UNIVERSITY OF MICHIGAN

Big Data in Macroeconomics

Matthew D. Shapiro
University of Michigan

Universidad de Chile – October 24, 2018

Naturally-Occurring Data
(Big data, non-designed data)

Data created in the normal course of activities of households, businesses, and governments

Contrast:

Designed data, e.g., surveys or experiments

Naturally-Occurring Data, Examples

Transactional and account data

Households

- Bank accounts, credit cards

Business

- Scanner data: price and quantity
- Payroll: employment and wages

Naturally-Occurring Data, Examples

Government:

- New claims for unemployment insurance
- Tax, Social Security records
- Benefits data
  - Food stamps, Social Security, Medicare
- Federal Contracts
- Tolls
- Airplane flight and load data
- Import duties
- Regulatory filings

Naturally-Occurring Data, Examples

Web scrapping:

- Prices
- Help wanted advertising
- Textual analysis/document contents

Naturally-Occurring Data, Examples

Social media:

- Twitter
- Facebook
- Craig’s list
- Linked-in
- Search activity

Naturally-Occurring Data, Examples

Other:

- Press reports
- Fitbit, etc
- Health club utilization
- Images
- Satellite data (ground cover, light)
- Property sales and assessment
- Box office sales
- U-Haul long-distance rentals

Conventional sources

Enumeration

- Census of population
- CPI price observations

Conventional sources

Government surveys

- Current Population Survey (unemployment)
- Census of manufacturing and other businesses
- Consumer Expenditure Survey

Academic surveys

- Health and Retirement Study
- Panel Study of Income Dynamics

Hybrid: Survey and Admin data

Government statistics

- Census survey use tax records for imputation
- GDP uses many commercial sources

Academic surveys

- Health and Retirement Study links Social Security and Medicare records

Billion Prices Project

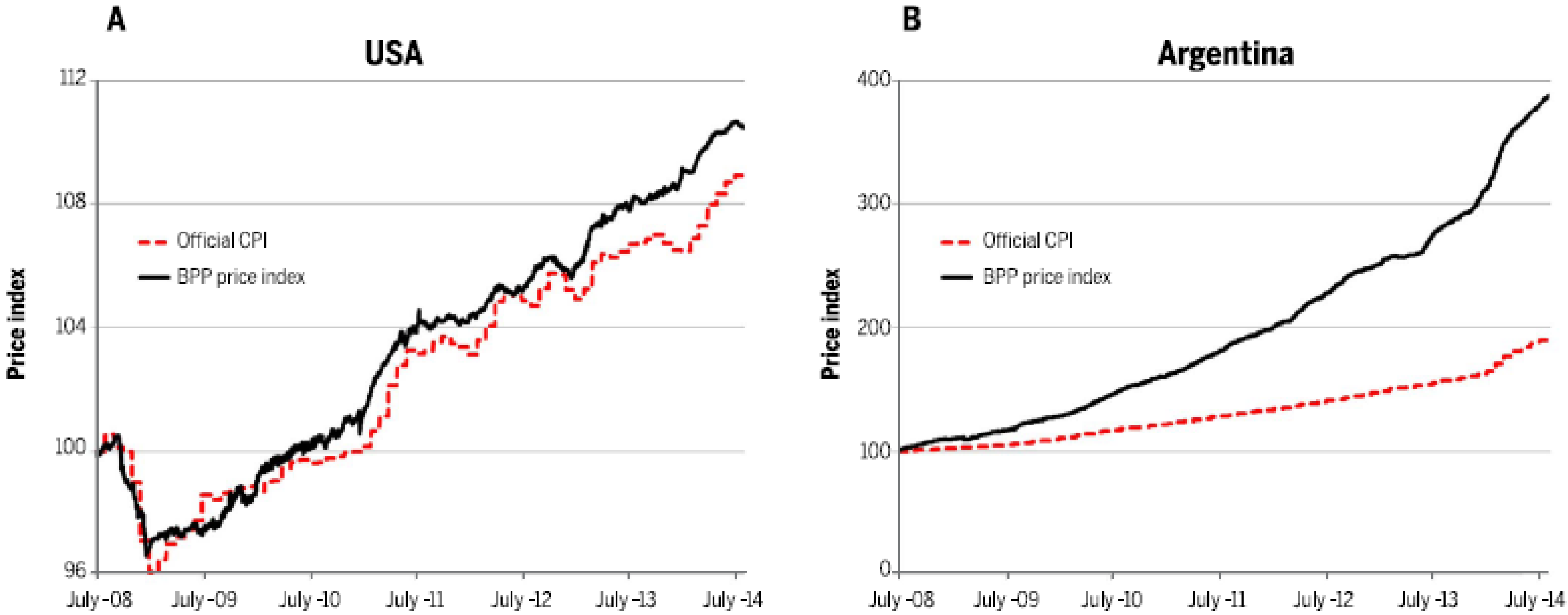
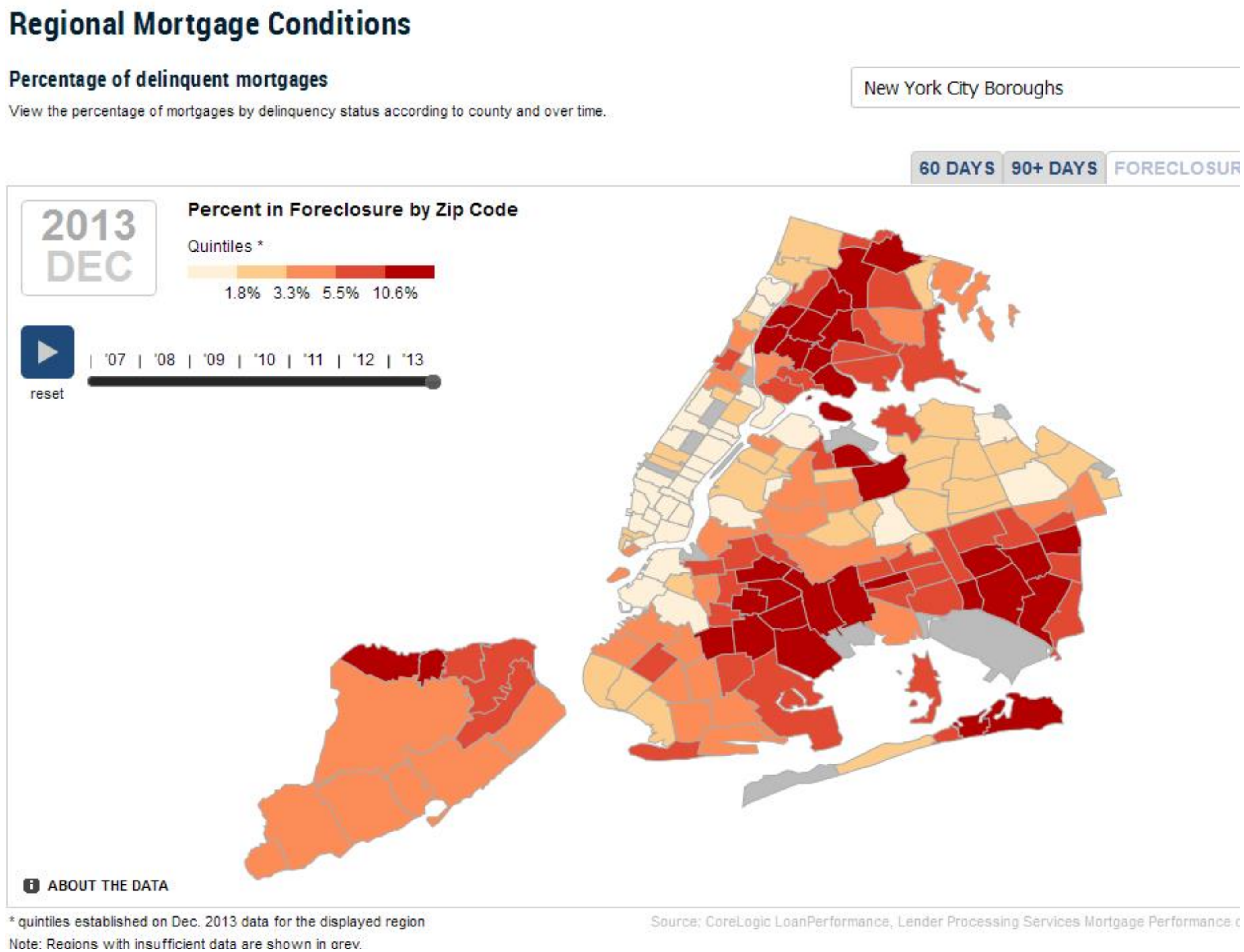


Fig. 2. BPP price index. Dashed red lines show the monthly series for the CPI in the United States (**A**) and Argentina (**B**), as published by the formal government statistics agencies. Solid black lines show the daily price index series, the "State Street's PriceStats Series" produced by the BPP, which uses scraped Internet data on thousands of retail items. All indices are normalized to 100 as of 1 July 2008. In the U.S. context, the two series track each other quite closely, although the BPP index is available in real time and at a more granular level (daily instead of monthly). In the plot for Argentina, the indices diverge considerably, with the BPP index growing at about twice the rate of the official CPI. [Updated version of figure 5 in (18), provided courtesy of Alberto Cavallo and Roberto Rigobon, principal investigators of the BPP]

FRB NY Financial Conditions

link



Light and economic activity

Fig. 1. Nighttime lights of North America. Nighttime stable lights for year 2006 in arc 30-s resolution are shown. The projected coordinate system of US contiguous Albers equal area conic projection is used and the image is generated with ArcGIS 9.3.

