Harnessing Commercial Data for Measurement: The Mint Bills Project

Michael Gelman, Michigan   Yuriy Gorodnichenko, UC Berkeley

Shachar Kariv, UC Berkeley   Matthew Shapiro, Michigan

Dan Silverman, Arizona State   Steve Tadelis, UC Berkeley

FESAC
June 2015

Challenges of Measuring Spending, Income, and Assets

Hard to get large samples to provide information that is:

- Accurate
- Comprehensive
    - Across individuals
    - Within individual
- High frequency
- Timely

Our Approach – Primary Advantages

- Partner with Intuit and its mobile banking app Mint Bills
- Data generated in the course of business
- Made available for research at no incremental cost
- High frequency and timely: daily updates from population of the app's users.
- Accurate: administrative records of transactions and balances
- Comprehensive
  - Across individuals: users are diverse
  - Within individual: all transactions and balances from an individual's checking, saving, credit card, investment accounts

Our Approach – Primary Challenges

- No direct information on demographics
  - ▸ What is a user, and how does it relate to the household?
  - ▸ How do users vary by gender, age, education, occupation, etc.?
- Data are raw, not organized for research
  - ▸ Spending, income, and assets are not pre-categorized
- Sample is not randomly selected

Background – What is *Mint Bills*?

Mint Bills

- *Mint Bills* is an app for mobile phones, tablets, and the web.
- Has registered more than 12 million registered users since its founding in 2007.
- Users can integrate information from nearly any financial account with a web-based portal.
- Users provide *Mint Bills* with the credentials necessary to access these portals and, every day, *Mint Bills* automatically logs into and scrapes the associated webpages.
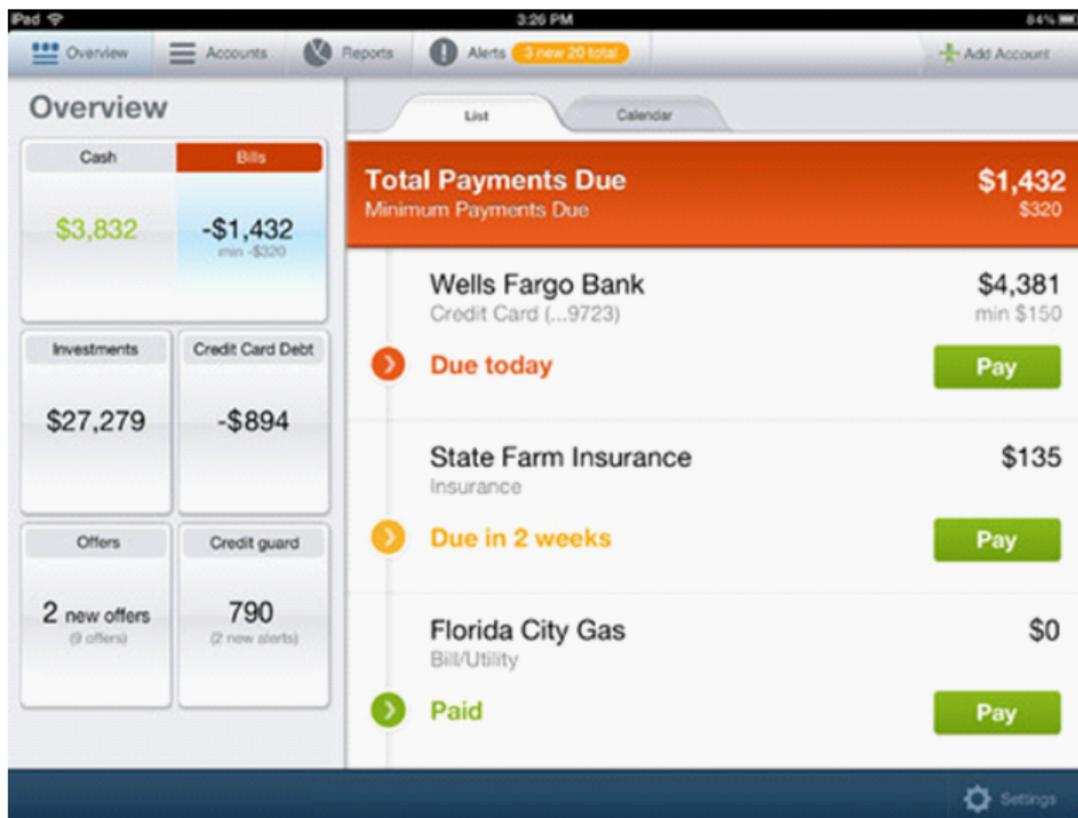
Mint Bills
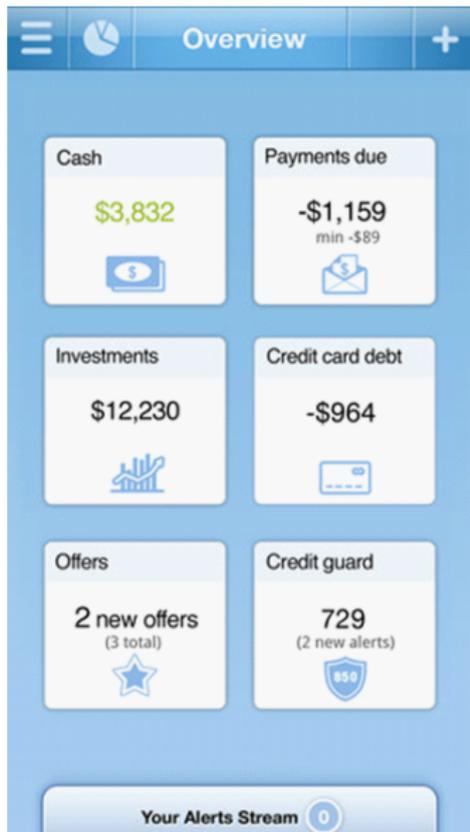
Mint Bills
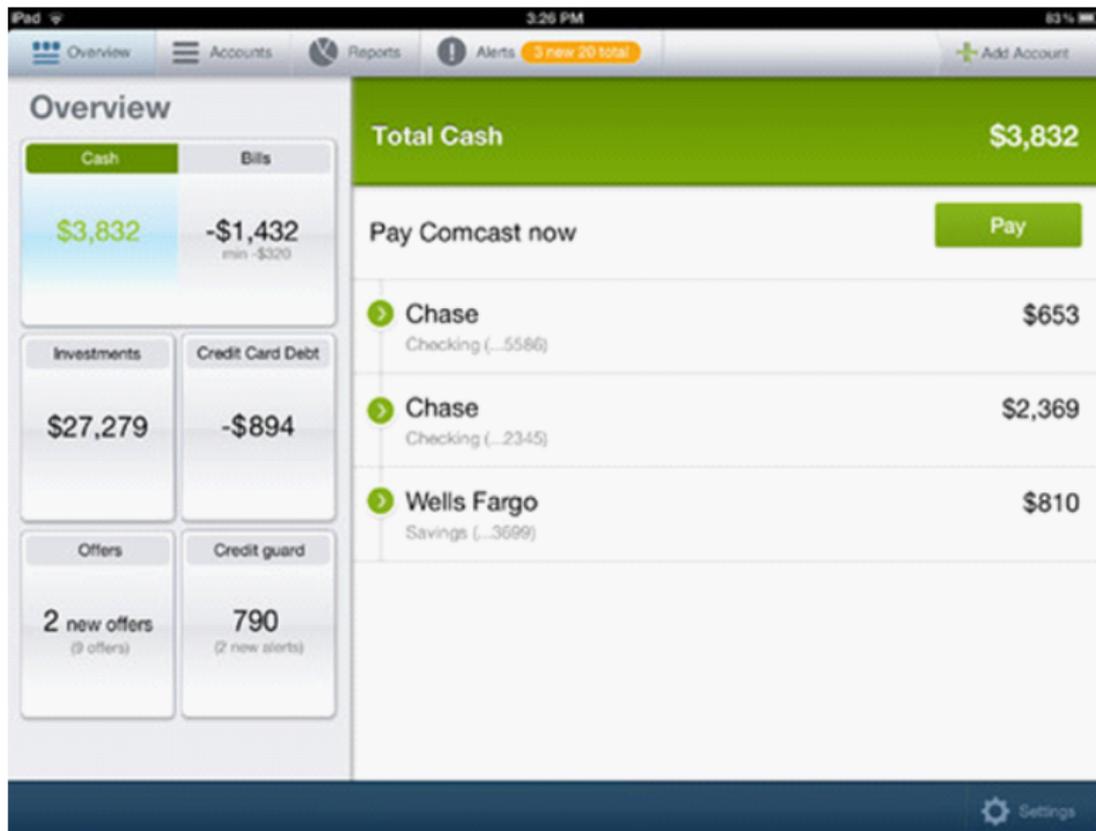
Mint Bills

Who is in the *Mint Bills* data?

