

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Harnessing naturally occurring data to measure the response of spending to income

### Permalink

<https://escholarship.org/uc/item/2mh786jw>

### Journal

Science, 345(6193)

### ISSN

0036-8075

### Authors

Gelman, Michael  
Kariv, Shachar  
Shapiro, Matthew D  
et al.

### Publication Date

2014-07-11

### DOI

10.1126/science.1247727

Peer reviewed



Published in final edited form as:

*Science*. 2014 July 11; 345(6193): 212–215. doi:10.1126/science.1247727.

## Harnessing naturally occurring data to measure the response of spending to income

Michael Gelman<sup>1</sup>, Shachar Kariv<sup>2</sup>, Matthew D. Shapiro<sup>1,3,\*</sup>, Dan Silverman<sup>3,4</sup>, and Steven Tadelis<sup>3,5</sup>

<sup>1</sup>Department of Economics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>2</sup>Department of Economics, University of California, Berkeley, Berkeley, CA 94720, USA

<sup>3</sup>National Bureau of Economic Research (NBER), Cambridge, MA 02138, USA

<sup>4</sup>Department of Economics, Arizona State University, Tempe, AZ 85287, USA

<sup>5</sup>Haas School of Business, University of California, Berkeley, Berkeley, CA 94720, USA

### Abstract

This paper presents a new data infrastructure for measuring economic activity. The infrastructure records transactions and account balances, yielding measurements with scope and accuracy that have little precedent in economics. The data are drawn from a diverse population that overrepresents males and younger adults but contains large numbers of underrepresented groups. The data infrastructure permits evaluation of a benchmark theory in economics that predicts that individuals should use a combination of cash management, saving, and borrowing to make the timing of income irrelevant for the timing of spending. As in previous studies and in contrast to the predictions of the theory, there is a response of spending to the arrival of anticipated income. The data also show, however, that this apparent excess sensitivity of spending results largely from the coincident timing of regular income and regular spending. The remaining excess sensitivity is concentrated among individuals with less liquidity.

---

Economic researchers and policy-makers have long sought high-quality measures of individual income, spending, and assets from large and heterogeneous samples. For example, when policy-makers consider whether and how to stimulate the economy, they need to know how individuals will react to changes in their income. Will individuals spend differently? Will they save at a different rate or reduce their debt, and when? There are many obstacles to obtaining reliable answers to these important questions. One obstacle is that existing data sources on individual income and spending have substantial limits in terms of accuracy, scope, and frequency.

---

\*Corresponding author. shapiro@umich.edu.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/345/6193/212/suppl/DC1](http://www.sciencemag.org/content/345/6193/212/suppl/DC1)

Materials and Methods

Figs. S1 and S2

Tables S1 to S3

References

This paper advances the measurement of income and spending with new high-frequency data derived from the actual transactions and account balances of individuals. It uses these measures to evaluate the predictions of a benchmark economic theory that states that the timing of anticipated income should not matter for spending. Like previous research, it finds that there is a response of spending to the arrival of anticipated income. The data show that, on average, an individual's total spending rises substantially above average daily spending on the day that a paycheck or Social Security check arrives, and remains high for at least the next 4 days. The data also allow the construction of variables that show, however, that this apparent excess sensitivity of spending results in large part from the coincident timing of regular income and regular spending. The remaining excess sensitivity is concentrated among individuals who are likely to be liquidity-constrained.

Traditionally, researchers have used surveys such as the Consumer Expenditure Survey (CEX) to measure individual economic activity. Such surveys are expensive to implement and require considerable effort from participants and are therefore fielded infrequently, with modest-sized samples. Researchers have recently turned to administrative records, which are accurate and can be frequently refreshed, to augment survey research. So far, however, the administrative records have typically represented just a slice of economic activities. They have not provided simultaneous information about various sources of income and forms of spending.

The data described here result from transactions that are captured in the course of business by Check (<https://check.me>), a financial aggregation and service application (app). The resulting income data are accurate and comprehensive, in that they capture income from several sources and can be linked to similarly accurate and comprehensive information on spending. These raw data present important technical and conceptual challenges. This paper describes protocols necessary for turning them into a data set with several features that are useful for research and policy analysis.

Check had approximately 1.5 million active users in the United States in 2012. Users can link almost any financial account to the app, including bank accounts, credit cards, utility bills, and more. The application logs into the Web portals for these accounts daily and obtains the user's primary financial data. The data are organized so that users can obtain a comprehensive view of their financial situation.

