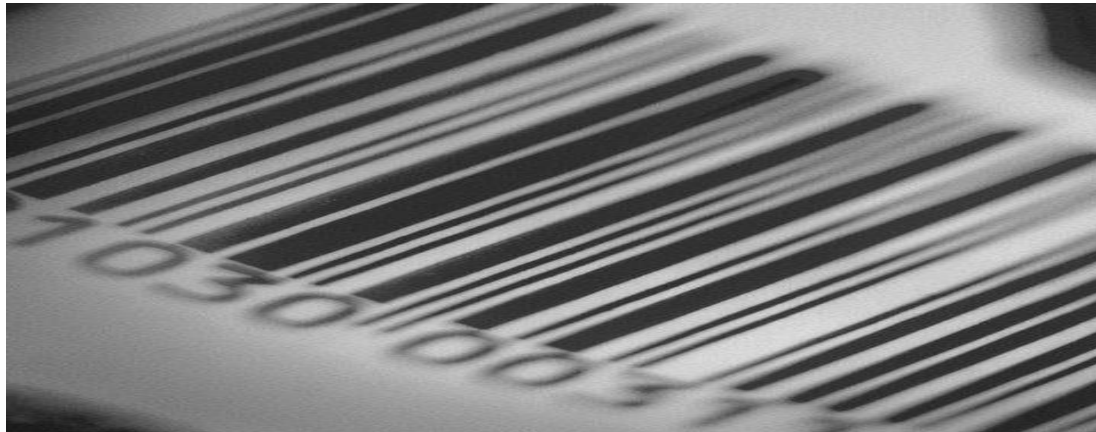




Institute for
Fiscal Studies



Using scanner technology to collect expenditure data

Andrew Leicester & Zoë Oldfield

Background and objectives

- Declining quality of traditional sources of spending data?
 - Using barcode scanners an alternative way to collect data
 - Questions over data quality, effect of collection method
- Key aims:
 - Assess the strengths and weaknesses of scanner data
 - Comparison to existing, well-understood data sources
 - How far are differences driven by scanner technology?
 - Sample composition, recruitment and retention
 - Impact on expenditure records
 - Inform researchers and raise awareness of data
- Outputs:
 - Leicester and Oldfield (2009), *Fiscal Studies*
 - IFS Working Paper 09/09 (www.ifs.org.uk)

Outline

- Description of scanner data
 - What is it?
 - How is it collected?
 - Why is it useful for research?
- Main findings
 - Demographic comparisons: cross-sectional and panel
 - Expenditure comparisons: scanner and diary data
 - Changes in spending over time: survey fatigue
 - Sample attrition
- Summary and conclusions

Consumer scanner data

- Market research organisation Kantar, **Worldpanel** data
 - Representative GB panel of 15,000 – 25,000 active households
 - Ongoing recruitment sampling approach
- Food & grocery purchases (FMCG), Nov 2001–Nov 2007
 - Recorded using in-home barcode scanner
 - Includes: off-sales alcohol, some non-food
 - Excludes: tobacco, baby food, food out of the home
 - In principle a complete record of household purchases
 - Non-barcoded items recorded via booklet of generic codes
 - Purchases from all stores
 - Prices collected via till receipts
 - Demographic data collected by telephone interview
 - Households participate for as long as they wish
 - Receive reward points for ongoing participation

What's in the data?

Media data

TV
Radio
Print media
Advertising
Attitudes

Product data

Size
Flavour
Packaging
Alcohol content
Low fat
Country of origin
Organic
Fair trade ...

External local area data

Deprivation
Area characteristics

Purchase data

Household id
Product id
Store id
Date
Expenditure
Volume
Number of units
Promotion
Nutrition

External store data

Store characteristics

Household data

Composition
Age, sex
Employment
Education
Postcode sector
Durables
BMI data
Income
Pets ...

Store data

Chain / fascia
Postcode

Why are the data useful?

- Advantages over traditional diary-based spending surveys:
 - Observe disaggregate prices and quantities
 - Observe household and store information
 - Panel data
 - Shorter lag in obtaining data
- But some disadvantages:
 - Uncertain data quality
 - Non-probability sample
 - Costly to obtain and unwieldy to analyse
 - Only a subset of total spending
- June 2006:
 - 2.32m purchase records, £3.39m expenditure
 - 18,835 households; 3,485 stores; 84,481 products

A word on sampling ...

- Worldpanel is a non-probability sample
 - Inference techniques are invalid
- Should we be using this data at all?
 - Rich, detailed data already available for researchers
 - Number of key research / policy questions need this kind of data
 - Costly to start from scratch using better sampling method
 - Potentially incorporate into existing probability surveys?

