METHODS ARTICLE

JOURNAL OF
INDUSTRIAL ECOLOGY WILEY

Estimating carbon footprints from large scale financial transaction data

Anna Trendl[1,2] | Anne Owen[3] | Lara Vomfell[1,2] | Lena Kilian[4,5] |
John Gathergood[6] | Neil Stewart[2] | David Leake[1]

[1]Behavioural Science, Lloyds Banking Group, London, UK

[2]Warwick Business School, University of Warwick, Coventry, UK

[3]School of Earth and Environment, University of Leeds, Leeds, UK

[4]Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

[5]Centre for Spatial Analysis and Policy, School of Geography, University of Leeds, Leeds, UK

[6]School of Economics, University of Nottingham, Nottingham, UK

Correspondence
Anna Trendl, Behavioural Science, Lloyds Banking Group, 25 Gresham Street, London EC2V 7HN, UK.
Email: Anna.Trendl@Lloydsbanking.com

Editor Managing Review: Junming Zhu

Abstract

Financial transactions are increasingly used by consumer apps and financial service providers to estimate consumption-based carbon emissions. This approach promises a low-resource, ultra-fast, and highly scalable approach to measuring emissions at different levels of potential policy intervention—spanning the national, subnational, local, and individual level. Despite this potential, there is a lack of research exploring the validity of this approach to carbon profiling. Here we address this oversight in three ways. First, we provide a step-by-step description of our approach toward estimating carbon footprints from micro-level transaction data generated by more than 100,000 customers of a large retail bank in the United Kingdom. Second, we quantitatively compare emission estimates obtained from transaction data with those calculated from a more standard data source used in carbon profiling, the largest household expenditure survey in the United Kingdom. Third, we offer a detailed qualitative comparison of the advantages and disadvantages of transactions versus alternative data sources (such as survey data), across key dimensions including data availability, data quality, and data detail. We find that financial transactions offer a credible alternative to survey-based sources and, if made more widely accessible, could provide important advantages for profiling emissions. These include objective, micro-level data on consumption behaviors, larger sample sizes, and longitudinal, frequent data capture.

KEYWORDS
big data, carbon footprint, consumption, greenhouse gas emissions, household expenditure, industrial ecology

1 | INTRODUCTION

In recent decades, accounting for greenhouse gas (GHG) emissions from the perspective of consumption has moved up the economic and political agenda. Working Group III's contribution to the Intergovernmental Panel on Climate Change's fifth assessment report has a chapter exploring consumption growth as a driver for GHG emissions (Blanco et al., 2015). In the United Kingdom, the Climate Change Committee (CCC) specifically

recommended that UK consumption emissions be "monitored….closely" (Climate Change Committee, 2019, p. 162). Measuring consumption-based emissions at different levels of potential policy intervention—national, subnational, local, and even at the individual level—is a critical requirement for the successful implementation of climate policies to deliver on Net Zero.

Despite this pressing need, clear consensus is yet to emerge on appropriate data sources and calculation methodologies for disaggregated, consumption-based emissions. We see three broad approaches currently in circulation. The first involves combining household spend data from national surveys with carbon emission factors derived from environmentally extended multiregional input–output (MRIO) databases (Tukker et al., 2018; Wiedmann et al., 2011). Household expenditure surveys, such as the Living Cost and Food Survey (LCFS) in the United Kingdom (Office for National Statistics, 2021), provide an annual view of consumption at the household level, and have been used to study how emissions vary by key demographic characteristics (Owen & Barrett, 2020). Whilst household surveys represent an important source of insights for climate policy, there are significant limitations, such as the low frequency of data collection and relatively modest sample sizes.

A second approach, also involving data collection via surveys, focuses instead on specific activities that generate known physical quantities of emissions (Carbon Footprint Calculator, 2022; Streamlined Energy and Carbon Reporting in the UK, UK Government, 2019; WWF UK, 2022). This approach can be highly accurate as it relies on measuring actual activity, rather than proxying activities with more general sources of information, such as data on household spending. It also provides emission measurements at the individual level. However, completing detailed surveys is time consuming and requires participants to track down a broad range of information that may not be readily available. These frictions and the associated costs to incentivize can result in low participation rates. Biased samples may also be a concern, due to participants self-selecting to take part (West et al., 2016).

A third approach, which we see becoming increasingly popular, involves estimating emissions based on peoples' spending activity as recorded via financial transactions captured by payments technology. As we describe in detail in this paper, by combining information from financial transactions with MRIO-based carbon emission factors, detailed carbon emission profiles can be constructed. A range of consumer apps are emerging using this core approach (see Svalna, 2022; CoGo, 2022; Earthchain, 2022; ecolytiq, 2022), promising low-resource, ultra-fast, and highly scalable emission measurements. Despite the potential opportunity and growing prevalence of the approach, to our knowledge, there are as-of-yet no formal research studies exploring the validity of transaction-based emission estimates and its advantages and disadvantages over alternative approaches.

Here, we begin to address this gap in the research literature in three ways. First, we provide a detailed overview of the steps involved in calculating carbon footprints from micro-level transaction data generated by more than 100,000 customers of a large retail bank in the United Kingdom. Second, we quantitatively compare carbon emission estimates produced by financial transaction data to those calculated from a household spend survey, demonstrating that transaction-based emission estimates correspond closely to those obtained from the largest household spend survey in the United Kingdom (LCFS). We provide these comparisons across several demographic variables and consumption categories.

Third, we offer a detailed qualitative comparison of the advantages and disadvantages of using transactions versus alternative sources, across dimensions including data availability, data quality, and data detail. This review finds that transaction data, if made more widely accessible, can offer important advantages over other methods, including objective, micro-level data on consumption behaviors; larger sample sizes; and longitudinal, more frequent data capture. In contrast, surveys often lack these characteristics and self-reported data may be biased by individual attention and preferences.

Based on these analyses, we conclude that transaction data offers a valuable additional data source for climate policy makers. Transaction data also offers individuals, businesses, and other organizations easier access to their own carbon footprints, without the frictions imposed by complex survey questionnaires or expensive engagements with third-party providers. This can encourage more informed consumer choices and support an evidence-based transition to Net Zero.

The paper proceeds as follows. Section 2 describes our data sources and methodology for measuring emissions from transactions. Section 3 quantitatively compares our results to emission estimates obtained from the LCFS. Section 4 provides a qualitative evaluation of the benefits and challenges involved in using transactions compared to other data sources, while Section 5 discusses options available to improve data access.

2 | DATA AND METHODS

To estimate carbon footprints from financial transactions, we collect 2018 spend data from an anonymous sample of 101,744 customers from one of the United Kingdom's largest retail banks (Lloyds Banking Group, 2022; LBG). A transaction is defined as any spend that occurs on a personal current or credit card account, including electronic transfers (automated or not), online transactions, in-store transactions, and ATM withdrawals. Full details on our sample selection procedure are provided in Section 2.2.1.

Our objective is to compare emissions calculated from these transactions to those from the LCFS, a long-running survey of the weekly expenditures of UK households (Office for National Statistics, 2021), that has been previously used for measuring UK emissions (Owen & Barrett, 2020).