The data we analyzed are derived from a sample of approximately 75,000 Check users, selected at random from the pool of U.S.-based users who had at least one bank or credit card account, and cover 300 consecutive days spanning 2012 and 2013. The data were de-identified and the analysis was performed on normalized and aggregated user-level data as described in the text and supplementary materials (SM). Check does not collect demographic information directly, and instead uses a third party that gathers both publicly and privately provided demographic data, anonymizes them, and provides aggregate tabulations of demographic characteristics of users. Table 1 compares the gender, age, education, and geographic distributions in the Check sample that matched with an email address to the distributions in the U.S. Census Bureau's American Community Survey (ACS), representative of the U.S. population in 2012.

Table 1 shows that the data overrepresent males and those aged 25 to 44. Education levels are broadly similar to those of the U.S. population, and the geographic distribution of Check users is reasonably consistent with that of the U.S. population. Overall, the sample contains large numbers of even the most underrepresented sociodemographic groups. For example, the sample contains about 3000 individuals aged 65 and older. At a point in time, the CEX contains information on approximately 1100 individuals aged 65 and older. We note, however, that the willingness to provide login credentials may select on personal characteristics or increased need for financial organization. The extent of this selection could be assessed with surveys of Check users, the results of which could be compared to those from existing surveys of representative populations. Alternatively, random samples of the population could be encouraged to link their financial accounts to the app, and the transaction and balances of this population could be compared with those of Check users.

Summary statistics for the raw data are provided in tables S1 and S2 of the SM. The data allowed us to calculate total income and to separately identify paychecks and Social Security payments using the description fields of transactions. Similarly, we measured total spending and subcategories of spending. We identified recurring and nonrecurring income and spending by looking for transactions that occurred at regular periodicity and had regular amounts.

We derived two measures of income: The first sums all transactions that represent credits to a user's non-credit card accounts, excluding transfers from one account to another. The second isolates only those transactions that credit paychecks and Social Security payments, using a list of keywords commonly found in the description field. Figure 1 shows the distribution of average monthly income measures at the user level.

Total monthly income depicted in Fig. 1A has a median of \$4800 and a mean of \$8923. The long and heavy right tail reflects income inequality and also includes large one-off transactions from asset sales. Paycheck and Social Security income, shown in Fig. 1B, is less skewed, with a median of \$2900 and a mean of \$3951. The figure also displays a kernel density estimate of the distribution of monthly incomes reported in the U.S. Census Bureau's ACS. The income concepts in the ACS and Check data have important differences. Figure 1A shows the distribution of ACS monthly pre-tax household income. The Check data shown in Fig. 1A are net of any (tax) withholding and may be aggregated from either individual or household income. Despite these differences, the ACS and Check distributions are qualitatively similar. Figure 1B shows the ACS distribution of wages, salaries, and the sum of wages and salaries and Social Security payments, which are more closely aligned with their analogs in the Check data. The shape of the ACS distribution is again similar to Check's.

For credit card accounts, we identify spending as transactions that post debits to the account. Non-credit card accounts are similar, but a sum of their debits will overstate spending, because some may represent credit card payments or transfers between accounts. Consequently, spending measures exclude debits we can identify as such payments or transfers either by amount or by transaction description.

We considered three measures of spending: (i) total spending, calculated using the method just described; (ii) nonrecurring spending; and (iii) spending on fast food and coffee shops. See fig. S1 in the SM for the distribution of these average weekly spending measures at the user level. Nonrecurring spending subtracts from total spending both ATM cash withdrawals and expenditures of at least \$30 that recur, in the exact same amount (to the cent), at regular frequencies, such as weekly or monthly. It isolates spending that, due to its irregularity, is not easily timed to match the arrival of income. This measure thus uses the amount and timing of spending rather than an a priori categorization based on goods and services, an approach made possible by the distinctive features of the data infrastructure. The fast food and coffee shop measure is identified using keywords from the transaction descriptions. This measure isolates an especially discretionary, nondurable, and highly divisible form of spending, which we used in the analysis of the spending response to anticipated income.

A benchmark theory indicates that the anticipated arrival of a payment should not affect the timing of spending. Specifically, spending should not rise after the arrival of a regular paycheck or Social Security payment. We estimated the excess sensitivity of total, nonrecurring, and coffee shop and fast food spending to the arrival of regular paychecks or Social Security payments. We thus evaluated the possibility that the benchmark theory describes behavior well and that excess sensitivity reflects either the convenience of coordinating recurring expenses with the arrival of regular income, or the intrinsic difficulty of smoothing some forms of spending. We also estimated the excess sensitivity of spending separately for users with different levels of liquidity and different levels of available credit. We thus evaluated the possibility that, as standard enhancements to the benchmark theory indicate, excess sensitivity is a phenomenon of those with inadequate liquidity or credit.

