Early-Stage Business Formation: An Analysis of Applications for Employer Identification Numbers

NBER-CRIW
Boston, July 17, 2017
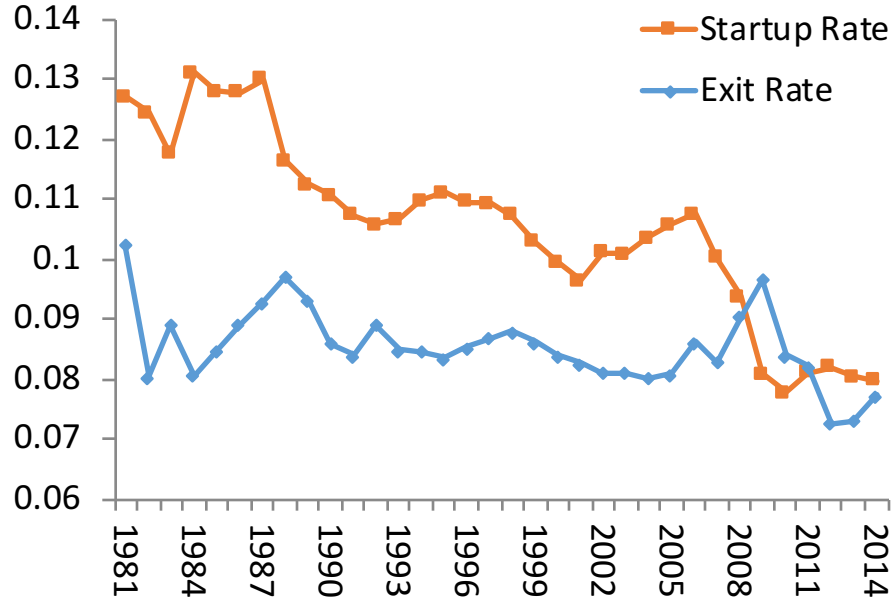
By

Kimberly Bayard	Federal Reserve Board
Emin Dinlersoz	US Census Bureau
Timothy Dunne	Federal Reserve Bank of Atlanta
John Haltiwanger	University of Maryland
Javier Miranda	US Census Bureau
John Stevens	Federal Reserve Board

Startup and Exit Rates in Nonfarm Private Sector, 1981-2014

Startup and Exit Rates

- Startup Rate
- Exit Rate

From the US Census Bureau Business Dynamics Statistics

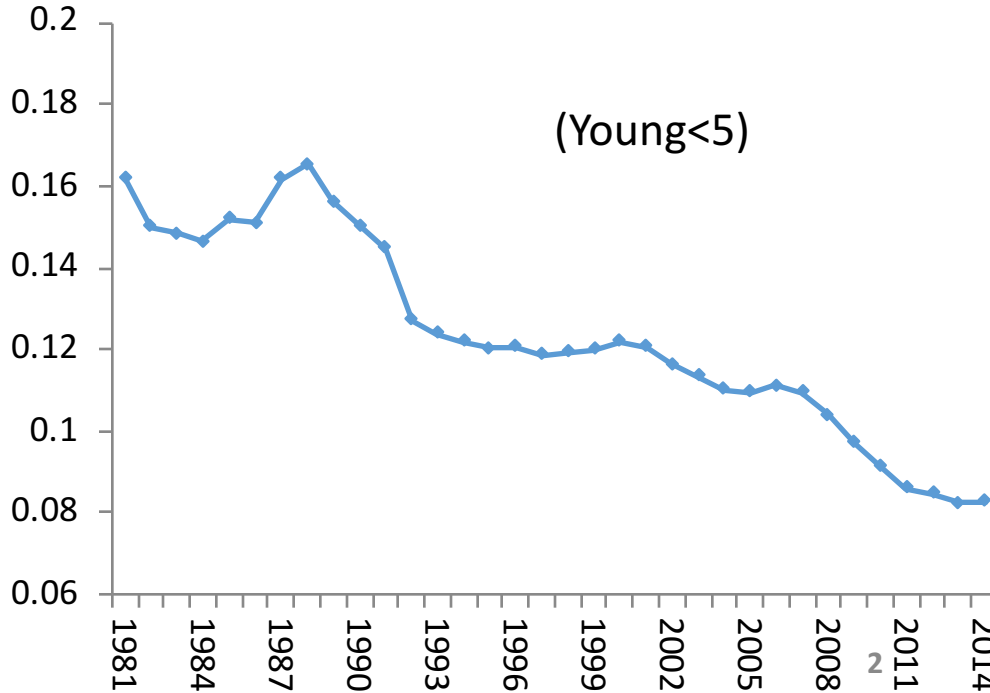## Share of Employment for Young Firms, 1981-2014, Nonfarm Private Sector

(Young<5)

Startup and Exit Rates in Nonfarm Private Sector, 1981-2014
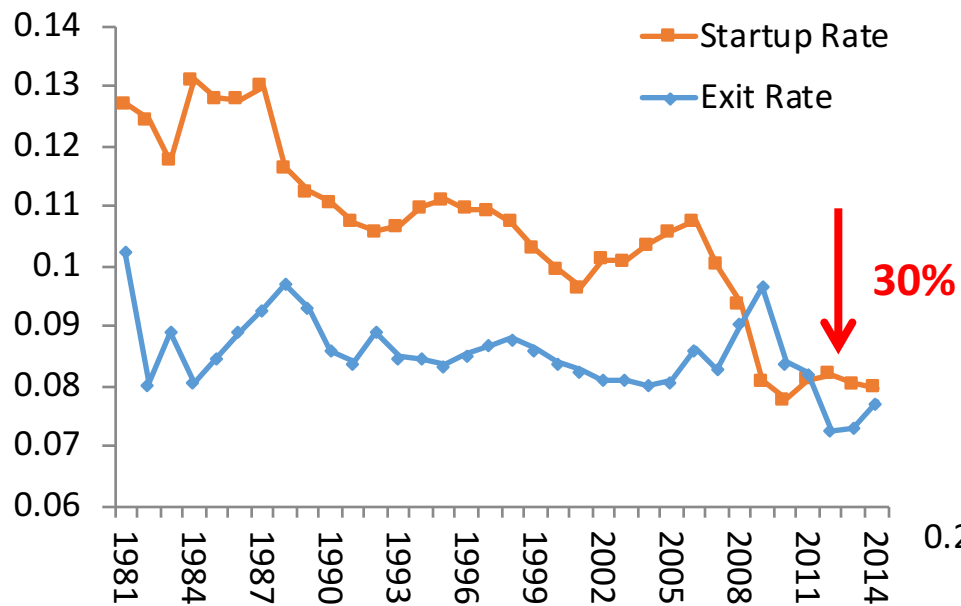
Startup and Exit Rates

From the US Census Bureau Business Dynamics Statistics

30%

Share of Employment for Young Firms, 1981-2014, Nonfarm Private Sector

(Young<5)