2.1 | LCFS data

Like other household spend surveys around the world, the LCFS is used to formulate retail price indices, analyze the effects of taxation and benefits, and to address other policy-relevant topics such as trends in nutrition. As described elsewhere in the literature (Bulman et al., 2017), approximately 5500 households take part in the survey each year, with survey results weighted to be representative of the UK population.

While the LCFS collects household-level spend data, it is possible to use these data to calculate individual-level estimates using a process known as equivalization (Gough et al., 2011). We apply a modified OECD equivalence scale, as preferred by UK government, to convert each survey entry into the equivalent spend if the household was a single adult[1] from the 2018 LCFS ($N = 5473$). We also obtain data on each household's income group (1–10), UK NUTS 2 region, and the age of the household reference person.

2.2 | LBG data

2.2.1 | LBG sample selection

While the LCFS obtains representative samples, LBG's customer base is not necessarily representative of the entire UK population, though LBG is one of the UK's largest retail banks. In particular, it is estimated that 22% of UK adults hold current accounts across multiple financial service providers (Moon et al., 2015). Therefore, to ensure a fair comparison to the LCFS profiles which capture complete household expenditures by design, we set three criteria for LBG sample inclusion with the aim to restrict the sample to individuals who undertake their main banking activity with LBG.

First, we identify customers who processed at least 12 customer-initiated transactions (e.g., non-automated payments) in every month of 2018 (across current accounts and credit cards, if applicable). This first criterion filters out customers who may not use LBG as their primary provider or have recently switched providers. This filtering is standard practice within the bank.

Second, we restrict our sample to those who have at least one payment classified as "Energy" by LBG's internal Transaction Classification System (TCS)[2]. This is another indicator of a customer's primary bank account status and ensures our sample is directly comparable with the LCFS, which explicitly collects data on household energy spending. In addition, domestic energy spend is the largest component of a household's carbon footprint and needs to be accounted for to give an accurate measure of emissions (Minx et al., 2013).

This second criteria may exclude renters whose rent payments are bundled with household bills, and therefore for whom we are unable to estimate energy-based emissions. However, this type of bundled bill structure is relatively rare in the United Kingdom (Property Reporter, 2022), and overall energy spend is unlikely to differ systematically by tenancy agreement (bills included/rent only) or tenure (renters/homeowners).

Third, we exclude individuals who have transfers to non-LBG current accounts in their name, or payments toward non-LBG credit cards exceeding 2.5% of their overall spending. Again, these criteria identify non-primary account status. The 2.5% threshold is selected to avoid unnecessarily limiting our sample but to ensure we capture the majority of an individuals' expenditure.

From this final sample of qualifying individuals, we draw an anonymous random sample of 101,744 aged 18 or over and identify their transactions associated with all open credit card and personal current accounts in 2018. Overall, 50% of individuals in this sample are found to hold a joint account ($N = 50,744$). The expenditure of individuals with joint accounts is likely to reflect household-level spending, and we take this distinction into account in subsequent analyses (see Section 3).

2.2.2 | LBG spend data

Of the 62 million transactions retrieved, the majority (86%) are categorized by the LBG TCS system or classified as ATM withdrawals. Of the 14% of transactions that are not classified, 51% are payments related to interest, charges, and fees, while 25% are inter-account transfers (transfers between two LBG accounts).

We exclude transfers and payments that are not classified by TCS (except domestic ATM withdrawals), and the following TCS categories: unclassified transactions, overseas cash withdrawal, and payments toward other current accounts, credit cards, and savings accounts. The bank's TCS is structured hierarchically across three levels, with increasing granularity. For example, the highest, level 1, category of "Motor" spend has level 2 sub-categories corresponding to "Petrol," "Maintenance/Repair," and "Parking/Tolls."

Finally, we aggregated spend across all three levels of remaining TCS categories for these 101,744 individuals. The final list of TCS categories can be found in Supporting Information S5 (Table SI1).

2.3 | Carbon multipliers by COICOP category

To calculate carbon emission estimates from transaction data, we use "carbon multipliers." These represent the emissions associated with £1 unit of spending in a given TCS category. We now describe their derivation in detail.

Carbon multipliers are constructed from two data sources. The first includes consumption-based GHG emission estimates for the UK economy, disaggregated into sub-categories (Department for Environment, Food & Rural Affairs, 2022). These originate from a MRIO framework. Here the UKMRIO is used, which reports carbon dioxide equivalents of GHGs.

Environmentally extended MRIO analyses enable indirect consumption-based emissions of UK households to be calculated at a product level (Miller & Blair, 2009; Owen et al., 2018; Wood et al., 2019). Where relevant, direct household emissions from fuel burning are also included to construct more complete, total consumption-based emissions.

The second data source contains total household-level spending on different goods, provided by the LCFS. Crucially, both these data sources share a common categorization scheme (referred to as COICOP, the Classification of Individual Consumption by Purpose, United Nations, 2014). Therefore, by dividing emission estimates for a given COICOP category by its corresponding total household spending (kg CO_2e/£), we can derive emissions associated with a £1 unit of spending.

These carbon multipliers by design account for both direct emissions associated with burning fuel to heat homes and driving cars, and indirect emissions, which consider the full domestic and international production supply chains of the goods and services consumed by UK households. A more detailed description of these two data sources can be found in the Supporting Information (SI).

Disaggregating national consumption-based accounts using spend data, as done here, can introduce uncertainty where the same products vary in price. To reduce this uncertainty, some researchers rely on physical data to estimate emissions (Girod & de Haan, 2010; Goldstein et al., 2017; Hendrie et al., 2014; Jones & Kammen, 2014). However, detailed data on functional unit consumption are unavailable in many countries, including the United Kingdom. Therefore, even research that employs some physical measurements of consumption often relies on expenditure data (Vita et al., 2019).

Whilst a spend-based approach may underestimate low footprints and overestimate high footprints, overall trends remain measurable. Therefore, despite some uncertainty in the estimates as a result of using expenditure-based multipliers, this type of analysis can still differentiate between high and low emitters. Consequently, emission estimates generated using expenditure-based carbon multipliers allow for useful assessments of emissions, including where they can be reduced and redistributed.

Overall, we generate carbon multipliers for 307 COICOP categories, each of which can be further categorized into 12 broad COICOP spend categories (e.g., Transport, Food, Housing, etc.; for the full list see Supporting Information S6, Table SI2).

2.4 | Calculation of carbon multipliers by TCS category (TCS-COICOP mapping)

Next, as spend data is on the TCS level, we need to generate carbon multipliers for the 252 TCS categories using the carbon multipliers established for the 307 COICOP categories.

For each TCS category, we review available COICOP categories to identify the corresponding COICOP categories. In cases where there is a simple one-to-one mapping, we simply carry over the available carbon multiplier directly. For example, the TCS category "Internet" maps directly to the COICOP category "Internet subscription fees," which carries a carbon multiplier of 0.17.

For those TCS categories where multiple corresponding COICOP categories exist, we construct a household spend-weighted average of the corresponding multipliers. For example, for TCS category "Gas & Electricity," COICOP categories "Gas" and "Electricity" are available with multipliers 6.966 and 2.249, respectively. According to the LCFS, 54.3% of the combined spend in these categories is spent on electricity, we therefore construct a spend-weighted multiplier average of 4.405 for the TCS category of "Gas & Electricity."

