A new year, a new you? Within-individual variation in food purchases

Laurens Cherchye, Bram De Rock, Rachel Griffith, Martin O’Connell, Kate Smith, Frederic Vermeulen

Contents lists available at ScienceDirect
European Economic Review
journal homepage: www.elsevier.com/locate/euroecorev

A R T I C L E   I N F O
Article history:
Received 4 September 2019
Revised 5 May 2020
Accepted 13 May 2020
Available online 26 May 2020

JEL classification:
C14
D12
D90
H12

Keywords:
Diet quality
Food purchases
Two-selves model
Collective model
Revealed preferences

A B S T R A C T

We document that within-individual variation in food choices is substantial and has potentially important consequences for nutrition, and hence well-being. We develop an approach that allows us to study the determinants of this within-individual variation within an economic framework and allow for across-individual preference heterogeneity. We show that around one-fifth of within-individual fluctuations in diet quality is explained by standard economic variables (prices and budgets), along with advertising and weather. The residual fluctuations are important and are larger for lower income and younger people, and individuals who state they are impulsive. We propose a two-selves model of food purchase behavior to structurally interpret these empirical patterns. We use nonparametric revealed preference techniques to show that this model rationalizes our food purchase data.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license.
(http://creativecommons.org/licenses/by/4.0/)

1. Introduction

The consequences of poor dietary choices for well-being and welfare are potentially profound, and governments around the world are struggling with how to address rising rates of obesity and diet related non-communicable diseases. The relationship between inequalities in diet quality and associated health outcomes across individuals is well documented. We focus, however, on the empirical determinants of within-individual variation in dietary choices. It is informative for policy design to understand whether successfully encouraging people to more consistently behave as they themselves do

* Corresponding author.
E-mail addresses: laurens.cherchye@kuleuven.be (L. Cherchye), bram.de.rock@ulb.ac.be (B. De Rock), rgriffith@ifs.org.uk (R. Griffith), martin.o@ifs.org.uk (M. O’Connell), kate.o@ifs.org.uk (K. Smith), frederic.vermeulen@kuleuven.be (F. Vermeulen).

1 See Finkelstein et al. (2005), Currie (2009), Currie et al. (2010) and Finkelstein et al. (2013).
2 See Baum and Ruhm (2009), Cutler and Lleras-Muney (2010), Bhattacharya and Sood (2011) and Allcott et al. (2017).

https://doi.org/10.1016/j.euroecorev.2020.103478
0014-2921/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license.
(http://creativecommons.org/licenses/by/4.0/)
in their relatively healthy periods could be a significantly more fruitful strategy than policies that focus more abstractly on persuading people to adopt more healthy choice behavior.

In this paper we make three contributions to the existing literature. First, we exploit new longitudinal data to document substantial within-individual variation in diet quality using information on individuals’ entire shopping baskets. We show that this variation is large and important from a nutritional perspective, and that it is of a similar magnitude to the well documented variation we see across people. Second, we empirically quantify the drivers of within-individual fluctuations in diet quality. We show that standard economic variables (prices, budgets), along with advertising and weather, are important and explain about one-fifth of within-individual variation in choices between healthy and unhealthy foods. The remaining four-fifths of variation is not explained by these variables. We show that the extent of this residual variation is correlated with observable characteristics of individuals, such as age and income, and with stated attitudes that reflect impulsiveness. Third, we develop a framework that allows us to structurally interpret these empirical patterns. We model individuals’ food choices as a bargaining process between two selves – a “healthy” and “unhealthy” self, which allows us to introduce time-varying preferences in a parsimonious way. This approach relies on insights from the collective household literature, and allows us to adapt revealed preference methods to show that our data are consistent with a two-selves model of behavior.

We use longitudinal data on the entire shopping basket of a sample of British individuals (the Kantar Worldpanel) to document substantial within-individual variation in diet quality over time. Looking over several years, we show that within-individual variation in diet quality is substantial from a nutritional perspective. For example, the within-individual standard deviation of the share of calories from added sugar (across months) is 4 percentage points; in comparison the UK government’s sugar reduction strategy targets a 5% reduction in sugar consumption in its first year (Public Health England, 2017). We show that the variation in diet quality that we see in our data can be parsimoniously described by shifts in the share of foods that are contained in a healthy (e.g. fruits, vegetables, whole grains, etc.) versus an unhealthy (e.g. soda, crisps, confectionery, etc.) basket. There is a sharp increase in the healthiness of foods purchased at the beginning of January, followed by a steady decline in healthiness over the calendar year. These declines are substantial: the median decline in the share of calories from healthy foods from the first to the final quarter of the calendar year is 5 percentage points, which is the same as the difference between the average purchases of a normal weight individual and an obese individual.

We empirically quantify the impact of standard economic variables and other factors in driving the healthiness of individuals’ food choices. We find that, in the cross-section, the mean preference for healthy foods is higher for higher income individuals, people with a lower body mass index, and those that state they try to eat a healthy diet. Our long panel allows us to estimate the relationship between the share of calories from healthy foods and explanatory factors separately for each individual, avoiding the need to impose homogeneous preferences. Looking within-individual over time, we see that, on average, around 20% of the within-individual variation in the share of calories from healthy foods is explained by responses to changes in standard economic variables such as prices, budgets, along with advertising and weather variation.

Nonetheless, there remains substantial within-individual variation that is not explained by these variables, which is broadly characterized by a decline in diet quality over the calendar year. We show how these fluctuations vary with demographics, with a particular focus on the correlation with age and income, as policy interventions are often targeted specifically at young people and lower income households. We find that individuals with lower income experience greater variation in the share of calories from healthy foods. This is true even after we control for potentially greater variation in the prices and food budgets that lower income individuals face, variation in their responsiveness to these changes, as well as the influence of advertising and weather. We find a similar relationship with age; the average within-individual variation in the residual share of calories from healthy foods is higher for younger people than older people.

An extensive psychological literature shows that individual choice behavior varies with context and time, and that individuals sometimes use self-regulation and behavior modification in an attempt to mitigate these influences (see references and discussion in Rabin, 1998 and Dellavigna, 2009). For example, experimental evidence suggests that individuals may be willing to impose (sometimes costly) commitments on themselves. New Years’ resolutions to eat a more healthy diet are a common form of self-regulation and behavior modification (Dai et al., 2014; Dai et al., 2015). We use information on individuals’ stated preferences and attitudes to investigate whether greater fluctuations in the share of calories from healthy food reflect impulsive behavior. We find that fluctuations are larger for individuals who state that they are more impulsive (e.g. spend money without thinking). Our results therefore build on a literature that finds empirical evidence of considerable within-individual variation in choice behavior in other settings, as well as in grocery purchases using alternative identification strategies.

In this spirit, we propose a two-selves model of food purchasing behavior as the lens through which to interpret our empirical findings. Our model is inspired by the multiple selves literature: we assume that individuals’ food choices are the outcome of an intra-personal bargaining process between a healthy and an unhealthy self, and we check whether this can explain the temporal shifts between healthy and unhealthy food baskets. More specifically, we capture individual specific

---

1 We use the term “self” to refer to a stable and rational preference ordering. More specifically, the healthy self has preferences over a healthy food basket and the unhealthy self over an unhealthy food basket. See Section 4.1 for details.
2 See Kaplan and Menzio (2015) and Kaplan and Schulhofer-Wohl (2017).
3 See Read and Van Leeuwen (1998), Read et al. (1999), Trope and Fishbach (2000), Ariely and Wertenbroch (2002) and Gilbert et al. (2002).
4 See Ashraf et al. (2006), Dellavigna and Malmendier (2006), Oster and Morton (2005), Bucciol (2012) and Hinnosaar (2016).
5 See Shapiro (2005), Milkman et al. (2010) and Sadoff et al. (2015).
non-monotonic changes in diet quality and quality resets throughout the years by a parsimonious structural model that draws on insights from the literature on collective household models.8 The model maps the observed share of spending on the healthy and unhealthy baskets to the theoretical bargaining weight between the selves, which can be affected by both standard and non-standard economic variables.

