

Tilburg University

## Beter meten en voorspellen met behulp van panel-gegevens

Wansbeek, T.J.; Kapteyn, A.J.

*Published in:*  
Economisch Statistische Berichten

*Publication date:*  
1978

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*  
Wansbeek, T. J., & Kapteyn, A. J. (1978). Beter meten en voorspellen met behulp van panel-gegevens. *Economisch Statistische Berichten*, 63, 43-47.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Beter meten en voorspellen met behulp van panel-gegevens

DR. IR. A. KAPTEYN\*  
DRS. T.J. WANSBEEK\*

*De recente discussies over de sociaal-economische problemen in Nederland, zoals de werkgelegenheid, de inkomensverdeling en de omvang van de collectieve voorzieningen, hebben eens te meer duidelijk gemaakt dat de gegevens waarop het beleid moet worden gebaseerd in veel opzichten te kort schieten. Verbetering in deze situatie kan, zo wordt in het onderstaande artikel betoogd, worden gebracht door op veel grotere schaal dan tot nu toe gegevens te verzamelen in panel-vorm en bij de analyse van die gegevens gebruik te maken van de mogelijkheden die de zich snel ontwikkelende econometrie van de panel-gegevens biedt. Het artikel geeft eerst een aantal voorbeelden van panel-gegevens die op allerlei plaatsen worden verzameld, waarbij overigens in een aantal gevallen het panel-karakter niet wordt uitgebuit. Vervolgens wordt het belang van dit soort gegevens voor micro- en macro-economisch beleid aannemelijk gemaakt. Ten slotte wordt in het kort ingegaan op de econometrische aspecten van het werken met panel-gegevens, en worden wat recente ontwikkelingen op dit gebied aangestipt. De aanleiding tot dit artikel is het congres over de econometrie van panel-gegevens dat in augustus 1977 in Parijs is gehouden.*

## Wat zijn panel-gegevens?

Men spreekt van panel-gegevens als van alle waarnemings-eenheden (bijvoorbeeld gezinnen, bedrijven, regio's) voor meer dan één tijdsperiode waarnemingen beschikbaar zijn van dezelfde variabelen. De waarnemingseenheden te zamen vormen het z.g. panel. Panel-gegevens bestaan dus uit tijdreeksen van dwarsdoorsnedegegevens. Wellicht ten overvloede: een dwarsdoorsnede is een steekproef van waarnemingseenheden waarvan de variabelen op hetzelfde tijdstip worden gemeten.

Panels hebben het meest toepassing gevonden in het markt-onderzoek. Zo zijn er panels van huisvrouwen die regelmatig rapporteren over onderwerpen als merk en prijs van de goederen die ze kopen en de winkels die ze bezoeken. De informatie die op die manier beschikbaar komt, is van belang voor onder andere de analyse van de effectiviteit van marketingstrategieën. Bekende huisvrouwenpanels in Nederland zijn die van „Attwood Statistics Nederland”, en de „Intomart Consumer Audit”, die beide een paar duizend leden

tellen. Ook voor wetenschappelijk onderzoek leveren dergelijke panels waardevolle informatie 1). Een ander bekend voorbeeld van een panel-onderzoek in Nederland is het kijk- en luisteronderzoek van Intomart. Het panel van televisiekijkers, ongeveer 1.500 deelnemers sterk, rapporteert iedere week welke programma's zij hebben gezien.

Naar wij in dit artikel duidelijk hopen te maken, kunnen met behulp van panel-gegevens resultaten worden verkregen die niet te behalen zijn met tijdreeks- of dwarsdoorsnedegegevens alleen. Speciaal voor dit type gegevens zijn in de econometrie analysemethoden ontwikkeld die van het speciale karakter van die gegevens gebruik maken. Het is dan ook te betreuren dat in nogal wat gevallen panel-gegevens wel beschikbaar zijn, zonder dat hun panel-karakter wordt uitgebuit. Dit geldt bijvoorbeeld voor het eerder genoemde kijk- en luisteronderzoek.

Er worden in Nederland veel meer gegevens op panel-basis vergaard dan de gebruikers veelal beseffen. Het CBS bijvoorbeeld baseert een deel van zijn omzetstatistieken op steekproeven van bedrijven die gedurende een reeks van jaren worden gevolgd. Naar in het vervolg zal blijken, werpen analyses vaak meer vrucht af indien in plaats van geaggregeerde cijfers gedesaggregeerde cijfers worden gebruikt. Het is de moeite waard na te gaan in hoeverre (met inachtneming van de geheimhoudingsplicht) het mogelijk is de gedesaggregeerde gegevens aan gebruikers ter beschikking te stellen zodat diepergaande analyses mogelijk zijn dan op grond van de geaggregeerde cijfers mogelijk is.