Startup and Exit Rates in Nonfarm Private Sector, 1981-2014

Startup and Exit Rates

From the US Census Bureau Business Dynamics Statistics

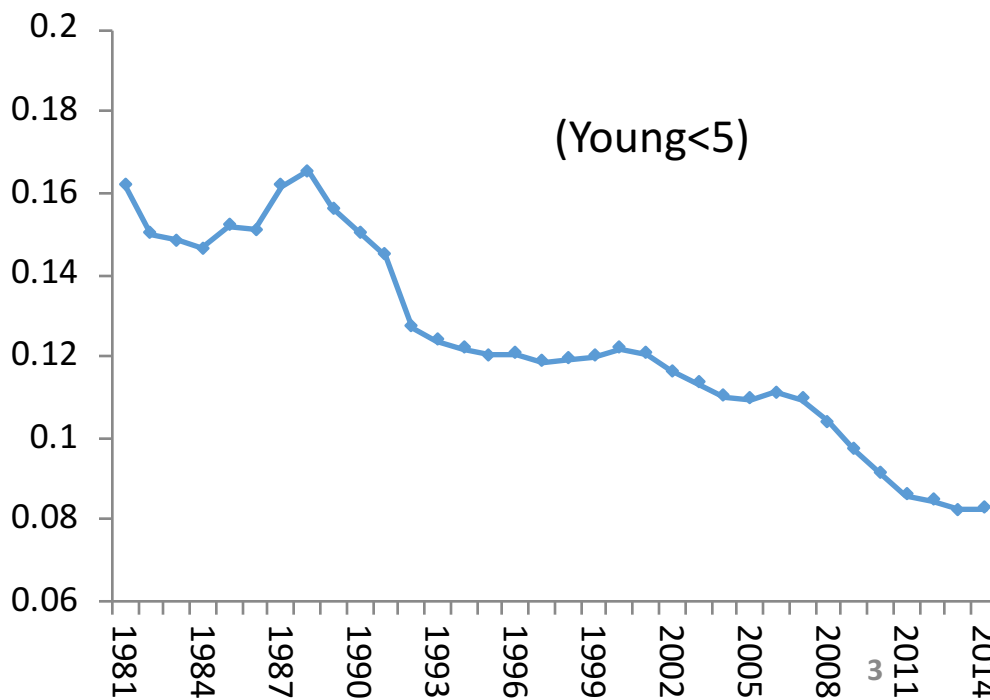**Share of Employment for Young Firms, 1981-2014, Nonfarm Private Sector**

(Young<5)

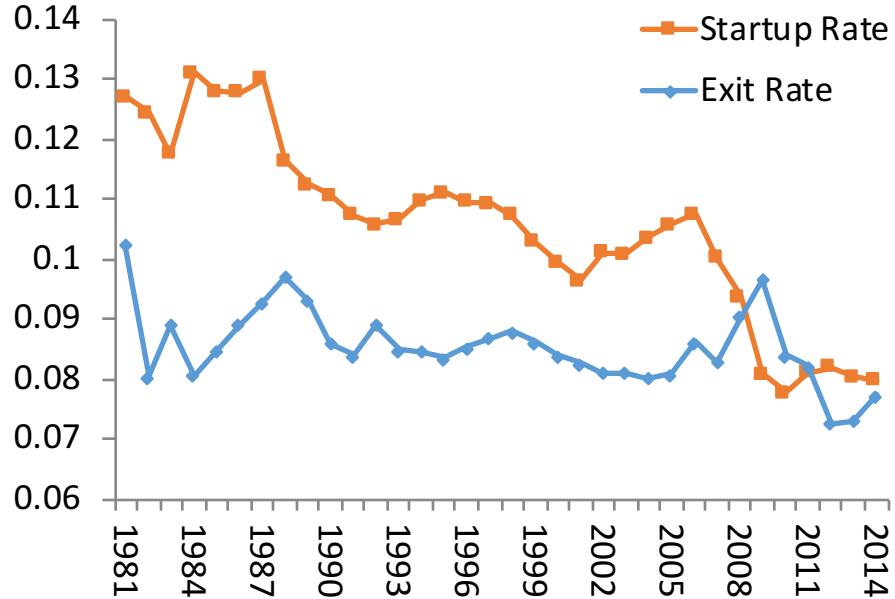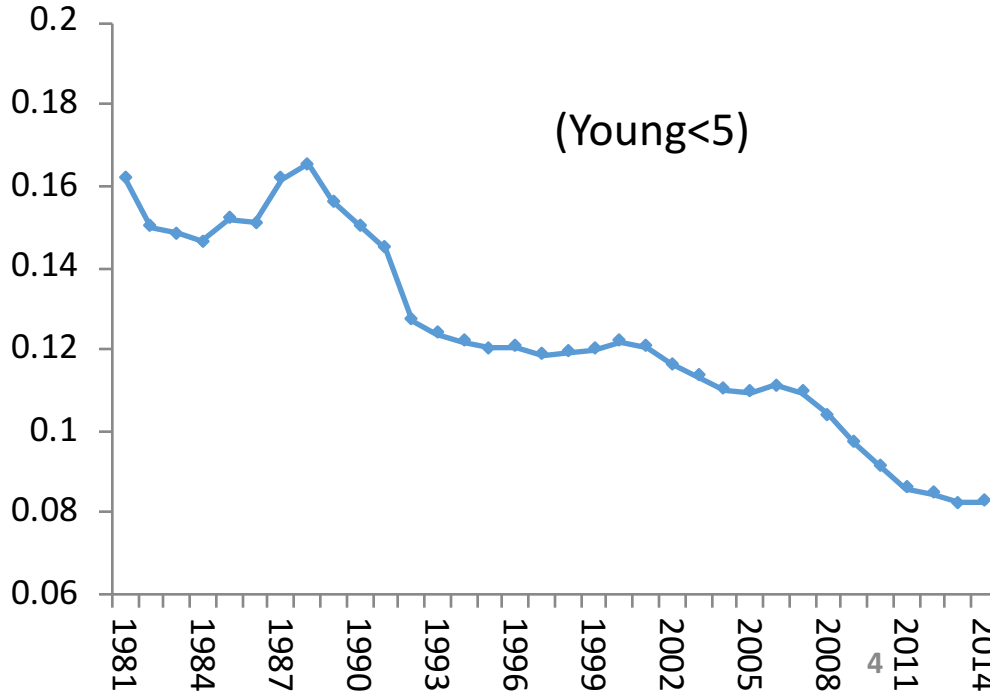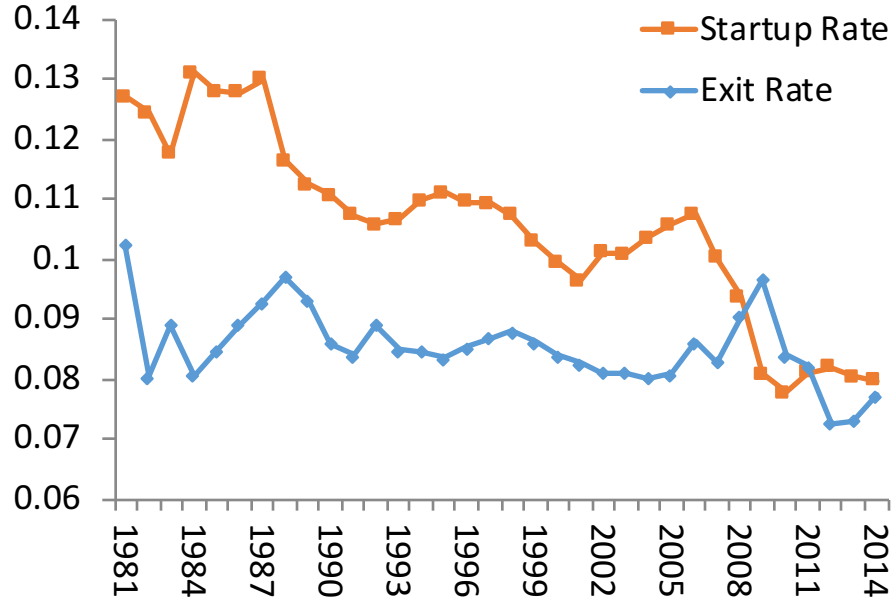**Startup and Exit Rates in Nonfarm Private Sector, 1981-2014**

Startup and Exit Rates

From the US Census Bureau Business Dynamics Statistics

Share of Employment for Young Firms, 1981-2014, Nonfarm Private Sector

(Young<5)