A specific advantage of this approach is that it allows us to make use of revealed preference methods to check whether the theoretical implications of our two-selves model are empirically satisfied for our data. We build on the work of Cherchy et al. (2007, 2011a), who developed revealed preference methods to analyze collective household behavior, to evaluate the consistency of the two-selves model with the data for each individual separately, and thus avoid assuming that individuals are characterized by homogeneous preferences. Our nonparametric revealed preference results suggest that the two-selves model does a good job at explaining the individual-specific variation in the data, and a better job than the more standard single-self model, in which each individual is characterized by a single, stable utility function. Our model allows us to structurally interpret our empirical findings and, therefore, forms a useful basis for the empirical investigation of temporal variation in food choices.

The rest of the paper is structured as follows. In the next section we provide evidence that there is substantial variation in food purchasing behavior within individuals across time. In Section 3 we quantify the extent to which this variation can be explained by prices, budgets and other factors, and how the amount of residual variation differs across individuals. In Section 4 we describe a two-selves model of food purchasing and show that it can rationalize observed purchasing patterns. In a final section, we conclude and discuss some avenues for future work.

2. Food purchasing behavior

Poor dietary choices are thought to be a leading cause of rising obesity and diet related diseases, with profound consequences for well-being and welfare,9 and of particular concern for low income households (Currie, 2009). The socio-economic gradient in diet quality and associated health outcomes across individuals is well documented. We bring new data to bear on the extent and importance of within-individual variation in diet quality, which is less well understood.

2.1. Data and measurement

We use data from the Kantar Worldpanel over the period 2005–2011, which we describe in detail in Online Appendix A.1. These data contain information on a representative sample of over 25,000 households. The data are longitudinal, with households remaining in the sample for, on average, two years. In this paper we focus on a sample of 3645 individuals in Britain who live on their own. We do this to avoid the confounding issue of intra-household allocation mechanisms; we show in Online Appendix B that the basic patterns we describe also hold in the full sample for all household types.

Individuals record (using handheld scanners in the home) all grocery purchases made and brought into the home. The data are at the transaction (i.e. the barcode or UPC) level and include all foods and drinks, as well as household goods such as cleaning supplies and toiletries. We know the exact products purchased, the price paid (including discounts and special offers), and we have information on the nutritional characteristics of each product. This type of data is increasingly widely used in research.10 There may be periods during which individuals do not record their grocery purchases or, for example, are away on holiday, which would overstate the degree of variability in diet quality that we measure. To account for this, we assume individuals are not recording properly if they record no grocery purchases for more than two weeks, and exclude individual-months that contain any part of a period of non-recording. Griffith and O’Connell (2009) show that removing periods of non-reporting is important for matching spending patterns recording in other consumer survey measures, such as the Living Costs and Food Survey (LCFS). We show in Section 3.3.4 that the aggregate patterns of diet quality within the calendar year are present in the LCFS, which is a cross-sectional study of households, and includes food eaten outside of the home. From now on we use “foods” as shorthand for foods and non-alcoholic drinks.11

Diet quality is a complex multi-dimensional object; whether a diet is “healthy” depends on the consumption of a whole range of nutrients. Policy in the UK, US and many other developed countries has focused on foods that are high in fat, salt and sugar.12 Eating too much of these types of foods is linked to increased risk of non-communicable diseases, such as heart disease, diabetes, and cancer, as well as to obesity.13 In the UK, 88% of adults consume more than recommended levels of sugar, 45% more than recommended levels of saturated fat, and 35% more than recommended levels of salt.14 There are other nutrients, such as protein and fiber, which are good for health and are under-consumed. We describe variation in purchases of these key nutrients below.

---

8 See Chiappori (1988), Chiappori (1992), Browning and Chiappori (1998), Chiappori and Ekeland (2009), Dunbar et al. (2013) and Browning et al. (2013).
9 See Cutler et al. (2003), Bleich et al. (2008), Finkelstein and Zuckerman (2008) and World Health Organization (2015).
10 See, for example, Hausman and Leibtag (2007), Aguilar and Hurst (2007), Broda and Weinstein (2010), Dubois et al. (2014) and Kaplan et al. (2016).
11 We exclude alcoholic drinks from our main analysis; we have information on alcoholic drinks purchased and brought into the home. Our main conclusions are robust to their inclusion.
12 See Public Health England (2017) and US Department of Agriculture (2015).
13 See Ascherio et al. (1999) and World Health Organization (2015).
14 Calculated using the National Diet and Nutrition Survey. UK government recommendations specify: no more than 5% of calories from added sugar, no more than 30g of saturated fat per day for men and 20g for women (aged 19–64), and no more than 6g salt per day.
Table 1
Variation in diet quality.

|                          | (1)    | (2)    |
|--------------------------|--------|--------|
| Purchases aggregated to: | Week   | Month  |
| % calories from added sugar | 12.19  | 12.02  |
| Mean                     | 11.93  | 5.51   |
| Standard deviation       | 11.25  | 3.90   |
| within individuals       | 4.04   | 3.94   |
| between individuals      |        |        |
| % calories from saturated fat | 14.34  | 14.40  |
| Mean                     | 6.60   | 3.95   |
| Standard deviation       | 6.01   | 2.85   |
| within individuals       | 2.75   | 2.75   |
| between individuals      |        |        |
| % calories from protein  | 14.96  | 14.26  |
| Mean                     | 9.68   | 3.85   |
| Standard deviation       | 9.24   | 2.76   |
| within individuals       | 2.97   | 2.74   |
| between individuals      |        |        |
| Salt intensity (g per 100g) | 0.48   | 0.47   |
| Mean                     | 0.49   | 0.32   |
| Standard deviation       | 0.15   | 0.17   |
| within individuals       |        |        |
| between individuals      |        |        |
| Fibre intensity (g per 100g) | 1.36   | 1.33   |
| Mean                     | 0.77   | 0.44   |
| Standard deviation       | 0.70   | 0.31   |
| within individuals       | 0.33   | 0.32   |
| between individuals      |        |        |
| % calories from healthy foods | 31.02  | 48.34  |
| Mean                     | 22.64  | 13.77  |
| Standard deviation       | 20.50  | 10.00  |
| within individuals       | 9.60   | 9.46   |
| between individuals      |        |        |

Notes: Column (1) shows the mean and variance decomposition for shopping baskets aggregated to the weekly level, column (2) shows the same for shopping baskets aggregated to the monthly level.

We show that this variation is well captured by dividing foods into “healthy” and “unhealthy” on the basis of an index used by the UK government. The nutrient profiling score (NPS) converts the multi-dimensional nutrient profile of each individual food product into a single-dimensional score.\(^{15}\) The NPS is used by the UK government, for example, to regulate advertising of food and drinks on children’s television programming. It is similar in spirit to the Healthy Eating Index used by the US government to measure how well a basket of foods align with key government recommendations. The popular “MyPlate” tool is designed to encourage people to consume diets in line with these recommendations. The NPS ranges from -15 to 40, with a lower score indicating that the product is more healthy.\(^{16}\) The products that have the lowest score are pulses and vegetables, with scores of around -10, and those with highest scores are solid fats, chocolates and biscuits, with scores over 20. We describe how nutrients are recorded in the data and provide details on the construction of the NPS in Online Appendix A.2.

Food products with an NPS below 4 and drinks products with an NPS below 1 are deemed “healthy” – this threshold is used by the UK Government to restrict advertising of unhealthy products to children. We use the government’s cutoff to classify foods as healthy or unhealthy.

### 2.2. Extent of within-individual variation in diet quality

We use the panel element of the data to look at the degree of within-individual (over time) variation in diet quality, in comparison with the cross-sectional variation. Table 1 shows that when purchases are aggregated to the weekly level, the within-individual intertemporal variation in diet quality is larger than the cross-sectional variation. This falls when we aggregate purchases to the monthly level, but within-individual intertemporal variation is still roughly the same as the well documented cross-sectional variation. For example, the within-individual standard deviation of the share of calories from

\(^{15}\) For more details see Rayner et al. (2005), Aranbepola et al. (2008) and Rayner et al. (2009).

\(^{16}\) The NPS gives positive points for the amount of saturated fat, sodium, sugar and calories per 100g and negative points for the amount of fibers, proteins and fruit, vegetables and nuts per 100g.
added sugar (computed across months) is 4%, which is similar to the between-individual standard deviation. The bottom panel of the table shows that this pattern holds for the share of calories from healthy foods.

