MS&E 231 Project Report:
Analysis of Randomized Control Trial on a Subset of SNAP Recipients

Oseas Ayerdi
Chris Bechler
Wendy De La Rosa
Alisa Yu

Abstract

Evidence from a randomized controlled trial shows that providing Supplemental Nutrition Assistance Program (SNAP) recipients with a weekly budget (vs. no weekly budget) does not alter their spending patterns. There were no differences in the rate of consumption throughout the month, number of transactions, size of transactions, and number of unplanned purchases. However, we do find supporting evidence that providing SNAP recipients with their current SNAP balance slows down their rate of consumption.

Introduction and Background

The Supplemental Nutrition Assistance Program (SNAP), formerly known as the Food Stamp Program, is the United States federally-funded food assistance program for low-income individuals and families. Approximately one in seven Americans participate in SNAP, making SNAP one of the most important government programs in the United States (USDA 2017). In 2016, 44.2 million people relied on SNAP benefits to supplement their food consumption (USDA 2017).

Although the deposits are supposed to last an entire month, overwhelmingly, the typical SNAP household tends to run out well before the month is over (Wilde and Ranney 2000; Shapiro 2005; Mastroubini and Weinberg 2009). According to a 2009 study conducted by the USDA, SNAP recipients spend 78.1% of their benefits in the first 14 days.

This accelerated consumption pattern is a cause for concern for many researchers. Many SNAP recipients go hungry at the end of the month. Studies have shown that caloric intake decreases throughout the month by 10 - 15%, contradicting claims that bulk buying is offsetting the front-loaded spending (Shapiro 2005). In addition, children are found to have lower test scores when their families' SNAP benefits run out (Cotti et al 2017). The authors conclude that their "results provide evidence that households do not sufficiently smooth consumption and that this has measurable effects on student performance."

Department of Management Science and Engineering,
Stanford University, 2017

Researchers have pointed to multiple reasons that could be causing this accelerated and non-smooth consumption pattern. First, the SNAP benefits are not enough. SNAP benefits are determined based on the Department of Agricultures Thrifty Food Plan (TFP), a diet plan intended to provide adequate nutrition at a minimal cost. The TFP assumes that families can prepare all of their meals at home, averaging just $1.40 per person per meal. Aside from SNAP benefits not being enough, other researchers have argued that the way that the benefits are dispersed fundamentally change the way families spend their SNAP.

Currently, SNAP benefits are usually deposited on a monthly basis, averaging $125 USD for an individual and $254 USD for a household. Deposit dates vary by person, based on their state and their SNAP case number. As of June 2004, all states have transitioned to loading the benefits onto an Electronic Benefits Transfer (EBT) card, similar to a debit card. Unlike a debit card though, the SNAP benefits on the EBT card cannot be redeemed for cash or used on non-food items like hygiene products or medicines.

This once a month lump-sum payment leads to a "windfall" or "payday" effect. In many other contexts, not just with SNAP, recipients of a large lump sum tend to make larger or more frequent discretionary purchases on "payday." We find evidence for this pattern across Social Security recipients in the U.S., paycheck recipients in the U.K., and pensioners in Japan (Stephens 2003, 2006; Stephens and Unayama 2011). Interestingly in one study, White and Basu (2015) examine what happens to the consumption patterns of cash assistance beneficiaries in Peru when the government changes their deposit schedule from twice a month to once a month. White and Basu found that this change to once a month deposits led to an increase in alcohol and sweet expenditures by 5580% and sweets by 1040%, respectively.

This lead us to our research question, could we help SNAP families smooth out their consumption? We couldn't change the way states disburse their income, but could providing SNAP recipients with a weekly recommended budget potentially change they way SNAP recipients perceived their deposit cycle and help smooth out their consumption?

Methodology

Data

We partnered with Propel, a startup who developed FRESHEBT, a smart-phone application that allows SNAP recipients to check their SNAP balance, see past transactions, locate stores that accept SNAP, and build a shopping list. Prior to FRESHEBT, SNAP recipients could only check their balance by making a transaction and seeing their balance on the transaction receipt or by calling a number on the back of their EBT card. FRESHEBT has over one million downloads on the Google Play store. We were able to ssh to their database.

Randomization

For our randomized controlled trial, starting on June 8, 2017, every new user was randomly placed either in the control group or the treatment group in a 5%/95% split, respectively. The treatment group are users that have a recommended weekly budget as part of the application's user interface. The control group has no weekly recommended budget.