In addition to calculating the relevant carbon multipliers, we determine the most relevant broad COICOP category for each TCS category (see Supporting Information S5 Table SI1). Supporting Information S7 (Table SI3) lists the five TCS categories with the highest carbon multipliers. Supporting Information S8 (Table SI4) contains the entire TCS-COICOP mapping table.

Out of 252 TCS categories, 55 were out of scope of the COICOP classification system as they are not included in the definition of household expenditure in the United Kingdom (Office for National Statistics, 2020) and are also not part of the carbon profiles estimated from the LCFS survey. These non-consumption expenditure categories include financial investments, loan, mortgage, council tax, charity, tax, and duty free spend (see Supporting Information S8, Table SI4), representing 26% of spend in the LBG sample.

After excluding these categories, ATM withdrawal constitutes 14% of overall spend in our sample. Drawing on information from the UK Finance's 2020 report on money spending habits in the United Kingdom (UK Finance, 2020), we use the mapping approach described above to determine the carbon multipliers associated with the top 20 cash spend categories (see Supporting Information S9, Table SI5) and corresponding cash spend proportions. More information on the steps involved in this calculation can be found in the SI.

2.5 | Calculation of carbon footprints for the LBG sample and the LCFS sample

We can now calculate TCS-level emissions (kg CO_2e) for each individual in the sample by multiplying TCS-level spend data (£) with TCS-level carbon multipliers (kg CO_2e/£). Using COICOP-level data, we repeat the parallel exercise for the LCFS sample ($N = 5473$).
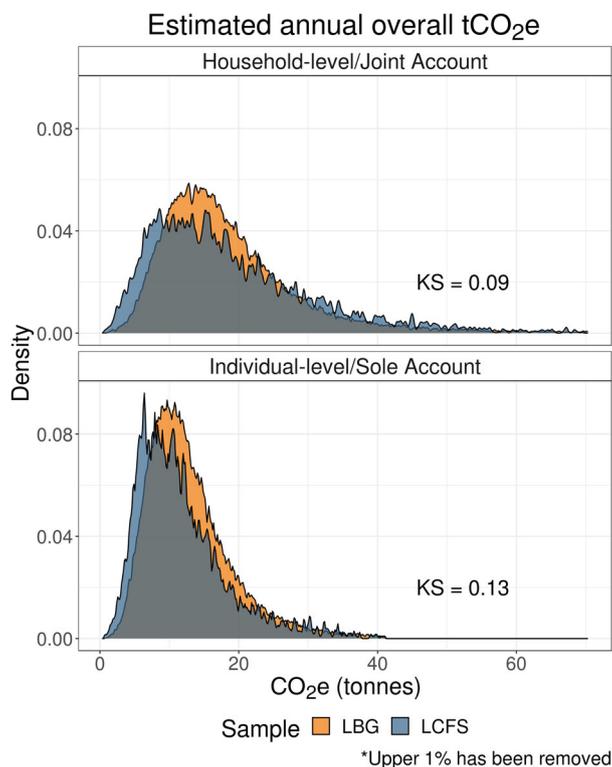
FIGURE 1 Distributions of overall carbon footprint estimates. Low-resolution data underlying this figure are available in Supporting Information S13 (Table SI9).

For the LBG transaction data, we sum TCS and ATM-related emissions by the corresponding broad COICOP category to calculate domain-specific emissions by individual. This allows us to compare LBG and LCFS emission estimates at the broad COICOP category level. Finally, by summing these for each individual, we derive total annual emission figures.

Supporting Information S4 (Figure SI1) summarizes the steps described throughout this section.

3 | QUANTITATIVE COMPARISON

The focus of our analysis is a quantitative comparison of carbon emission distributions derived from transaction data and LCFS across broad COICOP categories (e.g., housing, transport, food), and key demographic variables (age, income, region). To ensure a like-for-like comparison, we compare "sole account" holders in the LBG sample ($N = 51,000$) to "individuals" in LCFS (equivalized estimates); and separately, "joint account" holders in the LBG sample ($N = 50,744$) to "households" in LCFS. Supporting Information S10 (Table SI6) provides a summary of these distinctions.

To quantify the degree of overlap between corresponding distributions, we report the Kolmogorov–Smirnov (*KS*) statistics (for more detail see the SI). The *KS* statistic is bounded between 0 and 1, where the two extremes indicate no overlap and a perfect overlap, respectively. As a useful reference point, we note that the value of the KS statistic when comparing the emission distributions from the top and the bottom income deciles in the LCFS sample is 0.60.

3.1 | Emission groups: Overall, housing, transport, food

3.1.1 | Overall emissions distributions

Figure 1 shows the distributions of overall annual footprint estimates by sample (LBG vs. LCFS) and estimate type (Household/Joint vs. Individual/Sole account). Each figure reports the *KS* statistics.

We see a high degree of overlap between these distributions, as evidenced by the low KS statistics. Average household emissions from the LCFS are only slightly higher than the average household emissions from the LBG sample ($M_{LCFS} = 19.78$, 95% CI [19.37, 20.15];
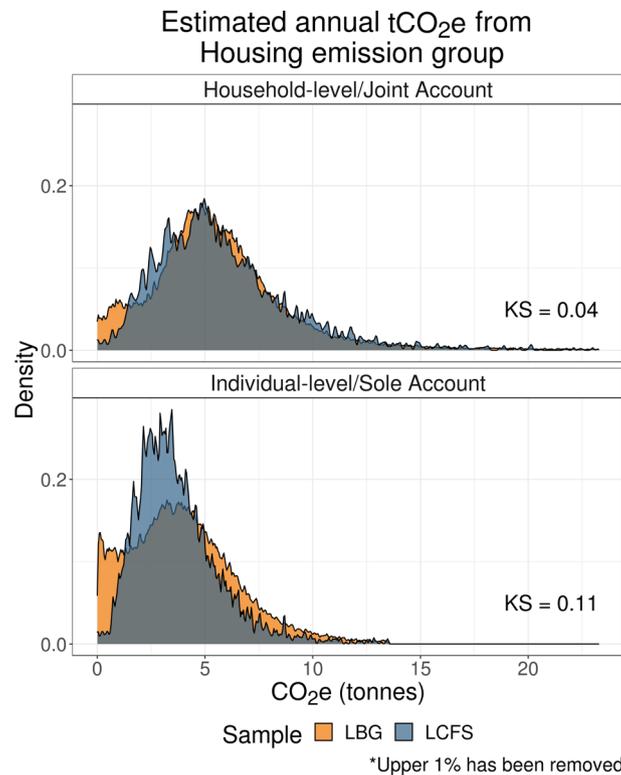
FIGURE 2 Distributions of housing-related carbon footprint estimates. Low-resolution data underlying this figure are available in Supporting Information S13 (Table SI9).

$M_{LBG} = 19.19$, 95% CI [19.06, 19.31]). Conversely, for individual-level carbon emissions, we find that the average LCFS emissions estimates are lower than the average LBG estimates ($M_{LCFS} = 12.28$, 95% CI [12.05, 12.46]; $M_{LBG} = 13.29$, 95% CI [13.20, 13.36]).

These results are consistent with a scenario where (1) LBG transaction data does not capture the entirety of household and individual-level expenditure, and (2) LBG sole account spending also captures some proportion of household spending. The effect of the former is minimized from the sample selection (see Section 2.2.1), and therefore the latter is likely to impact our estimates more. Notably, a significant proportion of households do not hold a joint account (Money Saving Expert, 2019).