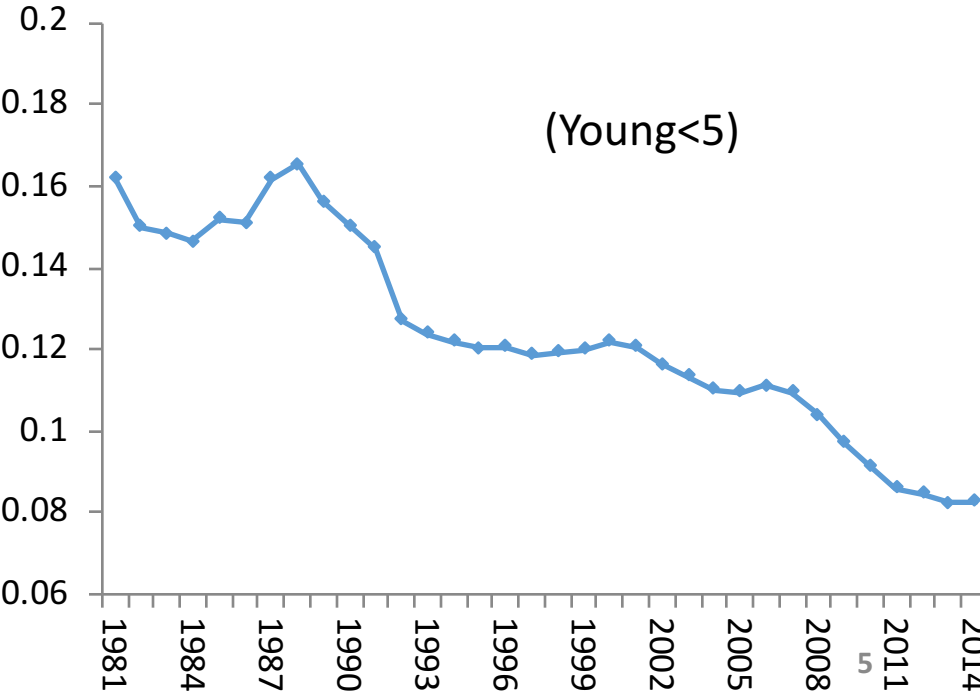Can we use high frequency administrative data to create timely (perhaps early) indicators of economic activity at the nation and local level?

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Data: Applications to New EINs

- Applications for Employer Identification Numbers (EINs)
  - Unique 9-digit numbers assigned by IRS to persons/entities for <u>tax purposes</u>
  - Most EIN applications made for business-related purposes
    - pay taxes (LLC, Partnerships, Corporations), Sole props with employees/purchase businesses/change LFO…
  - But many other non-business reasons (noisy):
    - estates, trusts, tax liens, REMIC, retirement and health care plan administration, non-profit and gov. organizations, etc…
  - Most business related applications never turn into businesses

- High Frequency
  - EIN applications transmitted continuously on a weekly basis to Census Bureau

- Contains valuable information on business intent and business characteristics
  - Name, address (mail/business), principal activity (sector box/write in), LLC, type of entity (15+ options), reason for application (8+ options), previous EIN, Wages paid,

- Coverage:
  - Geography: Nationwide, state, county, zip-code, census tract (80%)
  - Business: All business except sole proprietors without employees…
 - Years: 2004q3-present => 45.8 million applications

Questions

- Do EIN applications provide *useful information* about early-stage entrepreneurial activity?

- Can EIN applications and their characteristics be used to *predict new employer business formation* in a timely and geographically granular way?

- Can they serve as *economic indicators*?

- Are EIN applications correlated with *national and local* economic conditions?

United States™ **Census** Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Identifying Business Applications

- Filter all EIN applications to obtain core "Business Applications"
  - Exclude: *tax liens, estates, trusts, public entities, retirement/health plans, mortgage pools; some agricultural, certain financial, and private household businesses; missing geography, outside of 50 states and DC..* (72% of apps) [18% chance]
- Identify some important subsets of business applications:
  - High-Propensity Business Applications (1/2 of all business apps)
    - Business Applications that have high probability of turning into businesses with payroll based on application characteristics [30% chance]
    - Planned date for first wage payments/hiring employees/purchase business/change org. type/corporations/certain industries: mfg., retail, health care, restaurants
  - Business Applications for a Corporation (24% of apps) [33% chance]
  - Business Applications with Planned Wages (indicate a planned date for first wage payments; 24% of apps) [43% chance]

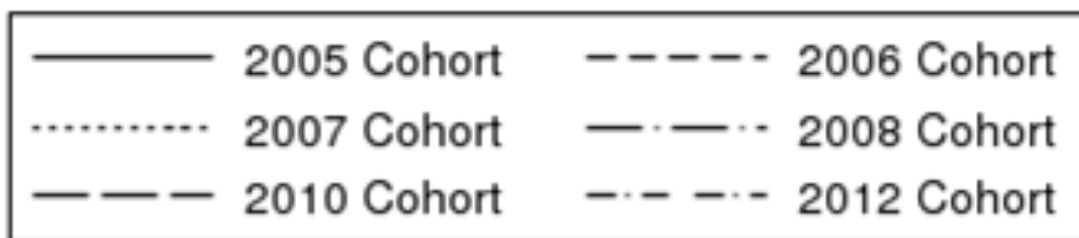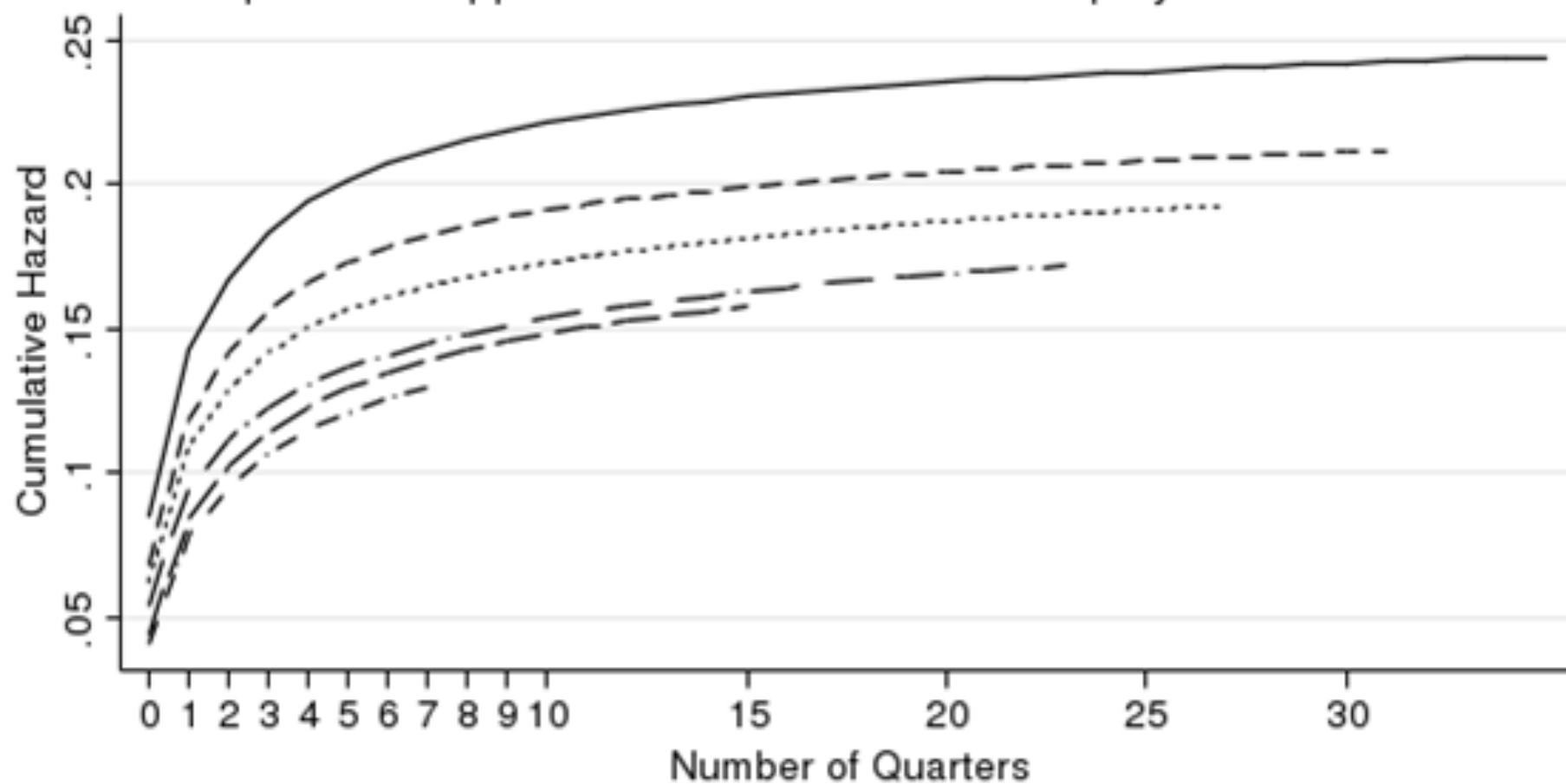EIN Applications

