STAC
BENCHMARK COUNCIL

STAC Update:
Big compute

Michel Debiche
Director of Analytics Research, STAC

michel.debiche@STACresearch.com

STAC-A2

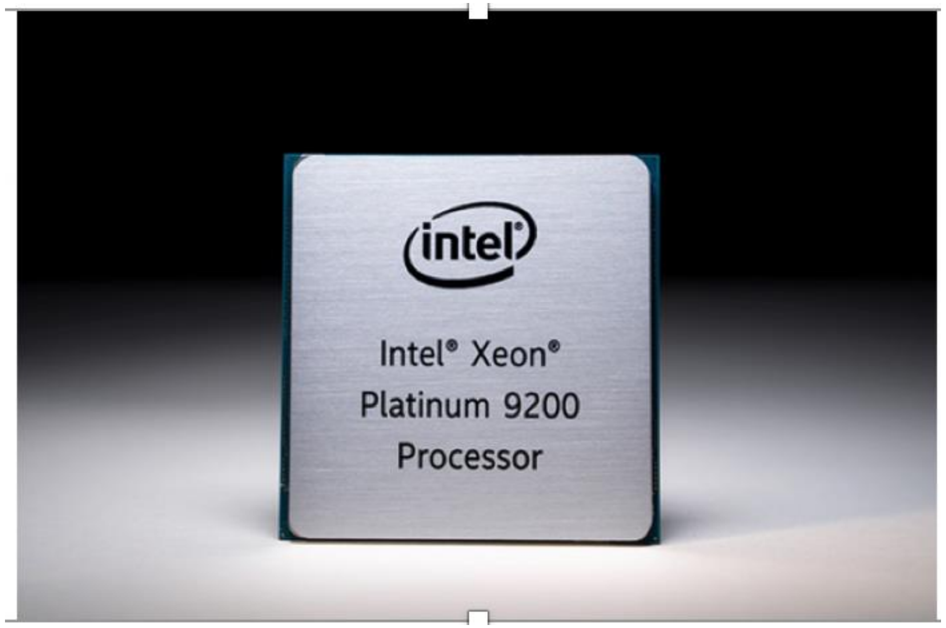
- Non-trivial Monte Carlo
  - Heston-based Greeks for multi-asset, path-dependent options with early exercise
  - Metrics: Speed, capacity, quality, efficiency
- Numerous reports
  - Some public, some in the STAC Vault
- Premium STAC members get:
  - Reports in STAC Vault
  - Detailed config info on public and private reports
  - Code from vendor implementations of the benchmarks

www.STACresearch.com/**a2**

Copyright © 2019 Securities Technology Analysis Center LLC

Intel 2S Server System S9200WK with Cascade Lake-AP CPUs

- SUT ID: INTC190903
- STAC Pack for Intel Parallel Studio XE (Rev M)
- Intel Server System S9200WK
- 2 x 48-core Intel Xeon Platinum 9242 (Cascade Lake-AP) processors
- RHEL 7.7
- 384GB DRAM

