

Institute for
Fiscal Studies

Using scanner technology to collect expenditure data

Andrew Leicester (IFS, UCL)

Zoë Oldfield (IFS)



European Survey Research Association: Warsaw, 1st July 2009

Outline

- Consumer panel expenditure data
 - What is it? How is it collected?
- Key objectives of our research
- Findings in two main areas
 - Survey fatigue
 - Attrition
- Some thoughts on this mode of data collection

Consumer panel expenditure data: TNS

- Data from market research company Taylor Nelson Sofres (TNS)
 - Fast-Moving Consumer Goods (groceries) sector, Great Britain
 - c. 15,000–25,000 participating households, 2001–2007
 - Ongoing recruitment, continuous participation from panellists
 - Quota sampling from a number of address sources in GB
 - Record purchases using in-home barcode scanner
 - Product characteristics
 - Store details
 - Prices paid matched in from till receipts
 - Panel data: households participate for as long as they wish to do so
 - Demographics at sign up and 're-assessed' every 9 months or so
- Enormous scope of data – w/c 4th June 2007:
 - 670,000+ purchases, 65,000+ products, 18,000+ households
 - 3,000+ stores
 - £1m+ of expenditure

Key objectives

- Assess the strengths and weaknesses of scanner technology
 - Detailed comparison of TNS to existing data sources
 - Expenditure and Food Survey, British Household Panel Study
 - Suggest how survey mode affects recorded data
 - Representativeness: recruitment and retention (attrition)
 - Expenditures: accuracy of records, changes over time (fatigue)
- Feed into future research using the data
 - Huge potential, so far little used for social science research
 - Understand potential biases
 - Existing surveys interested in using similar methods
- Raise awareness of data and provide guidelines for researchers

Survey fatigue

- Ongoing demands of continuous participation could change recorded spending over time
 - Problem potentially worse for some goods, trips, households
- Panellists supposed to scan *all* barcoded items brought home
- Evidence of strong decline in recorded spending even in two week, one-off survey
 - Ahmed et al, 2006: Canadian Food Expenditure diary (FoodEx)
 - Spending 9% lower in week 2 than week 1
- Better or worse in consumer scanner data?
 - Participation potentially indefinite
 - Easier to scan barcodes than to keep a written diary

