-0.217	0.008
$\beta_5$	0.179	0.008
$\beta_6$	-0.00004	0.0004





Gegeven deze bootstrap-standaardafwijkingen zouden eenvoudig standaard-betrouwbaarheidsintervallen voor  $\beta_1$  t/m  $\beta_6$  kunnen worden berekend (zie 4.16). Omdat echter bekend is dat de MSCORE-schatter asymptotisch geen normale verdeling volgt, is dit weinig zinvol. Daarom leiden we, op basis van de bootstrap-verdelingen voor de componenten van de MSCORE-schatter (dat wil zeggen de 1000 schattingen voor elke parameter), alleen betrouwbaarheidsintervallen af voor  $\beta_1$  t/m  $\beta_6$  met behulp van de percentiel- (PC) en de BC-methode. De eindpunten van de betrouwbaarheidsintervallen en de puntschattingen zijn weergegeven in tabel 6.4.

Tabel 6.4 95%-Percentiel- en 95%-BC-betrouwbaarheidsintervallen voor  $\beta_1$  t/m  $\beta_6$ .

Coëff.	Puntschat.	PC-ondergrens	PC-bovengrens	BC-ondergrens	BC-bovengrens
$\beta_1$	0.93768	0.93352	0.94065	0.90955	0.93859
$\beta_2$	-0.02049	-0.02422	-0.00902	-0.03474	-0.01726
$\beta_3$	-0.20364	-0.21592	-0.19276	-0.29015	-0.20129
$\beta_4$	-0.21668	-0.22880	-0.20972	-0.21816	-0.16233
$\beta_5$	0.17865	0.16690	0.18732	0.09162	0.18038
$\beta_6$	-0.00004	-0.00073	0.00018	-0.00017	0.00058

Als er een monotone transformatie van de MSCORE-schatter bestaat die normaal is verdeeld, zijn de BC-intervallen in elk geval beter dan de percentiel-intervallen; de "bias" ( $z_0\tau$  in 4.24) is namelijk ongelijk aan 0 bij de verdelingen van de MSCORE-schatters die niet symmetrisch zijn om de puntschatting (zie figuur 6.1 t/m 6.6). Het is niet eenvoudig om te achterhalen of genoemde monotone transformatie van de MSCORE-schatter bestaat. We weten derhalve ook niet welke waarde we mogen hechten aan deze intervallen. Gaan we uit van de juistheid van bovenstaande betrouwbaarheidsintervallen, dan zien we dat enkel de parameter-schatting voor  $\beta_6$  niet 'significant' is. De MLN-schattingen toonden een zelfde resultaat.

In deze toepassing hebben we gezien dat de MLN- en MSCORE-schattingen geen grote verschillen vertoonden. Indien de MSCORE-schattingen consistent zijn, lijken (ondanks de mogelijke misspecificatie van de verdeling van de

storingsterm) de MLN-schattingen ook consistent. Bovendien blijkt uit toetsen bij beide methoden dat dezelfde parameterschatting niet significant is. Schatten met MLN leidt in deze toepassing dus niet tot andere conclusies dan schatten met MSCORE.

## Toepassing 2 : "Sample selectivity".

In empirisch werk dat wordt gebaseerd op micro-gegevens, worden meestal waarnemingen weggelaten omdat ze onbruikbaar zijn; vaak ontbreken er voor de analyse belangrijke gegevens of blijken bepaalde gegevens onjuist. Er wordt dan verondersteld dat de steekproef na weglating van deze waarnemingen, nog steeds aselekt is. Met behulp van een zogenaamde selectie-vergelijking, kan dit worden gecontroleerd. Het al of niet opgenomen zijn van de waarnemingen uit de oorspronkelijke steekproef in de overgebleven, 'geschoonde' steekproef, tracht men te verklaren uit een aantal exogenen door middel van een latente variabele. Als de geschoonde steekproef niet aselekt is, is het mogelijk dat schatters gebruikt in het model dat gebaseerd is op deze selecte steekproef, niet langer consistent zijn. We lichten dit toe aan de hand van een vergelijking uit het model over "preference formation" van Kapteyn et al. (1980). In het model wordt een relatie gelegd tussen het inkomensniveau  $z$  en de 'waarde' die aan dat inkomen gehecht wordt, genoteerd als  $U(z)$ . Er wordt verondersteld dat de functie  $U(z)$ , die ook wel 'welzijnsfunctie' wordt genoemd, een lognormale verdeling  $\wedge(z, \mu, \sigma)$  volgt (met parameters  $\mu$  en  $\sigma$ ). De vergelijking die wij uit het model lichten is de lineaire verklaring van de zogenaamde 'welzijnsparameter'  $\mu$  van respondent  $t$  uit de logaritme van het inkomen, de logaritme van een functie van de leeftijd, een referentiegroep-gemiddelde inkomen en een referentiegroep-gemiddelde van de functie van de leeftijd van respondent  $t$ :

$$\mu_t = \sum_{i=1}^5 \alpha_i x_{it} + \varepsilon_t \quad t=1,2 \dots T \quad (6.2)$$