We restricted attention to approximately 23,000 users observed to receive paychecks or Social Security payments at a regular frequency and in regular amounts. A payment is classified as regular in frequency if the median number of days between its arrival is from 13 to 15 or from 26 to 34 and if its coefficient of variation is less than 0.5. The demographic characteristics of users who receive either regular paychecks or regular Social Security payments are remarkably similar to those of the entire sample, as are the distributions of their income, spending, and balances.

Our main econometric specification is

$$x_{ict} = \sum_{j=\text{Mon.}}^{\text{Sun.}} \delta_{jc} + \sum_{k=-7}^6 \beta_{kc} I_i(\text{Paid}_{t-k}) + \varepsilon_{ict} \quad (1)$$

where  $x_{ict}$  is the ratio of spending of individual  $i$  to  $i$ 's average daily spending in category  $c$ , at date  $t$ ,  $\delta_{jc}$  is a day-of-week fixed effect, and  $I_i(\text{Paid}_{t-k})$  is an indicator equal to 1 if  $i$  received a payment at time  $t-k$ , and equal to 0 otherwise. The  $\beta_{kc}$  coefficients thus measure the fraction by which individual spending in category  $c$  deviates from average daily spending in the days surrounding the arrival of a payment. The day-of-week dummies capture within-week patterns of both income and spending.

Figure 2 shows estimates of  $\beta_{kc}$  for the following categories of spending: (A) total, (B) non-recurring, and (C) coffee shop and fast food spending. The dashed lines are the bounds of the 95% confidence intervals of these estimates based on heteroskedasticity-robust standard errors, with clustering at the individual level. Figure 2A shows that, on average, a user's total spending rises about 70% above its daily average on the day that a regular paycheck or Social Security payment arrives, and remains high for at least the next 4 days.

Total spending includes, however, expenditures such as rent, cable bills, or tuition that are recurring and predictable and whose timing can be adjusted to match the arrival of regular income. Figure 2B shows the excess sensitivity of only nonrecurring spending, confirming that a substantial part (40%) of the excess sensitivity of total spending can be attributed to the convenience of paying major bills automatically and avoiding the bad consequences of temporary illiquidity. Given that we defined recurring spending conservatively (i.e., required that it be the same amount to the cent), this estimate is probably a lower bound on how much accounting for it reduces excess sensitivity.

Figure 2C provides still more evidence that the benchmark theory is a better description of behavior than the total spending estimates would suggest. For this imminently divisible and easily smoothed discretionary spending, we observe very modest excess sensitivity to the arrival of predictable income.

We find evidence of individual heterogeneity of excess sensitivity that is consistent with the theory that predicts such behavior among those with insufficient liquidity or available credit, perhaps due to imperfections in credit markets. Figure 3 plots estimates of  $\beta_{kc}$  for nonrecurring spending by terciles of liquidity. We define liquidity for each user as the average daily balance of checking and savings accounts over the entire sample period, normalized by the user's average daily spending. The average user in the lowest tercile has 5 days of spending in cash on hand; the average user in the highest tercile has 159 days. The estimates show that excess sensitivity is significantly more pronounced among those in the lowest tercile of the liquidity distribution.

Figure S2 plots estimates of excess sensitivity by terciles of the available credit utilization distribution. Excess sensitivity is concentrated among users near the limit of their ability to borrow with credit cards. Those who have little liquidity or take their debt levels very close to their limits may be poor at planning or optimizing. The evidence indicates that differences in liquidity and constraints drive heterogeneity of excess sensitivity among Check users.

Many prior studies of spending responses to income have used the CEX quarterly retrospective survey, which records self-reports of income but does not measure its timing precisely. Souleles, for example, uses it to estimate the spending response to the arrival of income tax refunds and overcomes the lack of timing information by calculating from aggregate statistics the likelihood of receiving a refund at various dates (1). Parker takes a similar approach and exploits anticipated changes in take-home pay when workers hit the annual cap on the Social Security payroll tax (2). Johnson *et al.* and Parker *et al.* measure the timing of some income more precisely by adding special questions to the CEX about tax rebates (3, 4).