In the rest of the paper we use monthly aggregation. This means we analyze broad changes in the healthiness of people’s diets rather than day-to-day fluctuations that may be driven by people eating cake on one day and compensating on the following day by abstaining, or by diets that entail high frequency changes (e.g. the 5:2 diet, which prescribes eating as normal for five days and eating a very restricted quantity of calories for two days).

2.3. Within-individual variation through time

One particularly salient dimension in which diet quality varies is that most people tend to make, and fail to keep, New Years’ resolutions to lead more healthy lifestyles, and specifically to eat a healthier diet. Fig. 1 shows data from Google Trends for both the US and UK; panel (a) shows the time trends in Google searches for the word “diet” and panel (b) shows the trends for searches for “healthy foods”. Both searches, in both countries, show spikes in January and a steady decline as the year progresses; this is consistent with the findings in Dai et al. (2014), Dai et al. (2015).

We show that this pattern is evident in our observational data, and has important implications for diet quality. However, it is important to note that this is only one of the dimensions in which diet quality varies within-individual. In Fig. 2, we show that, on average, the share of calories from healthy foods is highest in January and declines steadily over the year reaching a low in December; in Online Appendix B we show this pattern is evident for key nutrients, and for the full sample of households (which includes multiple as well as single occupancy households). In Fig. 2(b) we plot the distribution of changes in the share of calories from healthy food from the first to the final quarter of the year, across individuals.

This figure shows two things. First, there is considerable heterogeneity across individuals in the change in share of calories from healthy food. The majority of individuals see their diet quality decline from Q1 to Q4, but these declines vary from a 20 percentage point decline in the share of calories from healthy food to no change, and some individuals actually see an increase in the healthiness of their diets over the calendar year. This illustrates the importance of allowing for heterogeneity in purchasing patterns across individuals in order to understand the determinants of within-individual variation. Second, these observed changes over the year are substantial. For example, the median decline in the share of calories from healthy foods from Q1 to Q4 is approximately 5 percentage points. This is the same as the difference between the average share of calories from healthy foods bought by a normal weight individual and an obese individual.\footnote{We have information on the body mass index (BMI) for individuals in our sample. Individuals with a BMI of between 18 and 20 (normal weight is classified as a BMI of between 18.5 to 25) get, on average, 51% of their calories from healthy foods, in comparison with obese individuals with a BMI between 35 and 40, who get, on average, 46% of their calories from healthy foods.}

A striking feature of the data is that there is substantial variation in individuals’ diet quality over time, which is driven by changes in the choices they make over which foods to purchase. A second key feature of the data is heterogeneity in this behavior across individuals, in terms of both their average behavior and the variation in their choices over time. Our objective is to understand what drives the within-individual variation in behavior. In doing this it is important to allow for the rich heterogeneity in behavior so clearly apparent in the data.
Fig. 2. Variation over time in % calories from healthy foods. Notes: Panel (a) shows the percentage of calories from healthy foods purchased on each shopping trip for a sample of 3645 individuals over 2005–2011. Healthy foods are defined as those with an NPS of less than 4 (or drinks with an NPS below 1), see the text and Online Appendix B. The grey dotted line shows the mean across individuals on each day, the solid black line shows the mean across individuals in each week. Panel (b) shows the distribution of each individual’s mean change from Q1 to Q4.
3. Quantifying the determinants of variation in food choices

In this section we quantify the extent to which standard economic variables (prices, budgets), along with variation in advertising and weather, drive variation in diet quality. We describe differences in the mean share of calories from healthy foods across individuals, and we recover residual variation in this share (after conditioning on these observables) in order to study within-individual variation in diet quality. We show that variation is higher for poorer and younger households.

3.1. Effect of prices, budgets, advertising and weather

In what follows, we consider \( I \) individuals, indexed \( i \in \{1, \ldots, I\} \), and observe \( T_i \) grocery baskets purchased by each individual \( i \). In order to recover the variation in the share of calories from healthy foods, \( y_{it} \), that is driven by variation in the price vector of healthy foods \( \{p_{h}^{i,t}\} \), less healthy foods \( \{p_{d}^{i,t}\} \), the food budget \( (x_{it}) \), and variation that is due to other factors, \( \{z_{it}\} \), we impose the following restrictions.

First, we separate the vector \( z_{it} \) into a component with observable variables, \( \tilde{z}_{it} \), and an unobservable scalar, \( \epsilon_{it} \), which we assume is additively separable. The observable variables include monthly advertising expenditure on unhealthy foods (confectionery, snacks, soft drinks, prepared and convenience foods), which we obtain from the AC Nielsen Advertising Digest – a data set that records all monthly food and drink advertising expenditure at the brand level (Online Appendix A.4). The vector \( \tilde{z}_{it} \) also includes three variables that capture the weather conditions in individual \( i \)’s local area at time \( t \) – the minimum temperature in that month, the maximum temperature, and the amount of rain (Online Appendix A.5).

Second, we assume that the two price vectors can be approximated by two price indices: one for healthy foods (those with an NPS below the cutoff of 4), \( \Pi_{ht} \); and one for unhealthy foods (those with an NPS above the cutoff of 4), \( \Pi_{dt} \). These price indices are weighted averages of the prices for the goods which comprise each set. The weights are equal to the individual’s mean share of each good from the quantity purchased of each of the two sets of foods – healthy, unhealthy) see Online Appendix A.3 for details. In Section 3.3, we show robustness of our results to approximating the prices vectors with a larger set of price indices.

Under these assumptions we get:

\[
y_{it} = g_i(\Pi_{ht}, \Pi_{dt}, x_{it}, \tilde{z}_{it}) + \epsilon_{it}.
\]

We approximate \( g_i( \cdot \) \) with an expression that is linear in the log of the two price indices, the log of a deflated expenditure term, and the weather and advertising variables, and that has individual specific coefficients. We estimate:

\[
y_{it} = \alpha_i + \beta_i \ln \left( \frac{\Pi_{ht}}{\Pi_{dt}} \right) + \gamma_i \ln \left( \tilde{r}_{it} \right) + \theta_i \tilde{z}_{it} + \epsilon_{it},
\]

(3.1)

where \( \ln(\cdot) \) denotes that we normalize each variable by subtracting the individual specific mean, and:

\[
\ln \left( \tilde{r}_{it} \right) = \ln(x_{it}) - [\tilde{\eta}_i \ln(\Pi_{ht}^b) + (1 - \tilde{\eta}_i) \ln(\Pi_{dt}^b)]
\]

is the log of real food expenditure. Deflating by the price index for food ensures that the share of calories from healthy foods is homogeneous of degree zero in prices and expenditure. Normalizing the variables allows us to interpret \( \alpha_i \) as the individual’s preference for healthy food when she faces her average food budget, average relative prices, and average other observed factors (advertising and weather conditions).

We estimate Eq. (3.1) for each individual using OLS. In Section 3.3, we show that the results are robust to instrumenting log real food expenditure. The parameter estimate \( \tilde{\alpha}_i \) is consumer \( i \)’s “mean” share of calories from healthy food – i.e. evaluated at her mean real food expenditure, mean relative prices and mean values for the other observed factors (advertising and weather conditions). It therefore captures the consumer’s average propensity to choose healthy relative to unhealthy foods (i.e. their average share of calories from healthy foods). The median value of \( \tilde{\alpha}_i \) is 48%, the 25th percentile is 42% and the 75th percentile is 54%.

In Online Appendix D we summarise the mean \( \tilde{\alpha}_i \) across various demographic groups, and also show how the demographics of those in the bottom decile of \( \tilde{\alpha}_i \) compare with those in the bottom decile. High socioeconomic status individuals get a greater share of their calories from healthy foods than lower socioeconomic status individuals. This lines up with the literature, in which socioeconomic differences in health are well documented (e.g. Cutler and Lleras-Muney, 2010). People not in work tend to get a lower share of their calories from healthy foods than people in work. Smokers have less healthy diets than non-smokers: 34% of those in the bottom decile of calories from healthy foods are smokers, compared with 14% in the top decile. There are also intuitive patterns of \( \tilde{\alpha}_i \) with obesity: 74% of those in the bottom decile of calories from healthy foods are overweight or obese, compared with 56% in the top decile.

Kantar Wordpanel asks participants a selection of questions to gauge their attitude to a variety of lifestyle factors, including preferences for healthy and processed food. (see Online Appendix A.6 for details). Table 2 shows the correlation of

18 For instance, \( \ln(\Pi_{ht}^b) = \ln\left( \frac{\Pi_{ht}}{\Pi_{dt}} \right) - \frac{1}{2} \sum_{i} \ln\left( \frac{\Pi_{ht}}{\Pi_{dt}} \right) \).

