



Understanding Society

Working Paper Series

No. 2019 – 02

May 2019

**A review of new technologies and data sources for
measuring household finances: Implications for
total survey error**

Annette Jäckle (University of Essex)

Alessandra Gaia (University of Milano-Bicocca)

Carli Lessof (University of Southampton)

Mick P. Couper (University of Michigan)

Non-technical summary

There is much hype about the potential of process-generated data and new technologies to collect data for research. These include data generated by social media (e.g. Facebook or Twitter) or new technologies (e.g. smartphone apps or sensors) as well as administrative processes of private companies (e.g. credit rating data) or local and national government (e.g. health, education or benefit records). These new data sources are typically considered cheap to collect, or already exist, often include large volumes of data, may provide good quality objective data, may be measured passively, may measure concepts that cannot be measured with survey questions, or measure concepts in greater detail. In reality, access to process generated data is often difficult to obtain. Such data also have several limitations that can affect their suitability for research, most notably coverage of the population of interest, limited covariates and data that are often designed for a different purpose than is needed for research.

In this chapter we review different new technologies and process-generated data that could be used to enhance the measurement of household finances in surveys: data collected with barcode and till receipt scanning, or from financial aggregator websites, supermarket loyalty cards, credit cards and credit rating agencies.

The aim of this review is to contribute to a greater understanding of errors that may arise at different stages of data collection (with new technologies), or of the data generating mechanism (with process-generated data) and how resulting errors affect data quality. This will inform research and development into methods to reduce the likelihood and impact of errors.

For each of the data sources and technologies, we review existing published and grey literature focusing on what, if anything, is known about: (i) the content of what can be measured, (ii) which research questions have been addressed using these data, (iii) whether the data have been used as free-standing data sources or linked to probability sample surveys, and (iv) the quality of the data regarding representativeness and measurement quality. The review is structured around an adapted version of the Total Survey Error framework we have developed for evaluating these new data sources, and concludes with a discussion of implications for survey practice and research needs.

A review of new technologies and data sources for measuring household finances: Implications for total survey error

Annette Jäckle (University of Essex)

Alessandra Gaia (University of Milano-Bicocca)

Carli Lessof (University of Southampton)

Mick P. Couper (University of Michigan)

Abstract:

We review process generated data sources and new technologies that could be used to improve the measurement of household finances. For each of these we review what is known about (i) the content of what can be measured, (ii) examples of research for which these data have been used, (iii) whether the data have been used as free-standing data sources or linked to probability sample surveys, and (iv) the quality of the data regarding representativeness and measurement quality. The review is structured around an adapted version of the Total Survey Error framework and concludes with a discussion of implications for survey practice and research needs.

Keywords: financial aggregator, account aggregator, loyalty cards, barcode scanning, till receipt scanning, gamification

JEL classification: D14, C83

Acknowledgements: This work was funded by the UK Economic and Social Research Council Transformative Research scheme and the National Centre for Research Methods (grant number ES/N006534/1), as well as by an NCRM International Visitor Exchange Scheme grant funding research visits by Mick Couper to the University of Essex.

Corresponding author: Annette Jäckle, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, CO4 3SQ, UK, aejack@essex.ac.uk.

1. Introduction

Survey data about household finances are key to many policy areas. For example, expenditure data are used to calculate consumer price indices. Wealth data are used to estimate pension entitlements and to understand intergenerational transfers of wealth. Both expenditure data and income data are used for poverty analyses. Data on household finances are also used to assess how households respond to wealth shocks, income shocks, fiscal policy (tax cuts and rebates), monetary policy (interest rates), and other macroeconomic policy research questions, such as, whether households are credit constrained or whether they have time consistency problems.

There are two main limitations to the way data about household finances are currently collected in sample surveys: the perceived infeasibility of collecting data about the entire household budget within one survey, and measurement error in the aspects that are collected.

In this paper we review different data sources, technologies, and methods that have been used or could be used to improve the measurement of household finances. The aim is to identify ways of improving both the scope and the quality of survey data on household finances, by developing and possibly combining new data collection methods.

MEASURING THE HOUSEHOLD BUDGET

To date there are no data sources in the UK or other developed countries (except for the Canadian Survey of Household Spending, SHS) that measure the entire balance sheet of households and that would allow fundamental insights into financial behaviours. In the UK, the Family Resources Survey focuses on income and assets, the Living Costs and Food Survey focuses on expenditure and income, the Wealth and Assets Survey focuses on wealth, and general purpose surveys such as the UK Household Longitudinal Study focus on household income, with incomplete data on expenditure and infrequent collection of wealth data. Collecting data about the entire household budget would require a very long and overly burdensome questionnaire. The restrictions on the content severely limit the usefulness of survey data on household finances and as a result many important policy questions cannot be resolved.

For example, there is considerable uncertainty over identifying which types of people are poor. Currently, monitoring and policy interventions are focused on income-based measures of poverty. However survey data from the UK, the U.S., and Canada suggest that households with very low income spend more than households with moderately low income (Brewer, Etheridge and O'Dea 2017; Meyer and Sullivan 2003). We do not know whether this is because of (i) measurement error in surveys, with low-income households under-reporting income and/or over-reporting spending, (ii) low income households smoothing their consumption in periods of temporary low income by using savings, or borrowing against future incomes in a rational and sustainable way, or (iii) very low income households engaging in unsustainable borrowing. Which of these factors is at work has obvious implications both for the welfare of individual households and the sustainability of the current economic recovery, as well as for best practice in identifying who is poor for policy purposes.

A second unresolved question is whether the rich really save more. In cross-sectional survey data, high income households appear to save a larger share of their income than lower income households, yet there has been no increase in aggregate savings rates over time even though real per capita income has risen. It is unknown whether this inconsistency is caused by measurement error in

income or whether it reflects true behaviours. Friedman famously proposed that the resolution is that consumption depends on lifetime income rather than current income, so that the correlation between savings rates and current income reflects transitory income fluctuations. However, econometric attempts to confirm this have not been successful (Dyner, Skinner and Zeldes 2004). How saving varies with affluence is important for the transmission of inequality across generations, for assessing the distributional impact of switching from income tax towards expenditure based taxes (like VAT), and many other key social and economic issues.

MEASUREMENT ERROR IN IN SURVEY REPORTING OF FINANCES

Various studies have documented measurement error in the reporting of finances in surveys. For example, income tends to be under-reported, in particular income from self-employment (Hurst, Li and Pugsley 2014) and from State benefits (Lynn et al. 2012). Reports of income and spending tend to be inconsistent at the very bottom of the income distribution, where households tend to report higher spending than income, which appears mainly to be driven by under-reporting of income (Brewer, Etheridge and O'Dea 2017; Meyer and Sullivan 2003).

Reporting on household finances is a burdensome, tedious, and error prone task for respondents. Information on household finances may be fragmented, for example, the respondent might have multiple bank accounts, making it difficult to recall and report. Survey questions on household finances are sensitive and some respondents might not be willing to reveal information or might misreport their answers. Information on household finances might not be encoded in memory, at all or in detail: respondents might find it hard to recall aspects of their finances (e.g. their mortgage balance), might forget small and frequent purchases (e.g. supermarket spending), and might find it difficult to locate infrequent or irregular purchases (e.g. travel tickets) correctly in time.

NEW WAYS OF MEASURING HOUSEHOLD FINANCES

Technological changes mean there are increasingly alternative options for collecting data about household finances, other than asking questions in a sample survey. This includes both new technologies for collecting bespoke data about sample members, and harvesting existing data generated by government or other organisational or commercial processes.

In this paper we review different data sources, technologies, and methods that are being used or could be used to obtain (i) more detailed information that cannot be collected in a survey questionnaire (such as detailed transactional data), or (ii) better quality data (such as objective records that do not rely on respondent recall).

We review both 'organic' and 'designed' data (Groves 2011). 'Organic' data are generated as a by-product of an administrative or other process. These data can be used as standalone data sources or can be linked to individual level survey data. The organic data we review are data collected by financial aggregator services (Section 2), customer loyalty programmes (Section 3), credit and debit card transactions (Section 4), and credit rating agencies (Section 5).

'Designed' data are bespoke data collected as part of a survey, whereby sample members are equipped with a tool and asked to collect data about their finances using that. The new technologies for designed data collection that we review are scanning of shopping receipts (Section 6), scanning of barcodes (Section 7), and smartphone applications (Section 8).

We review journal articles and grey literature in the form of working papers and conference proceedings, to identify new methods being used in industry and research practice. The new possibilities for measuring household finances are evolving rapidly, as both the content and quality of data that new technologies can capture, and the penetration of new technologies in society continue to change.

For each data source and method we review what, if anything, is known about: (i) the content of the data, (ii) how the data or method has been used for research purposes, (iii) the quality of data obtained with this method, and (iv) the feasibility and cost of using these data or method. We base our review of data quality on the Total Survey Error framework.

TOTAL SURVEY ERROR

The Total Survey Error framework is a conceptual framework used to categories different types of statistical error of sample survey statistics. It has been adapted over time to include new methods of data collection (Groves and Lyberg 2010).

Our review of what is known, or not known, about data quality for each data source and method is organised by error source. There are some instances where a particular type of error could conceivably be classified in different ways. For example, failure to scan a particular shopping receipt could be classified as a form of non-response or a source of measurement error due to missing items. The terminology used in the survey methods literature to describe different types of errors also differs from terminology used in some other disciplines (Groves 1989). For example errors of coverage and non-response tend to be labelled as ‘selection errors’ by economists. For the purposes of this paper we have grouped the various sources of error in two broad categories: errors of representation and errors of measurement (see Table 1).

Errors of representation include:

- **Coverage error** is a source of survey error due to systematic exclusion of sections of the population of inference. For example credit card data exclude anyone who does not own a credit card.
- **Non-response or participation error** is a source of survey error arising when systematic sections of the population do not participate in a survey. For example, if sample members are invited to set up and use a smartphone app to scan their shopping receipts, participation error arises when sample members do not participate in this task. We define errors arising when sample members participate in the study, but fail to scan some of their purchases, as “measurement error due to missing items or episodes”.

Errors of measurement include:

- **Specification error** occurs if the concepts measured by the data or method do not coincide with the concepts of interest for research purposes. Specification error can for example be due to categories of items that are by the nature of the data or method not captured, or due to errors in the unit about which data are collected (e.g. individuals or households).
- **Missing or duplicate items or episodes** arise when some items of spending are ‘missed’ or ‘duplicated’. For example, missing data arises when respondents forget to scan an item they have bought, or forget to use their loyalty card when shopping, or if a scanned receipt is not codeable due to poor image quality.

- ***Errors in reports or data capture*** arise when information is incorrect or miscategorised due to human or technological errors.
- ***Coding and processing error*** arises in the process necessary to make the data useable for research purposes, for example by deriving indicators from unstructured large data sets, coding till receipts, or classifying item descriptions into spending categories.
- ***Error due to panel conditioning*** can arise when the use of the technology influences the sample member's behaviour. For example, financial aggregators are designed to help people manage their finances, and, thus, they may change the users' financial behaviour.

Each of the following sections is organised along these types of errors. Table 1 provides a high-level comparative overview.

Table 1: Comparison of Technologies and Methods – Sources of Total Survey Error

		Financial aggregators	Loyalty cards	Credit/debit cards	Credit rating	Barcode & price data	Receipt scanning
REPRESENTATION/SELECTIVITY							
Coverage	designed (linked to survey)	has bank account; has internet access	has loyalty card or is willing to join a loyalty scheme	has credit card (applied and not been refused); different card issuers attract different types of clients	has applied for credit with an organisation that reports to credit rating agencies	has own device; or device loaned	has own device; or device loaned
	organic	has financial aggregator account	has loyalty card; different shops attract different types of consumers			volunteers; often internet users only	volunteers
Participation	designed (linked to survey)	consent and set up account	Consent	consent	consent	set up and use for period of time	set up and use for period of time
	organic	n/a					
MEASUREMENT							
Specification	concepts not covered	payments deducted from gross income; cash income not deposited; categories of cash spending	spending with other shops or service providers	non-credit card spending; spending with other credit card	Income; cash not deposited; spending	non barcode or non-standard barcode items; price information separately	Non-receipt purchases

	unit of analysis	unclear if individual or joint accounts	unclear if individual or joint	credit card account	usually individual	unclear if individual or joint	unclear if individual or joint
Missing or duplicate items/episodes (conditional on specification)		transactions in accounts that are not linked	forgot to use the card	none	credit not reported to agencies; records that are unscorable	items or entire shops not scanned; duplicates	no receipt; forgot to scan; scan not codeable; duplicate receipt scans
Errors in data captured		misclassification of categories of income and spending	barcode error; point of sale (scanning) error	missing or misclassified categories of spending	linkage errors processing errors	barcode error; error in reported price; error in matched price if price not reported; misclassification of items	scanning errors
Coding and processing errors		coding of income and spending categories by account holder or aggregator service	none	coding error in assigned categories	coding error in assigned categories	non-unique barcodes	data entry errors; coding errors: misclassification of spending categories
Conditioning		changes in spending behaviour	vouchers may influence the purchasing behaviour	no		changes in purchase behaviour	changes in purchase behaviour

Note: ^a the decision to join a loyalty card scheme may depend on the shop characteristics (e.g. customers may prefer to join only large chains, or chains with branch(es) in their local area; ^bin answering some economic research questions (e.g. individuals' sensitivity to price rises) the researcher may be interested in the price advertised rather than the price effectively paid.

2. Financial aggregators / account aggregation services

Data from financial aggregators provide detailed measures of income and spending, based on the financial transactions recorded in bank statements. Financial aggregators, or account aggregators, are online platforms designed to help consumers manage their finances. Registered users upload the login details for their bank accounts (including current accounts, credit cards, savings, mortgages, and investments), permitting the platforms to automatically retrieve data on transactions. The data scraped from bank statements include the date, amount and reference for each individual transaction, and sometimes also the balance on the account. The platforms classify individual transactions in broad categories (e.g. groceries or electricity), provide users with summary information on their income, debts, and investments, and offer tools for budgeting and financial goal setting.

There are various providers of financial aggregator services: for example, Money Dashboard, Mint, OnTrees, Moneydance, Bankin', Linxo, PocketSmith, AceMoney. Some providers are active in different countries (e.g. Mint in the U.S. and Canada, Bankin' in France, UK, Germany, and Spain), while others are country specific (e.g. Money Dashboard in the UK). Also, some financial aggregators address specific financial needs. For example, ReadyForZero focuses on debt repayment, Quicken helps users produce tax reports. While most financial aggregators provide services directly to consumers, some (e.g. Yodlee in the U.S.) provide services to financial institutions, which then in turn offer these services to their customers.

There are other websites which access data from bank accounts for other purposes. For example, *people.io* recruits users to upload their banking and other online login data in exchange for "points" which can be used to purchase products and services online. The company sells the anonymized data scraped from their users' accounts to third parties. Another example is *miiCard*, which is an online identity verification service that checks whether online bank login details are correct and releases a virtual identity card that customers can use for online transactions. We are not aware of any uses of data generated by *people.io* or *miiCard* in research.