Some studies use higher-frequency data to estimate spending responses to income. The CEX diary survey records spending daily for 2 weeks but does not collect high-frequency income data. Stephens overcomes this limitation by studying the spending response to the receipt of Social Security benefits, which used to arrive on the same day of each month (5). The UK's Family Expenditure Survey collects the most recent paystub of respondents and asks them to track spending for 2 weeks. Stephens uses the paystub to infer the amount and timing of paychecks and estimates the spending response to them (6).

These prior studies use a variety of methods, but share an interest in estimating either an elasticity, defined as  $\frac{\partial \log(\text{spending})}{\partial \log(\text{income})}$ , or a marginal propensity to consume (MPC), defined as  $\frac{\partial(\text{spending})}{\partial(\text{income})}$ . Table S3 summarizes the key features of these prior estimates and compares them to analogous aspects of our study.

The studies differ in the time frame over which they measure spending changes in response to a change in income. This makes the levels of their estimated elasticities or MPCs difficult to compare. For our study, we present the point estimate of effects on the first day after the income arrives; that is  $\beta_{1c}$  from Eq. 1 for the elasticity of spending in category  $c$ . For the MPC we present the  $\gamma_{1c}$  from the equation

$$x_{ict} = \alpha_{ic} + \sum_{j=\text{Mon.}}^{\text{Sun.}} \delta_{jc} + \sum_{k=-7}^6 \gamma_{kc} \text{Payment}_{ic,t-k} + \varepsilon_{ict} \quad (2)$$

where  $x_{ict}$  is the ratio of spending of individual  $i$  to  $i$ 's average daily spending in category  $c$ , at date  $t$ ;  $\delta_{jc}$  is a day-of-week fixed effect;  $\alpha_{ic}$  is a user fixed effect; and  $\text{Payment}_{ic,t-k}$  is the ratio of the amount of the payment received by individual  $i$  divided by  $i$ 's average daily spending in category  $c$ , at date  $t-k$ . Analogously, table S3 presents only the shortest-run effects reported in all the other studies. Although our and other studies estimate larger impacts at longer horizons, the central conclusion of table S3 about the relative precision of the estimates is not affected by the choice of horizon.

The prior estimates are important and influential but, as table S3 shows, they often lack precision. Among studies of the quarterly CEX data, Hsieh is unusual in its precision (7). The last four rows of table S3 include the confidence intervals for our estimates of both the elasticity and the MPC. These intervals are small, both economically and relative to other studies. Only Broda and Parker provides estimates that are as precise as those from the Check data (8). That paper uses Homescan data and estimates precisely an MPC out of tax rebates near 0. These estimates rely on surveys to determine receipt of the rebate, however, and would be attenuated if those reports are subject to error. The Homescan spending data are also limited in scope, largely capturing only goods with Universal Product Codes. Moreover, the Check data allow estimates of the response to routine payments such as paychecks and Social Security payments, not just particular payments such as tax rebates.

Related studies of administrative data also provide accurate measures of spending but do not cover it comprehensively. For example, Agarwal *et al.* use data from a single credit card company to study the spending response to tax rebates; they can thus track the effects of the

rebate on a single account but not on overall spending (9). Kuchler makes use of more comprehensive administrative data collected from a debt management Web site, but the number of users (556) is relatively small (10). The financial application Mint (<https://www.mint.com/>) has a complementary data infrastructure that it is using to construct monthly time series of spending by types of goods (11). It has not been used for research along the lines of the estimates in this paper.

In policy discussions before the 2008 tax rebates, the Congressional Budget Office and others cited the point estimates of the effect of the 2001 rebate from Parker, Johnson, and Souleles, but not the substantial uncertainty about that estimate documented in that paper and in table S3 (3, 12). More generally, estimates of spending rates from different changes in income play a key role in the evaluation of the American Recovery and Reinvestment Act (13), making the stakes in getting credible and precise estimates of these parameters very high. This paper shows how economic theory and policy can benefit from analysis made possible with naturally occurring data such as those provided by Check.

