On the use of financial data as a random beacon
Overview

• We examine using the closing prices of stocks as a source for a true random seeds
• This approach has been used in binding E2E elections
• We conservatively estimate that over one trading day, the stocks in the Dow Jones have over 200 unpredictable bits
• We find the level of randomness is sufficient
Randomness in elections

• The detection of errors or fraud in elections can be achieved with audits
• In traditional elections, precincts can be randomly selected for manual recounts
• In end-to-end verifiable (E2E) elections, random challenges can prove the tally is correctly computed from a verifiable set of privacy-preserving receipts
• If the challenges were known in advance, the proof could be faked
Random challenges

• Two systems that require external randomness are Scantegrity II and Punchscan
• Both have run binding elections and both used financial market data for generating a seed
• The seed (or its pseudorandom expansion) is formatted to create challenges
• What properties should a random seed have for E2E elections?
  – Each bit should have a uniform probability of 0 or 1
  – Generated at the appropriate time
  – Appropriate length
  – Generation is observable by anyone
  – A high level of mathematics is tolerable
Price manipulation

• Since the price is determined by trades and anyone can trade, can’t anyone manipulate the closing price?
• In theory, yes, but…
• Widely considered to be difficult for liquid stocks on established exchanges
• There is empirical evidence for this
• Barrier options continue to be written, held and traded
• Other complexities: see paper
Method

Financial Model
• Choose a model to represent stock price movements

Historic Data
• Fit historic data to the model to estimate parameters

Monte Carlo Simulations
• Run simulations of price movements forward in time

Entropy Estimation
• Measure the resulting entropy

Extraction
• Determine how to extract random bits from prices
Modeling stock prices

- To estimate the randomness in a closing price, we need to assume a mathematical model holds for stock prices.
- These models do not predict prices.
- Models are used in real-life by banks to hedge against risky assets.
Black-Scholes

- We use the Black-Scholes model.
- This model is now widely considered to under-estimate market volatility: bad for banks when pricing options, good for us in estimating a lower-bound on the randomness in a closing price.
- Black-Scholes assumes that stock prices follow a stochastic process called geometric Brownian motion (GBM).
At each time-step, move up or down one unit
At each time-step, move up or down an amount drawn from a Normal distribution
Add a upward or downward drift
Geometric Brownian motion

• If we make it continuous in time, we get:

\[ dS_t = \mu S_t \, dt + \sigma S_t \, dW_t \]

- \( S_t \): Stock price at a given time
- \( \mu \): Drift term / rate of return / interest rate
- \( \sigma \): Diffusion term / volatility
- \( dW_t \): Increment of a Weiner process / stochastic term
Geometric Brownian motion

• With a series of prices for a specific stock, we can estimate its daily drift and diffusion rates

• Example: Microsoft over one year

• From March 23, 2009 (at $17.95) until March 23, 2010 (at $29.88)
MSFT closing prices
Logarithmic returns

- We are interested in the relative changes in the price, and need to fit it to an exponent
- For each price, we calculate its logarithmic return from the previous price:

\[
R_i = \ln \left( \frac{S_{i+1}}{S_i} \right), \quad 0 \leq i \leq T - 1
\]

where T is the number of prices in the period (T=251)
Histogram of log-returns
Estimator for drift/diffusion

• Under GBM, the log-returns should be normally distributed as:

$$R_i \sim N \left( \left( \mu - \frac{\sigma^2}{2} \right) \Delta t, \sigma^2 \Delta t \right)$$

• We can fit our historic data
• For MSFT during this period, daily drift was 0.23% and daily diffusion was 1.77%.
Monte Carlo

• Now that we have estimates for drift and diffusion, we simulate many possible paths for the stock price over the next day.
• We round the output price to the nearest cent.
• This gives a discrete probability distribution we can use to estimate the randomness.
• This approach has some bias: see paper.
Monte Carlo simulations
Histogram of outcomes
Entropy

• Randomness can measured: entropy
• A sequence of numbers with $N$ bits of (Shannon) entropy contains the same randomness as flipping a coin $N$ times
• We can generally extract some these random bits from the sequence but not necessarily all $N$ bits
• $M$ bits of min-entropy means we can (theoretically) extract $M \leq N$ coin tosses
Entropy Estimation

• Entropy is measured from histogram

• For MSFT over 1 day:
  – 7.76 bits of estimated Shannon entropy
  – 0.02 bits of estimated bias
  – 7.04 bits of estimated min-entropy