RESEARCH APPLICATIONS

The transaction data from financial aggregators can be used to derive a variety of aggregates or categories of income and spending, to suit specific research needs. Examples in the literature include measures of total income, total spending and subcategories of spending, recurring and non-recurring income and spending (Gelman et al. 2014), dates of tax refund payments, measures of financial constraints (Baugh, Ben-David and Park 2014), and information on planned repayment of debts and loans (Kuchler 2015).

Measures of income and spending derived from financial aggregators have so far mainly been used by economists to study the sensitivity of spending to changes in income. For example, using data from the U.S. financial aggregator *Check*, Gelman et al. (2014) test an economic theory predicting that individuals should manage their cash, savings and borrowing such that their spending is independent of the timing of income. The results however show that the timing of spending does depend on the timing of anticipated income. Similarly, using data on Icelandic households from the financial aggregator *Meniga*, Olafsson and Pagel (2018) analyse spending responses to the arrival of both regular and irregular income, finding that for most of the population consumption is sensitive to

the timing of income payments, few individuals hold little/no liquidity, and liquidity holdings are larger than predicted by economic models. Baugh et al. (2014) use data from an undisclosed financial aggregator service in the U.S. to examine how cash flows influence household consumption, examining three contrasting explanations: financial constraints, precautionary savings, and myopia. They find that on average households are financially constrained, show myopic behaviour, and do not act consistently with precautionary savings motives. Kuchler (2015) uses data from *ReadyForZero*, a financial aggregator specifically focused on debt repayment. The author shows that, due to preferences for present over future spending (so called “present bias”), many consumers fail to comply with the repayment plans they have set themselves.

Baker (2015) linked financial aggregator data to data on firms, using pay check descriptions from bank statements to identify where people worked. Three sources of firm-level data were used: Securities and Exchange Commission (SEC) data¹, quarterly profit reports from the Institutional Brokers’ Estimate System (IBES), and newspaper articles on layoffs, that is, temporary or permanent suspension of groups of employees for business reasons. The author studied whether the sensitivity of consumption to changes in income is higher among household with higher level of debt. Since households may adjust current spending by anticipating future changes in income, the author used data on shocks experienced by the firms in which household members are employed, to instrument for future changes in household income. The results indicated that households with high levels of debt are indeed more sensitive to income fluctuations and that these results are driven mainly by borrowing and liquidity constraints.

Most of the studies reviewed above make use of existing process-generated data from a single aggregator, without any additional information about account holders. There are however two independent research teams in the U.S. that are working on combining process generated financial data with designed survey data collection. A team of researchers at The Vanguard Group, the University of Michigan, and New York University are administering online questionnaires to samples of Vanguard account holders and combining the survey data with the process generated data the company already has (see http://ebp-projects.isr.umich.edu/VRI/survey_overview.html). Another research team at the University of Southern California is taking the opposite approach: they start with a probability-based online panel, the *Understanding America Study* (<https://uasdata.usc.edu/>), and invite panellists to create an account with a financial aggregator (Yodlee). The aggregator company then passes the financial data they collect about sample members on to UAS to be linked to their survey data.

DATA QUALITY: IMPLICATIONS FOR TOTAL ERROR

In terms of measuring financial transactions, aggregator data may be more comprehensive and detailed than survey data and in some respects also more accurate. However errors in estimates also depend on the extent to which account holders link all bank accounts to the aggregator service, the extent to which these accounts capture the full range of financial transactions, the quality of coding income and expenditure categories, and the extent to which the users of these services are representativeness of the population of interest. In the following we discuss and review evidence of different potential sources of error in data from financial aggregators.

¹ Firms are required to provide data to SEC at occasions such as firm or plant closures, or significant merger and acquisition activities, and to disclose information to stakeholders.

Errors of representation

A key concern is what proportion of the population use financial aggregators and how do those who do differ from those who do not. This is viewed as coverage error or selection error in the total survey error framework, and is discussed below. Another type of selection error is the coverage of financial accounts and transaction that are represented in these data. We discuss this source of error below under measurement error (missing data).

Existing users of financial aggregators are *self-selected*. The extent of inferential errors due to self-selection depends on the diffusion of aggregators among the studied population, and on the extent to which population members who sign up to financial aggregators are different, in characteristics relevant to the research purposes, from population members who do not sign up.

Financial aggregator data could alternatively be collected by asking members of a probability sample to register with an aggregator service, as done by the Understanding America Study. In this case, errors of representation could be due to a lack of internet access (*coverage error*) and/or *non-compliance* with the request. We discuss these different scenarios (self-selection, coverage, and non-compliance) and any empirical evidence in what follows.

Regarding *selection error*, Olafsson and Pagel (2018) reported that in 2014 one fourth of households in Iceland were using the app *Meniga*. The relatively high rate of use may be because the app was marketed through banks and the Icelandic population makes wide (nearly exclusive) use of electronic payments. In a survey in the Understanding America Study in the U.S. in late 2014, about 9% of panellists reported using an online personal financial management service. Recent data from wave 9 of the *Understanding Society* Innovation Panel suggest that less than 1% of respondents use any financial aggregator in the UK. More broadly, we expect the diffusion of financial aggregators to vary across countries and time, depending on several factors, including: regulatory climate, diffusion of technology, financial literacy, diffusion of electronic payments, as well as number of financial transactions and number of bank accounts per person (Olafsson and Pagel 2018). The benefits of using financial aggregator services are expected to increase with the number of financial transactions and the number of bank accounts an individual has. Therefore self-selected users are likely to over-represent those with high income, high wealth and high expenditure. In addition, individuals interested in subscribing to financial aggregators and willing to share their bank login credentials are likely to be a selected group of the population in terms of personal characteristics and/or “need for financial organization” (Gelman et al. 2014:213).

Several researchers using self-selected data have discussed differences between users of financial aggregators and the general population, to assess to what extent their results are generalizable. In terms of observed socio-demographic characteristics, two studies find that users of financial aggregators over-represent men and younger age groups, while education levels and geographic distribution of users are broadly similar to those of the U.S. population (Baker 2015; Gelman et al. 2014), as are marital status, household size, home ownership, profession and income level (Baker 2015). For the Icelandic population, the financial aggregator data are in line with national averages in terms of socio-demographic characteristics – age, gender, parenthood and employment status. Financial aggregator data match data from Statistic Iceland in terms of monthly total income, salary, overall spending and most categories of spending (i.e. groceries, alcohol, transportation, pharmacies, and clothing and accessories), while financial aggregator data overestimate spending on sports and activities, and underestimate spending on fuel, ready-made food, and home improvements.

The general consensus from U.S. studies is that unweighted aggregator data cannot be used to obtain unbiased estimates for the entire population. However, as the popularity of financial aggregators grows, the demographic characteristics of users may come closer to those of the population (Baker 2015). Given the number of different financial aggregators on the market, there may also be differences between users of different products. No study has compared data from different financial aggregators. Even though use of such products may increase, there will likely to continue to be selection biases in who chooses to use financial aggregators. In addition, aggregator data have the potential to capture spending from high-income earners, which are not adequately represented in other survey-based data, such as, for example, the Consumer Expenditure Survey (Baker 2015). However, if the objective is to link survey data and financial aggregator data for survey respondents, this group will still be missed. If members of a probability sample are invited to register with a financial aggregator, both coverage and non-compliance errors are potential issues.

Registering with a financial aggregator requires internet access. Sample members without internet access are therefore excluded, resulting in *coverage error* as these people are likely to have very different financial situations from people with internet access. Online probability panels that provide internet access to their panellists would not face this issue, conditional on participation in the panel.

Non-compliance (or non-response) is a further source of error. As discussed for self-selected samples above, it is likely that people with a more active interest in managing their finances are more willing to register with an aggregator service. There may also be motivational differences in compliance, with those who perceive some personal value to such a service being more likely to agree to use such a service and provide access to the data. For example they may be particularly organised and effective managers of their finances, using a service of this kind to keep on top of finances in place of or as a supplement to previously successful mechanisms, or they may be particularly disorganised and looking for new ways of addressing problematic behaviours such as overspending. These differences may have consequences, for example highly organised individuals may be less likely to show conditioning effects than those who have little awareness of their financial behaviours when they start to use a service of this kind. These issues have largely not yet been addressed in the literature.

In one exception, Angrisani, Kapteyn and Samek (2018) report on preliminary results from their study asking panellists to sign up for a financial aggregator. Of those invited, 65% consented to participate in the study. Of those who consented, 68% (or 32% of those invited) signed up for the financial aggregator service. Of those who signed up, 38% (or 12% of those invited) added one or more financial institutions to their account. Angrisani and colleagues found significant differences by age and education in each of these stages of participation, and differences in initial consent by income. In addition, those who already used the internet for banking were more likely to sign up. This study is likely to yield important insights on the participation and compliance errors associated with incorporating financial aggregator data into surveys.

Errors of measurement

Specification error may occur due to the definition of income and spending that is measured by financial aggregators: measured income is net take-home pay paid into a bank account, after taxes, social security contributions and any other payments deducted from gross income by the employer

such as healthcare or pension contributions (Baker 2015). Cash income that is not deposited in an account is not included. Correspondingly, spending only includes payments made from net income deposited in a bank account and cash spending cannot be classified. While deposits and withdrawals are observed in the bank statements, the income or expenditure categories are not recorded unless the user enters these manually. In other words, the existing data may not match what is needed for analysis. This may be different for different types or sources of income and for different types of expenditures; that is, financial aggregators are better for some measures than others.

Specification error may also occur because the unit of observation in financial aggregator data is not well defined: an individual person may create a financial aggregator account and link only their personal accounts, or link all (joint) accounts held by family members who make joint financial decisions. Baker (2015), for example, assumes that financial aggregator data represent household spending for married couples, but individual spending for unmarried roommates living in the same house. Identifying the correct unit of analysis is crucial both for valid inference to population characteristics and for comparisons with other data sources.

Reporting error for individual transactions in linked bank accounts is less likely than in survey data (Baker 2015). Measurement error in the aggregate value of (categories of) transactions may however result from bank accounts that are not linked, from missing coverage of cash and check transactions, and from coding and processing errors.

The type and number of linked accounts may lead to measurement error, if some accounts are not linked, or if only certain types of accounts are linked (Baker 2015). A survey of users of an undisclosed aggregator service in 2011 found that the vast majority of respondents (over 90%) reported linking all or almost all of their checking, savings and credit card accounts. In addition, 75% of respondents reported all or almost all of their equity accounts (Baker 2015). Among respondents who self-identified as homeowners, over 90% reported including information about their home value or mortgage. The coverage was less complete for unconventional asset accounts, such as retirement accounts: only 50% of respondents reported linking all or almost all of this type of account. This might result from the nature of the services offered by the financial aggregator, which is mainly used for tracking income and expenditures. Users are more likely to link accounts where transactions are more frequent: when asked about the reason for not linking certain accounts, respondents replied that there was little activity on these accounts, and they felt no need to track them. Unlinked accounts are therefore unlikely to account for an important share of transactions compared with linked accounts (Baker 2015), however unlinked accounts may account for an important share of stocks in savings or investments.

Unlinked accounts are a particular concern for new users, who may link their financial accounts gradually. If an account is added at a subsequent stage, increases in income and/or spending are likely to be observed, which do not reflect true changes. Such errors may be limited by cleaning and robustness check strategies, such as excluding data from a user's first months using the platform, excluding users with highly volatile numbers of accounts or few accounts (Baker 2015). Given that many financial aggregators are recent start-ups and their user bases are expanding rapidly, this may be more of a transitional concern.

Coding and processing errors are not discussed in the literature using data from financial aggregators. No information is provided about the reliability of the automated coding of income and expenditure classifications or on the format of the data or the resources necessary for cleaning and processing to

make them suitable for research purposes. Discussions with researchers using such services suggest that extensive processing is needed to make sense of these data. One aspect that is discussed, though, is the selection of cases included in analysis. Researchers using financial aggregator data tend to exclude data on users that are considered unreliable, such as incomplete profiles, inactive profiles, profiles with few accounts linked. For example, Gelman et al. (2014) selected users who had at least one bank or credit card account covering 300 consecutive days in a two year span. They do not report how many cases were excluded for not meeting these criteria. Similarly, Kuchler (2015) used only *ReadyforZero* users who had an account for at least 180 days, received regular biweekly pay checks, had linked their checking account and seemed to have linked all their active credit card accounts. Of the 3,653 users with linked accounts, 2,558 (or 70%) were observed for at least 180 days after sign-up; of these, 2,051 (or 80%) had all accounts linked at sign-up.

Another source of potential measurement error are *conditioning effects*. One of the stated benefits of using aggregator services is that they help individuals make better financial decisions. Therefore individuals using an aggregator service – whether of their own volition or in response to a survey request – may change their financial behaviours such that they are no longer representative of the general population. This may be a particular concern for members of ongoing panels encouraged to use such tools for the purpose of improved measurement. We are not aware of discussions of conditioning effects in the literature using financial aggregator data.

Total error in estimates

Some previous studies have compared estimates derived from self-selected aggregator data with external benchmarks, to identify the total error due to all errors of measurement and representation.

The quality of *expenditure data* from an undisclosed aggregator service is assessed by Baker (2015), by comparing categorized spending data with the Consumer Expenditure Survey and with the Census Retail Sales data (collected in a monthly survey of large and small retailers). The financial aggregator data for the period are highly correlated with annualized data from the Consumer Expenditure Survey for 2009-2012: the correlation exceeds 0.87 for all categories observed in both data sources. The aggregator data are also strongly correlated with the Census Retail data for expenditure categories other than motor vehicles, and after weighting the data to reflect the composition of the Current Population Survey (and hence the U.S. population). The divergence in expenditure for motor vehicles is explained by measurement error: many people purchase such goods with loans, which appear in user accounts on a monthly basis, while retailers record the sale as a one-off event (regardless of the modality of payment). Similar effects are observed, but less pronounced, for other types of durables which are often purchased with consumer credit or financing.

Housing wealth data from financial aggregators (weighted for observables) match national distributions by year and zip code (Baker 2015). *Liquid savings and credit* among households match estimates from the Survey of Consumer Finance closely, not only on weighted spending trends but also on observed asset and debt levels (Baker 2015).

All of these comparisons are at the aggregate level. Of most interest is the individual- or household-level data. Without matching survey data to data from financial aggregators, we do not know how large the differences are at the micro level. It is possible that there are few differences in average total error at the aggregate level, but large deviations at the individual level.

FEASIBILITY AND COSTS

Linking financial aggregator data to individual survey data requires collaboration with a financial aggregator service and inviting sample members to create a user account and link all their accounts to that service by providing their banking login details. While online platforms provide reassurances on the security of this procedure, some banking institutions discourage and/or prohibit the use of financial aggregators (see below). The ethics of asking respondents to participate in a procedure that could put their finances at risk therefore needs careful consideration.

The costs of accessing financial aggregator data are not discussed in the publications using such data. The costs of the services for individuals vary from free apps to a one-time app purchase fee, to monthly subscription services. If survey respondents were asked to sign up to these services for research purposes, the costs would be expected to be borne by the survey agency. For those researchers who have analysed data from financial aggregators, the data are usually provided free of charges, as the aggregators derive benefits from these analyses. As the demands for these data increase, this may well change.

