AUTOMATED & DISTRIBUTED **STATISTICAL ANALYSIS** of ECONOMIC **ABMs**

Andrea Vandin

Sant'Anna School of Advanced Studies – Pisa

Institute of Economics

Department of Excellence 2018 - 2022

EMbeDS Economics and Management in the era of Data Science

Joint work with

Daniele Giachini, Francesco Lamperti, Francesca Chiaromonte

LEM Working Paper available at: www.lem.sssup.it/WPLem/2020-31.html

Tool and models available at: https://bit.ly/MultiVeStATool

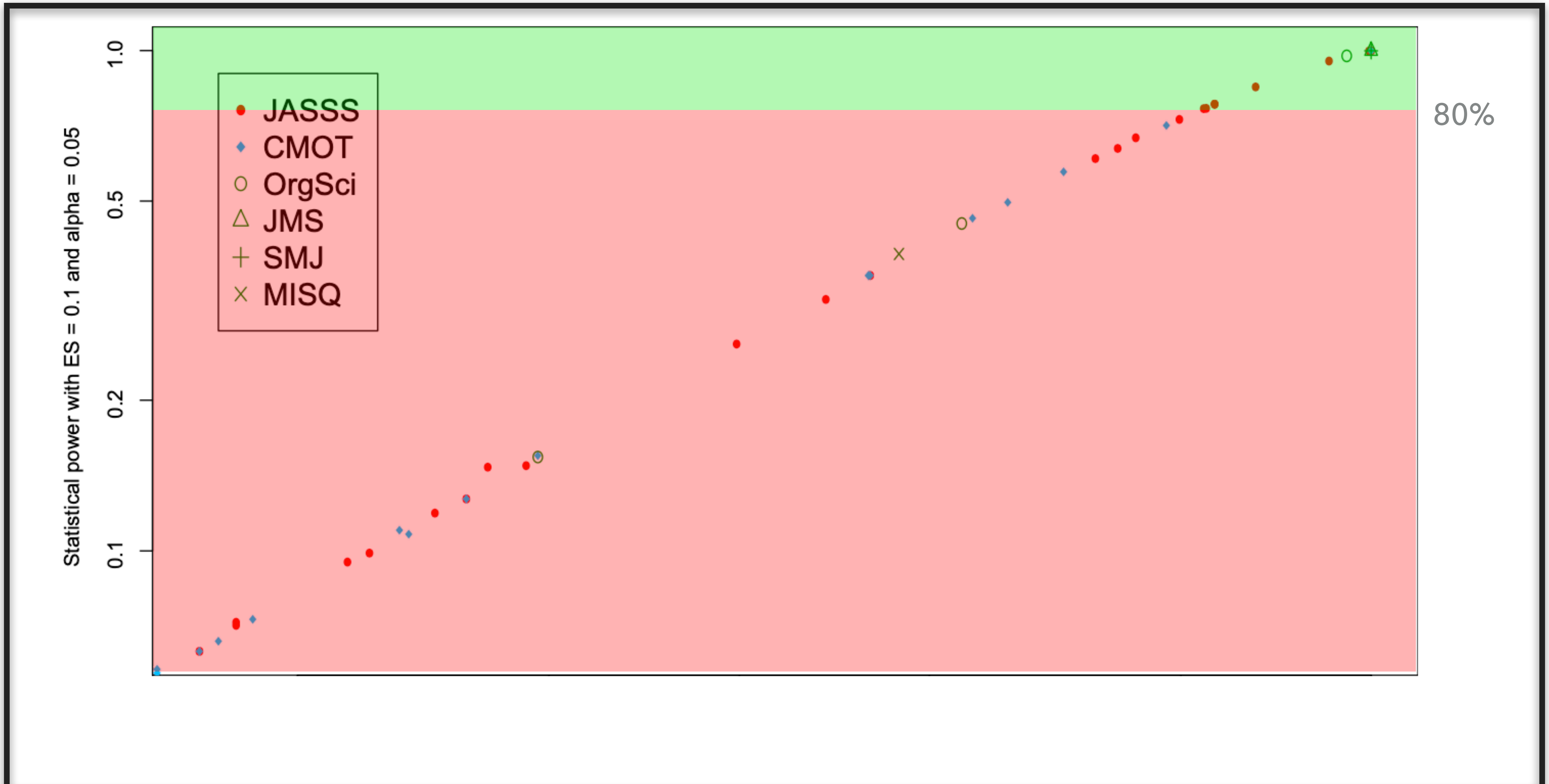
CEF 2021, Session F1: Agent-Based Models (II), 17/06/2021

