Een schattingsprocedure op basis van macro-data voor de (stationaire) overgangswaarschijnlijkheden in een Markovmodel ter beschrijving van de ontwikkeling van de omvangsverdeling van ondernemingen

Roemen, J.H.J.

Publication date:
1979

FEW

faculteit der economische wetenschappen

REEKS "TER DISCUSSIE"

[redacted]

[redacted]

Katholieke
Hogeschool
Tilburg

KATHOLIEKE HOGESCHOOL TILBURG

REEKS "TER DISCUSSIE"

No. 79.085 april 1979

Een schattingsprocedure op basis van macro-data voor de (stationaire) overgangswaarschijnlijkheden in een Markovmodel ter beschrijving van de ontwikkeling van de omvangsverdeling van ondernemingen.

J.H.J. Roemen.

FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
SUBFACULTEIT ECONOMETRIE

1. INLEIDING.

Naast het aantal ondernemingen vormt de verdeling van deze ondernemingen over een stelsel van omvangsklassen een belangrijk kenmerk voor de structuur van een bedrijfstak. De ontwikkeling in de tijd van deze omvangsverdeling kan worden beschreven door middel van een Markovmodel.[1)]

In deze studie gaan wij na hoe de parameters in een dergelijk model geschat kunnen worden, wanneer enkel geaggregeerde data[2)] m.b.t. de ondernemingen beschikbaar zijn en geen individuele gegevens.[3)] Daarbij beperken wij ons tot de situatie, waarbij hetzij het saldo van het aantal toe- en uittredende ondernemingen, maar niet de aantallen toe- en uittreders, bekend is hetzij het aantal ondernemingen in de loop van de tijd constant blijft, doordat toe- en uittreding uit de bedrijfstak niet mogelijk zijn.

We beginnen met de situatie voor een constant aantal ondernemingen.

---

1) Zie bijv.:

I. Adelman, A Stochastic Analysis of the Size Distribution of Firms, Journal of the American Statistical Association, 53, 1958, p. 893 e.v.

L. Engwall, Models of Industrial Structure, Heath, Lexington, 1973.

G. Mustert, De ontwikkeling van de omvangsverdeling van ondernemingen, Tilburg, 1976.

Y. Ijiri en H. Simon, Skew Distributions and the Size of Business Firms, North Holland, Amsterdam, 1977.

2) Zie bijv. ook:

T. Lee, G. Judge en A. Zellner, Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North Holland Publishing Company, Amsterdam, 1977.
Hoofdstuk 3, The estimation of transition probabilities from macro data, p. 31 e.v.
In hoofdstuk 9 van deze studie: "Bayesian analysis of the 'macro' model", p. 107, wordt een Bayesiaanse benadering voor de situatie met geaggregeerde data gegeven.

3) Zie voor de schattingsprocedure bij beschikbaarheid van individuele data bijv.:

T. Anderson en L. Goodman, Statistical Inference about Markov Chains, Annals of Mathematical Statistics, 28, 1957, p. 91 e.v.

T. Lee, G. Judge en A. Zellner, a.w.,

Hoofdstuk 2.2., Bayesian analysis of the micro model, p. 25. e.v.

2. UITGANGSPUNTEN EN VERONDERSTELLINGEN.

Stel, dat er zich een constant aantal van N ondernemingen in een bedrijfstak bevindt. Laat $\underline{x}_k(t)$, $k = 1,\ldots,N$, $t = 0,1,\ldots$, staan voor de omvangsklasse, waarin de k-de onderneming zich op tijdstip t bevindt. Bij m omvangsklassen wordt de uitkomstenruimte van $\underline{x}_k(t)$ gegeven door de verzameling $\{1,2,\ldots,m\}$. Stel dat $\{\underline{x}_k(t),\ t = 0,1,\ldots\}$ een enkelvoudige Markovketen is met overgangswhn.[1]

$$p_{ij}(t+1) = P\{\underline{x}_k(t+1) = j\,|\,\underline{x}_k(t) = i\},\ t = 0,1,2,\ldots,i,j=1,\ldots,m \quad (2.1)$$

De kans dat element k op tijdstip t+1 in toestand j zit, is afhankelijk van de toestand waarin element k zich op tijdstip t bevindt. Omdat deze kans kan variëren met de in beschouwing genomen periode, is tussen haken een index t opgenomen. Uiteraard geldt dat $\sum_{j=1}^{m} p_{ij}(t) = 1$ en $0 \leqq p_{ij}(t) \leqq 1$, $i,j = 1,\ldots,m$. Aangenomen wordt verder, dat $\underline{x}_k(t)$ en $\underline{x}_l(t)$, $k,l \in N$, onafhankelijk zijn: de ondernemingen bewegen zich onafhankelijk van elkaar door het stelsel van omvangsklassen.

We definiëren verder een stochast $\delta(\underline{x}_k(t)-j)$, die de waarde 1 aanneemt, als element k zich op tijdstip t in toestand j bevindt en de waarde 0, als dat niet het geval is. Met deze definitie staat $\underline{y}_j(t) := \frac{1}{N}\sum_{k=1}^{N}\delta(\underline{x}_k(t)-j)$ voor de fractie van de ondernemingen die zich ten tijde t in toestand j bevindt en $\underline{y}_{ij}(t+1) := \frac{1}{N}\sum_{k=1}^{N}\delta(\underline{x}_k(t)-i)\ \delta(\underline{x}_k(t+1)-j)$ voor de fractie van de ondernemingen, die in de periode van tijdstip t tot t+1 is overgegaan van toestand i naar toestand j.

Voor de verwachting van $\underline{y}_j(t)$ geldt:

$$E\{\underline{y}_j(t)\} = \frac{1}{N}\,E\,[\sum_{k=1}^{N}\{\delta(\underline{x}_k(t)-j)\}] = \frac{1}{N}\,N\,P\{(\underline{x}_k(t) = j)\} \quad (2.2)$$

met $P\{\underline{x}_k(t) = j\}$ de (onvoorwaardelijke) kans voor een onderneming zich ten tijde t in toestand j te bevinden. Voor deze kans geldt op basis van $P\{\underline{x}_k(t+1) = j \wedge \underline{x}_k(t) = i\} = P\{\underline{x}_k(t) = i\}\ P\{\underline{x}_k(t+1) = j\,|\,\underline{x}_k(t) = i\}$, dat

$$P\{\underline{x}_k(t+1) = j\} = \sum_{i=1}^{m} P\{\underline{x}_k(t) = i\}\ P\{\underline{x}_k(t+1) = j\,|\,\underline{x}_k(t) = i\},$$

of, met $P\{\underline{x}_k(t+1) = j\} = q_j(t+1)$,

---

1) Overgangswaarschijnlijkheden.

$$q_j(t+1) = \sum_{i=1}^{m} q_i(t)\, p_{ij}(t+1) \tag{2.3}$$

In matrixnotatie[1])

$$Q(t+1) = \Phi'(t+1)Q(t) \quad t = 0,1,2,\ldots \tag{2.4}$$

met $Q(t+1) = \{q_1(t+1),\ldots,q_m(t+1)\}'$ en $\Phi(t+1) = [\,p_{ij}(t+1)]$
Uitgeschreven op de aanvangsverdeling gaat (2.4) over in

$$Q(t+1) = \prod_{j=1}^{t+1} \Phi'(j)Q(0) \tag{2.5}$$

Bij stationariteit, waar wij in het vervolg van uitgaan, geldt

$$Q(t+1) = (\Phi')^{t+1}Q(0) \quad ^{2)} \tag{2.6}$$

De relatie (2.4) betitelen we in het volgende als een structurele relatie. Laat nu voor T opeenvolgende tijdstippen gegeven zijn de realisaties van de grootheden $\underline{y}_j(t)$. Er is dan sprake van de beschikbaarheid van zgn. macro- of geaggregeerde gegevens. Gegevens omtrent de loop van elk van de N elementen door het stelsel van toestanden, micro- of individuele gegevens, ontbreken echter, zo nemen we aan.