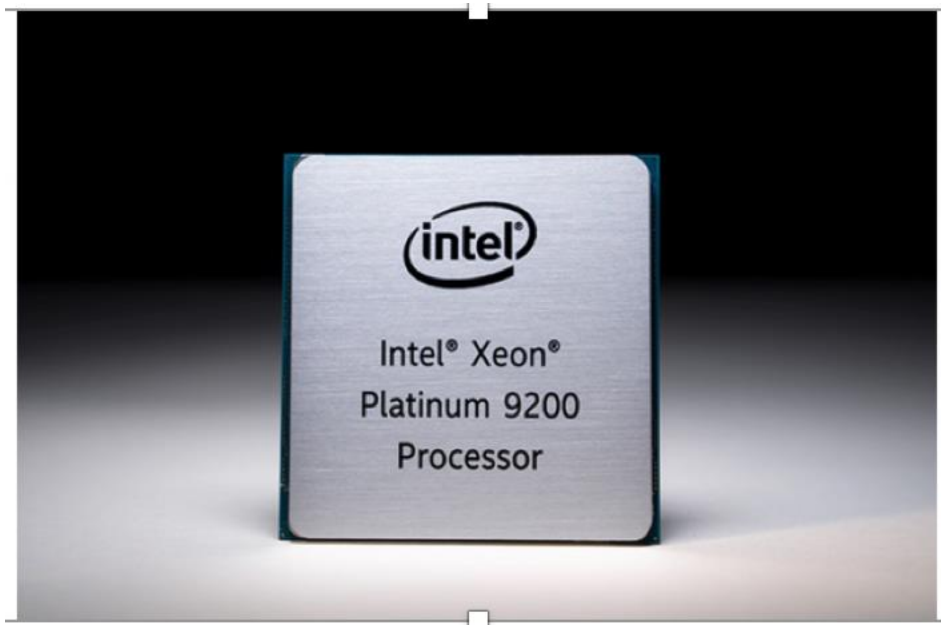


www.STACresearch.com/INTC190903

Copyright © 2019 Securities Technology Analysis Center LLC

Space efficiency vs. all other publicly disclosed results

- Solution based on Cascade Lake-AP set a new record for space efficiency (STAC-A2.β2.HPORTFOLIO. SPACE_EFF)
- This beats the previous record, also held by Intel (SUT ID INTC181012), by 8.5%



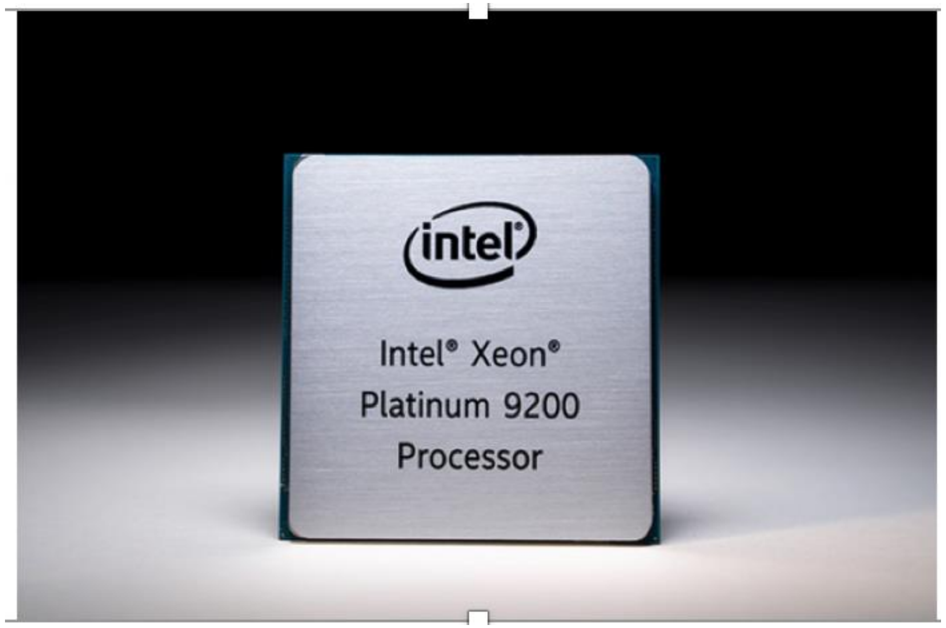
www.STACresearch.com/INTC190903

Copyright © 2019 Securities Technology Analysis Center LLC

Efficiency vs. most recently benchmarked using GPUs

Compared to the most recently benchmarked solution using GPUs (SUT ID NVDA181105), this system:

- Had over 1.8x the space efficiency (STAC-A2.β2.HPORTFOLIO.SPACE_EFF)
- Was within 20% of the energy efficiency (STAC-A2.β2.HPORTFOLIO.ENERG_EFF)