Waargenomen variabelen zijn:

- $\mu_t$  : 'welzijnsparameter' respondent  $t$ ;
- $x_{1t}$  : constante;
- $x_{2t}$  : log inkomen;
- $x_{3t}$  : log leeftijdsfunctie;

$x_{4t}$ : referentiegroep-gemiddelde inkomen;

$x_{5t}$ : referentiegroep-gemiddelde leeftijdsfunctie.

Een oorspronkelijke steekproef op basis waarvan het model geschat zou kunnen worden, bestond uit 1733 observaties. Na weglating van, voor de analyse niet bruikbare waarnemingen, bleven er nog 662 observaties over (dus  $T = 662$ ). We noteren  $y_t = 1$  als de  $t^e$  respondent in de geschoonde steekproef voorkomt en  $y_t = 0$  als dit niet zo is. Er wordt verondersteld dat respondent  $t$  is opgenomen in deze steekproef als latente variabele  $y_t^*$  groter of gelijk aan 0 is. Bovendien wordt er van uitgegaan dat de bruikbaarheid van waarneming  $t$  uit de oorspronkelijke steekproef middels  $y_t^*$  lineair te verklaren is uit een constante term ( $z_{1t}$ ), twee transformaties van de leeftijd ( $z_{2t}, z_{3t}$ ), het beroep van de respondent ( $z_{4t}, z_{5t}$ ) en een onafhankelijke storingsterm ( $\delta_t$ ). De selectievergelijking kan dan als volgt worden weergegeven:

$$\left. \begin{aligned} y_t^* &= \sum_{i=1}^5 \beta_i z_{it} + \delta_t \\ y_t &= 1 \text{ als } y_t^* \geq 0 \\ y_t &= 0 \text{ als } y_t^* < 0 \\ \delta_t &\text{ onafhankelijk van } \delta_s \quad s \neq t \\ \delta_t &\sim F_t \end{aligned} \right\} \quad t, s = 1, 2 \dots 1733 \quad (6.3)$$

Waargenomen variabelen zijn:

$y_t$ :  $y_t = 1$  als respondent  $t$  is opgenomen in de geschoonde steekproef  
 $y_t = 0$  als respondent  $t$  niet is opgenomen in de geschoonde steekproef;

$z_{1t}$ : constante;

$z_{2t}$ : log leeftijd respondent  $t$ ;

$z_{3t}$ : gekwadrateerde log leeftijd respondent  $t$ ;

$z_{4t}$ : dummy = 1 als respondent  $t$  werknemer is  
 dummy = 0 anders;

$z_{5t}$ : dummy = 1 als respondent  $t$  een hogere functie in een bedrijf heeft  
 dummy = 0 anders.

De consistentie van de gebruikelijke schatters (zoals de OLS-schatters) voor  $\alpha_1$  t/m  $\alpha_5$  hangt af van de eigenschappen van de storingsterm  $\varepsilon_t$ . In het



algemeen zal het voor consistentie noodzakelijk zijn dat  $E\{\epsilon_t\} = 0$ . Deze verwachting van storingsterm  $\epsilon_t$  moet echter worden berekend, gegeven het feit dat de observatie in de geschoonde steekproef zit:

$$E\{\epsilon_t | y_t^* \geq 0\} = E\{\epsilon_t | \delta_t \geq - \sum_{i=1}^5 \beta_i z_{it}\}. \quad (6.4)$$

Als de geschoonde steekproef niet aselekt is, dat wil zeggen als  $\epsilon_t$  en  $\delta_t$  niet onafhankelijk zijn, zal deze conditionele verwachting ongelijk aan 0 zijn en zullen dus de parameters  $\alpha_1$  t/m  $\alpha_5$  in het algemeen inconsistent worden geschat. Als de simultane verdeling van  $\epsilon_t$  en  $\delta_t$  bekend is, is (6.4) te berekenen en kan met Heckman's methode (Maddala (1983), blz 221 e.v.) worden gecorrigeerd voor een dergelijke "selectivity bias". Wij beperken ons tot de eerste stap van die correctieprocedure die bestaat uit het schatten van het binaire keuzemodel (6.3).