Uitgaande van de beschikbaarheid van geaggregeerde data zou nu met het oog op de schatting van de matrix Φ op basis van relatie (2.4) het volgende model gepostuleerd kunnen worden

$$\underline{Y}(t+1) = \Phi'\underline{Y}(t) \qquad t = 0,1,\ldots \tag{2.7}$$

met $\underline{Y}(t) = \{\underline{y}_1(t),\ldots,\underline{y}_m(t)\}'$.
De verdeling van de N ondernemingen over de m omvangsklassen op een tijdstip t+1 komt tot stand als resultaat van de verdeling op tijdstip t en de werking

1) Matrices zijn aangegeven met hoofdschrijfletters.
2) Onder bepaalde voorwaarden, zie daarvoor bijv. E. Seneta, Non-Negative Matrices, Allen and Unwin, London, 1973, p. 91, of, J. Kemeny en J. Snell, Finite Markov Chains, Van Nostrand, Princeton, 1960, p. 70, is de verdeling Q(t) invariant onder de transformatie volgens de matrix Φ. Deze invariante verdeling wordt gegeven door $\lim_{t\to\infty} \Phi^t$.
Wij besteden in het hierna volgende slechts terloops aandacht aan de consequenties van het bereiken van een zodanig tijdstip t, dat de verdeling Q(t) invariant is onder de transformatie volgens de matrix Φ.

van een verdelingsmechanisme, gegeven door de matrix van overgangswaarschijnlijkheden $\mathbf{P}$.
Voor de waarnemingen t.a.v. de vector $\underline{Y}(t)$ op T opeenvolgende tijdstippen zal het in het algemeen niet mogelijk zijn een matrix $\mathbf{P}$ te vinden, zodanig dat voor alle T tijdstippen (2.7) opgaat.[1]
Dat betekent, dat de relatie (2.7) aan verstoringen onderhevig is.
In verband met deze niet-waarneembare, als stochastische variabelen te beschouwen verstoringen voegen we aan (2.7) een vector van storingstermen toe; (2.7) gaat dan over in het volgende stelsel van stochastische lineaire differentievergelijkingen of eerste orde autoregressief schema

$$\underline{Y}(t+1) = \mathbf{P}'\underline{Y}(t) + \underline{E}(t+1), \quad t = 0,1,\ldots \tag{2.8}$$

met $\underline{E}(t+1) = \{\underline{\varepsilon}_1(t+1),\ldots,\underline{\varepsilon}_m(t+1)\}'$ een vector van storingstermen met $E\{\underline{E}(t+1)\} = 0$.
In termen van regressie-analyse geeft (2.8) een multivariaat, multipel regressiemodel met stochastische regressoren. Daarbij moeten de regressiecoëfficiënten p_{ij} voldoen aan

$$0 \leqq p_{ij} \leqq 1 \text{ en } \sum_{j=1}^{m} p_{ij} = 1, \; i = 1,\ldots,m.$$

V.w.b. de verdeling van de storingstermen merken we op, dat de elementen van de vector van contemporaine storingstermen zijn gecorreleerd. Uit

$$\sum_{j=1}^{m} \underline{y}_j(t+1) = \sum_{j=1}^{m}\sum_{k=1}^{m} p_{kj}\,\underline{y}_k(t) + \sum_{j=1}^{m} \underline{\varepsilon}_j(t+1) \tag{2.9}$$

volgt nl. dat $\sum_{j=1}^{m} \underline{\varepsilon}_j(t+1) = 0$.
Verder zijn $\underline{E}(t+k)$ en $\underline{Y}(t)$, $k = 1,2,\ldots$, ongecorreleerd, mits voor alle k $E\{\underline{E}(t+k)|\underline{Y}(t)\} = 0$. Er geldt nl.

$$\mathrm{Cov}\{\underline{Y}(t),\underline{E}(t+k)\} = E\{\underline{Y}(t)\{\underline{Y}(t+k)\}'\} - E\{\underline{Y}(t)\{\underline{Y}(t+k-1)\}'\mathbf{P}\} =$$

$$E[\underline{Y}(t)\{E\{\underline{Y}(t+k)|\underline{Y}(t)\}\}'] - E[\underline{Y}(t)\{E\{\underline{Y}(t+k-1)|\underline{Y}(t)\}\}']\mathbf{P} =$$

$$E\{\underline{Y}(t)\{\underline{Y}(t)\}'\mathbf{P}^k\} - E\{\underline{Y}(t)\{\underline{Y}(t)\}'\mathbf{P}^{k-1}\}\mathbf{P} = 0 \tag{2.10}$$

1) T. Lee, G. Judge en A. Zellner, a.w., p. 32.

Uit (2.10) volgt, dat

$$\mathrm{Cov}\{\underline{E}(t),\ \underline{E}(t+k)\} = \mathrm{Cov}\{\underline{Y}(t)-\varphi'\underline{Y}(t-1),\ E(t+k)\} = 0 \qquad (2.11)$$

De niet-contemporaine storingstermen zijn ongecorreleerd.
T.a.v. de (singuliere) (co)variantiematrix van de contemporaine storingstermen kan men veronderstellen, dat deze volledig onbekend is.
Lee e.a.[1)] volgend kan men er ook van uitgaan, dat de elementen van de vector van storingstermen elk, onafhankelijk van elkaar, een multinomiale verdeling volgen met $E\{\underline{E}(t)\} = 0$ en

$$V\{\underline{E}(t)\} = -\frac{1}{N(t)}\,[\,Q(t)\{Q(t)\}'-\mathrm{diag}\{Q(t)\}]$$

met $N(t)$ de omvang van de steekproef op tijdstip t - in ons geval N - en $\mathrm{diag}\{Q(t)\}$ een matrix met op de hoofddiagonaal de elementen van de vector $Q(t)$.
Op basis van de voorgaande veronderstellingen kan vervolgens een schattingsprocedure worden opgezet.
Beschouwen we echter $\underline{Y}(0)$ als gegeven, dan is naar onze mening geen veronderstelling omtrent de covariantiematrix van de contemporaine storingstermen nodig. In dat geval kan de waarde van elk element van deze matrix bepaald worden. Dat vindt zijn oorzaak daarin, dat het mogelijk is de covariantiematrix van $\underline{Y}(t)$ uit te drukken in $Y(0)$ en de matrix φ. Bij (bekende) matrix φ wordt de covariantiematrix van de contemporaine storingstermen dan gevonden uit

$$V\{\underline{E}(t+1)\,|\,\underline{Y}(0)\} = V\{\underline{Y}(t+1)\,|\,\underline{Y}(0)\}+\varphi' V\{\underline{Y}(t)\,|\,\underline{Y}(0)\}\varphi$$

$$-2\ \mathrm{Cov}\{\underline{Y}(t+1),\varphi'\underline{Y}(t)\,|\,\underline{Y}(0)\} \qquad (2.12)$$

Deze niet-hypothetische specificatie van de covariantiematrix van de storingstermen brengt met zich mee dat het aantal te schatten parameters van het model beperkt blijft tot m^2-m coëfficiënten p_{ij}.
De betekenis hiervan is duidelijk, wanneer men bedenkt dat voor een fenomeen als de omvangsverdeling van ondernemingen in het algemeen slechts waarnemingen voor een beperkt aantal tijdstippen beschikbaar zijn.

---

1) Lee, Judge en Zellner, a.w., bijv. p. 49 of p. 183.

3. DE SCHATTINGSPROCEDURE VOOR DE SITUATIE ZONDER TOE- EN UITTREDING.

Stel, dat aan het in de vorige § gegeven model

$$\underline{Y}(t+1) = \mathbf{P}'\underline{Y}(t) + \underline{E}(t+1) \qquad t = 0,1,\ldots \tag{3.1}$$

met $E\{\underline{E}(t+1)\} = 0$ en $E\{\underline{E}(t+1)|\underline{Y}(t)\} = 0$, zodat $\mathrm{Cov}\{\underline{E}(t+1), \underline{E}(s+1)\} = $ $= \delta_{t+1,s+1}\, V\{\underline{E}(t+1)\}$, $t,s = 0,1,\ldots,\delta_{t+1,s+1} = 1$ voor $t=s$ en 0 elders, op T opeenvolgende tijdstippen waarnemingen zijn verricht t.a.v. $\underline{Y}(t)$.

Met het oog op de schatting van de elementen van de matrix $\mathbf{P}$ brengen wij het voor de T waarnemingen geldende stelsel van multivariate, multipele regressievergelijkingen op de gedaante van het univariate stelsel.
Voor de i-de klasse, $i = 1,\ldots,m$, geldt bij een waarnemingsperiode $\{1,\ldots,T\}$

$$\begin{bmatrix} \underline{y}_i(2) \\ \underline{y}_i(3) \\ \cdots\cdots \\ \underline{y}_i(T) \end{bmatrix} = \begin{bmatrix} \underline{y}_1(1) & \underline{y}_2(1) & \cdots\cdots & \underline{y}_m(1) \\ \underline{y}_1(2) & \underline{y}_2(2) & \cdots\cdots & \underline{y}_m(2) \\ \cdots & \cdots & \cdots & \cdots \\ \underline{y}_1(T-1) & \underline{y}_2(T-1) & \cdots\cdots & \underline{y}_m(T-1) \end{bmatrix} \cdot \begin{bmatrix} p_{1i} \\ p_{2i} \\ \cdots \\ p_{mi} \end{bmatrix} + \begin{bmatrix} \underline{\varepsilon}_i(2) \\ \underline{\varepsilon}_i(3) \\ \cdots\cdots \\ \underline{\varepsilon}_i(T) \end{bmatrix} \tag{3.2}$$

In matrixnotatie:

$$\underline{Y}_i = \underline{\mathbf{Y}} P_i + \underline{E}_i \qquad i = 1,\ldots,m \tag{3.3}$$

Voor de (stochastische) steekproefmatrix $\underline{\mathbf{Y}}$ nemen we volledige (kolom)rang aan. Weliswaar leveren de elementen van $\underline{\mathbf{Y}}$ over de rijen gesommeerd steeds de waarde 1, maar dat impliceert geen lineaire afhankelijkheid van de kolommen van $\underline{\mathbf{Y}}$.
Voor alle klassen tesamen krijgen we nu het volgende stelsel:

$$\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \cdot\cdot \\ \underline{Y}_m \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{Y}} & 0 & \cdots & 0 \\ 0 & \underline{\mathbf{Y}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \underline{\mathbf{Y}} \end{bmatrix} \cdot \begin{bmatrix} P_1 \\ P_2 \\ \cdot\cdot \\ P_m \end{bmatrix} + \begin{bmatrix} \underline{E}_1 \\ \underline{E}_2 \\ \cdot\cdot \\ \underline{E}_m \end{bmatrix}, \tag{3.4}$$

of,

$$\underline{Y} = \underline{\mathbf{X}}\,\beta + \underline{E}, \tag{3.5}$$

met $\underline{Y}$ en $\underline{E}$ een $\{((T-1)\times m)\times 1\}$ vector, $\underline{X}$ een $(m\times(T-1))\times m^2$ matrix en β een $m^2\times 1$ vector.