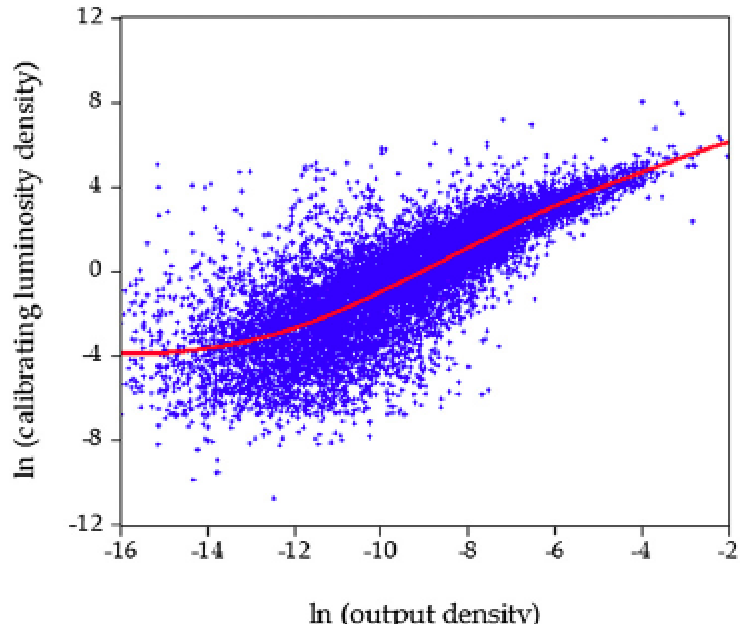


Fig. 2. Gross cell product (GCP) and luminosity data, all cells. Shown are the scatter plot of log calibrated luminosity for 2006 and log of gross cell product for all 1° × 1° grid cells. Output density is gross cell product (PPP in billions in 2005 international dollars) per square kilometer. Luminosity density per square kilometer is the radiance calibrated luminosity for 2006. All grid cells (n = 12,393) are included. The solid line is the kernel estimator using an Epanechnikov kernel and 100 grid points per kernel.

Source: Chen and Nordhaus *PNAS*

Data from Financial App

"Harnessing Naturally Occurring Data to Measure the Response of Spending to Income." *Science* (July 2014)

"How Individuals Respond to a Liquidity Shock: Evidence from the 2013 Government Shutdown." (NBER Working Paper 21025.) *Journal of Public Economics* (forthcoming).

Leading Methods: Surveys and Administrative Records

Surveys of individuals are comprehensive but ...

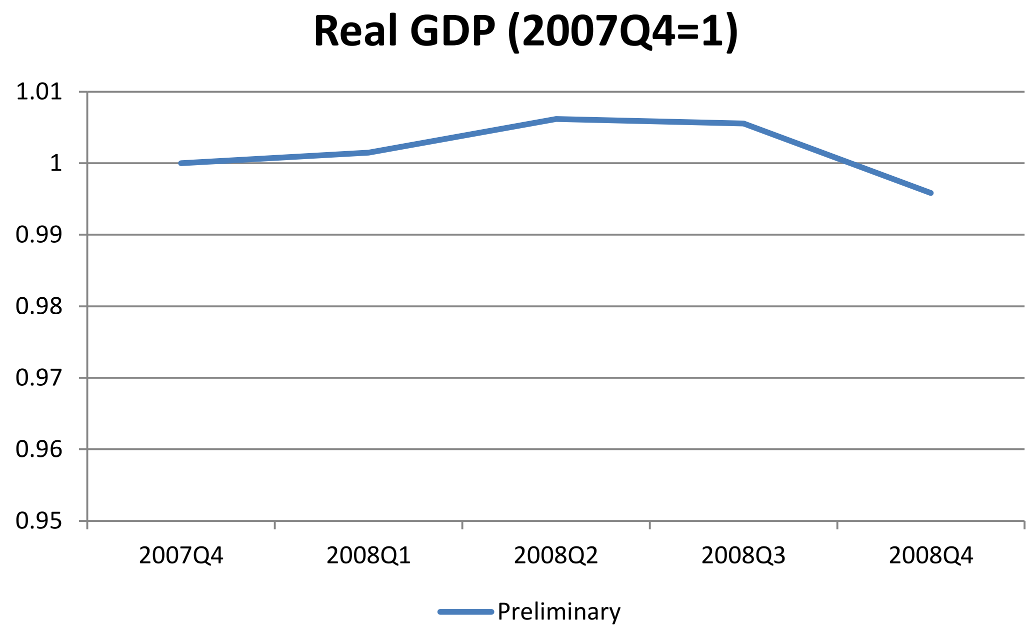
- Self-reported
- Typically low frequency
  - Long, varying, and staggered reporting intervals
  - Infrequent reports
  - Published with considerable lag

Administrative records are accurate, high frequency, and timely but ...

- Not comprehensive
- Large fractions of expenditure, portfolio, or income are missing

Current system gets it wrong exactly when it matters

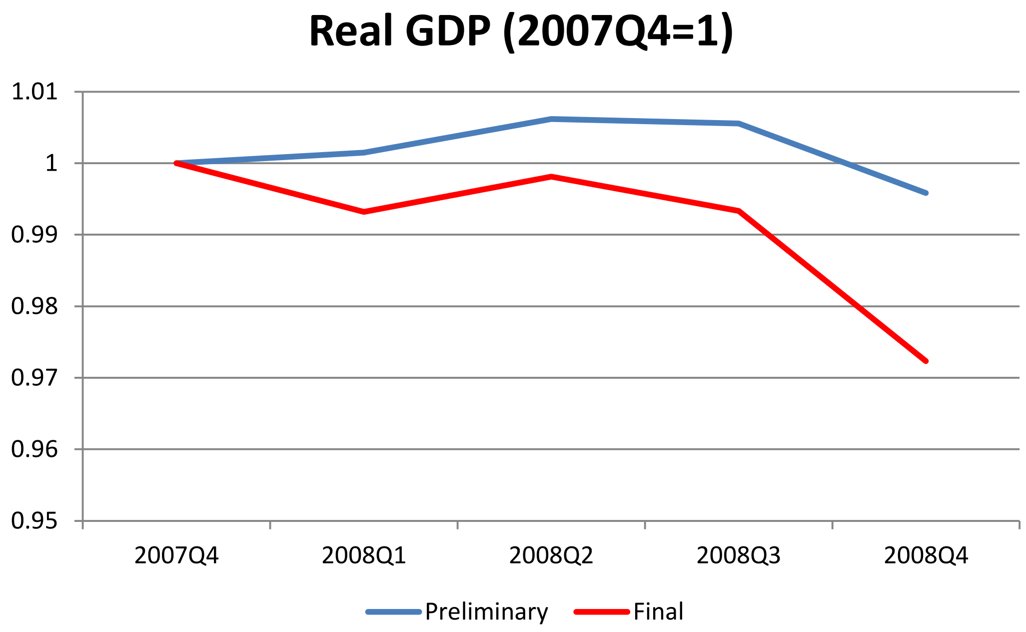
Key components of GDP extrapolated



Current system gets it wrong exactly when it matters

Key components of GDP extrapolated

→Huge miss in fourth quarter of 2008



Financial App

- App for mobile phones, tablets, and the web
- Has registered more than 10 million registered users 2007
  - Pilot sample of 75,000
  - Now following 1,000,000+ users
- Users can integrate information from nearly any financial account with a web-based portal
- Users provide app with the credentials necessary to access these portals and, every day, app automatically logs into and scrapes the associated webpages

Some Challenges of Data

- No direct information on demographics
- Data are raw, not organized for research
- Spending is not pre-categorized
- Sample is not randomly selected

Who is in App?