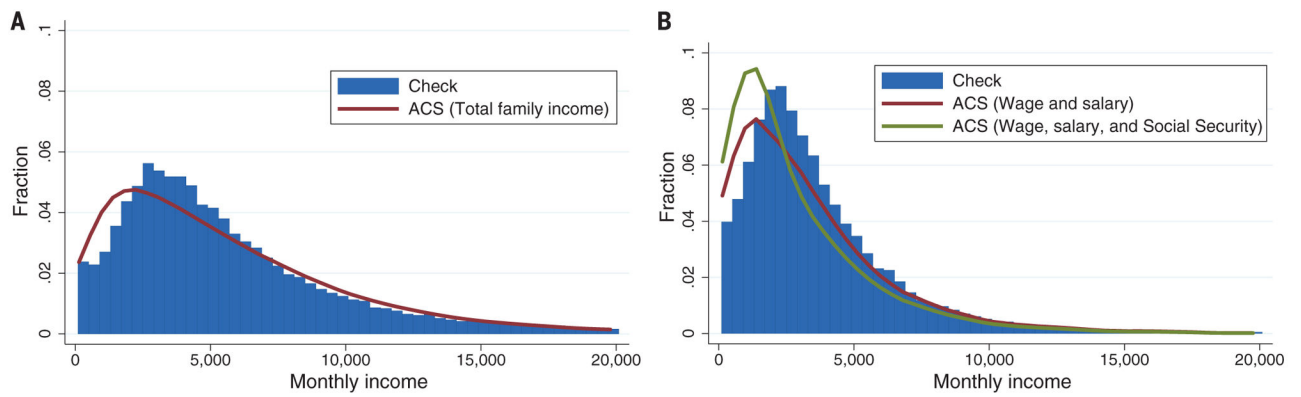
## Acknowledgments

This research was supported by a grant from the Alfred P. Sloan Foundation. M.D.S. acknowledges additional support through the Michigan node of the NSF-Census Research Network (NSF grant SES 1131500). This paper has benefited from suggestions by the participants of the NBER Summer Institute, the Conference on Economic Decisionmaking (Aspen, Colorado), and several seminars. A data set for replicating the results of this paper is available through the University of California Berkeley Econometrics Lab (EML) at <https://eml.berkeley.edu/cgi-bin/HarnessingDataScience2014.cgi>. To access the data, users must register with EML and agree to terms of use. The data set contains no personal or account identifiers. The data are aggregated and transformed so that they reveal no personally identifying information.

## REFERENCES AND NOTES

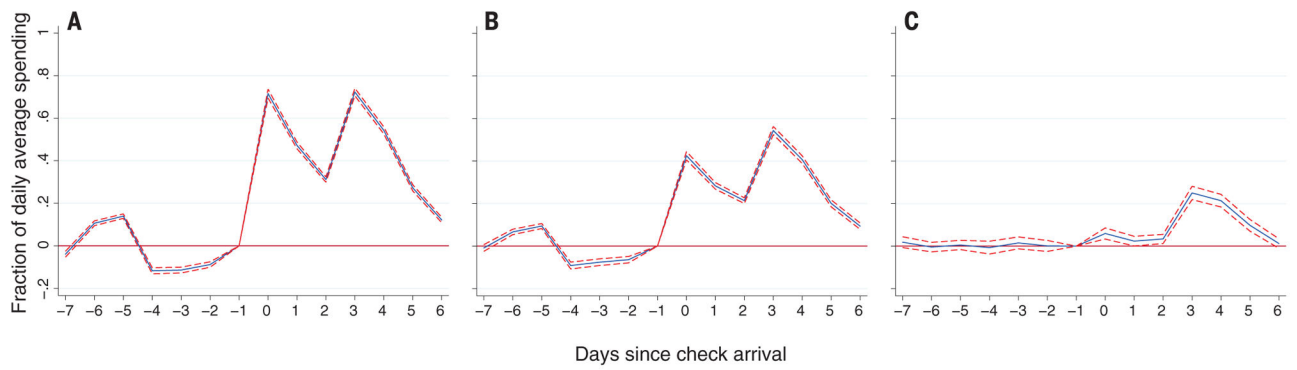
1. Souleles NS. *Am Econ Rev.* 1999; 89:947–958.
2. Parker JA. *Am Econ Rev.* 1999; 89:959–973.
3. Johnson DS, Parker JA, Souleles NS. *Am Econ Rev.* 2006; 96:1589–1610.
4. Parker JA, Souleles NS, Johnson DS, McClelland R. *Am Econ Rev.* 2013; 103:2530–2553.
5. Stephens M Jr. *Am Econ Rev.* 2003; 93:406–422.
6. Stephens M Jr. *Econ J.* 2006; 116:680–701.
7. Hsieh CT. *Am Econ Rev.* 2003; 93:397–405.
8. Broda, C.; Parker, JA. The economic stimulus payments of 2008 and the aggregate demand for consumption. Sloan School of Management; Cambridge, MA: 2014. manuscript
9. Agarwal S, Liu C, Souleles NS. *J Polit Econ.* 2007; 115:986–1019.
10. Kuchler, T. Sticking to your plan: Hyperbolic discounting and credit card debt paydown. Stern School of Business; New York: 2013. manuscript
11. Intuit. Intuit consumer spending index. Intuit; Mountain View, CA: 2013.
12. Congressional Budget Office. Options for Responding to Short-Term Economic Weakness. U.S. Congressional Budget Office; Washington, DC: 2008.
13. Congressional Budget Office. Estimated Impact of the American Recovery and Reinvestment Act on Employment and Economic Output from July 2010 Through September 2010. U.S. Congressional Budget Office; Washington, DC: 2010.