19 This is the log of nominal food expenditure deflated with a food price index (the weights in the index, \( \tilde{\eta}_i, 1 - \tilde{\eta}_i \) are the individual’s mean share of spending on healthy and unhealthy foods).
Table 2
Variation in mean share of calories from healthy food by stated preferences.

| (1)                                    | (2) Mean | (3) Difference | (4) 95% CI for diff. |
|----------------------------------------|----------|----------------|-----------------------|
| Above median preferences for healthy food | 50.3     |                |                       |
| Below median preferences for healthy food | 46.1     | -4.2           | [-4.8, -3.6]          |
| Above median preferences for processed food | 47.1     |                |                       |
| Below median preferences for processed food | 49.4     | 2.3            | [1.6, 2.9]            |

Notes: The numbers in column (3) are the difference in means from the first row in each group. Confidence intervals reflect uncertainty arising from the sample of individuals, but not over the time series variation for each individual.

Table 3
Partial R$^2$s from regression of $y_{it}$.

| Regressors                        | Mean 25th | Mean 50th | Mean 75th |
|-----------------------------------|-----------|-----------|-----------|
| Price and budget                  | 0.10      | 0.03      | 0.07      | 0.14      |
| Advertising                       | 0.05      | 0.01      | 0.02      | 0.06      |
| Weather                           | 0.09      | 0.03      | 0.06      | 0.11      |
| Price, budget, advertising, weather | 0.20    | 0.12      | 0.18      | 0.26      |

Notes: Table reports moments of the distribution of partial R$^2$ of 3645 separate individual level regressions of Eq. [3.1].

3.2. Residual variation in the share of calories from healthy food

To show the importance of observables in explaining within-individual variation in the share of calories from healthy foods, we compute several partial R$^2$ indices. One of these measures is the variation in this share that is explained by variation in the economic environment (relative prices and food budgets). The others measure the variation in the share that is explained by variation in the weather or advertising. The total R$^2$ measures the amount of the variation in the share of calories from healthy food that is explained by variation in the economic environment (relative prices and food budgets), and other factors such as advertising and weather. Table 3 summarizes these partial R$^2$s.

The mean partial R$^2$ with respect to prices and food budgets across individuals is 0.10 – on average, around 10% of the within-individual variation in the share of calories from healthy food is explained by variation in price and food budgets. The mean total R$^2$ across individuals is 0.20 – on average, around 20% of within-individual variation in the share is explained by prices, food budgets, advertising and weather. This illustrates that individuals’ response to variation in these variables does explain a portion of the variation in the share of calories from healthy food, but a considerable fraction of the fluctuations in the share of calories from healthy foods over the year remains unexplained.

3.2.1. Variation over time

In Fig. 3 we summarize the average (across individuals) variation in $\hat{e}_{it}$ (i.e. the residual share of calories from healthy food) over the year. We plot the deviation in the mean relative to January (pooled over years). We also show the mean (relative to January) of the observed share of calories from healthy foods, $y_{it}$, over the year, which captures the average variation in the share that is driven both by observable and unobservable variables.

Fig. 3 shows that the observed share of expenditure on healthy goods declines from January to March, plateaus from March to August, and then deteriorates from August onward. Once we control for prices, food budgets, advertising and weather variation, the decline in the share of calories from healthy food over the final few months of the year is lessened, but still present. However, this aggregate decline masks a great deal of variation across individuals. Although the majority of people have diets that deteriorate over the calendar year, the magnitude of this decline, and when the decline starts to occur, varies across individuals.

We therefore construct the standard deviation of the residual share of calories from healthy food for each individual $i$ over $t$: $\hat{e}_{it} = \text{sd}(\hat{e}_{it})$; this captures variation over time in the share around the mean that is not driven by that individual’s responses to changes in food prices, food budgets, weather, or advertising. We also construct the standard deviation of the observed share of calories from healthy food for each individual $i$ over $t$: $\hat{e}_{it} = \text{sd}(y_{it})$, which measures the total variation in the individual’s share, driven by both unexplained changes and changes in the economic environment, weather, and advertising.

Fig. 4 compares the distributions of $\hat{e}_{it}$ and $\hat{e}_{it}$ across individuals. The leftward shift of the distribution shows the extent to which accounting for individuals’ responses to variation in prices, food budgets, advertising and weather explains variation...
Fig. 3. Deviations from the mean share of calories from healthy foods over the calendar year. Notes: Black dots show the mean share of calories from healthy food $y_{it}$, in each calendar month, pooled across years. Red squares show the mean residual share of calories from healthy food, $\tilde{\epsilon}_{it}$, which removes the effect of variation in prices, food budgets, advertising, and weather, in each calendar month, pooled across years. Both are measured relative to the mean values in January. 95% confidence intervals are shown. Confidence intervals reflect uncertainty arising from the sample of individuals, but not over the time series variation for each individual.

Fig. 4. Standard deviation of the observed and residual share of calories from healthy food. Notes: The white bars show the distribution of individuals’ standard deviation in the share of calories from healthy food, $\tilde{\sigma}_{i}$. The grey bars show the distribution of individuals’ standard deviation in the residual share of calories from healthy food, $\tilde{\sigma}_{i}$, which takes out the responses to changes in prices, food budgets, advertising and weather.
(a) Standard deviation of residual share of calories from healthy food

(b) Effect of responses to changes in prices, food expenditure, advertising and weather

Fig. 5. Variation in \( \hat{\sigma}_i \) with age. Notes: Age is the individual’s median age while in the sample. The left panel shows the mean standard deviation in the residual share of calories from healthy food across individuals at each age. For each individual we calculate the difference in the standard deviation in the share of calories from healthy food and the standard deviation in the residual share of calories from healthy food, \( \hat{\sigma}_i - \hat{\sigma}_i \); the right panel shows the conditional mean of this difference across ages. 95% confidence intervals are shown. Confidence intervals reflect uncertainty arising from the sample of individuals, but not over the time series variation for each individual.

in their spending on healthy food over time. The average standard deviation of the observed share of calories from healthy food, \( \bar{\sigma}_i \), is 9.9 percentage points, but the average standard deviation in the residual share, \( \hat{\sigma}_i \), is 8.7 percentage points. This means that a substantial proportion of the within-individual variation in the share of calories from healthy food is unexplained by prices, budgets, advertising or weather, even allowing for individuals’ heterogeneous responses to changes in these variables.

3.2.2. Variation with demographics

We consider how the within-individual variation in the share of calories from healthy food differs with age and income. We focus on these demographics because they are the focus of policy intervention.\(^{20}\) For example, the UK government restricts advertising of unhealthy foods targeted at young people, and has recently introduced a tax on sugary soft drinks, motivated by a desire to reduce sugar consumption, and ultimately obesity, in young people. In the US many policies, such as the Supplemental Nutrition Assistance Program, are aimed at improving the diets of low income households. Policies that aim to improve the nutrition of lower income and younger people have been shown to have long run impacts (e.g. Hoynes et al., 2016). In addition, Mani et al. (2013) show that poverty directly impedes cognitive function, thereby leading to poor choices.

Fig. 5 (a) shows how the standard deviation in the residual share of calories from healthy food, \( \hat{\sigma}_i \), varies across age groups. Panel (b) shows how the difference in variation in the observed and residual share, \( (\hat{\sigma}_i - \hat{\sigma}_i) \), varies with age. Together these graphs show two things. First, there is an age gradient. Young people, on average, have more unexplained variation in the share of calories from healthy food. Second, failing to account for the effects of the economic environment, advertising and weather conditions would lead to an over-statement of this gradient. For individuals aged below 40, variation in these observables is, on average, responsible for 1.5 percentage points of the standard deviation in their share of calories from healthy food; for individuals aged over 70, 1.05 percentage points is explained by responses to these variables. One possible explanation for why the quality of older individuals’ diets does not react as strongly to variation in prices and incomes could be that they have more time to shop around and more scope to engage in home production due to a lower opportunity cost of time relative to younger individuals (Aguirau and Hurst, 2007; Aguirau and Hurst, 2013).