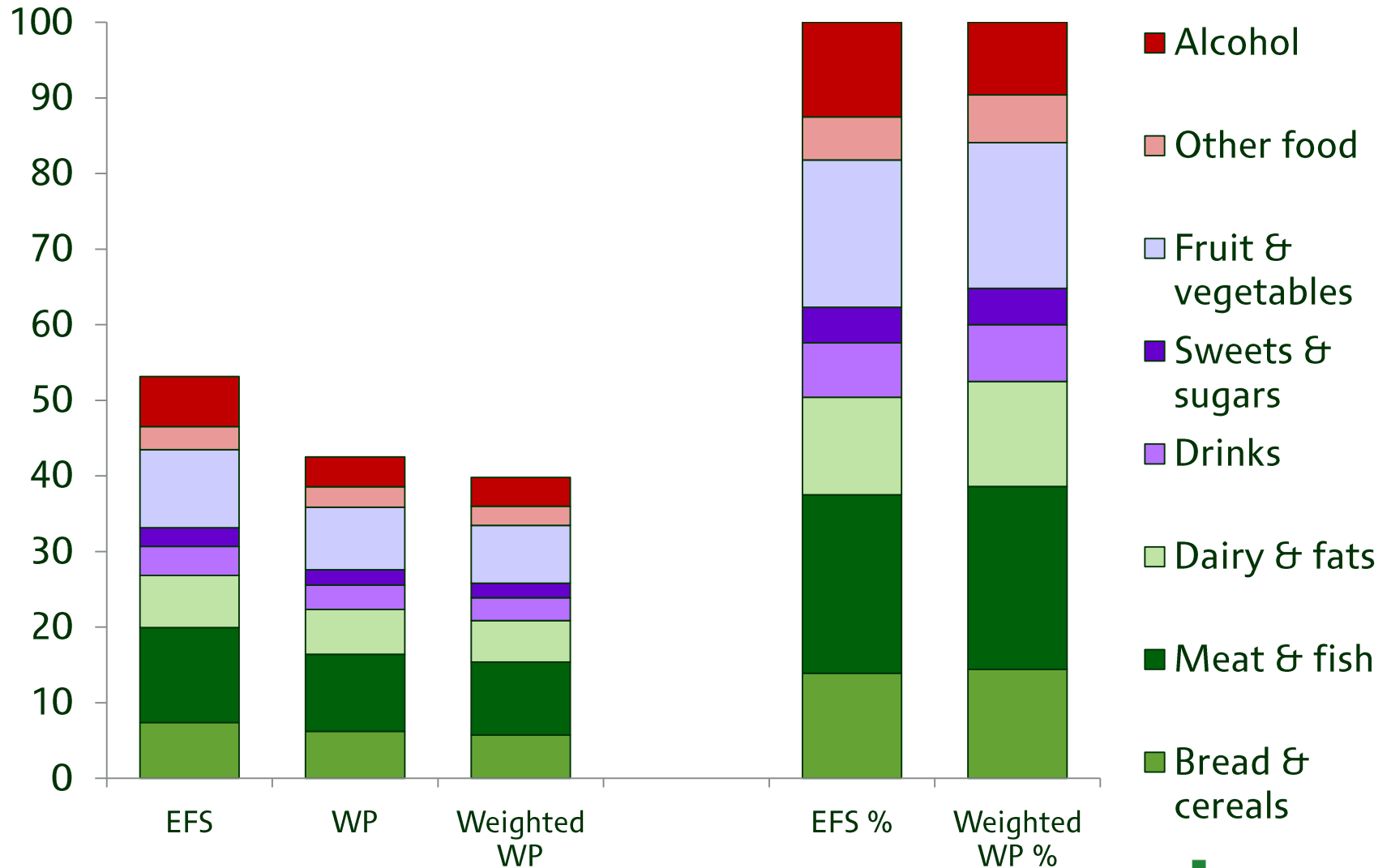
Cross-sectional demographic comparisons (2006)

- Compare sample in Worldpanel data to Expenditure & Food Survey
 - Differences may or may not be attributable to survey method
- Key comparative findings:
 - Scanner data oversamples multi-person households
 - Single adult households: 33% EFS, 23% Worldpanel
 - Deliberate – difficulties in recording ‘non-main shopper’ spending
 - Fewer very elderly households in scanner data
 - Households containing someone 80+: 8% EFS, 4% Worldpanel
 - Incomes substantially lower in scanner data
 - Household gross annual income > £60,000: 13% EFS, 5% Worldpanel
- Differences in sample could affect expenditure comparisons
- Calculate ‘propensity weight’ for scanner data to correct for this

Demographic transitions

- Demographics supposedly updated every 9 months or so
 - Useful to analyse expenditure response to demographic shocks
- Compare Worldpanel transitions to British Household Panel Study
- Evidence that Worldpanel records transitions poorly
 - Limits usefulness of panel aspect of data
- Age 50+ employed; probability of not working on year on:
 - BHPS 11%, Worldpanel 3%
- Does transition cause attrition?
 - Kantar admit problems with maintaining demographic data