Wij hebben dit gedaan met MLN, MSCORE zonder opgave van startwaarden (MSCORE 1) en MSCORE met opgave van de genormeerde MLN-schattingresultaten als startwaarden (MSCORE 2). De schattingresultaten staan vermeld in tabel 6.5. De waarden tussen haakjes geven de geschatte standaardafwijkingen. Bij de uitvoering van MSCORE zijn, net als in toepassing 1, 5 extra iteraties uitgevoerd en 10 bootstrap-steekproeven genomen.

Tabel 6.5 Schattingresultaten model (6.3).

Variabele	MLN	MSCORE 1	MSCORE 2
constante	-11.365 (3.816)	-0.089 (0.509)	-0.891 (0.005)
log leeftijd	5.684 (2.049)	0.154 (0.492)	0.446 (0.007)
(log leeftijd) <sup>2</sup>	-0.745 (0.273)	-0.093 (0.141)	-0.059 (0.003)
dummy werknemer	0.618 (0.078)	0.920 (0.444)	0.048 (0.016)
dummy hogere functie	0.352 (0.235)	-0.338 (0.659)	0.028 (0.012)

Volgens de MLN-schattingen zou de kans dat een respondent wordt verwijderd uit de steekproef, kleiner zijn als de respondent een werknemer is dan wanneer deze een hogere functie in een bedrijf heeft. Bovendien stijgt deze kans naarmate de respondent jonger is. Dit laatste spreekt eerder gevonden

resultaten tegen (zie bijv. Kapteyn et al. (1988)). Daarbij moet bovendien worden opgemerkt dat wij significante schattingen vinden, op de dummy-variabele voor de hogere functie na, terwijl in het genoemde artikel (waarin o.a. vergelijking (6.2) en (6.3) simultaan worden geschat) enkel deze dummy-variabele significant is.

Vergelijken van de MLN- en MSCORE-resultaten wordt weer mogelijk door  $g(\beta_{ML})$  en  $stdev\{g(\beta_{ML})\}$  te berekenen (zie toepassing 1, tabel 6.2).

Tabel 6.6 Getransformeerde MLN - schattingsresultaten.

Variabele	$g(\beta_{ML})$	$stdev\{g(\beta_{ML})\}$
$z_1$	-0.891	0.005
$z_2$	0.446	0.012
$z_3$	-0.058	0.003
$z_4$	0.048	0.019
$z_5$	0.028	0.022

Ook hier is  $g(\beta_{ML})$  als startwaarde gebruikt voor de tweede MSCORE (tabel 6.5). Het eindpunt wijkt hierbij niet veel af van dit startpunt, terwijl het eindpunt bij MSCORE 1 een geheel andere schatting voor  $\beta$  geeft. Het negatieve teken in MSCORE 1 bij de vijfde (bij MLN niet significante) coëfficiënt lijkt vrij aannemelijk: respondenten met hogere functies tonen wel eens vaker minder bereidheid om vragen met betrekking tot inkomen e.d. te beantwoorden. De kans op verwijdering uit de steekproef zal derhalve groter zijn.

In tegenstelling tot toepassing 1 geldt hier dat het MSCORE-algoritme met opgave van startwaarden in elk geval geen globaal maximum bereikt; de maximum score in het eindpunt is 0.2626 terwijl deze in het eindpunt van MSCORE 1 0.2718 is. Omdat de scores zo dicht bij elkaar liggen, hebben we beide MSCORE-algoritmes nog eens uitgevoerd met opgave van een groot aantal, namelijk 100, extra iteraties (in plaats van 5). Bovendien hebben we bootstrap-standaardafwijkingen op basis van de, door Efron geadviseerde, 50 bootstrap-steekproeven genomen. Bij het berekenen van de MSCORE-schattingen op basis van de bootstrap-steekproeven zijn wel slechts 5 extra iteraties uitgevoerd. De resultaten staan vermeld in tabel 6.7; MSCORE 1 is de uitvoering van MSCORE zonder opgave van startwaarden en MSCORE 2 is de

uitvoering met als startwaarden de genormeerde MLN-schattingen. De waarden tussen haakjes zijn de geschatte bootstrap-standaardafwijkingen.

