Moneyball for Startups

Tauhid Zaman

Joint Work with David Scott Hunter
Picking Winners

Nutanix
- Virtualization
- Previous Experience: 38 Years Old
- Masters

stiQRd
- Mobile Apps
- Previous Experience: 8 Previous Companies
- PhD

Checkd.In
- Big Data/Advertising
- Previous Experience: “Top” School
- 26 Years Old

Sanuthera
- Medical Devices
- No Previous Experience
- MD

IMRSV
- Computer Vision
- “Top” School
- PhD
- 28 Years Old

IPO: $2.5 B
100x
Venture Capital Investments

- The average early-stage VC investment produces a return of 31%

- Yet, most VC firms lose money on these investments
  - 80-90% of early-stage startups do not achieve an exit
  - 5-10% achieve exits with returns of 10-20X
  - 1% achieve exits with returns greater than 100X
Venture Capital Investments

• How do they make these investment decisions?
Quantitative Approach

• Scholars
  – Most academic studies have focused on what factors are correlated with startup success
  – No academic study has considered a fully quantitative approach to VC investment
Quantitative Approach

• Dollars

– Some VCs have developed analytical tools to assist in the investment decision-making process
Our Contribution

1. New data on startups

2. New model for startup success (based on random walks)

3. Analytics based portfolio construction method (“Moneyball”)
Data

crunchbase

LinkedIn

PitchBook
Data

- **Crunchbase** (from 1981 to 2016, public user)
  - 83,000 startup companies
  - 48,000 investors
  - 147,000 investment rounds
  - 558,000 employees
Data

• **Pitchbook (Privately maintained)**
  – 774,000 companies
  – Investing rounds information
  – Valuation at these rounds
Data

- LinkedIn
  - 200,000 employees
  - Employment history
  - Education

![LinkedIn Profile with information on Tauhid Zaman, Assistant Professor at Sloan School of Management, MIT, and details on his skills, education, and patents, including a research on information propagation probability for a social network.](image-url)
Dataset for Analysis

- US companies founded after 2000
- 24,000 companies
Funding Rounds Data

The graph shows the timeline of funding rounds for various companies, categorized by different rounds such as Series A, Series B, Series C, Series D, Series E, Series F, and Exit. The horizontal axis represents the year, ranging from 2004 to 2016, while the vertical axis represents the funding round. The graph includes data points for Twitter, Yelp, Facebook, and Etsy, indicating their progression through different funding rounds over the years.
Maximum Funding Round (as of 2016)
Time of Maximum Funding Round

![Box plot](image)
Sector Data

- 59 sector indicators
Leadership Data

• Using Crunchbase:
  – Previous startup experience for the founders, employees, and advisors

• Using LinkedIn:
  – Previous startup experience
  – Education
  – Academic major
  – Age of the founders
Investor Data

Companies

Investors
Investor Data

• Network features
  – investor neighborhood size
  – investor IPO/acquisition fraction

• Dynamic company-investor networks
  • Each edge has a time stamp
No Cheating Condition

• Funding round data
• Sector data
• Investor network data
• Leadership data

Could you have known this information when deciding to invest in the company?
Random Walk Drift and Diffusion

- **Drift** – avg. rate of increase of random walk
- **Diffusion** – how erratic the random walk is
Modeling Drift and Diffusion

Slow down over time
Temporal Behavior of Drift and Diffusion

Constant for a while

Then start decaying

\[ C \left( \frac{t - \nu}{\tau} \right) \]
Modeling Drift and Diffusion

• For a company \( i \) that is founded in year \( y \) with feature vector \( \mathbf{x}_i \) we have:

\[
\text{Drift: } \mu_i = \beta_y^T \mathbf{x}_i \\
\text{Diffusion: } \sigma_i = \gamma^T \mathbf{x}_i
\]

• Time varying strength of drift features:

\[
\beta_{i,y+1} = \beta_{i,y} + \epsilon \\
\epsilon = \text{noise}
\]
Building Portfolios

• Given a predictive model, how can we select companies?
  – If at least one company exits, we make a huge profit. Otherwise, we lose money.
  – Let $E_i$ correspond to the event that company $i$ exits.

$$\max_{S \subseteq [m], |S| = k} \mathbb{P}\left( \bigcup_{i \in S} E_i \right)$$

“Picking Winners” Portfolio
Picking Winners

• Venture capital

• Romance

• Fantasy sports
$55K Sniper Payoff Structure

26% of the money in the top 10 lineups
Were we able to win?

November 15, 2015

November 16, 2015

November 17, 2015

November 23, 2015

200 lineups
Scott Hunter – Current MIT student

Jason Robbins – CEO DraftKings

Tauhid Zaman – Former MIT student, Compulsive gambler
Policy Change

200 lineups -> 100 lineups
Performance in Baseball

<table>
<thead>
<tr>
<th>Rank</th>
<th>Player</th>
<th>Performance</th>
<th>Prize</th>
<th>PIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>zlisto</td>
<td>232.20</td>
<td>$50,000.00</td>
<td>0</td>
</tr>
<tr>
<td>2nd</td>
<td>geotico</td>
<td>228.05</td>
<td>$10,000.00</td>
<td>0</td>
</tr>
<tr>
<td>3rd</td>
<td>woppen1</td>
<td>227.50</td>
<td>$5,000.00</td>
<td>0</td>
</tr>
<tr>
<td>4th</td>
<td>eggjockee</td>
<td>218.25</td>
<td>$3,000.00</td>
<td>0</td>
</tr>
</tbody>
</table>