	Mint Bills	ACS
Female	40.07	51.41
Age		
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	18.39
55-64	7.76	16.06
65+	3.48	17.95
Highest degree		
Less than College	69.95	62.86
College	24.07	26.22
Graduate School	5.98	10.92
Census Bureau Region		
Northeast	20.61	17.77
Midwest	14.62	21.45
South	36.66	37.36
West	28.11	23.43

Table 1: Mint Bills vs. ACS Demographics

Note: The sample size for Mint Bills 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

What's in the *Mint Bills* data?

Pilot Sample of Transactions and Accounts

Table 2: Transactions and Accounts

	Mean		P_5	P_{25}	P_{50}	P_{75}	P_{95}
	Mint Bills	SCPC			Mint Bills		
Daily Transactions	4.54	2.33	1	2	3	6	13
Credit card	1.23	0.50	0	0	1	2	5
Checking account	3.03	0.91	0	0	2	4	11
Saving account	0.22	n/a	0	0	0	0	1
Accounts	5.84		2	3	5	8	12
Credit card	3.58	4.00	1	2	3	5	9
Checking account	1.35	1.60	0	1	1	2	3
Saving Account	0.79	1.80	0	0	1	1	2

Note: In total, 72,902 Mint Bills users genearated 57,731,354 transactions over the 300 days of the study period. The Boston Fed's Survey of Consumer Payment Choice (SCPC) is conducted with a sample of approximately 2,100 members of the online survey, the American Life Panel.

Balance Data

Table 3: Account Balances

Panel (a) Bank		Median		Mean	
		Mint Bills	SCF	Mint Bills	SCF
All		$2,200	$4,000	$14,415	$30,493
	Checking	$1,400	$1,900	$6,969	$8,338
	Saving	$400	$3,500	$6,476	$17,312
	Money Market	$900	$22,500	$12,076	$90,033
	C.D.	$500	$16,000	$12,734	$63,859

| Panel (b) Credit Card	Median		Mean	
	Mint Bills	SCF	Mint Bills	SCF
Current Balance	$3,600	n/a	$7,228	n/a
Credit Limit	$11,900	$15,000	$23,019	$40,197
Utilization Ratio	0.45	n/a	0.48	n/a
Revolving Debt	$3,500	$2,500	$5,828	$5,707
APR	18.00%	13.90%	18.46%	13.57%

Notes: Mint Bills data are from the sample of 75,000, in the study period covered by Gelman et al. (2014). Survey of Consumer Finance analogues are weighted with population weights. Amounts are conditioned on a user/respondent having that type of account. Last three rows of Panel (b) condition on having revolving debt.

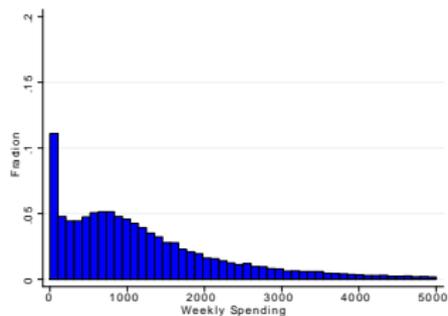Measuring Income and Expenditure from Transactions