Next, we examine the emission distributions for the top three broad COICOP categories. Descriptive statistics for all emissions groups are reported in Supporting Information S11 (Table SI7), while emission shares by emission category, sample, and estimate types are reported in Supporting Information S12 (Table SI8).
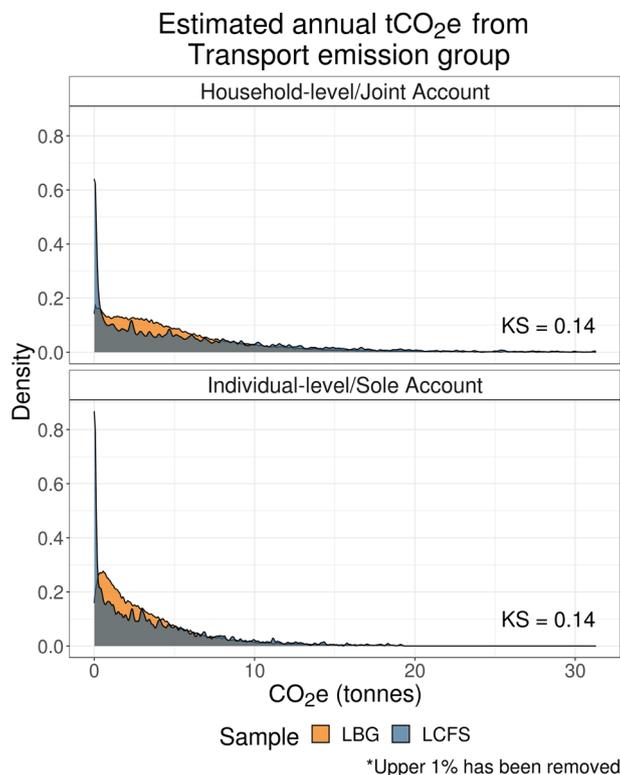
3.1.2 | Housing emissions distributions

Housing-related emissions primarily capture emissions arising from spend on gas, electricity, and water charges, constituting about 33–37% of overall carbon footprints across samples.

While Figure 2 shows a high level of correspondence between household-level and joint account housing-related estimates, we see larger differences between individual-level and sole account emission distributions.

First, we see that low emission estimates are more common in the LBG sample. This might stem from the fact that a relatively high proportion of the sole account sample shows evidence for "pay as you go" energy payments. These prepaid energy meters are often preferred by landlords, and it is possible that there are relatively more renters in the sole account sample. Compared to monthly direct debit, prepaid energy payments are more likely to be spread across multiple accounts within the household or paid by cash, which could explain why we see a higher proportion of individuals with relatively low energy payments in the LBG sample.

Second, we also note that a higher share of the LBG sole account sample has relatively high housing-related emission estimates (above 5 tonnes of CO_2e). As discussed in Section 3.1.1, this pattern is consistent with the assumption that some proportion of the sole account sample captures household-level spend.

Estimated annual tCO$_2$e from Transport emission group

FIGURE 3 Distributions of transport-related carbon footprint estimates. Low-resolution data underlying this figure are available in Supporting Information S13 (Table SI9).

3.1.3 | Transport emissions distribution

Transport-related estimates include emissions originating from spend on all modes of transport, including petrol, air and rail travel, public transport, and car finance. It constitutes about 24–26% of overall household and individual-level emissions across samples.

As Figure 3 and Supporting Information S11 (Table SI7) demonstrate, although the medians of transport-related emission estimates show a reasonable degree of correspondence, the average LCFS estimates are higher than LBG estimates.

This difference may result mechanically from how spend data is collected in the LCFS. Specifically, annual petrol spend estimates are extrapolated from the shorter, 2-week spend diary used by the LCFS (see SI). This may fail to capture more infrequent petrol spending occurring outside of this critical 2-week period and overestimate spend for those who did purchase petrol during this period.

This feature of the LCFS approach could also explain why LCFS transport-related emission distribution peak at zero, and have a long tail. The LBG distribution may, in fact, be a more truthful approximation of transport-related carbon emissions.

3.1.4 | Food emissions distributions

Food-related LBG estimates are calculated as emissions arising from general supermarket spend, while for the LCFS sample, they are calculated from all spend on specific foodstuff and non-alcoholic drinks consumed at home. These constitute about 15–18% of overall emissions across samples.

Using general supermarket spending to proxy specific food spending is a crude but necessary assumption, given the lack of basket-level information available in financial transactions. Supermarket spending figures include expenditure on both food and non-food items available. Similarly, carbon multipliers for supermarket-related TCS categories reflect a weighted average of relevant food and non-food COICOP categories. As a result, LBG food emission estimates will necessarily exceed LCFS estimates that are restricted to food-related COICOP categories.

As expected, average food-related emission estimates are significantly higher in the LBG sample than in the LCFS sample, as shown by Figure 4 and Supporting Information S11 (Table SI7). In Section 4, we will return to the challenge associated with a lack of basket-level information in financial transaction data.
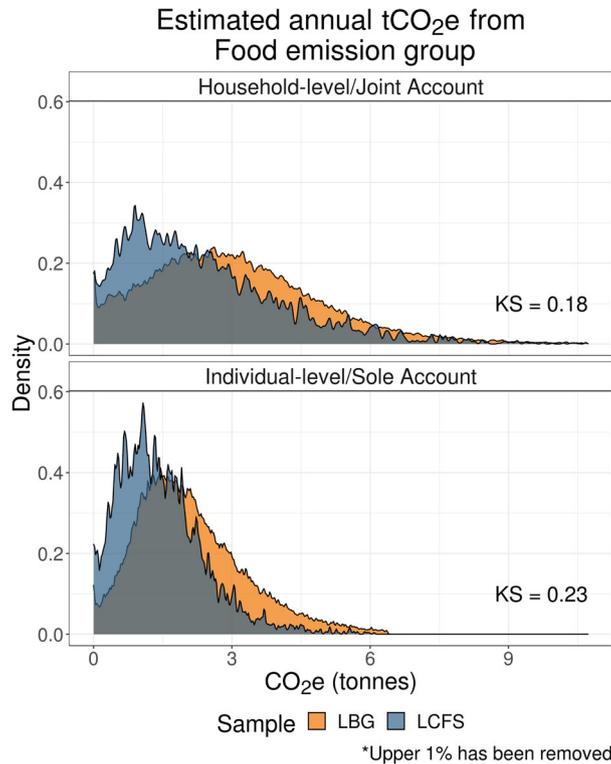
FIGURE 4 Distributions of food-related carbon footprint estimates. Low-resolution data underlying this figure are available in Supporting Information S13 (Table SI9).

3.2 | Demographics: Age, income, region

For climate policy insights, it is useful to break down emissions by demographic characteristics (Owen & Barrett, 2020). We now compare the overlap between LCFS and LBG overall carbon footprint distributions across different levels of age, income, and region to demonstrate that transaction-based emission estimates represent a valuable source of insights. We note that the LCFS emission distributions show more volatility due to the relatively small sample sizes in different demographic categories.

3.2.1 | Age

Figure 5 shows overall emission distributions across six different age groups. Both the LBG and LCFS distributions suggest an inverse U-shaped relationship between emissions and age groups, where emissions peak in the 40–50 and 50–60 age groups.