Het verzamelen en analyseren van gegevens voor het voorbereiden van sociaal-economisch beleid is in de Verenigde Staten verder ontwikkeld dan in ons land. Mede doordat de groeiende algemene welvaart niet voldoende bleek te zijn om alle sociaal-economische problemen (met name werkloosheid en armoede) op te lossen, is men daar al vroeg gaan inzien dat gedetailleerde kennis van maatschappelijke processen nodig is. Nadat men zich eerst nog tevreden stelde met onderzoek gebaseerd op uitsluitend dwarsdoorsneden is men langzamerhand de noodzaak gaan inzien van het verwerven van gegevens op panel-basis. Befaamde panel-onderzoeken met een beleidsgericht karakter zijn onder andere het „New Jersey negative income tax experiment” (gehouden van augustus 1968 tot en met september 1972) en de „Seattle and Denver income maintenance experiments” (aangevangen in november 1971) 2). In deze experimenten worden met groepen

\* Economisch Instituut, Rijksuniversiteit Leiden.

1) Een voorbeeld is het Wageningse proefschrift van Dr. B. Wierenga, *An investigation of brand choice processes*, Rotterdam University Press, Rotterdam, 1974.

2) Meer gegevens met betrekking tot deze experimenten zijn onder andere te vinden in: O. Kershaw, J. Fair, *The New Jersey income-maintenance experiment, Vol. I*, Academic Press, New York, 1976 en M. Kurz en R. G. Spiegelman, *The Seattle experiment: the combined effect of income maintenance and manpower investment*, *American Economic Review*, vol. 61, 1971, papers and proceedings, blz. 22-29.

gezinnen contracten gesloten die uitkeringen garanderen in geval van werkloosheid. Door met verschillende gezinnen verschillende contracten te sluiten en het arbeidsmarktgedrag van gezinsleden in de tijd te volgen, krijgt men een indruk van bijvoorbeeld het effect van de uitkeringshoogte op de werkloosheid.

Andere panel-onderzoeken zijn de „Panel study of income dynamics” (aangevangen in 1968) van het Survey Research Center van de Universiteit van Michigan (Ann Arbor) en de „National longitudinal surveys” (sinds het midden van de jaren zestig) van het Ohio State University Center for Human Resource Research. Met name op grond van de Michigan-gegevens zijn inmiddels resultaten van tientallen onderzoeken gepubliceerd in de internationale wetenschappelijke bladen.

Dat men ook buiten de Verenigde Staten steeds meer de unieke analysemogelijkheden van panel-gegevens gaat inzien, blijkt uit een onlangs door het INSEE (het Franse CBS) georganiseerde symposium over „The econometrics of panel data” dat van 22 tot en met 24 augustus 1977 in Parijs werd gehouden. Door de ongeveer 35 deelnemers werden de recente ontwikkelingen op dit gebied besproken. Aan het eind van dit artikel vindt men een lijst van de daar gehouden lezingen. Medio volgend jaar zullen de „proceedings” van dit congres door het Centre National de la Recherche Scientifique worden uitgegeven.

#### Het gebruik van panel-gegevens voor sociaal-economisch beleid

Panel-gegevens kunnen een oplossing bieden in twee gevallen: ten eerste in het geval dat men informatie zoekt die zonder panel-gegevens niet te achterhalen is, en ten tweede in het geval dat men de gezochte informatie wel op andere wijze kan verkrijgen, maar met minder nauwkeurigheid of tegen hogere kosten. We zullen beide gevallen kort toelichten.

Het eerste geval treedt onder andere op als men welvaartsaspecten van de inkomensverdeling wil bestuderen. Het is dan niet alleen van belang te weten dat in een gegeven jaar een zeker percentage van bijvoorbeeld de zelfstandigen beneden het minimumloon zit. Een dergelijk gegeven krijgt meer perspectief als men weet in hoeverre dat percentage ieder jaar dezelfde mensen betreft. Iets dergelijks geldt ook voor het verschijnsel werkloosheid. De evaluatie van een werkloosheidspercentage wordt mede bepaald door de vraag of de populatie van werklozen snel wisselt dan wel min of meer constant blijft. In het eerste geval behoeft men aan het sociale aspect van werkloosheid minder zwaar te tillen dan in het tweede geval. Betrouwbare informatie over dit soort vragen kan men alleen met behulp van panel-gegevens verkrijgen.