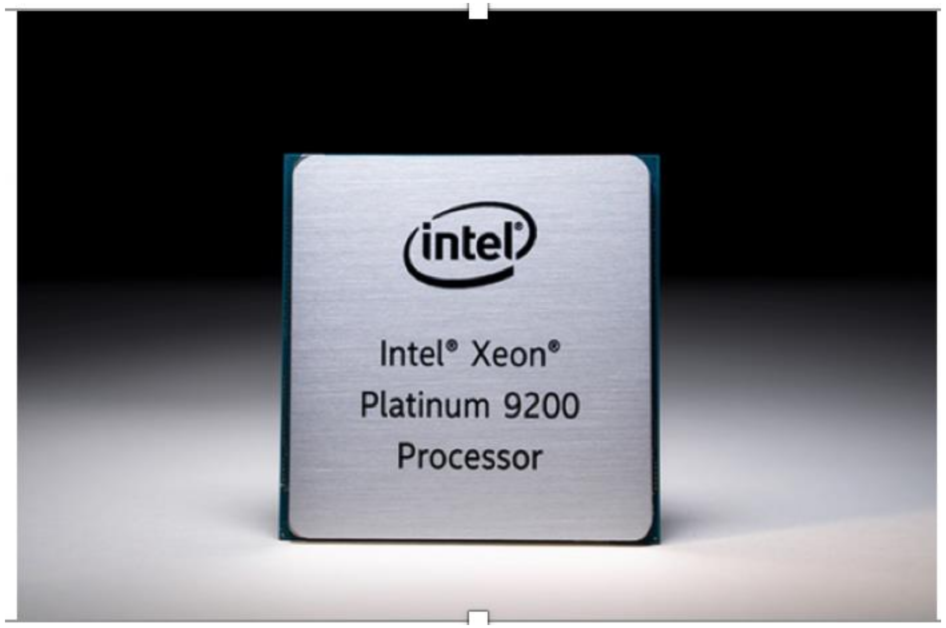
www.STACresearch.com/INTC190903

Copyright © 2019 Securities Technology Analysis Center LLC

Versus a "non-AP" Cascade Lake based solution

Compared to a similar 2S system using Platinum 8280 ("Cascade Lake") processors (SUT ID INTC190402), this system had:

- 2.67x the space efficiency (STAC-A2.β2.HPORTFOLIO.SPACE_EFF)
- 1.84x the throughput (STAC-A2.β2.HPORTFOLIO.SPEED)
- 1.35x the energy efficiency (STAC-A2.β2.HPORTFOLIO.ENERG_EFF)



www.STACresearch.com/INTC190903

Copyright © 2019 Securities Technology Analysis Center LLC

AI Benchmark PoCs

- Deep time series (training)
  - Small number of largely homogeneous data types per symbol
  - Long, dense time series
- Wide time series (training)
  - Large collection of heterogenous data types per symbol
  - Often have regular but different frequencies
- Entity matching (inference)
  - Match descriptions of companies to unique identifiers
- NLP (training)
  - Topic modeling of business description in annual report required from U.S. firms
  - STAC Study in Vault: "*Scaling a common machine learning workload in the cloud*"
  - Excerpts on STAC website: STAC Study Excerpts - NLP topic modeling 2018

