Spending profiles predict savings*

Fabian Gunzinger Neil Stewart

September 29, 2022

Contents

*We are grateful to Redzo Mujcic and and Zvi Safra for helpful comments. The research was supported by Economic and Social Research Council grant number ES/V004867/1. WBS ethics code: E-414-01-20. Gunzinger: Warwick Business School, fabian.gunzinger@warwick.ac.uk; Stewart: Warwick Business School, neil.stewart@wbs.ac.uk.

1 Introduction

Ignore this section for now. Below are notes.

This paper documents variations in spending profiles and payments into emergency savings for a large set of users of a financial management app and shows that spending profiles predict emergency savings.

We define emergency savings as inflows into savings accounts. These savings will be made up of savings for particular goals – a new car, a holiday, a wedding – and savings directed towards building up a buffer for financial emergencies. Because in our data we cannot distinguish between these two cases, we refer to all of them as emergency savings.[1] These short-term savings are distinct from long-term savings aimed to build up funds for retirement, either through individually owned pension and investment vehicles or employer-linked pension schemes. Both kinds of saving are important for financial well-being, yet while there is a large literature on pension savings, little is known about how individuals save for the short-term.[2]

Studying the determinants of emergency savings is important because around a quarter of adults in the UK and the US are unable to cover irregular expenses like car and medical bills: In the UK, 25 percent of adults would be unable to cover an unexpected bill of £300 (Philipps et al. 2021), while in the US, about 30 percent would be unable to cover a $400 bill (Governors of the Federal Reserve System 2022). Similarly, resarch in the UK has shown that having £1000 in savings reduces by more than half a household's chances of falling into debt that leads to financial problems (Philipps et al. 2021).

Studying spending profiles is of interest because:

- Our understanding of how individuals spend their money is based on survey data.
- Large-scale transaction-level data offers the possibility to study spending

[1] MDB allows users to create custom tags and some users use them to indicate the intended use for their savings transactions (e.g. "wedding", "holidays"). But only a very small number of transactions have such tags, and we do not pursue this further.

[2] Well-documented behavioural biases that help explain undersaving for pensions are, among others, present bias (Laibson 1997, Laibson and Marzilli-Ericson 2019), inertia (Madrian and Shea 2001), over-extrapolation (Choi et al. 2009), and limited self-control and willpower (Thaler and Shefrin 1981, Benhabib and Bisin 2005, Fudenberg and Levine 2006, Loewenstein and O'Donoghue 2004, Gul and Pesendorfer 2001). One danger of viewing low savings mainly as a result of behavioural biases is that while these biases likely do play some role and designing environments and tools to help correct them are thus part of the solution, it is at least conceivable that this is an area where the focus on behaviour-level solutions distracts from an effort to find more effective society-level solutions, a danger inherent in behavioural science research convincingly highlighted in Chater and Loewenstein (2022): if the main problem is that many individuals are unable to earn enough to save, then the effectiveness of helping them manage their low incomes more effectively pales in comparison with efforts to help them earn more.

behaviour based on real-time data that are automatically collected for a large number of individuals. Such data has only become available ver recently and have not, thus far, been used to investigate systematically how individuals spend their money.

- Research in psychology suggests that disorder is maladaptive and associated with a range of negative outcomes such as impaired executive function (Vernon-Feagans et al. 2016), lower cognitive inhibition (Mittal et al. 2015), and activation of anxiety-related neural circuits (Hirsh et al. 2012). In the study of human behaviour, more chaotic behaviour has been found to predict the a higher number of visits to and higher spend in supermarkets (Guidotti et al. 2015), higher calorie intake (Skatova et al. 2019) and financial distress (Muggleton et al. 2020).

We hypothesise that less predictable spending patterns are associated with a lower probability for making payments into emergency savings accounts. Possible channels:

- Disorder (personal life or environment): leads to more impulsive shopping behaviour and makes forgetting to save more likely.
- Scarcity: life challenges focus attention away from deliberate shopping, causing hore impulse purchases, and make fortetting to save more likely.

What we do:

- Systematically documenting emergency savings patterns.
- Systematically documenting variation in spending profiles.
- Showing that unpredictability in spending profiles is associated with lower emergency savings.

Contribution to literatures:

- Understanding emergency savings behaviour (nest, aspen reports), (Sabat and Gallagher 2019) for sources on short-term savings literature, Colby and Chapman (2013) for lit on savings goals. See Colby and Chapman (2013) has useful literature review on short-term savings and suggests that subgoals can increase willingness to forego short-amounts in the present because they move the reference point in a prospect-theory framework. Philipps et al. (2021) present results from an employer-linked initiative that offers employees to have a portion of their salary automatically transferred into a savings pot. Policy literature: (CAN 2019, CFPB 2017, MPS 2018). Older literature:Savings lit: Lunt and Livingstone (1991) and Oaten and Cheng (2007)

- Understanding effect of behavioural entropy - eliciting useful personality characteristics from large-scale data.
- Use of high-frequency transaction data (itself a sub-literature of use of newly available large-scale datasets).

2 Methods

2.1 Dataset description

We use data from Money Dashboard (MDB), a financial management app that allows its users to link accounts from different banks to obtain an integrated view of their finances.[3] The full dataset contains more than 500 million transactions made between 2012 and June 2020 by about 270,000 users, and provides information such as date, amount, and description about the transaction as well as account and user-level information.

The main advantages of the data for the study of consumer financial behaviour are its high frequency, that it can be cheaply collected for a very large number of users, that collection is automatic and thus less prone to errors and unaffected by biases that bedevil survey measures, and that it offers a view of consumers' entire financial life across all their accounts, rather than just a view of their accounts held at a single bank, provided they added all their accounts to MDB. The main limitation is the non-representativeness of the sample relative to the population as a whole. Financial management apps are known to be used disproportionally by men, younger people, and people of higher socioeconomic status (Carlin et al. 2019). Also, as pointed out in Gelman et al. (2014), a willingness to share financial information with a third party might not only select on demographic characteristics, but also for an increased need for financial management or a higher degree of financial sophistication. Because our analysis does not rely on representativeness, we do not address this.[4] Another major limitation is imperfect transaction labelling. Money Dasbhoard automatically classifies transaction into different categories based on transaction descriptions and merchant information. While this classification is quite mature and reliable for transactions with major retailers, it is imprecise and at times incorrect for less common transactions. We discuss this further in the next section.

[3]https://www.moneydashboard.com.

[4]For an example of how re-weighing can be used to mitigate the non-representative issue, see Bourquin et al. (2020).