• Scantegrity II used the 30 stocks in the Dow Jones
<table>
<thead>
<tr>
<th>Stock</th>
<th>$S_T$</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\hat{N}$</th>
<th>$\hat{M}$</th>
<th>$H(P)$</th>
<th>$B$</th>
<th>$H_B(P)$</th>
<th>$H_\infty(P)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>14.50</td>
<td>0.00338956</td>
<td>0.0354406</td>
<td>386</td>
<td>440</td>
<td>7.72544</td>
<td>0.0022</td>
<td>7.73</td>
<td>6.99</td>
</tr>
<tr>
<td>AXP</td>
<td>41.24</td>
<td>0.00512444</td>
<td>0.0365912</td>
<td>1071</td>
<td>1305</td>
<td>9.2823</td>
<td>0.006525</td>
<td>9.29</td>
<td>8.50</td>
</tr>
<tr>
<td>BA</td>
<td>72.18</td>
<td>0.00313975</td>
<td>0.0219116</td>
<td>1112</td>
<td>1406</td>
<td>9.34279</td>
<td>0.00703</td>
<td>9.35</td>
<td>8.57</td>
</tr>
<tr>
<td>BAC</td>
<td>17.13</td>
<td>0.00455058</td>
<td>0.0468486</td>
<td>588</td>
<td>699</td>
<td>8.37453</td>
<td>0.003495</td>
<td>8.38</td>
<td>7.62</td>
</tr>
<tr>
<td>CAT</td>
<td>62.41</td>
<td>0.00352696</td>
<td>0.027272</td>
<td>1173</td>
<td>1540</td>
<td>9.4536</td>
<td>0.0077</td>
<td>9.46</td>
<td>8.69</td>
</tr>
<tr>
<td>CSCO</td>
<td>26.64</td>
<td>0.00200486</td>
<td>0.0167037</td>
<td>347</td>
<td>396</td>
<td>7.52981</td>
<td>0.00198</td>
<td>7.53</td>
<td>6.79</td>
</tr>
<tr>
<td>CVX</td>
<td>74.77</td>
<td>0.00565046</td>
<td>0.0136131</td>
<td>730</td>
<td>844</td>
<td>8.70798</td>
<td>0.00422</td>
<td>8.71</td>
<td>7.95</td>
</tr>
<tr>
<td>DD</td>
<td>38.31</td>
<td>0.0025213</td>
<td>0.0219181</td>
<td>603</td>
<td>751</td>
<td>8.43244</td>
<td>0.003755</td>
<td>8.44</td>
<td>7.69</td>
</tr>
<tr>
<td>DIS</td>
<td>34.01</td>
<td>0.00271648</td>
<td>0.0199576</td>
<td>506</td>
<td>596</td>
<td>8.13347</td>
<td>0.00298</td>
<td>8.14</td>
<td>7.39</td>
</tr>
<tr>
<td>GE</td>
<td>18.33</td>
<td>0.00264998</td>
<td>0.0239698</td>
<td>335</td>
<td>391</td>
<td>7.50284</td>
<td>0.001955</td>
<td>7.50</td>
<td>6.76</td>
</tr>
<tr>
<td>HD</td>
<td>32.59</td>
<td>0.00166033</td>
<td>0.0161739</td>
<td>404</td>
<td>475</td>
<td>7.76791</td>
<td>0.002375</td>
<td>7.77</td>
<td>7.03</td>
</tr>
<tr>
<td>HPQ</td>
<td>53.15</td>
<td>0.00234904</td>
<td>0.015783</td>
<td>615</td>
<td>758</td>
<td>8.43501</td>
<td>0.00379</td>
<td>8.44</td>
<td>7.69</td>
</tr>
<tr>
<td>IBM</td>
<td>129.37</td>
<td>0.00124652</td>
<td>0.0124436</td>
<td>1121</td>
<td>1460</td>
<td>9.36931</td>
<td>0.0073</td>
<td>9.38</td>
<td>8.60</td>
</tr>
<tr>
<td>INTC</td>
<td>22.67</td>
<td>0.0019257</td>
<td>0.0176758</td>
<td>302</td>
<td>352</td>
<td>7.37295</td>
<td>0.00176</td>
<td>7.37</td>
<td>6.63</td>
</tr>
<tr>
<td>JNJ</td>
<td>65.36</td>
<td>0.00101973</td>
<td>0.00811278</td>
<td>406</td>
<td>472</td>
<td>7.7723</td>
<td>0.00236</td>
<td>7.77</td>
<td>7.03</td>
</tr>
<tr>
<td>JPM</td>
<td>44.58</td>
<td>0.00261719</td>
<td>0.0318482</td>
<td>992</td>
<td>1190</td>
<td>9.18918</td>
<td>0.00595</td>
<td>9.20</td>
<td>8.43</td>
</tr>
<tr>
<td>KFT</td>
<td>30.49</td>
<td>0.00134673</td>
<td>0.0129888</td>
<td>314</td>
<td>333</td>
<td>7.35464</td>
<td>0.001665</td>
<td>7.36</td>
<td>6.62</td>
</tr>
<tr>
<td>KO</td>
<td>55.30</td>
<td>0.0010976</td>
<td>0.0111199</td>
<td>460</td>
<td>570</td>
<td>7.98678</td>
<td>0.00285</td>
<td>7.99</td>
<td>7.21</td>
</tr>
<tr>
<td>MCD</td>
<td>67.35</td>
<td>0.00111279</td>
<td>0.0113681</td>
<td>569</td>
<td>732</td>
<td>8.3043</td>
<td>0.00366</td>
<td>8.31</td>
<td>7.55</td>
</tr>
<tr>
<td>MMM</td>
<td>82.35</td>
<td>0.00235099</td>
<td>0.0148201</td>
<td>854</td>
<td>1075</td>
<td>8.97369</td>
<td>0.005375</td>
<td>8.98</td>
<td>8.22</td>
</tr>
<tr>
<td>MRK</td>
<td>38.50</td>
<td>0.00162879</td>
<td>0.0166847</td>
<td>486</td>
<td>554</td>
<td>8.05935</td>
<td>0.00277</td>
<td>8.06</td>
<td>7.29</td>
</tr>
<tr>
<td>MSFT</td>
<td>29.88</td>
<td>0.0022737</td>
<td>0.0176583</td>
<td>394</td>
<td>449</td>
<td>7.76265</td>
<td>0.002245</td>
<td>7.76</td>
<td>7.04</td>
</tr>
<tr>
<td>PFE</td>
<td>17.54</td>
<td>0.00120496</td>
<td>0.01571</td>
<td>216</td>
<td>243</td>
<td>6.82701</td>
<td>0.001215</td>
<td>6.83</td>
<td>6.10</td>
</tr>
<tr>
<td>PG</td>
<td>64.53</td>
<td>0.00146004</td>
<td>0.0125241</td>
<td>587</td>
<td>703</td>
<td>8.37914</td>
<td>0.003515</td>
<td>8.38</td>
<td>7.64</td>
</tr>
<tr>
<td>T</td>
<td>26.55</td>
<td>0.000357228</td>
<td>0.0121909</td>
<td>251</td>
<td>289</td>
<td>7.05851</td>
<td>0.001445</td>
<td>7.06</td>
<td>6.31</td>
</tr>
<tr>
<td>TRV</td>
<td>53.90</td>
<td>0.00154645</td>
<td>0.0188065</td>
<td>734</td>
<td>926</td>
<td>8.7059</td>
<td>0.00463</td>
<td>8.71</td>
<td>7.96</td>
</tr>
<tr>
<td>UTX</td>
<td>73.09</td>
<td>0.00232501</td>
<td>0.0159515</td>
<td>835</td>
<td>1015</td>
<td>8.9016</td>
<td>0.005075</td>
<td>8.91</td>
<td>8.14</td>
</tr>
<tr>
<td>VZ</td>
<td>30.98</td>
<td>0.000367966</td>
<td>0.0117435</td>
<td>279</td>
<td>320</td>
<td>7.22926</td>
<td>0.0016</td>
<td>7.23</td>
<td>6.48</td>
</tr>
<tr>
<td>WMT</td>
<td>55.89</td>
<td>0.000497465</td>
<td>0.010295</td>
<td>431</td>
<td>512</td>
<td>7.89168</td>
<td>0.00256</td>
<td>7.89</td>
<td>7.16</td>
</tr>
<tr>
<td>XOM</td>
<td>66.95</td>
<td>0.000317968</td>
<td>0.012391</td>
<td>604</td>
<td>752</td>
<td>8.41962</td>
<td>0.00376</td>
<td>8.42</td>
<td>7.65</td>
</tr>
</tbody>
</table>
• We also isolated the effect of each parameter on entropy: drift, diffusion, initial price, and elapsed time
• See paper
Correlated stocks

