

# Kauffman Firm Survey

Alicia Robb, Kauffman Foundation & UC Santa Cruz

Academy of Entrepreneurial Finance

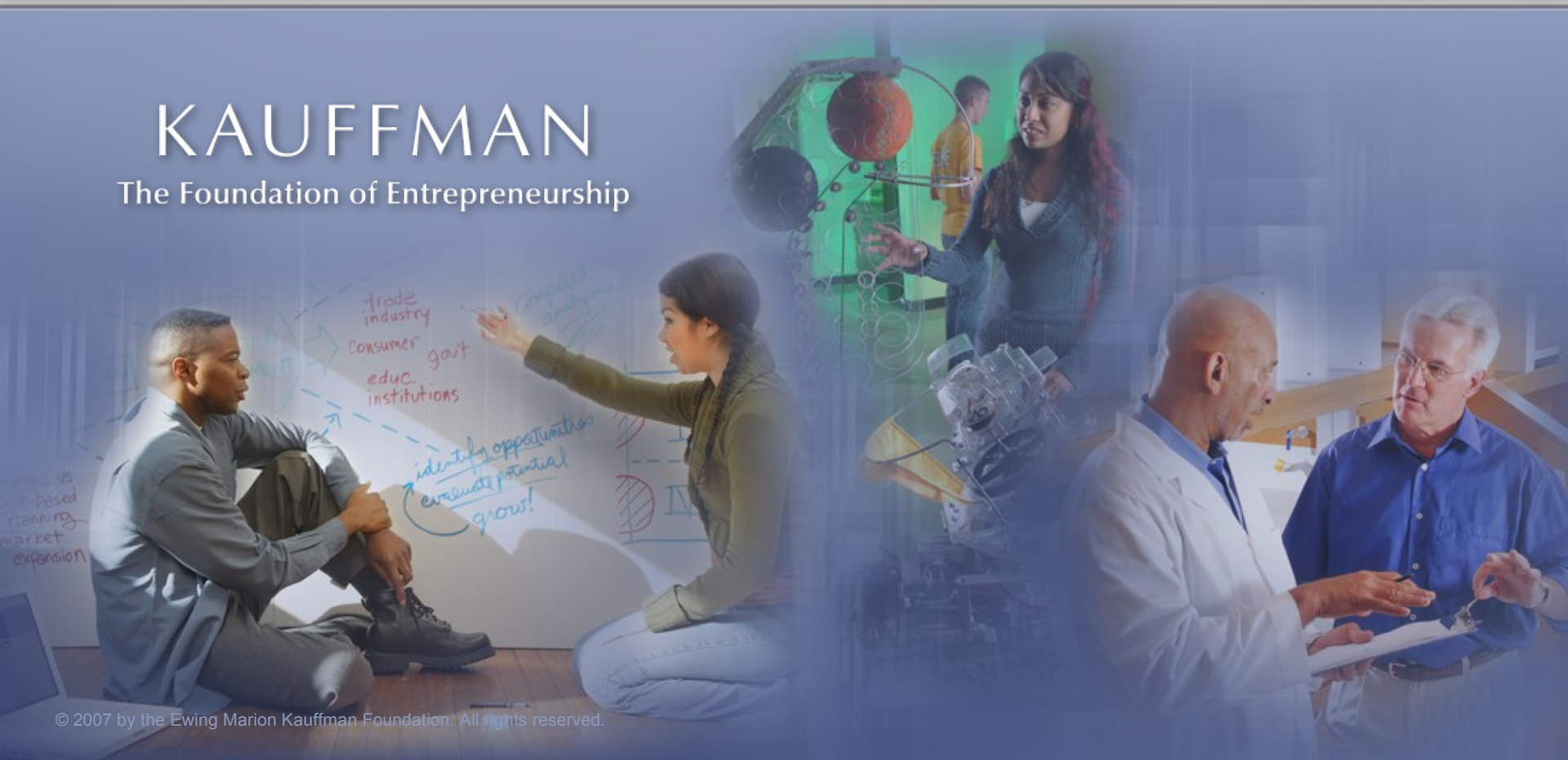
September 25<sup>th</sup>, Las Vegas

“Every individual that we can inspire, that we can guide, that we can help to start a new company, is vital to the future of our economic welfare.”