2.2 Preprocessing and sample selection

We use the dataset described above for a number of projects, and perform a number of steps to create a minimally cleaned version of the dataset that is the basis for all such projects. These steps are performed in a dedicated data repository and not run as part of this project.[5] Here, we briefly describe the main cleaning steps and their rationale. We drop all transactions with a missing description string because these cannot be categoriesed, and all transactions that are not automatically categoriesed by the app. Dropping these transactions makes it likely that we will underestimate amounts spent and saved, but minimises the risk of incorrectly classified transactions. We also group transactions into transfer, spend, and income subgroups, following Muggleton et al. (2020) to define spend subgroups and Hacıoğlu-Hoke et al. (2021) to define income subgroups.[6] Finally, we classify as duplicates and drop transactions with identical user ID, account ID, date, amount, and transaction description. This will drop some genuine transactions, such as when a user buys two identical cups of coffees at the same coffee shop on the same day. However, data inspection suggests that in most cases, we remove genuine duplicates.

To minimise the influence of outliers, we winsorise all variables at the 1 percent level or – if we winsorise on both ends of the distribution – at the 0.5 percent level.[7] We rely on winsorisation (replacing top values with percentile values) instead of trimming (replacing top values with missing values) because data inspection suggests that in most cases, very large (absolute) values are not the result of data errors, which would call for trimming, but reflect genuine outcomes, which makes winsorising appropriate because it leaves these observations in the data while lowering their leverage to influence results.

The overall aim of sample selection is to restrict our sample to users for whom we can observe a regular income, can be reasonably sure that they have added all their bank account to MDB, and for whom we observe at least six months of data. Table 1 summarises the sample selection steps we applied, associated data losses, and the size of our final sample.

We start by dropping the first and last month of data for each user because we are unlikely to observe users' complete data in these months, which might affect selection for the criteria below. Next, we keep only users whom we observe for at least 6 months to ensure that we have a suitable amount of data for each user. We also require users to have at least one current and one savings accounts to be able to oberve their spending and savings behaviour. The next few steps

[5] The module with all cleaning functions is available on Github.

[6] The precise list used to classify transactions is available on Github.

[7] The code that performs the winsorisation are available on Github.

Table 1: Sample selection

	Users	User-months	Txns	Txns (m£)
Raw sample	271,856	7,948,520	662,112,975	124,573
Drop first and last month	265,760	7,406,482	643,851,490	121,098
At least 6 months of data	231,888	7,318,202	638,056,402	120,155
At least one savings account	144,309	4,788,799	445,037,084	90,541
At least one current account	141,514	4,723,402	439,698,170	89,629
At least £5,000 of annual income	54,248	1,636,887	172,146,281	33,921
At least 10 spend txns each month	40,596	1,186,503	133,867,104	24,998
At least 4 grocery txns each month	18,468	472,656	60,646,753	9,472
At least £200 of monthly spend	18,254	467,639	60,189,168	9,435
Complete demographic information	14,704	394,377	50,895,612	7,923
Working age	14,539	389,790	50,410,659	7,774
Final sample	14,539	389,790	50,410,659	7,774

Notes: Number of users, user-months, transactions, and transaction volume in millions of British Pounds left in our sample after each sample selection step.

are all aimed at selecting users about whom we can be reasonably sure that they have added all their financial accounts. We thus select on criteria we would expect to see for all such users: an annual income of at least £5,000, at least 10 monthly spending transaction, at least 4 monthly grocery transactions, and a total monthly spend of at least £200. Finally, we keep only users for whom we observe all relevant demographic characteristics, and who are between 18 and 65 years old. The reason for the last step is that younger users and retirees will plausibly have different fianncial objectives than people in their working age, which is the population we are interested about.[8]

2.3 Dependent variables

Our outcome variable is biniry indicator for whether of not a user made a payment into any of their savings account in a given month. We classify as payments into savings accounts all savings account credits of £5 or more.[9]

The reason we focus on a binary indicator is twofold: first, becaus we hypothesis that chaotic or difficult life circumstances that are also reflected in spending entropy might make it harder to remember to save, for which the accurate thest is to see whether spending entropy is related to the likelihood of making ang savings transctions. Second, research by MPS (2018) suggests that to build up sufficient emergency funds over time forming a habit of saving regularly is more important than the specific amounts saved month to month.

[8]The code that implements the selection criteria is available on Github.

[9]While standing order transactions are unlikely to be related to entropy in the short-run, we do not exclude such transactions since, best we can tell, the only account for a small fraction of total transactions.

2.4 Spending profiles

We define a user's spending profile as the distribution of the number of spending transactionas across different spend categories. To summarise these distributions, we calculate spending entropy, based on the formula proposed by Shannon (1948), who defines entropy as $H = -\sum p_i log(p_i)$, which sums, for all possible events, the product of the probability of an event i occuring with the logarithm of that probability.[10] The base of the logarithm is often chosen to be 2, though other choices are possible. Entropy is a cornerstone of information theory, where it measures the amount of information contained in an event. In the behavioural sciences, behavioural entropy has recently been shown to predict the frequency of grocery visits and the per-capita spend per visit (Guidotti et al. 2015), the amount of calories consumed (Skatova et al. 2019), and the propensity for financial distress (Muggleton et al. 2020). In our context, we define the entropy of a user's spending profile in a particular period as (we omit individual and time subscripts to keep the notation simpler):

$$H = -\sum_{c \in \mathcal{C}} p_c log(p_c), \tag{1}$$

where $\mathcal{C}$ is the set of all spending categories, p_c the probability that an individual makes a purchase in spending category c, and log the base 2 logarithm.

Higher entropy means that transactions are more equal across different spending categories, which makes it hard to predict the next transaction, whereas low entropy profiles have the bulk of transactions in a few dominant categories (such as groceries and transportation) and have relatively few transactions in other categories.[11] For simpler interpretation of our regression coefficients below, we standardise entropy scores to have a mean of 0 and a standard deviation of 1.

One limitation introduced by the imperfect transaction labelling in the MDB data is that entropy scores for high-entropy individuals will be biased downwards. This happens because unlabelled transactions tend to be transactions that are rare (i.e. not grocery or Amazon purchases), and it is high-entropy individuals that are more likely to engage in rare transactions. Because our analysis mainly relies on relative entropy levels, this is not of major consequence and we do not pursue this further.

We calculate entropy based on three sets of spend categories. The first measure is based on 9 spending categories used by Muggleton et al. (2020). The second measure is based on our own, more fine-grained, categorisation into 48

[10] Shannon entropy is customarily denoted as H following Shannon's own naming after Ludwig Boltzman's 1872 H-theorem in statistical mechanics, to which it is analagous.

[11] For further discussion on how to interpret Equation 1, see Appendix A.

different categories.[12] The third measure is based on merchant names, as labelled by Money Dashboard.