Since the start of the experiment, over 464k users have been allocated to either of the aforementioned groups across the US. Of these, our analysis is restricted to California residents (the state with the most users our sample), with a time span between the start of the experiment, up until November 28, 2017; the variables contained in the initial, tabulated data include:

- *Account_id*: unique ID number identifying each user of the FRESHEBT app
- *transaction_type*: deposit or purchase
- *transaction_date*: date of a specific transaction
- *amount_cents*: total number of cents of the transaction, negative for purchases and positive for deposits
- *experiment_condition*: either 5A or 5B depending on whether the user was allocated to the control or treatment group
- *next_deposit_date*: date of next deposit

In addition to those listed above, the following variables were also part of the dataset: *user_download date, state, last_balance_amount, last_balance_check, transaction_id, experiment_joined_date, transaction_month*.

The total amount of transactions included are over 2.5 million for $55,515$ users.

Data Cleaning

As a first step, all transactions before the first deposit following a user's entry into the experiment were removed, as we are interested in capturing the spending behavior between deposits. Additionally, users that were both in the control and treatment groups were excluded. This last item happened because anytime a user changed phones, or reinstalled the application, the allocation to a group within the experiment was redone. Finally, because we were interested in understanding in how large lump sums affect consumption over time, we decided to move any deposits lower than 10 USD as they were unlikely to affect spending behavior, given their negligible size compared to the initial deposit (usually ranging from 192 to 650 USD). In a brief analysis of all deposit amounts, we found that 10 USD was the second most frequently occurring deposit, account for 10% of all deposits. (Note: 192 USD was the most frequently occurring deposit, accounting for 10.4% of all deposits). We later learned that the 10 USD deposit was a Work Incentive Nutritional Supplement Benefit that was provided to working families. Because these deposits of 10 USD were so frequently occurring, yet was such a small supplement, compared to their monthly deposit, we thought it was a natural cutoff point to use. At the start of our analyses, we found that only 50% of deposits occurred in regular monthly intervals. However, after excluding deposits of 10 USD or less, we found that about 80% of remaining deposits did occur in regular monthly intervals. Thus, we decided to proceed with our exclusion criteria. This removed 17835 of the 175162 deposits in our dataset (10.2%). With these adjustments, the number of transactions were still in the order of 2.5 million, but the number of users in the analysis dropped to $55,093$.

Addition of Variables

To aid the analysis, some additional columns were added to the original dataset, these being:

- *days_after_deposit*: starting from 0 on the date of the deposit
- *current_balance*: the amount of cents remaining from the initial month's deposit following all purchases up to the current row's
- *perc_of_dep*: the fraction of the *current_balance* divided by the month's deposit.

We were interested in analyzing the behavior of users that received one deposit at the beginning of the month, rather than multiple ones; as such, we removed any instance where the days between deposits was lower than 28 days. Additionally, the purchases within one day were aggregated to have total spent per day. For days with a deposit, there were two entries, one for the total deposit amount and another one for the total spent (if any). With these changes, the transactions available for review were 1.8 million for $48,893$ users.

Since some users won't make purchases on a daily basis, a script was created to carry-over the *current_balance* and *per_of_dep* for dates in between days with purchases. Adding these rows increased the total number to 3.2 million (from 1.8 million); the number of users remained unchanged.

Aggregation

With this final dataset, the information can be summarized by group. More specifically, we want the average balance by

day from day of deposit, 0, to day before next deposit (up to 30) for both the treatment and control group.

Other Comparisons

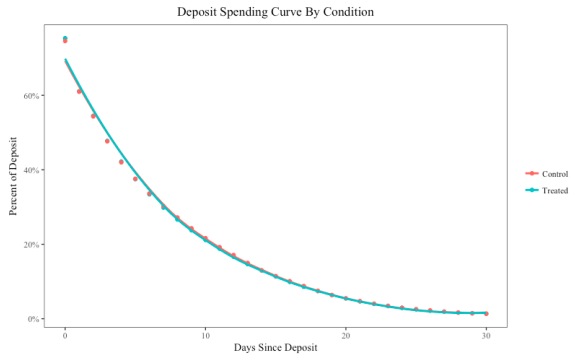
In addition to exploring potential differences in the groups' spending curves, we also wanted to review if there were any statistical differences in four items:

- *Number of transactions under* 6 *USD*: this serves as a proxy for unplanned purchases
- *transactions under* 6 *USD as a % of total transactions*: similar as above but normalizing by the total transactions in the month
- *Number of purchases in month*: a proxy for bulk purchases, which in turn, serves as an indication of proactive purchasing behavior
- *Transaction size in USD*: an alternative proxy for bulk purchases