Comparison of demographics

- App: Third-party data based on email
- ACS: American Community Survey from Census Bureau

Age

	App	ACS
18–20	0.6	5.7
21–24	5.3	7.4
25–34	37.9	17.5
35–44	30.1	17.0
45–54	15.0	18.4
55–64	7.8	16.1
65+	3.5	18.0

Gender

	App	ACS
Male	59.9	48.6
Female	40.1	51.4

Education

	App	ACS
Less than college	70.0	62.9
College	24.1	26.2
Graduate school	6.0	10.9

Region

	App	ACS
Northeast	20.6	17.8
Midwest	14.6	21.5
South	36.7	37.4
West	28.1	23.4

Transactions and accounts

	Mean	P_5	P_{25}	P_{50}	P_{75}	P_{95}
Daily transactions	4.54	1	2	3	6	13
Credit card	1.23	0	0	1	2	5
Checking account	3.03	0	0	2	4	11
Saving account	0.22	0	0	0	0	1
Accounts	5.84	2	3	5	8	12
Credit card	3.58	1	2	3	5	9
Checking account	1.35	0	1	1	2	3
Saving account	0.79	0	0	1	1	2

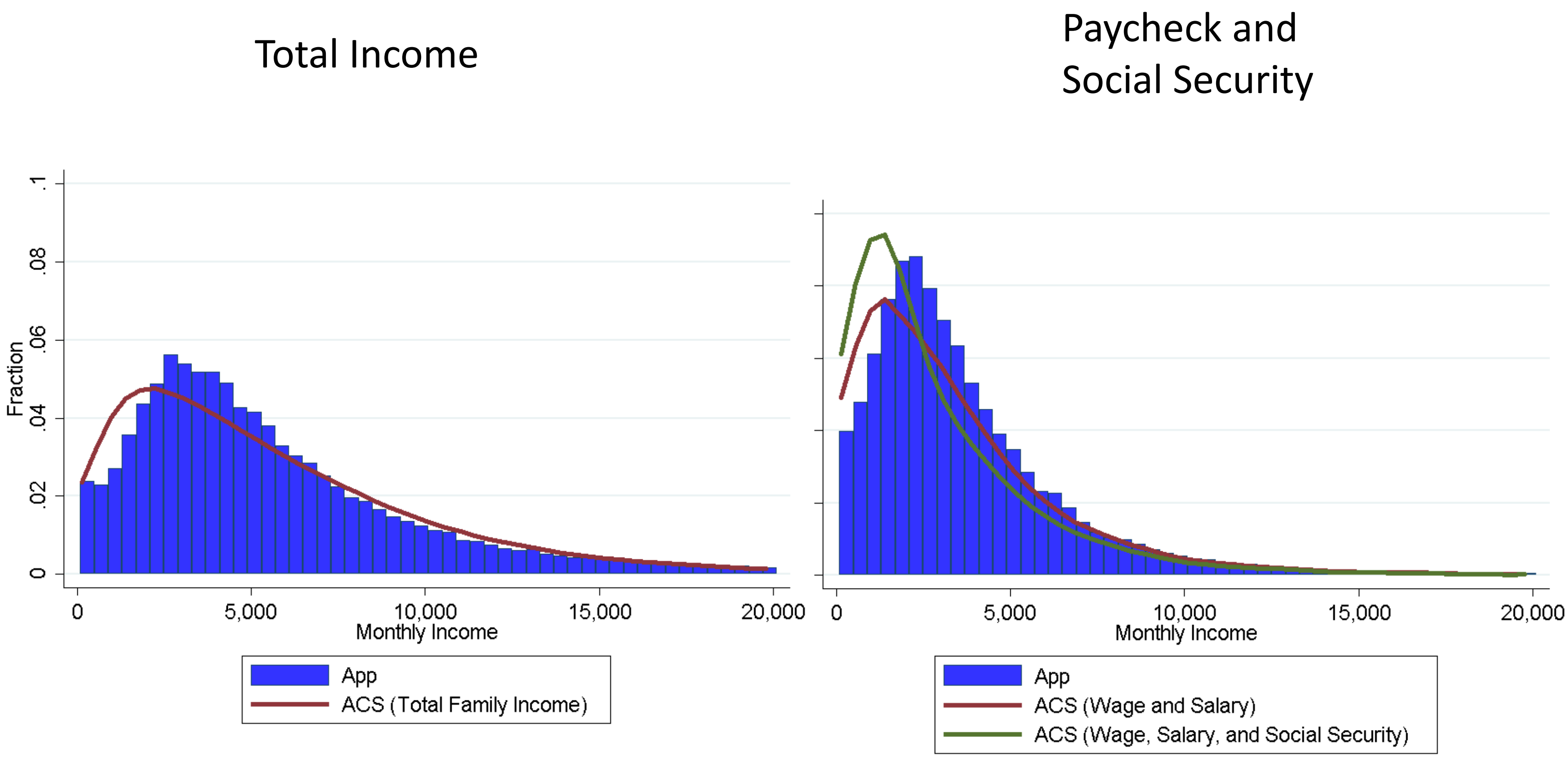
Notes: In total, the 57,731,354 transactions are generated from 72,902 unique users over the study period.

Account balance

Panel (a): Bank	Mean	P_5	P_{25}	P_{50}	P_{75}	P_{95}
All	\$14,415	\$100	\$700	\$2,200	\$7,900	\$55,400
Checking	\$6,969	\$100	\$500	\$1,400	\$3,800	\$23,100
Saving	\$6,476	\$0	\$0	\$400	\$2,500	\$25,200
Money Market	\$12,076	\$0	\$100	\$900	\$7,700	\$57,400
C.D.	\$12,734	\$0	\$0	\$500	\$4,000	\$39,200

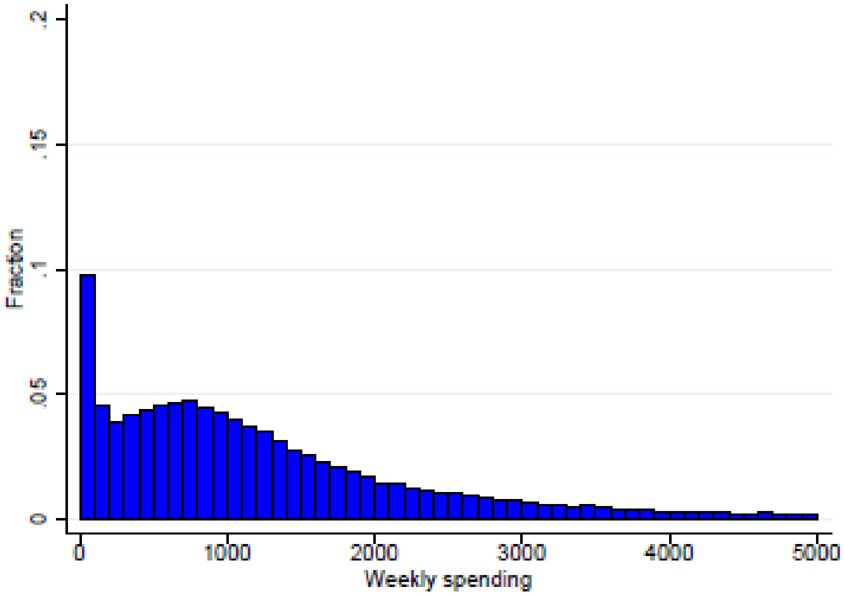
Panel (b): Credit Card	Mean	P_5	P_{25}	P_{50}	P_{75}	P_{95}
Balance	\$7,228	\$200	\$1,400	\$3,600	\$8,500	\$26,100
Credit Limit	\$23,019	\$800	\$4,200	\$11,900	\$29,500	\$81,800
Utilization Ratio	0.48	0.02	0.15	0.45	0.78	1.00
Revolving Debt	\$5,828	\$1,200	\$2,100	\$3,500	\$6,700	\$18,000
APR	18.46%	10%	15%	18%	23%	27%