OUTLINE

1. Motivation, vision, and proposal
    1. Automated analysis with statistical guarantees for ABMs
    2. The MultiVeStA Statistical Model Checker
2. Transient Analysis of a large-scale financial macro ABM
    1. Estimation of expected outcome and Confidence Interval
    2. Counterfactual analysis for different model configurations
3. Steady-state analysis of a prediction market model
    1. Steady-state analysis by Replication and Deletion (RD)
    2. Warmup estimation
    3. Steady-state analysis by Batch Means (BM)
    4. A methodology for ergodicity analysis based on RD and BM
4. Conclusions & Future works

‘Quality’ of Statistical Analysis on 55 ABM from Management & Organisational Research

Adapted from Secchi, Seri, Computational and Mathematical Organization Theory, 2017[Secchi,Seri2017]



- The importance of designing well simulation-based analysis.
  - Power analysis on ‘are the expected outcomes of different configurations of parameters the same’?
- Power is 1 - P(Type II error)
  - Roughly, P(test=‘outcomes are different’ | outcomes are different)
  - “The value that seems to be more commonly accepted is 80%”
- “We need to encourage researchers to be more precise in the determination of the number of runs”

Similar studies can be found also in other communities

A systematic review of statistical power in software engineering experiments

Tore Dybå [a,b,*], Vigdis By Kampenes [a], Dag I.K. Sjøberg [a]

[a] *Simula Research Laboratory, P.O. Box 134, NO-1325 Lysaker, Norway*
[b] *SINTEF ICT, NO-7465 Trondheim, Norway*

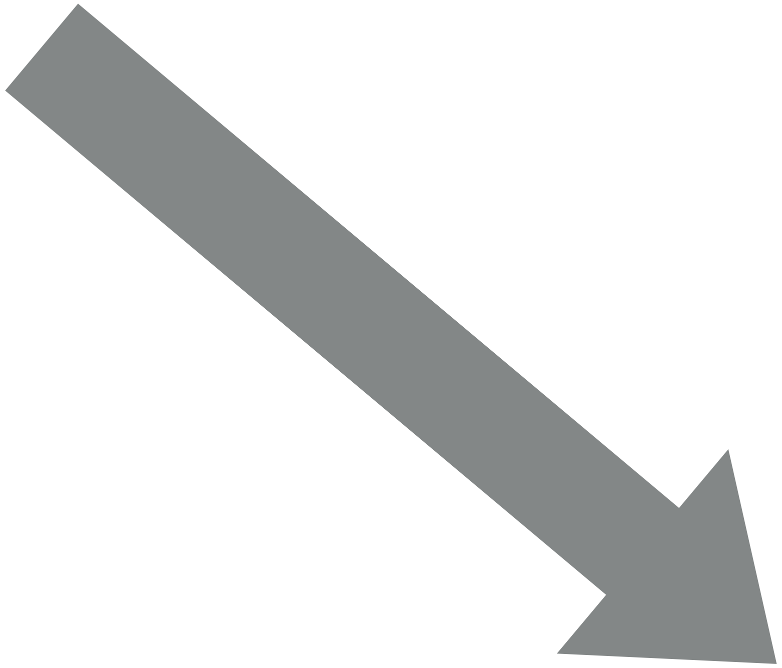
Received 11 May 2005; revised 24 August 2005; accepted 31 August 2005
Available online 3 November 2005