- From chart: MSFT has 7.76 bits and IBM has 9.38 bits
- If we concatenate their prices, do we get $7.76 + 9.38 = 17.14$ bits?
- No. The price movements are correlated
- See the paper for modeling correlated stocks
Bottom line

• We estimate the randomness in the DJIA portfolio to have 218 bits of Shannon entropy and 192 bits of min-entropy.
Useful form

• Consider taking a set of closing prices and concatenating them together into a large binary string
• Some of the individual bits in this string will be nearly random while others will be almost deterministic
• Can we convert it into a smaller bitstring where each individual bit is uniform random?
• Yes. We require an extractor
Extractors

• Can we just hash it?
• No. A hash function (ideal compression & Merkle-Damgaard) does not make a good extractor [DGHKR’04]
• However we can use a standard cryptographic primitive: block cipher (ideal PRP) in CBC-MAC mode [DGHKR’04]
Producing a seed

- In summary, to make a random seed: take closing prices, concatenate them together, and extract.
- This is minimal: seeds rely on only that day and rely fully on the market’s randomness.
- We present a general protocol for a beacon service provider that offers some additional security properties: see paper.
Concluding Remarks

• The approach of using closing prices for post-election audits in E2E elections is sound
• Using a portfolio such as the Dow Jones will produce enough bits for a cryptographically strong seed
• This seed can be used directly or expanded with a PRG
Questions?