Fig. 6 shows a similar pattern for the relationship between \( \hat{\sigma}_i \), \( \hat{\sigma}_i - \hat{\sigma}_i \) and the income distribution. People with low income exhibit more variation in their share of calories from healthy food. This is partly driven by their responses to changes in observable factors – see panel (b). However, it is also the case that lower income individuals have more variation in their residual share of calories from healthy food. This difference is meaningful: individuals in the bottom quintile have a standard deviation in their residual share of calories from healthy food that is more than 2 percentage points larger than individuals in the top quintile.

Failure to account for individuals’ responses to changes in prices, budgets, advertising and weather leads to an over-estimate of the gradient with age and income. This is primarily due to differences in the fluctuations in individuals’ food budgets and how they respond to them. Younger and lower income individuals have larger fluctuations in their food bud-

---

\(^{20}\) In Online Appendix D.2 we show how the residual variation in the share of calories from healthy foods varies with other demographics.
(a) Standard deviation of residual share of calories from healthy food

(b) Effect of responses to changes in prices, food expenditure, advertising and weather

![Graphs](image)

**Fig. 6.** Variation in $\hat{\sigma}_i$ with income. Notes: For all individuals we calculate the mean total spending on fast moving consumer goods (food, alcohol, household supplies, toiletries etc.) across the period they are in the sample, which we use as a proxy for income. In the Online Appendix A.7 we provide further evidence on this measure of income. Percentiles are based on the distribution of this variable across individuals. The left panel shows the conditional mean standard deviation in the residual share of calories from healthy food across income percentiles. For each individual we calculate the difference in the standard deviations in the observed and residual shares of calories from healthy food, $\sigma_i - \hat{\sigma}_i$; the right panel shows the conditional mean of this difference across income percentiles. 95% confidence intervals are shown. Confidence intervals reflect uncertainty arising from the sample of individuals, but not over the time series variation for each individual.

**Table 4**

| Specification                  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|-------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
|                               | Mean $\hat{\sigma}_i$ | $\delta_i$ by age group | $\delta_i$ by income quint. |
| Baseline                      | 8.68 | 9.66 | 8.35 | 1.16 | 10.19 | 7.51 | 1.36 |
| Controls for holidays, birthdays | 8.20 | 9.00 | 7.93 | 1.14 | 9.67 | 7.05 | 1.37 |
| Controls for four price indices | 8.30 | 9.21 | 8.01 | 1.15 | 9.76 | 7.17 | 1.36 |
| Instrumenting food expenditure | 8.84 | 9.80 | 8.52 | 1.15 | 10.37 | 7.65 | 1.35 |

Notes: Column (1) lists the different robustness specifications for which we estimate Eq. (3.1) and construct $\hat{\sigma}_i$ as the individual’s standard deviation in the residual from each regression. Column (2) shows the mean $\hat{\sigma}_i$ across individuals. Columns (3) and (4) show the mean $\hat{\sigma}_i$ for individuals aged under 40 and over 70, respectively, and column (5) shows the ratio of the two means. Columns (6) and (7) show the mean $\hat{\sigma}_i$ for individuals in the lowest and highest income quintiles, respectively, and column (8) shows the ratio of the two means.

gets: the standard deviation of logged real food expenditure is 30% higher for individuals aged under 40 compared with individuals aged over 70, and it is 70% higher for individuals in the bottom income quintile compared with individuals in the top. Across the income gradient, this is amplified by the fact that real food expenditure statistically significantly affects the share of calories from healthy food for more individuals in the bottom quintile (28%) than the top quintile (15%).

### 3.3. Robustness

In this section we show robustness to our specification choices. Table 4 shows that for a range of specifications (details below) there remains substantial within-individual time-series fluctuations in diet quality, and that these fluctuations are larger for younger and lower income individuals.

#### 3.3.1. Holidays and birthdays

Christmas, Easter and birthdays are times at which individuals’ diet quality deteriorates and then again improves (see also Dai et al. (2014)). We estimate a variant of Eq. (3.1) including dummy variables for Christmas, Easter and the individual’s birthday. These celebrations explain a small fraction of the residual share of calories from healthy food – the average standard deviation in the residual share of calories from healthy food falls from 8.7 percentage points to 8.2 percentage points when we control for holidays and birthdays; there remains substantial unexplained variation in diet quality over the year. We find that the correlations with age and income are robust to taking out the variation due to holidays and birthdays.
3.3.2. Four price indices

One possible concern is that, in our estimation of Eq. (3.1), we use two price indices to capture variation in the prices of healthy and unhealthy foods over time. It is not possible to include the full vector of price indices for all goods in Eq. (3.1) while allowing for individual heterogeneity in price effects. However, we do consider a more flexible specification in which we assume that the movements in the price vector can be approximated by four price indices: one for very healthy foods, one for healthy foods, one for unhealthy foods, and one for very unhealthy foods. Including additional price indices does reduce the variation in the residual share of calories from healthy food, but only slightly—the average standard deviation falls from 8.7 percentage points to 8.3 percentage points, and Table 4 shows that the correlations with age and income are robust to including more controls for price variation.

3.3.3. Instrumenting food expenditure

In Section 3.1 we partial out the effect of variation in food prices and budgets. However, it may be the case that unobserved shocks to demand for healthy food also lead to changes in total food budgets. To deal with this possibility, which could affect our interpretation of the residual share of calories from healthy food, we instrument for the expenditure term in Eq. (3.1) using variables that are likely to drive total spending on food, but are less plausible shapers of the preference for healthy food. Our instrument set includes a set of prices from the consumer price index (CPI) and individual monthly spending on non-food items (cleaning products, toiletries, cosmetics). We expect non-food spending and the relative price of food and non-food to influence individuals’ allocation of their total budget between food and other commodities, but not to be correlated with preferences for healthy vis-a-vis unhealthy food. The results are very similar across the OLS and IV specifications. Slightly more of the observed variation in the share of calories from healthy food is explained by prices and budgets in the OLS specification than in the IV specification (the mean standard deviation in the residual share is 8.7 percentage points in the OLS specification, compared with 8.8 percentage points in the IV specification). However, the relationships with age and income are qualitatively similar; see Table 4.

3.3.4. Food outside the home

Our analysis focuses on grocery purchases, which constitute 85% of calories purchased (calculated using the Living Costs and Food Survey, described below), and are therefore key to understanding the nutritional implications of people’s food choices. However, there may be concern that variation in the nutritional quality of people’s at home food purchases is partly offset by changes in their purchases made for consumption outside of the home.

We show that the aggregate patterns presented in Section 2 are neither an artefact of the Kantar Worldpanel data, nor are affected by food consumed outside the home. We use data from the Living Costs and Food Survey (LCFS), which is the main UK expenditure survey and similar to the CEX in the US. It is a repeated cross-sectional two-week expenditure diary and records information on the quantities of foods consumed in and outside of the home. The fact that it is a repeated cross-section means that we cannot look at within-individual time series variation, but we can check whether the aggregate decline in diet quality over the calendar year holds when we look at food in and out of the home. Fig. 7 shows that there is clear decline in the share of calories from healthy foods from January through to December. In addition, the share of spending on food out is stable over the calendar year, which suggests that substitution to food out over the year is unlikely to be confounding our results.

3.4. Does within-individual variation reflect impulsiveness?

Our results show that within-individual fluctuations in diet quality over time are considerable, and are only partly explained by observable factors such as prices, food budgets, advertising and weather changes. We look at whether these fluctuations are correlated with other aspects of individuals’ behavior.

We investigate the relationship between the standard deviations of individuals’ residual share of calories from healthy food, \( \hat{\sigma}_i \), and individuals’ stated attitudes. We use responses to the stated preference and attitude questions from the Kantar Worldpanel on: the tendency to buy things on offer, shopping commitment, and stated impulsiveness (see Online Appendix A.6 for details). Table 5 shows the correlation between these variables and \( \hat{\sigma}_i \). Individuals who have higher stated impulsiveness (i.e., state that they spend money without thinking, or spend more on their credit card than they should) have larger fluctuations in their residual share of calories from healthy foods over the year. Individuals who state that they commit to buying the same brands experience smaller deviations, while those who say that they are more spontaneous and, for example, are influenced by promotions, experience larger deviations in their residual share of calories from healthy foods.