We also calculate spending category probabilities in two different ways. To calculate what we call "unsmoothed" entropy scores, we calculate the p_cs in Equation 1 as simple frequentist probabilities

$$p_c = \frac{f_c}{F}, \tag{2}$$

where f_c is the number of transactions in spend category c (the frequency with which c occurs) and $F = \sum_{c \in \mathcal{C}} f_c$ the total number of spending transactions. To avoid taking the log of zero for categories with zero transactions, the sum in Equation 1 is taken over categories with positive transaction counts only.[13] To calculate "smoothed" entropy scores, we apply additive smoothing to calculate propabilities as

$$p_c^s = \frac{f_c + 1}{F + |\mathcal{C}|}, \tag{3}$$

where the size of set $\mathcal{C}$, $|\mathcal{C}|$, is the number of unique spending categories. Hence, additive smoothing simply adds one to the numerator and the number of unqiue spending categories to the denominator of the unsmoothed probabilities. Because categories with a zero transaction count will have a numerator of 1, the sum in Equation 1 will be taken over all categories.

2.5 Summary statistics

Table 2 provides summary statistics and Figure 1 distributions for the key variables used in our analysis as well as for some sample demographics. Income and spending distributions are being compared to those of the Living Cost and Food Survey, which is the ONS's main survey on household spending. We can see that the sample distributions broadly match the UK-wide ones but that we tend to underestimate incomes and overestimate spend. As discussed above, our sample is biased towards high-earning individuals, so that we would expect both income and spend to be higher than across the whole population. The main reason this is not the case is likely due to the imperfect identification of income transactions in the Money Dasbhoard data.

Individuals in our data make at least one savings transaction in half of all observed periods, and have at least some income in 98 percent of those periods.

[12] The precise mapping from MDB transaction tags into 9 and 48 categories is available on Github here and here, respectively.

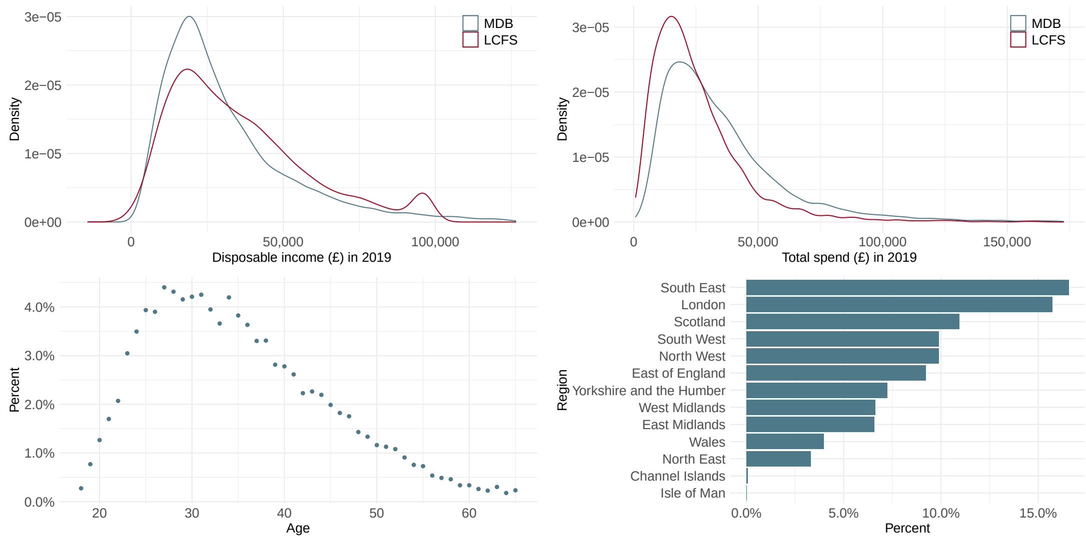
[13] This is automatically handled by the entropy implementation of Python's SciPy package, which is what we use to calculate entropy scores.

Table 2: Summary statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Month income	2.77	2.23	0.00	1.45	2.18	3.43	13.69
Has income in month	0.98	0.13	0	1	1	1	1
Has savings	0.50	0.50	0	0	1	1	1
Month spend	2.90	2.50	0.20	1.37	2.20	3.49	16.05
Age	35.72	9.74	18	28	34	42	65
Female	0.43	0.49	0	0	0	1	1
Urban	0.85	0.36	0	1	1	1	1
Unique categories (9)	7.84	1.05	1	7	8	9	9
Unique categories (48)	16.54	4.13	1	14	16	19	35
Unique categories (Merch.)	26.78	9.35	2	20	26	33	85

Notes: Summary statistics for main variables used in analysis and user characteristics. Unique categories refer to the number of distinct spending categories that invididuals made purchases in within a given month for spending transaction categorisation based on 9 categories, 48 categories, and merchant names.

Figure 1: Demographic characteristics of Money Dashboard users

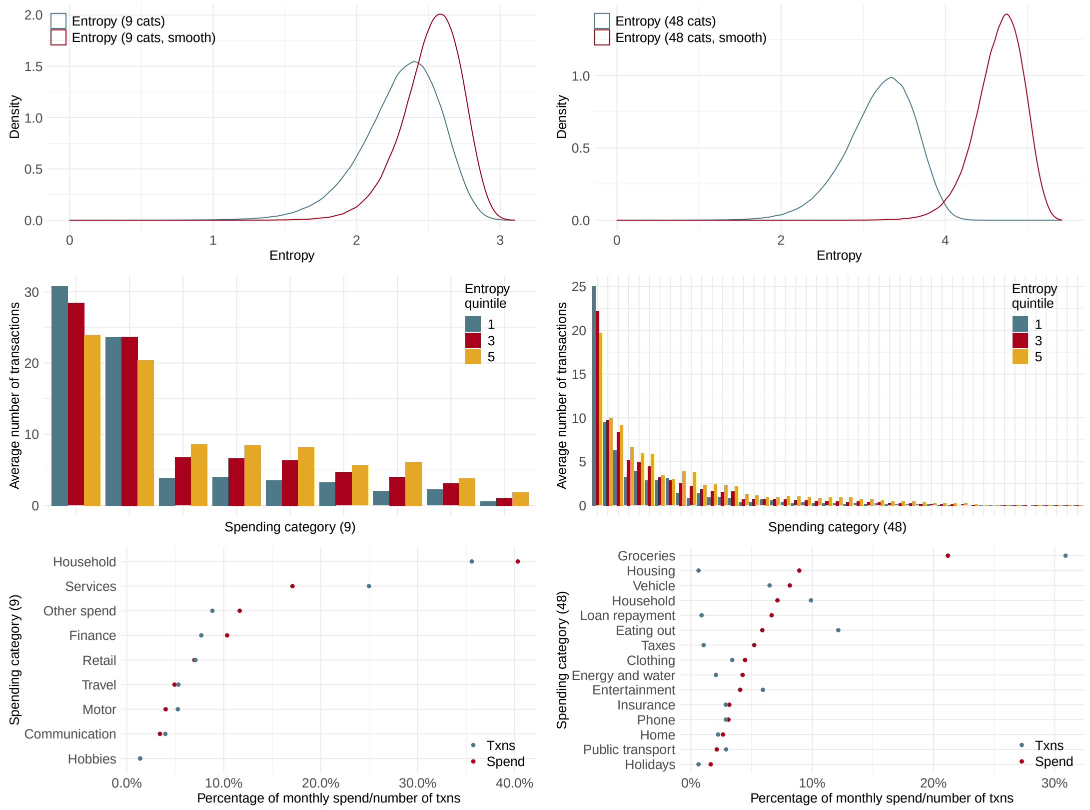


Notes: The top left and top right panels show the distribution of disposable income and total spending in 2019, respectively, benchmarked against the 2018/19 wave of the ONS Living Cost and Food Survey (LCFS). The bottom left panel shows the distribution of age, the bottom right panel that of the regions.