Een ander voorbeeld van een situatie waarin panel-data noodzakelijk zijn, is dat van een model waarin variabelen optreden die niet of niet zonder fout te meten zijn, maar die wel door de tijd heen constant zijn. Een dergelijk meetprobleem kan ontstaan als het moeilijk is de variabele in kwestie goed te meten, of als het principieel onmogelijk is dat te doen omdat er geen direct waarneembare grootheid is die met een bepaalde variabele in het model correspondeert. Dit laatste treedt bijvoorbeeld op in het volgende geval.

Een onderzoeker wil voor melkveehouderijbedrijven een produktiefunctie schatten, dat wil zeggen een verband bepalen tussen enerzijds de hoeveelheid produkt van een bedrijf, en aan de andere kant de verschillende „inputs”, zoals de hoeveelheid arbeid, de verzameling kapitaalgoederen en de kwaliteit van de bedrijfsleiding. Met enige fantasie nemen we aan dat het voor de eerste twee produktiefactoren nog wel mogelijk is om per bedrijf de hoeveelheden te meten. De kwaliteit van de bedrijfsleiding is echter moeilijk rechtstreeks te meten.

Voor de hand liggende oplossingen in een dergelijk geval zijn het dan maar weglaten van zo'n lastige variabele, of het invullen van een „proxy”, een andere variabele die er een beetje op lijkt. De econometrische theorie leert echter dat

beide oplossingen leiden tot inconsistente schattingen van de coëfficiënten in het model en dat betekent in het voorbeeld van de melkveehouderij dat het effect van de bedrijfs grootte op het bedrijfsresultaat wordt overschat, omdat (naar alle waarschijnlijkheid) de „goede” bedrijfsleider geneigd is relatief veel kapitaal en arbeid in te zetten<sup>3)</sup>. Als gevolg hiervan zou een landbouwbeleid gericht op verbetering van de inkomenspositie van boeren ten onrechte bedrijfsvergroting kunnen stimuleren en pas na verloop van jaren zou men ontdekken dat het beoogde doel niet wordt bereikt.

De beschikbaarheid van panel-gegevens brengt hier uitkomst, want de onderzoeker kan nu het volgende model specificeren:

$$x_{it} = \alpha \cdot a_{it} + \kappa \cdot k_{it} + \beta + \mu_i + u_{it}, \quad (1)$$

waarin  $x_{it}$ ,  $a_{it}$  en  $k_{it}$  de (logaritmen van) hoeveelheden produkt, arbeid en kapitaal zijn van bedrijf  $i$  in jaar  $t$ ;  $\alpha$ ,  $\beta$  en  $\kappa$  zijn te schatten parameters en  $u_{it}$  is de storingsterm, terwijl  $\mu_i$  de onbekende kwaliteit van de leiding van bedrijf  $i$  vertegenwoordigt. Als we beschikken over waarnemingen over meer dan één periode en als we veronderstellen dat de  $\mu$ 's in de verschillende perioden niet veranderen, kunnen de  $\mu$ 's tegelijk met  $\alpha$ ,  $\beta$  en  $\kappa$  door middel van regressierekening worden geschat. In dit geval zijn de schattingen van de coëfficiënten wel consistent (en ook zuiver).

Met het oog op de discussie verderop is het nuttig erop te wijzen dat de volgende uitbreiding van model (1) voor de hand ligt:

$$x_{it} = \alpha \cdot a_{it} + \kappa \cdot k_{it} + \beta + \mu_i + \gamma_t + u_{it}. \quad (2)$$

$\gamma_t$  vertegenwoordigt het jaareffect, dat wil zeggen de invloed van voor alle bedrijven in een bepaald jaar gelijkelijk geldende, maar niet direct kwantificeerbare omstandigheden als „de conjunctuur”, „de afzetmogelijkheden” of „het weer”.

Zoals gezegd, zijn er ook situaties waarin panel-gegevens niet strikt noodzakelijk zijn, maar wel goedkopere en nauwkeuriger informatie geven dan een tijdreeks van verschillende dwarsdoersneden. Stel bijvoorbeeld dat men de verandering wil meten in de waardering die mensen voor hun eigen inkomen hebben. Indien men dat doet door telkens nieuwe steekproeven te trekken, is één van de foutenbronnen dat toevallige verschillen in de streekproefsamenstelling verschillen in gemeten verdelingen van individuele waarderungen veroorzaken. Ten einde het effect van deze foutenbron zo klein mogelijk te houden, heeft men betrekkelijk grote steekproeven nodig. Wanneer een panel gebruikt wordt, is deze foutenbron afwezig en kan bij dezelfde gewenste nauwkeurigheid worden volstaan met een kleinere steekproef, of kan bij eenzelfde steekproefomvang een grotere nauwkeurigheid worden bereikt.