REGULATORY ENVIRONMENT AND OUTLOOK

The literature using financial aggregator data so far stems from the U.S. We are not aware of any study using UK financial aggregator data, which might be due to the limited diffusion of such services in the UK. Currently UK banking institutions are discouraging the use of financial aggregators, officially on the grounds of security concerns and technical issues, but potentially also to limit competition (Sidel 2015). Financial institutions have argued that it is not possible to distinguish legitimate logins from third party applications and hacking attempts (Peters 2016), and that encouraging consumers to hand out banking login details to third parties expose them to phishing attacks (Crosman 2015). Financial institutions have also stressed that the activity of financial aggregators generates web traffic overload, which in some cases resulted in the banks temporarily limiting access (Crosman 2015). Some commentators (e.g. Sidel 2015), however argue that banking institutions cut off access to their data or provide their own services to limit competition.

The regulatory environment is now changing and as a result the diffusion of financial aggregator services in the UK may increase. In an attempt to increase competition in the UK banking sector, the UK Financial Conduct Authority is seeking to lower the barriers faced by financial aggregators, by examining the security of sharing personal information of bank customers and intervening to encourage wider use of aggregator services (Financial Conduct Authority 2015). This shift in regulation may lead to a wider diffusion of financial aggregator services, providing new opportunities for research use in the UK.

3. Loyalty card data

Data from loyalty cards provide detailed measures of spending at a defined set of outlets, based on purchases recorded by the company's loyalty scheme. Loyalty cards (also called rewards cards, points cards or club cards) are personalized cards which are usually issued by a retail chain or travel-related company (air, rail, hotels, etc.). Loyalty cards are used for research purposes by matching the card owner's demographic data with transactional data (point-of-sale scanner data, flight booking information, etc.), thereby creating consumer panel data on expenditures linked to the loyalty card

(see Tin, Mhurchu and Bullen 2007). Most major supermarket chains have loyalty card schemes, and their use by consumers is widespread.

Already in the late 1990s, 150 loyalty card schemes were implemented in the UK with over 40 million cards being issued (Byrom et al. 2001). One example of loyalty card scheme is Tesco Clubcard launched by the Supermarket chain Tesco in 1995.

Loyalty card data include the information provided by the customer when applying for the card, information about purchases made using the card at the point of sale, information about redemptions obtained using the rewards of the loyalty programme, and responses to surveys or other schemes conducted by the loyalty programme (Worthington and Fear 2009). For example these data include demographic information (e.g. age, sex), location of stores visited by the customer, products purchased (category, brand, price), frequency of purchase, and transaction value (average basket size).

RESEARCH APPLICATIONS

Loyalty card data have been used in research in the area of nutrition. For example, Andreyeva et al. (2012) used data from the loyalty card system of a large supermarket chain with stores in New England to study refreshment beverages purchases among participants in federal assistance programmes aimed at reducing food insecurity and addressing malnutrition in low-income individuals. The authors find sugar-sweetened refreshment beverages to contribute significantly to grocery expenditures in the studied population, and advise reconsidering the programme's practices to adjust public spending to public health goals. Hornibrook, May and Fearn (2015) analyse consumers' responses to carbon footprint labels, finding that carbon labels do not shift the demand to lower carbon products on the three product categories analysed (orange juice, potatoes and washing detergent).

A variety of other studies have used loyalty card data, for example to study the effect of product promotions on purchases (Felgate et al. 2012), purchase of organic food (Juhl, Fenger and Thogersen 2017), grocery store choice behaviours (Sturley, Newing and Heppenstall 2018) and grocery spending by tourists (Newing, Clarke and Clarke 2014). Many of these studies are based on limited data (a single store, a small geographical area or a limited set of products). The relative paucity of papers based on large-scale loyalty card data points to the potential difficulty of getting access to such data for research purposes.

Loyalty card data can potentially be used as sources of statistical data in two ways. Firstly, they can be used as a standalone data source of organically generated data. Secondly, they can be linked to survey data: survey sample members may be asked to join a loyalty card scheme and data from their purchases may be linked to their survey responses. While there are a few examples of the first use, we have not found any examples of the second. Further, we know of no studies that have explored the use of generic cards (e.g. loyalty cards that work across networks of stores or credit cards that award points for all types of purchases).

DATA QUALITY: IMPLICATIONS FOR TOTAL ERROR

Errors of representation

In terms of representation and selectivity, while organic data collection require participants to be members of a loyalty card scheme; designed data collection require sample members to either already be part of a loyalty card scheme or be willing to join one.

Andreyeva et al. (2012) estimated that 90% of transactions in a large supermarket chain in New England were made using a loyalty card. In the UK, the population of loyalty card users in the largest UK shopping retailer amounts to 14 million shoppers (Hornibrook, May and Fearne 2015); in the supermarket chain Tesco, 80% of transactions are covered by the Tesco Clubcard programme (Worthington and Fear 2009). In a survey conducted in Australia, Worthington and Fear (2009) found that 83% of adults reported having at least one loyalty card, with the most common card being held by 59% of respondents. Women held on average 2.02 loyalty cards, compared with 1.29 for men. Loyalty card possession also increased with age.

Even though the share of transactions covered by loyalty cards is very high, errors in representation may arise in the form of *coverage error* if loyalty scheme members are different or have different shopping behaviour from other customers. This type of error may be because loyalty card schemes may attract frequent shoppers and/or customers who are sensitive to prices and discounts (Cortiñas, Elorz and Múgica 2008).

Cortiñas et al. (2008) analyse customer responses to product characteristics such as discounts, brands, and sizes between loyalty scheme members and non-members. They find that card holders are less sensitive to usual prices, but more sensitive to promotions in some product categories. The authors argue that loyalty scheme members may behave differently from non-members as a result of marketing programmes associated with the loyalty card, such as special offers, discounts and gifts, or due to self-selection in which types of customers decide to join the scheme.

Participation (or non-response) error may arise in loyalty card data when customers are actively asked to provide some information but decide not to. Data on product purchases (category, brand, price), location of stores visited by the customer, and frequency of purchases can be gleaned passively from loyalty card data without any action required from the customer. However other data are collected actively by asking the customer enrolled in a loyalty scheme to provide additional information. This happens, for example, in the registration form when signing up for a loyalty card, or in any survey of loyalty card users that the company may run. Data collected actively are subject to non-response error, because the respondent may decide not to participate in the survey. In a survey starting with a sample of loyalty card holders, only 19.2% of those invited to the survey completed it (Panzone et al. 2016).

Errors of measurement

Specification error may arise in loyalty card data when the unit of analysis of interest for the researcher does not coincide with the loyalty card user. For example, researchers are usually interested in analysing either single individuals or groups of people that take financial decision jointly, typically, a household. However, loyalty cards do not necessarily identify an individual or a household; several household members might share the same card, or different household members

might use different loyalty cards. Linking cards within households may be difficult or impossible (Andreyeva et al. 2012).

Another form of specification error arises for purchases not covered by the loyalty card; loyalty cards are chain (or company) specific, and thus do not cover purchases in other stores outside the chain issuing the loyalty card. Similarly, large retailers are more likely to issue loyalty cards; purchases made in non-affiliated stores (small business owners) or in other settings (e.g. farmers' markets) are likely to be missed. Even if the purchases are made in the stores covered by the card, customers may forget to use the card, or choose not to use the card for certain transactions. We have found no literature that addresses these measurement issues.

Total error in estimates

The papers we reviewed focused on a narrow range of behaviour or products purchased, and were limited to a small geographical area. Given this, none of the papers compared aggregate estimates from loyalty card data to external data sources.

FEASIBILITY AND COSTS

The limited use of loyalty card data in the literature is an indication of potential difficulty acquiring such data for research purposes. The data are viewed as proprietary and used mainly for marketing purposes. Given the availability of alternative data sources (e.g. scanner data), and the limited scope of loyalty card data in terms of both persons and products covered, little effort seems to have been invested in getting access to loyalty card data for research purposes. The recent introduction of the General Data Protection Regulation (GDPR) in Europe may further limit research access to loyalty card data. Existing panel members or loyalty card holders would likely need to provide explicit consent for any use of such data. The GDPR rules apply similarly to other existing data sources.

4. Credit and debit card data

Data from credit and debit cards provide detailed measures of all spending, based on purchases made using card payments. The content of credit card data and the frequency with which the information is collected varies between credit card issuers. The information collected can be grouped into four main types: 1) data from the monthly billing statement for the account (i.e. balances, payments, credit limits, interest charges and interest rates); 2) data obtained by the issuers from credit bureaus (e.g. external credit scores, number of other credit cards, and combined balances on other credit cards); 3) administrative data related to each account (e.g. internal credit scores, changes in credit limits and interest rates); and 4) socio-demographic data from the credit application such as age, income, and marital status (Agarwal, Liu and Souleles 2007; Gross and Souleles 2002a).

Credit card data have the structure of panel data: individual accounts are typically tracked monthly allowing fine-grained analyses of dynamics (Agarwal, Liu and Souleles 2007). The panel data are unbalanced: accounts are tracked for a different number of years depending on the card issuer and how long the card holder uses the account. Gross and Souleles (2002a) note that on average, in their dataset, individual accounts are observed for 24 months.

A further advantage of credit card data is that they contain all information that the card issuer uses to evaluate accounts, which makes it possible to disentangle credit supply and credit demand (Gross

and Souleles 2002b). A disadvantage is however the lack of information on other components of household finances, such as assets or spending using payment methods other than credit cards (Agarwal, Liu and Souleles 2007).

RESEARCH APPLICATIONS

Credit card data have been used extensively for research purposes. Researchers typically use random samples of credit card data, either from a single card issuer (e.g. Agarwal, Liu and Souleles 2007), or combining data from more than one issuer (e.g. Gross and Souleles 2002a). Gross and Souleles (2002b) use credit card data to study credit card delinquency, personal bankruptcy, and the stability of credit risk models. They also use credit card data to study how people respond to changes in credit supply (Gross and Souleles 2002a). The authors find that increases in credit limits lead to rising debt. The proportion of additional disposable income which is spent on consumption (that is, the marginal propensity to consume) is larger for individuals near their credit limit, but also significant for people well below their limits. The authors also find strong effects on total borrowing of changes in account-specific interest rates. In another example, Agarwal, Liu and Souleles (2007) use credit card data that include the penultimate digit of the account holder's social security number to analyze consumers' response to the 2001 U.S. federal income tax rebates. They found that while consumers initially increased savings as a result of the rebate, they increased spending soon afterwards. This evidence is contrary to the prevailing predictions of the permanent income hypothesis (Friedman 1957), that consumption is insensitive to temporary changes in income that do not affect lifetime income.

We are not aware of research where the process generated data held by credit card companies have been combined with designed data collection.

DATA QUALITY: IMPLICATIONS FOR TOTAL ERROR

Errors of representation

Coverage error in credit card data arises since the data from card issuers may not represent the target population of interest for research purposes. Generalising to the population is problematic since the general population includes individuals who do not hold credit cards, either because they do not choose this form of credit or because they are refused credit. Even generalising to the population who hold credit cards can be problematic, as different issuers attract consumers with different earning profiles and spending behaviours. Therefore the data from one credit card issuer do not necessarily represent the variety of profiles of the credit active population. Some authors (e.g. Gross and Souleles 2002b) have attempted to decrease this source of error by pooling data from different issuers.

Non-response and non-consent error would be additional sources of error affecting representation, if attempts were made to link individual credit card data to survey data.

Errors of measurement

Specification error arises depending on the concepts of interest. Spending is measured only partially by credit card data, as it excludes information about spending with all other payment methods: cash, cheque, debit card, standing orders, direct debits, bank transfers, PayPay, mobile wallets, etc.

The unit of analysis can be a further source of specification error. The unit of observation in the credit card data is the credit card account, rather than the individual or household level (Agarwal, Liu and Souleles 2007), which are the units of analysis usually of interest for research purposes. This might be problematic because, within households, multiple people may share the same account (Gross and Souleles 2002a), multiple cards may be issued for the same account (Gross and Souleles 2002a), and individuals may have multiple accounts (Agarwal, Liu and Souleles 2007).

Reporting error in credit card data is less problematic than in survey data (Agarwal, Liu and Souleles 2007; Gross and Souleles 2002a; Gross and Souleles 2002b): spending data are captured objectively, without relying on the respondent to recall the required information and report it correctly. Other types of data contained in credit card data, such as the information supplied for credit applications, or the data provided by credit bureaus, can however be affected by errors.

Processing error may arise in classifying categories of spending, and in pooling together data from different data sources in a single dataset. In this case the variables of interest need to be defined consistently across credit card accounts from different issuers (Gross and Souleles 2002a).

Total error in estimates

We are aware of no studies that have compared data from credit and debit cards to external benchmarks to examine total error in estimates.

FEASIBILITY AND COSTS

The main obstacle to using credit card data for research purposes is that the data are proprietary. The commercial owners grant only limited, if any, access to their data and are likely to attempt to monetize their data by charging fees.

A further complication is that the organic nature of credit card data can lead to very large datasets. This is an advantage as it allows the study of low probability events, such as bankruptcy and delinquency (Gross and Souleles 2002b) but may also lead to computational problems. Researchers therefore tend to use random subsamples of cases for computational tractability (e.g. Gross and Souleles 2002a) have in some cases selected a random subsample of cases for computational tractability.

5. Credit rating data

Credit rating data provide detailed measures of credit applications, loan payments, and credit scores, based on information provided to potential lenders (see Avery et al. 2003). Credit rating data are created by credit reporting or credit reference agencies, which collect data on credit applications and credit payments, derive credit scores, and report these to potential lenders.

In the UK, there are several credit reporting agencies, providing records on over half of the economically active population (Jentzsch 2007). In the U.S. three nationwide credit reporting agencies (Equifax, Experian, and TransUnion) collect credit data. Credit records include the following types of information: identifying information (e.g. name, address, social security number), information reported by creditors on credit accounts (loan, lease, and non-credit-related bills – including utility and medical bills), information derived from money-related public records (e.g. records of bankruptcy, foreclosure, tax liens), information on credit accounts and non-credit-related

bills, identities of individuals and/or companies that request information from an individual's credit records (Avery et al. 2003).

Credit rating data have been used for research purposes on their own, or as sampling frames for surveys on consumers (e.g. Bucks and Couper 2018). For example, credit rating data have been used to create two anonymized administrative data panels and in the Home Mortgage Disclosure Act database. The New York Fed Consumer Credit Panel is a panel containing quarterly data on 5% of U.S. consumers with a social security number and a credit history, as well as the members of their households with a credit report, from 1999 to date (Lee and Klaauw 2010). The database includes information on mortgage installments and revolving credit accounts, as well as information on other loan types, including car purchase loans, bankcard loans, student loans and consumer finance loans. In addition, the Consumer Credit Panel includes information from public records, such as records of bankruptcy and tax liens, information reported by collection agencies and individual characteristics such as the year of birth and individual credit score at the end of each quarter.