Measuring income and expenditures from transactions

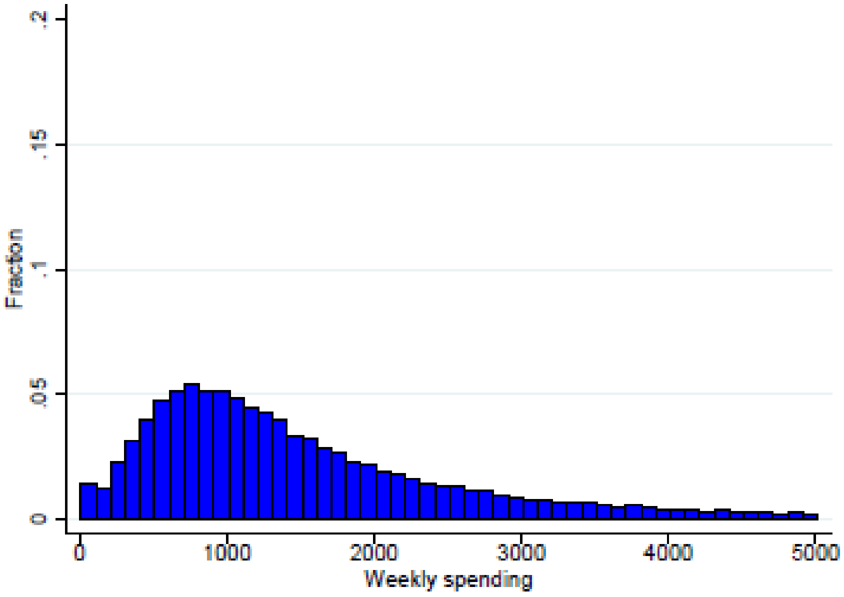


Weekly Spending: Total

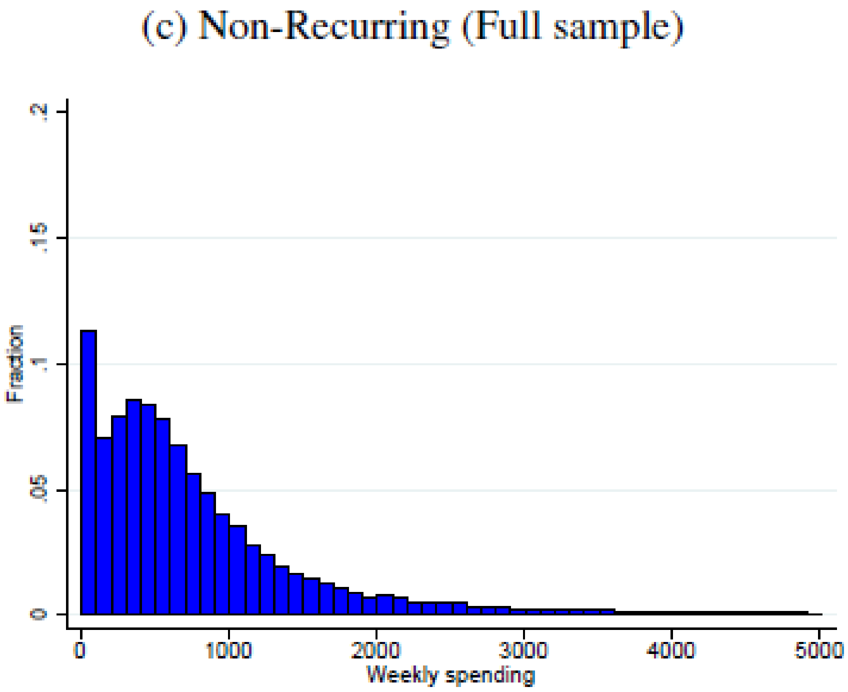
(a) Total (Full sample)

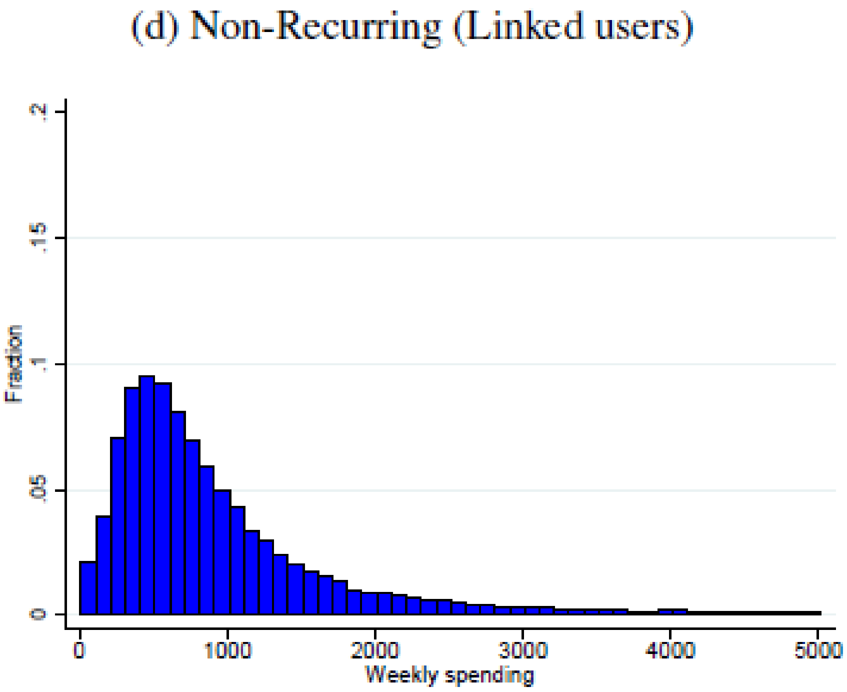


(b) Total (Linked users)



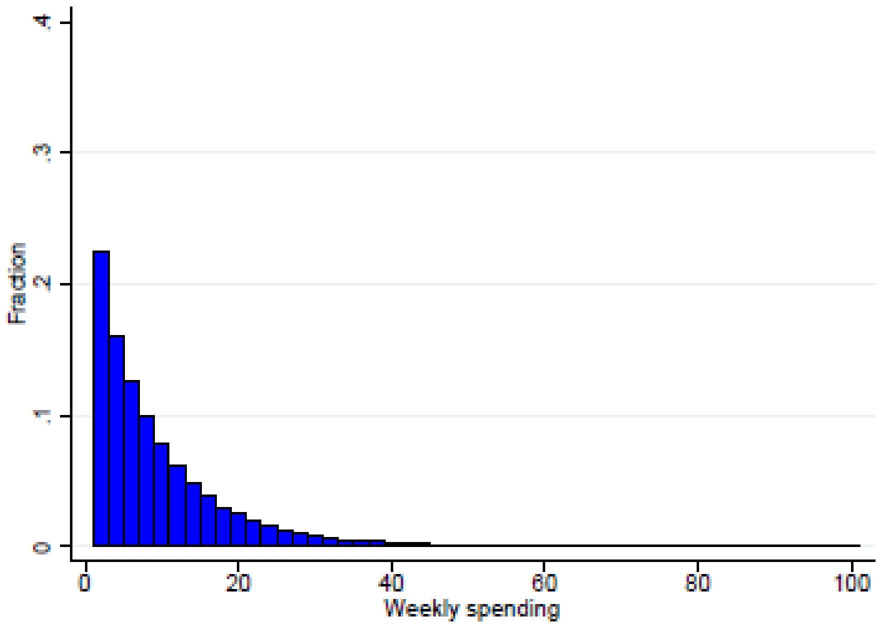
Weekly Spending: Non-Recurring



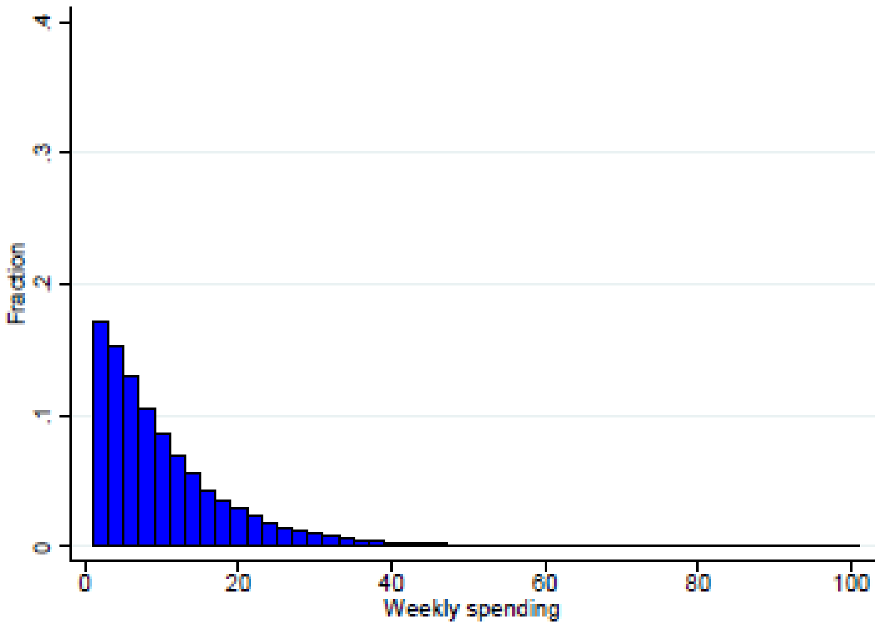


Weekly Spending: Coffee Shop and Fast Food

(e) Fast Food and Coffee Shop (Full sample)



(f) Fast Food and Coffee Shop (Linked users)



Using the data to answer macroeconomic questions

What is the MPC from predictable income, from unpredictable income?

- Does the standard lifecycle/permanent income fit the data?
- What MPC should be used to calibrate models?
- How will households respond to economic stimulus payments such as tax rebate?

Using the data to answer macroeconomic questions

Many individuals have *very low liquidity*.

- Are they vulnerable to shocks?
- How do they manage an adverse shock, e.g., a car repair or temporary income loss?
- What buffers consumption for those with low cash-on-hand?
- Is low cash-on-hand a mistake?

Using the data to answer macroeconomic questions

Application 1:

How does spending respond to <u>predictable, recurring income</u> (e.g., paycheck)?

Application 2:

How does spending respond to a <u>temporary shock to liquidity</u> (government shutdown)?