Tabel 6.7 Schattingsresultaten model (6.3).

Variabele	MSCORE 1	MSCORE 2
constante	-0.089 (0.519)	-0.889 (0.410)
log leeftijd	0.154 (0.441)	0.451 (0.193)
(log leeftijd) <sup>2</sup>	-0.093 (0.129)	-0.060 (0.024)
dummy werknemer	0.920 (0.459)	0.046 (0.095)
dummy hogere functie	-0.338 (0.545)	0.031 (0.077)

Door het toevoegen van extra iteraties is de score bij MSCORE 1 (het MS-algoritme uitgevoerd zonder opgave van startwaarden) niet veranderd; de resultaten zijn hetzelfde gebleven. De score bij MSCORE 2 is daarentegen gestegen van 0.2626 tot 0.2833 en daarmee net boven de score van MSCORE 1 uitgekomen. De bootstrap-standaardafwijkingen bij MSCORE 2 zijn allen kleiner dan bij MSCORE 1. Op grond van deze resultaten concluderen wij dat het redelijk aannemelijk is dat schatten op basis van de genormeerde MLN-schattingen beter is.

Een dergelijk probleem speelde in toepassing 1 niet omdat daar meteen duidelijk was dat de schatting op basis van de MLN-resultaten veel beter was dan de schatting gebaseerd op de door het MSCORE-algoritme zelf bepaalde startwaarden.

De MSCORE-resultaten wijken, net als in toepassing 1, weinig af van de getransformeerde MLN-resultaten. Berekenen van 'betrouwbaarheidsintervallen' voor de vijf parameters met behulp van de standaardafwijkingen en de parameterschattingen uit tabel 6.6, geeft als resultaat:

(-0.901;-0.881)	'BI' voor $\beta_1$
(0.422;0.470)	'BI' voor $\beta_2$
(-0.064;-0.052)	'BI' voor $\beta_3$
(0.01;0.086)	'BI' voor $\beta_4$



(-0.016;0.072) 'BI' voor  $\beta_5$

Er zou dus kunnen worden verwacht dat de LM-toets normaliteit van de storingsterm niet zal verwerpen; de MSCORE-resultaten vallen immers duidelijk in deze intervallen. De toetsingsgrootte van de LM-toets is:

$$\xi_{LM} = 1733 * 0.0066 = 11.4378$$

De  $H_0$ -hypothese: ' $\epsilon_t \sim N(0,1)$ ' wordt dus, net als in toepassing 1, verworpen. Ook hier lijkt het verwerpen van normaliteit op basis van de resultaten van schatten met MLN en schatten met MSCORE (met opgave genormeerde MLN-schattingen) niet aannemelijk.

In de figuren 6.7 tot en met 6.11 zijn MSCORE-schattingen voor respectievelijk  $\beta_1$  tot en met  $\beta_5$  op basis van 500 bootstrap-steekproeven uit de oorspronkelijke steekproef uitgezet. We hebben slechts 500 steekproeven in plaats van de door Efron geadviseerde 1000 bootstrap-steekproeven genomen omdat uitvoering van het MSCORE-algoritme vrij tijdrovend is vanwege de grote steekproef. Bij het berekenen van de MSCORE-schattingen zijn als startwaarden de genormeerde MLN-schattingen genomen. Wij hebben 50 uitschieters (25 naar boven en 25 naar beneden) geëlimineerd. De verdelingen van de schattingen om de puntschattingen zijn weliswaar niet symmetrisch, maar nemen wel af naarmate de afwijking ten opzichte van de puntschatting toeneemt. Berekenen van percentiel- en BC-betrouwbaarheidsintervallen op basis van deze verdelingen geeft de volgende resultaten:

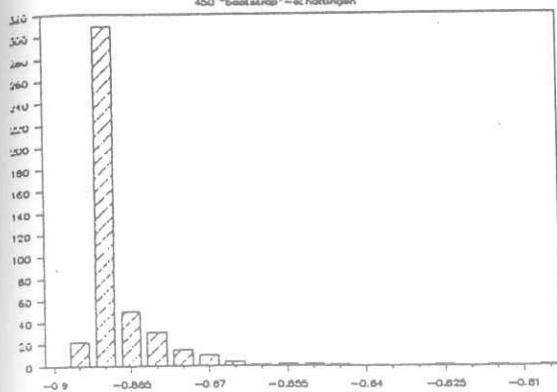
Tabel 6.8 95%-Percentiel- en 95%-BC-betrouwbaarheidsintervallen voor  $\beta_1$  t/m  $\beta_5$ .