In line with the assumption that some proportion of the LBG sole account sample captures household-level spend, we see consistently higher LBG estimates in the age groups where individuals are most likely to live in a multi-person household (20–50 age groups).

Interestingly, we also see a slight divergence between household-level estimates in older age groups (60+), where LCFS estimates are generally lower, driven by differences in transport and restaurant-related emission groups. It is possible that the LCFS data collection methodology fails to adequately capture the relatively higher emission share of non-regular travel and hotel spend in this age group.

3.2.2 | Income

Figure 6 shows overall carbon footprint distributions across income groups, where groups 1 and 10 correspond to the lowest and highest income groups, respectively. For both household- and individual-level estimates, we see a strong positive relationship between income and emissions.

However, for household-level estimates, we observe some divergence: in the highest income group, LCFS estimates exceed LBG, whereas in the lower income groups, LBG slightly exceed LCFS. Again, this is partly driven by corresponding differences in transport-related emissions between the two samples.
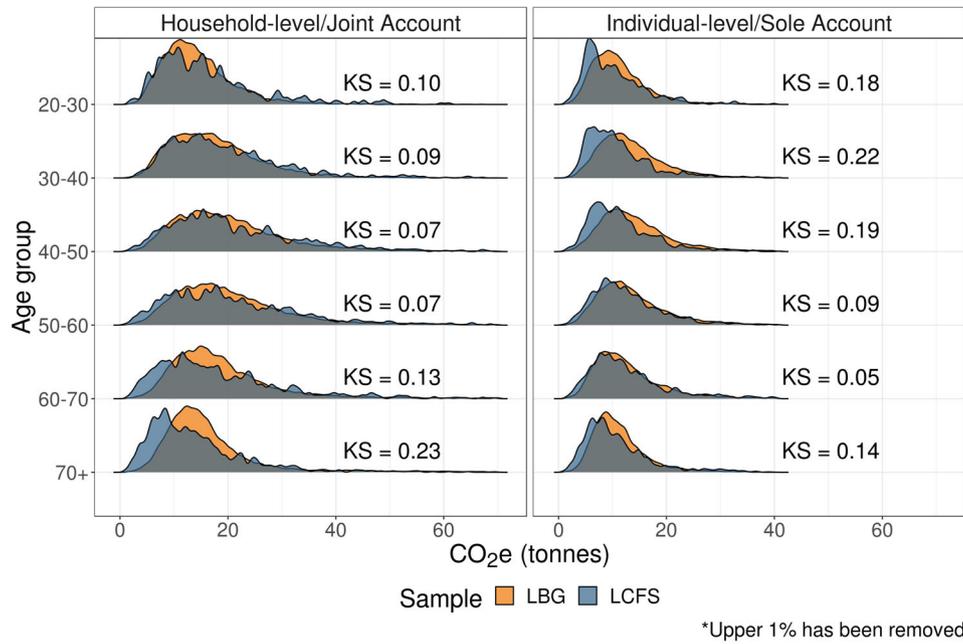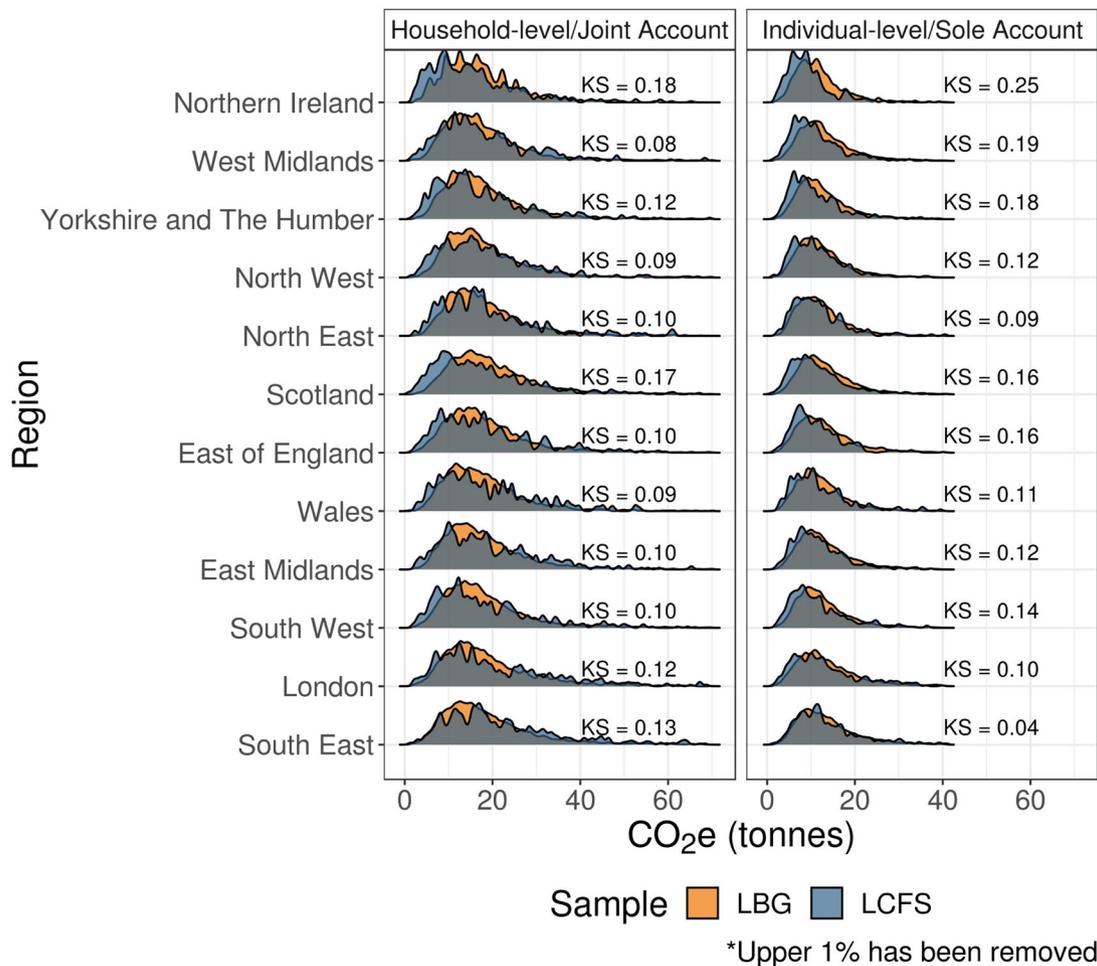
FIGURE 5 Carbon footprint distributions by age group. Low-resolution data underlying this figure are available in Supporting Information S13 (Table SI9).
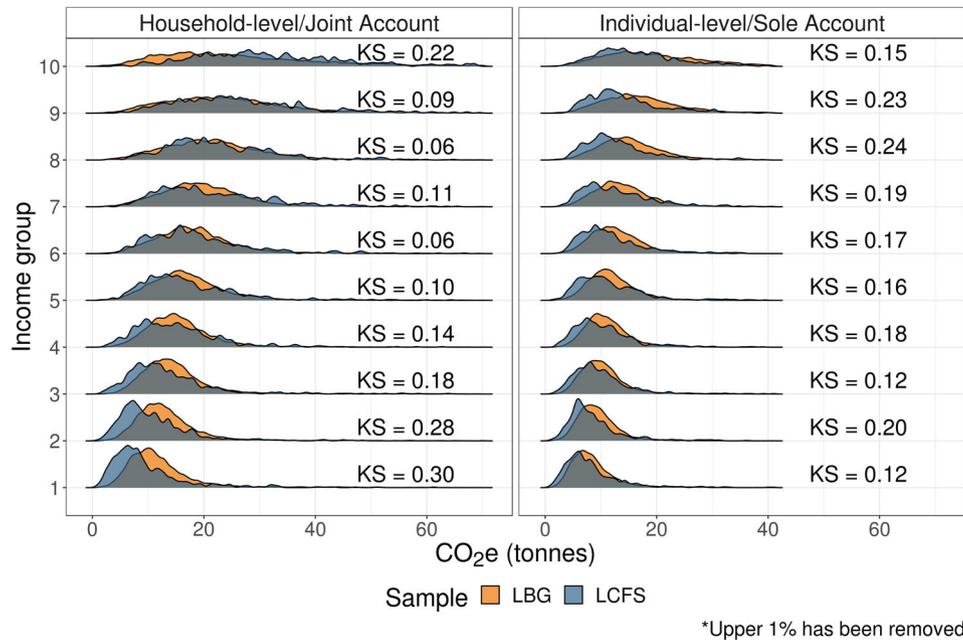
FIGURE 6 Carbon footprint distributions by income group. Low-resolution data underlying this figure are available in Supporting Information S13 (Table SI9).