Application 1: Excess sensitivity to paycheck and Social Security

Data allow

- Identification of payments
- Estimate of response by
  - Type novel classification of spending
  - By liquidity, etc

Economic Model

Paycheck/Social Security predictable, so receipt of income should not affect timing of spending

Econometric model

$$x_{ict} = \sum_{j=Mon.}^{Sun.} \delta_{jc} + \sum_{k=-7}^{6} \beta_{kc} I_i \left(Paid_{t-k} \right) + \varepsilon_{ict},$$

x_{ict} = daily spending/average daily spending,

i=individual, c=type of spending, t=time

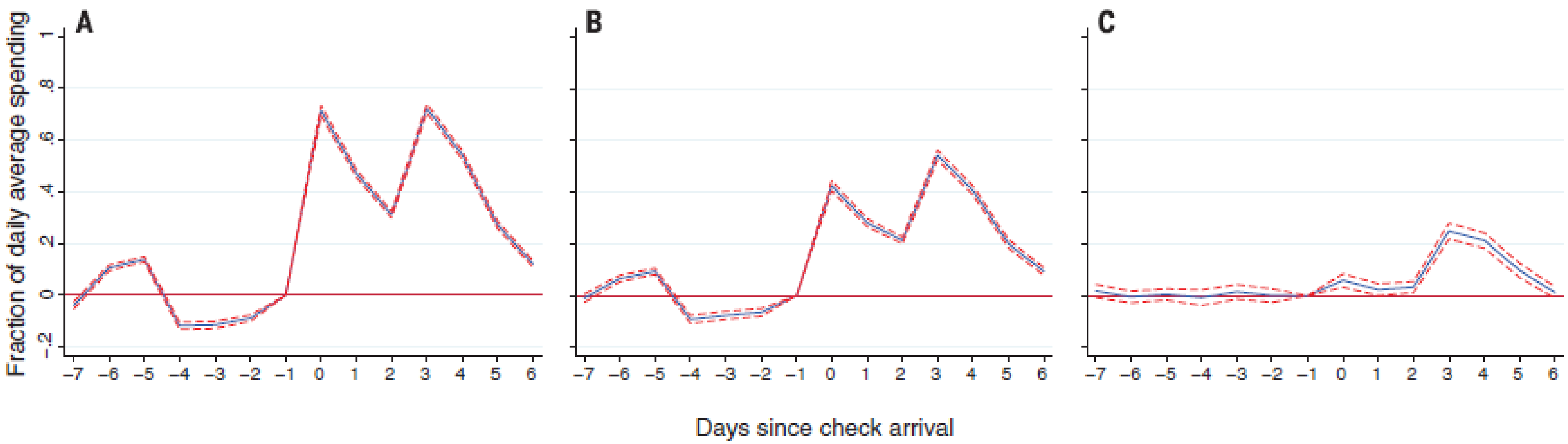
$\mathrm{I}_i(Paid_{t-k})$ = dummy for getting paid on date t - k

Response of spending to paycheck

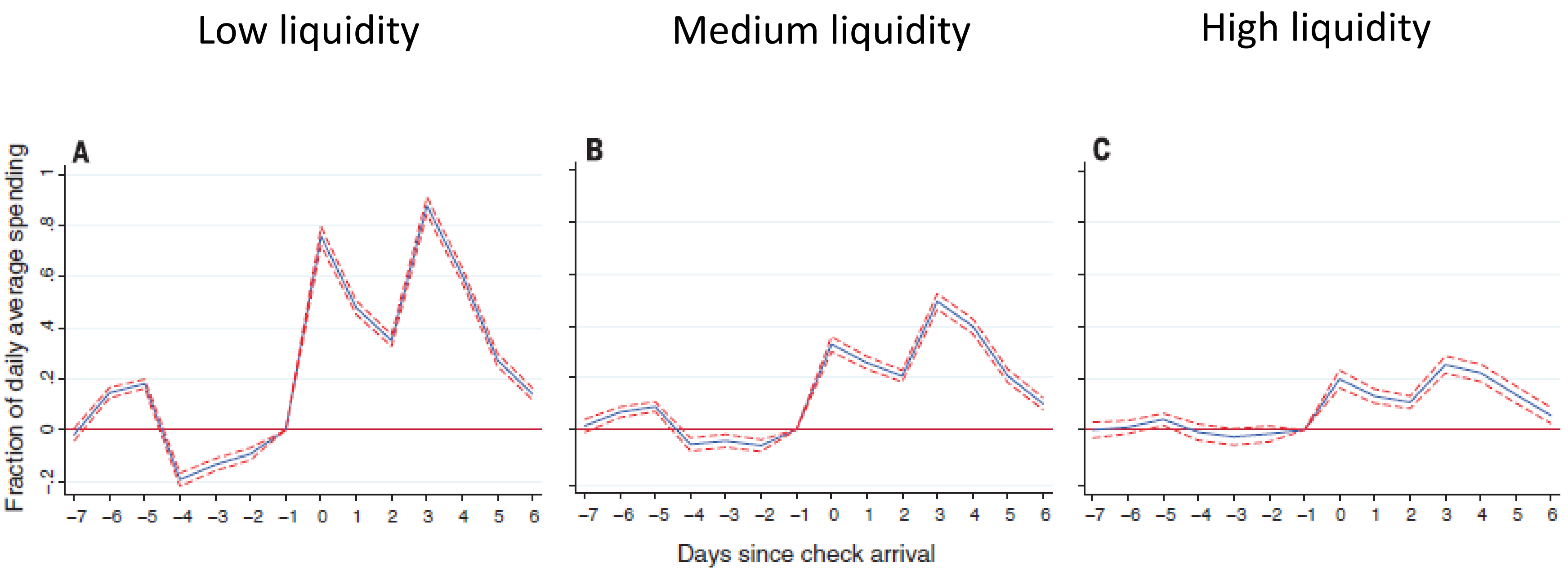
Total

Non-recurring

Coffee shop and
fast food



Response of spending to paycheck
Non-recurring spending by liquidity



Application 1. Conclusions

- Much of high-frequency "excess sensitivity" owes to rational timing of payments
- Low liquidity individuals do display some excess sensitivity

Application 2:
2013 US government shutdown

Workers subject to shutdown

- lost 40% of pay in one pay period
- reimbursed in next pay period

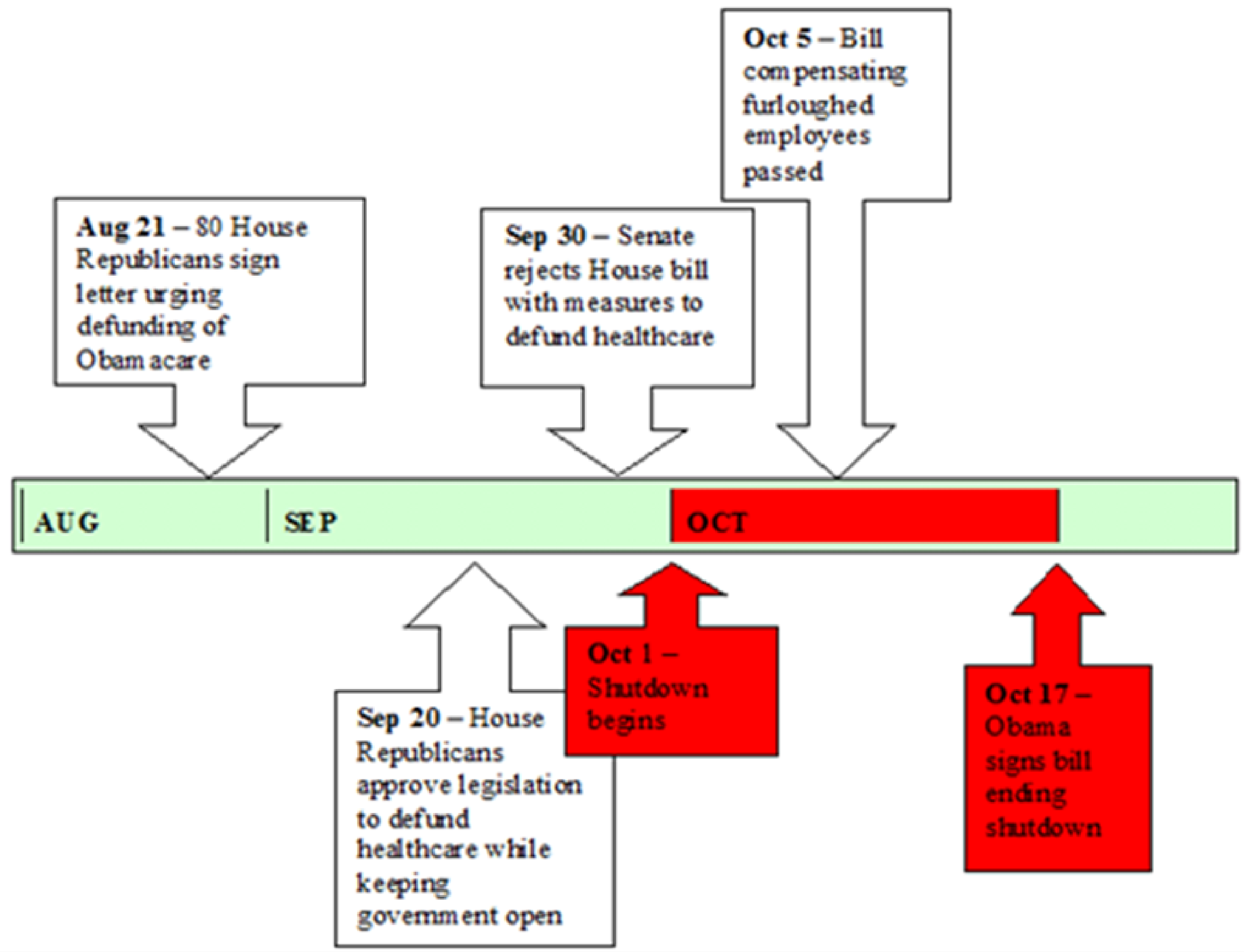
Distinctive experiment:

Timing of income only

- Liquidity shock
- No wealth effects

Government Shutdown of Oct 2013

	Sunday	Monday	Tuesday	Wednes day	Thursday	Friday	Saturday
Pay period	Sept 22	23	24	25	26	27	28
	29	30	~~Oct 1 Shutdown Begins~~	~~2~~	~~3~~	~~4~~	5
Pay period	6	7	8	9	10 First pay date affected by shutdown	11	12
	13	14	15	16	17 Shutdown Ends	18	19
	20	21	22	23	24 Typical pay date after shutdown	25	26


Oct 5 – Bill compensating furloughed employees passed
Aug 21 – 80 House Republicans sign letter urging defunding of Obamacare
Sep 30 – Senate rejects House bill with measures to defund healthcare
AUG
SEP
OCT
Oct 1 – Shutdown begins
Sep 20 – House Republicans approve legislation to defund healthcare while keeping government open
Oct 17 – Obama signs bill ending shutdown

Treatment and control

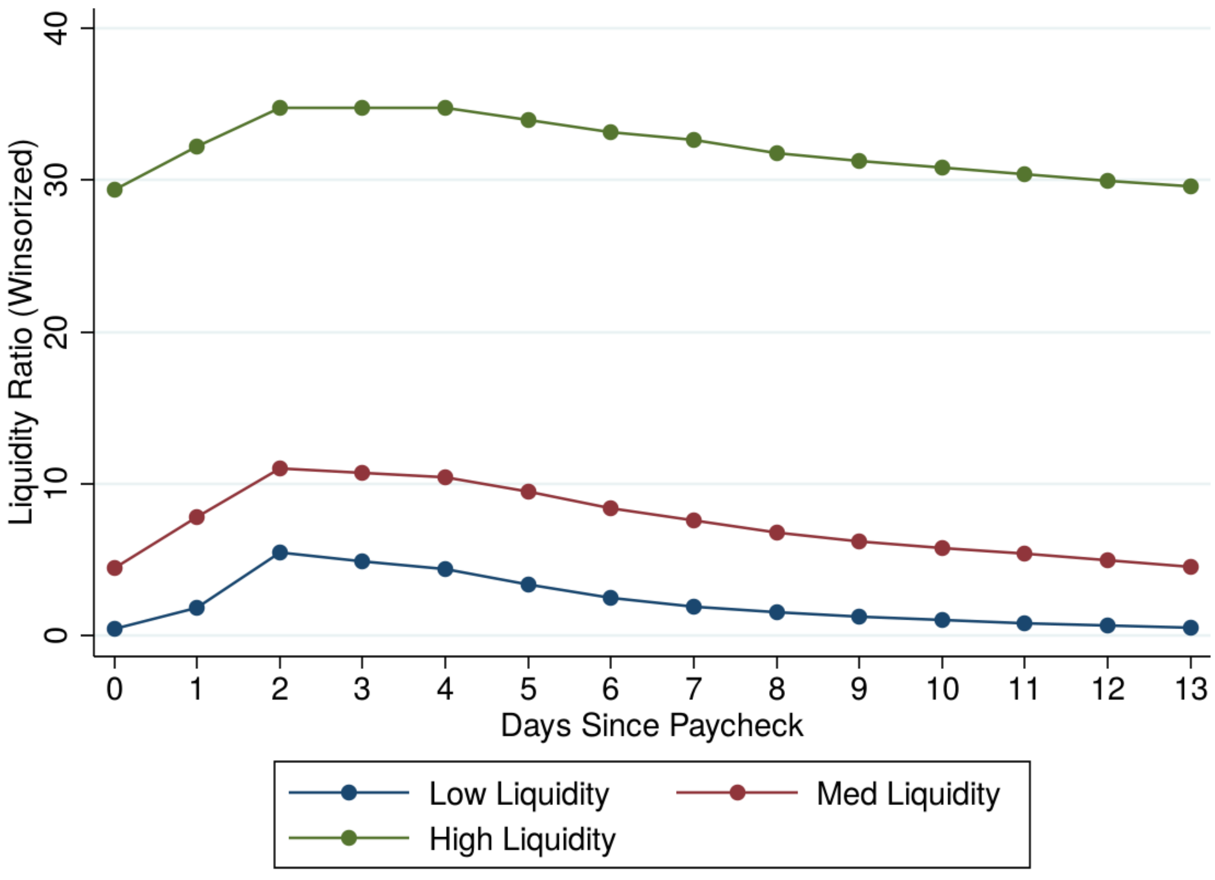
Treatment

- Federal worker (paycheck memo), and
- Decline in paycheck consistent with shutdown

Control

- Other worker on same biweekly pay schedule as government

Pre-Shutdown Median Liquidity over the Paycheck Cycle

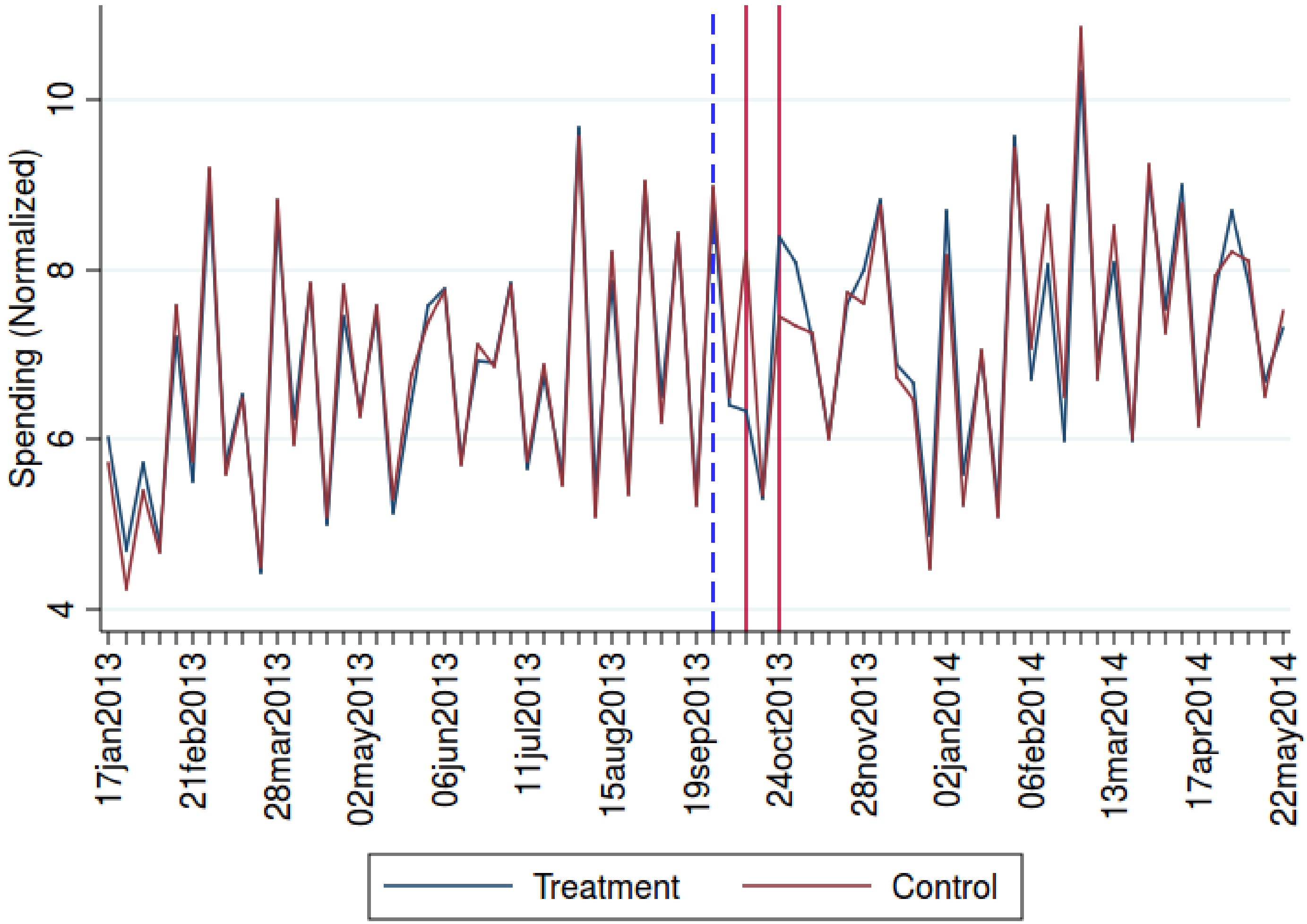


“Seasonal” interactions

- Day of week
  - Spending
  - Clearing of payments
- Beginning of pay period effects
- Seasonal/holiday/macro effects