As discussed above, our sample is skewed towards men in their 20s, 30s, and 40s, who live in urban areas, mainly the South East of England. Mean monthly spend is £2,000 and the median is £2,200, and the average invididual spends money across 27 different merchants in 16 different categories (of the 48 category classification).

Figure 2 shows the distributions of the unsmoothed and smoothed version of entropy based on 9 and 48 spending categories, in the top row, the associated average spending profiles for individuals in the first, third and fifth quintiles in the middle row, and an overview of the main categories individuals spend on in the bottom row. The entropy distributions are bounded on the right by the maximum entropy value, given by the value for a discrete uniform distribution

Figure 2: Entropy distributions



Notes: The top row shows distributions of unsmoothed and smoothed versions of entropy. The middle row shows associated spending profiles for individuals in the first, third, and fifth quintile of tne entropy distribution. Spending category labels are deliberately omitted to not overcrowd the plots and to focus on the distribution of transaction counts. The bottom row shows the distributions of spend and number of transactions that goes to the top categories. The plot on the left shows all 9 categories, that on the right the top 15 out of 48 categories. The left column shows data for the 9 category classification, the right column data for the 48 categories classification.

with 9 or 48 distinct values,[14] and on the left by zero, the value for cases where an individual makes all spending transactions in a single category. The different effect of smoothing on the two variables is a result of the different number of categories with counts of zero.

The spending profiles in the middle row show that higher entropy individuals make both fewer transactions in high-count categories and more transactions in low-count categories. The spend profile based on the 48 category classification also shows that spend counts follow approximately a power law distributions, with very few of the 48 categories accounting for the large majority of transactions.

The relative ordering of the top spending categories based on the 48 category classification is about as expected, as are the relationship between the number of transactions and spend: there are, for instance, a large number of relatively low value transactions and meals eaten out (which includes snacks and coffee) and a very small number of high value transactions for housing. The proportion of spend that goes towards housing, less than 10%, is clearly an underestimate,

[14]For instance, the upper bound for the 9 category variable is $9 \times \left(\frac{1}{9}\right) log_2 \left(\frac{1}{9}\right) \approx 3.17$.

however, and reflects the challenge and imperfect nature of automatic transaction labelling in our data that we discussed above.

2.6 Estimation

To estimate the relationship between spending entropy and savings behaviour we estimate fixed-effect models of the form:

$$y_{i,t} = \alpha_i + \lambda_t + \beta H_{i,t} + x'_{i,t}\delta + \epsilon_{i,t}, \tag{4}$$

where $y_{i,t}$ is an indicator variable equal to one if individual i made at least one transfers to any of their savings account in year-month period t and zero otherwise, H_{it} is i's spending entropy in year-month period t, $x_{i,t}$ a vector of control variables, α_i an individual fixed effect, λ_t a year-month fixed effect, and $\epsilon_{i,t}$ the error term.

The vector of controls includes month spend, month income, an indicator for whether a user had positive income in a given month, and income variability, calculated as the standard deviation of month income over the previous 12 months.

Note that while we might in principle be worried about reverse causality, since making payments into savings accounts might lead to a non-zero count in an additional spend category and thus change entropy, this is not a concern here. As discussed in Section 2.3 and Section 2.4, we define savings as inflows into savings accounts and define entropy based on the classification of spend transactions on current accounts. If a user pays money from their current into one of their savings account, such a transaction will usually be labelled in their current account as a transfer rather than a spending transaction, and thus not enter the calculation of their entropy score.

3 Spending profiles predict emergency savings

Table 3 shows the effect of entropy on the probability of building emergency savings in a given month. Columns (1)-(3) show results for unsmoothed entropy based on 9 categories, 48 categories, and merchant names, respectively. Columns (4)-(6) show results for smoothed entropy based on the same variables. All models include user and year-month fixed effects, and standard errors are clustered at the user-level. 95% confidence intervals are shown in brackets.[15]

[15] In line with an emerging consensus for how to avoid over-reliance on statistical significance, we deliberately do not show significance stars in our regression tables and instead emphasise the implications of the bounds of the confidence intervals when discussing our results. For two recent discussions, see Imbens (2021) and Romer (2020).

Results for unsmoothed entropy suggest that a one standard-deviation increase in spending entropy is associated with an increase in the probability of a user making at least one transfer into their savings accounts of between 1.1 and 2.3 percentage points – an effect equal to an increase in monthly income of between £1000 and £2000. Combined, the results in the three columns present very strong evidence that entropy captures something about users's spending distribution that is related to their likelihood for making payments into their savings accounts. Furthermore, entropy variables defined based on a larger number of unique categories, that allow for a more precise segmentation of spending behaviour, capture features of spending behaviour that are as strongly or more strongly related to savings behaviour as monthly income.

The results for smoothed entropy are similar but a tend to be smaller in magnitude, and – together – also provide strong evidence that smoothed entropy captures features of spending behaviour that is related to savings behaviour in a statistically significant and economically relevant way. The main difference to the results for unsmoothed entropy is the reversal of the sign of the coefficients: across all three measures, an increase in entropy is estimated to be associated with a decrease in the likelihood of any savings transactions. The effect of the measure based on 9 categories is not meaningfully different from zero, but the estimates for the measures based on 48 categories and merchants are almost identical and suggest that the magnitude of the effect is between 1.2 and 2.1 percentage points – equalling the effect of a reduction of monthly income of between £1000 and £2000.

Table 3: Effect of entropy on P(savings transactions)

	(1)	(2)	(3)	(4)	(5)	(6)
Entropy (9 cats)	0.011					
	[0.005; 0.017]					
Entropy (48 cats)		0.023				
		[0.017; 0.029]				
Entropy (merchant)			0.017			
			[0.011; 0.023]			
Entropy (9 cats, smooth)				-0.004		
				[-0.007; -0.000]		
Entropy (48 cats, smooth)					-0.017	
					[-0.021; -0.012]	
Entropy (merchant, smooth)						-0.017
						[-0.021; -0.013]
Month spend (£'000)	0.010	0.009	0.010	0.010	0.009	0.008
	[0.009; 0.011]	[0.008; 0.011]	[0.008; 0.011]	[0.009; 0.011]	[0.008; 0.010]	[0.006; 0.009]
Month income (£'000)	0.012	0.012	0.012	0.012	0.012	0.012
	[0.011; 0.014]	[0.011; 0.014]	[0.011; 0.014]	[0.011; 0.014]	[0.010; 0.014]	[0.010; 0.013]
Has income in month	0.086	0.085	0.086	0.087	0.086	0.086
	[0.070; 0.102]	[0.069; 0.101]	[0.069; 0.102]	[0.071; 0.103]	[0.070; 0.102]	[0.070; 0.102]
Income variability	0.001	0.001	0.001	0.001	0.001	0.001
	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]
Observations	389,790	389,790	389,790	389,790	389,790	389,790
R^2	0.46824	0.46838	0.46830	0.46820	0.46844	0.46870
Within R^2	0.00621	0.00648	0.00633	0.00615	0.00659	0.00709
User fixed effects	✓	✓	✓	✓	✓	✓
Year-month fixed effects	✓	✓	✓	✓	✓	✓

Notes: Results from estimating Equation 4. The dependent variable in all columns is a dummy variable indicating whether a user made at least one transaction into any of their savings accounts in a given period. Terms in brackets denote the upper and lower ends of a 95% confidence interval. Standard errors are clustered at the user-level.