Zien we voorlopig af van de voorwaarden waaraan de coëfficiënten p_{ij} moeten voldoen, dan resulteert op basis van het kleinste kwadratencriterium[1] als schatter voor β:

$$\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}\ \underline{Y} \quad ^{2)} \tag{3.6}$$

T.a.v. de eigenschappen van de schatter $\hat{\underline{\beta}}$ merken we op, dat we op basis van (3.6) niet kunnen aantonen, dat aan een (minimale) eis als zuiverheid voldaan wordt.[3]

Uit de uitschrijving van (3.6) blijkt, dat het niet uitmaakt, of de elementen van de vector β afzonderlijk of tesamen worden geschat,[4] ofschoon de storingstermen contemporain zijn gecorreleerd. Dat hoeft niet langer te gelden, wanneer bij de afleiding van een schatter voor β de voorwaarden op de regressiecoëfficiënten worden meegenomen. Teneinde de - gegeven een startwaarde bekende - covariantiematrix van de storingstermen mee te kunnen nemen gaan we nu over op een schattingsprocedure waar niet de uitdrukking:

---

1) Voor het gebruik van andere dan het kleinste kwadratencriterium zij verwezen naar Lee, Judge en Zellner, a.w.

2) De inverse bestaat, omdat voor $\underline{Y}$ volledige rang werd aangenomen.

3) Voor modellen met vertraagde endogene variabelen zoals (3.1) kan onder bepaalde voorwaarden consistentie van de schatter van de modelparameters worden bewezen. Voor het model (3.1) houden deze voorwaarden in, dat:

   1. het proces $\{\underline{Y}(t),\ t=0,1,\ldots\}$ covariantiestationair zou moeten zijn;
   2. de vectoren van storingstermen $\underline{E}(k)$ en $\underline{E}(s)$, $k,s = 1,2,\ldots$, identiek en onafhankelijk verdeeld zouden moeten zijn met verwachting 0 en $\mathrm{Cov}\{\underline{E}(k), \underline{E}(s)\} = \delta_{ks}\boldsymbol{\Sigma}$;
   3. de matrices $\underline{M}_{xx}^{(T)} := \frac{1}{T}(\underline{X}'_{(T)}\underline{X}_{(T)})^{-1}$, waarbij $\underline{X}_{(T)}$ de bij T waarnemingen behorende steekproefmatrix, convergeren voor $T\to\infty$ in waarschijnlijkheid naar een limiet.

   [E. Malinvaud, Statistical Methods of Econometrics, North-Holland, 1970, p. 546]. De betekenis van dit asymptotisch resultaat is, dat in modellen met vertraagde endogene variabelen de steekproefmatrix $\underline{X}$ voor een voldoend groot aantal waarnemingen als deterministisch mag worden beschouwd.
   [H. Theil, Principles of Econometrics, Wiley, 1971, p. 413].

4) H. Theil, a.w., p. 310. Van de singulariteit van de covariantiematrix van de storingstermen zien we daarbij een moment af.

$$\sum_{j=1}^{T-1} \{\underline{Y}(j+1)-\mathcal{P}'\underline{Y}(j)\}'\{\underline{Y}(j+1)-\mathcal{P}'\underline{Y}(j)\}$$

wordt geminimaliseerd als functie van $\mathcal{P}$, maar

$$\sum_{j=1}^{T-1} \{\underline{Y}(j+1)-E\{\underline{Y}(j+1)|\underline{Y}(j)\}\}'\{\underline{Y}(j+1)-E\{\underline{Y}(j+1)|\underline{Y}(j)\}\}.$$

Uitgaande van (3.1) geldt:

$$E\{\underline{Y}(t+1)|\underline{Y}(t)\} = \mathcal{P}'Y(t)+E\{\underline{E}(t+1)|\underline{Y}(t)\} = \mathcal{P}'Y(t), \tag{3.7}$$

wanneer we veronderstellen, dat $E\{\underline{E}(t+1)|\underline{Y}(t)\} = 0$, en

$$V\{\underline{Y}(t+1)|\underline{Y}(t)\} = V\{\underline{E}(t+1)|\underline{Y}(t)\} \tag{3.8}$$

Voor de niet-contemporaine storingstermen blijft de ongecorreleerdheid gehandhaafd:

$$Cov\{\underline{E}(k), \underline{E}(s)|\underline{Y}(t)\} = \delta_{ks}\, V\{\underline{E}(k)|\underline{Y}(t)\}, \tag{3.9}$$

mits $E\{\underline{E}(k)|\underline{Y}(t)\} = 0$ en k en s groter zijn dan t, $t = 0,1,\ldots, k,s = 1,2,\ldots$
De afleiding van de voorwaardelijke covariantiematrix van contemporaine storingstermen (3.8) verloopt als volgt. Er geldt:

$$V\{\underline{Y}(t)\} = E[E\{(\underline{Y}(t)-E\{\underline{Y}(t)\}\ (\underline{Y}(t)-E\{\underline{Y}(t)\})'|\underline{Y}(t-1)\}] =$$

$$E[E\{\underline{Y}(t)\{\underline{Y}(t)\}'|\underline{Y}(t-1)\}]-E[E\{\ \underline{Y}(t)\}(E\{\underline{Y}(t)\})'\ |\underline{Y}(t-1)] \tag{3.10}$$

Voor de eerste term aan de rechterkant van (3.10) geldt, elementsgewijs:

$$E[E\{\underline{y}_i(t)\underline{y}_j(t)|\underline{Y}(t-1)\}] = E[E\{\sum_{k=1}^{m} \underline{y}_{ki}(t).\sum_{l=1}^{m} \underline{y}_{lj}(t)|\underline{Y}(t-1)\}]$$

$$= \sum_{k=1}^{m}\sum_{l=1}^{m} E[E\{\underline{y}_{ki}(t)\underline{y}_{lj}(t)|\underline{Y}(t-1)\}], \tag{3.11}$$

waarbij

$$\underline{y}_{ki}(t) = \frac{\sum_{h=1}^{N} \delta(\underline{x}_h(t-1)-k)\ \delta(\underline{x}_h(t)-i)}{N}$$

Bedenken we, dat, wanneer de waarde van de elementen van de vector $\underline{Y}(t-1)$ vastligt, de elementen van $\underline{Y}(t)$ een multinomiale verdeling volgen,[1] dan volgt m.b.v. de momenten genererende functie van de multinomiale verdeling:

$$E[E\{\underline{y}_i(t)\underline{y}_j(t)|\underline{Y}(t-1)\}] = \sum_{k=1}^{m}\sum_{l=1}^{m} p_{ki}p_{lj}\ E\{\underline{y}_k(t-1)\underline{y}_l(t-1)\} +$$

$$-\frac{1}{N}\sum_{k=1}^{m} p_{ki}p_{kj}\ E\{\underline{y}_k(t-1)\}+\frac{1}{N}\sum_{k=1}^{m} \delta_{ij}p_{ki}\ E\{\underline{y}_k(t-1)\} =$$

$$\sum_{k=1}^{m}\sum_{l=1}^{m} p_{ki}p_{lj}\ E\{\underline{y}_k(t-1)\underline{y}_l(t-1)\} +$$

$$\frac{1}{N}\sum_{k=1}^{m} (\delta_{ij}p_{ki}-p_{ki}p_{kj})E\{\underline{y}_k(t-1)\}^{2)}\ ,\ i,j = 1,\dots,m$$

met $\delta_{ij} = 1$ voor $i = j$ en 0 voor $i \neq j$. (3.12)

In matrixnotatie luidt (3.12)

$$E[E\{\underline{Y}(t)\{\underline{Y}(t)\}'|\underline{Y}(t-1)\}] =$$

$$\mathcal{P}'E\{\underline{Y}(t-1)\{\underline{Y}(t-1)\}'\}\mathcal{P} - \frac{1}{N}\mathcal{P}'\ \mathrm{diag}\{Q(t-1)\}\mathcal{P} +$$

$$+\frac{1}{N}\mathrm{diag}\{\mathcal{P}'Q(t-1)\}, \qquad (3.13)$$

waarbij $\mathrm{diag}\{Q(t-1)\}$ en $\mathrm{diag}\ \{\mathcal{P}'Q(t-1)\}$ m×m-matrices met op de hoofddiagonaal de elementen van de vector $Q(t-1)$ resp. $\mathcal{P}'Q(t-1)$.
Voor de tweede term aan de rechterkant van (3.10) geldt, in matrixnotatie,

$$E[E\{\underline{Y}(t)\}(E\{\underline{Y}(t)\})'|\underline{Y}(t-1)\}] = E[E\{\underline{Y}(t)|\underline{Y}(t-1)\}(E\{\underline{Y}(t)|\underline{Y}(t-1)\})'] =$$

$$E\{\mathcal{P}'\underline{Y}(t-1)\}[E\{\mathcal{P}'\underline{Y}(t-1)\}]' = \mathcal{P}'E\{\underline{Y}(t-1)\}[E\{\underline{Y}(t-1)\}]'\mathcal{P} \qquad (3.14)$$

---

1) T. Anderson en L. Goodman, Statistical Inference about Markov Chains, Annals of Mathematical Statistics, 28, 1957, p. 91.