2005-2017: First quarter only
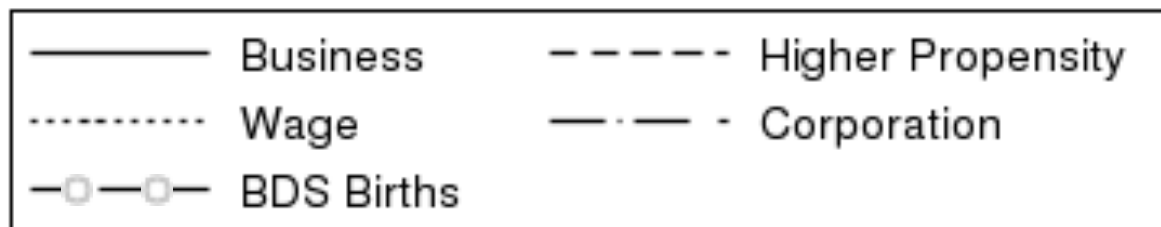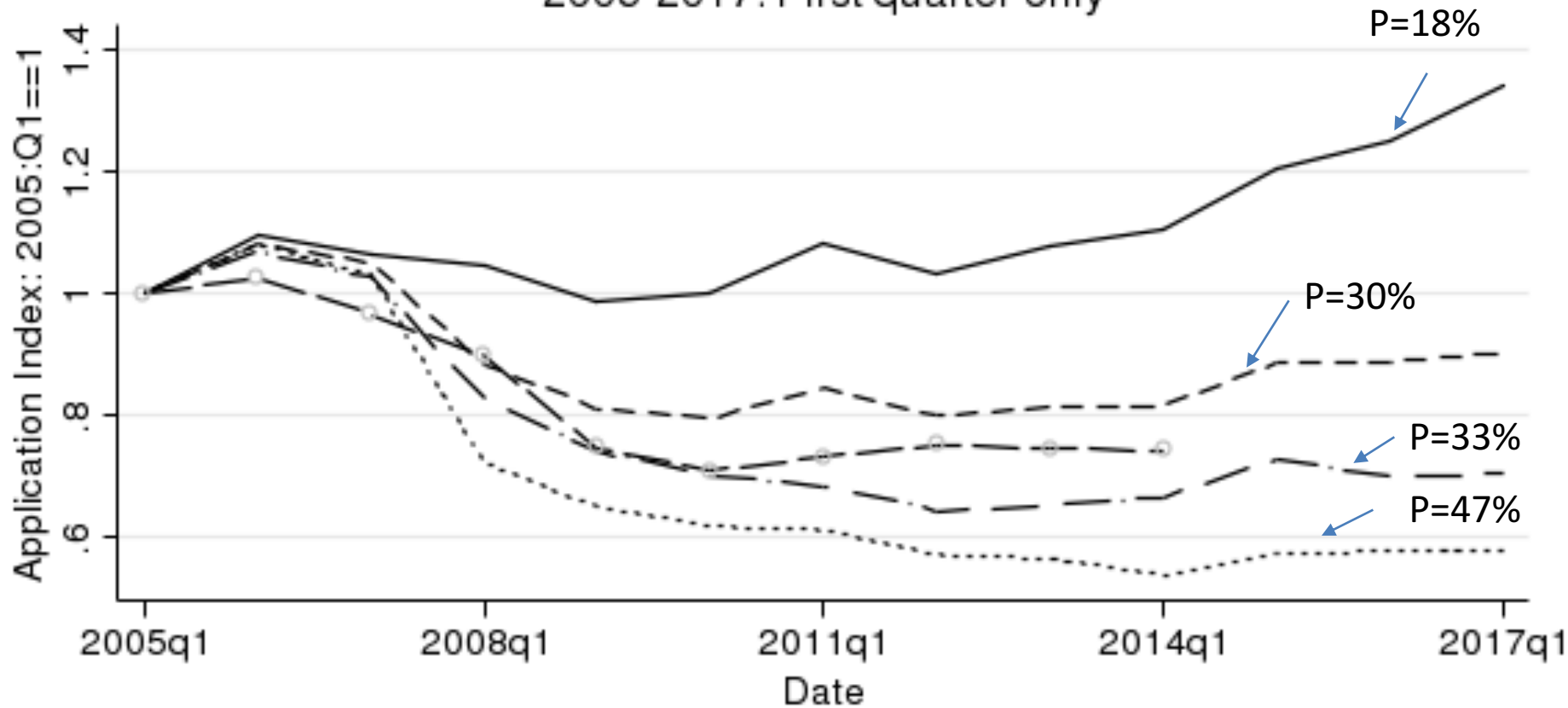
Source: EIN Applications File, US Census Bureau.