→ Having controls with same pay schedule valuable

Average weekly spending



Diff-in-diff effects of shutdown

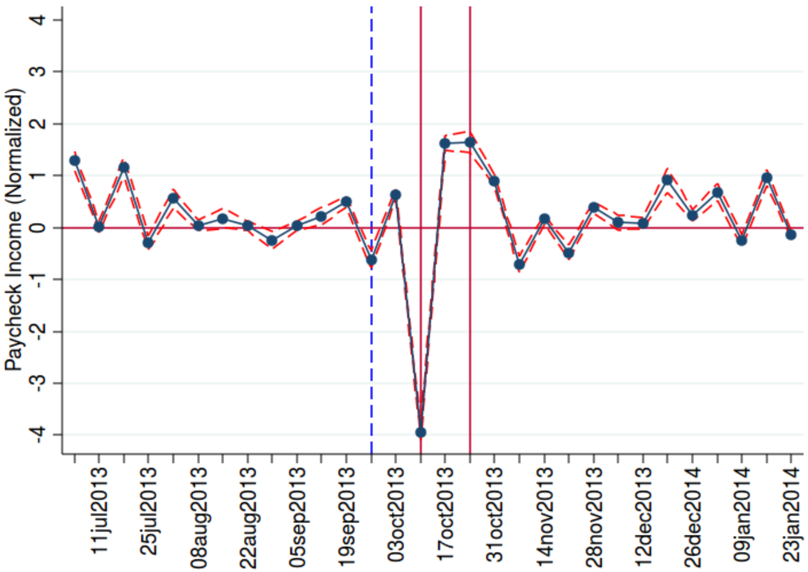
$$y_{i,t} = \sum_{k=1}^{T} \delta_k \times Week_{i,k} + \sum_{k=1}^{T} \beta_k \times Week_{i,k} \times Shut_i + \Gamma' X_{i,t} + \varepsilon_{i,t}$$

Specification:

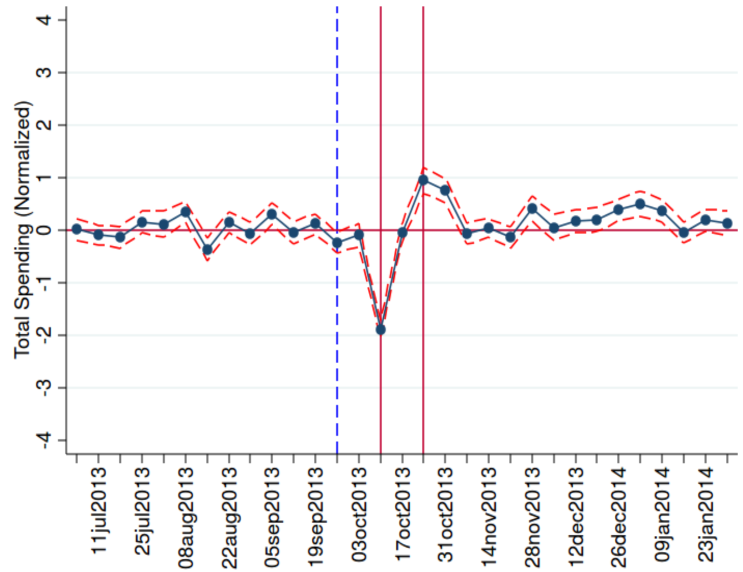
LHS = variable of interest (income, category of spending)
Normalized by average individual spending (daily rate)