**Fig. 1. Distribution of monthly income**

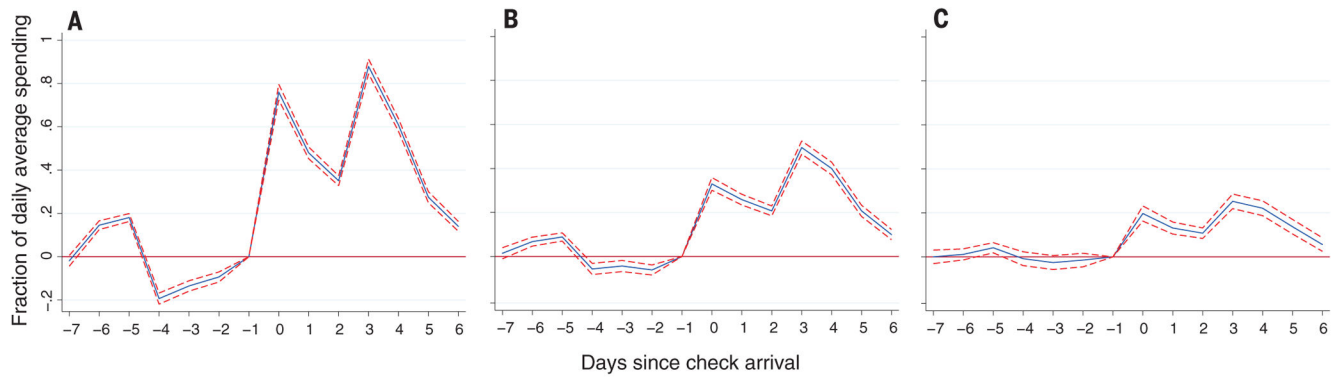
(A) Total income. (B) Paycheck and Social Security income. The figure shows average monthly income across users. Any month in which the user had fewer than 20 days of data was dropped from computation of the average. In (A), the Check distribution represents 61,184 users who have at least one checking or savings account. In (B), both Check and ACS distributions are conditional on having paycheck and Social Security income (47,050 users).



**Fig. 2. Response of spending to income: Alternative components of spending**

(A) Total spending. (B) Nonrecurring spending. (C) Fast food and coffee shop spending.

The solid line represents regression coefficients from Eq. 1. The dashed lines are 95% confidence intervals. Estimates are based on 5,371,244, 5,371,244, and 5,173,594 total observations from 23,985, 23,985, and 23,021 users for panels (A), (B), and (C), respectively.



**Fig. 3. Response of nonrecurring spending to income: Liquidity ratio**

(A) Low liquidity. (B) Medium liquidity. (C) High liquidity. The solid line represents regression coefficients from Eq. 1. The dashed lines are 95% confidence intervals. Estimates are based on 1,784,460, 1,809,839, and 1,769,968 total observations from 7956, 7956, and 7955 users for panels (A), (B), and (C), respectively. The liquidity ratio is defined as the average daily balance of checking and savings accounts normalized by daily average spending.

**Table 1**  
**Check versus ACS demographics (percent)**

The sample size for Check is 59,072, 35,417, 28,057, and 63,745 for gender, age, education, and region, respectively. The sample size for ACS is 2,441,532 for gender, age, and region and 2,158,014 for education.

	Check	ACS
<i>Sex</i>		
Male	59.93	48.59
Female	40.07	51.41
<i>Age</i>		
18–20	0.59	5.72
21–24	5.26	7.36
25–34	37.85	17.48
35–44	30.06	17.03
45–54	15.00	18.39
55–64	7.76	16.06
65+	3.48	17.95
<i>Highest degree</i>		
Less than college	69.95	62.86
College	24.07	26.22
Graduate school	5.98	10.92
<i>Census Bureau region</i>		
Northeast	20.61	17.77
Midwest	14.62	21.45
South	36.66	37.36
West	28.11	23.43