Panel attrition: earlier Dutch experiences with relevance for current practice

Henk Meurs, Radboud University Nijmegen-MuConsult

Radboud Universiteit

Panels: promising paradise for researchers

Panel data: observations on large N and small T
track individuals over time

Many advantages, well-known (Kitamura,1990; Meurs, 1991) :

- Smaller sample sizes to detect changes (Fixed Effects models)
  Identify (unobserved) heterogeneity across individuals (RE models)
- Analyse dynamics: adjustment processes (dynamic models)
- Allow for unraveling ‘cause and effects’ (simultaneous models)
- Stability of elasticities over time (varying parameter models)

These advantages led to institution of the Dutch Mobility Panel (1984-1989):

- Assess effects of fare increases in Public Transport
- Understanding the dynamics in travel behavior after recession 1980s

BUT……..

Mobility was declining (too much…….)
(not plausible compared to RC-data from Bureau of of Statistics)

Wave	Total	BTM	Car
1	100	100	100
3	94	88	92
5	94	94	96
7	89	92	101
change	-11%	-8%	+1%

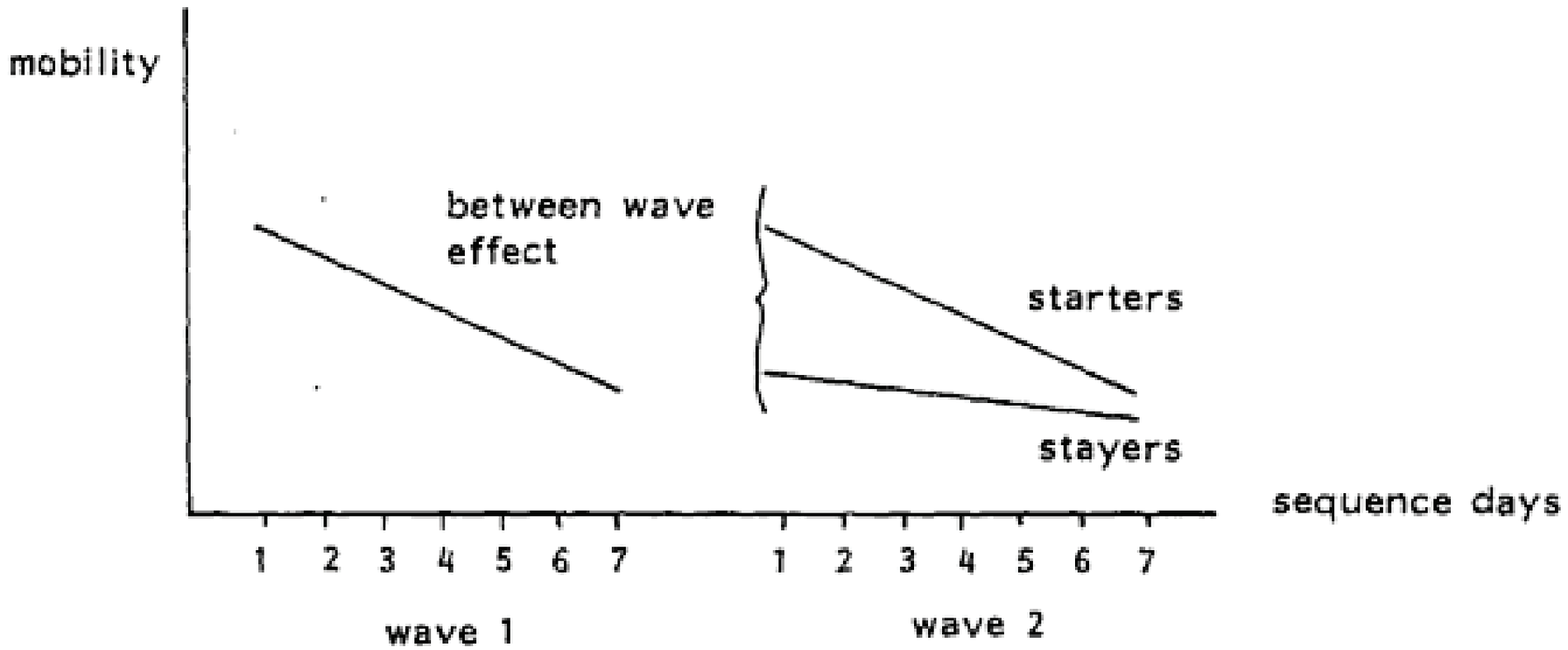
Hypothesis explaining declining trends

- **Reporting errors:**
  - Omitting reporting of certain trips
  - Respondent behavior: fatique and motivation (lots of work)
  - May differ among days and waves→incorrect levels in weekly totals and trends

  Cannot be distinguished from panel conditioning. Changes in behavior and/or attitudes due to panel participation.