Copyright © 2019 Securities Technology Analysis Center LLC

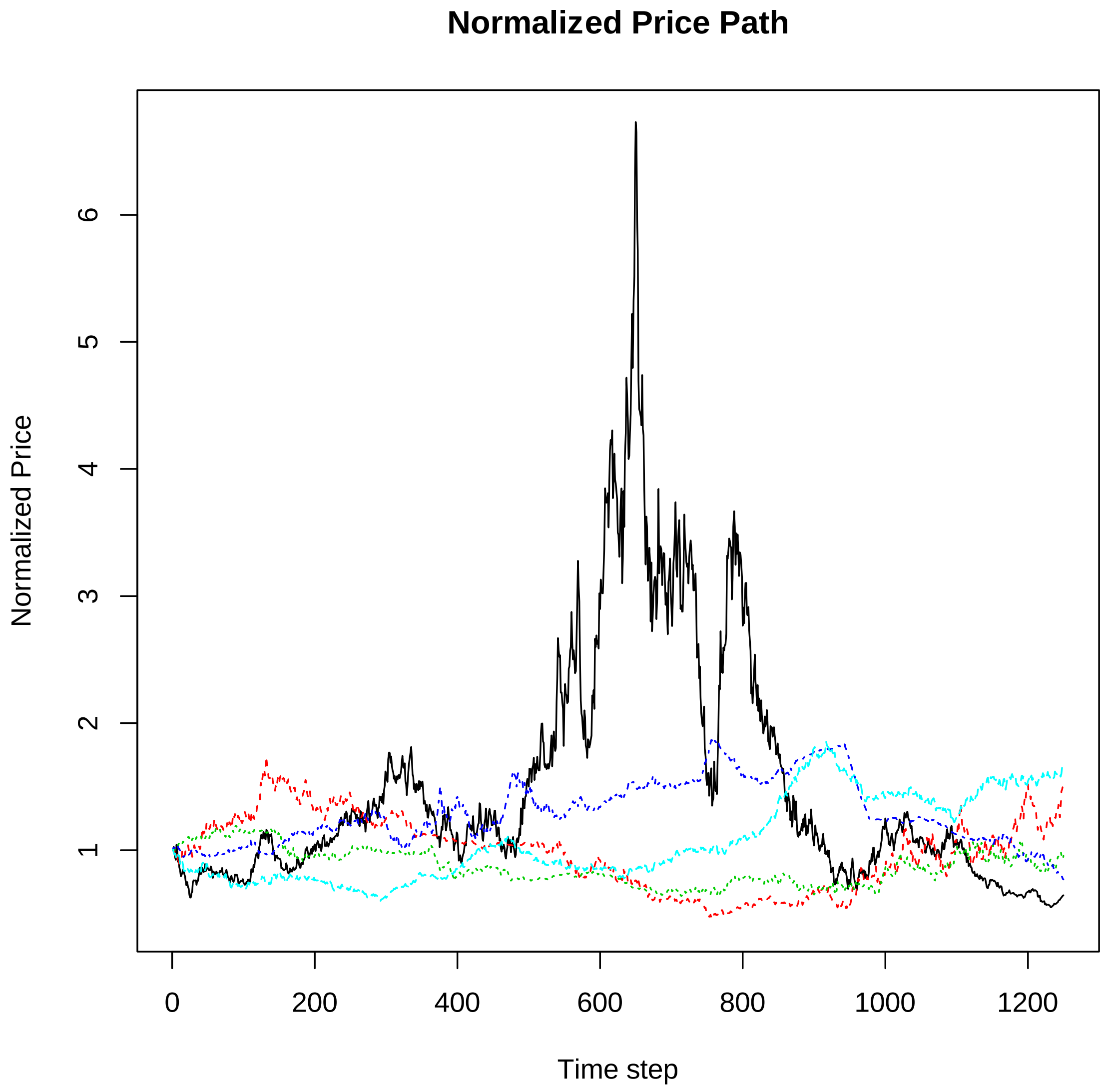
AI Benchmark PoCs

- Deep time series (training)
  - Small number of largely homogeneous data types per symbol
  - Long, dense time series
- Wide time series (training)
  - Large collection of heterogenous data types per symbol
  - Often have regular but different frequencies
- Entity matching (inference)
  - Match descriptions of companies to unique identifiers
- NLP (training)
  - Topic modeling of business description in annual report required from U.S. firms
  - STAC Study in Vault: "*Scaling a common machine learning workload in the cloud*"
  - Excerpts on STAC website: STAC Study Excerpts - NLP topic modeling 2018

Copyright © 2019 Securities Technology Analysis Center LLC

AI benchmarks in finance: the data problem

- Core challenge with real data
  - Don’t know underlying ”real” signal
  - Neural net models are opaque
  - How to judge relative performance?
- Simulated data advantages
  - Specify signal
  - Vary attributes e.g. signal:noise ratio
  - Compare model performance to reality