Identifying Business Formations

- Track their employment activity through the *US Census Bureau Business Register* (BR).
    - List of employer business from IRS payroll filings (941/944). Employers are required to report wages, tips and other compensation quarterly. When the first employee is hired/payroll tax paid, the EIN enters the administrative data for employer businesses (Business Register - BR) => quarterly frequency activity data
    - Use the incidence and timing of first appearance of the EIN in the BR to identify employer "Business Formation" from an EIN application (the first quarter with payroll for each startup EIN)
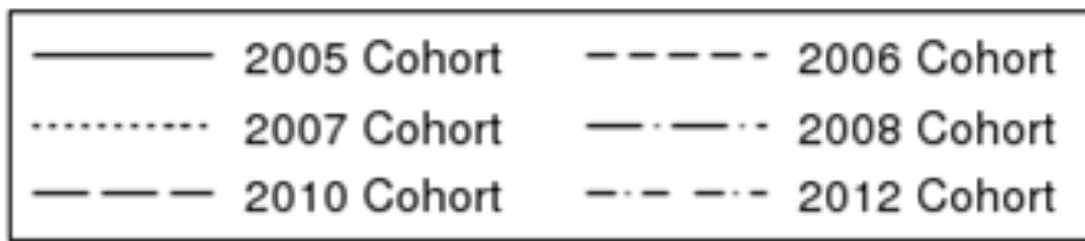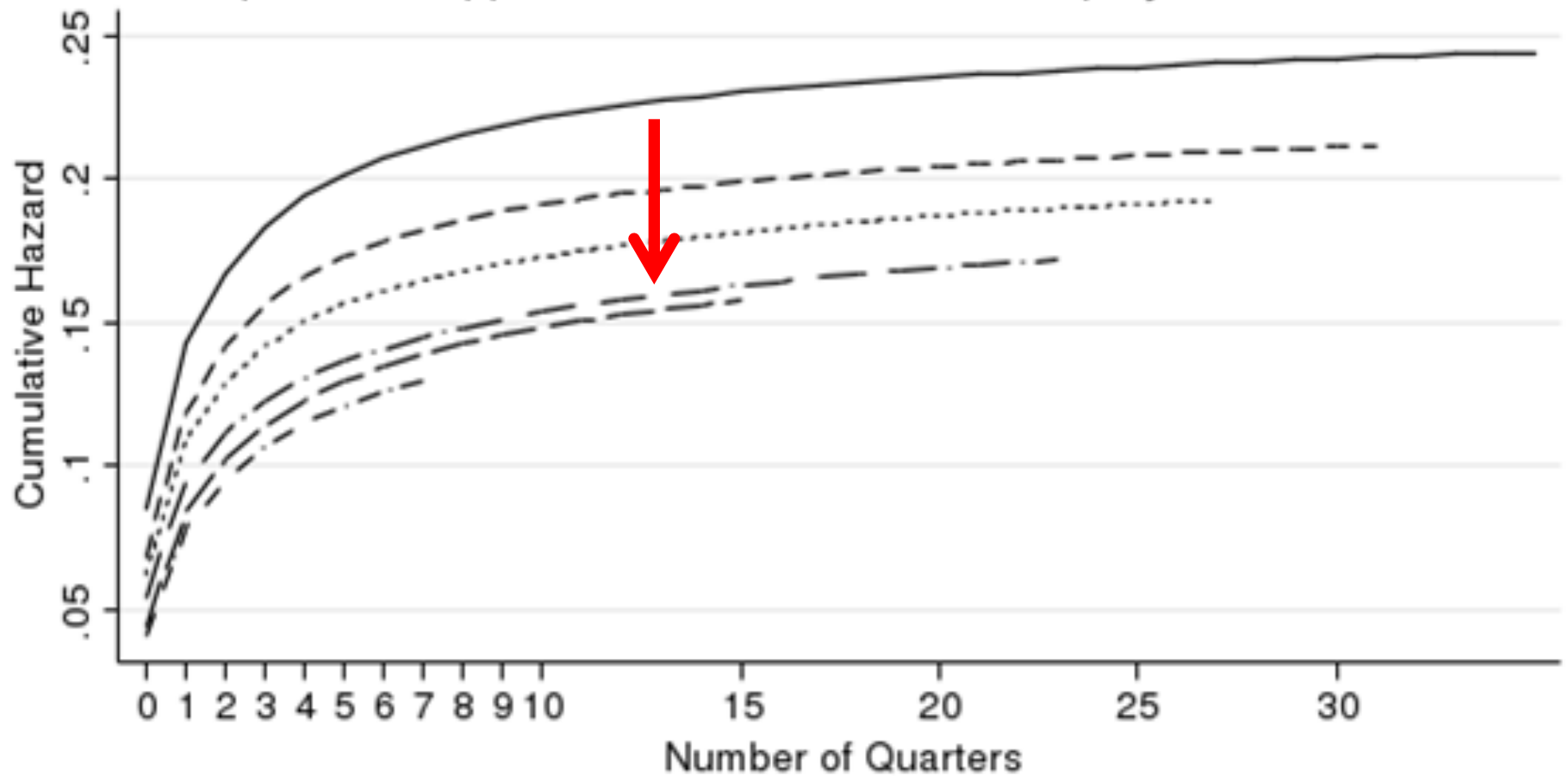   - EIN applications from 2004q3 onwards are matched to payroll observations 2004q3-2014q4 in BR (last year BR=2015)

Cumulative Transitions by Application Cohort
Proportion of applications that transition to employer business status

Cumulative Hazard (y-axis): .05, .1, .15, .2, .25

Number of Quarters (x-axis): 0 1 2 3 4 5 6 7 8 9 10 15 20 25 30

Legend:
— 2005 Cohort
– – – 2006 Cohort
······· 2007 Cohort
–·–·– 2008 Cohort
– – – 2010 Cohort
–··–··– 2012 Cohort

Cumulative Transitions by Application Cohort
Proportion of applications that transition to employer business status

Y-axis: Cumulative Hazard (.05, .1, .15, .2, .25)
X-axis: Number of Quarters (0 1 2 3 4 5 6 7 8 9 10 15 20 25 30)

Legend:
—— 2005 Cohort
– – – 2006 Cohort
········· 2007 Cohort
—··— 2008 Cohort
— — 2010 Cohort
—·—·· 2012 Cohort

Model

- Model the probability an application turns into an employer business
  - Forward looking: Does EIN application *i* received in quarter *q* turns into an employer business within the next 4 quarters

$$P_{iqt+k} = F(X'\beta)$$

- Type of entity	- ∃ Prior EIN
- Reason for application	- ∃ Trade name
- Week of application	- ∃ Executor name
- 6-Digit NAICS	- ∃ Business address
- Wage date	- Rich interactions
- LLC	

- *X* = application characteristics
- β = estimated probabilities

- We estimate both Probit and LPM
- Nation (with state effects) and fully interacted by state (25,000 params)
- With and without time trends and economic indicators
- Using different estimation samples (including/excluding pre recession)

- Use predicted probabilities to obtain the expected number of new business formations within the next 4 quarters (sum of probability weighted applications): nationwide, state, county, census tract…

Model Evaluation

- RMSD: to evaluate predictions out of sample (2013)
- No large differences between models but
  - More recent estimation window works better: 2010-2012 rather than longer or earlier
  - By state
  - With rich interactions
  - No time trends or auxiliary economic indicators
  - Week of application rather than quarter dummies
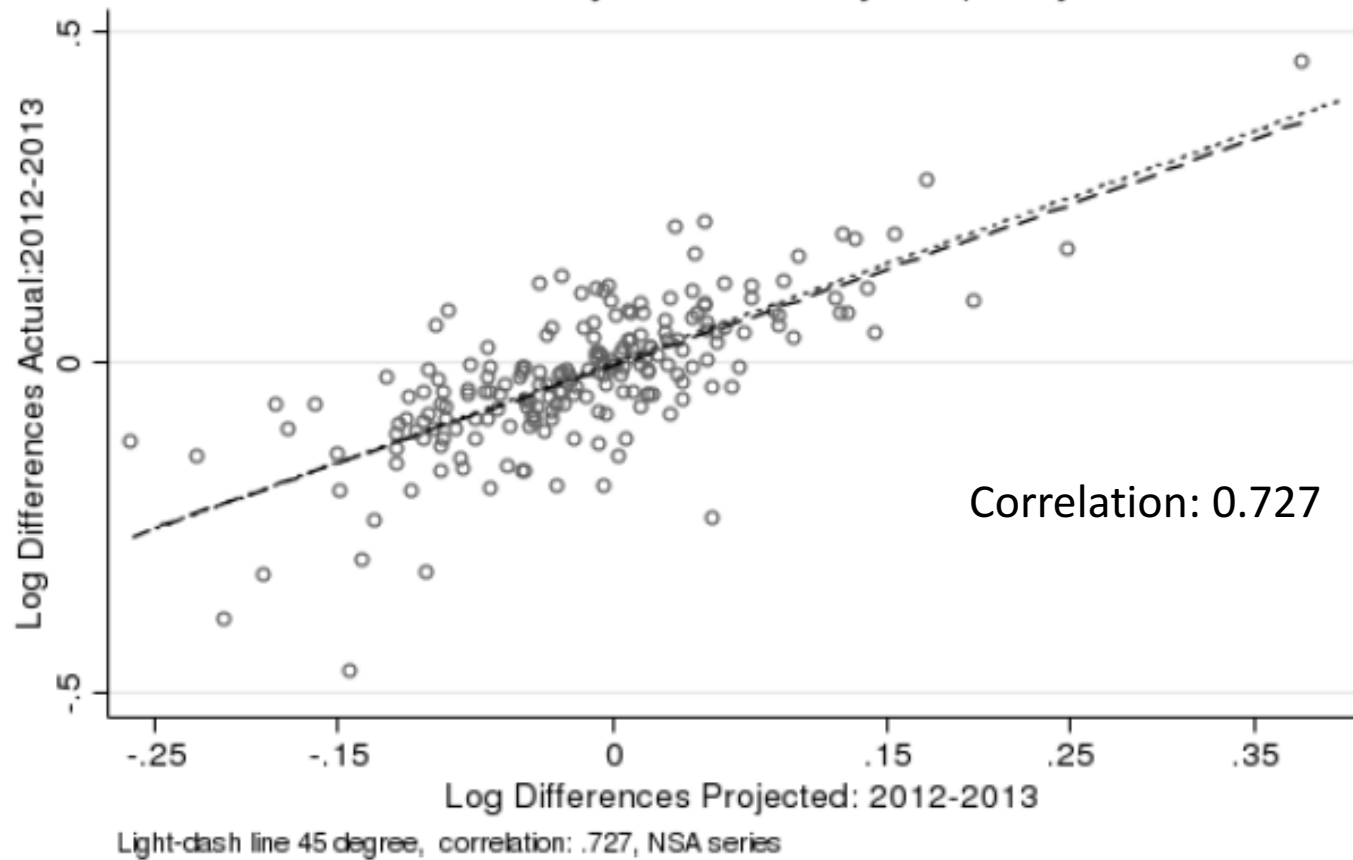 - Probit rather than LPM