The CFPB Consumer Credit Panel is a random sample of approximately 2% of credit records initially drawn in 2012 (Carroll 2015). The dataset includes information about the name of businesses where the consumer has financial records, dates of account opening, credit limits, types of accounts, balances owed, and payment histories. The panel does however not include information about mortgages, such as loan purpose, owner-occupancy, pricing, loan-to-value ratio, income and borrower demographics (Avery, Courchane and Zorn 2014). Finally, the Home Mortgage Disclosure Act (HMDA) data collected 14.4 million records from 6,913 financial institutions in 2015 (Consumer Financial Protection Bureau n.d.). These records include information on loans (including successful and declined mortgage applications, loan amounts, the type of loan, the loan purpose and, in some cases, reason for loan denial), demographic information on the applicant (e.g. race, ethnicity and sex), information on the lender (name of the lender and which agency regulates them), and information about the property (type of property, whether the owner intends to live there) and location (census tract). The HMDA data do not contain information on loan performance and has limited information on borrower credit-worthiness (Avery, Courchane and Zorn 2014). The data can be analyzed as a standalone data source, accessible from an online platform (see Essene and Byrne 2014). The data however have up to a 21-month delay in release (Avery, Courchane and Zorn 2014).

RESEARCH APPLICATIONS

Credit rating data have been used for various research projects.² For example, Bhardwaj and Sengupta (2008) use the New York Fed Consumer Credit Panel to compare subprime mortgages originating from early years (2000-2002) with subprime mortgages originating from the boom years (2004-2006), and conclude that "early years" mortgages were no less vulnerable to the environment than mortgages from 2004-2006. Brevoort and Cooper (2013) analyse the consequences for borrowers of mortgage foreclosure, that is, losing possession of a mortgaged property as a result of the borrower's failure to keep up mortgage payments. They find that while the paths of borrowers whose mortgages entered foreclosure in the period 2007-2009 is similar to that of borrowers who foreclosed earlier (2000-2006), credit score recovery is taking longer for those entering foreclosure more recently.

² For a list of papers based on the New York Fed Consumer Credit Panel please see <https://www.newyorkfed.org/microeconomics/ccp.html> .

The CFPB Consumer Credit Panel was used by Brevoort and Kambala (2015). The authors found that medical debts and other debts are not equally predictive of the subsequent credit performance of consumers: consumers with more medical than non-medical debts are more likely to over-perform their credit scores than consumers with more non-medical than medical debts. Brevoort, Grimm and Kambala (2016) use data from the CFPB Consumer Credit Panel in conjunction with data from the 2010 Decennial Census and the 2008-2012 American Community Survey, to estimate the number of credit invisible consumers and consumers with unscored credit records (i.e. consumers having insufficient credit history or lacking of a recent credit history) and their demographic characteristics. The authors estimate that in 2010, 26 million U.S. consumers (approximately 11% of the adult population) were credit invisible, and 19 million consumers (8.3% of the adult population) had credit records which were considered “unscorable” by commercially-available credit scoring models. Credit invisibility seemed associated with living in low income neighbourhoods and with Black and Hispanic ethnicity.

The CFPB Consumer Credit Panel has also been used extensively as a sampling frame, for example for the Consumer Survey of Debt Collection Experiences (CFPB, 2017). This is a national survey which collects information on the contact of consumers with debt collectors (timeframe, frequency of contacts, and nature of debt).

DATA QUALITY: IMPLICATIONS FOR TOTAL ERROR

Errors of representation

In terms of *coverage error*, that is, whether systematic sections of the target population are not included in the data, one main concern is credit invisibility and unscored credit ratings. While the target population for the New York Fed Consumer Credit Panel is U.S. residents with credit history (Lee and Klaauw 2010), researchers are usually interested in generalizing their results to the general population. The research by Brevoort, Grimm and Kambala (2016) estimated that, in 2010, the CFPB’s Consumer Credit Panel did not cover 11% of the U.S. population because of credit invisibility and 8.3% because credit records were considered “unscorable” by credit agencies. This under-coverage led to individuals with Black or Hispanic ethnicity being under-represented. Credit invisibility and prevalence of unscored credit is also associated with age: while over 80% of young adults aged 18 or 19 are credit invisible or have unscored credit, this percentage drops to 40% for adults aged 20-24 years hold, and further decreases until age 60-64.

When credit rating data are used as a stand-alone data source, data are originated passively and, thus, the respondents’ participation is not required. Conversely, when credit rating data are used as a sampling frame for surveys then the sample member must be approached and asked to actively participate. Inevitably, some sample members will decide not to participate in the survey and they may be different in terms of some characteristics from those that participate, leading to error of selectivity due to lack of participation.

Of course, an advantage of using data from credit reporting agency as a sampling frame is that the administrative records on credit rating are available both for survey respondents and non-respondents. It is therefore possible to compare respondents and non-respondents on key variables of interest for the researcher, modelling non response patterns and evaluating and eventually reducing non response bias, with the possibility of using administrative data in the construction of non-response weights (Carroll 2015).

When credit data are used as a sampling frame, relevant information can be collected through survey questions; this is an important feature, since, while some variables are more reliably obtained from credit rating data (such as, for example, credit scores or mortgage balance), other information can only be collected through survey data (e.g. attitudes, intentions, etc.).

Errors of measurement

In terms of *specification error* due to categories of items missing data, not all credit accounts are represented in credit agency reporting (Avery et al. 2003). Small retail, mortgage and finance companies, and some government agencies might not report to credit reporting agencies; furthermore, “loans extended by individuals, employers, insurance companies, and foreign entities typically are not reported” (Avery et al. 2003). In addition, creditors might not report or keep up to date the credit accounts of those borrowers who make payments as scheduled, and/or creditors may fail to notify to the credit reporting agency that an account has been closed.

In terms of *specification error* due to a different unit of analysis, in The New York Fed Consumer Credit Panel (CCP) the definition of household coincides with cohabitation. Indeed, the CCP has information on a sample of 5% of U.S. consumers and all other individuals residing at the same address. Nevertheless, the unit of analysis of interest for the researchers might be individuals who make their financial decision jointly, and this group may not reside at the same address: for example, couples may make financial decisions jointly but live at different addresses (e.g. for working related reasons), and young adults may share their house with lodgers and students with peers or friends with whom they don't take financial decisions jointly. It should be noted, however, that if individuals are jointly responsible for a credit (e.g. mortgage) a record of that credit account appears in each individual records, and an indicator is included to signal that the credit is a joint account (Avery et al. 2003). Researcher may need to aggregate individual level information at the household level, or at any other desired level of analysis.

Another example of specification error due to the unit of analysis is the presence of duplicate records in the Federal Reserve Bank of New York Consumer Credit Panel (Avery, Courchane and Zorn 2014).

In terms of measurement error due to *missing items or episodes*, credits may not appear in credit files, if creditors that are not represented in collection agency reporting do not report the credit to the reporting companies.

One example of measurement error due to *errors in reports or data capture*, is the credit reports' sensitivity to the date when the credit account is forwarded to the credit agency: a credit report will have a larger outstanding balance if forwarded to the agency just before (as opposed to just after) a payment (Avery et al. 2003).

It is not clear to which extent consumer credit reports are consistent across different credit reporting companies, and credit reports may contain errors. Early studies have estimated the number of errors in credit reports to vary widely, from 0.2 to 70% (for a review see Consumer Federation of America 2002). However the accuracy of these studies has been questioned (Jentzsch 2007), and the methodological rigor of the lower estimate of 0.2% of errors in credit reports found by a 1992 survey conducted by the consulting firm Arthur Andersen is questioned since only the records that have been amended as a consequence of a revision request on denied credit applications were considered erroneous. More recent evidence (Smith et al. 2013c) has shown that in a sample of 1000 U.S.

consumers, 26% had at least one error in their credit reports. While some of these errors are minor, others are major and, thus, may lead to the denial of credit or to less favorable credit terms. Ultimately, this phenomenon leads to more erroneous credit scores.

Measurement error due to *coding and processing error* may arise in the coding and transcribing process of administrative data. For example, in early 2000s, Avery et al. (2003) report that credit records include information from public bodies, which are not always computerized; it may require labor-intensive transcribing and recording processes to transmit this information and credit records are not always easily obtainable. Nevertheless, we expect the level of computerization of records held at public bodies to have increased in the last 15 years, and to be subject to further improvement in the future, with a resulting reduction of coding and processing error.

Total error in estimates

We are aware of no studies that have compared credit rating data to external benchmarks to examine total error in estimates.

FEASIBILITY AND COSTS

As is the case with other proprietary data, access to credit rating data is restricted. For example, access to the New York Fed Consumer Credit Panel and to the CFPB Consumer Credit Panel is limited to staff at the Federal Reserve Board or the Consumer Financial Protection Bureau (Avery, Courchane and Zorn 2014). Access to credit rating data for the purposes of sampling is presumably even more restricted as this would require handing over identifying information that would be needed to contact the panel members.

6. Scanner technologies – in-home barcode scanner data

Data from in-home barcode scanning provides detailed measures of products purchased; associated spending data are derived from receipts or prices reported by the consumer. In-home or household-based barcode scanner data are collected by asking a sample of households to scan the barcodes of purchased items over an extended period of time (Leicester 2015; Zhen et al. 2009). The products scanned are generally food and grocery items and the scanning is normally carried out after each shopping trip, using a barcode reader installed in the sample member's home (Leicester 2015). Non-barcode items such as loose fruit and vegetables need to be entered manually. New developments include the use of smartphone apps, to reduce the costs of devices and to encourage householders to scan on-the-go purchases such as snacks (Kantar Worldpanel 2016 personal communication). Shoppers are asked to provide additional information, such as when and where the products were bought, as well as transmitting the associated receipts using a digital camera or scanner and email. Some panels, such as the Consumer Network panel, provide scanner cards which can be presented at certain shops at the point-of-sale so that barcode data from purchases are uploaded automatically. In-home barcode scanner data is intended to provide a complete and detailed record of purchasing behaviours from multiple outlets visited by the householder (Zhen et al. 2009). In contrast, loyalty card data, which also relies on barcode technology, provides point-of-sale information from a specific chain of stores.

The collection of in-home scanner data has been driven by the market research industry and is relatively recent (Zhen et al. 2009) in comparison to point-of-sale scanner data, which uses similar

principles and technology but is collected in-store. Examples of in-home scanner databases are HomeScan (delivered by Nielsen and known as the Nielsen Panel in the UK and the National Consumer Panel in the U.S.), Consumer Network (delivered by Information Resources Inc. or IRI) and ShopandScan (delivered by Kantar Worldpanel). These panels are substantial: ShopandScan has a sample of approximately 30,000 households in the UK (Kantar Worldpanel personal communication, 2016) and Homescan in the U.S. includes 125,000 core panel households and 15,000 fresh foods panel members (Zhen et al. 2009). The datasets made available to analysts contain detailed information about products purchased, quantity purchased, the physical characteristics of the items and their universal product code or UPC. The product information that is scanned can be linked to additional sources such as nutritional databases which provide information such as the amount of fat and salt in food products. The data should also contain price paid, provided by the shopper, which may differ from product-price database information, as well as information about promotions or discounts associated with the purchase. The data also include information about the consumer collected through questionnaires, such as household demographics (Leicester 2015).

By its nature, in-home barcode scanning generates large volumes of longitudinal data which cover much longer periods than time-limited diary-based surveys. The panels have very large samples and capture many transactions, making it possible to observe dynamic purchasing behaviour even for low frequency products.

RESEARCH APPLICATIONS

There are a number of research applications for in-home barcode scanner data across several disciplines. Although these are related to household spending and consumption, they are not directly concerned with understanding household finances per se. Researchers interested in food consumption have used in-home barcode scanning as an alternative route to gathering home food inventories (Weinstein et al. 2006). Barcode scanning has been trialed as part of a government funded survey to measure food purchase behavior in the Food Acquisition and Purchase Survey (Leicester 2015). Economists have used the detailed longitudinal data provided by barcode scanning to analyse shocks and policy interventions such as the impact of changes in cigarette taxes on prices and consumption (Harding, Leibtag and Lovenheim 2012). Aguiar and Hurst (2007) used the 1994 Homescan data for Denver, combined with household time diaries, to show that households that shop more intensively pay lower prices for goods. Hausman and Leibtag (2007) used the 1998–2001 Homescan data to measure the effect of increased competition in shopping outlets from Walmart on consumer welfare and Shum (2004) used in-home barcode scanner data to look at the effect of advertising on brand loyalty.

Consideration has been given to whether in-home barcode scanning could help estimate household spending. Scanner data records for household expenditures over long periods of time could be used to explore how the time-limited nature of budget surveys affects the spending patterns which are observed (Leicester 2015). Scanner data has been proposed for imputation: respondent burden could be reduced by asking limited questions about aggregate category-level expenditure in budget surveys and supplementing this with more detailed information from scanners (Tucker 2011). Imputation could take advantage of known information such as which shop the goods were purchased from, because spending patterns (even within the same broad aggregate commodity such as food) may vary by store (Leicester 2015). Their more extensive use as a data collection tool is proposed in

budget surveys alongside or in place of paper diaries and recall questions (Mathiowetz, Olson and Kennedy 2011).

DATA QUALITY: IMPLICATIONS FOR TOTAL ERROR

Errors of representation

In-home barcode scanner panels depend on volunteers³ and use quota sampling methods (Harris 2005; Tucker 2011; USDA 2009; Zhen et al. 2009). They have been criticized for providing limited information about sampling methods, response rates and attrition (Leicester 2015; Perloff and Denbaly 2007; Westat 2011). As a result it is difficult to comment meaningfully on coverage error and participation error, a task made more difficult because the demographic data that is collected in panels are less comprehensive than in budget surveys (Leicester 2015).

However, since recruitment for and participation in barcode scanning panels takes place at least partly online this necessarily introduces some degree of *coverage error*, with offline groups excluded. Households also need a reasonable level of technical skills to carry out the scanning and transmission of shopping and pricing information. That said, these studies normally provide households with scanning equipment, rather than relying on smartphone ownership as is the case with receipt scanning, and this minimizes the exclusion of those who have limited access to the newest technologies.

Caveats aside, some differences have been observed between scanner panels and budget surveys although these are somewhat inconsistent. For example, in the U.S., more smaller households take part in Homescan compared to the Consumer Expenditure Survey (Huffman and Jensen 2004) but in the UK, Kantar households have more members on average than the Living Costs and Food Survey (Leicester and Oldfield 2009a). There are likely to be unobservable differences in the characteristics of households in different surveys. Lusk and Brooks' (2011) observation that scanner households are more price responsive than the general population may reflect the fact that households who agree to collect scanner data are more price conscious than the general population (Leicester 2015), rather than being explained by panel conditioning suggested earlier.

In terms of income, there is conflicting evidence in the U.S. where Homescan and IRI households have been found to have, respectively, higher and lower incomes than comparators (Huffman and Jensen 2004; Perloff and Denbaly 2007) as well as a mix of demographic differences (Zhen et al. 2009). In the UK, Worldpanel households were found to be poorer on average and more likely to be unemployed or part-time workers (Leicester and Oldfield 2009a). Leicester (2015) notes that U.S. and UK accounts reach different conclusions about the extent to which demographic variables account for the differences in spending. Panel companies derive weights to adjust over a particular period of observation to ensure that the households are representative, in the case of Kantar Worldpanel accounting for household size, housewife age, social class and region.

³ Academic studies have also been based on volunteer samples, including varying socioeconomic strata and races/ethnicities and are relatively small (Byrd-Bredbenner and Bredbenner 2010; Byrd-Bredbenner and Maurer Abbot 2009; Cuite et al. 2008a; Cuite et al. 2008b).