As discussed in Section 2.6, these results are not a results of reverse causality. While we might think that making a savings transactions might change some or all of the components of entropy discussed in Section 2.4 – the number of unique spending categories with positive frequency count, the standard deviation of these counts, and the total number of spend transactions – and thus change entropy, this is not the case because of the way we define entropy and savings, and the way spending transactions are categorised. We define entropy based on all current account debits that are identified as spends, while we define savings transactions as the sum of all savings accounts credits. If a user transfers money from their current account to their savings account, this will be identified as a savings transaction, but be identified as a transfer on their current account and thus not considered when calculating their entropy score.

The estimates of our control variables are largely as expected, with the exception of monthly spend, which one might have expected to be negatively correlated with savings. Also, it is evident that the strongest predictor among the included controls for whether a user makes any savings transfer is whether they receive any income in that month. Income variability, in contrast, is not correlated with savings behaviour in any economically significant way, suggesting that people with variable incomes do not build savings cushions for periods where they have no income.

Overall, the effect of entropy in spending profiles is statistically and economically significant, and robust across different definitions. In other words, the scores seem to pick up a feature of the spending distribution that is predictive of savings behaviour. The obvious question raised by the results is why smoothing entropy scores flips the direction of the effect of entropy. We address this next.

3.1 Why does smoothing flip the direction of the effect

One way to think about the sign change in Table 3 is to realise that it implies that at least for some individuals, the relative rank of smoothed and unsmoothed entropy must differ considerably – there must be some individuals that have low unsmoothed entropy but high smoothed entropy or some that have high unsmoothed entropy and low smoothed entropy or both. Understanding who those individuals are might thus help us understand the sign flip.

To understand rank differences between unsmoothed and smoothed entropy scores it is useful to rewrite Equation 1 in a way that makes it easy to see its component parts. Remember from Section 2.4 that f_c is the number of transactions made by a user in a given period in spending category c, $\mathcal{C}$ the set of all spending categories, and F the total number of transactions made by a given user in a given period. Additionally, let $\mathcal{C}^+ = \{c : f_c > 0\}$ be the set of all spending categories with positive frequency counts (i.e. with at least one transaction) and $\mathcal{C}^0 = \{c : f_c = 0\}$ the set of all spending categories with a zero frequency count, so that $\mathcal{C} = \mathcal{C}^0 \cup \mathcal{C}^+$. Then, using our definitions of unsmoothed and smoothed probabilities, we can write unsmoothed entropy as

$$H = -\sum_{c \in \mathcal{C}^+} \left(\frac{f_c}{F}\right) log \left(\frac{f_c}{F}\right), \tag{5}$$

and smoothed entropy as:

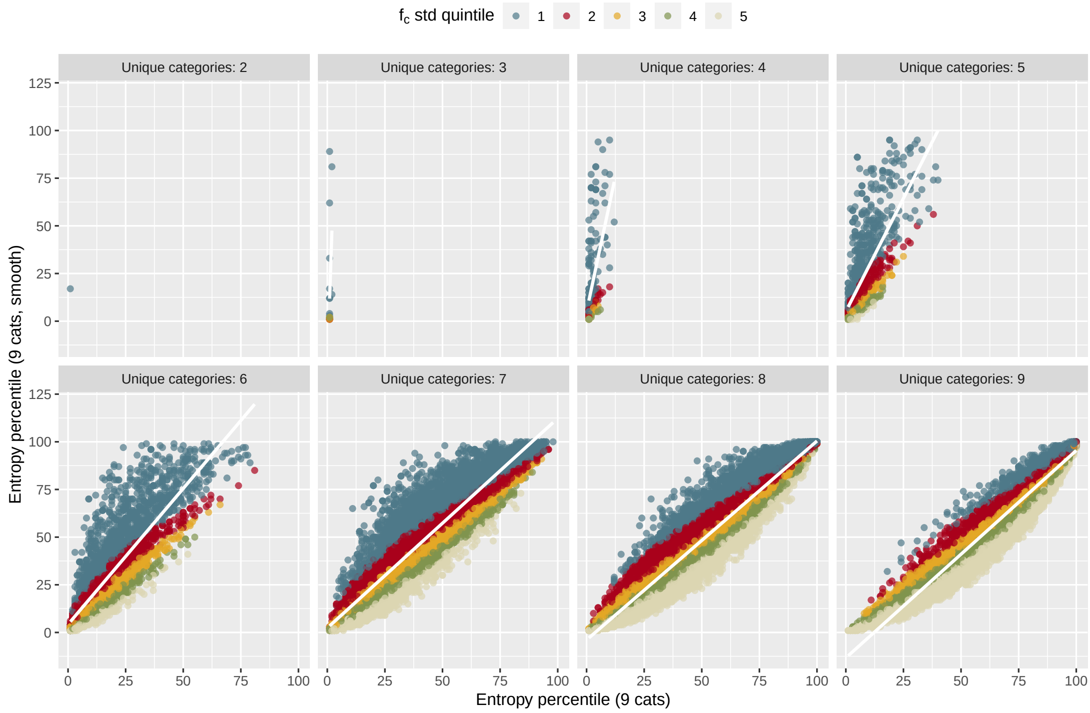
$$H^s = -\sum_{c \in \mathcal{C}^+} \left(\frac{f_c + 1}{F + |\mathcal{C}|}\right) log \left(\frac{f_c + 1}{F + |\mathcal{C}|}\right) - |\mathcal{C}^0| \left(\frac{1}{F + |\mathcal{C}|}\right) log \left(\frac{1}{F + |\mathcal{C}|}\right), \tag{6}$$

where the size of set $\mathcal{C}^0$, $|\mathcal{C}^0|$, is the number of all spending categories in which a user makes no transactions in a certain period. These expressions make clear that, by definition, unsmoothed entropy is a function of frequency counts of categories with positive counts only while smoothed entropy has two parts: the sum over all additively smoothed frequency counts of categories with positive counts, plus the same sum for additively smoothed probabilities of categories with zero counts, which reduces to a constant term that is multiplied by the number of such categories.