2) D. Bartholomew, Stochastic Models for Social Processes, Wiley, New York, 1967, p. 22.
U. Narayan Bhat, Elements of Applied Stochastic Processes, Wiley, New York, 1972. p. 301.

Op basis van (3.13) en (3.14) volgt nu, dat

$$V\{\underline{E}(t)\} = V\{\underline{Y}(t)\} + V\{\mathbf{P}'\underline{Y}(t-1)\} - 2\,\mathrm{Cov}\{\underline{Y}(t), \mathbf{P}'\underline{Y}(t-1)\} =$$

$$\mathbf{P}'E\{\underline{Y}(t-1)\{\underline{Y}(t-1)\}'\}\mathbf{P} - \frac{1}{N}\mathbf{P}'\mathrm{diag}\{Q(t-1)\}\mathbf{P} + \frac{1}{N}\mathrm{diag}\{\mathbf{P}'Q(t-1)\} - \mathbf{P}'E\{\underline{Y}(t-1)\}[E\{\underline{Y}(t-1)\}']\mathbf{P} +$$

$$\mathbf{P}'E\{\underline{Y}(t-1)\{\underline{Y}(t-1)\}'\}\mathbf{P} \qquad - \mathbf{P}'E\{\underline{Y}(t-1)\}[E\{\underline{Y}(t-1)\}']\mathbf{P} +$$

$$-2\mathbf{P}'E\{\underline{Y}(t-1)\}\{\underline{Y}(t-1)\}'\}\mathbf{P} \qquad + 2\mathbf{P}'E\{Y(t-1)\}[E\{\underline{Y}(t-1)\}']\mathbf{P}$$

$$= -\frac{1}{N}\mathbf{P}'\,\mathrm{diag}\{Q(t-1)\}\mathbf{P} + \frac{1}{N}\,\mathrm{diag}\{\mathbf{P}'Q(t-1)\} \qquad (3.15)$$

De uitdrukking (3.15) is in zoverre van betrekkelijke waarde, dat daarvoor de vectoren van verwachtingswaarden, $Q(t-1)$, $t = 2,3,\ldots$ en de matrix $\mathbf{P}$ bekend moeten zijn.

Beschouwen we echter $\underline{Y}(t-1)$ als gegeven, dan is voor de berekening van de covariantiematrix van storingstermen, naast de gegeven $Y(t-1)$, enkel de matrix $\mathbf{P}$ nodig.[1]
(3.15) gaat dan over in

$$V\{\underline{E}(t)|\underline{Y}(t-1)\} = -\frac{1}{N}(\mathbf{P}')\,\mathrm{diag}\{Y(t-1)\}\mathbf{P} + \frac{1}{N}\,\mathrm{diag}\{(\mathbf{P}')Y(t-1)\} \qquad (3.16)$$

Uitgeschreven luidt (3.16):

$$V\{\underline{E}(t)|\underline{Y}(t-1)\} = \sum_{i=1}^{m} \frac{y_i(t-1)}{N} \begin{bmatrix} p_{i1}(1-p_{i1}) & -p_{i1}p_{i2} & \cdots & -p_{i1}p_{im} \\ -p_{i2}p_{i1} & p_{i2}(1-p_{i2}) & \cdots & -p_{i2}p_{im} \\ \cdots & \cdots & \cdots & \cdots \\ -p_{im}p_{i1} & -p_{im}p_{i2} & \cdots & p_{im}(1-p_{im}) \end{bmatrix} \qquad (3.17)$$

---

1) Met kennis van alleen de matrix $\mathbf{P}$ kan ook volstaan worden voor de bepaling van (3.15), wanneer we er vanuit gaan, dat het aantal malen, dat volgens model (3.1) overgangen hebben plaatsgevonden nadert naar oneindig. In dat geval geldt, dat

$$V\{\underline{E}(t)\} = -\frac{1}{N}\mathbf{P}'\,\mathrm{diag}\{Q\}\mathbf{P} + \frac{1}{N}\,\mathrm{diag}\{Q\},$$

waarbij $Q = [q_1,\ldots,q_m]'$, de vector van verwachtingswaarden voor de stationaire toestand, gegeven wordt door $\lim_{t\to\infty}\{\mathbf{P}^t\}$. Wij gaan er in het hierna volgende niet vanuit dat de steekproefmatrix $[\underline{Y}\,\mathbf{X}]$ uit (3.5) betrekking heeft op zodanige tijdstippen waarvoor $Q = \mathbf{P}'Q$ met Q de invariante verdeling.

$$= \sum_{i=1}^{m} \frac{y_i(t-1)}{N} \cdot \mathcal{P}_i \tag{3.18}$$

Voor de covarianties van de storingstermen behorend bij een waarnemingsperiode $\{1,\ldots,T\}$ geldt op basis van (3.16), elementsgewijs,

$$\mathrm{Cov}\{\underline{\varepsilon}_j(g),\underline{\varepsilon}_k(g)\,|\,\underline{Y}(g-1)\} = \frac{1}{N}\sum_{l=1}^{m} (\delta_{jk}p_{lj}-p_{lj}p_{lk})y_l(g-1),\ g = 2,\ldots,T, \tag{3.19}$$

met $\delta_{jk} = 1$ voor $j = k$ en 0 voor $j \neq k$.
De bij de gehele waarnemingsperiode $\{1,\ldots,T\}$ behorende covariantiematrix wordt gegeven door het Kronecker-product

$$\Sigma = \mathrm{Cov}\{\underline{E},\underline{E}\,|\,\underline{\mathbf{Y}}\} = -\frac{1}{N}\sum_{i=1}^{m} \{\mathcal{P}_i \otimes (\mathrm{diag}\ Y_i^{*})\},^{1)} \tag{3.20}$$

waarbij $\mathcal{P}_i$ gegeven is in (3.18) en

$$\mathrm{diag}\ Y_i^{*} = \begin{bmatrix} y_i(1) & & 0 \\ & \ddots & \\ 0 & & y_i(T-1) \end{bmatrix}$$

In een figuur ziet (3.20) er als volgt uit:

---

1) Voor de stationaire toestand wordt de onvoorwaardelijke covariatiematrix bij een waarnemingsperiode $\{1,\ldots,T\}$ gegeven door

$$\mathrm{Cov}\{\underline{E},\underline{E}\} = -\frac{1}{N}\,(\mathcal{P}'\ \mathrm{diag}\{Q\}\,\mathcal{P} - \mathrm{diag}\{Q\})\otimes I_{T-1,T-1},$$

met $\mathrm{diag}\ Q = \begin{bmatrix} q_1 & & 0 \\ & \ddots & \\ 0 & & q_m \end{bmatrix}$ en $I_{T-1,T-1}$ een $(T-1)\times(T-1)$ eenheidsmatrix.

FIGUUR 3.1. Cov{$\underline{E}$,$\underline{E}$|$\mathbf{Y}$}

$$
-N.\boldsymbol{\Sigma} =
\begin{array}{c|cccc|}
 & \underline{\varepsilon}_1(2)\ldots\ldots\ldots) & \underline{\varepsilon}_1(T) & \underline{\varepsilon}_2(2)\ldots\ldots & \underline{\varepsilon}_2(T) \\
\hline
\underline{\varepsilon}_1(2) & \sum_i p_{i1}(1-p_{i1})y_i(1) & 0 & \sum_i p_{i1}p_{i2}y_i(1) & 0 \\
\vdots & 0 & \ddots & 0 & \ddots \\
\underline{\varepsilon}_1(T) & & \sum_i p_{i1}(1-p_{i1})y_i(T-1) & & \sum_i p_{i1}p_{i2}y_i(T-1) \\
\underline{\varepsilon}_2(2) & \sum_i p_{i1}p_{i2}y_i(1) & 0 & \sum_i p_{i2}(1-p_{i2})y_i(1) & 0 \\
\vdots & 0 & \ddots & 0 & \ddots \\
\underline{\varepsilon}_2(T) & & \sum_i p_{i1}p_{i2}y_i(T-1) & & \sum_i p_{i2}(1-p_{i2})y_i(T-1) \\
\underline{\varepsilon}_3(2) & \sum_i p_{i3}p_{i1}y_i(1) & 0 & & \\
 & 0 & \ddots & & \\
\hline
\end{array}
$$

Omdat de covariantiematrix van de storingstermen (3.20) vanwege $\sum_{j=1}^{m} \underline{\varepsilon}_j(g)=0$ geen inverse bezit, kan voor de schatting van de vector van regressiecoëfficiënten β met inachtneming van de correlatiestructuur op de storingstermen niet zonder nadere voorzieningen gebruik worden gemaakt van de gegeneraliseerde kleinste kwadraten of Aitken - schatter. In verband met de singulariteit dient overgegaan te worden op het gebruik van de Moore-Penrose inverse of het aantal te schatten parameters te worden verminderd.