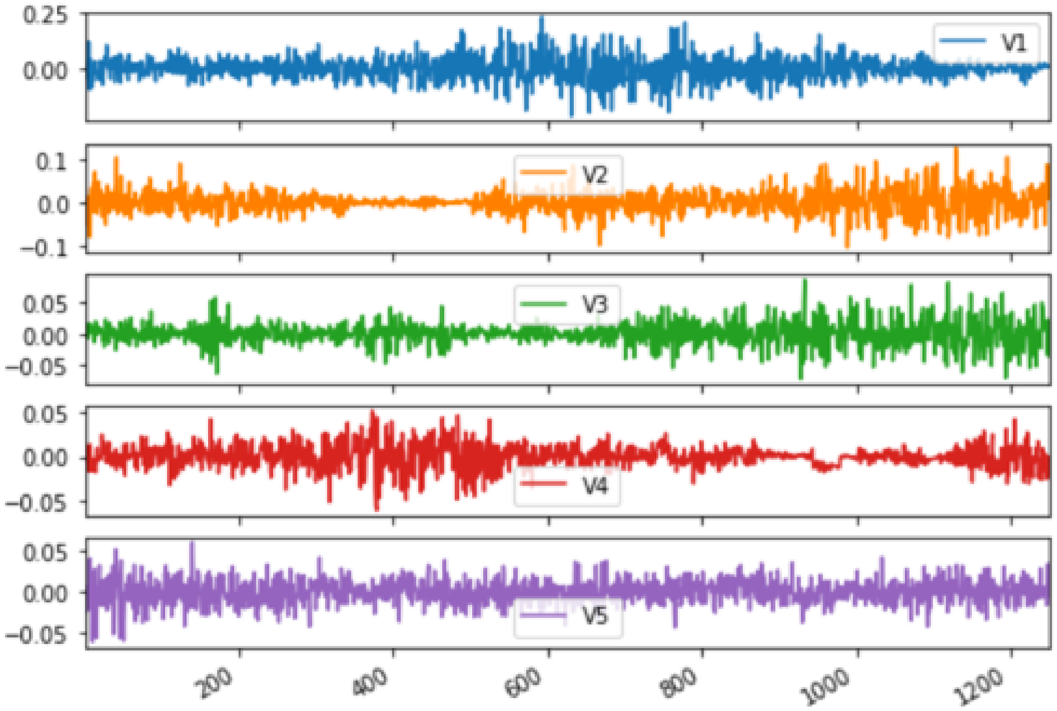


Copyright © 2019 Securities Technology Analysis Center LLC

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

Deep Time Series: Status of crawl, walk, run approach

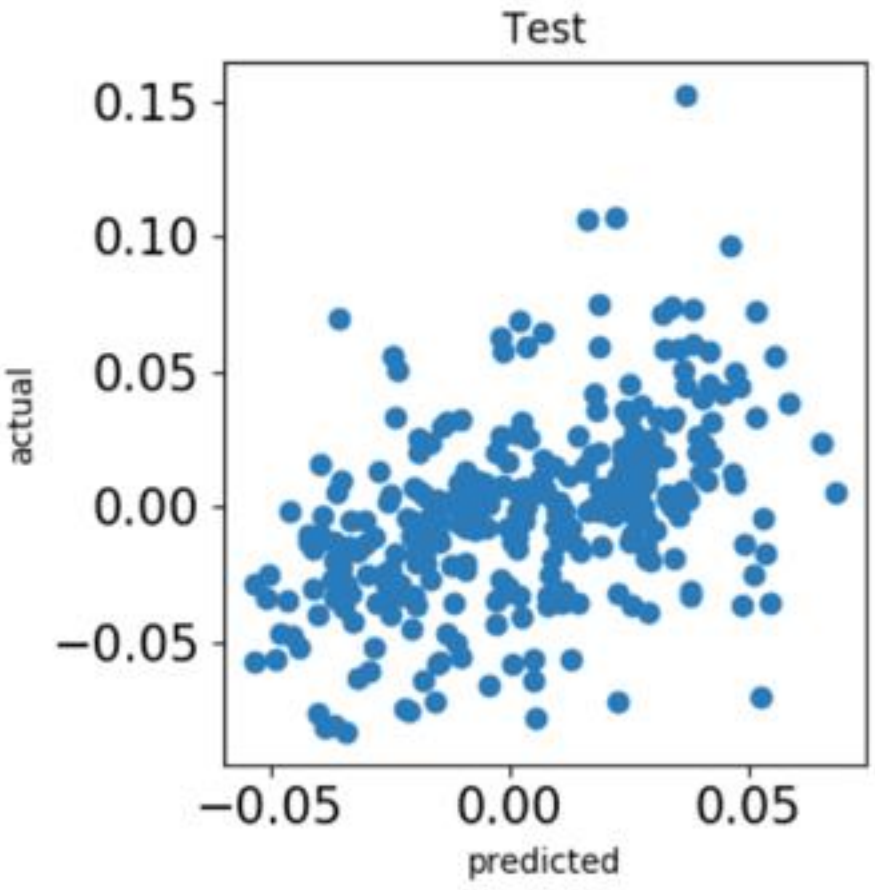
- Leveraged existing STAC-A2 path generation
- Generate simulated market data with known stochastic properties