Coëff.	Puntschat.	PC-ondergrens	PC-bovengrens	BC-ondergrens	BC-bovengrens
$\beta_1$	-0.88871	-0.89583	0.16096	-0.90868	-0.87760
$\beta_2$	0.45113	-0.00672	0.49129	-0.50831	0.46768
$\beta_3$	-0.06019	-0.09324	-0.04548	-0.06754	0.07090
$\beta_4$	0.04570	0.02129	0.41965	-0.13700	0.06585
$\beta_5$	0.03103	-0.08963	0.09425	0.00023	0.30048

Uitgaan van de percentiel-methode zou heel andere conclusies opleveren dan uitgaan van de BC-methode. Omdat duidelijk is dat de verdelingen van de

### Toepassing 2

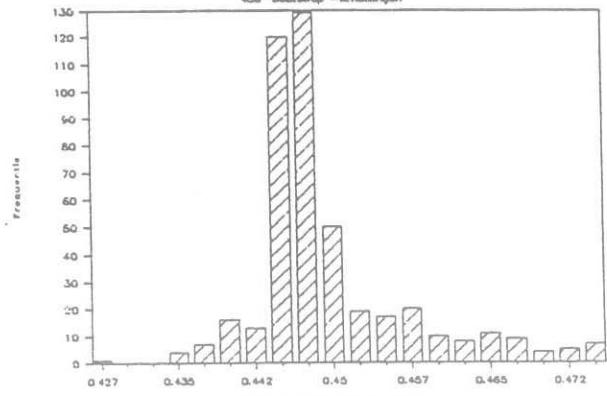
450 "bootstrapping"-schattingen



Figuur 6.7 Schattingen voor beta1

### Toepassing 2

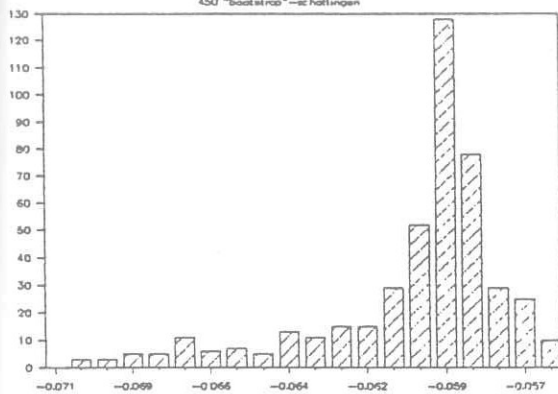
450 "bootstrapping"-schattingen



Figuur 6.8 Schattingen voor beta2

### Toepassing 2

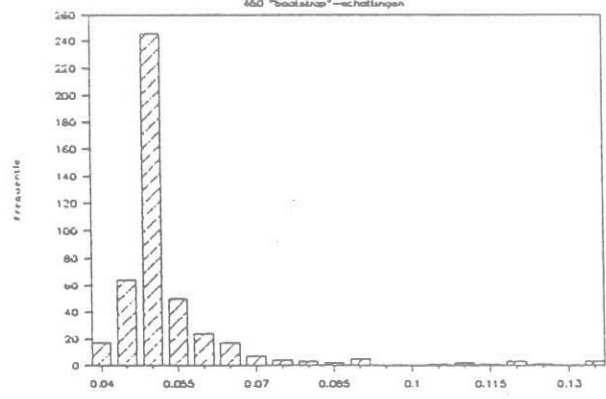
450 "bootstrapping"-schattingen



Figuur 6.9 Schattingen voor beta3

### Toepassing 2

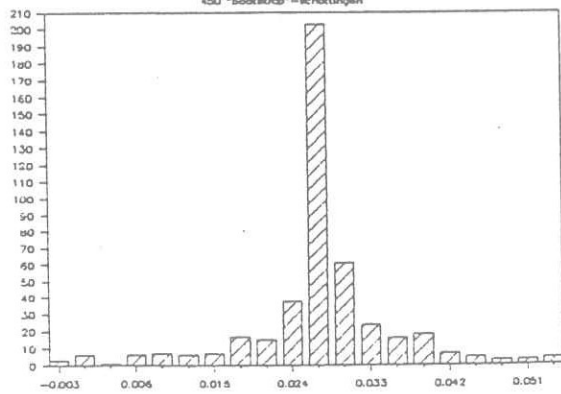
450 "bootstrapping"-schattingen



Figuur 6.10 Schattingen voor beta4

### Toepassing 2

450 "bootstrapping"-schattingen



Figuur 6.11 Schattingen voor beta5

MSCORE-schattingen niet symmetrisch zijn (en derhalve de percentielintervallen afwijken van de BC-intervallen), is de BC-methode in elk geval beter dan de percentiel-methode als er tenminste een normaal verdeelde monotone transformatie van de MSCORE-schatter bestaat. Volgens de BC-resultaten zijn drie verklarende variabelen 'insignificant'. Het hebben van een hogere functie in een bedrijf heeft daarentegen wel invloed op het al of niet opgenomen worden in de geschoonde steekproef. Deze resultaten spreken de MLN-schattingen tegen maar sluiten beter aan bij eerder gevonden resultaten in Kapteyn et al. (1988). De berekende betrouwbaarheidsintervallen kunnen enigszins verklaren waarom MSCORE 1 en MSCORE 2 scores opleveren die zo dicht bij elkaar liggen en toch zo verschillende parameterschattingen geven. Voor deze steekproef zouden volgens de MSCORE-resultaten immers hele andere variabelen moeten worden genomen voor de verklaring van het al of niet worden opgenomen in de geschoonde steekproef. Het feit dat de LM-toets normaliteit verwerpt, lijkt gezien deze resultaten dan ook minder onaannemelijk.

Natuurlijk is het noodzakelijk om betrouwbaarheidsintervallen die rekening houden met de scheefheid van de verdeling van de MSCORE-schatters (zoals bijv.  $BC_a$ -intervallen) e.d. te berekenen alvorens echte conclusies kunnen worden getrokken met betrekking tot dit model. Wij beperken ons echter tot deze methoden en zullen nader ingaan op conclusies die zijn gebaseerd op onder andere deze toepassingen.