Distributions of User Income
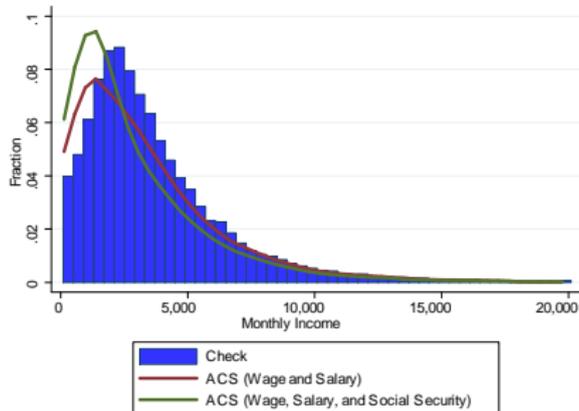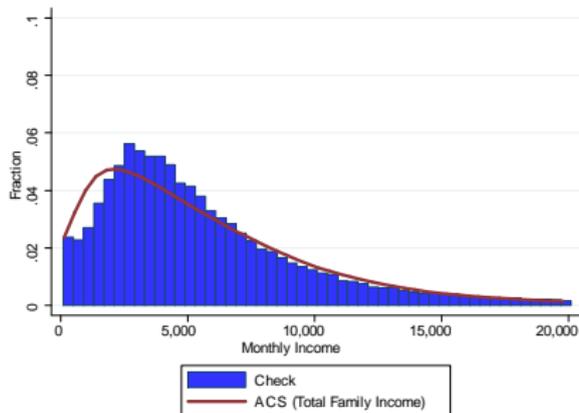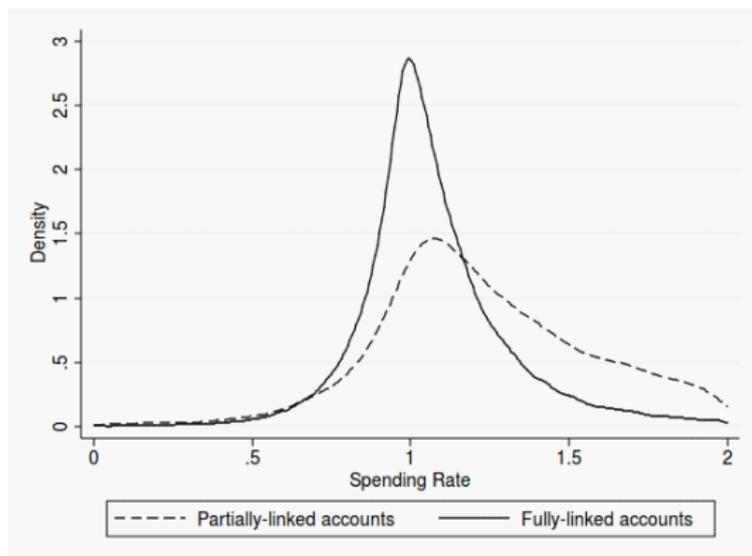
Total Weekly Spending
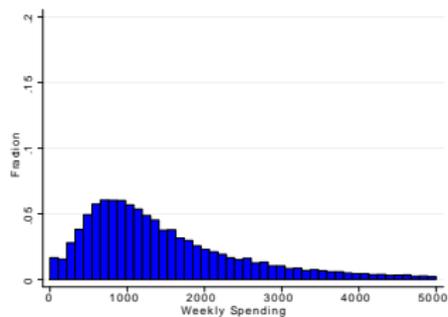
Total (well-linked)

Total (all users)

Integrating Income and Spending Measures

Integrating Measures: The Spending Rate

Annual after tax average in the 2013 CE Survey: 0.91

- 1.14 for hhlds earning $<\$70K$
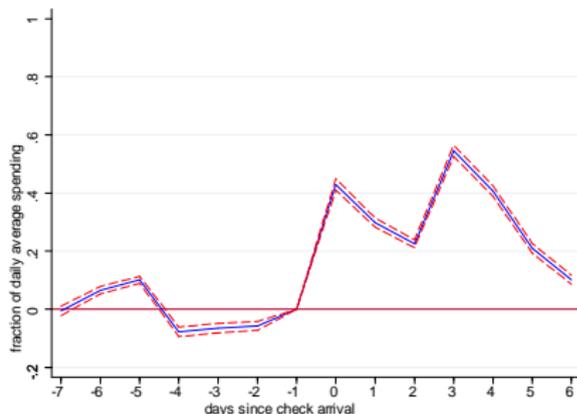- 0.73 for hhlds earning $\geq \$100K$

Integrating Measures: Spending Responses to the Arrival of Anticipated Income

- Gelman et al., *Science* (2014)
- Does spending respond to the arrival of anticipated income?
- If they know the money is coming, why don't consumers smooth their spending?