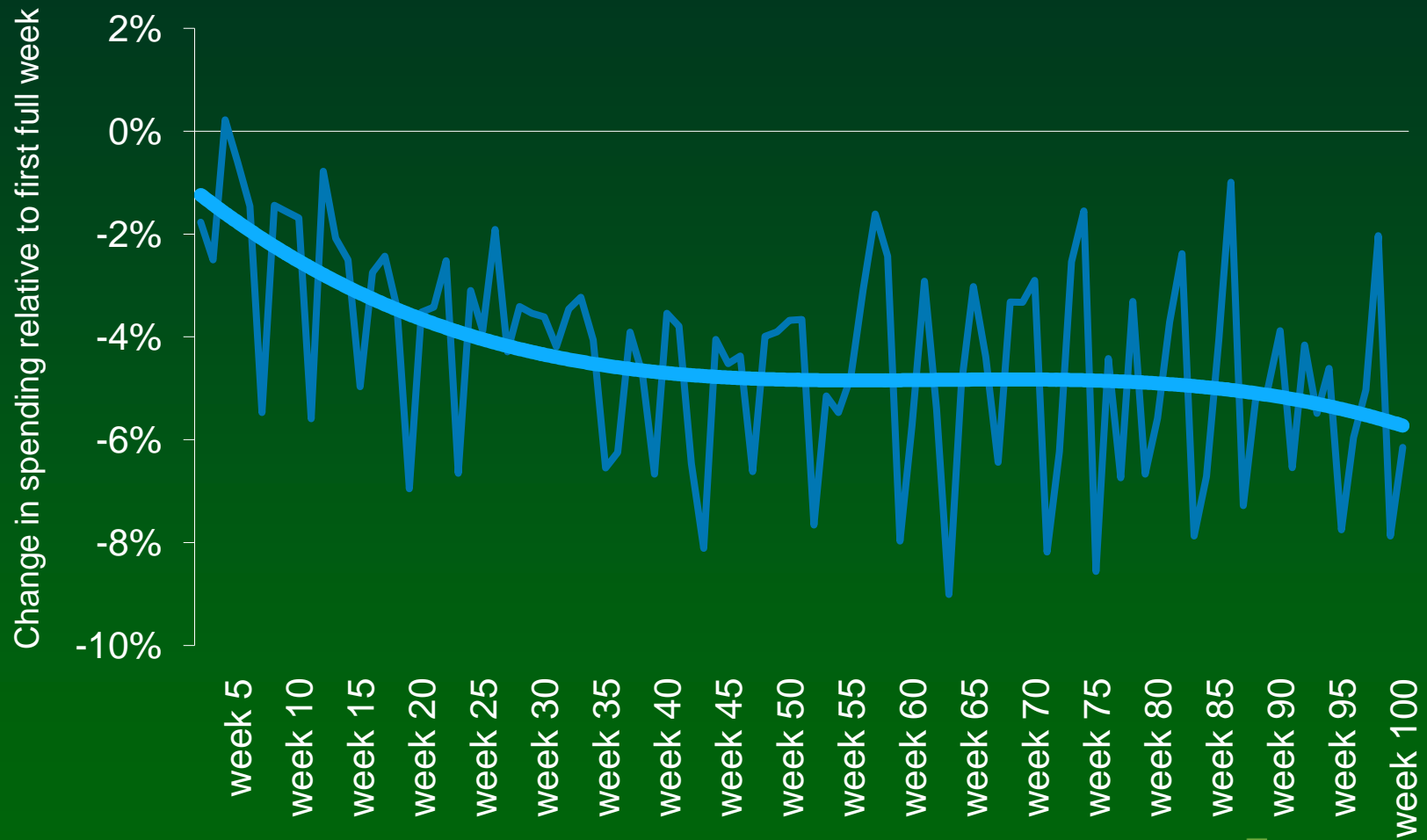
Survey fatigue: modelling approach

- Sample of households that *begin* participating in our data period
- Record (log) total spending per week
- Regress spending on dummy variables for the number of weeks since beginning to participate
 - First full week is excluded period, limit to 100 weeks
 - Coefficient on week dummy then tells us approx. change in spending in each week relative to week one
- Households who stay in a long time experience less fatigue?
 - Upward bias
 - Use household fixed effects: removed unobserved household characteristic correlated with long participation and low fatigue

$$\ln x_t^h = \alpha^h + \sum_{n=2}^{100} \beta_n w_n + \gamma Z_t^h + \varepsilon_t^h$$



Fatigue results



Fatigue results

- Spending around 5% lower on average after 6 months
- Less than found in FoodEx data after one week
 - Though hard to make direct comparisons given nature of surveys
- Variation across goods and households
 - Households with children: higher early fatigue
 - Childless households: no early fatigue, then more sustained decline
 - Pensioner households: no evidence of fatigue
 - Fatigue greater for alcohol, sweets & chocolates
 - Smaller for fish, fruit
- Patterns consistent with Canadian diary evidence

Attrition

- Sample of households that we observe *begin* participating
- Estimate non-parametric survival function:



- 7% drop out within 4 weeks
- 39% drop out within 1 year
- 54% drop out within 2 years
- 18% survive for 5 years or more
- Average duration is 48 weeks where we observe both start and end

Attrition

- TNS: probability of new household being observed 1 year later 63%
- BHPS: 86% of wave 1 sample gave full interview in wave 2
- Hard to make direct comparison but TNS attrition rate not bad ...
- TNS attrition varies with observable household characteristics
- Results of semiparametric duration model show:

Significantly higher risk of attrition

Households aged under 30
Households with any children
Missing employment status (compliance?)
Lone parents
Household without a car

Significantly lower risk of attrition

Households aged over 30
Single adult households
Childless households
Having new scanner technology

Conclusions and thoughts on scanner technology

- Data offers considerable advantages for research
 - Need to be aware of the potential biases and problems that arise
- We may expect issues like fatigue and attrition to be severe ...
 - But our findings are actually quite reassuring on these fronts
 - Problems may be particularly acute for some goods and households
- Suggestive evidence of modal effects
 - Recorded expenditure patterns and sample composition
 - Demographic differences do not explain spending differences
- Data collected for market research, not social science research
 - Transitions poorly recorded, limits value of panel aspect
 - But also some advantages; non-traditional data
 - Useful to try to integrate the two