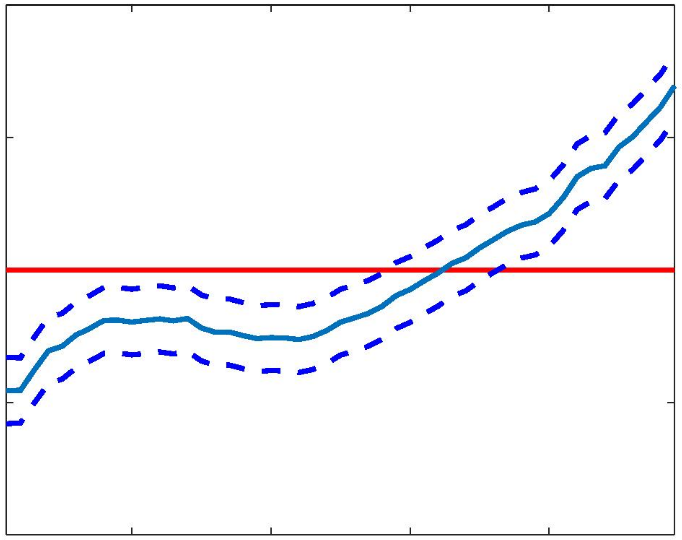
Abstract

Statistical power is an inherent part of empirical studies that employ significance testing and is essential for the planning of studies, for the interpretation of study results, and for the validity of study conclusions. This paper reports a quantitative assessment of the statistical power of empirical software engineering research based on the 103 papers on controlled experiments (of a total of 5,453 papers) published in nine major software engineering journals and three conference proceedings in the decade 1993–2002. The results show that the statistical power of software engineering experiments falls substantially below accepted norms as well as the levels found in the related discipline of information systems research. Given this study's findings, additional attention must be directed to the adequacy of sample sizes and research designs to ensure acceptable levels of statistical power. Furthermore, the current reporting of significance tests should be enhanced by also reporting effect sizes and confidence intervals.

Our Proposed Approach to Simulation-Based Analysis

newstalkzb.co.nz/news/education/modern-lego-sets-more-complex-less-inspiring/





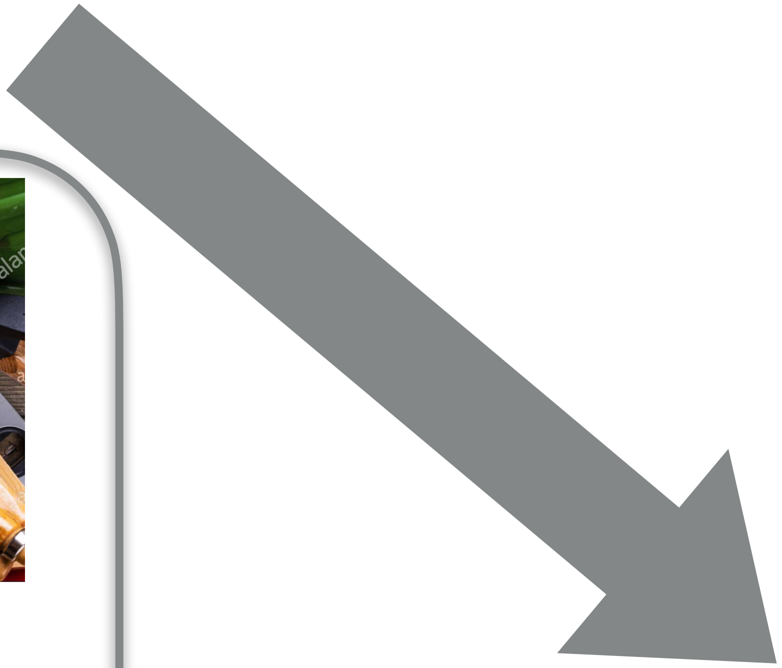
Our Proposed Approach to Simulation-Based Analysis

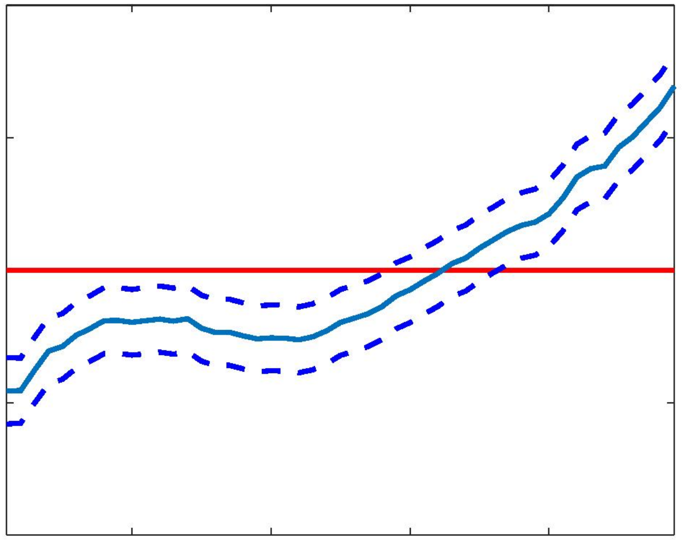
newstalkzb.co.nz/news/education/modern-lego-sets-more-complex-less-inspiring/

https://www.alamy.com/

Handcrafted

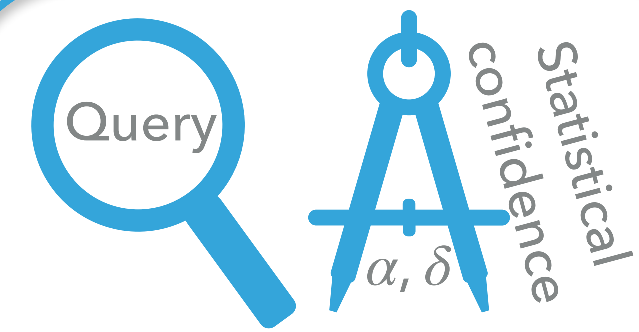
- Mainly manual process
  - Time-consuming
  - Problems with replicability
  - Error-prone
    - Modify model, interpret CSV
- Ad-hoc implementations
  - Reliability? Efficiency?