Results

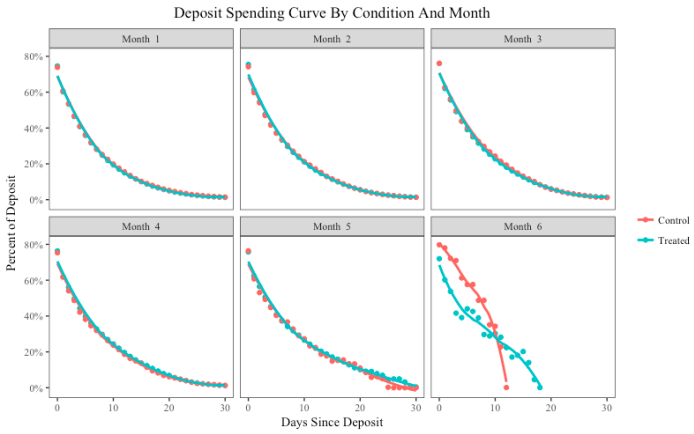
Spending Curves

On average, the control group's balance on day 0 is 74.58% of the month's deposit; in comparison, the treatment groups balance is 75.32%. By day 10 both groups have spent nearly 80% of the month's deposit, with the control group's balance at 21.62% and the treatment group at 21.17%.



Spending Curves, by month

To review potential changes in spending behavior of the treatment group through time, the data was aggregated by treatment group and month, with a bit over 5 months available for analysis; the reduction in available data for month 6 explains its unique shape vs. the rest of the months (i.e. the entries for month 6 are in the order of 10s vs. in the 1,000s for months 1-5). In month 1 the treatment and control group % balance of the deposit was 73.84% and 74.56%, respectively; by month 5 these figures were 75.75% (treatment) and 76.39% (control). In both groups, there is a drop in percent spent between months 1 and month 5.



To review if this difference is significant, we compare all the balances for day 0 between months 1 and 5 for users in both groups. The results are summarized in the table below for the control group:

Month	$n_{observations}$	μ_{month}	σ_{month}
1	1,729	73.84%	0.2896%
5	116	76.39%	0.2688%

The resulting t-statistic is -0.9875 (*p-value* = 0.3252), with a 95% confidence interval containing 0, meaning that we can't conclusively point to this observed difference as significant.

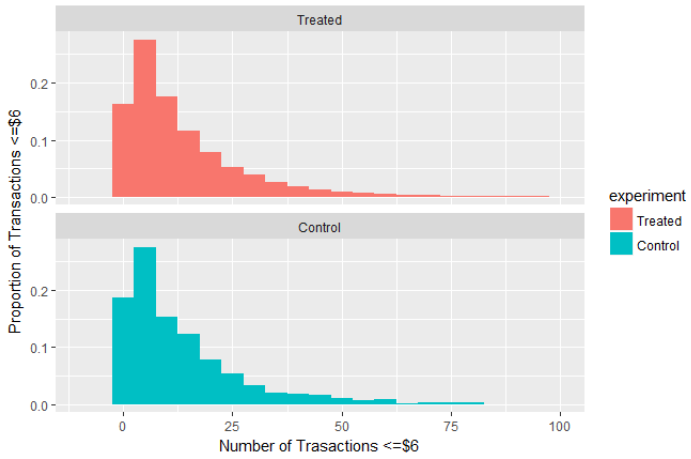
We carry out the same analysis for the treatment group, with the results summarized below:

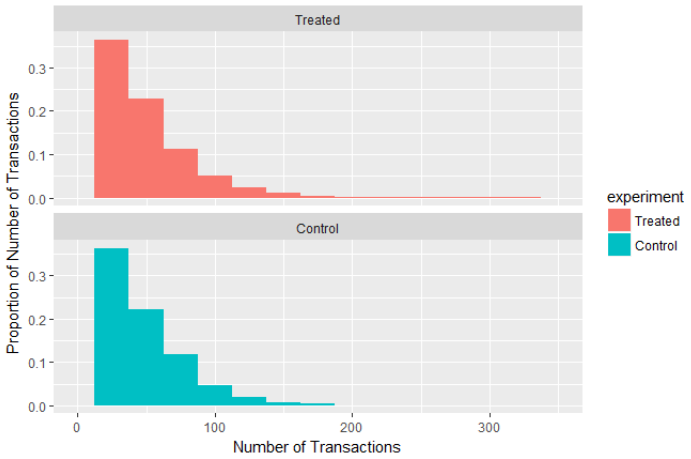
Month	$n_{observations}$	μ_{month}	σ_{month}
1	45,561	74.56%	0.2893%
5	3,616	75.75%	0.2859%