Errors of measurement

The potential for in-home barcode scanning to help understand household expenditure is constrained by a lack of alignment between receipts and total spending, in other words, *specification error*. The approach provides rich detail about a large number of items purchased by households, but is limited to food and grocery products including cleaning products and personal care items. However, even this is somewhat inconsistent; for example alcohol purchased off-license is included but tobacco is not. The restriction to food and groceries is not simply a reflection of the fact that many purchased goods are not barcoded. On the contrary, these panels exclude some products that do have barcodes while including some non-barcoded items such as loose fruit and vegetables. Instead, their scope is largely defined by their purpose, which is to support market research into spending on fast-moving consumer goods (FMCG). The discrepancy between the spending included in in-home barcode scanning and total spending is considerable. Leicester (2015) estimates that the set of products contained in the Worldpanel data make up something like 18% of all non-housing expenditure (falling from around 24% in 1996) and that “at best just over a third of total expenditures (by CPI weight) appear to be readily amenable to in-home scanner technology”. Duly et al. (2003) compared Nielsen Homescan data in the U.S. to Consumer Expenditure Survey (CE) diary data from 2000, and found that only 45% of the CE diary items (but 83% of food items) were covered in the Homescan data (Leicester 2015).

In addition to the specification error associated with the large number of missing categories of items, in-home barcode scanning may also be subject to specification error in terms of the unit of analysis. Ostensibly, these panels cover household spending, but it is not clear that there is a complete overlap between the products that are captured and those that are either purchased or consumed by the individuals within the household. See the discussion in Section 7 on receipt scanning.

Measurement error may further arise from households forgetting to scan individual items, particularly if they are minor purchases at convenience stores or consumed “on the go” (Leicester and Oldfield 2009b). Complete shopping trips may be missed, and under-reporting of bread and milk, which are traditional ‘top-up’ items, is given as evidence that participants are particularly likely to forget to report “top-up trips” (Leicester and Oldfield 2009b) or smaller supermarket shopping trips (Eyles, Jiang and Mhurchu 2010).

The extent of missing items or shops depends to some extent on the burden placed on respondents: burden is higher for products that are not barcoded, and also varies between different scanning technologies. Some fresh food purchases might be not entered if the household shops at farmers’ markets, butcher shops, or bakeries due to the additional burden associated with recording them manually (Zhen et al. 2009). Until 2006, all Kantar Worldpanel households were, for example, asked to report non-barcoded food and grocery purchases using a booklet of generic barcodes. Details of the product characteristics for these items (such as weight, country of origin, flavour and so on) were also entered manually via the scanner device. Compliance with reporting varies even with advanced scanner technology. For example, IRI report lower compliance with key panelists (who scan products at all retailers) compared to card panelists (who show a card at the checkout each time they visit a participating retailer) and key+card panelists who only scan purchases for retailers who do not support the card system (Kruger and Pagni 2008; Zhen et al. 2009). Where electronic data can be collected directly from the retailer this minimizes participant burden and makes this method suitable

for use with all age groups, those with poor memory and literacy levels “placing little or no burden on participants, and in being an efficient and objective measure” (Eyles, Jiang and Mhurchu 2010).

There is some evidence that compliance with scanning changes over time. Homescan only includes records after a few weeks of monitoring returns, as participating households need time to adjust to the technology (Harris 2005). However Kantar Worldpanel includes all purchases reported by households irrespective of length of participation (Leicester and Oldfield 2009b). They found expenditures were highest in the first few weeks of participation but fell away slightly such that after about 6 months households spent, on average, about 5% less than in their first week. This might be evidence of fatigue or a settling-in process in which early errors such as multiple recording which may inflate expenditures (i.e. measurement error) are reduced, or may be a genuine behavioural reaction to participation (i.e. panel conditioning). The decline in spending over a 6 month period is slight, relative to a 9% reduction between the first and second week of the Canadian Food Expenditure Survey (Ahmed, Brzozowski and Crossley 2006).

The literature also identifies several additional sources of *measurement error*. Demographic information in commercial in-home barcode scanner datasets is less comprehensive than in budget surveys. For example, Kantar WorldPanel collects basic characteristics of the main shopper in a baseline telephone interview and recollects this approximately every nine months. Income was not collected routinely until after 2008, and even then used a banded measure of gross total income collected from the main shopper (Leicester 2015). Education and employment are not asked consistently for all household members and housing tenure is not always reported. Leicester and Oldfield (2009b) also identified poor reporting of demographic transitions over time, though these could also have resulted from sampling issues. Furthermore, information from additional household members may be incomplete (Leicester 2015). In addition, imperfect information about produce prices is also identified as a source of *measurement error*. For example, with ShopandScan, pricing information is obtained from till receipts which are mailed (or emailed) to Kantar Worldpanel who match the price to the purchase record. Where no receipts are available, prices are taken from centralized databases of store- and product-specific prices, or are otherwise imputed, which is inevitably imperfect.

There are also several sources of potential *processing or coding error*. Barcode scanner data is subject to many of the same food composition database errors as traditional dietary assessment methods (Eyles, Jiang and Mhurchu 2010). Although barcodes do not need to be transcribed, reducing potential processing error, some products do not have standard UPCs, instead having retailer or synthetic UPCs, and some barcodes are not unique. As a result, information about the product may be missing or inaccurate. There may also be errors in how an item has been barcoded which could result in it being mis-categorised. The process of transcribing price information and linking it to the purchased items may also be imperfect.

Although many academic studies use commercial datasets which are already processed, a lot of processing is still necessary, particularly if linkage to nutritional or price databases needs to be carried out. Analysts report several challenges. For example, in order to make comparisons to aggregate data, analysts follow a complex and inevitably imperfect set of processes to map food and drink purchases using Kantar Worldpanel data product characteristics to the equivalent expenditure codes in the Living Cost and Food Survey records and then into commodity groups to match those defined in the UK Consumer Price Index (Leicester 2015; Leicester and Oldfield 2009b).

Barcode scanning may lead to panel conditioning and result in changes in purchase behaviour: participants may be less inclined to buy unhealthy foods or more likely to buy healthy ones and their shopping behaviour may adapt to the incentive schemes offered by panels. Lusk and Brooks (2011) carried out a web survey with a random sample from the Homescan and IRI panels and a random-digit-dialling survey of the general population and found that scanner households are more price responsive than the general population. One possible explanation is that participating in scanning may make households more aware of their purchasing behaviour and thus more price sensitive, a form of panel conditioning (Leicester 2015).

Total error in estimates

In the U.S. and UK, comparisons between household scanner data and budget surveys shows that scanner studies significantly underestimate expenditure (Duly et al. 2003; Leicester and Oldfield 2009b; Zhen et al. 2009). The short-fall is particularly noticeable in categories such as alcohol, cigarettes and those with a high proportion of non-barcoded items such as fruit and vegetables. Based on analysis of nutritional content, under-estimation may partly be explained by insufficient attention to the treatment of weeks with zero purchases (Griffith and O'Connell 2009) and may be lessened by observing purchases over a full year rather than a short period. Differences in estimates may partly result from the demographic composition of the samples although without an experimental approach it is hard to identify the key drivers (Leicester 2015).

A direct comparison of Homescan household scanner data with loyalty card data by Einav, Leibtag, and Nevo (2008) found that approximately 20% of shopping trips were not matched, suggesting inaccurate store or date information. Based on best estimates, around half the trips that were reported in the store data were not observed in Homescan and, where trips could be matched, the Homescan record appeared to miss on average 10-15% of the items purchased in the store record, most often small consumables like soft drinks which may be consumed on the way home. Substantial mismatches between Homescan reports of food expenditure and data from the Consumer Expenditure Survey were also reported by Zhen et al. (2009). Specifically, Zhen and colleagues found that higher-income households and households with more members have larger expenditure differences between datasets, suggesting higher opportunity costs for these households, leading to more skipping of purchases or trips.

Historically, the largest errors in barcode datasets were found with price records: on matched trips, the price reported in the Homescan record failed to match the loyalty card recorded price about half the time. However this seems to have been driven mostly by the way prices were imputed, relying on a database which held the price of each product in each week, which ignored store-specific discounts, offers and individual loyalty card or coupon discounts. This was less likely to be evident in studies such as Kantar Worldpanel where receipts are returned and prices matched to individual purchases and this approach was recommended for Homescan (Einav, Leibtag and Nevo 2008; Leicester 2015).

FEASIBILITY AND COSTS

Some authors argue that mobile Universal Product Code (UPC), or barcode scanning technology provides an efficient, accurate and comprehensive method for conducting home food inventories

intended to describe the household's nutrient supply, saving time over traditional written inventories (Byrd-Bredbenner and Bredbenner 2010; Byrd-Bredbenner and Maurer Abbot 2009; Stevens et al. 2008). However this approach remains resource intensive, requiring a home-visit of about two hours during which the interviewer scans the household's inventory (Weinstein et al. 2006), as well as interviewer training and equipment, adapting commercial software, purchasing database licenses, developing and maintaining a nutrient database and linking products to nutrient data (Byrd-Bredbenner and Bredbenner 2010). Costs could be reduced by giving householders hand-held scanners and the facility to transmit data electronically, but this might generate less comprehensive or accurate data.

Given the high level of investment needed, dedicated social science projects using scanning technology remain scarce. Commercial panels have grown rapidly to meet the demand for market research into shopper behavior and, in time, these datasets have become available for economic and social research. IRI made both its store and household scanner data available to academic researchers at a heavily discounted price (Bronnenberg, Kruger and Mela 2008) although the most up to date datasets retain their high premium. Kantar Worldpanel data made its household scanner data available to analysts (Leicester 2015) following the same principles. These panels minimize costs by having householders collect all the necessary data. They have, as needed, reduced the burden involved, for example for scanning non-barcoded goods. Academic users have been able to secure funding to analyse these data and consideration has been given to making the data available more widely for social scientists. Nevertheless, this approach does introduce certain limitations since the researchers do not have control over issues such as sample selection, the extent and quality of demographic data, and the conduct of methodological experiments (say, of mode effects) or policy experiments.

7. Scanner technologies – receipts

Data from scanned receipts provide detailed spending data on purchases of goods and services. Receipt scanning involves individuals or households retaining receipts which have been given to them at a point of sale and transmitting them to a central database. Receipt scanning provides information about purchases of a broad range of goods and services and basic information about each item purchased, though not the supplementary detail provided by barcodes. In the past, paper receipts were returned by post, or scanners were used to capture images which were then sent by email. More recent examples depend on smartphone apps which use the mobile device's camera to photograph receipts and automatically transmit the images collected. Supplementary activities may be required such as annotating receipts before they are scanned or providing information about non-receipted purchases. The scanned receipts are then data-entered or machine-read and coded for analysis.

Receipt scanning has been developed to serve two main commercial markets, neither of which is directly concerned with understanding household finances per se. The first commercial use is to support employees, the self-employed and businesses to manage refundable expenses and collate data required for accounting purposes and tax returns. A number of apps such as Receipt Bank, Expensify, 1Tap Receipts, Concur and Receipt Catcher, are emerging to replace low-tech approaches to this task. The motivation to complete receipt scanning for expenses is both financial (individuals cannot claim reimbursement without providing evidence of purchase) and regulatory. The detailed data contained in the receipt is not exploited, and the process is barely mentioned in the academic

literature, aside from the possible impact that the introduction of technologies may have on business efficiency (Kepczyk 2011).

The second main use of receipt scanning is to provide the market research industry with detailed expenditure data to understand shopping behaviours such as responses to special offers or advertising. We did not find published information about this emerging area, but it seems likely that this approach is being developed as a less burdensome and lower cost way of gathering shopping data on a wider range of products than full barcode scanning. The apps in this market include Receipt Hog, ReceiptPal, Yaarlo, Shopprize app and the Kantar Worldpanel snacking app. Participation is incentivized with small financial rewards for transmitting receipts. In the case of ReceiptPal these incentives are only available for receipts which have been validated, and an additional, non-financial incentives involve offering panel members access to their stored receipts which can be seen as a form of feedback. The more sophisticated versions of these apps seek permission to link to the participant's email address and extract information about online purchases from major retailers; for example, Receipt Hog links to Amazon.com. This app also incentivises participants to give permission to link data on other app users in the household, to link to Twitter and Facebook accounts. They alter the adverts that participants are exposed to and measure ad exposure, making it possible to evaluate the effect of campaigns on purchasing behaviours.

There is little information on the market penetration of these receipt-scanning apps. However, Ibotta claims to be the most frequently-used app in the U.S., with nearly 22 million downloads since it was launched in 2012 (see <https://ibotta.com/>).

RESEARCH APPLICATIONS

Until recently, the main academic research application of receipt scanning seems to have been small-scale, paper-and-post based studies which used receipts to analyse foods and beverages purchased and consumed by household members over a defined period as a method of supplementing or replacing dietary surveys (Becker 2001; DeWalt et al. 1990; French et al. 2008; Rankin et al. 1998; Ransley et al. 2001; Sekula et al. 2005). For these applications, receipts may require annotation and may need to be collected alongside other data activities such as individual dietary intake surveys (French et al. 2008) and manual records of visitors attending household meals, meals taken away from home, and foods purchased from shops not providing itemized bills (Greenwood et al. 2006). Elsewhere in this literature, receipts have been used as a "source of verification" providing objective measures of food purchases (Tang et al. 2016) and dietary reports (Eyles, Jiang and Mhurchu 2010).

As far as we are aware, the *Understanding Society* Innovation Panel spending study is the first, population representative academic survey to carry out a receipt scanning exercise to measure household spending on goods and services. Participants were asked to download an app onto their smartphone and scan receipts, directly enter spending without a receipt, or record 'no spend' days over a one month period (Jäckle et al. 2019).

The data items collected through receipt scanning vary, depending on the purpose of the activity. Business applications may ask the user to manually enter information such as date of payment, amount paid, expenditure category, VAT rate applicable and payment method and do not generally make use of more detailed information contained in the receipt itself. For market research and academic studies, on the other hand, the transcribed or coded data items provide information such as time and location of purchase, name of store, value and product details for each item, information

about price reductions or multi-purchases and whether a loyalty card was used. Where exact information is collected about the purchased product, or a barcode is embedded in the receipt, or is scanned as a supplementary activity, additional information may be linked from external databases. Commercial apps such as Receipt Hog are also able to manipulate exposure to online advertising, measure exposure and relate this to subsequent shopping behaviours.

DATA QUALITY: IMPLICATIONS FOR TOTAL ERROR

Errors of representation

Only a limited amount of evidence about representation error can be gained from receipt-based dietary studies since they have relatively small sample sizes, may be in localized geographical area (French et al. 2008) or based on purposive samples of population sub-groups. Ransley et al. (2001), for example, based their study on a random sample of Tesco Clubcard holders in one city in the UK, most likely building in selection bias from the start. Card holders were invited to participate in a study involving the collection of receipts and completion of a shopping diary for non-receipt purchases over 28 days. Of those invited, 52% expressed initial interest, of whom 63% (or 34% of those invited) reported spending the majority of their food expenses at Tesco and were willing to take part in the study. Of those meeting the criteria, 75% (or 27% of invitees) participated in the study, but the degree of participation is not specified. The authors concluded that the responding sample was broadly representative of their consumer base but noted that people aged 30-59, women, and Social Class II were over-represented. Similarly, Smith et al. (2013b) report a 71% participation rate over four weeks in a group of targeted volunteers from low income households. They also report that low food security households, where food expenditure would likely have been lower, were less likely to participate.