The expressions also make transparent the three main components of both types of entropy that are determined by user behaviour. The first is the number of spending categories with a non-zero frequency count, $|\mathcal{C}^+|$, which determines the number of elements summed over in Equation 5, and partitions the categories into either contributing to the sum on the left hand side of Equation 6 or – given that for an exogenously fixed $|\mathcal{C}|$, $|\mathcal{C}^0| = |\mathcal{C} \backslash \mathcal{C}^+|$ – to the constant term on the right hand side. The second component is the variation of the frequency counts, f_c, and the third component is the number of total spend transactions, F. Together, these two elements determine the probabilities associated with a spend transaction occurring in a given category. The number of total spending categories, $|\mathcal{C}|$, also determines smoothed entropy and, implicitly, also unsmoothed entropy since it "scales" the number of categories with a positive frequency count, $|\mathcal{C}^+|$, as a given number of spending transactions are categorised into finer or coarser categories. But it is exogenously given and does not depend on user behaviour.

These components help us reason about rank differences between unsmoothed and smoothed entropy. From equations 5 and 6 we can see that the first part of smoothed entropy that sums over all spending categories with positive frequency counts is very similar to the entire expression of unsmoothed entropy –

Figure 3: Effect of smoothing on entropy



Notes: Percentile ranks of 9-category-based unsmoothed and smoothed entropy separated by the number of categories with positive frequency counts. White reference lines indicate equal percentile ranks; colours, frequency count standard deviation quintiles. Figure is based on a 10% sample of the dataset used for analysis. There are only 8 panels since there are cases where a user has spends in only a single category.

it is that same expression but with additively smoothed probabilities. Hence, all else equal, the higher the number of categories with positive counts, the more smoothed entropy is determined by that first part, and the more similar it will be to unsmoothed entropy. As a result, we would expect to find large (rank) differences between entropy scores among cases with few positive-counts categories. Furthermore, given that unsmoothed entropy is increasing in the number of categories with positive counts, we expect these cases to have relatively low unsmoothed entropy.

Given that we expect high rank difference cases to make spends across a small number of categories, and expect these cases to have low unsmoothed entropy, high rank differences will occur among the subset of those cases that have high smoothed entropy. Remember from Section 2.4 that entropy is higher the more equal the spending category probabilities are. Hence, for a given number of zero-count categories, smoothed entropy will be higher if the (additively smoothed) probabilities of all positive-count categories are close to the (additively smoothed) probabilities of the zero-count categories, which will be the case (i) if there are few overall transactions, such that frequency counts (f_c) are close to zero and (ii) if the counts are similar.

Figure 3 visualises this entire line of reasoning for our 9-category based entropy variable: it shows scatter plots of the percentile ranks of unsmoothed and smoothed entropy with a reference line indicating identical rank, separated by the number of categories with positive frequency counts and coloured based on the quintile of the frequency count standard deviation.[16] First, ignoring the colouring and focusing on the shape of the dots only we can see that, as expected, the relationship between the two entropy measures is tighter the higher the number of non-zero spending categories is. Cases with large entropy rank differences are thus to be found among cases with fewer positive spend categories. Among these, the colouring makes clear that, as expected, the cases with the largest rank differences – those with the furthest vertical difference to the reference line – have low variation in their frequency counts. However, it is also clear that the reverse is not true: there are cases with low count frequency variation that experience little or even a negative rank difference. Hence, while counts variation is some help in identifying cases with high entropy rank differences, it does not do so perfectly.

One reason the relationship is not perfect is that while cases with few total transactions that are evenly spread across a small number of categories will have high smoothed entropy, such will also have high unsmoothed entropy. What we are looking for, more precisely, are cases where this pattern holds broadly, but that also have a small number of frequency counts that dominate the others but are still relatively close to zero. This would lead to low unsmoothed entropy (because of the dominant counts) but high smoothed entropy (because all counts are still relatively close to zero, so that the additively smoothed distribution resembles a uniform distribution). Further research will be necessary to more fully classify such cases, and to inquire what it is about their behaviour that might be related to savings behaviour.

3.2 Is entropy informative beyond its component parts?

Another question that arises once we think of entropy as comprising three main component parts is whether its non-linear nature captures anything about the relationship between spending and saving behaviour that is not captured already by the three simpler components.

[16] Colouring based on the quintile of the total number of transactions leads to a very similar plot and is thus not shown.

Table 4: Controlling for components

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Entropy (9 cats)	0.011	0.004						
	[0.005; 0.017]	[-0.003; 0.011]						
Entropy (9 cats, smooth)			-0.004	-0.004				
			[-0.007; -0.000]	[-0.009; 0.001]				
Entropy (48 cats)					0.023	0.020		
					[0.017; 0.029]	[0.005; 0.034]		
Entropy (48 cats, smooth)							-0.017	-0.021
							[-0.021; -0.012]	[-0.031; -0.010]
Unique categories ($\|\mathcal{C}^+\|$)		0.004		0.004		0.002		0.004
		[0.002; 0.005]		[0.003; 0.005]		[0.001; 0.004]		[0.003; 0.005]
Category counts std.		-0.001		-0.003		0.004		-0.010
		[-0.005; 0.003]		[-0.008; 0.001]		[-0.002; 0.010]		[-0.016; -0.005]
Number of spend txns (F)		0.001		0.001		0.000		0.001
		[0.000; 0.001]		[0.000; 0.001]		[-0.000; 0.001]		[0.000; 0.001]
Month spend	0.010	0.006	0.010	0.006	0.009	0.006	0.009	0.006
	[0.009; 0.011]	[0.005; 0.007]	[0.009; 0.011]	[0.005; 0.007]	[0.008; 0.011]	[0.005; 0.007]	[0.008; 0.010]	[0.005; 0.007]
Month income	0.012	0.011	0.012	0.011	0.012	0.011	0.012	0.011
	[0.011; 0.014]	[0.009; 0.013]	[0.011; 0.014]	[0.009; 0.013]	[0.011; 0.014]	[0.009; 0.013]	[0.010; 0.014]	[0.009; 0.013]
Has income in month	0.086	0.082	0.087	0.082	0.085	0.081	0.086	0.081
	[0.070; 0.102]	[0.066; 0.098]	[0.071; 0.103]	[0.066; 0.098]	[0.069; 0.101]	[0.065; 0.097]	[0.070; 0.102]	[0.065; 0.097]
Income variability	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]
Observations	389,790	389,790	389,790	389,790	389,790	389,790	389,790	389,790
R^2	0.46824	0.46908	0.46820	0.46908	0.46838	0.46910	0.46844	0.46914
Within R^2	0.00621	0.00778	0.00615	0.00779	0.00648	0.00783	0.00659	0.00790
User fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Year-month fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Terms in brackets denote the upper and lower ends of a 95% confidence interval. Standard errors are clustered at the user-level.