Omdat van de vectoren P_i, i = 1,...,m, er tengevolge van de somvoorwaarde m-1 onafhankelijk zijn, schrappen wij in (3.4) de vector P_m en de daarmee corresponderende vergelijkingen en proberen daarmee te bereiken, dat de singulariteit van de covariantiematrix niet optreedt.

In plaats van (3.4) nemen we nu het volgende stelel als uitgangspunt.

$$\underline{Y}_{*} = \underline{\mathbf{X}}_{*}\, \beta_{*} + \underline{E}_{*}, \qquad (3.21)$$

waarbij

$$\underline{Y}_* = \begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \vdots \\ \underline{Y}_{m-1} \end{bmatrix} \quad \underline{E}_* = \begin{bmatrix} \underline{E}_1 \\ \underline{E}_2 \\ \vdots \\ \underline{E}_{m-1} \end{bmatrix} \quad \beta_* = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_{m-1} \end{bmatrix} \quad \mathsf{X}_* = \begin{bmatrix} \underline{\mathbf{Y}} & & & \\ & \underline{\mathbf{Y}} & & \\ & & \ddots & \\ & & & \underline{\mathbf{Y}} \end{bmatrix}$$

met $\underline{\mathsf{X}}_*$ een $\{(m-1)(T-1)\}\times m(m-1)$ matrix

$$\Sigma_* = \mathrm{Cov}\{\underline{E}_*, \underline{E}_* | \underline{\mathbf{Y}}\} = -\frac{1}{N}\sum_{i=1}^{m} \{\Phi_{i*} \otimes (\mathrm{diag}\ Y_i^*)\},$$

$$\Phi_{i*} = \begin{bmatrix} p_{i1}(1-p_{i1}) - p_{i1}p_{i2} & \cdots & -p_{i1}p_{i,m-1} \\ -p_{i2}p_{i1} & \cdots & -p_{i2}p_{i,m-1} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ -p_{i,m-1}p_{i1} & \cdots & p_{i,m-1}(1-p_{i,m-1}) \end{bmatrix}$$

Daarbij gaan we er vanuit dat $\mathbf{\Sigma}_*$ een inverse bezit.

Zien we af van de (ongelijkheids)restricties op de regressiecoëfficiënten, $\sum_{j=1}^{m-1} p_{ij} \leq 1$ en $0 \leq p_{ij} \leq 1$ en beschouwen we (3.21) als een multivariaat multipel, regressiestelsel dan levert minimalisatie naar Φ_* van

$$\underline{F} = \sum_{g=2}^{T} \{\mathbf{K}_*(g)\underline{Y}_*(g) - \mathbf{K}_*(g)E\{\underline{Y}_*(g)|\underline{Y}_*(g-1)\}\}'\{\mathbf{K}_*(g)\underline{Y}_*(g) - \mathbf{K}_*(g)E\{\underline{Y}_*(g)|\underline{Y}_*(g-1)\}\}$$

$$= (\mathbf{K}_*\underline{Y}_* - \mathbf{K}_*\mathsf{X}_*\beta_*)'(\mathbf{K}_*\underline{Y}_* - \mathbf{K}_*\mathsf{X}_*\beta_*),$$

waarbij

$$\Phi_* = [P_1, \ldots, P_{m-1}], \quad \underline{Y}_*(g) = [\underline{y}_1(g), \ldots, \underline{y}_{m-1}(g)],$$

$$\mathbf{K}_*(g)'\mathbf{K}_*(g) = V\{\underline{E}_*(g)|\underline{Y}_*(g-1)\}^{-1} \text{ en}$$

$$\mathbf{K}_*'\mathbf{K}_* = \mathbf{\Sigma}_*^{-1}$$

$$\underline{\hat{\beta}}_{*} = (\mathbf{X}_{*}' \Sigma_{*}^{-1}\mathbf{X}_{*})^{-1}\mathbf{X}_{*}' \Sigma_{*}^{-1} \underline{Y}_{*}, \quad ^{1)} \tag{3.22}$$

aangenomen dat $(\mathbf{X}_{*}' \Sigma_{*}^{-1}\mathbf{X}_{*})^{-1}$ bestaat en $\mathbf{\Sigma}_{*}$ bekend is. De schatter voor de overige elementen van de matrix $\mathbf{P}$, de vector P_m, wordt gevonden uit:

$$\underline{\hat{P}}_m = 1_m - \sum_{i=1}^{m-1} \underline{\hat{P}}_i \tag{3.23}$$

met 1_m een m×1 vector van enen.
De formule (3.22) is in zoverre van betrekkelijke waarde dat daarin de onbekende $\mathbf{\Sigma}_{*}$ voorkomt, die een functie is van de te schatten matrix $\mathbf{P}$. Verder kan niet aangetoond worden dat (3.22) voldoet aan de ongelijkheidsrestricties $0 \leqq p_{ij} \leqq 1$ en $\sum_{j=1}^{m} p_{ij} \leqq 1$, die tot hieraan buiten beschouwing werden gelaten. Omdat aan deze restricties moet zijn voldaan, wil het model (3.1) zinvol zijn en omdat niet bekend is hoe deze voorwaarden gehanteerd dienen te worden bij de afleiding van een schatter, resteert als enige mogelijkheid een herformulering van het schattingsprobleem, waarin de ongelijkheidsrestricties expliciet worden meegenomen. Uitgaande van het kleinste kwadratencriterium wordt dan de formulering:

$$\min(\underline{Y}_{*} - \mathbf{X}_{*} \beta_{*})' \hat{\mathbf{\Sigma}}_{*}^{-1}(\underline{Y}_{*} - \mathbf{X}_{*} \beta_{*}) \tag{3.24}$$

onder de voorwaarden

$$\sum_{j=1}^{m-1} p_{ij} \leqq 1, \; i = 1,\dots,m$$

---

1) Op basis van 1) pag. 11, d.w.z. het systeem heeft de stationaire toestand bereikt, kan deze uitdrukking vereenvoudigd worden. Bedenken we dat de steekproefmatrix $\underline{\mathbf{X}}$ geschreven kan worden als Kronecker-product van matrices, dan geldt uitgaande van (3.15)

$$\underline{\hat{\beta}}_{*} = [\,(I_{m-1,m-1} \otimes \underline{\mathbf{Y}}')(-\tfrac{1}{N}\{\mathbf{P}_{*}' \operatorname{diag}\{Q\}\mathbf{P}_{*} - \operatorname{diag}\{Q\}\} \otimes I_{T-1,T-1}),$$

$$(I_{m-1,m-1} \otimes \underline{\mathbf{Y}})]^{-1}(I_{m-1,m-1} \otimes \underline{\mathbf{Y}})(-\tfrac{1}{N}\{\mathbf{P}_{*}' \operatorname{diag}\{Q\}\mathbf{P}_{*} - \operatorname{diag}\{Q\})\underline{\mathbf{Y}}$$

$$= I_{m-1,m-1} \otimes (\underline{\mathbf{Y}}'\underline{\mathbf{Y}})^{-1}\, \underline{\mathbf{Y}}\, \underline{Y} = \begin{bmatrix} \underline{\hat{P}}_1 \\ \vdots \\ \underline{\hat{P}}_{m-1} \end{bmatrix},$$

waarbij $[\underline{\hat{P}}_1,\dots,\underline{\hat{P}}_{m-1}]'$ staan voor de kleinste kwadraten schatters voor ieder afzonderlijk vergelijkingenstelsel. Dat betekent dat het combineren van de (m-1) vergelijkingenstelsels niet meer oplevert dan de kleinste kwadraten procedure, ofschoon de vectoren $\underline{E}_1$ t/m $\underline{E}_{m-1}$ gecorreleerd zijn.
Cf. Theil, Principles of Econometrics. Wiley 1971, p. 310.

$$0 \leqq p_{ij} \leqq 1$$

Dit kwadratisch programmeringsprobleem kan worden opgelost m.b.v. één van de beschikbare algorithmen. De geschatte covariantiematrix $\hat{\Sigma}_*$ wordt daarbij verkregen op basis van een OLS-schatting voor P onder de voorwaarden

$$\sum_{j=1}^{m} p_{ij} = 1, \; i = 1,\ldots,m, \text{ en } 0 \leqq p_{ij} \leqq 1.$$

In een iteratieve procedure kan nu de optimale waarde van de elementen van de vector β worden gezocht, aangenomen dat in iedere iteratie de inverse van de covariantiematrix bestaat.
Tenslotte merken we op dat de eigenschappen van een schatter onder ongelijkheidsrestricties als (3.24) niet bekend zijn.[1)]

---

1) H. Theil, Principles of Econometrics, Wiley, 1971, p.353.