Diff-in-diff effects of shutdown

Paycheck income

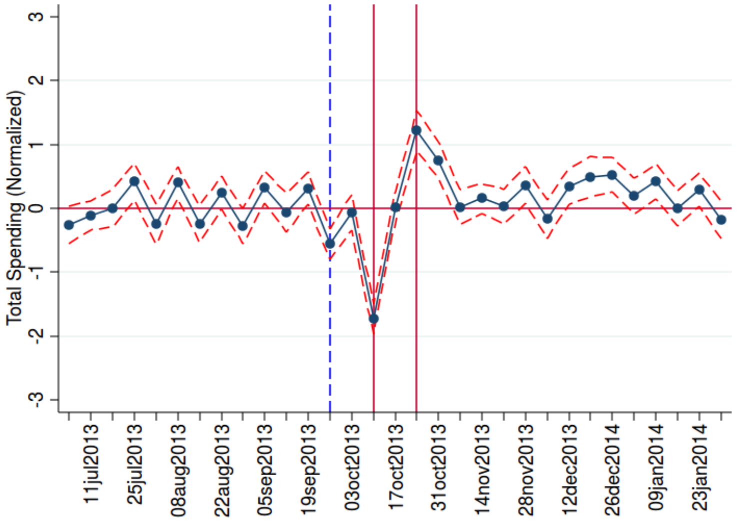


Total spending

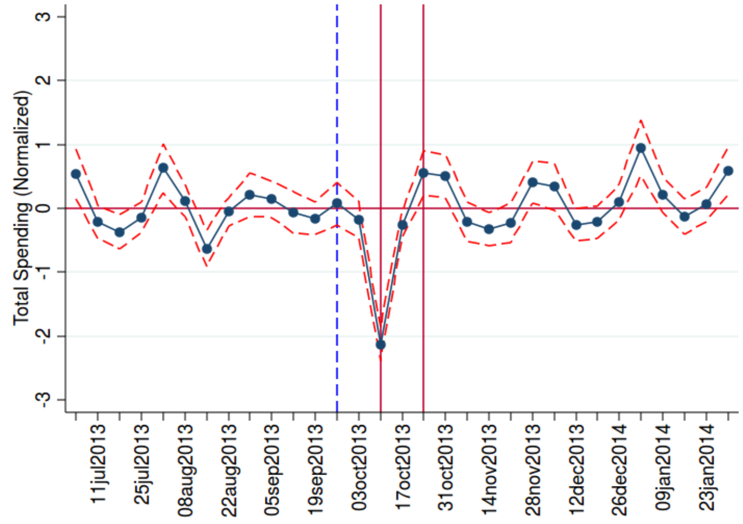


Diff-in-diff effects of shutdown

Non-recurring spending

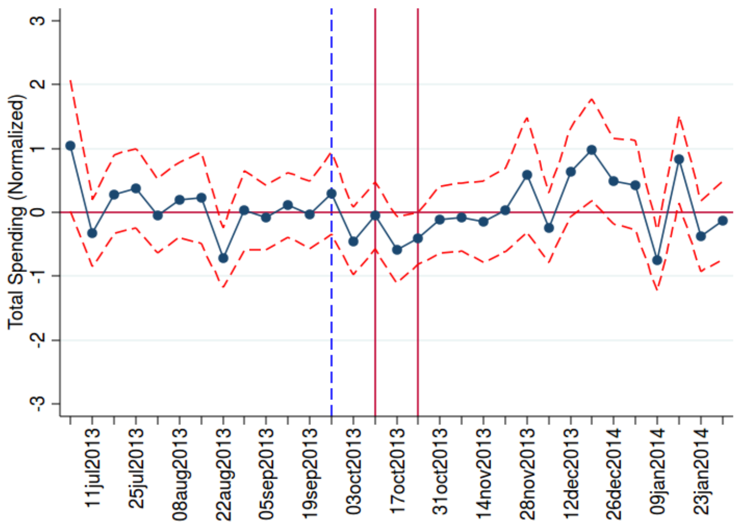


Recurring spending

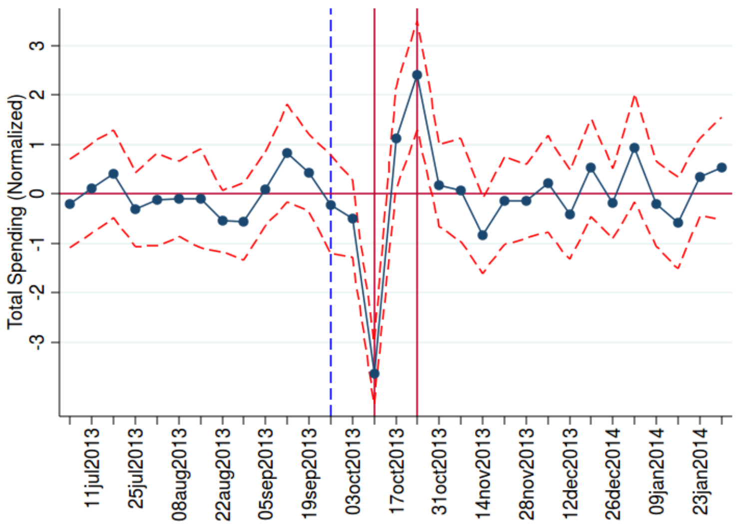


Diff-in-diff effects of shutdown

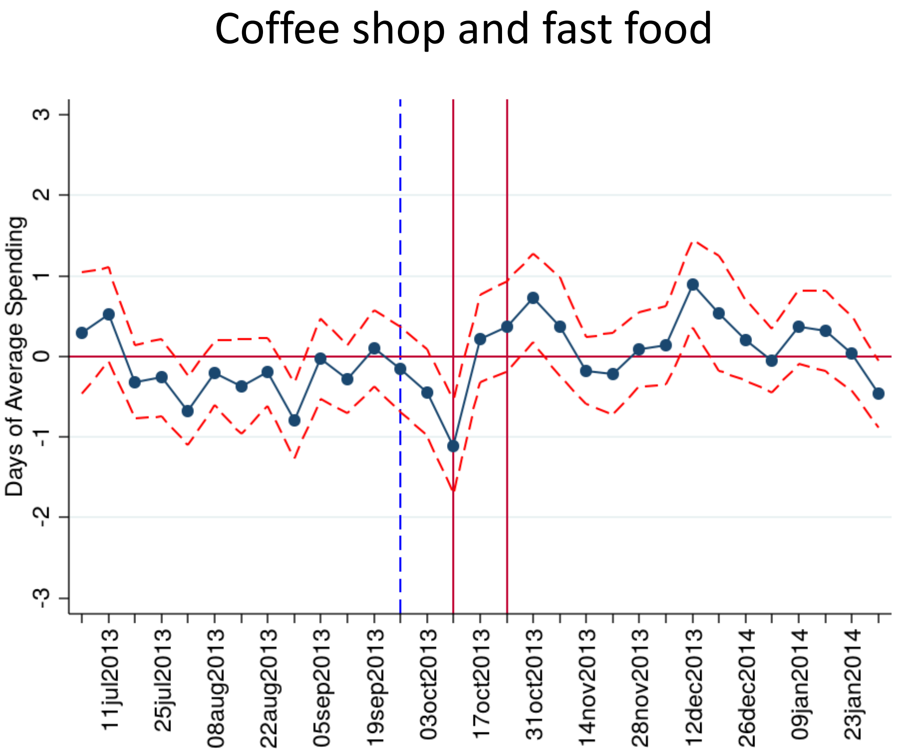
Credit card spending



Credit card balance payments

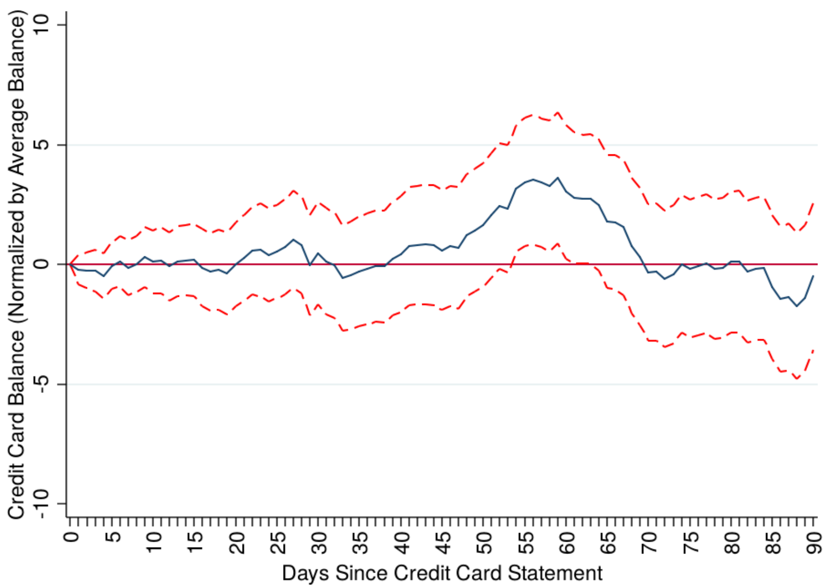


Diff-in-diff effects of shutdown

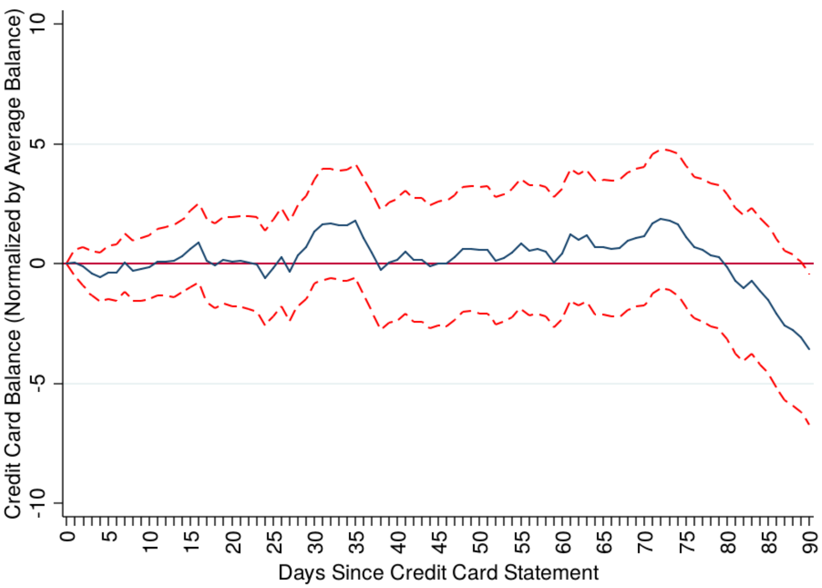


Diff-in-diff effects of shutdown, Credit Card Balance by "Liquidity Risk"

Credit Card Balance,
Accounts at Liquidity Risk



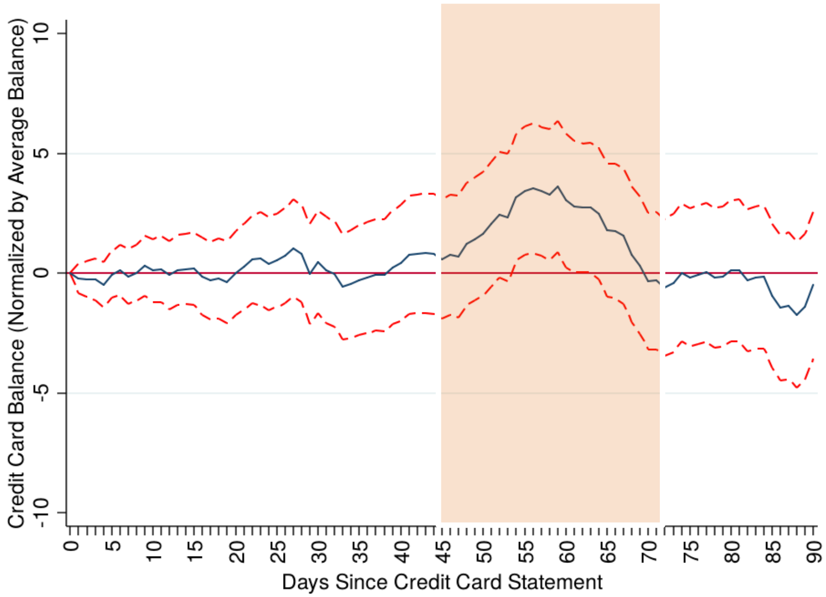
Credit Card Balance,
Accounts Not at Liquidity Risk



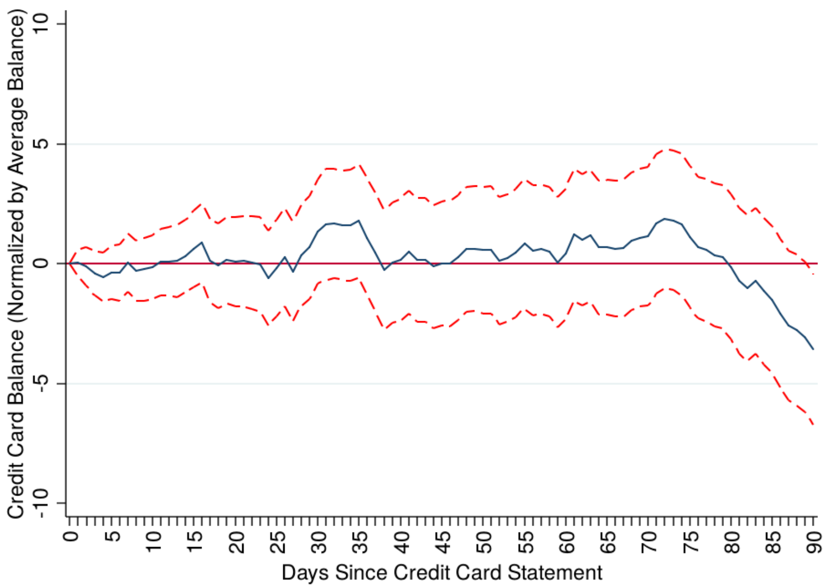
Notes: Horizontal axis is days since August 2013
Revolvers only
Data are by individual credit card account levels
"At risk" accounts have payment due dates in pay period affected by shutdown

Diff-in-diff effects of shutdown, Credit Card Balance by "Liquidity Risk"

Credit Card Balance,
Accounts at Liquidity Risk



Credit Card Balance,
Accounts Not at Liquidity Risk



Notes: Horizontal axis is days since August 2013
Revolvers only
Data are by individual credit card account levels
"At risk" accounts have payment due dates in pay period affected by shutdown

Lessons from shutdown

Puzzle for standard models:

- Very sizeable spending response to a two-week delay in pay

Success for standard models:

- Rearrangement of payments, not consumption

New data essential