3.2.3 | Region

Finally, Figure 7 shows overall emission distributions across regions in the United Kingdom. Again, we see generally better alignment for households. We observe the largest differences for regions with the highest (London, South East) and lowest (Northern Ireland, Scotland) average incomes.

FIGURE 7 Carbon footprint distributions by region. Low-resolution data underlying this figure are available in Supporting Information S13 (Table SI9).

3.3 | Summary of quantitative sample comparisons

Comparing the carbon footprint distributions derived from LBG transaction data with those from the LCFS household survey, we find broad agreement between the two data sources across key emission areas and demographic characteristics.

Where the distributions differ, we identify methodological differences in how spend is captured. These include prepaid utilities potentially paid by cash (see housing-related emissions), a 2-week spend diary not capturing all fuel spend (transport), and lack of basket-level information in transaction data (food). We also observe some differences where "sole accounts" potentially capture spend for multi-person households, rather than just a single person.

A future study could explore the similarities and differences between carbon emission estimates from financial transactions and those from an activity-based carbon calculator capturing physical quantities.

4 | QUALITATIVE COMPARISON

We now explore a broader, qualitative comparison of the advantages and disadvantages of using transactions versus household surveys and carbon footprint calculators. As described in Table 1, this comparison considers a range of dimensions, including data availability, quality, and detail.

4.1 | Data availability

Based on their relative availability, only household spend surveys are currently publicly available of the three data sources. Physical activity surveys require specific data to be collected and then made available, and transaction data, like other data assets generated by commercial activity, are

TABLE 1 Overview of strengths and weaknesses of carbon footprint data sources

	Method		
Criteria	Household expenditure survey (e.g., LCFS)	Activity-based carbon footprint calculator	Financial transactional data (e.g., LBG)
Data availability			
Data Access	✓	✗	✗
Big Data	✗	✗	✓
Scalability	✗	✗	✓
Representative sample	✓	✗	✓
Longitudinal view	✗	✗	✓
Frequency of data points	✗	✗	✓
Demographic variables	✓	✗	✓
Data quality			
Completeness (capturing all spend/activities)	✓	✗	✗
Reliability (self-reported vs objective data)	✗	✗	✓
Data detail			
Granularity (itemized list of spend/activities)	✓	✓	✗
Physical volumes	✗	✓/✗	✗
Differentiation by product brand/product life cycle	✗	✗	✗

subject to data access constraints and data privacy protection. We recognize this represents a major barrier to benefiting from transaction data to support climate policy and return to this topic in Section 5.

4.1.1 | Big data, scalability, representative sample

Assuming data access is available, transactional data-based carbon measurements could in principle be constructed for the majority of the adult population in developed economies. This is a significant advantage over surveys and physical carbon footprint calculators, which rely on active participation.

While both transaction and survey data allow for a representative sample (the former by data volumes, the latter by survey design), those using physical carbon footprint calculators are almost certainly self-selecting to take part which may bias results (e.g., they are likely to be individuals who are more conscious of their carbon footprint; West et al., 2016).

Also, since transactional data captures the behavior of a significant proportion of the population, it is ideally suited to address a key limitation of survey-based methods by identifying and modeling the behavior of the upper 1% of the income distribution.

4.1.2 | Longitudinal view, frequency of data points

Household surveys and physical carbon calculators only provide a point-in-time snapshot of carbon emissions. In contrast, transactional data allows for measurement over time, making it ideally suited for analyzing the effects of policy changes. Furthermore, the high frequency of transaction

data supports much more regular carbon profiling, unlike infrequent survey data collection that may result in biased spending estimates (see Section 3.1.3).

4.1.3 | Demographic variables

Carbon footprint data combined with demographic variables (e.g., age, income, and region) enables analyses of subgroup effects. Household surveys and transaction data already include this information; however, carbon calculators are not a feasible data source due to sample selection effects.

4.2 | Data quality

4.2.1 | Completeness

Surveys can provide a detailed, comprehensive picture of a household's spending profile. In contrast, 22% of UK adults hold accounts across several financial service providers (Moon et al., 2015). Therefore, using transaction data from just one provider may underestimate true expenditure.

Open Banking technology could address this completeness problem, providing a complete view of spending activity across multiple providers (Open Banking Implementation Entity, 2022). But these solutions still require explicit customer opt-ins and active participation which reduces sample sizes.

Physical carbon calculators are also limited by the number of survey questions being posed and it is recognized that there is a trade-off between complexity and usability (West et al., 2016).

4.2.2 | Reliability

Household surveys and carbon footprint calculators rely on self-reported data, whereas transactions represent an objective view of an individual's spending behavior, capturing actual outgoing flows.

4.3 | Data detail

4.3.1 | Granularity

A key limitation of using financial transactions when compared to survey methods is the lack of basket-level information. For example, as discussed in Section 3.1.4, whilst transactions provide the total amount of spending in a supermarket, the types and quantities of food items such as beef or vegetables are unknown.

A related point concerns cash spending, where financial transactions cannot identify categories of spending (see Section 3.1.2). We discuss the granularity issue further in Section 4.4.

4.3.2 | Physical volumes

Physical carbon calculators capture volumes directly for selected emission categories. Whereas carbon multipliers used by spend data-based methods are constructed from a monetary input–output system and rely on spending as a proxy for volume. This is a limitation as some spend differences will inevitably reflect price differences rather than volume differences.

This so-called "Proportionality Assumption Uncertainty" (Lenzen, 2000, p. 138) remains an issue across all footprint calculation methods which rely on CO_2e/spend-based multipliers, including transaction and survey data-based estimates.

4.3.3 | Differentiation

Neither the COICOP categories used for LCFS and LBG calculations, nor physical carbon footprint calculators are sufficiently granular to recognize more subtle differences within the same product category. These differences may include brands or slight variations of products.