4. DE SCHATTINGSPROCEDURE VOOR DE SITUATIE MET TOE- EN UITTREDING.

In deze paragraaf ontwikkelen we op basis van de resultaten van de vorige paragraaf een schattingsprocedure voor de situatie met toe- en uittreding. We gaan er niet langer vanuit, dat toetreding tot de bedrijfstak van nieuwe ondernemingen en uittreding uit de bedrijfstak van bestaande ondernemingen onmogelijk zijn. Dientengevolge hoeft het aantal ondernemingen in de bedrijfstak niet langer constant te blijven in de loop van de tijd.
Net als in de vorige paragraaf nemen we aan, dat voor T opeenvolgende waarnemingstijdstippen gegeven is het totaal aantal ondernemingen in de bedrijfstak en de opdeling daarvan over de omvangsklassen, maar niet de gang van elke individuele onderneming door het stelsel van omvangsklassen in deze T perioden. We beschikken zodoende enkel over de per saldo mutatie van het aantal ondernemingen van waarnemingstijdstip tot waarnemingstijdstip, maar niet over de uitsplitsing van dit saldo naar toe- en uittreders en evenmin over de saldi van toe- en uittreders per omvangsklasse.

Om het fenomeen van toe- en uittreding in een model als (2.8) op te kunnen nemen zouden we het aantal onderscheiden omvangsklassen met één uit kunnen breiden tot m+1. In deze omvangsklasse verdwijnen de uittredende ondernemingen en uit deze omvangsklasse treden nieuwe ondernemingen tot de bedrijfstak toe. We noemen deze omvangsklasse de klasse van potentiële toetreders.

Als model voor de ontwikkeling in de tijd van de verdeling van de ondernemingen over een stelsel van omvangsklassen zouden we dan uit kunnen gaan van

$$\underline{F}(t+1) = \mathbf{O}'\underline{F}(t) + \underline{E}(t+1),\ t = 0,1,... \qquad (4.1)$$

waarbij de kolomvector

$$\underline{F}(t+1) = \{f_0(t+1),...,\underline{f}_m(t+1)\}'$$

de aantallen ondernemingen in de (m+1) omvangsklassen op tijdstip t+1 geeft, de vector van storingstermen $\underline{E}(t+1) = \{\underline{\varepsilon}_0(t+1),...,\underline{\varepsilon}_m(t+1)\}$ dezelfde functie vervult als in (2.8) en

$$\mathbf{O} = \begin{bmatrix} o_{00} & o_{01} & o_{02} & \cdots & o_{0m} \\ o_{10} & o_{11} & o_{12} & \cdots & o_{1m} \\ o_{20} & o_{21} & o_{22} & \cdots\cdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ o_{m0} & o_{m1} & o_{m2} & \cdots\cdots & \end{bmatrix} = \begin{bmatrix} o_{00} & o_{01} & o_{02} & \cdots \\ o_{10} & (1-o_{10})p_{11} & (1-o_{10})p_{12} & \cdots \\ o_{20} & (1-o_{20})p_{21} & (1-o_{20})p_{22} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ o_{m0} & (1-o_{m0})p_{m1} & (1-o_{m0})p_{m2} & \cdots \end{bmatrix}$$

met o_{0j}, $j = 0,\ldots,m$, de kans op toetreding in klasse j voor een nieuwe onderneming en o_{j0}, $j = 1,\ldots,m$, de kans op uittreding uit klasse j voor een bestaande onderneming.

De verdeling van de ondernemingen over de (m+1) omvangsklassen op tijdstip t+1 komt tot stand als resultaat van de verdeling op tijdstip t, de werking van een verdelingsmechanisme, gegeven door de matrix van overgangswaarschijnlijkheden $\mathbf{P}$, en de vectoren van kansen $\{o_{01},\ldots,o_{0m}\}$ en $\{o_{10},\ldots,o_{m0}\}$.

Op basis van (4.1) zou nu voor de afleiding van een schattingsprocedure eenzelfde redenering kunnen worden gevolgd als in § 3 is gegeven voor de situatie met een constant aantal ondernemingen.

Aan een schattingsprocedure op basis van (4.1) is echter het bezwaar verbonden, dat voor de klasse van potentiële toetreders geen waarnemingen beschikbaar zijn. Dit bezwaar kan ondervangen worden door de omvang van de klasse van potentiële toetreders vast te leggen op grond van overwegingen buiten het model om. Daarbij zou men zich kunnen laten leiden door de marktverhoudingen waardoor de in beschouwing genomen bedrijfstak wordt gekenmerkt. Bij marktverhoudingen die lijken op volledige mededinging wordt dan voor een grote omvang van de klasse van potentiële toetreders gekozen en bij marktverhoudingen die lijken op een oligopolie of monopolistische concurrentie voor een kleine omvang.[1)]

Omdat de schattingsresultaten afhankelijk zijn van de gekozen omvang van de klasse van potentiële toetreders[2)], biedt bovenstaand uitgangspunt naar onze mening onvoldoende houvast voor een schattingsprocedure op basis van (4.1).

Om deze reden zien we af van het model (4.1) en nemen in plaats daarvan het volgende model als uitgangspunt.

1) B.Stanton en L. Kettunen, Potential Entrants ans Projections in Markov Process Analysis, Journal of Farm Economics, 49, 1967, p. 639.

2) G. Mustert, De ontwikkeling van de omvangsverdeling van ondernemingen, Tilburg, 1976, p. 131.

$$\underline{F}(t+1) = \mathrm{P}'\,\underline{F}(t) + \underline{u}(t+1)S + \underline{H}(t+1),\ t = 0,1,\ldots \qquad (4.2)$$

waarbij de kolomvector

$$\underline{F}(t+1) = \{\underline{f}_1(t+1),\ldots,\underline{f}_m(t+1)\}'$$

de aantallen ondernemingen in de m omvangsklassen op tijdstip t+1 geeft, de vector van storingstermen $\underline{H}(t+1)$ dezelfde functie vervult als in (2.8), $\underline{u}(t+1) = \sum_{j=1}^{m} \underline{f}_j(t+1) - \sum_{j=1}^{m} \underline{f}_j(t)$ en met behulp van de kolomvector $S = \{s_1,\ldots, s_m\}'$ de verdeling van de per saldo groei van het aantal ondernemingen over de m omvangsklassen plaatsvindt. De verdeling van de ondernemingen over de m omvangsklassen op tijdstip t+1 komt tot stand als resultaat van de verdeling op tijdstip t en de werking van matrix P alsmede van een correctie op deze herverdeling i.v.m. de per saldo groei van het aantal ondernemingen in de bedrijfstak. Daarbij geldt voor de coëfficiënten p_{ij} en s_j dat

$$\sum_{j=1}^{m} p_{ij} = 1,\quad \sum_{j=1}^{m} s_j = 1,\quad 0 \leqq p_{ij}, s_j \leqq 1.$$

Voor een handiger voorstelling voegen we in (4.2) de matrix P en de vector S samen. We krijgen dan:

$$\begin{bmatrix} \underline{f}_1(t+1) \\ \underline{f}_2(t+1) \\ \vdots \\ \underline{f}_m(t+1) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{m1} & s_1 \\ p_{12} & p_{22} & \cdots & p_{m2} & s_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{1m} & p_{2m} & \cdots & p_{mm} & s_m \end{bmatrix} \cdot \begin{bmatrix} \underline{f}_1(t) \\ \underline{f}_2(t) \\ \vdots \\ \underline{f}_m(t) \\ \underline{u}(t+1) \end{bmatrix} + \begin{bmatrix} \underline{n}_1(t+1) \\ \underline{n}_2(t+1) \\ \vdots \\ \underline{n}_m(t+1) \end{bmatrix} \qquad (4.3)$$

In matrixnotatie:

$$\underline{F}(t+1) = \mathcal{B}' \, \underline{F}^{*}(t) + \underline{H}(t+1)^{1)}, \quad t = 0,1,\ldots \tag{4.4}$$

met $\underline{F}^{*}(t) = \{\underline{f}_1(t),\ldots,\underline{f}_m(t),\ \underline{u}(t+1)\}'$.
T.a.v. de vector van storingstermen $\underline{H}(t+1)$ veronderstellen we, dat

$$E\{\underline{H}(t+k)\,|\,\underline{F}^{*}(t)\} = 0, \quad k = 1,2,\ldots \tag{4.5}$$

Op basis van deze veronderstelling geldt, dat

$$\mathrm{Cov}\{\underline{F}^{*}(t),\ \underline{H}(t+k)\} = 0,^{2)} \tag{4.6}$$

$$\mathrm{Cov}\{\underline{H}(t),\ \underline{H}(t+k)\} = 0,^{3)} \tag{4.7}$$

en

$$\mathrm{Cov}\{\underline{H}(k),\ \underline{H}(s)\,|\,\underline{F}^{*}(t)\} = \delta_{ks}\ V\{\underline{H}(k)\,|\,\underline{F}^{*}(t)\},^{4)} \tag{4.8}$$

voor k en s groter dan t, t = 0,1,...,k,s = 1,2,...