#### Het belang van panels voor macro-economisch beleid

Macro-economische modellen beschrijven relaties tussen geaggregeerde grootheden. Meestal worden deze modellen geschat op grond van tijdreeksen van geaggregeerde grootheden. In het nu volgende, sterk vereenvoudigde, voorbeeld laten we zien dat macro-economische voorspellingen aan nauwkeurigheid kunnen winnen door het gebruik van panel-gegevens.

Bezie een maatschappij bestaande uit twee individuen. Het consumptieniveau van deze individuen in jaar  $t$  wordt beschreven door

$$c_{it} = a_i y_{it} + b_i y_{1,t-1} + d_i. \quad (3)$$

3) Zie onder andere L.C. Zachariasse, *Boer en bedrijfsresultaat*, dissertatie, Wageningen, 1974.

$$c_{2t} = a_2 y_{2t} + b_2 y_{2,t-1} + d_2, \quad (4)$$

waarin  $c_{1t}$  en  $c_{2t}$  de consumptieniveaus van individuen 1 en 2 zijn in jaar  $t$ ;  $y_{1t}$  en  $y_{2t}$  zijn hun respectieve inkomens in datzelfde jaar;  $y_{1,t-1}$  en  $y_{2,t-1}$  zijn de inkomens in het jaar daarvoor. De grootheden  $a_1$ ,  $b_1$ ,  $d_1$ ,  $a_2$ ,  $b_2$  en  $d_2$  zijn parameters. Men kan de vergelijkingen (3) en (4) beschouwen als een zeer vereenvoudigde versie van Friedmans „permanent income hypothesis“: het consumptieniveau van een individu wordt geacht zowel af te hangen van zijn huidige inkomen als van zijn inkomen in het vorig jaar.

De geaggregeerde consumptie  $C_t$  in jaar  $t$  in de tweepersons maatschappij is uiteraard gedefinieerd als

$$C_t \equiv c_{1t} + c_{2t}. \quad (5)$$

Evenzo is het geaggregeerde inkomen  $Y_t$  gedefinieerd als

$$Y_t \equiv y_{1t} + y_{2t} \quad (6)$$

Definieer nu de aandelen van de individuen 1 en 2 in het geaggregeerde inkomen,  $w_{1t}$  en  $w_{2t}$ , door

$$w_{1t} \equiv \frac{y_{1t}}{Y_t} \quad (7)$$

$$w_{2t} \equiv \frac{y_{2t}}{Y_t}. \quad (8)$$

De geaggregeerde consumptiefunctie is dan

$$C_t \equiv (a_1 w_{1t} + a_2 w_{2t}) Y_t + (b_1 w_{1,t-1} + b_2 w_{2,t-1}) Y_{t-1} + d_1 + d_2 \quad (9)$$

Men ziet dat in (9) de coëfficiënten van  $Y_t$  en  $Y_{t-1}$  afhangen van de inkomensverdeling in de jaren  $t$  en  $t-1$ . Indien de inkomensverdeling in de loop van de tijd constant blijft, is het gerechtvaardigd de relatie (9) te schatten met behulp van geaggregeerde cijfers. Men kan vervolgens (9) gebruiken om, onder zekere veronderstellingen ten aanzien van het verloop van het geaggregeerde inkomen in de toekomst, de geaggregeerde consumptie te voorspellen. Mocht echter in de voorspelperiode de inkomensverdeling veranderen, dan verliest het model zijn voorspellende waarde.

Met andere woorden: zolang de structuur van een economie niet te veel verandert (in het voorbeeld: de inkomensverdeling), dan zal voorspelling met geaggregeerde tijdreeksen redelijk nauwkeurige uitkomsten leveren. In die situatie is het voorspellen echter niet zo interessant. De overheid wil macro-economische modellen met name gebruiken om beleid te voeren. Maar: het voeren van beleid impliceert meestal dat de structuur van een economie wordt beïnvloed (denk aan de vrij sterke verschuiving welke in de afgelopen kabinetsperiode is opgetreden in de inkomensverdeling).