These types of errors can be characterized as aggregation and allocation uncertainties (Lenzen, 2000). Efforts to accurately capture emissions associated with the entire life cycle of a particular product (Meinrenken et al., 2020) will prove useful in addressing this problem.

4.4 | Summary of qualitative comparison

In summary, whilst these data sources have different strengths and weaknesses, the transaction data approach offers the most advantages (see Table 1). A comprehensive solution could build on transactional data whilst addressing some of its methodological shortcomings, in particular, concerns around data access and new innovations to improve data completeness and data granularity.

We return to a discussion on data access in Section 5. On challenges related to data completeness and lack of basket-level information, one consumer app (Evocco, 2022) addresses these issues directly by providing a service to scan food shopping receipts. Whilst this is highly innovative, the approach still requires active participation and is therefore subject to similar criticisms as surveys, for example, low sample sizes.

Alternative innovations are needed to address these key methodological issues. A promising direction is to develop hybrid approaches for calculating carbon multipliers. These could incorporate additional information from population segments or product supply chains to improve data granularity, and statistical techniques to infer missing data and help improve data quality.

5 | CONCLUSION

In this study, we provide a detailed overview of our approach toward calculating carbon footprints from transaction data, and quantitatively compare our emission estimates to those obtained from the LCFS household spend survey, finding broad correspondence between emissions estimated from the two data sources. Additionally, we provide a detailed qualitative comparison of the advantages and disadvantages of using transactions versus alternative data sources. Based on these analyses, we conclude that transaction data offers a valuable additional data source for climate policy makers, providing key advantages over survey-based data sources including detailed and objective measurements, larger sample sizes, and longitudinal, frequent data capture.

Transaction data also promises to provide individuals, businesses, and other organizations easier access to their own detailed carbon footprint profiles, without the frictions imposed by complex survey questionnaires or expensive engagements with third-party providers.

However, the restricted data availability of financial transactions remains a key challenge to realizing these benefits. Transactions are stored by banks and other commercial organizations, and are therefore confidential and protected. New data sharing solutions and cultural norms are required to bring these resources safely and securely to policy makers and other concerned parties.

We see two broad routes forward. The first is via personal data donation. With the introduction of new General Data Protection Regulation (European Union, 2016), the general public can now access data collected about them by organizations. With the additional ease of access introduced with standards of Open Banking technology (Open Banking Implementation Entity, 2022), financial transactions can—in principle—be donated by the public at a scale that is useful to carbon policy.

Recent research finds over half of those individuals surveyed would be willing to donate their data to benefit society (Skatova & Goulding, 2019). However, in reality, only 30% of users may choose to share their anonymized carbon footprint data for research purposes (Andersson, 2020). In practice, new, dedicated services would also be required to manage data sharing and the underlying complex logistics—and these must be funded, incentivized, and developed at scale.

The second route focuses instead on commercial data donation. With a number of banks and credit card providers starting to introduce new carbon tracking features into digital banking (Gausden, 2021) and dedicated start-ups developing new carbon apps and calculators, these organizations should recognize a social responsibility to make carbon insights more widely available.

This cultural shift toward sharing carbon data need not jeopardize consumer data privacy. With appropriate data anonymization and aggregation practices, standards can be safely maintained. The formatting for disclosures could follow existing standards (UK Data Service, 2021) and regular reporting frequencies could be established, paralleling quarterly business performance disclosures.

As carbon profiling from transactions inevitably becomes much more commonplace across the financial services industry, a commensurate culture of data sharing and transparency is also required to ensure benefits are fully realized by policymakers, customers, and other key stakeholders.

CONFLICT OF INTEREST

The authors declare no conflict of interest. The views and opinions expressed are those of the authors and do not necessarily reflect the views of LBG, its affiliates or its employees.

ETHICS STATEMENT

The LBG Privacy Risk and Impact Assessment process recognizes the ethical basis for this study on aggregated, anonymous data as part of a strategy to help customers' transition to a more sustainable future. Upon opening an account, LBG customers consent to their data to be used for research: https://www.lloydsbank.com/help-guidance/customer-support/privacy-explained/data-privacy-notice.html

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from LBG but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are available from the authors upon reasonable request and with permission of LBG.

ORCID

Anna Trendl https://orcid.org/0000-0003-1994-4207

Anne Owen https://orcid.org/0000-0002-3872-9900

Lara Vomfell https://orcid.org/0000-0001-5264-9897

Lena Kilian https://orcid.org/0000-0002-0885-1198

John Gathergood https://orcid.org/0000-0003-0067-8324

Neil Stewart https://orcid.org/0000-0002-2202-018X

David Leake https://orcid.org/0000-0003-2361-9394

NOTE

[1] The OECD equivalence scale reference case (weight = 1) is a two-adult household with no children. Each household in the LCFS is assigned an equivalence factor. Households with a single adult score 0.67, each subsequent adult adds 0.33, and children under the age of 14 score 0.2. For example, a household of two adults and two children scores 1.4. Income and expenditure is first divided by the recorded equivalence factor so that each entry is equivalent to a household of two adults and no children. Finally, the entries are further multiplied by 0.67 to convert to a solo adult.

[2] It is standard practice within industry for banks to maintain internal, proprietary Transaction Categorization Systems (TCS). This system underpins several crucial analytical processes and is continuously improved using a variety of private and public sources.

REFERENCES

Andersson, D. (2020). A novel approach to calculate individuals' carbon footprints using financial transaction data – App development and design. *Journal of Cleaner Production*, *256*, 120396. https://doi.org/10.1016/j.jclepro.2020.120396

Blanco, G., Gerlagh, R., Suh, S., Barrett, J., de Coninck, H. C., Diaz Morejon, C. F., Mathur, R., Nakicenovic, N., Ofosu, A. A., Pan, J., Pathak, H., Rice, J., Richels, R., Smith, S. J., Stern, D. I., Toth, F. L., & Zhou, P. (2015). Drivers, trends and mitigation. In O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlomer, C. von Stechow, T. Zwickel, & Minx (Eds.), *Climate change 2014 mitigation of climate change. Working group III contribution to the fifth assessment report of the intergovernmental panel on climate change.* (pp. 351–412). Cambridge University Press. https://doi.org/10.1017/CBO9781107415416

Bulman, J., Davies, R., & Carrel, O. (2017). Living costs and food survey - Technical Report for survey year: 2015–2016. *Office for National Statistics*, https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey/livingcostsfoodtechnicalreport2015.pdf

Carbon Footprint Calculator - carbon footprint calculator for individuals and households. (2022). https://www.carbonfootprint.com/calculator.aspx

Climate Change Committee. (2019). *Net zero the UK's contribution to stopping global warming.* https://www.theccc.org.uk/publication/net-zero-the-uks-contribution-to-stopping-global-warming/

Cogo | Calculate The Carbon Footprint of Your Transactions. (2022). https://www.cogo.co/about

Department for Environment, Food & Rural Affairs. (2022). *UK full dataset 1990–2019, including conversion factors by SIC code.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1085184/2019_Defra_results_UK.ods

Earthchain. (2022). https://earthchain.co.uk/

Ecolytiq. (2022). *Climate action infrastructure for sustainable finance.* https://ecolytiq.com/

European Union. (2016). General data protection regulation. *Official Journal of the European Union*, L 119, 4.5.2016, p. 1–88. https://gdpr-info.eu/eli/reg/2016/679/oj