- **Attrition:** respondents stop participating, usually non-random
  - Smaller sample sizes
  - Selection problems:
    - Trends do not reflect changes in population
    - Bias in parameters of models

Effects of reporting errors



Source: Meurs et al, 1989)

Effects of reporting errors on mobility per wave (weekly totals)

	Observed				Corrected			
Wave	1	3	5	7	1	3	5	7
Total trips	25.42	23.82	23.77	22.57	27.69	29.52	31.07	30.92
Segments	29.70	27.32	26.85	25.84	31.60	32.80	33.99	34.07
Travel time	235.21	228.83	238.98	216.96	261.66	297.05	236.62	317.39
Modes								
BTM	0.84	0.74	0.79	0.77	0.95	0.85	0.91	0.89
Train	0.31	0.33	0.29	0.29	0.31	0.33	0.29	0.29
Car driver	7.89	7.33	7.60	8.03	8.44	8.50	9.05	9.67
Car passenger	3.16	2.94	2.94	3.17	3.26	3.38	3.54	3.88
Bicycle	8.76	8.14	7.60	6.44	9.57	10.00	9.94	9.10
Walk	6.66	5.97	5.87	5.55	7.69	8.60	9.24	9.39

Source: Meurs et al,

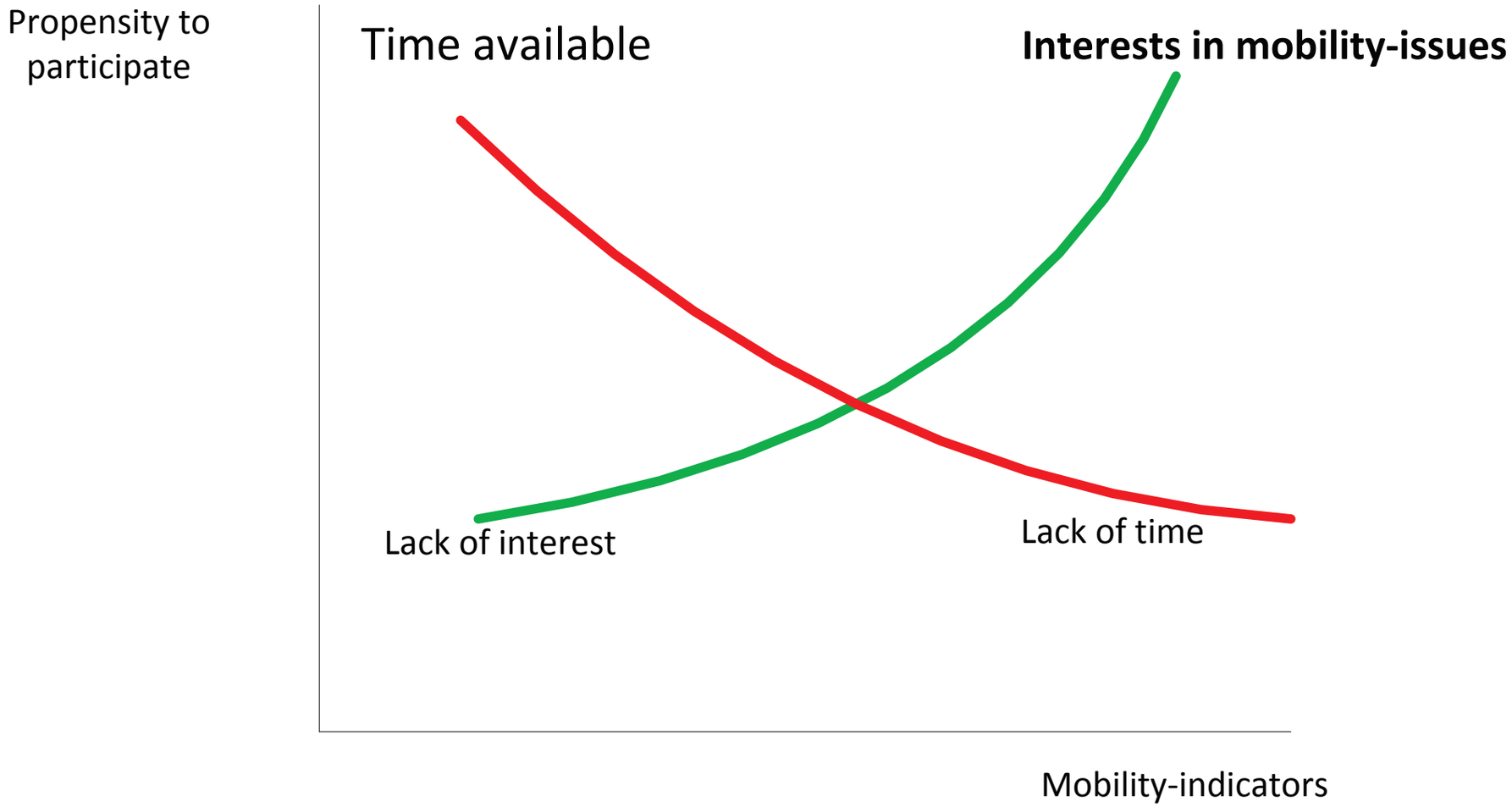
Attrition as a source of biases (1)