Ons voorbeeld laat zien dat in veel gevallen, waarin vanuit beleidsoogpunt aan een modelmatige voorspelling het meest behoefte is, een op geaggregeerde gegevens gebaseerd macro-economisch model het zal laten afweten!

Het is interessant na te gaan in hoeverre in ons voorbeeld een dwarsdoorsnede soelaas zou kunnen bieden. Een dwarsdoorsnede lijkt bij uitstek geschikt om effecten van een inkomensverdeling te kwantificeren. Bij dit type gegevens echter ontbreekt kennis van  $y_{1,t-1}$  en  $y_{2,t-1}$ , zodat het meten van de relaties (3) en (4) onmogelijk wordt (4). Veelal worden in zo'n geval de termen  $b_1 y_{1,t-1}$  en  $b_2 y_{2,t-1}$  in (3) en (4) weggelaten. Men verkrijgt dan echter onzuivere schattingen van  $a_1$  en  $a_2$ , en voorspellingen op grond van een dwarsdoorsnede zijn daarmee hoogst dubieus geworden.

De ideale oplossing voor de geschetste problemen is ieder jaar een volledige telling te houden. Vanzelfsprekend is zo'n operatie zeer kostbaar, terwijl de ervaring met onder andere de volkstelling leert dat de relevante gegevens eerst met een

vertraging van een aantal jaren ter beschikking komen. De enige haalbare oplossing is ook hier het opzetten van een panel-onderzoek. Men kan dan betrekkelijk goedkoop alle relevante informatie vergaren en die bovendien snel ter beschikking hebben.

Het bovenstaande voorbeeld heeft betrekking op het meten van een macro-economische consumptiefunctie. Op andere gebieden zijn analoge voorbeelden te bedenken. Naar eerder werd opgemerkt, verzamelt het CBS een gedeelte van zijn gegevens via panels van bedrijven, doch worden de resultaten alleen in geaggregeerde vorm gepubliceerd. Ook hier ligt het voor de hand dat een analyse van de originele panelgegevens tot betrouwbaarder voorspellingen aanleiding zou geven dan thans beschikbaar zijn.

Ons vermoeden vindt onder andere steun in een recent onderzoek van Prof. Donald D. Hester van de Universiteit van Wisconsin-Madison 5). Hester poogt de samenstelling en omvang van de activa van Amerikaanse handelsbanken te verklaren en te voorspellen. Hij heeft daartoe zowel de beschikking over panel-gegevens van 320 Amerikaanse handelsbanken gedurende 287 weken als over geaggregeerde cijfers. Hester gebruikt zijn materiaal vervolgens om de samenstelling van de activa van alle Amerikaanse handelsbanken te voorspellen. In het begin van de voorspelperiode van twee jaar blijkt een op de geaggregeerde cijfers gebaseerd model beter te voorspellen. Dat ligt voor de hand daar de parameters in het geaggregeerde model worden geschat volgens hetzelfde criterium dat ook wordt gebruikt om voorspellingen te evalueren. Aldus krijgt het geaggregeerde model een kunstmatige voorsprong op het model dat op panel-gegevens is gebaseerd. Naarmate men echter probeert wat verder vooruit te voorspellen (verder dan één jaar) blijkt het geaggregeerde model slechter te gaan voorspellen, terwijl het panel-model qua voorspelkwaliteit constant blijft en al spoedig superieur wordt aan het geaggregeerde model. Hester merkt op dat dergelijke resultaten waarschijnlijk ook verkregen zouden worden in andere sectoren van de economie dan het bankwezen. Als dat vermoeden juist is, zou het op grote schaal gebruik maken van panel-gegevens kunnen leiden tot een drastische verbetering van macro-economische voorspellingen op termijnen langer dan een jaar.

#### Iets over analysemethoden

Voordat we dit artikel besluiten met een zeer korte beschrijving van de recente ontwikkelingen op het gebied van de econometrie van panel-gegevens, vertellen we eerst iets over de methoden om deze gegevens te analyseren.

Voor de eenvoud nemen we aan dat de onderzoeker is geïnteresseerd in het volgende causale verband tussen twee grootheden  $x$  en  $y$ :

$$y = \alpha x + \beta, \quad (10)$$

en dat hij op grond van panel-gewijze waarnemingen over  $x$  en  $y$  de parameters  $\alpha$  en  $\beta$  zo goed mogelijk wil schatten. Voor bedrijf  $i$  in jaar  $t$  wordt dan gepostuleerd:

$$y_{it} = \alpha x_{it} + \beta + u_{it} \quad (11)$$

waarin  $u_{it}$  weer de storingsterm is. Een gebruikelijke veron-

4) Men zou kunnen overwegen de respondenten in de steekproef te vragen wat hun inkomen in het vorig jaar was. Het is evenwel bekend dat dit soort retrospectieve informatie onbetrouwbaar is. Dat geldt *a fortiori* voor de informatie die nodig zou zijn voor realistischere modellen dan (3) en (4).