Our Proposed Approach to Simulation-Based Analysis

newstalkzb.co.nz/news/education/modern-lego-sets-more-complex-less-inspiring/



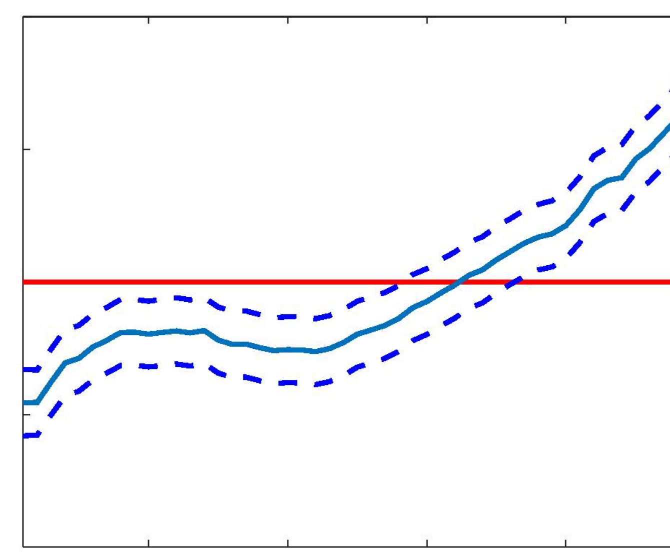
Statistical Model Checking

- Automatic
  - Time-saving and Reproducible
  - Promotes use of *standard* analysis
- Reference implementation
  - Reliable and Efficient