Liked by mel.a.chen, sirtedb and 9 others
zlisto Still a Draft God. #dkmafia

View all 2 comments
kevincturner Which sport???
# Performance in Football

## MY CONTESTS

[View Last 30 Days] [Search] [Download Entry History]

### ENTRIES

<table>
<thead>
<tr>
<th>Contest Name</th>
<th>Place</th>
<th>$ Won</th>
<th>FPTS</th>
<th>Completed</th>
<th>Total Prizes</th>
<th>Top Prize</th>
<th>Places Payed</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>23</td>
<td>$3,000</td>
<td>219.12</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>391</td>
<td>$300</td>
<td>202.22</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>1123</td>
<td>$150</td>
<td>193.62</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>1414</td>
<td>$150</td>
<td>191.42</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>1887</td>
<td>$125</td>
<td>188.82</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>1918</td>
<td>$125</td>
<td>188.62</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>3889</td>
<td>$75</td>
<td>182.02</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>3889</td>
<td>$75</td>
<td>182.02</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>4681</td>
<td>$60</td>
<td>180.32</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>5906</td>
<td>$60</td>
<td>177.92</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>6829</td>
<td>$50</td>
<td>176.32</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
<tr>
<td>NFL $5M Fantasy Football Millionaire ($1M to $100M)</td>
<td>7216</td>
<td>$50</td>
<td>175.72</td>
<td>09/19/2016</td>
<td>$5,000,000</td>
<td>$1,000,000</td>
<td>81475</td>
</tr>
</tbody>
</table>
Fantasy Sports to Venture Capital

• Colleagues and reviewers wanted us to apply our “picking winners” technique to something more “business oriented”

• “We don’t play games at Sloan!”

• So we were “forced” to apply it to venture capital
Back to Startups

• Given a predictive model, how can we select companies?

\[
\max_{S \subseteq [m], |S| = k} P \left( \bigcup_{i \in S} E_i \right)
\]

“Picking winners” portfolio
Building the Picking Winners Portfolio

• Choose observation date $t_{obs}$

• Estimate model using data before $t_{obs}$

• For companies founded the year after $t_{obs}$ solve

$$\max_{S \subseteq [m], |S| = k} P \left( \bigcup_{i \in S} E_i \right)$$
Drift and Diffusion by Funding Round

![Graph showing Drift and Diffusion by Funding Round](image-url)
# Non-Sector Parameter Values

<table>
<thead>
<tr>
<th>Non-sector feature</th>
<th>$\beta_{2010}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive acquisition</td>
<td>0.80</td>
</tr>
<tr>
<td>Executive IPO</td>
<td>0.80</td>
</tr>
<tr>
<td>Advisory IPO</td>
<td>0.26</td>
</tr>
<tr>
<td>Leadership age</td>
<td>0.25</td>
</tr>
<tr>
<td>Maximum acquisition fraction</td>
<td>0.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-sector feature</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job IPO</td>
<td>3.31</td>
</tr>
<tr>
<td>Previous founder</td>
<td>3.22</td>
</tr>
<tr>
<td>Job acquisition</td>
<td>2.88</td>
</tr>
<tr>
<td>Top school</td>
<td>2.45</td>
</tr>
<tr>
<td>Maximum acquisition fraction</td>
<td>2.43</td>
</tr>
</tbody>
</table>
## Sector Parameter Values

<table>
<thead>
<tr>
<th>Sector feature</th>
<th>Drift $\beta_{2010}$</th>
<th>Diffusion $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-learning</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Ride sharing</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>Open source</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Cloud computing</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Social media</td>
<td></td>
<td>4.88</td>
</tr>
<tr>
<td>Messaging</td>
<td></td>
<td>3.47</td>
</tr>
<tr>
<td>Social network</td>
<td></td>
<td>3.46</td>
</tr>
<tr>
<td>Apps</td>
<td></td>
<td>2.47</td>
</tr>
<tr>
<td>Cloud computing</td>
<td></td>
<td>1.94</td>
</tr>
<tr>
<td>2011 Company</td>
<td>Maximum funding round</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------</td>
<td></td>
</tr>
<tr>
<td>SHIFT</td>
<td>Acquired</td>
<td></td>
</tr>
<tr>
<td>Jibbigo</td>
<td>Acquired</td>
<td></td>
</tr>
<tr>
<td>Sequent</td>
<td>Series B</td>
<td></td>
</tr>
<tr>
<td><strong>Nutanix</strong></td>
<td>IPO</td>
<td></td>
</tr>
<tr>
<td>PowerInbox</td>
<td>Series A</td>
<td></td>
</tr>
<tr>
<td>Friend.ly</td>
<td>Acquired</td>
<td></td>
</tr>
<tr>
<td>Jybe</td>
<td>Acquired</td>
<td></td>
</tr>
<tr>
<td>MediaRoost</td>
<td>Seed</td>
<td></td>
</tr>
<tr>
<td>CloudTalk</td>
<td>Series A</td>
<td></td>
</tr>
</tbody>
</table>
Papers and Code

Picking Winners:
A Framework For Venture Capital Investment
https://arxiv.org/abs/1706.04229

Picking Winners Using Integer Programming
https://arxiv.org/abs/1604.01455

DraftKings baseball code
https://github.com/zlisto/dailyfantasybaseball
Thank You!