5) Zie de lijst van rapporten aan het eind van dit artikel.

6) In de voorbeelden zullen we voornamelijk spreken over bedrijven, maar in het geval van andere waarnemingseenheden (individuen, regio's, gezinnen enz.) kunnen we volstrekt analoge modellen opstel-

derstelling over de  $u$ 's is dat ze onderling ongecorrleerd zijn. Als dat zo is, leidt de toepassing van de methode van de kleinste kwadraten tot nauwkeurige schattingen van  $\alpha$  en  $\beta$ . Maar in veel gevallen is dit model niet erg realistisch: behalve  $x_{it}$  zijn op  $y_{it}$  nog allerlei jaar- en bedrijfsspecifieke effecten werkzaam. De volgende modificatie ligt dan ook voor de hand:

$$y_{it} = \alpha x_{it} + \beta + \mu_i + \gamma_t + u_{it}, \quad (12)$$

waarin  $\mu_i$  het bedrijfsspecifieke, en  $\gamma_t$  het jaarspecifieke effect vertegenwoordigt.

Wat die effecten zijn, hoeft ons niet per se te interesseren. We hebben in de paragraaf over het gebruik van paneelgegevens voor sociaal-economisch beleid echter gezien dat het negeren van de  $\mu$ 's en de  $\gamma$ 's tot inconsistente schattingen van  $\alpha$  en  $\beta$  kan leiden. Het toepassen van de methode van de gewone kleinste kwadraten op het model (12) levert onmiddellijk consistente schattingen op voor alle parameters ( $\alpha$ ,  $\beta$ ,  $\mu_1, \dots, \mu_I$ ;  $\gamma_1, \dots, \gamma_T$ ) in het model. Deze veelgebruikte methode van schatten staat bekend als *covariantie-analyse*.

Wanneer de onderzoeker alleen is geïnteresseerd in de waarden van  $\alpha$  en  $\beta$ , zal deze methode hem maar matig bevallen: de meeste schattingen, die van de  $\mu$ 's en de  $\gamma$ 's, interesseren hem niet, maar krijgt hij toch gepresenteerd. Deze redundantie heeft zijn prijs;  $\alpha$  en  $\beta$  kunnen nauwkeuriger worden geschat door af te zien van de informatie omtrent de  $\mu$ 's en  $\gamma$ 's. Deze worden dan niet beschouwd als vaste parameters, maar als kansgrootheden en dus als onderdeel van de storingsterm. Die wordt dan  $\mu_i + \gamma_t + u_{it}$ , en krijgt aldus een zekere structuur. Door met deze structuur expliciet in de schattingsmethode rekening te houden (bijvoorbeeld via de methode van de gegeneraliseerde kleinste kwadraten), kan men  $\alpha$  en  $\beta$  nauwkeuriger schatten dan met behulp van covariantie-analyse. Deze methode staat bekend als die van

de *variantie-componenten* („error-components”).

De variantie-componentenmethode is alleen toegestaan wanneer de  $\mu$ 's en de  $\gamma$ 's niet correleren met de  $x$ -en. Als dat wel het geval is, verdient covariantie-analyse de voorkeur.

We zien dus dat er twee manieren zijn om bedrijfsspecifieke en jaarspecifieke effecten te behandelen. Als men die effecten wil weten, past men covariantie-analyse toe en beschouwt men de specifieke effecten als vast en schat die met behulp van regressierekening op het model (12). Men spreekt in dat geval van een „fixed effects” (FE)-model. Wanneer men alleen is geïnteresseerd in  $\alpha$  en  $\beta$ , beschouwt men de specifieke effecten als onderdeel van de storingsterm en gebruikt de structuur die daardoor aan de storingsterm wordt opgelegd om  $\alpha$  en  $\beta$  nauwkeuriger te schatten. Deze laatste benadering beschouwt de specifieke effecten als kansvariabelen en dan noemt men model (12) een „random effects” (RE)-model.