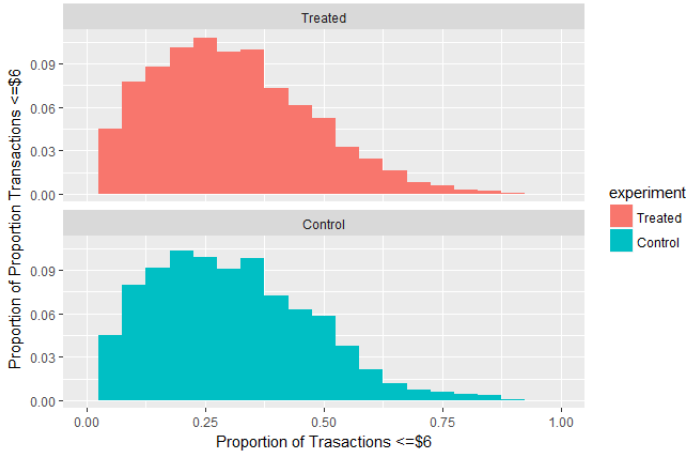
The resulting t-statistic is -2.4016 (*p-value* = 0.0164), with a 95% confidence interval between -0.0216 and -0.0022, meaning that we can reject the null that the differences are due to chance, and gain more confidence in the hypothesis that the users are pacing their spending as a result of having the 'weekly balance' feature in their app.

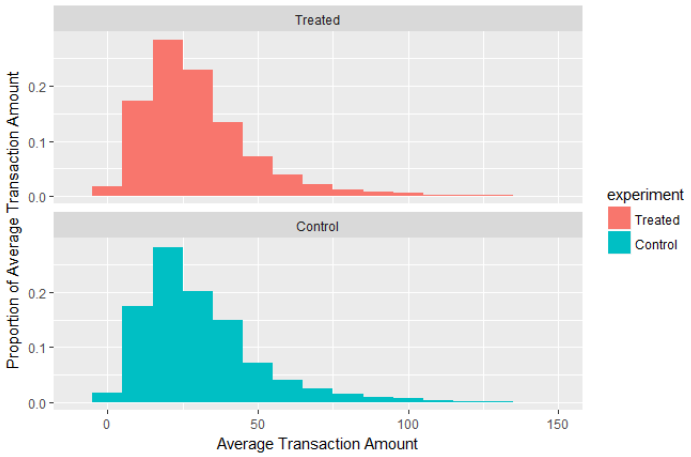
Unplanned Transactions

Small-value purchases may be an indicator for reactive buying behavior, and as such, we wanted to see if there was a difference in the number of purchases below a certain threshold for both groups; we chose this to be 6 USD because the median transaction at 7/11 (which we deemed to be unnecessary purchases) was 5.88 USD. The resulting pairs of histograms for the number of purchases below 6 USD (top) and as a % of total purchases (bottom) are shown below:









We run two separate t-tests to review if the slight differences are statistically significant; they are not, as the figures in the table below show:

Test	$\mu_{control}$	$\mu_{treatment}$	t-statistic	p-value
Number	14.0	14.3	-0.36	0.348
Percent	28.6%	28.9%	-0.00	0.507

Bulk Purchases

On the other side of the spectrum, we want to view if there is a difference in the degree of how proactive the groups' respective purchasing behaviors are. To gauge this, we review two items, as discussed before: *Number of purchases in month*, and *Transaction size*. We expect the more proactive group to make less purchases per month, and as a consequence, spend more per purchase. These results are summarized in the following two pairs of histograms below, with the first looking into the distribution of the number of transactions per group; the second histogram looks at the distribution of the amount (in USD) spent per transaction:

As we did for the unplanned transaction section, we check for significance (if any) between the differences observed; the results of the two t-tests are summarized below:

Test	$\mu_{control}$	$\mu_{treatment}$	t-statistic	p-value
Number	39.4	40.8	-1.43	0.063
Size	31.55 USD	31.00 USD	0.55	0.296

Though there is no sign of significant difference in the size of the transaction, the results do point out to a significant difference in the number of purchases, with the control group having fewer of these.

