Slides on

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

Science 2014

Michael Gelman, Shachar Kariv, Matthew D. Shapiro, Dan Silverman, Steven Tadelis
“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.”

Why You Must Register

The public release data are stored in the facilities of the University of California, Berkeley’s Econometrics Lab (EML). They are protected by password and you must register to acquire the credentials. When you register to gain access, you are agreeing to the Conditions of Use for these data.

After registering you will obtain a link to download the Gelman, Kariv, Shapiro, Silverman, and Tadelis public release data, as well as replication code for Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014, Science). Before downloading the public release data and replication code you will be asked to agree to the following:

By downloading these Materials, I agree to the following:

I will not use the Materials:

- For any commercial purpose.
- To obtain information that could directly or indirectly identify individuals whose data have been aggregated in these Materials.
- To obtain information about, or further contact with, individuals known to me and whose data have been aggregated in these Materials.

I agree not to download any Materials where prohibited by applicable law.

I agree not to use the Materials in any way prohibited by applicable law.

I agree not to redistribute the Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014) public release data or replication code to any third parties.

I agree to secure the Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014) public release data on a password protected machine to prevent its accidental redistribution.
Our main econometric specification is

$$x_{ict} = \sum_{j=\text{Mon.}}^{\text{Sun.}} \delta_{jc} + \sum_{k=-7}^{6} \beta_{kc} I_i(Paid_{t-k}) + \epsilon_{ict}$$  \hspace{1cm} (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 $I_i(Paid_{t-k})$ 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.

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.

Figure 3 plots estimates of $\beta_{kx}$ 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.