### Enige recente ontwikkelingen

Model (12) kan ook worden opgevat als een model waarbij de constante term wordt opgesplitst in een jaarspecifiek, een bedrijfsspecifiek en een algemeen deel. Men kan zich voorstellen dat ook de parameter  $\alpha$  in de tijd en per bedrijf varieert. Daarmee ontstaat een structuur waarbij zowel  $\alpha$  als  $\beta$  uit drie componenten zijn opgebouwd: één component die per jaar verschillend is, één component die per bedrijf verschillend is en een component die constant is voor alle bedrijven voor alle jaren. Wederom kunnen dan twee soorten veronderstellingen worden gemaakt: wanneer de bedrijfsspecifieke en jaarspecifieke componenten als vaste parameters worden beschouwd, spreekt men weer van een „fixed effects” (FE)-model; wanneer zij als trekkingen uit een kansverdeling worden beschouwd, spreekt men weer van een „random effects” (RE)- of ook „random coefficients” (RC)-model.

Met name Swamy heeft veel bijgedragen aan de ontwikkeling van RC-modellen 7). De voordelen van RE-modellen boven FE-modellen (zie boven; ruwweg: RE-schattingen zijn „zuiniger“) hebben ertoe geleid dat de laatste jaren vooral de RE-achtige methoden verder zijn ontwikkeld. Daarbij zijn twee richtingen aan te wijzen. In de eerste plaats worden de econometrische methoden om maximaal profijt te trekken van panel-gegevens verder verfijnd. Daarnaast is er een toenemende aandacht voor de problemen die ontstaan bij het in de praktijk gebruiken van panels. We zullen op beide kort ingaan.

De econometrische technieken om panel-gegevens optimaal te benutten richten zich op drie aspecten. In de eerste plaats poogt men een zo algemeen mogelijke formulering te geven van het type modellen dat op grond van panel-gegevens kan worden gehanteerd. Voorbeelden van dergelijke studies zijn de congresbijdragen van Jöreskog, Wansbeek en Kapteyn, Mazodier en Trognon, en Harvey (zie de lijst aan het eind van dit artikel). In de tweede plaats wordt onderzoek gedaan naar de statistische eigenschappen van allerlei schattingsmethoden voor dit type modellen; zie onder andere de bijdragen van Trognon, Glejser en Jacquemin, en Mundlak. In de derde plaats worden computerprogramma's ontwikkeld die de drempel voor het in de praktijk gebruiken van panel-modellen lager maken. Zo heeft Jöreskog laten zien dat zijn bekende programma LISREL kan worden gebruikt voor het schatten van sommige panel-modellen. In Harvard wordt door Hall algemene programmatuur ontwikkeld.

Op het congres werd ook een groot aantal toepassingen besproken. Een belangrijke bijdrage aan de empirie van de volledige vraagstelsels werd geleverd door Salvas-Bronsard, die liet zien hoe het gebruiken van panel-gegevens in de analyse van het consumentengedrag een oplossing biedt voor het aggregatieprobleem. Een groot aantal lezingen was gewijd aan bedrijfstatonderzoek, zoals die van Eisner, Mairesse en Lipsey. De lezing van Paelinck, die liet zien welke complicaties optreden bij het in de ruimtelijke economie toepassen van panel-methoden, trok sterk de aandacht.

Uiteraard brengt het gebruik van panel-gegevens ook zijn eigen specifieke problemen mee. Wanneer panel-gegevens worden verzameld door een groep mensen of bedrijven van tijd tot tijd te enquêteren, ontstaan gewoonlijk complicaties omdat niet alle leden van een panel steeds meedoen. Meestal is een panel aan een zekere slijtage („attrition“) onderhevig. Analyses die alleen zijn gebaseerd op volledige waarnemingen over alle jaren, leiden tot een vertekening in de uitkomsten wanneer de non-respons niet willekeurig optreedt, maar een verband vertoont met variabelen die voorkomen in het model dat de onderzoeker op zijn panel-gegevens toepast. Dit probleem kan in principe worden opgelost door het model uit te breiden met een of meer relaties waarin dat verband expliciet wordt gemaakt 8). Door het aldus uitgebreide model te schatten kan de „attrition bias“ worden opgeheven. Voorbeelden van studies hiervan zijn te vinden in de congresbijdragen van Maddala, Griliches, Hall en Hausman, en Chamberlain. Wat ruw geformuleerd kan men zeggen dat bij het gebruik van panel-gegevens de mogelijkheid bestaat een oplossing te vinden voor het bij sociaal-economisch onderzoek maar al te bekende probleem van de non-respons.