---

21 Very healthy foods are those with an NPS less than 0, \( \Pi^H \); healthy foods are those with an NPS between 0 and 4, \( \Pi^H \); unhealthy foods are those with an NPS between 4 and 10, \( \Pi^H \); and very unhealthy foods are those with an NPS greater than 10, \( \Pi^H \). These price indices are weighted averages of the prices for the goods which comprise each set, and are constructed in an analogous way to the two price indices described in Online Appendix A.3.

22 The prices consist of the all-items CPI, which captures the general price level in the economy, and the CPI component indices for the set of non-housing goods (food, alcohol and tobacco, furniture and equipment, health care, transport, communications, recreation, education, restaurants and hotels, other goods and services).

23 Pooling in the first stage across individuals results in an F-statistic for a test of the joint significance of the instruments of over 700. Estimating the first stage individual-by-individual results in lower F-statistics, and, for some individuals, weak instruments.
In Online Appendix C we show that these correlations are robust to conditioning on gender, age, socioeconomic status, employment status, BMI, whether the individual is a smoker or vegetarian and income. We also show that individuals with lower stated impulsiveness get a higher share of their calories from healthy foods.

Lack of self-control can manifest itself in many ways; for example, it may lead people to consistently eat unhealthy foods, or it may lead to larger deviations from average behavior as individuals try to commit themselves to healthy diets, and then succumb to temptation. We cannot separately identify whether average differences in food choices across individuals reflect lack of self-control or preference heterogeneity. However, by looking within individuals across time, we can measure the extent to which people deviate from their long-run behavior. Thus, one interpretation of the fluctuations in diet quality are that they capture an individual’s ability to act in accordance with her long-run preferences, and that larger fluctuations reflect, at least in part, a lack of self control.

With this interpretation in mind, we return to the demographic correlations described in Section 3.2. Fig. 5 indicates that younger people have larger fluctuations in their residual share of calories from healthy foods than older people; this is consistent with the findings in Ameriks et al. (2007) that show that the young suffer more from self-control problems than older people. Fig. 6 accords with evidence that low income people are more susceptible to self-control problems. Indeed, a number of papers point to low income being causally related to self-control problems. For example, Haushofer and Fehr (2014) and Mani et al. (2013) suggest that the stress and cognitive loads of being in poverty means people are more likely to make unwise decisions and underweight the future. Bernheim et al. (2015) argue that poverty can perpetuate itself by undermining the capacity for self-control: low initial wealth precludes self-control, and hence asset accumulation, creating a poverty trap. Banerjee and Mullainathan (2010) take an alternative approach by assuming that “temptation goods”

![Graph](image-url)

**Fig. 7.** Calorie share of healthy food over the calendar year for food in and out, LCFS 2005-11. Notes: We use data on the food and drink purchases (consumed in and out of the home) made by 42,062 households over 2005-11 recorded in the Living Costs and Food Survey. We construct the share of calories purchased from healthy foods (foods with an NPS below 4, and drinks with an NPS below 1); the line shows the mean across households and years.

**Table 5**

| Stated Preference | (1) Mean | (2) Difference | (3) 95% CI for diff. |
|-------------------|----------|----------------|---------------------|
| Above median stated impulsiveness | 8.70     |                |                     |
| Below median stated impulsiveness  | 8.42     | -0.28          | [-0.50, -0.06]      |
| Above median tendency to buy on promotion | 8.98     | -0.41          | [-0.61, -0.25]      |
| Below median tendency to buy on promotion | 8.38     | -0.61          | [-0.77, -0.45]      |
| Above median shopping commitment  | 8.50     |                |                     |
| Below median shopping commitment  | 8.91     | 0.41           | [0.25, 0.57]        |

Notes: The numbers in column (3) are the difference in means from the first row in each group. Confidence intervals for $\hat{\sigma}_i$ reflect uncertainty arising from the sample of individuals, but not over the time series variation for each individual.
are inferior goods, which leads to a similar conclusion that self-control problems give rise to asset traps. Mastrobuoni and Weinberg (2009) find that retired individuals who have accumulated lower savings over their life cycle are less likely to smooth their food consumption over their Social Security pay periods.

4. A two-selves model of food choice

To structurally interpret the empirical patterns in the previous sections, we set out a model in which an individual’s food choices are driven by the interplay between two different selves. The theoretical literature typically models temptation and self-control problems in a multi-period framework. A number of papers model self-control through dynamically inconsistent preferences, where individuals are characterized by multiple selves across different time periods (for instance a short-sighted present self and a patient future self), and make decisions they later regret because their current preferences differ from their future preferences (see, for instance, Strotz (1955) Peleg and Yaari (1973), Laibson (1997) and O’Donoghue and Rabin (1999)). Gul and Pesendorfer (2001, 2004) instead model temptation and self-control through dynamically consistent preferences. An individual has commitment preferences and temptation preferences; the individual’s choice reflects a compromise between them, and if the individual resists temptation it comes at a utility cost. Bénabou and Pycia (2002) demonstrate that Gul and Pesendorfer (2001)’s canonical model, which is framed as a two-period model, can be reformulated in terms of a costly intra-temporal and -personal conflict between two contemporneous selves, which they term a Planner and a Doer (as in Thaler and Shefrin (1981)). These selves are involved in a game, which stochastically determines whether the individual resists or succumbs to temptation (see also Fudenberg and Levine (2006)).

The average patterns of diet quality deterioration over the calendar year, with resets each January, do not directly accord with the aforementioned theoretical models. Therefore, we choose to interpret our results through the lens of a parsimonious model that is similar in nature and easy to implement empirically. We assume that an individual is characterized by a healthy self and an unhealthy self; both selves are characterized by their own stable and well-behaved preferences. The healthy self has preferences over a basket of healthier foods and the unhealthy self over a basket of less healthy foods. These selves are involved in a cooperative bargaining process that determines the individual’s observed food choices. We impose minimal structure on the interaction between these two selves by only assuming that an individual’s food purchases are Pareto efficient. The influence of the unhealthy relative to the healthy self in the intra-personal bargaining process may vary over time and determines an individual’s food choices. We do not impose additional structure (for example, that may be implied by the more sophisticated two-selves models proposed in the literature) about the nature of the bargaining process between the selves; we discuss this further at the end of the section.

Our model is strongly influenced by the literature on collective household models; see Chiappori (1988, 1992) for seminal work. These models explicitly account for the fact that multi-person households consist of different individuals with their own preferences. Our collective two-selves model is easily brought to data and empirically flexible, meaning that it can readily capture the individual specific non-monotonic changes in diet quality and the quality resets throughout the years. The empirically driven nature of the model differs from many of the theoretical models mentioned above. It allows us to apply a revealed preference approach to analyzing food choices on an individual-by-individual basis; as a direct implication, we can thus naturally account for the rich heterogeneity in food purchases that we documented above.

4.1. Model set-up

We model an individual’s choices over what foods to consume. A first set of foods, \( H \), is associated with a healthy lifestyle and contains items such as fruits, vegetables and whole grains. A second set, \( L \), contains less healthy foods, such as soda, crisps and confectionery. These goods together make up the entire food and non-alcoholic drink grocery basket and are each an aggregate of nutritionally similar food or drink products (UPCs or barcodes). See Online Appendix A.2 for a more detailed discussion.

Individuals are indexed by \( i \in \{1, \ldots, I\} \), and we observe \( T_i \) grocery baskets purchased by each individual \( i \). For each observation \( t \in \{1, \ldots, T_i\} \), we denote the quantities of the healthy food items by \( q^h_{it} \in \mathbb{R}^H_t \) and the quantities of the less healthy food items by \( q^l_{it} \in \mathbb{R}^L_t \). The market prices associated with these goods are denoted by \( p^H_t \in \mathbb{R}^H_{t+1} \) and \( p^L_t \in \mathbb{R}^L_{t+1} \), respectively. The individual’s food budget spent on healthy food items is denoted by \( x^h_i \) and is equal to \( p^H_t q^h_{it} \), the food budget spent on less healthy food is denoted by \( x^l_i \) and is equal to \( p^L_t q^l_{it} \). The food budget of consumer \( i \) at time \( t \) is denoted by \( x_i \), where \( x_i = x^h_i + x^l_i \). To summarize, for each individual \( i \in \{1, \ldots, I\} \) we observe the data \( S^i = \{ (p^H_t, p^L_t, q^h_{it}, q^l_{it}), t = 1, \ldots, T_i \} \).