Since the receipt scanning activities carried out for the market research industry are not based on a sampling frame and instead rely on volunteer samples, they also do not provide a good basis from which to assess coverage, response rates or bias. No published information could be found about their participation and attrition rates and it is not possible to report on how representative these samples may be. That said, we know that taking part relies on owning a smartphone which has a suitable operating system (generally Apple or Android), having sufficient storage for the app, and having an adequate data plan, so under-representation resulting from *coverage error* is inevitable, despite the increasing availability of smartphones resulting in more households having the technology needed for receipt scanning (Steele 2015). Even among individuals who do have the necessary technology, there is also likely to be a degree of *participation error*, because people are unwilling to take part, or lack the necessary skills to download the app, or the confidence or competence to successfully do so. In addition, failure to test the compatibility of apps on all devices may result in problems uploading data (Volkova et al. 2016) and participants may change their phones resulting in them falling out of the study. Taken together this increases the risk of sample selection bias and means that the generalizability of these approaches must be examined carefully. It is reasonable to assume that coverage and non-response error may be related to socio-economic factors (ownership, phone quality, data plan), region (some rural areas have poorer connectivity) or individual characteristics associated with mobile use such as age, cognitive function, and manual dexterity. That said, in the *Understanding Society* spending study, where the sample is representative of the population and characteristics of non-responders are known, there was some evidence of bias related to socio-demographic characteristics, to use of mobile technologies, and to whether sample

members already use apps or other computerized technologies to monitor their finances, but no evidence of bias related to levels of income or household spending (Jäckle et al. 2019) .

Errors of measurement

When evaluating whether receipt scanning could be a useful technology for researching aspects of household finances, an important source of *specification error* is the fact that some types of spending do not routinely generate a receipt so not all spending will be accounted for. There are several reasons for this. Regular payments made by standing order or direct debit, whether for routine housing and utility costs or for other spending such as subscriptions for leisure activities, are not linked to a readily available receipt. Spending outside structured shopping environments, for example at market stalls, and informal expenditures on items such as music lessons, payments to a domestic cleaner or money put into a charity box often take place without a receipt being provided. Some receipts are provided by email, most commonly because they result from an online purchase, and increasingly because some shopping outlets are reducing their reliance on paper. Indeed, the limitations of receipts as a record of spending have been exacerbated in the UK by an overall decline in the production of paper receipts with a growing number of shops, including major supermarkets such as Sainsbury, only giving receipts on request. For these reasons, relying solely on receipts will result in gaps in spending records, and the accuracy of estimates of expenditure will depend on how much effort is put into gathering information about direct debits and standing orders separately, asking for summary information about non-receipted items and requesting e-receipts as well as paper receipts.

Receipt scanning is also subject to *specification error* because receipts do not make clear whether purchased items constitute spending for the individual or spending for the household, except in the simple case where a household is comprised of one person only. This issue may also be of less concern for non-food items such as leisure which are more likely to be reported at the individual level (Ransley et al. 2001) but even here there is ambiguity, for example when establishing whether clothes are bought for the purchaser or for a child in the household.

Inevitably, the impact of any specification error will vary depending on the nature of spending in a household. For example, receipt scanning will more accurately reflect total spending for small households which rely on a single weekly supermarket shop, particularly if the store used provides detailed receipts. The corollary is that it will be harder to estimate spending in larger households which have more complex shopping behaviours or which use small, local shops where receipts may provide limited information.

Measurement error can also occur when receipts are produced but not collected in-store, or are lost or forgotten before scanning (Smith et al. 2013a). Participants may fail to scan single receipts or may miss days or even whole weeks of reporting and this may be exacerbated if incentive schemes demotivate study participants from recording multiple receipts in each time-period or scanning missed receipts from the previous day or week. Completeness of data will also vary by number of individuals in the household: where this is more than one there is a greater risk of ambiguity and confusion about who is providing data and less engaged household members may fail to provide full information. With receipt scanning it is not possible to accurately separate ‘missingness’ from periods of no or low spending, although this has been attempted by asking individuals to report “no spend days” and to estimate the spending of other adult household members (Jäckle et al. 2019). That said, evidence from dietary studies suggests that receipt scanning may also *reduce* reporting

error relative to food diaries where item by item reporting of harmful foods may be under-reported due to social desirability bias (Ransley et al. 2003) and may overcome the unintentional mis-reporting of food intake (Macdiarmid and Blundell 1998) that may occur in obese people (Ransley et al. 2003). We did not find evidence showing whether this bias applies to non-food expenditure, but we could hypothesise that participants may be *less* likely to scan receipts which reveal spending on sensitive items such as alcohol or parking fines relative to summary reports which only identify the category of spending (such as food and drink or travel).

Dietary studies provide evidence about decreasing compliance over time, resulting in a decline in reported expenditure and in the number of receipts returned (Smith et al. 2013b; Weerts and Amoran 2011).

In addition, since receipts must be transmitted, transcribed and coded, there is an opportunity for *coding and processing* error, particularly given the very high volumes of data generated by each panel member. A receipt may, for example, not have sufficient detail to allow for an item to be correctly categorized, or there may be failures in the transmission or data storage. We are not aware of any literature on the challenges of consistently coding receipts and validating these activities.

Finally, *panel conditioning* may occur when study participants change their spending behavior because of social desirability bias where they seek to reduce expenditure on certain products (such as alcohol) or increase spending on others (such as fruit and vegetables). Panel members may also change their behavior in response to study incentives, for example by making smaller, more frequent shopping trips to increase the number of events they can report or by avoiding shopping trips to reduce their reporting burden, although in practice the incentives given are generally so low that this seems unlikely.

Total error in estimates

Analyses of the biases in estimates of spending based on receipt data are still ongoing for the *Understanding Society* spending study. We are not aware of other studies that have examined the combined impact of different error sources on estimates of spending based on receipts.

FEASIBILITY AND COSTS

The cost and feasibility of receipt scanning needs to be assessed in comparison with alternative approaches that meet the same research purpose. In the dietary field, annotated receipts and records are considered feasible ways of collecting comprehensive, detailed measures of household food purchasing for population-based samples and likely to be lower in cost and complexity than Universal Product Code scanning (French et al. 2008). They may also be less burdensome than asking study participants to carry out detailed self-report diaries of expenditure and food consumption. On the other hand, these methods are clearly far more arduous and time consuming than asking a set of recall questions within a survey. Regardless of the burden and costs of collecting receipts, “unless total per-shop expenditure is sufficient, this approach requires very extensive manual coding and data entry, which is time-consuming and resource intensive” (Cullen et al. 2007; French et al. 2009; Martin et al. 2006; Volkova et al. 2016:2). This is also the case with the *Understanding Society* spending study.

8. Smartphone applications

The final set of new technologies we discuss are not included in the summary table, as they mostly enable methods or approaches we have already discussed or potentially enhance tools that have already been in use for some time.

Smartphone, tablet, and PC apps are an emerging methodology for social research (Volkova et al. 2016). Smartphone apps can collect data with the active involvement of the respondent, which may mean administering survey questions or actively collecting other data types such as photographs using the additional capabilities built-in to smartphones (Lessof and Sturgis 2018), or passively, which could involve collecting movement or location data. The built-in capabilities of mobile devices to measure location and movement (GPS) and activity (accelerometry) make them ideal for passive measurement of activity in health- and travel- related research. Financial behaviour is harder to track passively, and requires more active involvement from participants, either directly (entering data) or indirectly (providing access to transaction data used for other purposes).

The market for mobile apps for managing and/or monitoring household finances is relatively recent but has grown rapidly and is already very extensive, although the published literature on these developments inevitably lags behind. Apps can broadly be divided into four main categories:

- Mobile wallets – i.e. apps that store credit and debit cards, coupons and loyalty cards, and are used for in-store payments. Examples of mobile wallets are Apple Pay, PayPal, and Google Wallet.
- Spending diaries and budgeting apps that address a similar need to the financial aggregators, but may focus on an element of the household budget such as spending and rely on manually entered information. Examples include Dollarbird, Fudget, and Goodbudget (see Sharf 2016).
- Mobile versions of financial or account aggregators – i.e. apps that scrape data from registered users' bank accounts and summarize income and expenses.
- Mobile apps that allow uploading and sharing of receipts (e.g. Receipt Bank, Expensify, Concur, Wally) and "Cashback" apps (e.g. Top Cashback), which allows the upload of till receipts in exchange for cashback on certain products purchased (Sharf 2016).

We discussed the latter two in the context of financial aggregators and receipt scanning. Mobile wallets include data on expenditures, shop check-in, coupons, and accumulated loyalty points (Fundinger 2016). We are not aware of any research studies using mobile wallets, so focus on spending diaries here.

Diary studies are self-report instruments to collect data repeatedly on ongoing experiences (Bolger, Davis and Rafaeli 2003). Diaries are used to collect data on a variety of different topics, for example, travel, time use, food intake and spending.

The spending diary implemented in the U.S. Consumer Expenditure Survey collects expenditure data on small, frequently purchased items, including food and beverages, housekeeping supplies and services, non-prescription drugs, and personal care products and services (Bureau of Labour Statistics 2016). Paper and pencil spending diaries are still the norm nowadays though they co-exist and, in some cases, are gradually being supplemented with or replaced by electronic formats (e.g. Erhard et al. 2016; Ralph and Manclossi 2016). The growth of smartphone ownership has also encouraged the implementation of spending diaries using mobile apps. Mobile apps offer additional technical

features that cannot be implemented with paper diaries. For example respondents can be prompted to enter data in their diary when sensors on their device (such as GPS) detect a certain situation, such as entering a particular store (Iida et al. 2012).

RESEARCH APPLICATIONS

Several large scale national surveys such as the U.S. Consumer Expenditure Survey and the UK Living Cost and Food Survey are considering moving to electronic diaries (Erhard et al. 2016; Ralph and Manclossi 2016) and Statistics Austria is currently developing an app-based diary for their Household Budget Survey.

DATA QUALITY: IMPLICATIONS FOR TOTAL ERROR

Errors in representation

Coverage error is a key concern in studies that rely on participants installing apps on their own smartphones: not everyone has access to a smartphone and those that do not could be different from those that do on key variables of interest. Some studies have loaned participants a mobile device (e.g. Antoun et al. 2018; Scherpenzeel 2017) to overcome coverage problems, however this approach is not common. Smartphone coverage is however still increasing. According to Ofcom statistics, the proportion of adults in the UK with a smartphone has continued to increase, from 71% (Q1 of 2016) to 76% (Q1 of 2017).

Non-response error is a further source of error among those who do have a smartphone, whether their own or loaned. Non-response may occur for several reasons, some of which are the same as in any survey: refusal due to lack of time, lack of interest, or privacy and data security concerns. Additional reasons for non-response are related to ability and willingness to download and use smartphone apps, availability of storage space on devices, and constraints of mobile data plans.

Indeed not all smartphone owners use all technical features of their devices, and not all use apps. According to data from the UK *Understanding Society* Innovation Panel in 2017, 69% of smartphone users installed new apps on their devices, 60% used smartphones for online shopping, and 59% for online banking. Even if they do use apps, they may not be willing or have the storage space to install one for a survey. When asked whether they would be willing to answer survey questions via a smartphone app 46% of Innovation Panel respondents in 2017 said they would be very or somewhat willing to do so. In the *Understanding Society* spending study that used an app to take pictures of shopping receipts, participation in the study was clearly related to usage of mobile devices. For example, sample members who used apps to check bank statements were over-represented by 20 percentage points (Jäckle et al. 2019; Wenz, Couper and Jäckle 2019).

Errors of measurement

Measurement error may arise in spending diaries if the respondent misreports the information in the diary, either by misreporting amounts or spending categories, or by omitting items. This can result from diary fatigue (Crossley and Winter 2014) or from failure to report spending in the diary close to the time of making the purchase. Ideally respondents fill in their spending diaries close in time to making purchases. However in practice this is not always the case; some respondents complete the diary retrospectively at the end of the reporting period, rather than completing it every day (Silberstein and Scott 1991). App-based diaries are expected to support real-time reporting:

participants can, in principle, enter their purchases in the app at the point of payment. This should reduce the likelihood of omissions and recall errors in the reporting of amounts and item descriptions. The extent to which participants comply as intended however depends on the usability of the app. In studies where participants use their own devices, differences in the technical specification, such as processing speed or storage space, might also affect participants' compliance with the task. We are not aware of any comparative studies of reporting behaviours in paper versus app-based diaries.

Panel conditioning is another potential source of error. Whether on paper or in an app, continuous reporting of spending over time might change the participants' spending or reporting patterns (Bolger, Davis and Rafaeli 2003; Iida et al. 2012). We are however not aware of any comparative studies that have examined panel conditioning in paper versus app-based diaries.

Processing error can be reduced in electronic diaries. The coding of item descriptions in paper diaries relies on respondents looking up and entering classification codes, or on coding text descriptions handwritten by participants. Electronic diaries can be programmed with look-up functions or can make use of barcode scanning linked to product data bases.

Total error in estimates

We are aware of no studies that have compared data collected with app-based diaries to external benchmarks to examine total error in estimates.

FEASIBILITY AND COSTS

If the data collection is based on bespoke smartphone apps programmed for the study, feasibility and costs are associated with programming, testing and implementing the app. If data collection is based on already existing smartphone apps, feasibility and costs are associated with accessing data from the app owners.

As smartphone ownership and use of apps is not universal, it may currently still be necessary to offer browser-based alternatives to app diaries. This implies additional costs for programming, testing and implementation.

9. Conclusion and outlook

Interest in finding new tools and methods for measuring household budgets continues to grow. For example, in 2017 Eurostat commissioned a Task Force on "Innovative Tools and Sources" for European Household Budget Surveys⁴, focusing on modernising data collection tools. Among the strands of modernisation to be considered by the Task Force are the following:

- Collection of survey data from households and individuals using various modern IT tools, smart devices and e-technologies
- Direct data collection from new sources such as administrative data or big data (including privately owned microdata)

⁴[https://circabc.europa.eu/webdav/CircaBC/ESTAT/hbs/Library/Task%20Force%20on%20Innovative%20Tools%20and%20Sources%20for%20HBS/TERMS%20OF%20REFERENCE%20_HBS%20-%20%20final\(0\).pdf](https://circabc.europa.eu/webdav/CircaBC/ESTAT/hbs/Library/Task%20Force%20on%20Innovative%20Tools%20and%20Sources%20for%20HBS/TERMS%20OF%20REFERENCE%20_HBS%20-%20%20final(0).pdf)

- Possibility of using data from credit/ debit cards, loyalty cards and cash register data
- Collection and scanning of receipts to relieve the burden of data entry from the respondent

This literature review addresses many of the issues that are the focus of this Task Force. Key motivations for exploring these alternatives are 1) the desire for richer detail and better quality data on the household budget, and 2) the perceived infeasibility of collecting data about the entire household budget within one survey.