## §7 Conclusies.

In deze scriptie hebben we een semi-parametrische schattingsmethode voor het binaire keuzemodel besproken en vergeleken met een veel gebruikte methode voor dit model, nl. Probit Maximum Likelihood. De Probit-Maximum-Likelihood-methode heeft enkele 'prettige' eigenschappen, maar is gebaseerd op de vrij sterke veronderstellingen dat de storingen onafhankelijk, normaal verdeeld en homoskedastisch zijn. Vaak wordt in toepassingen deze methode gebruikt zonder rekening te houden met de mogelijkheid dat de genoemde veronderstellingen niet juist kunnen zijn en daardoor eigenschappen als consistentie, asymptotische efficiëntie en asymptotische normaliteit niet zonder meer gelden.

Bij enige onzekerheid omtrent de verdelingsveronderstelling van de storingsterm in binaire keuzemodellen en in geval van grote steekproeven zou de semi-parametrische schattingsmethode, Maximum Score van Manski, geadviseerd kunnen worden. De methode is consistent en stelt alleen een zwakke eis aan de mediaan van de storingsterm. Met betrekking tot deze consistentie moet wel opgemerkt worden dat de Maximum-Score-schatter vrij langzaam convergeert naar de ware parameter ( $T^{1/3}$ -consistentie in plaats van  $\sqrt{T}$ -consistentie zoals bij Maximum-Likelihood-schatters), zodat grotere steekproeven nodig zijn dan bij Maximum Likelihood om een goede puntschatting voor de parameter te verkrijgen. Een ander, niet onbelangrijk, verschil tussen beide methoden betreft de vorm van de criteriumfunctie: bij Maximum Likelihood is de likelihoodfunctie concaaf zodat altijd het globale maximum wordt gevonden. De scorefunctie zal daarentegen slechts in modellen met twee onbekende parameters met zekerheid het globale maximum bereiken.

Maximum Score is een algemeen bruikbare methode, maar heeft ook enkele beperkingen. Zo is het voor de identificatie van de parameters van belang, dat (in geval van onafhankelijke exogenen) minstens één van de verklarende variabelen continu is. Bovendien ontbreekt bij de Maximum-Score-methode een bruikbare asymptotische verdeling van de schatters zodat standaardafwijkingen, toetsen en betrouwbaarheidsintervallen, zoals bij Maximum Likelihood, niet op deze manier zijn af te leiden. Om enige uitspraak over de nauwkeurigheid van de Maximum-Score-schatters te doen, kunnen "bootstrap"-methoden worden gebruikt. Deze methoden zijn niet

eenduidig, maar wel algemeen; ze zouden bijv. ook kunnen worden gebruikt voor het berekenen van standaardafwijkingen e.d. van de Maximum-Likelihood-schatters in kleine steekproeven. Bekend is dat de beste resultaten met behulp van "bootstrapping" echter worden verkregen in grote steekproeven. Helaas is nog niet helemaal duidelijk welke waarde precies aan "bootstrap"-methoden mag worden gehecht.