We propose a two-selves model in which we assume that individual \( i \) is characterized by two selves, each with stable and rational preferences. The first self is associated with a healthy lifestyle and derives utility from only the healthy food

\[ (p^H_t, p^L_t, q^h_{it}, q^l_{it}), t = 1, \ldots, T_i \]
items $q_i^h$. The second self derives utility from only the unhealthy food items $q_i^u$. The preferences of each self are represented by the well-behaved utility functions $u^h(q_i^h)$ and $u^l(q_i^l)$. The two selves enter into a bargaining process that is different for every individual and that may not be stable over time.\footnote{Note that we assume that food purchases are separable from other purchases.}

More formally, Pareto efficiency implies that individual $i$'s observed food purchase behavior $(q_i^h, q_i^l)$ can be represented by the solution of the following maximization problem:

$$
\max_{q^h, q^l} \mu_{it} u^h(q_i^h) + (1 - \mu_{it}) u^l(q_i^l)
$$

subject to

$$
p_i^h q_i^h + p_i^l q_i^l \leq x_{it}.
$$

In this representation the parameter $\mu_{it} \in [0, 1]$ is a Pareto weight that represents the bargaining weight of the healthy self in consumer $i$'s optimization problem in period $t$. If $\mu_{it}$ equals one, then the individual behaves according to the healthy self's preferences, while if $\mu_{it}$ equals zero the allocation of the food budget is entirely determined by the unhealthy self's preferences. We note that the Pareto weight, $\mu_{it}$, depends, in general, on the food prices $p_i^h$ and $p_i^l$, and on the food expenditure $x_{it}$, as well as on other factors (that may vary across time and individuals, such as advertising or the weather).

An appealing feature of our framework is that it gives a structural interpretation to the share of spending on healthy foods. The "sharing rule" gives the relative share of food spending by the healthy self in total food expenditure, $\eta_{it} = x_{ih}/x_{it}$. There is a one-to-one relation between the Pareto weight $\mu_{it}$ and the sharing rule for our two-selves model: for given prices and food budget, the share of spending on healthy food is a monotone transformation of the bargaining weight of the healthy self. Hence, the sharing rule is a function of the same variables that affect the Pareto weight. In other words, the model allows us to interpret variation in the share of spending on healthy foods as variation in the bargaining weight of the healthy self. In Section 3 we describe the determinants of variation in the share of calories from healthy foods – the results are very similar if the share of spending is used instead. We can therefore interpret the greater residual variation in this share among lower income and younger individuals as greater variation in the bargaining weight of the healthy vis-à-vis the unhealthy self.

4.2. Implications of the model

4.2.1. Rationalizability

Browning and Chiappori (1998) and Chiappori and Ekeland (2009) characterize the testable implications of the collective model, which has a similar structure to our two-selves model. Cherchye et al. (2007, 2011a) characterized similar conditions in a revealed preference setting à la Samuelson (1938), Afriat (1967) and Varian (1982). In what follows, we sketch the basic intuition of the testable revealed preference implications of our collective two-selves model. We refer to Online Appendix E.1 for more formal details.

Let $S_i^h = \{ (p_i^h; q_i^h), t = 1, \ldots, T_i \}$ be the set of observations of individual $i$'s healthy self. Similarly, $S_i^l = \{ (p_i^l; q_i^l), t = 1, \ldots, T_i \}$ contains the set of observations of the less healthy self. A self $j (j = h, l)$ of individual $i$ can be shown to have rational preferences if and only if the set $S_i^j$ satisfies a series of so-called Afriat-inequalities, which we refer to as the Afriat Condition.\footnote{Varian (1982) has shown that this Afriat Condition for a given set $S_i^h$ can equivalently be formulated as a GARP (Generalized Axiom of Revealed Preference) condition. This GARP condition is more popularly used in the applied literature on empirical revealed preference analysis. However, for the purpose of the current study, the Afriat Condition is more practical than its equivalent GARP formulation.} Given that all goods are exclusively associated with only one self, a necessary and sufficient condition for Pareto efficiency of the bargaining process between the two selves is that both sets $S_i^h$ and $S_i^l$ independently satisfy the Afriat Condition (see Cherchye et al. (2011a)). These conditions do not impose any explicit structure on the bargaining process. In particular, the impact of the factors that affect the bargaining process is implicitly modeled via the sharing of the budget: we do not need an exhaustive list of all variables that influence the bargaining process when empirically checking the testable implications of our two-selves model. The Afriat Condition does not entail any particular restrictions on the evolution of the bargaining process between the two selves; how the bargaining weights evolve over the year is an empirical question. The pattern observed in Fig. 3, which shows that the observed share of expenditure on healthy goods deteriorates from January to December, with an important reset in January, can be interpreted as a decreasing relative weight of the healthy self throughout the calendar year, while that self’s bargaining weight is associated with a discrete increase in January (which may be associated with the observed New Year’s resolutions).

The Afriat Condition can be checked by means of standard linear programming techniques. We can apply this approach separately to each individual $i$, so it does not impose homogeneity of preferences across consumers. This is an important feature in light of our empirical evidence showing considerable variation across consumers in their food purchasing behavior. Further, we note that the Afriat Condition provides a pass/fail test of rational preferences: either the data satisfy the condition or they do not. It is therefore useful to measure how close the observed behavior is to exact rationalizability in the case that one or both of the sets $S_i^h$ and $S_i^l$ violates the Afriat Condition for a given individual $i$. For this purpose, we use (a two-selves, weighted, version of) the Afriat index (Afriat (1973)); see Online Appendix E.2 for details. This index measures
the fraction by which observed expenditures must be adapted for the data to be rationalized by the model. The Afriat index is frequently used in revealed preference applications to assess the goodness-of-fit of a given model.

4.2.2. Implementation of revealed preference tests

We use the above described conditions to check how well the two-selves model rationalizes the data. We aggregate the more than 100,000 UPCs recorded in our data to 85 goods based on their nutritional characteristics; see Online Appendix A.2. As in Section 3, we use the government’s cutoff to classify foods as healthy or unhealthy. This results in 51 healthy goods (i.e. $H = 51$) and 34 unhealthy goods (i.e. $L = 34$). We construct a price index for each good that is a weighted average of the product level region-month specific prices, where the weights reflect the quantity share of products; see Online Appendix A.3. To summarize, for each individual, we use data on the prices and quantities purchased of 85 goods, partitioned in 51 healthy and 34 unhealthy goods, in $T_i \in [24, 84]$ months.

Fig. 8 (a) shows the distribution of the weighted Afriat index for the two-selves model. The Afriat indices are very high, indicating that only small perturbations (1-2% on average) of the budget are needed to ensure purchase behavior is rationalized by the two-selves model. To compute a measure of the power of the revealed preference test we construct Afriat indices for random draws from budget sets for each individual, as described in Online Appendix E.3. We calculate the proportion of random draws that have Afriat indices greater than the true Afriat index computed with the actual data for the given individual. This can be interpreted as the probability that the true Afriat index is below that implied by random behavior. Fig. 8(b) shows the distribution of the probabilities – they are concentrated around zero, indicating that the test has sufficient power to discriminate between observed and random behavior. In Online Appendix E we also present the distributions of the Afriat Indices separately for the healthy and unhealthy self; the Afriat Indices are slightly higher for the unhealthy self, however, once we account for the differences in power between the two tests, the empirical fit for the healthy and unhealthy self is similar.

A natural alternative to the two-selves model is a single-self model, in which an individual has a single stable utility function defined over all 85 goods. Data consistency with the two-selves model may well be more demanding than data consistency with the single-self model. More precisely, the single-self model imposes consistency with the Afriat Condition for (only) the full data set containing 85 goods. In contrast, the two-selves model imposes data consistency with the Afriat Condition for 2 separate data sets simultaneously, i.e. the data set containing the 51 healthy goods and the data set containing the 34 unhealthy goods. However, as the data sets considered for the single-self and two-selves models are different, we cannot determine a priori whether one model is more restrictive than the other; in general, this will depend on the specific data. A fair comparison of the two models must simultaneously consider their empirical fit and their discriminatory power (see, for instance, Beatty and Crawford (2011) for more discussion). Therefore, for each of the two models we calculate the fraction of individuals for which the Afriat index for the observed behavior exceeds the Afriat index for random behavior (which we define as the average Afriat index over the random draws from the individuals’ budget sets). A higher fraction reveals a better empirical performance of the behavioral model under consideration. The fraction equals 87% for the single-self model and 91% for the two-selves model. We interpret this as evidence that the two-selves model does a better job in explaining the observed behavior than the single-self model.