Discussion

Limitations

There are a couple of limitations in our analysis. To begin with, we set out to determine if there was a difference in the spending behavior between control and treatment group; however, we did put a framework together to determine what purchase pattern would constitute a better one vs. an alternative. For example, some people may decide to purchase in bulk at the beginning of the month (or at some point during the month), leading to a low balance in the start, but an overall lower total-spent for the month. Others, on the other

hand, may space evenly their purchases across the month following the initial deposit, as to take advantage of specific deals that might appear in their nearest grocery store, etc. What constitutes an optimal spending strategy is beyond the scope of the paper, but would've been a beneficial thing to have in order to look for these patterns in the data. Additionally, we don't control for any other sources of revenue for the users. For some, the deposits observed in the dataset may be a significant portion of their disposable income; for others, it could be purely supplemental; this will have a significant effect in the spending behavior we observe. Furthermore, our proxies may not measure what we are looking for, in particular the 6 USD threshold to represent 'unplanned' transactions. This figure may hold for some areas of California, but for others, with higher costs of living, any unplanned purchase could easily land above this threshold and be out of our radar.

Future Steps

In order to tackle some of the limitations outlined above, future analysis would need to consider several items. First, total wages and other sources of income for the users in the analysis; this would give us a better understanding of how significant the deposits in the study are for a specific user. Secondly, instead of focusing on an absolute threshold to define an 'unplanned' purchase, we could look into the amount of the purchase as a % of the initial deposit amount, say 5% or lower, this would allow us to control for different amounts of deposits to users. In addition, we could also incorporate the venue in which the purchase was made. If we see stores like *Sam's Club* or *Costco*, irrespective of the dollar amount, we should be fairly certain that this is a planned purchase, as oppose to one done in a gas stations or a *7-11*. Thus, we would need to look further into the available data to come up with a better proxy for 'unplanned transactions.'

Beyond addressing the limitations of the current study, future work should increase this report's scope. As discussed earlier, the focus was on California residents, but information is readily available for the rest of the US, and this should be looked into. In addition to increasing the scope geographically, future work should also look into longer timescales; this report already has some results hinting at a change for the treatment group over time, and this should be looked into with more months in the data, in order to ascertain if the observed trend continues.

Conclusion

We set out to find differences in purchasing behavior between two groups of users of the app FRESHEBT. The treatment group had a slight modification in its user interface that prompted users with a recommended weekly budget, whereas the control didn't. We found that there were few instances where both groups differed, at least in the items we investigated, these being: i) balance of their initial deposit throughout the month, ii) number of purchases under 6 USD, iii) these same purchases but as a % of their total transactions during the month, iv) total number of purchases, and v) transaction size (in USD) of purchases. Aside from the difference observed in the number of purchases (the control group seems to have a fewer amount of them), we found one interesting shift in behavior in the treatment group: between months 1 and 5 of being exposed to the app, their consumption amount on the first day of the month—as a % of their deposit—was significantly lower in month 5, than in month 1, suggesting that having the monthly balance may lead them to pace their purchasing. However, this is far from conclusive, and we have listed some additional work that would need to be looked into in order to make more definite statements.

References

Cotti, C. D., Gordanier, J., & Ozturk, O. D. (2017). When Does it Count? The Timing of Food Stamp Receipt and Educational Performance.

Mastrobuoni, G., & Weinberg, M. (2009). Heterogeneity in intra-monthly consumption patterns, self-control, and savings at retirement. *American Economic Journal: Economic Policy*, 1(2), 163-189.

Shapiro, J. M. (2005). Is there a daily discount rate? Evidence from the food stamp nutrition cycle. *Journal of public Economics*, 89(2), 303-325.

Stephens Jr, M. (2003). " 3rd of tha Month": Do Social Security Recipients Smooth Consumption Between Checks?. *American Economic Review*, 93(1), 406-422.

Stephens, M. (2006). Paycheque receipt and the timing of consumption. *The Economic Journal*, 116(513), 680-701.

Stephens, J., & Unayama, T. (2011). The consumption response to seasonal income: Evidence from Japanese public pension benefits. A*merican Economic Journal: Applied Economics*, 3(4), 86-118.

USDA (2007). *Characteristics of Supplemental Nutrition Assistance Program Households: Fiscal Year 2016*.

White, J. S., & Basu, S. (2016). Does the benefits schedule of cash assistance programs affect the purchase of temptation goods? Evidence from Peru. *Journal of health economics*, 46, 70-89.

Wilde, P. E., & Ranney, C. K. (2000). The Monthly Food Stamp Cycle: Shooping Frequency and Food Intake Decisions in an Endogenous Switching Regression Framework. *American Journal of Agricultural Economics*, 82(1), 200-213.