## Conclusie

In dit artikel hebben we willen laten zien welke voordelen het werken met panel-gegevens kan opleveren. We hebben gepoogd een idee te geven van de mogelijkheden van panel-analyse bij beleidsondersteunend onderzoek, en van de methoden die bij het onderzoek van panel-data worden gebruikt 9).

Natuurlijk is het panel geen panacee voor alle sociaal-economische kwalen. Maar er is een aantal gebieden waarop het gebruik van panels een verbetering zou betekenen ten opzich-

te van de huidige praktijk, waarin men (noodgedwongen) dwarsdoorsneden of geaggregeerde tijdreeksen gebruikt. Voor veel doeleinden is het panel goedkoper, efficiënter en soms onmisbaar.

Hoewel in Nederland nogal wat cijfers in panel-vorm beschikbaar zijn, worden zij zelden als zodanig geanalyseerd. Dat is jammer. Anderzijds zijn er gebieden waarop panel-gegevens niet bestaan en waar het wenselijk is dat er ook in ons land voor sociaal-economische beleidsdoelinden panels worden opgezet.

A. Kapteyn  
T.J. Wansbeek

## Lijst van congresbijdragen

- M. Atkinson en J. Mairesse, *Length of life of equipment in French manufacturing industries: tentative estimates based on individual time series.*  
M.J. Boskin en M.O. Hurd, *The effect of social security on early retirement.*  
G. Chamberlain, *Omitted variable bias in panel data: estimating the returns to schooling.*  
J.C. Chevaillier, J. Paelinck, H. Smit en H. Stijnen, *Parameter-component models in spatial econometrics.*  
R. Eisner, *Cross section and time series estimates of investment functions.*  
H. Glejser en J.-c. Jacquemin, *Truncation remainders in small sample panel models with distributed lags and auto-correlated disturbances.*  
M. Gort en Rao Singamsetti, *Deriving the personality profiles of firms from published panel data.*  
Z. Griliches, B.W. Hall en J.A. Hausman, *Missing data and self-selection in large panels.*  
B. H. Hall, *A general framework for time series-cross section estimation.*  
A.C. Harvey, *The estimation of time-varying parameters from panel data.*  
J.J. Heckman, *Statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence.*  
P.-Y. Henin, *Effets individuels et temporels dans le comportement de distribution de dividendes des entreprises. Essai de spécification et d'estimation.*  
D. Hester, *Dynamic portfolio behavior and market clearing by weekly reporting banks.*  
A. Jonsson en A. Klevmarke, *On the relationship between cross-sectional and cohort earnings profiles.*  
K.G. Jöreskog, *An econometric model for multivariate panel data.*  
L. A. Lillard, *Estimation of permanent and transitory response function in panel data: A dynamic labor supply model.*  
R. Lipsey, *The creation of a microdata set for enterprises and establishments.*  
G.S. Maddala, *Selectivity problems in longitudinal data.*  
J. Mairesse, *Technical progress in French manufacturing industries 1966-1975: a study on individual time series.*  
P. Mazodier en A. Trognon, *Heteroscedasticity and stratification in error components models.*  
Y. Mundlak, *Models with variable coefficients — integration and extension.*  
G. Oudiz, *Investment behavior of French industrial firms: a study on panel data.*  
L. Salvas-Bronsard, *Estimating systems of demand equation from French time-series of cross-sections data.*  
D. Shapiro en F.L. Mott, *Labor force attachment during the early childbearing years: An econometric study based on the national longitudinal surveys of young women.*  
J.P. Smith en F. Welch, *The „new depression“ in higher education: ulcers or indigestion?*  
A. Trognon, *Miscellaneous asymptotic properties of ordinary least squares and maximum likelihood estimators in dynamic error component models.*  
T.J. Wansbeek en A. Kapteyn, *The separation of individual variation and systematic change in the analysis of panel survey data.*

7) P.A.V.B. Swamy, *Statistical inference in random coefficients regression models*, Springer, Berlijn, 1971.

8) Men kan ook het panel aanvullen met nieuwe leden die op de uitvallers lijken, maar dit tast uiteraard het panel-karakter van de cijfers aan, waardoor de bestaande analysemethoden niet meer kunnen worden toegepast.

9) Wie zich ter zake verder wil oriënteren, kan ook terecht in een aantal recente econometrische leerboeken, zoals J. Kmenta, *Econometrics*, McMillan, New York, 1971, R. S. Pindyck en D. L. Rubinfeld, *Econometric models and economic forecasts*, McGraw-Hill, New York, 1976 en G. S. Maddala, *Econometrics*, McGraw-Hill, 1977.