It is worth noting that the two-selves model and the single-self model are not nested (see, for example, Section 3.B.2 in Chiappori (1988)). Our two-selves model assumes two stable utility functions for both types of goods that are summed in the
individuals’ maximization process (see Section 4.1). This only nests a single-self model with a strong separability structure related to the healthy and unhealthy goods. Imposing the additional structure of strong separability on the single-self model decreases further its goodness-of-fit and thus its Afriat indices.

We conclude that our two-selves model provides a good fit of the data. We cannot, of course, rule out that alternative models would also explain the observed food purchasing behavior of individuals. However, the appeal of our model is that it introduces time-varying preferences in a parsimonious way, while retaining an economic interpretation and allowing for the observed individual heterogeneity in the previous sections. It draws on the literature on multi-selves models, by interpreting changes over time as resulting from the interaction between the two selves. However, our distinguishing feature is that we do not impose any restrictions on this interaction process (e.g. commitment, long term versus short term aspirations, externalities, etc.) in order to fully capture the observed differences across individuals. Extending the structure of our model is a promising avenue for further research. Such an extension may involve a more specific modeling of the mechanisms driving the interplay between the healthy and unhealthy selves, which may draw from the theoretical literature on temptation and self-control cited above.

5. Summary and concluding comments

We show that within-individual, across time fluctuations in food purchasing are large, and have first-order implications for diet quality. The variation follows a clear within-year pattern, where diet deteriorates over the course of the calendar year, and resets each January. We provide empirical evidence using observational data on the entire grocery basket that is consistent with other findings in the behavioral literature. We also document substantial heterogeneity across people, both in their average purchasing patterns and the extent to which their food choices vary over time.

We demonstrate that the fluctuations in the share of calories from healthy food across time are not fully explained by responses to the economic environment or other factors, such as advertising and the weather. We recover the “residual” share of calories from healthy food, which removes individuals’ responses to changes in these other factors. Using data on individuals’ stated impulsiveness, we provide suggestive evidence that fluctuations are larger for individuals with higher impulsiveness. We also find that variation in the residual share is larger for younger and lower income individuals.

To structurally interpret these empirical findings, we present a model that is based on the literature on multiple selves. We use nonparametric revealed preference conditions to show that our data on food purchases can be rationalized by a model in which food choices are a compromise between a healthy and an unhealthy self. This model outperforms the standard single-self model, and provides an economic interpretation to the observed variation in our purchasing data. Specifically, the sharing rule, or share of spending on healthy foods, can be interpreted as the bargaining weight of the healthy self.

Our results are informative for governments struggling to deal with diet related disease. There remains a challenge in how to design policy that encourages individuals to behave as they do at their healthiest times. One potentially fruitful avenue for further research is the possible welfare enhancing role of commitment devices, which can provide a way for people to achieve their long-run preferred outcome and avoid self-control problems (see Bryan et al. (2010) for a survey). It is unclear whether the market will provide such commitment devices (Gottlieb, 2008), and government policy can compensate for the inability of the market to efficiently provide commitment devices. For example, “sin taxes”, which raise the price of tempting goods, can increase welfare if enough people are time inconsistent (O’Donoghue and Rabin, 2003). Determining the optimal sin tax level entails trading off the welfare gain to those suffering from internalities against the welfare loss of those that do not suffer from internalities but who face higher prices (Griffith et al., 2017). Our results suggest this trade-off may also apply within individual over time.

Acknowledgements

The authors gratefully acknowledge financial support under ERC-2015-AdG-694822 and ERC-2013-CoG-614221, the ESRC grant number ES/M010147/1, the British Academy under pF160093, the FWO and FNRS. Data supplied by Kantar. We also thank the editor and two anonymous referees for helpful comments. The use of Kantar data in this work does not imply the endorsement of Kantar in relation to the interpretation or analysis of the data. All errors and omissions remained the responsibility of the authors.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.euroecorev.2020.103478.

References

Afriat, S.N., 1967. The construction of utility functions from expenditure data. Int. Econ. Rev. 8 (1), 67–77.
Afriat, S.N., 1973. On a system of inequalities in demand analysis: an extension of the classical method. Int. Econ. Rev. 14 (2), 460–472. doi:10.2307/2525934.
Aguiar, M., Hurst, E., 2007. Life-cycle prices and production. Am. Econ. Rev. 97 (5), 1533–1559. doi:10.1257/aer.97.5.1533.
Aguiar, M., Hurst, E., 2013. Deconstructing life cycle expenditure. J. Polit. Econ. 121 (3), 437–492. doi:10.1086/670740.
Mastrobuoni, G., Weinberg, M., 2009. Heterogeneity in intra-monthly consumption patterns, self-control, and savings at retirement. Am. Econ. J. 1 (2), 163–189. doi:10.1257/pol.1.2.163.

Milkman, K.L., Rogers, T., Bazerman, M.H., 2010. I'll have the ice cream soon and the vegetables later: a study of online grocery purchases and order lead time. Mark. Lett. 21 (1), 17–35. doi: 10.1007/s11002-009-9087-0.

O'Donoghue, T., Rabin, M., 1999. Doing it now or later. Am. Econ. Rev. 89 (1), 103–124.

O'Donoghue, T., Rabin, M., 2003. Studying optimal paternalism, illustrated by a model of sin taxes. Am. Econ. Rev. 93 (2), 186–191.

Oster, S.M., Morton, F.M.S., 2005. Behavioral biases meet the market: the case of magazine subscription prices. BE J. Econ. Anal. Policy 5 (1).

Peleg, B., Yaari, M.E., 1973. On the existence of a consistent course of action when tastes are changing. Rev. Econ. Stud. 40 (3), 391. doi:10.2307/2296458.

Public Health England, 2017. Sugar reduction: Achieving the 20% - GOV.UK. https://www.gov.uk/government/publications/sugar-reduction-achieving-the-20.

Rabin, M., 1998. Psychology and economics. J. Econ. Lit. 36 (1), 11–46.

Rayner, M., Scarborough, P., Boxer, A., Stockley, L., 2005. Nutrient Profiles: Development of Final Model, Final Report for the Food Standards Agency. British Heart Foundation Health Promotion Research Group, Department of Public Health, University of Oxford, Oxford.

Rayner, M., Scarborough, P., Lobstein, T., 2009. The UK Ofcom Nutrient Profiling Model: Defining Healthy and Unhealthy Foods and Drinks for TV Advertising to children.

Read, D., Loewenstein, G., Kalyanaraman, S., 1999. Mixing virtue and vice: combining the immediacy effect and the diversification heuristic. J. Behav. Decis. Mak. 12 (4), 257–273.

Read, D., Van Leeuwen, B., 1998. Predicting hunger: the effects of appetite and delay on choice. Organ. Behav. Hum. Decis. Process. 76 (2), 189–205.

Sadoff, S., Samek, A.S., Sprenger, C., 2015. Dynamic inconsistency in food choice: experimental evidence from a food desert. mimeo.

Samuelson, P.A., 1938. A note on the pure theory of consumer’s behaviour. Economica 5 (17), 61–71. doi:10.2307/2548836.

Shapiro, J.M., 2005. Is there a daily discount rate? Evidence from the food stamp nutrition cycle. J. Public Econ. 89 (2–3), 303–325. doi:10.1016/j.jpubeco.2004.05.003.

Strotz, R.H., 1955. Myopia and inconsistency in dynamic utility maximization. Rev. Econ. Stud. 23 (3), 165–180.

Thaler, R.H., Shefrin, H.M., 1981. An economic theory of self-control. J. Polit. Econ. 89 (2), 392–406. doi:10.1086/260971.

Trobe, Y., Fishbach, A., 2000. Counteractive self-control in overcoming temptation. J. Pers. Soc. Psychol. 79 (4), 493–506. doi:10.1037/0022-3514.79.4.493.

US Department of Agriculture, 2015. Dietary Guidelines for Americans 2015–2020, 8.

Varian, H.R., 1982. The nonparametric approach to demand analysis. Econometrica 50 (4), 945–973.

World Health Organization, 2015. Healthy diet.