Projected Business Formations

2006-2016

Out of sample
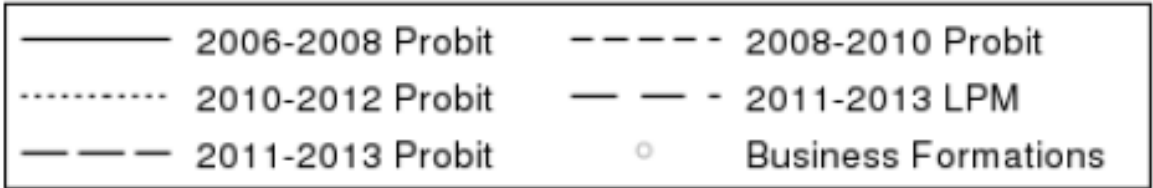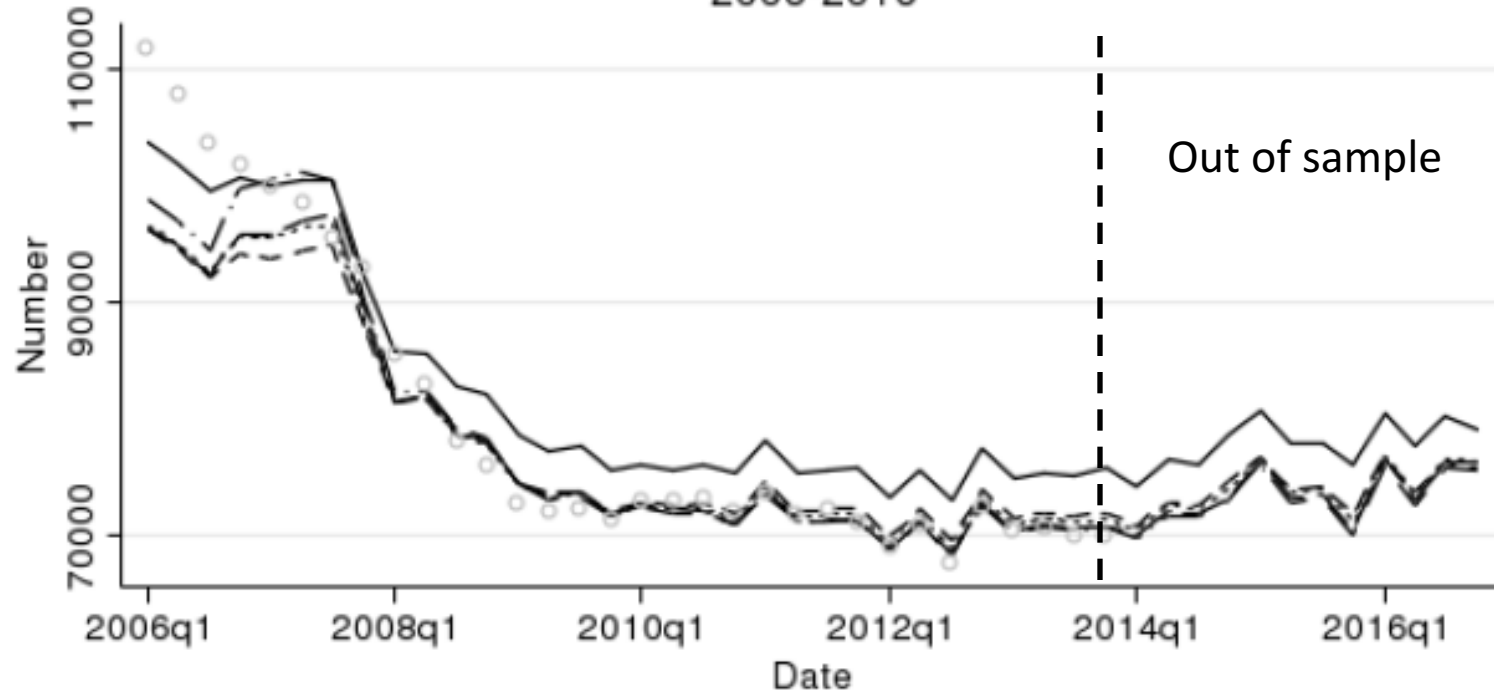
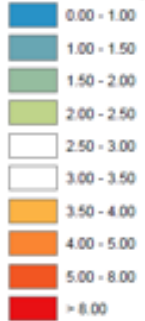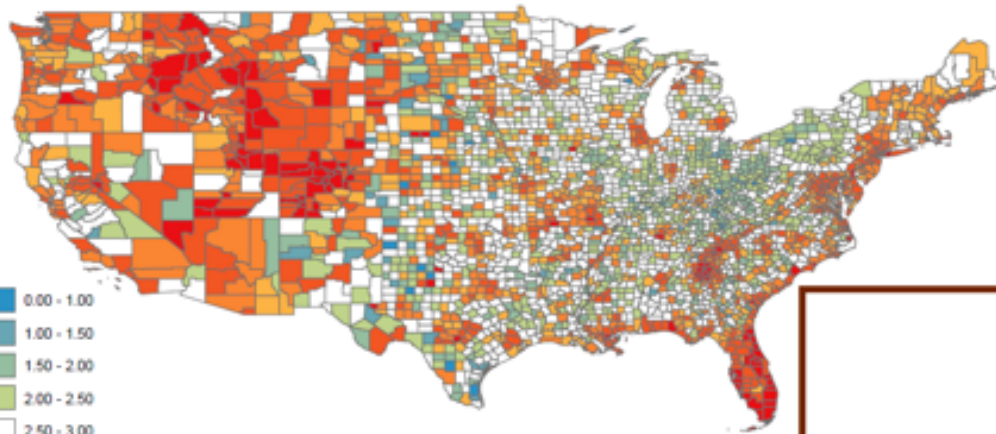Source: EIN Applications File, US Census Bureau.