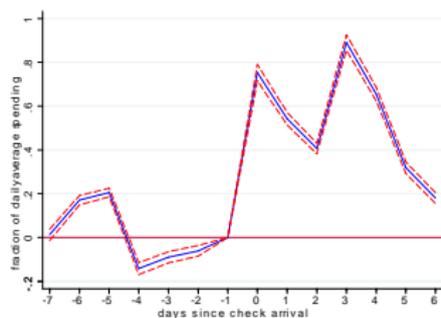# Response of Total and Non-Recurring Spending to Paycheck Income
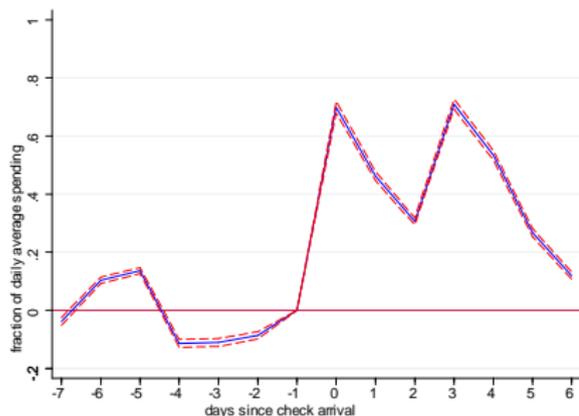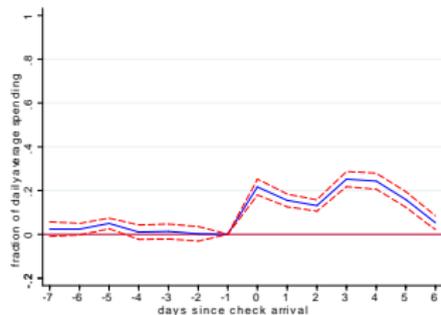

Total

Non-Recurring

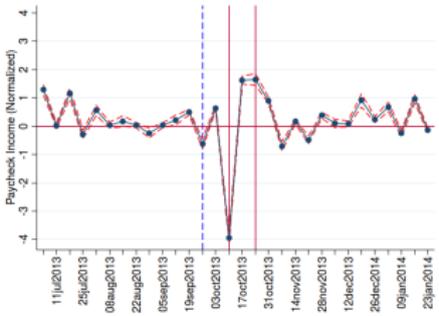# Further Integration: Response Heterogeneity By Liquidity

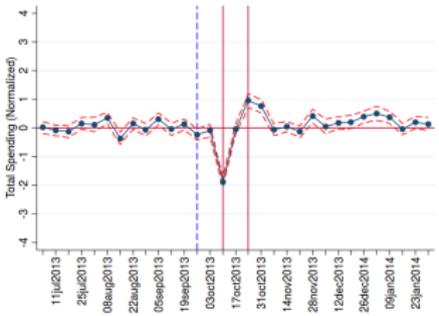Bottom Third of the Liquidity Distribution

Top Third of the Liquidity Distribution

Responses to Income Shock: The 2013 Govt. Shutdown

Paycheck Income

Total Spending

Making These Data More Useful for Official Statistics

- Mint Bills data have proved useful for research
- Work in progress should make the data more useful for official statistics
- Survey Mint Bills users
  - ▶ Assess representativeness
  - ▶ Understand the relationship between users and hhlds.
  - ▶ Evaluate integration with the app
- Surveys will thus facilitate weighting and aggregation
- Allow real-time income, spending, and debt indices to complement official statistics

Some Challenges of Generating Time Series

- Complex "seasonals"
  - Day of week
  - Point in month
  - Paycheck effects
  - Trading day and reconciliation effects
  - Interactions of all of the above
- Short sample (Mint can help)
- Changing composition of users
- Evolving features and uses of the app

Making These Data More Useful for Official Statistics

- Use the Mint Bills apparatus as a tool of survey research
- Bring the apparatus to representative samples
  - American Life Panel
  - PSID
  - NLSY
- Complement existing income and spending measures
- Decrease the burden on survey respondents without sacrificing accuracy.
- Use app to replace components of CE Survey or Diary.

Making These Data More Useful for Official Statistics

- Providing responsible access to the data is a challenge.
- Priority is on protecting the confidentiality of users and the business interests of the firm
- Careful aggregation has allowed registered researcher access to data associated with Gelman et al. (2014).
- How can we facilitate better cooperation between researchers, firms, and official statistics?