We can test this by checking whether the relationship between spending entropy and the probability of making a savings transaction remains economically and statistically significant once we control for the three components. Columns (1) and (3) in Table 3 replicate the results for the 48-category-based unsmoothed and smoothed entropy measures presented in Table 3 for reference. In columns (2) and (4) we additionally control for the three entropy components. Including these components has some effect: the coefficients change slightly – decreasing in absolute magnitude in the case of unsmoothed entropy, increasing in the case of smoothed entropy – while the width of the confidence intervals about double in both cases, reflecting the strong collinearity amount the component and entropy. However, both coefficients remain statistically significant and their confidence intervals cover values that are also economically significant. Hence, the results make clear that the results in Table 3 cannot be attributed simply to the effect of one or more of entropy's simple components.

4 Discussion

Ignore this section for now. Below are just notes. There are a number of alternative ways to characterise spend profiles. We could calculate profiles based on the distribution of transaction values rather than counts. We could also calculate profiles based on inter-temporal rather than intra-temporal distributions, focusing on consistency of purchasing behaviour over time rather than on predictability at any given time (Krumme et al. 2013). Further, we could focus on time-based rather than category-based measures, focusing, for instance, on whether purchases of the same type tend to occur on the same day of the week (Guidotti et al. 2015). Finally, one could also create composite measures based on principal component analysis, an approach used in Eagle et al. (2010). We leave these extensions for future research.

References

Benhabib, Jess and Alberto Bisin (2005). "Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions". In: *Games and economic Behavior* 52.2, pp. 460–492.

Bourquin, Pascale, Isaac Delestre, Robert Joyce, Imran Rasul, and Tom Walters (2020). "The effects of coronavirus on household finances and financial distress". In: *IFS Briefing Note BN298*.

CAN, Commonwealth Bank of Australia (2019). "Improving the Financial Wellbeing of Australians". Tech. rep. URL: https://www.commbank.com.au/content/dam/caas/newsroom/docs/CWM0375_Financial%20Wellbeing%20Report_v4.pdf.

Carlin, Bruce, Arna Olafsson, and Michaela Pagel (2019). "Generational Differences in Managing Personal Finances". In: *AEA Papers and Proceedings*. Vol. 109, pp. 54–59.

CFPB, Consumer Financial Protection Bureau (2017). "Financial Well-being in America". Tech. rep. URL: https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-america/.

Chater, Nick and George Loewenstein (2022). "The i-frame and the s-frame: How focusing on the individual-level solutions has led behavioral public policy astray". In: *Available at SSRN 4046264*.

Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick (2009). "Reinforcement learning and savings behavior". In: *The Journal of finance* 64.6, pp. 2515–2534.

Colby, Helen and Gretchen B Chapman (2013). "Savings, subgoals, and reference points". In:

Eagle, Nathan, Michael Macy, and Rob Claxton (2010). "Network diversity and economic development". In: *Science* 328.5981, pp. 1029–1031.

Fudenberg, Drew and David K Levine (2006). "A dual-self model of impulse control". In: *American economic review* 96.5, pp. 1449–1476.

Gelman, Michael, Shachar Kariv, Matthew D Shapiro, Dan Silverman, and Steven Tadelis (2014). "Harnessing naturally occurring data to measure the response of spending to income". In: *Science* 345.6193, pp. 212–215.

Governors of the Federal Reserve System, Board of (2022). "Economic Well-Being of U.S. Households in 2021". Tech. rep.

Guidotti, Riccardo, Michele Coscia, Dino Pedreschi, and Diego Pennacchioli (2015). "Behavioral entropy and profitability in retail". In: *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, pp. 1–10.

Gul, Faruk and Wolfgang Pesendorfer (2001). “Temptation and self-control”. In: *Econometrica* 69.6, pp. 1403–1435.

Hacıoğlu-Hoke, Sinem, Diego R Känzig, and Paolo Surico (2021). “The distributional impact of the pandemic”. In: *European Economic Review* 134, p. 103680.

Hirsh, Jacob B, Raymond A Mar, and Jordan B Peterson (2012). “Psychological entropy: a framework for understanding uncertainty-related anxiety.” In: *Psychological review* 119.2, p. 304.

Imbens, Guido W (2021). “Statistical significance, p-values, and the reporting of uncertainty”. In: *Journal of Economic Perspectives* 35.3, pp. 157–74.

Krumme, Coco, Alejandro Llorente, Manuel Cebrian, Alex Pentland, and Esteban Moro (2013). “The predictability of consumer visitation patterns”. In: *Scientific reports* 3.1, pp. 1–5.

Laibson, David (1997). “Golden eggs and hyperbolic discounting”. In: *The Quarterly Journal of Economics* 112.2, pp. 443–478.

Laibson, David and Keith Marzilli-Ericson (2019). “Intertemporal choice”. In: *Handbook of Behavioral Economics* 2.

Loewenstein, George and Ted O’Donoghue (2004). “Animal spirits: Affective and deliberative processes in economic behavior”. In: *Available at SSRN 539843*.

Lunt, Peter K and Sonia M Livingstone (1991). “Psychological, social and economic determinants of saving: Comparing recurrent and total savings”. In: *Journal of economic Psychology* 12.4, pp. 621–641.

Madrian, Brigitte C and Dennis F Shea (2001). “The power of suggestion: Inertia in 401 (k) participation and savings behavior”. In: *The Quarterly journal of economics* 116.4, pp. 1149–1187.

Mittal, Chiraag, Vladas Griskevicius, Jeffry A Simpson, Sooyeon Sung, and Ethan S Young (2015). “Cognitive adaptations to stressful environments: When childhood adversity enhances adult executive function.” In: *Journal of personality and social psychology* 109.4, p. 604.

MPS, Money and Pension Service (2018). “Building the Financial Capability of UK Adults”. Tech. rep. URL: https://moneyandpensionsservice.org.uk/2019/02/06/adult-financial-capability-building-the-financial-capability-of-uk-adults-survey/.

Muggleton, Naomi K, Edika G Quispe-Torreblanca, David Leake, John Gathergood, and Neil Stewart (2020). “Evidence from mass-transactional data that chaotic spending behaviour precedes consumer financial distress”. Tech. rep. DOI: 10.31234/osf.io/qabgm. URL: psyarxiv.com/qabgm.

Oaten, Megan and Ken Cheng (2007). “Improvements in self-control from financial monitoring”. In: *Journal of Economic Psychology* 28.4, pp. 487–501.

Philipps, Jo, Annick Kuipers, and Will Sandbrook (2021). "Supporting emergency savings: early learnings of the employee experience of workplace sidecar savings". Tech. rep.

Romer, David (2020). "In praise of confidence intervals". In: *AEA Papers and Proceedings*. Vol. 110, pp. 55–60.

Sabat, Jorge and Emily Gallagher (2019). "Rules of thumb in household savings decisions: Estimation using threshold regression". In: *Available at SSRN 3455696*.