State level analysis

Actual vs. Projected: State-Level Growth in Firm Births
Year-over-year at Quarterly Frequency

Correlation: 0.727

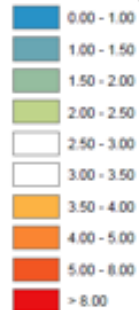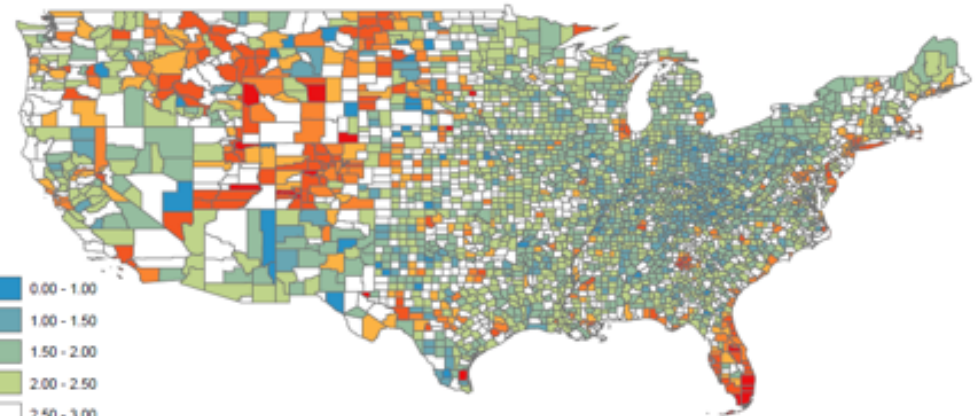Light-dash line 45 degree, correlation: .727, NSA series

Number of High-Propensity Business Applications per Capita by County

2006

2015

Next Steps

- Release nation and state series in the near future
  - Actual series and modelled projections
- Model number of business formations at a point in time
- Explore lower levels of geography, industry and higher frequency
- Model expansion of existing business and non employers
- Research:
  - Business formation and policy impact (local tax, labor, subsidies, finance)
  - Business formation and local conditions
  - Entrepreneurial gaps (demographics)
  - Entrepreneurial quality and high growth
  - Explore other modelling approaches: machine learning
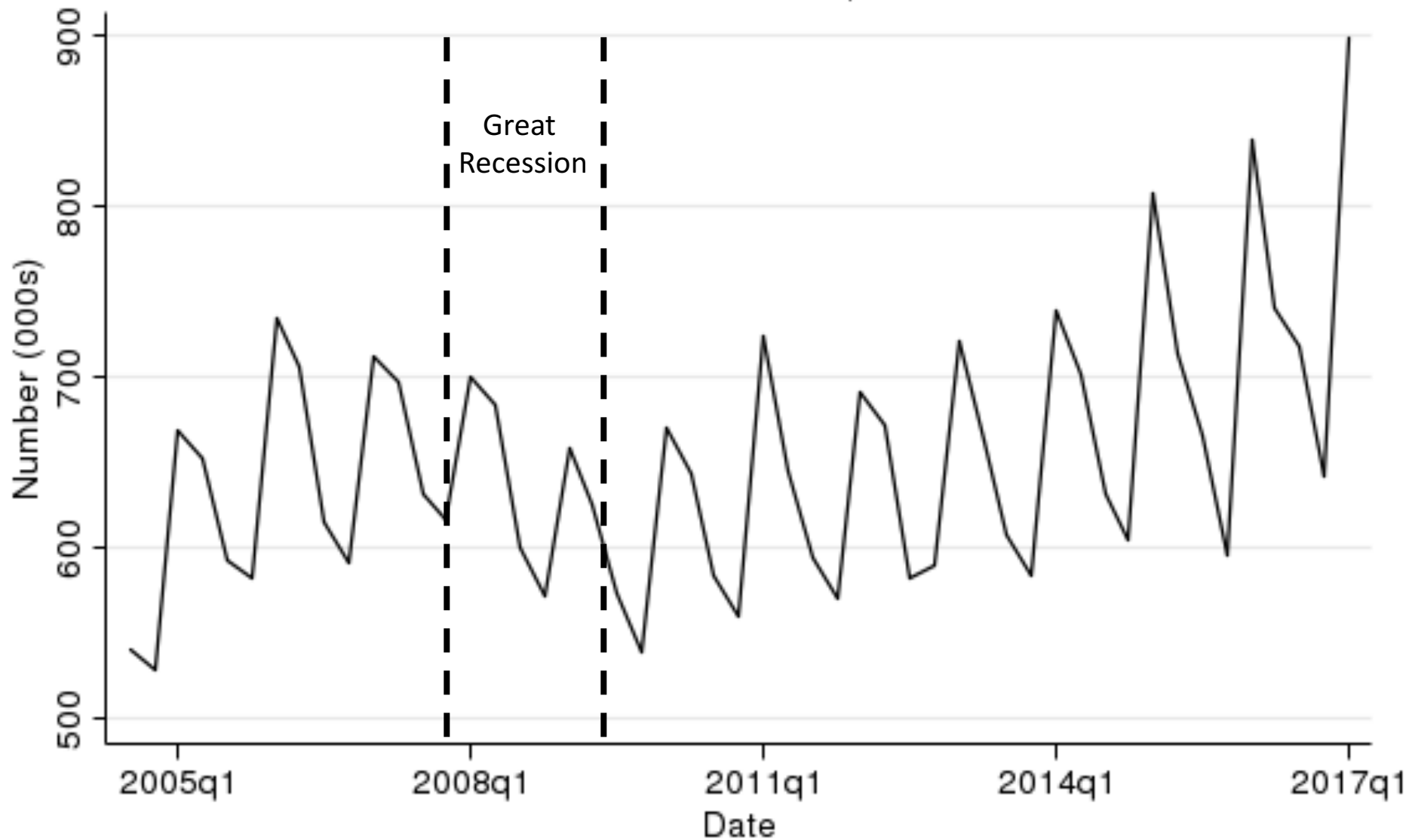
Thank you for listening!