- Attrition: not all respondents participate in all waves: unbalanced panels
- A sample selection problem arises in case attrition is based on factors that are related to variable of interest, even after conditioning on explanatoty variables such as age, income or eduction
- Reasons:
  - The field agency did not get into contact with the target person.
  - The target person may be ill and therefore not able to respond.
  - The target person is no longer willing to cooperate
- Distinguish from panel attrition:
  - Demographic losses due to death.
  - People no longer belong to the sampling-frame
  - Losses induce by sampling design: rotation groups with restricted participation length or “nonsample” persons who are not followed if they separate from the “sample” persons

Attrition as a source of biases (2)

- In that case the mechanism that causes missing observations must be taken into account
- In DMP:
- Higher dropout:
  - Low income families,
  - families with a head with little education,
  - households living in a large city,
  - households with more than one car
  - households with few driving licenses

Decision making wrt participation in surveys
an economic model of trade-offs



Note selectivity-effects wrt response-increasing strategies!!!

Consequences of nonrandom selection

- In the econometric literature attrition biases is associated with selection bias (Heckman, 1979)
  Let a^* be the latent propensity to participate and y be a mobility variable

$$a^* = \gamma_0 + \gamma_1' x + \gamma_2' z - v$$

$$a = 1 \qquad \text{if } a^* \geq 0$$

$$a = 0 \qquad \text{if } a^* < 0.$$

The conditional mean of y given x in the sample is given by

$$E(y \mid x, a = 1) = \beta_0 + \beta_1' x + E(u \mid v \leq \gamma_0 + \gamma_1' x + \gamma_2' z, x).$$

Failure to take the conditional mean of u into account will:

- Give wrong estimates of trends
- Bias coefficients β if correlated with x

Approaches to attrition

- **Testing** for consequences of attrition, see Nijman and Verbeek (1992)
  - They proposed simple methods to assess consistency of most widely used estimators in panel data

  -→important: not always problems with regression coefficients, although trends appear to be incorrect

- How to **correct** for non-respons (econometric approaches):
  - Hausman and Wise (1979): first econometric model using “Heckman-type”
  - Ridder (1992) extending the HW-model making a distinction between permanent mobility levels (heterogeneity) and random shocks.
  - Meurs and Ridder (1997) used a random effects stochastic censoring model and exploratory methods to estimate the size of the attrition bias in the DNM
  - Hirano, Imbens, Ridder and Rubin (2001): general approach using refreshment samples taking duration of participation into account:
    - Allows for more precise estimates
    - Allows for more general models, less assumptions required
    - Provides additional information about the attrition process

Some results of attrition models of DMP (1)

- GHW model for non-random attrition makes things worse. The estimated negative trend is even more pronounced: A consequence of the estimated positive effect of the total number of trips on the probability of participation in another wave of the panel. Hence the model predicts that households that make few trips are likely to leave the panel.

- Ridder (1992) found that this was related to two opposite effects:
  - mobile families are less likely to leave the panel,
  - families who experience a transitory positive shock in their mobility are more likely to leave the panel.

Hence, we should distinguish between heterogeity and random shocks in controlling for the effects. This does improve outcomes, but not completely

Effects of attrition(2):

- The approach by Hirano et al shows that:
  - Taking refreshment into account does improve the models
  - Considerable different elasticities of trip making wrt income
  - trends better estimated
  - Not only attrition: both measurement errors and attrition important

Conclusions

- Panel data have a lot of advantages
- Also problems that may cause:
  - Problems with estimating parameters
  - Estimating trends in data
- Econometrics provides:
  - Methods to test for effects: not always problematic
  - Methods to deal with attrition→ refreshment may be advisable
- Conventional advices wrt improving survey methods may or may not work

Efforts by analysts to deal with attrition crucial