- Apply random correlation matrix
- Overlay simple signals
- Can this data be modeled?



Copyright © 2019 Securities Technology Analysis Center LLC

Deep Time Series: Status of crawl, walk, run approach

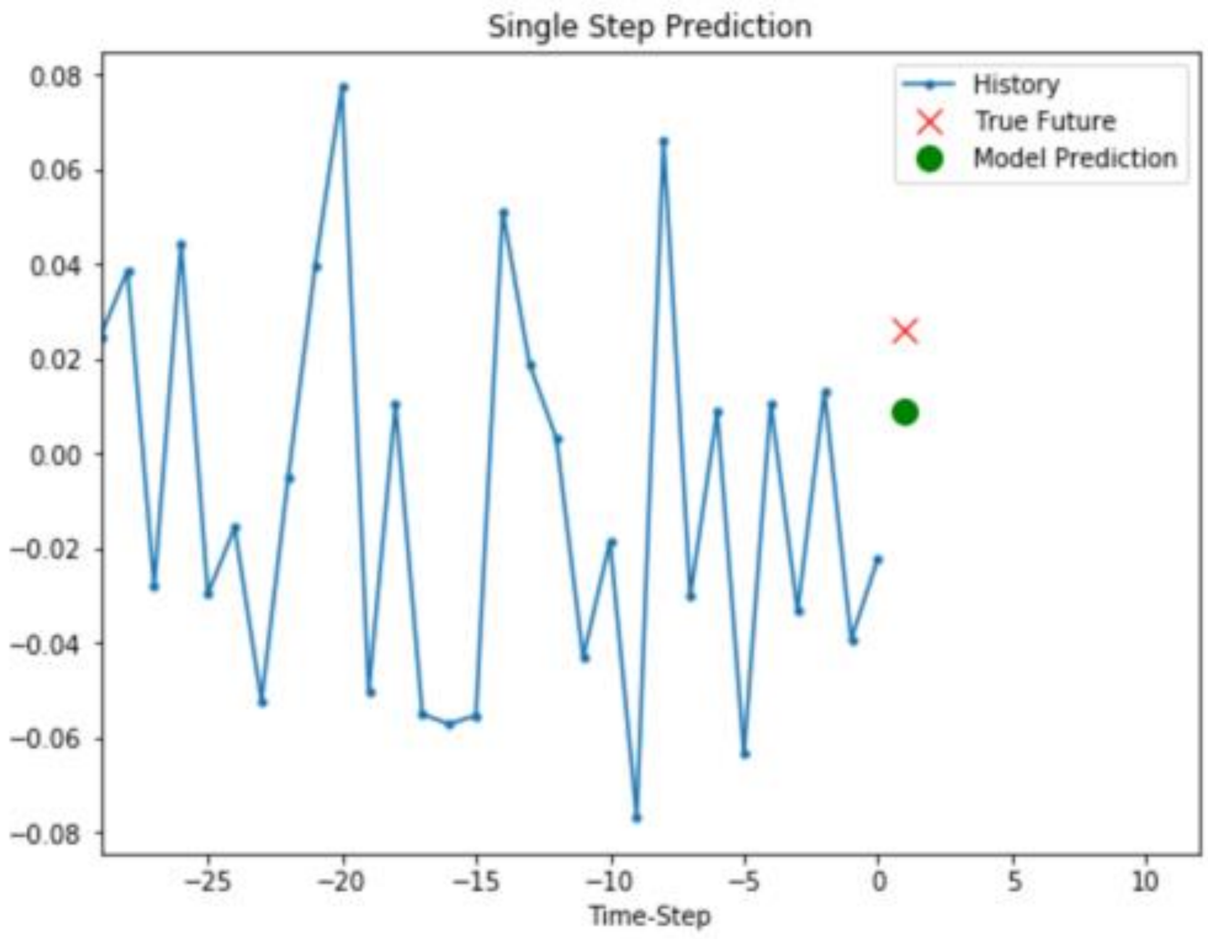
- On “tiny” problem, an AutoML tool finds models with tunable accuracy
- First does automated feature engineering
- Runs for a couple of hours