— Ewing Kauffman

## KAUFFMAN

The Foundation of Entrepreneurship



# Motivation for the KFS

To better understand:

- New firm dynamics
- High-tech firms
- Firm innovation
- Firm financing
- Entrepreneurs (owners of high growth firms)

We need a longitudinal survey that tracks firms over time

# Motivation

- PSED is the only source of data on business formation for a sample of firm founders representative of an underlying population that has information on individual and firm-level variables that match theories used by entrepreneurship scholars
- PSED is great, but...
  - A surprisingly small portion of people “in the process of starting a business” manage to create a business
  - Potentially high performing businesses are not evenly distributed across industries
  - Financing activities are an important part of firm formation, but we don't have a lot of data on them

# The Approach

- Project began with the decision to use Dun & Bradstreet (D&B) data on new businesses rather than population sampling approach
- Take D&B data on “new businesses” and ensure that those that were listed as new businesses were truly new
- Could be new businesses that did not make D&B list that met screening criteria, but they are not observed

# Eligibility Screening

- The business did not have a legal form before 2004
- The business did not establish an EIN number before 2004
- The respondent did not file a Schedule C on their personal income tax before 2004
- The business did not pay state unemployment insurance taxes before 2004
- The business did not pay FICA before 2004
- The business did perform at least one of the above activities for the first time in 2004

# Sample Design

- Stratified random sample based on D&B start year (2004)
- Oversampling on technology status based on BLS definitions (Hadlock, Hecker and Gannon [1991])
- Three technology strata: high-tech, medium-tech, and non-tech businesses
- All high-tech businesses in frame selected

# KFS Milestones

- Survey development began May 2004
- Large scale pilot test conducted in early 2005
- Baseline Survey conducted in 2005-2006
- First Follow-Up Survey conducted in 2006
- Second Follow-Up Survey conducted in 2007
- Third Follow-Up Survey began in June 2008

# Baseline Survey Results

- Data collection period: July 2005 – July 2006
- Total sample size = 32,469
- 53% screened for eligibility, producing an eligibility rate of 35%
- 4,928 interviews completed mainly by CATI (77%), 23% completed by web
- 43% response rate (AAPOR response rate 3)

# Follow up Years

- First Follow up (Calendar Year 2005):
  - 3,998 completed interviews: 59% by web, 41% by CATI
  - 369 identified as out of business between Baseline and First Follow-Up (7.5%)
  - 89% response rate
- Second Follow up (Calendar Year 2006):
  - 3,390 interviews: 63% by web, 37% by CATI
  - 408 identified as out of business (9%)
  - 82% response rate

We have more than 4,100 firms w/data through 2006 or that have been identified as having gone out of business in 2005 or 2006.

# Detailed Data: Baseline & Over Time

- Firm characteristics  
Industry, Legal Form, # of Owners, # of Employees (PT/FT),  
Types of Customers, Location
- Firm strategy and innovation  
Product/Service Offerings, Intellectual Property, Licensing In  
and Licensing Out, R&D
- Detailed financial information  
Equity & Debt Financing, Income Statement Info (Revenue,  
Expenses, Profits), Balance Sheet Info (Assets, Liabilities,  
Equity)
- Employees  
Types of Benefits Offered, Task/Work Structure
- Owner characteristics and work behaviors  
(Information on up to 10 owners)  
Education, Age, Race, Ethnicity, Gender, Citizenship,  
Immigrant Status, Hours Worked, Previous Years of Work  
Experience, Previous Start-up Experience (same/different  
industry as this firm)

# Accessing the Data: Two Approaches

- Public-use Microdata
  - Available on the web
  - Identifying features omitted
  - Can be used to get familiarity with data and some research
  - [www.kauffman.org/kfs](http://www.kauffman.org/kfs)
- Data Enclave
  - Secure server available via remote access
  - More identifying variables available
  - Teams of geographically distributed researchers can access and share code, results and work in collaborative environment
  - <http://dataenclave.norc.org/>

# Contact Us

- David DesRoches (Survey Methodology)  
[ddesroches@mathematica-mpr.com](mailto:ddesroches@mathematica-mpr.com)
- Tim Mulcahy (NORC Enclave)  
[mulcahy-tim@norc.uchicago.edu](mailto:mulcahy-tim@norc.uchicago.edu)
- Alicia Robb (Data and Research)  
[arobb@ucsc.edu](mailto:arobb@ucsc.edu)