Extra slides

Motivation

- Startups and young firms disproportionally contribute to job and productivity growth
    - Haltiwanger, Jarmin and Miranda (2013), Haltiwanger, Jarmin, Kulick and Miranda (2016)
- Particularly relevant in innovation sectors
    - Decker, Haltiwanger, Jarmin and Miranda (2016) Acemoglu, Akcigit, Bloom, Kerr (2013)
- Display secular decline tied to slow recoveries and are very sensitive to business cycles
    - Fort, Haltiwanger, Jarmin and Miranda (2013), Adelino, Schoar and Severino (2015), Pugsley and Sahin (2014)
- Policy makers want to get the pulse of startup activity as an indicator of activity but timely information on early-stage business dynamics is hard to come by
    - Surveys of new businesses are difficult to implement are costly and offer little granularity
    - Administrative data typically used in construction of business dynamics is at low frequencies (e.g. IRS Business Income/Payroll, W2, UI) => 1-2 year lags
- Can we use high frequency administrative data to provide timely (perhaps early) indicators of economic activity at nation and local level?
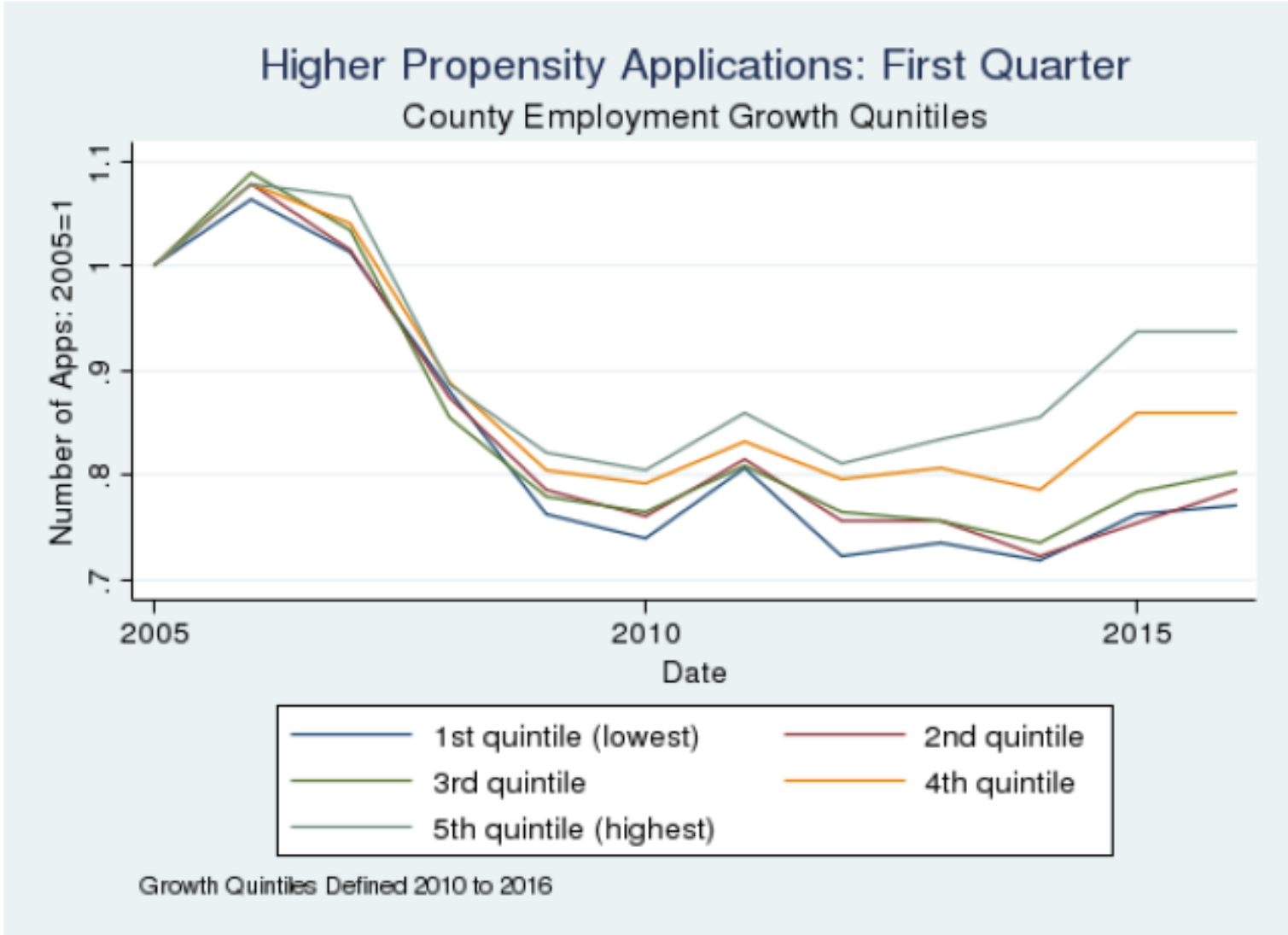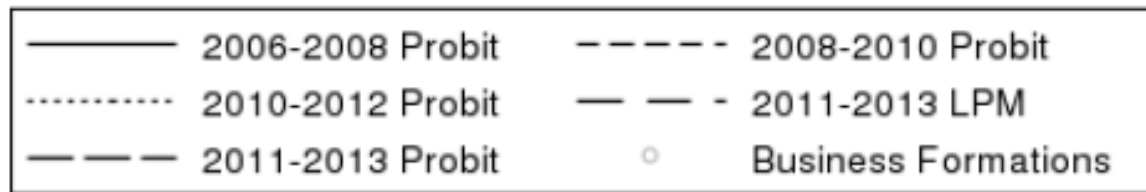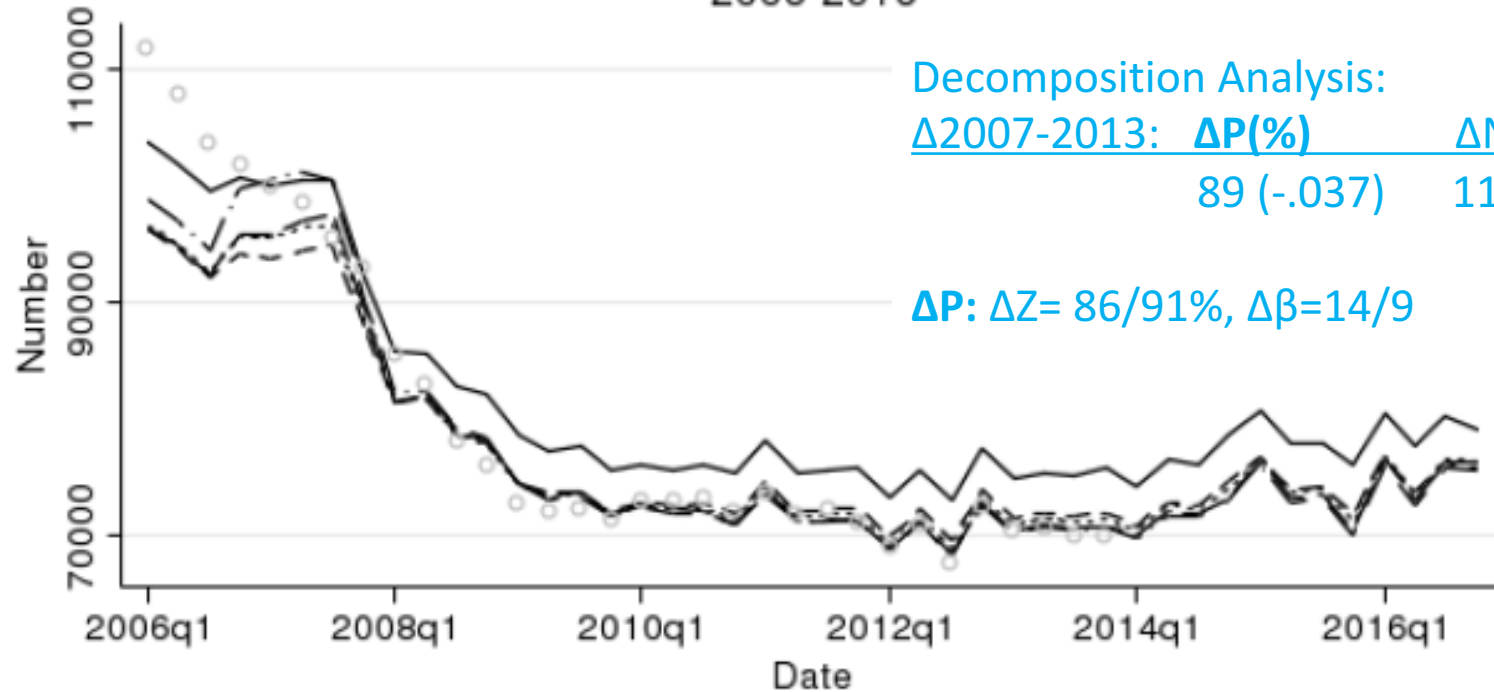
Business Applications
2004:Q3-2017:Q1, NSA

Great
Recession

Number (000s)

2005q1   2008q1   2011q1   2014q1   2017q1
Date

Source: EIN Applications File, US Census Bureau.

22

Applications as Indicators of Local Economic Activity?

Projected Business Formations
2006-2016

Decomposition Analysis:
Δ2007-2013:   **ΔP(%)**   ΔN(%)
                    89 (-.037)   11

ΔP: ΔZ= 86/91%, Δβ=14/9

Legend:
- 2006-2008 Probit
- 2008-2010 Probit
- 2010-2012 Probit
- 2011-2013 LPM
- 2011-2013 Probit
- Business Formations

Source: EIN Applications File, US Census Bureau.