The Link between R&D and Entrepreneurship

Nathan Goldschlag[1]   Ron Jarmin[1]   Julia Lane[2]   Nikolas Zolas[1]

[1]U.S. Census Bureau

[2]New York University

American Economic Association Meetings, 2017

Contribution & Motivation

- Contribution:
  - New measure of human capital, research training, enhanced with machine learning
  - Correlate human capital composition and startup performance

- Findings: ⇑ research trained workers at startups
  - ⇒ ⇑ survival and
  - ⇒ ⇑ growth

- Motivation:
  - Economic impacts of grant funded research are not well understood
  - High skill human capital and specialization important for growth
  - Entrepreneurship is a mechanism for translating grant research into the economy

Framework

Consider startup outcomes as being determined by

$$Y_{ft} = f(K_{ft}, L_{ft}, X_{ft}) \qquad (1)$$

- Measures of human capital, L_{ft}
    - Research training
    - Experience in: R&D performing firm, High Tech industries, research universities

Data

- Grant-funded researchers
    - UMETRICS

- Individuals
    - Individual Characteristics File (ICF)

- Firms and establishments
    - Longitudinal Business Database (LBD) and Business Register (BR)

- Jobs
    - Longitudinal Employee-Household Dynamics (LEHD) and W2

- University attributes
    - Carnegie Institute University data and Integrated Postsecondary Education System (IPEDS)

Modeling Research Training

- Machine learning to identify research trained individuals outside of UMETRICS universities

- Training set: 1.4M person-university pairs from 14 UMETRICS universities

- Out-of-sample set: 6.8M person-university pairs from 130 top research universities

- Feature set includes measures of pre, post, during employment history, demographics, and university characteristics

Why it Works

- Research trained individuals are significantly different (Zolas et al., 2015)

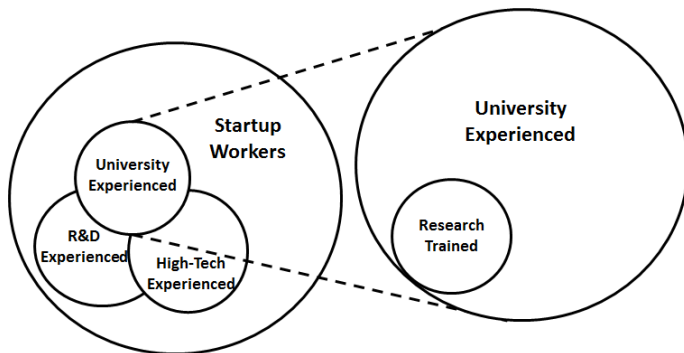|	Research Trained	Non-Research Trained
Pct Female	50.5	54.1
Mean Birth Year	1977.7	1975.6
Pct Foreign Born	21.8	11.4
Pct White	73.2	77.2
Pct Black	5.7	9.3
Pct Hispanic	4.3	4.9
Pct Asian	14.1	6.2
Post University Earnings (2014)	63,600.0	51,950.0
Observations	141K	1.26M

Notes: All differences significant at $p < 0.01$. Numbers are rounded for disclosure purposes.

Learning Results

- Accuracy scores across feature sets > 97

- Precision scores > 98 and recall scores > 77

- Quality holds up in cross validation, mean 10-fold precision > 70

- Results likely under-count research trained individuals

Basic Survival Facts
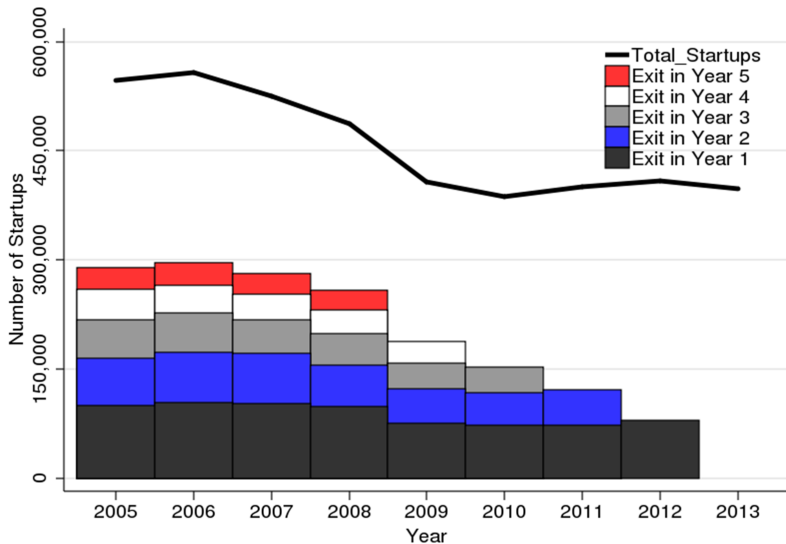
Overlapping Human Capital