This review has focused on a number of specific data sources and approaches. We have used the Total Survey Error (TSE) framework as an organising tool, but note that there are some extensions of the framework necessary to accommodate non-traditional data sources and approaches. Some types of errors become more salient when considering organic data (e.g. specification error), while others may be less relevant (e.g. sampling error). Sometimes it is difficult to distinguish between coverage error (does not have access to the technology) and non-response error (chooses not to use the technology), and treating both as selection errors or participation errors may be appropriate in some circumstances. In other cases an elaboration of non-response error may be necessary for more active users of new technology. There are many more points at which non-response may occur, from initial willingness and consent, to initial compliance in downloading or activating the tool, to ongoing adherence with the protocol. In this way, the distinction between unit non-response as an error of representation and missing data (episodes, transactions, etc.) as measurement error is more blurred when considering the new data sources. Nonetheless, with some modification, the TSE framework remains useful for organising a consideration of the data quality and inferential implications of these new data sources and methods.

Process generated data and new technologies present exciting new opportunities to rethink the way data about household finances are collected in surveys. Depending on the data source or technology they can provide a broad range of financial information (such as income and spending data from bank accounts), or a narrow view of particular aspects (such as spending in a particular outlet from loyalty card data). In all cases they can provide more granular detail over longer periods of time than can be collected with survey questions or paper diaries. The financial transactions that are recorded by a particular process or technology are also likely to be recorded more accurately than when respondents are asked to recall their transactions.

This literature review highlights emerging examples of how these new data sources and approaches are being used to address substantive research questions. In several cases they offer opportunities for analyses that may not be possible with existing (survey) data sources. However, our primary focus is on how these approaches can be used to replace or supplement survey data, in particular to obtain good quality estimates of key household financial behaviours. A key conclusion is that while they may be useful supplements to survey data, these alternative sources are not yet ready to replace survey data.

Focusing on the supplementary role of these new data sources are technologies, we have identified the broad approaches in the literature review:

- Linking survey data with process generated data
- Using process generated data as sampling frames for surveys
- Asking sample members to use new technologies to collect or provide supplementary data.

In all cases the implications for total survey error require careful thought and examination.

A key challenge for representation is that of access to or use of the technologies or data sources described here. In contrast to health surveys, significantly fewer people use mobile devices to keep track of their expenditures than use health tracking apps or devices. In addition, people use a wide variety of instruments to make financial transactions, and measuring financial behaviour is more complex than, say, measuring physical activity or travel. Further, financial behaviour is viewed as more sensitive than many other things we ask about (Couper et al. 2010), presenting challenges for gaining consent to measure these activities, whether using active or passive approaches.

A key challenge for measurement is that the survey data and process generated are often designed for very different purposes, and using one data source to replace another requires knowledge of the data generation process in each case. Understanding inclusion or exclusion rules and definitions of key concepts are critical steps in combining data sources.

As the digital economy continues to grow, and as every aspect of peoples' financial lives is increasingly leaving a digital trace, some of these concerns may diminish. While these new data sources and technologies may not be ready to replace designed data collection in surveys in the short run, continued research on the approaches we reviewed here will help identify the type and magnitude of the errors that may arise, and identify ways to mitigate those errors.

We have identified a number of key gaps in the literature. Our knowledge of the various error sources associated with these alternative approaches to measuring financial behaviour is very limited. Only recently have there been attempts to try new ways of gathering such data (Angrisani, Kapteyn and Samek 2018; Jäckle et al. 2019; Wenz, Couper and Jäckle 2019), while in other areas (e.g. in-home scanning) a lot of the data is proprietary and little is known about the error sources. There is much work that remains to be done to understand how those who are willing to use these technologies may differ from those who are not, how to increase the consent, compliance, and adherence rates for using these tools, and how participating in such additional measurement activities may affect ongoing survey participation (in terms of both attrition and panel conditioning). Finally, the measurement properties of these new tools and technologies are not yet well understood, and careful comparisons of the data to alternative measures or benchmarks is needed. Nonetheless, there are many potential uses of such data other than fully replacing survey estimates for general populations. It is important to evaluate these data sources in terms of their fitness for use (Biemer and Lyberg 2003), rather than against some absolute standard. That is, decisions about utility and value of alternatives should not only consider quality, but also costs – both human and financial.

As these technologies continue to evolve and are used by larger segments of the population, and as our knowledge of how best to implement these methods grows, we are likely to see increased adoption of these new data sources and technologies to enhance and extend the measurement of a range of financial behaviours in surveys. At the current time, much work remains to be done before we can fully exploit these new data sources and technologies. It is hoped that this review of the extant literature will help facilitate that work.

References

- Agarwal, S., C. Liu, and N.S. Souleles (2007) The Reaction of Consumer Spending and Debt to Tax Rebates - Evidence from Consumer Credit Data. *Journal of Political Economy*, 115(6):986-1019.
- Aguiar, M., and E. Hurst (2007) Life-Cycle Prices and Production. *American Economic Review*, 97(5):1533-59.
- Ahmed, N., M. Brzozowski, and T.F. Crossley (2006) Measurement Errors in Recall Food Consumption Data. IFS Working Paper 06/21 [online]. London: Institute for Fiscal Studies. Available at: www.ifs.org.uk/publications/3752 [Accessed: 25/01/2018].
- Andreyeva, T., J. Luedicke, K.E. Henderson, and A.S. Tripp (2012) Grocery Store Beverage Choices by Participants in Federal Food Assistance and Nutrition Programs. *American Journal of Preventive Medicine*, 43(4):411-18.
- Angrisani, M., A. Kapteyn, and S. Samek (2018) Real Time Measurement of Household Electronic Financial Transactions in a Population Representative Panel. Paper prepared for the 35th IARIW General Conference, Copenhagen [online]. Available at: <http://www.iariw.org/copenhagen/angrisani.pdf> [Accessed 18/12/2018].
- Antoun, C., F.G. Conrad, M.P. Couper, and B.T. West (2018) Simultaneous Estimation of Multiple Sources of Error in a Smartphone-Based Survey. *Journal of Survey Statistics and Methodology*, published online first, DOI: <https://doi.org/10.1093/jssam/smy002>.
- Avery, R.B., P.S. Calem, G.B. Canner, and R.W. Bostic (2003) An Overview of Consumer Data and Credit Reporting. Federal Reserve Bulletin 2003, February [online]. Washington: The Federal Reserve Board. Available at: <https://www.federalreserve.gov/pubs/bulletin/2003/0203lead.pdf>. [Accessed: 25/01/2018].
- Avery, R.B., M.J. Courchane, and P. Zorn (2014) The Creation of the National Mortgage Database. in *What counts. Harnessing data for America's communities*, edited by N. Cytron, K.L.S. Pettit, and G.T. Kingsley: Federal Reserve Bank of San Francisco and Urban Institute.
- Baker, S.R. (2015) Debt and the Consumption Response to Household Income Shocks. [online]. Available at SSRN: <https://ssrn.com/abstract=2541142> [Accessed: 18/12/2018]:
- Baugh, B., I. Ben-David, and H. Park (2014) Disentangling Financial Constraints, Precautionary Savings, and Myopia: Household Behaviour Surrounding Federal Tax Returns. NBER working paper No. 19783 [online]. Cambridge: National Bureau of Economic Research. Available at: <http://www.nber.org/papers/w19783.pdf>. [Accessed: 25/01/2018].
- Becker, W. (2001) Comparability of Household and Individual Food Consumption Data – Evidence from Sweden. *Public Health Nutrition*, 4(5b):1177-82.
- Bhardwaj, G., and R. Sengupta (2008) Subprime Loan Quality. Working Paper 2008-036E [online]. St. Louis, MO: Federal Reserve Bank of St. Louis. Available at: <https://files.stlouisfed.org/files/htdocs/wp/2008/2008-036.pdf>. [Accessed: 25/01/2018].
- Biemer, P.P., and L.E. Lyberg (2003) *Introduction to Survey Quality*. New York: Wiley.
- Bolger, N., A. Davis, and E. Rafaeli (2003) Diary Methods: Capturing Life as It Is Lived. *Annual Review of Psychology*, 54(1):579-616.
- Brevoort, K.P., and C.R. Cooper (2013) Foreclosure's Wake: The Credit Experiences of Individuals Following Foreclosure. *Real Estate Economics*, 41(4):747-92.
- Brevoort, K.P., P. Grimm, and M. Kambala (2016) Credit Invisibles and the Unscored. *Cityscape*, 18(2):9-34.
- Brevoort, K.P., and M. Kambala (2015) Are All Collections Equal? The Case of Medical Debt. *Journal of Credit Risk*, 11(4):73-97.
- Brewer, M., B. Etheridge, and C. O'Dea (2017) Why Are Households That Report the Lowest Incomes So Well-Off. *Economic Journal*, 127(605):F24-F49.
- Bronnenberg, B.J., M.W. Kruger, and C.F. Mela (2008) Database Paper—the Iri Marketing Data Set. *Marketing Science*, 27(4):745-48.

- Bucks, B., and M.P. Couper (2018) The Fine Print: The Effect of Legal/Regulatory Language on Mail Survey. *Survey Practice*, 11(2).
- Bureau of Labour Statistics (2016) Consumer Expenditures in 2016. BLS Report 1073 [online]. Washington DC: U.S. Bureau of Labor Statistics. Available at: <https://www.bls.gov/opub/reports/consumer-expenditures/2016/pdf/home.pdf> [Accessed: 11/03/2019].
- Byrd-Bredbenner, C., and C.A. Bredbenner (2010) Assessing the Home Food Environment Nutrient Supply Using Mobile Barcode (Universal Product Code) Scanning Technology. *Nutrition & Food Science*, 40(3):305-13.
- Byrd-Bredbenner, C., and J. Maurer Abbot (2009) Differences in Food Supplies of U.S. Households with and without Overweight Individuals. *Appetite*, 52(2):479-84.
- Byrom, J., T. Hernandez, D. Bennison, and P. Hooper (2001) Exploring the Geographical Dimension in Loyalty Card Data. *Marketing Intelligence & Planning*, 19(3):162-70.
- Carroll, C. (2015) The Cfpb Consumer Credit Panel: Direct Use and as a Sampling Frame. Paper presented at the Federal Economic Statistics Advisory Committee Meeting, June, Washington. Available at: https://www2.census.gov/about/partners/fesac/2015-06-12/Carroll_Presentation.pdf [Accessed: 25/01/2018].
- Consumer Federation of America (2002) Credit Score Accuracy and Implications for Consumers. [online]. Washington: Consumer Federation of America National Credit Reporting Association. Available at: https://consumerfed.org/pdfs/121702CFA_NCRA_Credit_Score_Report_Final.pdf [Accessed: 25/01/2018].
- Consumer Financial Protection Bureau (n.d.) The Home Mortgage Disclosure Act.
- Cortiñas, M., M. Elorz, and J.M. Múgica (2008) The Use of Loyalty-Cards Databases: Differences in Regular Price and Discount Sensitivity in the Brand Choice Decision between Card and Non-Card Holders. *Journal of Retailing and Consumer Services*, 15(1):52-62.
- Couper, M.P., E. Singer, F.G. Conrad, and R.M. Groves (2010) Experimental Studies of Disclosure Risk, Disclosure Harm, Incentives, and Survey Participation. *Journal of Official Statistics*, 26(2):287-300.
- Crosman, P. (2015) The Truth Behind the Hubbub over Screen Scraping. in *American Banker*.
- Crossley, T.F., and J.K. Winter (2014) Asking Households About Expenditures: What Have We Learned? Pp. 23-50 in *Improving the Measurement of Consumer Expenditures*: University of Chicago Press.
- Cuite, C.L., S.D. Schefske, A. Bellows, T. Vivar, C. Byrd-Bredbenner, and W.K. Hallman (2008a) Emergency Preparedness and Food Storage among Mexican Families in New Jersey. Paper presented at the Health across Borders, 2nd Annual Conference.
- Cuite, C.L., S.D. Schefske, C. Byrd-Bredbenner, A. Bellows, T. Vivar, E.M. Randolph, and W.K. Hallman (2008b) Auditing Kitchens of New Jersey Families: One Methodology for Diverse Populations. Presented at Restoring our Urban and Rural Communities with Food, 12th Annual Conference of the Community Food Security Coalition.
- Cullen, K., T. Baranowski, K. Watson, T. Nicklas, J. Fisher, S. O'Donnell, J. Baranowski, N. Islam, and M. Missaghian (2007) Food Category Purchases Vary by Household Education and Race/Ethnicity: Results from Grocery Receipts. *Journal of the American Dietetic Association*, 107(10):1747-52.
- DeWalt, K., S. D'Angelo, M. McFadden, F. Danner, M. Noland, and J. Kotchen (1990) The Use of Itemized Register Tapes for Analysis of Household Food Acquisition Patterns Prompted by Children. *Journal of the American Dietetic Association*, 90(4):559-62.
- Duly, A., T. Garner, E. Keil, S. Reyes-Morales, and C. Wirth (2003) The Consumer Expenditure Survey and Acielsen Survey: A Data Comparison Study. Bureau of Labor Statistics, Washington DC.
- Dynan, K.E., J. Skinner, and S.P. Zeldes (2004) Do the Rich Save More? *Journal of Political Economy*, 112(2):397-444.