Evocco. (2022). *Eat within planetary boundaries.* https://www.evocco.com/

Gausden, G. (2021). Reduce your carbon footprint: Six apps and tools that highlight the emissions you're responsible for - and how to cut down. *This is Money*, https://www.thisismoney.co.uk/money/bills/article-9126055/Six-tools-track-personal-carbon-footpring.html

Girod, B., & de Haan, P. (2010). More or better? A model for changes in household greenhouse gas emissions due to higher income. *Journal of Industrial Ecology*, *14*(1), 31–49. https://doi.org/10.1111/j.1530-9290.2009.00202.x

Goldstein, B. P., Hauschild, M. Z., Fernández, J. E., & Birkved, M. (2017). Contributions of local farming to urban sustainability in the Northeast United States. *Environmental Science and Technology*, *51*(13), 7340–7349. https://doi.org/10.1021/acs.est.7b01011

Gough, I., Abdallah, S., Johnson, V., Ryan-Collins, J., & Smith, C. (2011). *The distribution of total greenhouse gas emissions by households in the UK, and some implications for social policy.* CASEpapers (CASE/152). Centre for Analysis of Social Exclusion. https://sticerd.lse.ac.uk/case/_new/publications/abstract/?index=3863

Hendrie, G. A., Ridoutt, B. G., Wiedmann, T. O., & Noakes, M. (2014). Greenhouse gas emissions and the Australian diet—Comparing dietary recommendations with average intakes. *Nutrients*, 6(1), 289–303. https://doi.org/10.3390/NU6010289

Jones, C., & Kammen, D. M. (2014). Spatial distribution of U.S. household carbon footprints reveals suburbanization undermines greenhouse gas benefits of urban population density. *Environmental Science and Technology*, 48(2), 895–902. https://doi.org/10.1021/es4034364

Lenzen, M. (2000). Errors in conventional and input-output—based life—cycle inventories. *Journal of Industrial Ecology*, 4(4), 127–148. https://doi.org/10.1162/10881980052541981

Lloyds Banking Group. (2022). *Lloyds Banking Group Annual Report and Accounts 2022*. https://www.lloydsbankinggroup.com/assets/pdfs/investors/annual-report/2020/2020-lbg-annual-report.pdf

Meinrenken, C. J., Chen, D., Esparza, R. A., Iyer, V., Paridis, S. P., Prasad, A., & Whillas, E. (2020). Carbon emissions embodied in product value chains and the role of life cycle assessment in curbing them. *Scientific Reports*, 10, 6184. https://doi.org/10.1038/s41598-020-62030-x

Miller, R. E., & Blair, P. D. (2009). *Input-output analysis: Foundations and extensions* (2nd ed.). Cambridge University Press.

Minx, J., Baiocchi, G., Wiedmann, T., Barrett, J., Creutzig, F., Feng, K., Förster, M., Pichler, P.-P., Weisz, H., & Hubacek, K. (2013). Carbon footprints of cities and other human settlements in the UK. *Environmental Research Letters*, 8(3), 35039. https://doi.org/10.1088/1748-9326/8/3/035039

Do you have a joint bank account with your partner? (2019). Money saving expert. https://www.moneysavingexpert.com/poll/2019/do-you-have-a-joint-bank-account-with-your-partner-18-06/

Moon, N., McHugh, S., & Cass, N. (2015). *Personal current account investigation - A report for the competition and markets authority by GfK NOP*. https://assets.publishing.service.gov.uk/media/555cabd0ed915d7ae2000007/PCA_Banking_Report.pdf

Office for National Statistics. (2020). *Definitions and conventions for UK household final consumption expenditure*. https://www.ons.gov.uk/economy/nationalaccounts/satelliteaccounts/methodologies/definitionsandconventionsforukhouseholdfinalconsumptionexpenditure

Office for National Statistics, Department for Environment, Food and Rural Affairs. (2021). *Living* (4th ed.). UK Data Service. https://doi.org/10.5255/UKDA-SN-8686-4

Open Banking Implementation Entity. (2022). https://www.openbanking.org.uk/about-us/

Owen, A., & Barrett, J. (2020). Reducing inequality resulting from UK low-carbon policy. *Climate Policy*, 20(10), 1193–1208. https://doi.org/10.1080/14693062.2020.1773754

Owen, A., Scott, K., & Barrett, J. (2018). Identifying critical supply chains and final products: An input-output approach to exploring the energy-water-food nexus. *Applied Energy*, 210, 632–642. https://doi.org/10.1016/j.apenergy.2017.09.069

Skatova, A., & Goulding, J. (2019). Psychology of personal data donation. *PLoS ONE*, 14(11), e0224240. https://doi.org/10.1371/journal.pone.0224240

Svalna. Calculate your climate impact. (2022). https://www.svalna.se/web/en

Tukker, A., de Koning, A., Owen, A., Lutter, S., Bruckner, M., Giljum, S., Stadler, K., Wood, R., & Hoekstra, R. (2018). Towards robust, authoritative assessments of environmental impacts embodied in trade: Current state and recommendations. *Journal of Industrial Ecology*, 22(3), 585–598. https://doi.org/10.1111/jiec.12716

UK Data Service. (2021). Living Costs and Food Survey, 2018–2019, Documentation, Volume F. https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8686#!/documentation

UK Finance. (2020). *UK Consumer Payments 2020*. https://www.ukfinance.org.uk/policy-and-guidance/reports-publications/uk-consumer-payments-2020

UK Government. (2019). *Environmental reporting guidelines: Including streamlined energy and carbon reporting guidance*. www.nationalarchives.gov.uk/doc/open-government-licence/

United Nations. (2014). *COICOP: Detailed structure and explanatory notes*. http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5

Vita, G., Lundström, J. R., Hertwich, E. G., Quist, J., Ivanova, D., Stadler, K., & Wood, R. (2019). The environmental impact of green consumption and sufficiency lifestyles scenarios in Europe: Connecting local sustainability visions to global consequences. *Ecological Economics*, 164, 106322. https://doi.org/10.1016/j.ecolecon.2019.05.002

Warnings raised over the true cost of bills inclusive renting. (2022, May 3). *Property reporter*. https://www.propertyreporter.co.uk/finance/warnings-raised-over-the-true-cost-of-bills-inclusive-renting.html

West, S. E., Owen, A., Axelsson, K., & West, C. (2016). Evaluating the use of a carbon footprint calculator: Communicating impacts of consumption at household level and exploring mitigation options. *Journal of Industrial Ecology*, 20(3), 396–409. https://doi.org/10.1111/jiec.12372

Wiedmann, T., Wilting, H. C., Lenzen, M., Lutter, S., & Palm, V. (2011). Quo vadis MRIO? Methodological, data and institutional requirements for multi-region input–output analysis. *Ecological Economics*, 70(11), 1937–1945. https://doi.org/10.1016/j.ecolecon.2011.06.014

Wood, R., Neuhoff, K., Moran, D., Simas, M., Grubb, M., & Stadler, K. (2019). The structure, drivers and policy implications of the European carbon footprint. *Climate Policy*, 20:sup1, S39-S57. https://doi.org/10.1080/14693062.2019.1639489

WWF UK. (2022). *WWF footprint calculator*. https://footprint.wwf.org.uk/#/

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.