Startup Affiliated Workers	% R&D Experience	% High Tech Experience	% University Experience	% Research Trained
5.2M	20.82	12.88	2.50	0.04

Note: Unit of observation is individual. Averages 2005-2013.

Descriptive Regressions

$$
\begin{aligned}
Y_f = \quad \alpha \quad & + \beta_1 \ln RD_f + \beta_2 \ln HT_f + \beta_3 \ln UNI_f + \beta_4 \ln RSCH\_TRND_f \\
& + \sum_{k=1}^{9} \delta_k SIZE_{fk} + \beta_5 \ln \overline{EARN}_i + \beta_6 \ln \overline{AGE}_i \\
& + \beta_7 \ln FEMALE_i + \beta_8 \ln FOREIGN_i + \epsilon
\end{aligned}
$$

- For startup f, outcomes Y_f include:
    - **Survival** to age 5 with $10+$ employees in year 5
    - **High Growth** - top decile of employment growth distribution to 5

- Other notes:
    - All worker type counts $+=1$
    - All specifications include industry fixed effects and zip-year fixed effects

Startup Survival
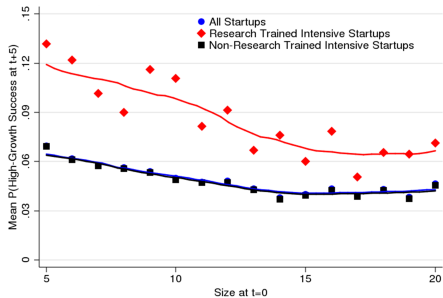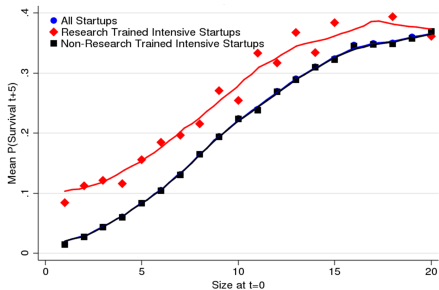
Mean Percent of Startups that Survive: 6.6%

Outcome Variable	Coefficient	Mean	Doubling Worker Type
ln $RSCH\_TRND_{i0}$	0.83***	0.01	8.7%
ln RD_{i0}	0.63***	1.38	6.6%
ln HT_{i0}	0.47***	0.54	4.9%
ln UNI_{i0}	0.62***	0.17	6.5%

Notes: Robust standard errors are in parentheses. * $p < 0.05$,
** $p < 0.01$, *** $p < 0.001$. Unit of observation is the startup
establishment.

Startup Growth

Mean Percent of Startups with High Grow: 1.2%

Outcome Variable	Coefficient	Mean	Doubling Worker Type
ln $RSCH\_TRND_{i0}$	0.42***	0.01	24.6%
ln RD_{i0}	0.17***	1.38	9.7%
ln HT_{i0}	0.15***	0.54	8.9%
ln UNI_{i0}	0.22***	0.17	12.9%

Notes: Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Unit of observation is the startup establishment.

Concluding Remarks

- Introduce new measures of human capital–research training, extended with machine learning

- Investigate human capital composition of startups

- Findings: doubling research trained workforce
    $\Rightarrow$ 8.7% increase survival and
    $\Rightarrow$ 24.6% increase high growth

- Next steps:
    - Additional robustness, linking to productivity, explore mechanisms

thank you

Learning Results

Feature Set	Random Forest			
	Chi Squared	Decision Tree	Gini Impurity	Hand Curated
Accuracy	99.703	97.277	99.950	99.924
Precision	99.990	98.971	99.991	99.981
Recall	97.443	77.248	99.580	99.365
Mean 10-Fold Precision	99.849	70.816	99.806	62.222
Mean Unv-Fold Precision	51.900	79.570	41.646	42.934

Note: The hand-curated set includes demographic variables and demeaned earnings for individuals during their time at the university.