Om te onderzoeken of de Maximum-Score-methode in kleine steekproeven ook een betere schattingsmethode is dan Probit-Maximum-Likelihood als de verdeling van de storingsterm geen normale verdeling is, hebben we enkele Monte-Carlo-experimenten uitgevoerd. Daaruit bleek onder andere dat bij een model met uniform verdeelde, homoskedastische storingen de resultaten van Probit Maximum Likelihood veel beter waren dan de Maximum-Score-resultaten. Bij gesimuleerde modellen met heteroskedastische storingen kwam Maximum Score daarentegen duidelijk als de betere methode naar voren. Deze resultaten komen overeen met een uitgebreid onderzoek van Manski en Thompson (1986) tussen Maximum Score en Logit Maximum Likelihood. In het genoemde onderzoek blijkt dat Maximum Score niet beter is dan Maximum Likelihood bij verkeerd gespecificeerde, homoskedastische storingen en dat het omgekeerde geldt als de storingen heteroskedastisch zijn.

Een belangrijk punt dat ook bij onze Monte-Carlo-experimenten naar voren kwam, was het feit dat bij homoskedasticiteit een misspecificatie van de verdelingsfunctie in bepaalde gevallen helemaal geen gevolgen heeft voor de consistentie van de Maximum-Likelihood-schatters (Ruud, 1983). In deze gevallen en vanzelfsprekend ook in de gevallen dat er geen sprake is van misspecificatie, is Maximum Likelihood altijd een betere schattingsmethode dan Maximum Score.

De twee binaire keuzemodellen met betrekking tot de participatie van vrouwen op de arbeidsmarkt (toepassing 1) en "sample-selection" (toepassing 2), met empirische gegevens geschat met Maximum Likelihood en Maximum Score, lieten geen significant grote verschillen in schattingsresultaten zien. Bij toepassing 1 zou om deze reden en de berekende "bootstrap"-betrouwbaarheidsintervallen ook niet worden verwacht dat in het probitmodel een verkeerde verdelingsveronderstelling wordt gemaakt. Enigszins verbazingwekkend is daarom toch het feit dat de Lagrange-Multiplier-toets de

normaliteit van de storingsterm verworpt. Een plausibele verklaring hiervoor is moeilijk te vinden. Het is mogelijk dat het model aan bepaalde voorwaarden voldoet, waardoor schatten met Probit Maximum Likelihood consistent is, ondanks een misspecificatie van de verdeling van de storingsterm.

Bij de tweede toepassing kwamen de schattingsresultaten van beide methoden weliswaar ook vrij sterk overeen, maar toetsen bij beide methoden gaven zeer uiteenlopende resultaten. Het verwerpen van normaliteit in deze toepassing zou daarom iets aannemelijker kunnen zijn. De voorkeur bij schatten van dit model zou op basis van de resultaten dan ook beter aan Maximum Score kunnen worden gegeven. Indien normaliteit inderdaad niet geldt, zijn de standaardfouten bij Probit Maximum Likelihood (en daarmee de berekende betrouwbaarheidsintervallen) ook niet correct, waardoor verkeerde conclusies kunnen worden getrokken.

Bij de bespreking van de Maximum-Score-methode en de vergelijking met Probit Maximum Likelihood, hebben we ons op bepaalde punten beperkt tot eenvoudigere gevallen dan het algemene geval. Zo hebben we ons bij Maximum Score slechts bezig gehouden met de mediaan in plaats van een willekeurig kwantiel van de storingsterm. Bij de behandeling van de theoretische achtergrond van de Maximum-Score- en de Maximum-Likelihood-methode zijn wij ook niet nader ingegaan op allerlei aspecten die mogelijk meer beperkingen stellen aan het gebruik van beide methoden. Het vergelijken van Maximum Score met Probit Maximum Likelihood in kleine steekproeven hebben we gedaan door een heel beperkt aantal modellen met slechts twee verklarende variabelen te simuleren. Monte-Carlo-experimenten zouden veel meer informatie kunnen geven door factoren zoals de steekproefgrootte, aantal replicaties, verdelingsveronderstellingen etc. te laten variëren. Op basis van dergelijke experimenten zouden algemene conclusies met betrekking tot Maximum Score in vergelijking met een Maximum-Likelihood-methode kunnen worden getrokken.