---

1) Beschouwen we de elementen van $\underline{F}(t+1)$ als te verklaren variabelen en die van $\underline{F}(t)$ als verklarende, dan resulteert, wanneer we alle te verklaren variabelen overbrengen naar het linkerlid, als structurele vorm van (4.2)

$$\mathcal{A}_0\ \underline{F}(t+1) = \mathcal{A}_1'\ \underline{G}(t) + \underline{H}(t+1), \quad t = 0,1,\ldots,$$

waarbij

$$\mathcal{A}_0 = \begin{bmatrix} (1-s_1) & -s_1 & -s_1 \ldots -s_1 \\ -s_2 & (1-s_2) & -s_2 \ldots -s_2 \\ \ldots & \ldots & \ldots \\ -s_m & -s_m & -s_m \ldots (1-s_m) \end{bmatrix}, \quad \mathcal{A}_1' = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{m1} & -s_1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ p_{1m} & p_{2m} & \cdots & p_{mm} & -s_m \end{bmatrix}$$

en $\underline{G}(t) = \{\underline{f}_1(t),\ldots,\underline{f}_m(t),\ \sum_{j=1}^{m} \underline{f}_j(t)\}'$.

In een schattingsprocedure op basis van het kleinste kwadraten criterium voor deze structurele vorm komen de regressiecoëfficiënten niet lineair voor. Om die reden maken wij geen gebruik van deze structurele vorm.

2) Verg. (2.10)

3) Verg. (2.11)

4) Verg. (3.9).

Op basis van eenzelfde redenering als gegeven in § 3 wordt de voorwaardelijke covariantiematrix van contemporaine storingstermen nu gegeven door

$$V\{\underline{H}(t)|\underline{F}^{*}(t-1)\} = -\mathcal{B}'\,\mathrm{diag}\{F^{*}(t-1)\}\mathcal{B} + \mathrm{diag}\{\mathcal{B}'F^{*}(t-1)\}^{1)} \qquad (4.9)$$

$$= [\sum_{i=1}^{m+1} f_i(t-1)\varphi_i]^{2)}$$

waarbij

$$\varphi_i = \begin{bmatrix} p_{i1}(1-p_{i1}) & \cdots & -p_{i1}p_{im} \\ -p_{i2}p_{i1} & \cdots & -p_{i2}p_{im} \\ \text{- - -} & & \text{- - -} \\ -p_{im}p_{i1} & \cdots & p_{im}(1-p_{im}) \end{bmatrix},\ i=1,\ldots,m,\ \varphi_{m+1} = \begin{bmatrix} s_1(1-s_1) & -s_1s_2 & \cdots & -s_1s_m \\ -s_2s_1 & s_2(1-s_2) & \cdots & -s_2s_m \\ \text{- - -} & & & \text{- - -} \\ -s_ms_1 & -s_ms_2 & \cdots & s_m(1-s_m) \end{bmatrix}$$

en $f_{m+1}(t-1) = u(t)$.

Laat nu voor T opeenvolgende tijdstippen waarnemingen zijn gegeven t.a.v. de aantallen ondernemingen in de m omvangsklassen.
Voor de i-de klasse, i = 1,...,m, geldt bij een waarnemingsperiode {1,...,T}

$$\begin{bmatrix} \underline{f}_i(2) \\ \underline{f}_i(3) \\ \vdots \\ \underline{f}_i(T) \end{bmatrix} = \begin{bmatrix} \underline{f}_1(1) & \underline{f}_2(1) & \cdots & \underline{f}_m(1) & \underline{u}(2) \\ \underline{f}_1(2) & \underline{f}_2(2) & \cdots & \underline{f}_m(2) & \underline{u}(3) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \underline{f}_1(T-1) & \underline{f}_2(T-1) & \cdots & \underline{f}_m(T-1) & \underline{u}(T) \end{bmatrix} \cdot \begin{bmatrix} p_{1i} \\ p_{2i} \\ \vdots \\ p_{mi} \\ s_i \end{bmatrix} + \begin{bmatrix} \underline{n}_i(2) \\ \underline{n}_i(3) \\ \vdots \\ \underline{n}_i(T) \end{bmatrix} \qquad (4.10)$$

In matrixnotatie

$$\underline{F}_i = \underline{\mathcal{F}}\, B_i + \underline{H}_i,\ i = 1,\ldots,m \qquad (4.11)$$

We nemen aan, dat de stochastische steekproefmatrix $\underline{\mathcal{F}}$ van volledige rang is.
Voor alle klassen tesamen krijgen we het volgende stelsel.

---

1) Verg. (3.16)

2) Verg. (3.18)

$$\begin{bmatrix} \underline{F}_1 \\ \underline{F}_2 \\ \vdots \\ \underline{F}_m \end{bmatrix} = \begin{bmatrix} \underline{\mathsf{F}} & 0 & \dots & 0 \\ 0 & \underline{\mathsf{F}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \underline{\mathsf{F}} \end{bmatrix} \cdot \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{bmatrix} + \begin{bmatrix} \underline{H}_1 \\ \underline{H}_2 \\ \vdots \\ \underline{H}_m \end{bmatrix} , \qquad (4.12)$$

of;

$$\underline{F} = \underline{\mathbf{L}}\, \gamma + \underline{H} \qquad (4.13)$$

met $\underline{F}$ en $\underline{H}$ een $\{((T-1)\times m)\times 1\}$ vector, $\underline{\mathbf{L}}$ een $\{(m\times(T-1))\times(m(m+1))\}$ matrix en γ een $\{(m(m+1))\times 1\}$ vector.

I.v.m. de singulariteit van de covariantiematrix van contemporaine storingstermen (4.9) schrappen we in (4.3) de vector B_m en de daarmee corresponderende vergelijkingen en nemen aan, dat daarmee de singulariteit van de covariantiematrix is opgeheven.
Als gevolg van deze reductie gaat (4.13) over in

$$\underline{F}_{\star} = \underline{\mathbf{L}}_{\star}\, \gamma_{\star} + \underline{H}_{\star} \qquad (4.14)$$

waarbij $\underline{F}_{\star}$, $\underline{\mathbf{L}}_{\star}$, $\gamma_{\star}$ en $\underline{H}_{\star}$ analoog met (3.21) zijn gedefinieerd.
De bij de gehele waarnemingsperiode $\{1,\dots,T\}$ behorende voorwaardelijke covariantiematrix wordt gegeven door het Kronecker-product

$$\Xi := \mathrm{Cov}\{\underline{H}_{\star},\underline{H}_{\star}|\underline{\mathbf{L}}_{\star}\} = \sum_{i=1}^{m+1} \{\Phi_{i\star} \otimes (\mathrm{diag}\ F_{i\star})\} \qquad (4.15)$$

waarbij

$$\Phi_{i\star} = \begin{bmatrix} p_{i1}(1-p_{i1}) & -p_{i1}p_{i2} & \cdots & -p_{i1}p_{i,m-1} \\ -p_{i1}p_{i2} & p_{i2}(1-p_{i2}) & \cdots & -p_{i2}p_{i,m-1} \\ & & & \\ -p_{i1}p_{i,m-1} & -p_{i2}p_{i,m-1} & \cdots & p_{i,m-1}(1-p_{i,m-1}) \end{bmatrix}$$

met $p_{m+1,j} = s_j$, en

$$F_{i\star} = \begin{bmatrix} f_i(1) & & 0 \\ & \ddots & \\ 0 & & f_i(T-1) \end{bmatrix}$$

met $f_{m+1}(g) = u(g+1)$ $g = 1,\ldots,T-1$.
Uitgaande van het kleinste kwadratencriterium wordt de formulering van het schattingsprobleem met inachtneming van de restricties op de coëfficiënten nu:

$$\min\ (\underline{F}_* - \mathbf{L}_*\, \gamma_*)'(\hat{\Xi}_*)^{-1}\ (\underline{F}_* - \mathbf{L}_*\gamma_*)^{1)} \qquad (4.16)$$

onder de voorwaarden

$$\sum_{j=1}^{m-1} p_{ij} \leq 1, \quad i = 1,\ldots,m$$

$$\sum_{j=1}^{m-1} s_j \leq 1$$

$$0 \leq p_{ij},\ s_j \leq 1.$$

Daarbij wordt $\hat{\Xi}_*$ verkregen op basis van een OLS schatting voor $\mathcal{B}$ onder de restricties

$$\sum_{j=1}^{m} p_{ij} = 1 \quad i = 1,\ldots,m, \quad \sum_{j=1}^{m} s_j = 1, \quad 0 \leq p_{ij},\ s_j \leq 1$$

en vervolgens in een iteratieve procedure gezocht naar de optimale waarde van de regressiecoëfficiënten.