Copyright © 2019 Securities Technology Analysis Center LLC

Deep Time Series: Status of crawl, walk, run approach

- Preliminary work with LSTM in Tensorflow
- Experimenting with attributes of the overlaid signal
- Experimenting with attributes of the model
- Results will be provided to the Working Group



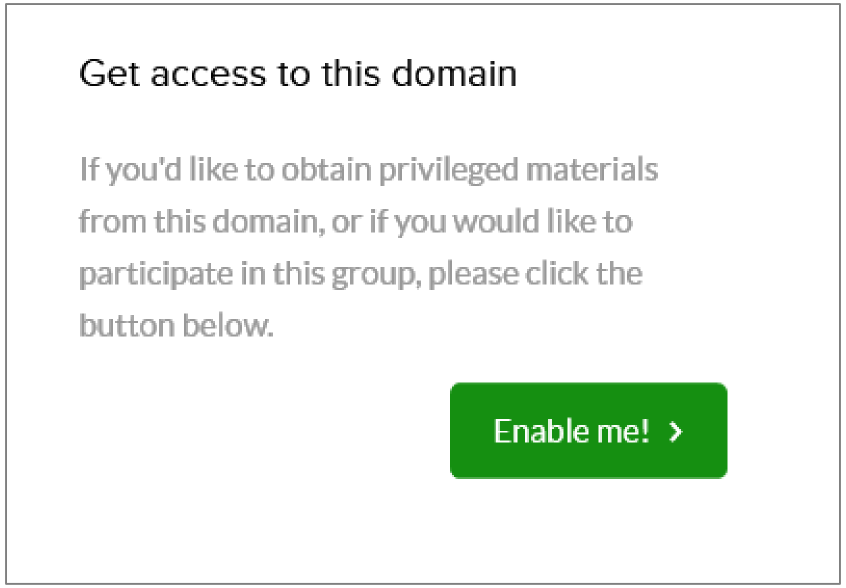
Copyright © 2019 Securities Technology Analysis Center LLC

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

Deep Time Series: Next steps and call to action

- We have data generator and outlines of the work load
- Will refine in the next couple of months
- Will post outstanding issues and ways to help in online forum
- To join the STAC-AI Working Group:
  - Go to www.STACresearch.com/ai, right side of the page
  - If you see the “Group Members” list, you’re are already in it
  - If not, click on the “Enable Me” button

www.STACresearch.com/ai



Get access to this domain

If you'd like to obtain privileged materials from this domain, or if you would like to participate in this group, please click the button below.

Copyright © 2019 Securities Technology Analysis Center LLC