Shannon, Claude Elwood (1948). "A mathematical theory of communication". In: *The Bell system technical journal* 27.3, pp. 379–423.

Skatova, Anya, Neil Stewart, Edward Flavahan, and James Goulding (2019). "Those Whose Calorie Consumption Varies Most Eat Most". In:

Thaler, Richard H and Hersh M Shefrin (1981). "An economic theory of self-control". In: *Journal of political Economy* 89.2, pp. 392–406.

Vernon-Feagans, Lynne, Michael Willoughby, and Patricia Garrett-Peters (2016). "Predictors of behavioral regulation in kindergarten: Household chaos, parenting, and early executive functions." In: *Developmental psychology* 52.3, p. 430.

A Interpreting entropy

To see how we can interpret entropy as the predictability of a user's spendin that we discussed above behaviour, it is useful to have a more complete understanding of Equation 1. The building blocks of entropy is the information content of a single event. The key intuition Shannon (1948) aimed to capture was that learning of the occurrence of a low-probability event is more informative than learning of the occurrence of a high-probability event. The information of an event $I(E)$ is thus inversely proportional to is probability $p(E)$. One way to capture this would be to define the information of event E as $I(E) = \frac{1}{p(E)}$. Yet this implied that an event that is certain to occur had information 1, when it would make sense to have information 0. To remedy this (and also satisfy additional desireable characteristics of an information function), Shannon proposed using the log of the expression. Hence, the information of event E, often called *Shannon information*, *self-information*, or just *information*, is defined as:

$$I(E) = log\left(\frac{1}{p(E)}\right) = -log(p(E)). \tag{7}$$

The choice of the base for the logarithm varies by application and determines the units: base 2 means that information is expressed in bits; the natural logarithm, another popular choice, expresses information in *nats*.

Entropy, often called *Information entropy*, *Shannon entropy*, or just *entropy*, is the information of a random variable, X, and captures the expected amount of information of an event drawn at random from the probability distribution of the random variable. It is calcualted as:

$$H(X) = -\sum_x p(x) \times log(p(x)) = \sum_x p(x)I(x) = \mathbb{E}I(x). \tag{8}$$

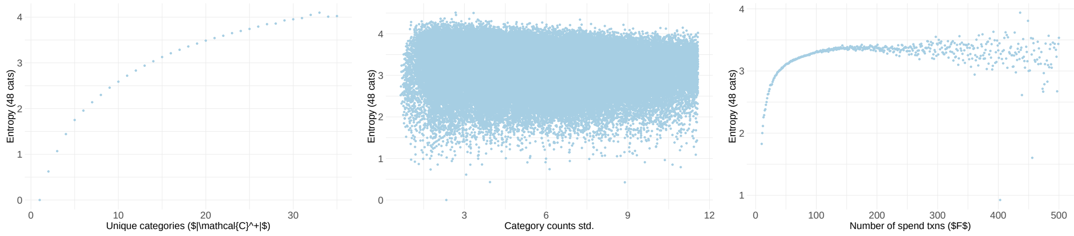
For a single event, the key intution was that the less likely an event, the more information is conveyed when it occurs. The related idea for distributions is similar: the less skewed a distribution of a random variable, the less certain the realised value of a single draw from the distribution, the higher is entropy - the maximum entropy distribution is the uniform distribution.

B Additional results

B.1 Entropy components

Figure 4 shows the empirical relationship with our 48-categories-based unsmoothed entropy variable and these three components.[17] We can see that for the values we observe in the dataset, entropy increases monotonically in the number of unique spending categories with positive frequency counts, has no clear relationship with the standard deviation of those counts, and increases in the number of total spending transactions up to about 175 transaction, before being increasingly determined by other elements thereafter.

Figure 4: Correlation of entropy with its components

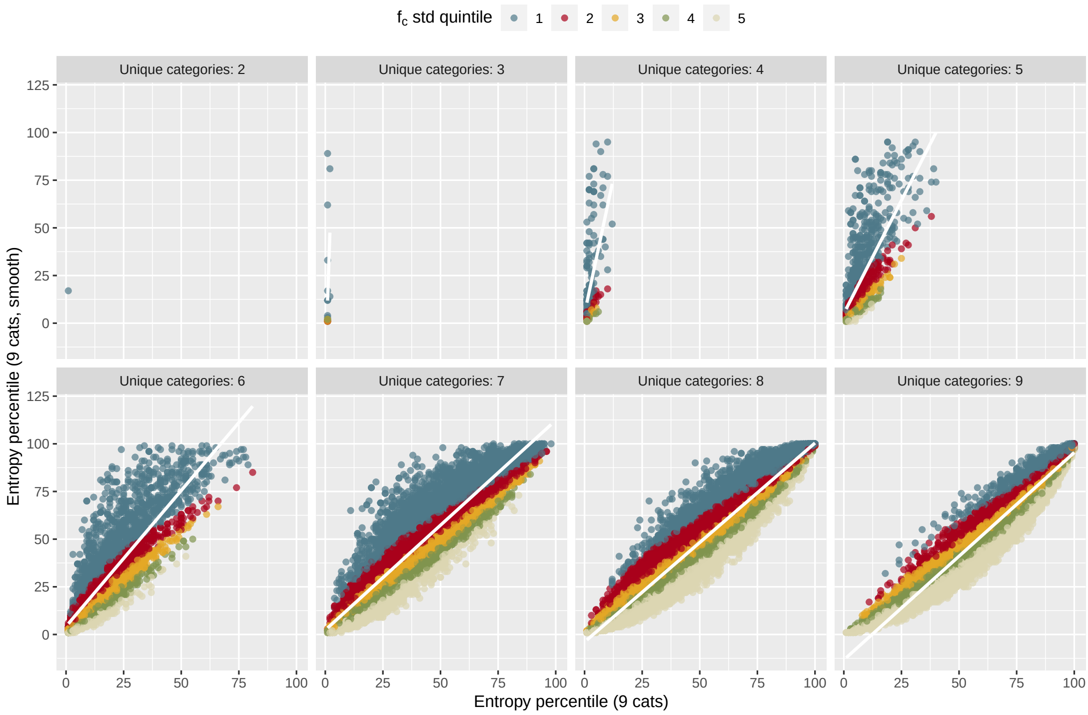


Notes: Correlation of 48-categories-based unsmoothed entropy with its three main components: the number of unique spending categories with positive frequency counts (left), the standard deviation of those frequency counts (middle), and the number of total spend transactions (left).

B.2 Effect of smoothing on entropy (by number of spend quintile)

[17]To highlight the main features of the relationships we have trimmed the component values at the 95th percentile.

Figure 5: Effect of smoothing on entropy



Notes: Percentile ranks of 9-category-based unsmoothed and smoothed entropy separated by the number of categories with positive frequency counts. White reference lines indicate equal percentile ranks. Colours indicate the quintile of the total number of spending transactions.