1) Verg. (3.25)

In de Reeks ter Discussie zijn verschenen:

1.H.H. Tiggelaar	Spectraalanalyse en stochastische lineaire differentievergelijkingen.	juni '75
2.J.P.C.Kleijnen	De rol van simulatie in de algemene econometrie.	juni '75
3.J.J. Kriens	A stratification procedure for typical auditing problems.	juni '75
4.L.R.J. Westermann	On bounds for Eigenvalues	juni '75
5.W. van Hulst J.Th. van Lieshout	Investment/financial planning with endogenous lifetimes: a heuristic approach to mixed integer programming.	juli '75
6.M.H.C.Paardekooper	Distribution of errors among input and output variables.	augustus '75
7.J.P.C. Kleijnen	Design and analysis of simulation Practical statistical techniques.	augustus '75
8.J. Kriens	Accountantscontrole met behulp van steekproeven.	september '75
9.L.R.J. Westermann	A note on the regula falsi	september '75
10.B.C.J. van Velthoven	Analoge simulatie van ekonomische modellen.	november '75
11.J.P.C. Kleijnen	Het ekonomisch nut van nauwkeurige informatie: simulatie van ondernemingsbeslissingen en informatie.	november '75
12.F.J. Vandamme	Theory change, incompatibility and non-deductibility.	december '75
13.A. van Schaik	De arbeidswaardeleer onderbouwd?	januari '76
14.J.vanLieshout J.Ritzen J.Roemen	Input-ouputanalyse en gelaagde planning.	februari '76
15.J.P.C.Kleijnen	Robustness of multiple ranking procedures: a Monte Carlo experiment illustrating design and analysis techniques.	februari '76
16.J.P.C. Kleijnen	Computers and operations research: a survey.	februari '76
17.J.P.C. Kleijnen	Statistical problems in the simulation of computer systems.	april '76
18.F.J. Vandamme	Towards a more natural deontic logic.	mei '76
19.J.P.C. Kleijnen	Design and analysis of simulation: practical, statistical techniques.	juni '76
20.H.H. Tigelaar	Identifiability in models with lagged variables.	juli '76
21.J.P.C. Kleijnen	Quantile estimation in regenerative simulation: a case study.	augustus '76
22.W.Derks	Inleiding tot econometrische modellen van landen van de E.E.G.	augustus '76
23.B. Diederen Th. Reijs W. Derks	Econometrisch model van België.	september '76
24.J.P.C. Kleijnen	Principles of Economics for computers.	augustus '76
25.B. van Velthoven	Hybriede simulatie van ekonomische modellen.	augustus '76.

26.F. Cole	Forecasting by exponential smoothing, the Box and Jenkins procedure and spectral analysis. A simulation study.	september '76
27.R. Heuts	Some reformulations and extensions in the univariate Box-Jenkins time series analysis.	juli '76
28.W. Derks	Vier econometrische modellen.	
29.J. Frijns	Estimation methods for multivariate dynamic models.	oktober '76
30.P. Meulendijks	Keynesiaanse theorieën van handelsliberalisatie.	oktober '76
31.W. Derks	Structuuranalyse van econometrische modellen met behulp van Grafentheorie. Deel I: inleiding in de Grafentheorie.	september '76
32.W. Derks	Structuuranalyse van econometrische modellen met behulp van Grafentheorie. Deel II: Formule van Mason.	oktober '76
33. A. van Schaik	Een direct verband tussen economische veroudering en bezettingsgraadverliezen.	september '76
34. W. Derks	Structuuranalyse van Econometrische Modellen met behulp van Grafentheorie. Deel III. De graaf van dynamische modellen met één vertraging.	oktober '76
35. W. Derks	Structuuranalyse van Econometrische Modellen met behulp van Grafentheorie. Deel IV. Formule van Mason en dynamische modellen met één vertraging.	oktober '76
36. J. Roemen	De ontwikkeling van de omvangsverdeling in de levensmiddelenindustrie in de D.D.R.	oktober '76
37. W. Derks	Structuuranalyse van Econometrische modellen met behulp van grafentheorie. Deel V. De graaf van dynamische modellen met meerdere vertragingen.	oktober '76
38. A. van Schaik	Een direkt verband tussen economische veroudering en bezettingsgraadverliezen. Deel II: gevoeligheidsanalyse.	december '76
39. W. Derks	Structuuranalyse van Econometrische modellen met behulp van Grafentheorie. Deel VI. Model I van Klein, statisch.	december '76
40. J. Kleijnen	Information Economics: Inleiding en kritiek	november '76
41. M. v.d. Tillaart.	De spectrale representatie van mutivariate zwak-stationaire stochastische processen met discrete tijdparameter.	november '76
42. W. Groenendaal Th. Dunnewijk	Een econometrisch model van Engeland	december '76
43. R. Heuts	Capital market models for portfolio selection	september '76

44. J. Kleijnen en P. Rens	A critical analysis of IBM's inventory package impact.	december '76
45. J. Kleijnen en P. Rens	Computerized inventory management: A critical analysis of IBM's impact system.	december '76
46. A. Willemstein	Evaluatie en foutenanalyse van econometrische modellen. Deel I. Een identificatie methode voor een lineair discreet systeem met storingen op input, output en structuur.	januari '77
47. W. Derks	Structuuranalyse van econometrische modellen met behulp van grafentheorie. Deel VII. Model I van Klein, dynamisch.	februari '77
48. L. Westermann	On systems of linear inequalities over $\mathbb{R}^n$.	februari '77
49. W. Derks	Structuuranalyse van econometrische modellen met behulp van Grafentheorie. Deel VIII. Klein-Goldberger model.	februari '77
50. W.v. Groenendaal en Th. Dunnewijk	Een econometrisch model van het Verenigd Koninkrijk	februari '77
51. J. Kleijnen en P. Rens	A critical analysis of IBM's inventory package "IMPACT"	februari '77
52. J.J.A. Moors	Estimation in truncated parameter-spaces	maart '77
53. R.M.J. Heuts	Dynamic transfer function-noise modelling (Some theoretical considerations)	december '76
54. B.B. v.d. Genugten	Limit theorems for LS-estimators in linear regression models with independent errors.	mei '77
55. P.A. Verheyen	Economische interpretatie in modellen betreffende levensduur van kapitaalgoederen	juni '77
56. W.v.den Bogaard en J.Kleijnen	Minimizing wasting times using priority classes	juni '77
57. W. Derks	Structuuranalyse van Econometrische Modellen met behulp van Grafentheorie. Deel IX. Model van landen van de E.E.G.	juni '77
58. R. Heuts	Capital market models for portfolio selection (a revised version)	juni '77
59. A.P. Willemstein	Evaluatie en foutenanalyse van econometrische modellen. Deel II. Het Model I van L.R. Klein.	aug. '77
60. Th. Dunnewijk W. van Groenendaal	An econometric Model of the Federal Republic of Germany 1953-1973	aug. '77
61. A. Plaisier A. Hempenius	Slagen of zakken. Een intern rapport over de studieresultaten propedeuse-economie 1974/1975	aug. '77

62. A. Hempenius	Over een maat voor de juistheid van voorspellingen	aug. '77
63. R.M.J. Heuts	Some reformulations and extensions in the univariate Box-Jenkins time series analysis approach (a revised version)	sept. '77
64. R.M.J. Heuts	Applications of univariate time series modelling of U.S. monetary and business indicator data	sept. '77
65. A. Hempenius en J. Frijns	Soorten van prijsheteroskedasticiteit in marktvraagfunkties.	okt. '77
66. H.H. Tigelaar	Identifiability in Multiple Time Series	okt. '77
67. J. van Lieshout en P. Verheyen	Levensduur in een jaargangenmodel	nov. '77
68. Pieter J.F.G. Meulendijks	"De macro-economische betekenis van geïnduceerde technische ontwikkeling; een meer-sektoren model met jaargangentheorie"	nov. '77
69. R.M.J. Heuts en P.J. Rens	A Monte Carlo study to obtain the percentage points of some goodness of fit tests in testing normality, when observations satisfy a certain low order ARMA-scheme	dec. '77
70. J.C.P. Kleijnen	Generalizing Simulation Results through Metamodels	dec. '77
71. Th. van de Klundert	Winstmaximalisatie in het Jaargangenmodel met vaste technische coëfficiënten; een inventarisatie van de problematiek	dec. '77
72. B.B. van der Genugten	A central limit theorem with applications in regression analysis	jan. '78
73. A.P. Willemstein	Evaluatie en foutenanalyse van econometrische modellen. DEEL III Stochastische fluctuaties op de parameters en heteroscedasticiteit in een lineair model	jan. '78
74. Pieter J.F.G.Meulendijks	Some reflections on macro-economic planning, policy and development in the Netherlands: 1918-1978	apr. '78
75. J.H.J. Roemen	Overeenkomsten en verschillen in uitgangspunten tussen lognormale en Pareto-verdeling.	juli '78

:

76.	J.D. Sylwestrowicz	Applications of Hybrid Computers in Econometrics - Part I.	juli '78
77.	Pieter J.F.G. Meulendijks	On a disequilibrium analysis of the labour market. Review of and comments upon R.S.G. Lenderink and J.C. Siebrand, A disequilibrium analysis of the labour market, Rotterdam University Press, '76.	aug. '78
78.	B.B. van der Genugten	Asymptotic normality of 2SLS-estimators in simultaneous equation systems.	aug. '78
79.	J.M.G. Frijns	Imperfect competition of the labour market(s) and the adjustment of factor inputs.	okt. '78
80.	H.H. Tigelaar	The identifiability problem in dynamic simultaneous equations with moving average errors.	nov. '78
81.	Jack P.C. Kleijnen	Bayesian information economics: an evaluation.	dec. '78.
82.	J.M.G. Frijns	Instrumental Variable Estimators Applied on Pooled Time Series Cross Section Models.	maart '79
83.	M.H.C. Paardekooper	Inverse perturbations for approximations of least squares solutions.	maart '79
84.	J.H.J. Roemen	Kostenverbijzondering op basis van een input-output-model voor een situatie met vaste en variabele kosten	maart '79