Met behulp van Monte-Carlo-experimenten zou de betrouwbaarheid van de behandelde "bootstrap"-methoden ook kunnen worden onderzocht. Zo kan bijv. de "bootstrap"-methode voor standaardfouten van Maximum-Likelihood-schatters in grote steekproeven worden vergeleken met berekende standaardfouten op

basis van de informatiematrix. Na dergelijke experimenten kan pas worden geconcludeerd welke waarde aan "bootstrap"-methoden kan worden gehecht.

Om het schatten met Maximum Likelihood en Maximum Score in toepassingen te vergelijken, zou het wellicht beter zijn geweest als we modellen hadden kunnen kiezen die op voorhand met grote waarschijnlijkheid verkeerd zijn gespecificeerd door een probitmodel. De Lagrange-Multiplier-toets op normaliteit zou de  $H_0$ -hypothese zeker moeten verwerpen en de verschillen tussen de Maximum-Likelihood- en de Maximum-Score-parameterschattingen zouden dan duidelijk significant moeten zijn. Met de twee toepassingen die wij besproken hebben, hebben we geen grote verschillen tussen schatten met Probit Maximum Likelihood en Maximum Score aan kunnen tonen.

Een mogelijkheid om te onderzoeken of in toepassingen de Probit-Maximum-Likelihood-methode ten onrechte wordt gebruikt, is in elk geval het toetsen op normaliteit en het vergelijken van het Maximum-Likelihood- met het Maximum-Score-resultaat. Door bij het schatten met Maximum Score een groot aantal extra iteraties uit te voeren en meerdere startwaarden te kiezen (onder andere de genormaliseerde Maximum-Likelihood-schattingen), kan de beste schatting worden vergeleken met de Maximum-Likelihood-schatting. Op basis van de vergelijking en met behulp van de Lagrange-Multiplier-toets kunnen dan conclusies worden getrokken voor het schatten van het binaire keuzemodel.



## Referenties.

Amemiya, T., 1985, Advanced econometrics, Oxford: Basil Blackwell

Bera, K., C. Jarque and L. Lee, 1984, 'Testing the normality assumption in limited dependent variable models', International Economic Review 25, 563-578

Chesher, A. and M. Irish, 1987, 'Residual analysis in the grouped and censored normal linear model', Journal of Econometrics 34, 33-61

Efron, B., 1985, 'Bootstrap confidence intervals for a class of parametric problems', Biometrika 72, 45-58

Efron, B. and R. Tibshirani, 1986, 'Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy', Statistical Science 1, 54-77

Van der Genugten, B., 1988, Inleiding tot de waarschijnlijkheidsrekening en mathematische statistiek deel 2, Leiden/Antwerpen: H.E. Stenfert Kroese B.V.

Kapteyn, A., T. Wansbeek and Jeannine Buyze, 1980, 'The dynamics of preference formation', Journal of Economic Behavior and Organization 1, 123-157

Kapteyn, A., P. Kooreman and R. Willemse, 1988, 'Some methodological issues in the implementation of subjective poverty definitions', Journal of Human Resources 23, 222-242

Kim, J. and D. Pollard, 1989, 'Cube root asymptotics' (Working paper), Statistics Department Yale University

Kleijnen, J., en W. van Groenendaal, 1988, Simulatie: technieken en toepassingen, Schoonhoven: Academic Service

- Maddala, G.S., 1983, Limited-dependent and qualitative variables in econometrics, Cambridge: Cambridge Press
- Manski, C., 1985, 'Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator', Journal of Econometrics 27, 313-333
- Manski, C. and T. Thompson, 1986, 'Operational characteristics of maximum score estimation', Journal of Econometrics 32, 85-108
- Manski, C. and T. Thompson, 1987, MSCORE: A program for maximum score estimation of linear quantile regressions from binary response data with NPREG: A program for kernel estimation of univariate nonparametric regression functions, University of Wisconsin-Madison Department of Economics.
- Robinson, P., 1988, 'Semiparametric econometrics: a survey', Journal of Applied Econometrics 3, 35-51
- Ruud, A., 1983, 'Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models', Econometrica 51, 225-228