Worldpanel and EFS expenditures (2005)



Expenditure comparisons (2005)

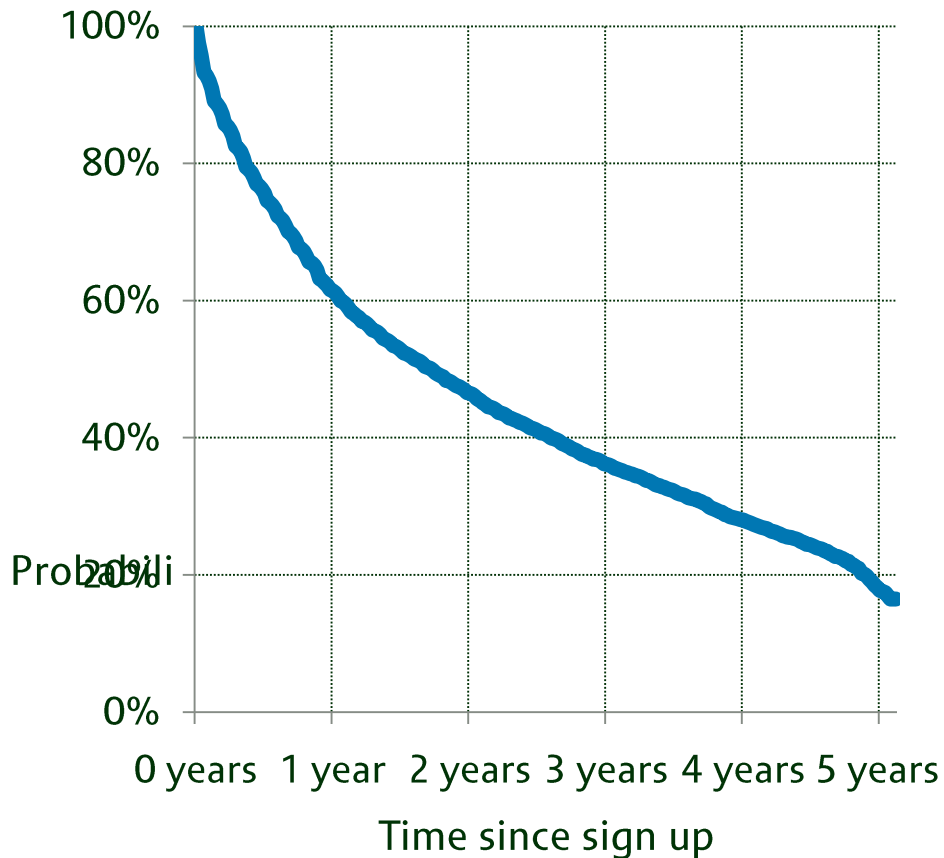
- Average weekly spending about 20% lower in scanner data
- Not accounted for by demographic differences
- Overall expenditure *patterns* match well
 - Extent of ‘under-record’ similar across broad spending groups
 - Alcohol an important exception
 - Embarrassment? (why worse in scanner data?)
 - Purchases by ‘non-main shoppers’?
 - Disaggregate comparisons suggest effects of scanner technology:
 - Gap bigger for categories where non-barcoded items important (meat, fruit & veg)
- Variation across household groups
 - Relatively higher spending for younger, single, childless households
 - Also for poorer, inactive/unemployed
 - Effects of time on ability to record?

Fatigue: changing spending within household

- 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
- Main results:
 - Spending 5% lower on average after 6 months
 - No further evidence of fatigue
 - Variation across goods and households
 - Does not fully explain spending gap with EFS

Attrition

- 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 48 weeks where complete spell observed

Attrition

- Worldpanel: 63% of new households observed 1 year later
- BHPS: 86% of wave 1 sample gave full interview in wave 2
 - Hard to compare directly ...
- Attrition risk varies with observable household characteristics:

Significantly higher risk of attrition

Households aged under 30

Households with any children

Lone parents

Significantly lower risk of attrition

Households aged over 30

Single adult households

Childless households

Conclusions

- Scanner data offers considerable advantages for research
 - Need to be aware of the potential biases and problems that arise
- Understanding the implications of data collection method vital
 - Sample composition differences at least partly driven by known reporting issues (e.g. multiple adult households)
 - Demographics and fatigue do not explain expenditure differences
 - Effects of attrition and fatigue vary across households and goods
 - Top-up shopping, time constraints have effects on spending
- Data collected for market research, not social science research
 - Non-probability sample
 - ‘Standard’ demographics not always available or maintained
- We offer a first comparative look at the data
 - Experimental methods may be needed for more definitive evaluation
 - Potential to be included in existing surveys?