- Einav, L., E. Leibtag, and A. Nevo (2008) Not-So-Classical Measurement Errors: A Validation Study of Homescan. NBER Working Paper 14436 [online]. Cambridge, MA: National Bureau of Economic Research. Available at: <https://www.nber.org/papers/w14436> [Accessed: 18/12/2018]:
- Erhard, L., B. McBride, P. S., and T. L. (2016) Proof of Concept Test for the Consumer Expenditure Survey. Results on Respondent Cooperation, Select Interview and Diary Characteristics, and Respondent Experience. [online]. Washington DC: Bureau of Labor Statistics. Available at: https://www.bls.gov/cex/research_papers/pdf/proof-of-concept-test-final-report-public-revised.pdf [Accessed: 11/03/2019].
- Essene, R., and M. Byrne (2014) Applying Technology Advances to Improve Public Access to Mortgage Data. in *What counts. Harnessing data for America's communities*, edited by N. Cytron, K.L.S. Pettit, and G.T. Kingsley: Federal Reserve Bank of San Francisco and Urban Institute.
- Eyles, H., Y. Jiang, and C.N. Mhurchu (2010) Use of Household Supermarket Sales Data to Estimate Nutrient Intakes: A Comparison with Repeat 24-Hour Dietary Recalls. *Journal of the American Dietetic Association*, 110(1):106-10.
- Felgate, M., A. Fearne, S. Di Falco, and M. Garcia Martinez (2012) Using Supermarket Loyalty Card Data to Analyse the Impact of Promotions. *International Journal of Market Research*, 54(2):221-40.
- Financial Conduct Authority (2015) Cash Savings Market Study Report: Part I: Final Findings Part II: Proposed Remedies. Market Study Report No. MS14/2.3 [online]. London: Financial Conduct Authority Available at: <https://www.fca.org.uk/publication/market-studies/cash-savings-market-study-final-findings.pdf>. [Accessed: 25/01/2018].
- French, S.A., S.T. Shimotsu, M. Wall, and A.F. Gerlach (2008) Capturing the Spectrum of Household Food and Beverage Purchasing Behavior: A Review. *Journal of the American Dietetic Association*, 108(12):2051-58.
- French, S.A., M. Wall, N.R. Mitchell, S.T. Shimotsu, and E. Welsh (2009) Annotated Receipts Capture Household Food Purchases from a Broad Range of Sources. *International Journal of Behavioral Nutrition and Physical Activity*, 6(1):37-48.
- Friedman, M. (1957) The Permanent Income Hypothesis. Pp. 20-37 in *A Theory of the Consumption Function*: Princeton University Press.
- Funderinger, D. (2016) Mobile Wallet Analytics and Personalization: The Value of Data. in *Mobile Business Insights [online]*. 10 August. Available at: <https://mobilebusinessinsights.com/2016/08/mobile-wallet-analytics-personalization-value-data/>. [Accessed: 25/01/2018].
- Gelman, M., S. Kariv, M.D. Shapiro, D. Silverman, and S. Tadelis (2014) Harnessing Naturally Occurring Data to Measure the Response of Spending to Income. *Science*, 345(6193):212-15.
- Greenwood, D.C., J.K. Ransley, M.S. Gilthorpe, and J.E. Cade (2006) Use of Itemized Till Receipts to Adjust for Correlated Dietary Measurement Error. *American Journal of Epidemiology*, 164(10):1012-18.
- Griffith, R., and M. O'Connell (2009) The Use of Scanner Data for Research into Nutrition. *Fiscal Studies*, 30(3-4):339-65.
- Gross, D.B., and N.S. Souleles (2002a) Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data. *Quarterly Journal of Economics*, 117(1):149-85.
- (2002b) An Empirical Analysis of Personal Bankruptcy and Delinquency. *Review of Financial Studies*, 15(1):319-47.
- Groves, R.M. (1989) *Survey Errors and Survey Costs*. New York, NY: Wiley & Sons.
- (2011) Three Eras of Survey Research. *Public Opinion Quarterly*, 75(5):861-71.
- Groves, R.M., and L. Lyberg (2010) Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, 74(5):849-79.

- Harding, M., E. Leibtag, and M.F. Lovenheim (2012) The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data. *American Economic Journal: Economic Policy*, 4(4):169-98.
- Harris, J.M. (2005) Using Homescan Data and Complex Survey Design Techniques to Estimate Convenience Food Expenditures. Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting.
- Hausman, J., and E. Leibtag (2007) Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart. *Journal of Applied Econometrics*, 22(7):1157-77.
- Hornibrook, S., C. May, and A. Fearne (2015) Sustainable Development and the Consumer: Exploring the Role of Carbon Labelling in Retail Supply Chains. *Business Strategy and the Environment*, 24(4):266-76.
- Huffman, S.K., and H.H. Jensen (2004) Demand for Enhanced Foods and the Value of Nutritional Enhancements of Food: The Case of Margarine. Paper presented at the American Agricultural Economics Association Annual Meeting, Denver, Colorado (August 1-4). Available at: <http://ageconsearch.umn.edu/bitstream/20205/1/sp04hu05.pdf>. [Accessed 25/01/2018].
- Hurst, E., G. Li, and B. Pugsley (2014) Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed. *Review of Economics and Statistics*, 96(1):19-33.
- Iida, M., P.E. Shrout, J.P. Laurenceau, and N. Bolger (2012) Using Diary Studies in Psychological Research. Pp. 277-305 in *Apa Research Method in Psychology: Foundations, Planning, Measures and Psychometrics*, edited by H. Cooper. Washington DC: American Psychological Association.
- Jäckle, A., J. Burton, M.P. Couper, and C. Lessof (2019) Participation in a Mobile App Survey to Collect Expenditure Data as Part of a Large-Scale Probability Household Panel: Coverage and Participation Rates and Biases. *Survey Research Methods*, 13(1):23-44.
- Jentsch, N. (2007) Financial Privacy. An International Comparison of Credit Reporting Systems. New York: Springer.
- Juhl, H.J., M.H.J. Fenger, and J. Thøgersen (2017) Will the Consistent Organic Food Consumer Step Forward? An Empirical Analysis. *Journal of Consumer Research*, 44(3):519-35.
- Kepeczyk, R.H. (2011) Digital Administrative Capture. *Practicing CPA*, 36(7):6-7.
- Kruger, M.W., and D. Pagni (2008) Iri Academic Data Set Description Version 1.314. Information Resources Incorporated, Chicago.
- Kuchler, T. (2015) Sticking to Your Plan: Empirical Evidence on the Role of Present Bias for Credit Card Paydown. SIEPR Discussion Paper 12-025 [online]. Stanford, CA: Stanford Institute for Economic Policy Research. Available at: <https://ssrn.com/abstract=2629158> [Accessed: 25/01/2018].
- Lee, D., and W.v.d. Klaauw (2010) An Introduction to the Frbny Consumer Credit Panel. Federal Reserve Bank of New York Staff Report no. 479 [online]. New York: Federal Reserve Bank of New York. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1719116. [Accessed: 25/01/2018].
- Leicester, A. (2015) The Potential Use of in-Home Scanner Technology for Budget Surveys. Pp. 441 – 91 in *Improving the Measurement of Consumer Expenditures*, edited by C.D. Carroll, T.F. Crossley, and J. Sabelhaus: University of Chicago Press.
- Leicester, A., and Z. Oldfield (2009a) Household Panel Data, Scanner Data, Expenditure, Food, Duration Models, Attrition. IFS Working Paper 09,09 [online]. London: Institute for Fiscal Studies. Available at: <https://www.econstor.eu/handle/10419/47514> [Accessed: 18/12/2018].
- Leicester, A., and Z. Oldfield (2009b) Using Scanner Technology to Collect Expenditure Data. *Fiscal Studies*, 30(3-4):309-37.
- Lessof, C., and P. Sturgis (2018) New Kinds of Survey Measurement. in *The Palgrave Handbook of Survey Research*, edited by D. Vannette and J. Krosnick: Palgrave Macmillan.
- Lusk, J.L., and K. Brooks (2011) Who Participates in Household Scanning Panels? *American Journal of Agricultural Economics*, 93(1):226-40.

- Lynn, P., A.E. Jäckle, S.P. Jenkins, and E. Sala (2012) The Impact of Questioning Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation Study. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 175(1):289-308.
- Macdiarmid, J., and J. Blundell (1998) Assessing Dietary Intake: Who, What and Why of under-Reporting. *Nutrition Research Reviews*, 11(2):231-53.
- Martin, S.L., T. Howell, Y. Duan, and M. Walters (2006) The Feasibility and Utility of Grocery Receipt Analyses for Dietary Assessment. *Nutrition Journal*, 5(1):10-16.
- Mathiowetz, N., K. Olson, and C. Kennedy (2011) Redesign Options for the Consumer Expenditure Survey. Report prepared for the National Academy of Sciences, Washington DC.
- Meyer, B.D., and J.X. Sullivan (2003) Measuring the Well-Being of the Poor Using Income and Consumption. *Journal of Human Resources*, 38 Special Issue on Income Volatility and Implications for Food Assistance Programs:1180-220
- Newing, A., G. Clarke, and M. Clarke (2014) Exploring Small Area Demand for Grocery Retailers in Tourist Areas. *Tourism Economics*, 20(2):407-27.
- Olafsson, A., and M. Pagel (2018) The Liquid Hand-to-Mouth: Evidence from a Personal Finance Management Software. *Review of Financial Studies*, 31(11):4398-446.
- Panzone, L., D. Hilton, L. Sale, and D. Cohen (2016) Socio-Demographics, Implicit Attitudes, Explicit Attitudes, and Sustainable Consumption in Supermarket Shopping. *Journal of Economic Psychology*, 55:77-95.
- Perloff, J.M., and M. Denbaly (2007) Data Needs for Consumer and Retail Firm Studies. *American Journal of Agricultural Economics*, 89(5):1282-87.
- Peters, B. (2016) If Banks Fear Screen Scraping, Why Are They Fighting the Alternative? in *American Banker*.
- Ralph, J., and S. Manclossi (2016) Living Costs and Food Survey. National Statistics Quality Review: Series (2) Report 3 [online]. Newport: Office for National Statistics. Available at: file:///C:/Users/aejack/AppData/Local/Temp/lcfnsqreport.pdf [Accessed: 11/03/2019].
- Rankin, J.W., R.A. Winett, E.S. Anderson, P.G. Bickley, J.F. Moore, M. Leahy, C.E. Harris, and R.E. Gerkin (1998) Food Purchase Patterns at the Supermarket and Their Relationship to Family Characteristics. *Journal of Nutrition Education*, 30(2):81-88.
- Ransley, J.K., J.K. Donnelly, H. Botham, T.N. Khara, D.C. Greenwood, and J.E. Cade (2003) Use of Supermarket Receipts to Estimate Energy and Fat Content of Food Purchased by Lean and Overweight Families. *Appetite*, 41(2):141-48.
- Ransley, J.K., J.K. Donnelly, T.N. Khara, H. Botham, H. Arnot, D.C. Greenwood, and J.E. Cade (2001) The Use of Supermarket Till Receipts to Determine the Fat and Energy Intake in a Uk Population. *Public Health Nutrition*, 4(6):1279-86.
- Scherpenzeel, A. (2017) Mixing Online Panel Data Collection with Innovative Methods. Pp. 27-49 in *Methodische Probleme Von Mixed-Mode-Ansätzen in Der Umfrageforschung*, edited by S. Eifler and F. Faulbaum. Wiesbaden: Springer.
- Sekula, W., M. Nelson, K. Figurska, M. Oltarzewski, R. Weisell, and L. Szponar (2005) Comparison between Household Budget Survey and 24-Hour Recall Data in a Nationally Representative Sample of Polish Households. *Public Health Nutrition*, 8(4):430-39.
- Sharf, S. (2016) 12 Free Apps to Track Your Spending and How to Pick the Best One for You. in *Forbes [online]*. 2 March. <https://www.forbes.com/sites/samanthasharf/2016/03/02/12-free-apps-to-track-your-spending-and-how-to-pick-the-best-one-for-you/#15e812a54453> [Accessed: 23/04/2018].
- Shum, M. (2004) Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market. *Journal of Economics & Management Strategy*, 13(2):241-72.
- Sidel, R. (2015) Big Banks Lock Horns with Personal-Finance Web Portals. J.P. Morgan, Wells Fargo Are Snarling the Flow of Data to Popular Websites That Help Consumers Manage Their Finances. in *The Wall Street Journal [online]*. 4 November. <https://www.wsj.com/articles/big-banks-lock-horns-with-personal-finance-web-portals-1446683450> [Accessed 25/01/2018].

- Silberstein, A.R., and S. Scott (1991) Expenditure Diary Surveys and Their Associated Errors. Pp. 303–26 in *Measurement Error in Surveys*, edited by P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman. Hoboken, NJ: Wiley.
- Smith, C., W.R. Parnell, R.C. Brown, and A.R. Gray (2013a) Balancing the Diet and the Budget: Food Purchasing Practices of Food-Insecure Families in New Zealand. *Nutrition & Dietetics*, 70(4):278-85.
- Smith, C., W.R. Parnell, R.C. Brown, and A.R. Gray (2013b) Providing Additional Money to Food-Insecure Households and Its Effect on Food Expenditure: A Randomized Controlled Trial. *Public Health Nutrition*, 16(8):1507-15.
- Smith, L.D., M. Staten, T. Eysell, M. Karig, B.A. Freeborn, and A. Golden (2013c) Accuracy of Information Maintained by Us Credit Bureaus: Frequency of Errors and Effects on Consumers' Credit Scores. *Journal of Consumer Affairs*, 47(3):588-601.
- Steele, R. (2015) An Overview of the State of the Art of Automated Capture of Dietary Intake Information. *Critical Reviews in Food Science and Nutrition*, 55(13):1929-38.
- Stevens, J., M. Bryant, J. Borja, V. Antony, and M. Bentley (2008) Symposium: Household Food Purchase Behaviors, Development of an Exhaustive Home Food Availability Inventory. Paper presented at the International Society for Behavioral Nutrition and Physical Activity conference.
- Sturley, C., A. Newing, and A. Heppenstall (2018) Evaluating the Potential of Agent-Based Modelling to Capture Consumer Grocery Retail Store Choice Behaviours. *International Review of Retail Distribution and Consumer Research*, 28(1):27-46.
- Tang, W., A. Aggarwal, Z. Liu, M. Acheson, C.D. Rehm, A.V. Moudon, and A. Drewnowski. (2016) Validating Self-Reported Food Expenditures against Food Store and Eating-out Receipts." 70, No. 3 (2016): 352. *European Journal of Clinical Nutrition*, 70(3):352-57.
- Tin, S.T., C.N. Mhurchu, and C. Bullen (2007) Supermarket Sales Data: Feasibility and Applicability in Population Food and Nutrition Monitoring. *Nutrition Reviews*, 65(1):20-30.
- Tucker, C. (2011) Using Multiple Data Sources and Methods to Improve Estimates in Surveys. Paper presented at BLS Household Survey Producers Workshop, Washington. Available at: http://www.bls.gov/cex/hhsrvywrkshp_tucker.pdf [Accessed 25/01/2018].
- USDA (2009) The Consumer Data and Information Program: Sowing the Seeds of Research. Washington DC United States Department of Agriculture.
- Volkova, E., N. Li, E. Dunford, H. Eyles, M. Crino, J. Michie, and C.N. Mhurchu (2016) "Smart" Rcts: Development of a Smartphone App for Fully Automated Nutrition-Labeling Intervention Trials. *JMIR mHealth and uHealth*, 4(1):e23.
- Weerts, S.E., and A. Amoran (2011) Pass the Fruits and Vegetables! A Community-University-Industry Partnership Promotes Weight Loss in African American Women. *Health Promotion Practice*, 12(2):252-60.
- Weinstein, J.L., V. Phillips, E. MacLeod, M. Arsenault, and A.M. Ferris (2006) A Universal Product Code Scanner Is a Feasible Method of Measuring Household Food Inventory and Food Use Patterns in Low-Income Families. *Journal of the American Dietetic Association*, 106(3):443-45.
- Wenz, A., M.P. Couper, and A. Jäckle (2019) Willingness to Use Mobile Technologies for Data Collection in a Probability Household Panel. *Survey Research Methods*, 13(1):1-22.
- Westat (2011) Redesign Options for the Consumer Expenditure Survey. Rockville, Maryland: Westat. Available at: http://www.bls.gov/cex/redwrkshp_pap_westatrecommend.pdf. [Accessed: 25/01/2018].
- Worthington, S., and J. Fear (2009) The Hidden Side of Loyalty Card Programs. Retail Therapy [online] Clayton, Victoria: The Australian Centre for Retail Studies. Available at: <http://tapchibanle.org/retail-lib/hidden-side-of-loyalty.pdf>. [Accessed: 25/01/2018].
- Zhen, C., J.L. Taylor, M.K. Muth, and E. Leibtag (2009) Understanding Differences in Self-Reported Expenditures between Household Scanner Data and Diary Survey Data: A Comparison of Homescan and Consumer Expenditure Survey. *Review of Agricultural Economics*, 31(3):470-92.