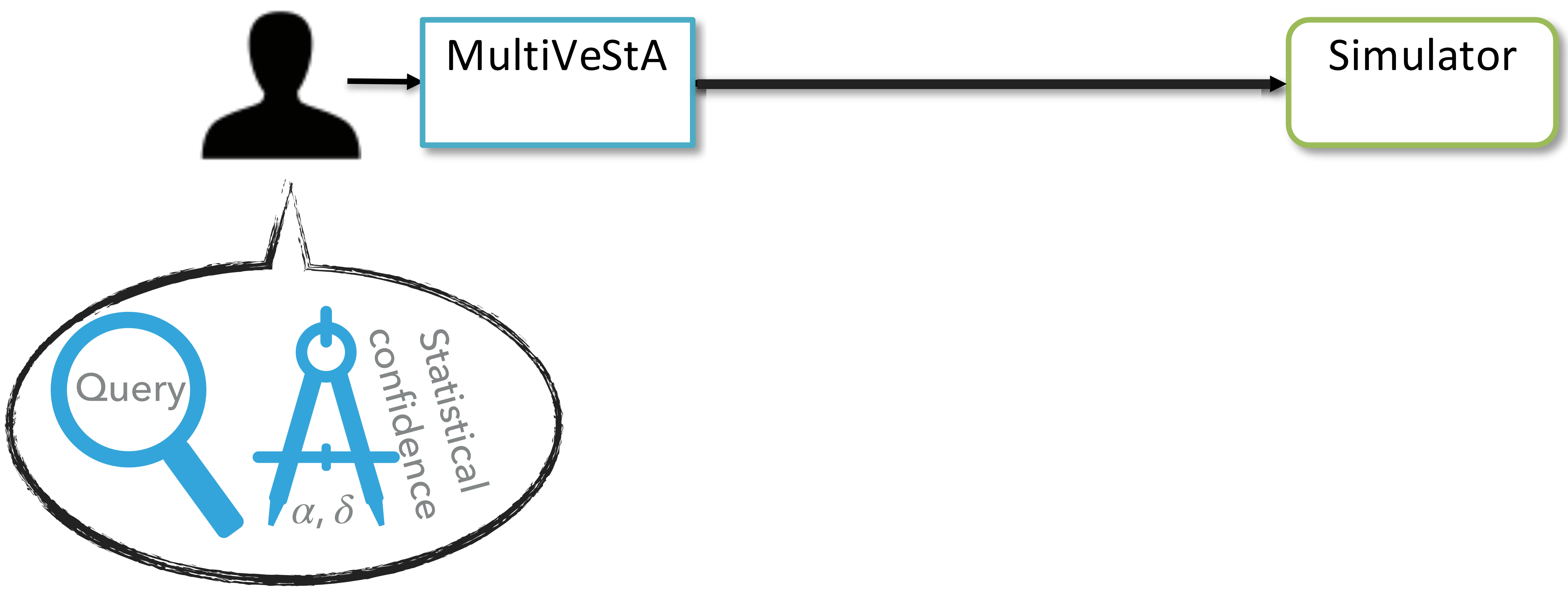
https://www.alamy.com/

Handcrafted

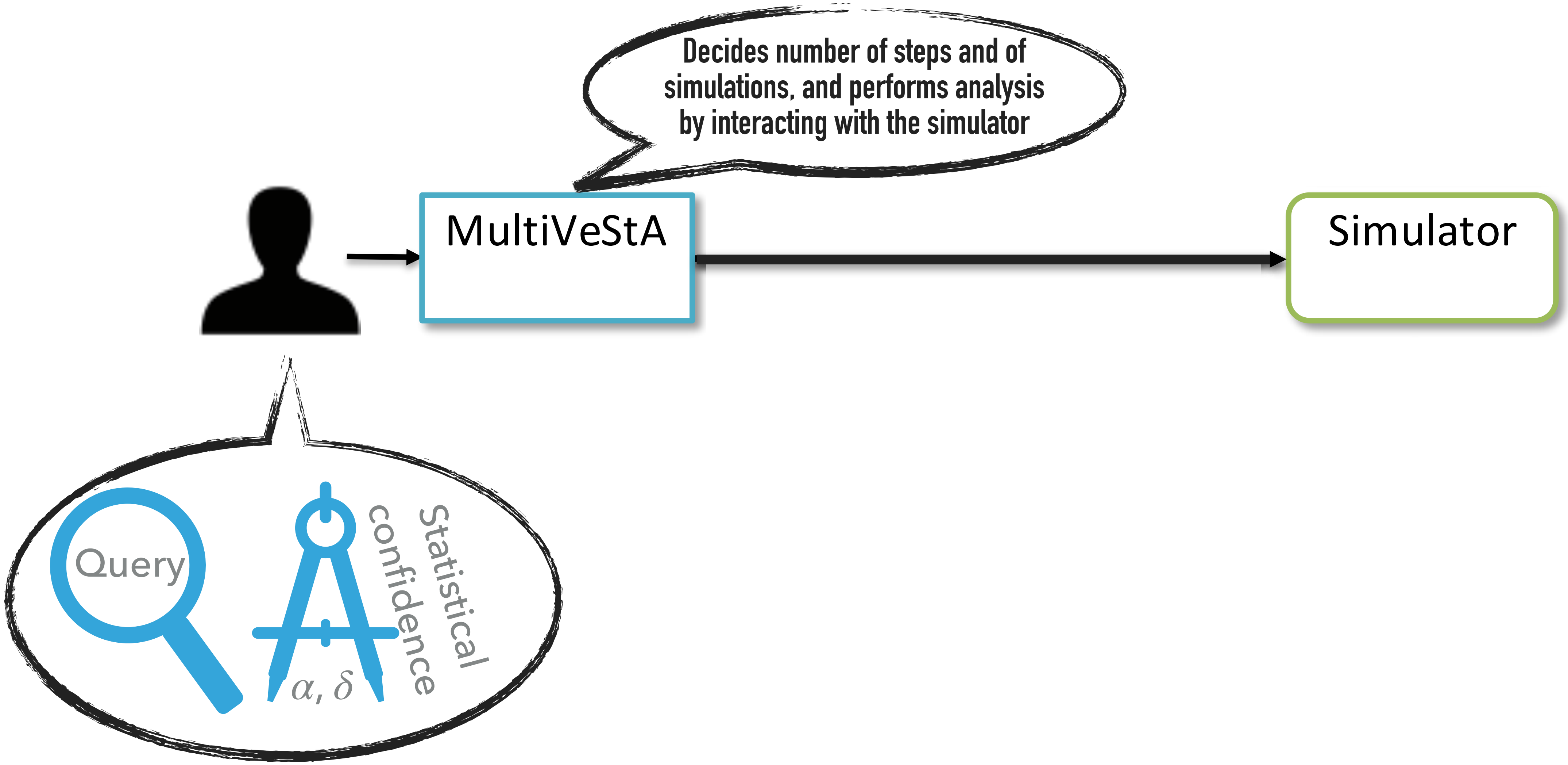
- Mainly manual process
  - Time-consuming
  - Problems with replicability
  - Error-prone
    - Modify model, interpret CSV
- Ad-hoc implementations
  - Reliability? Efficiency?



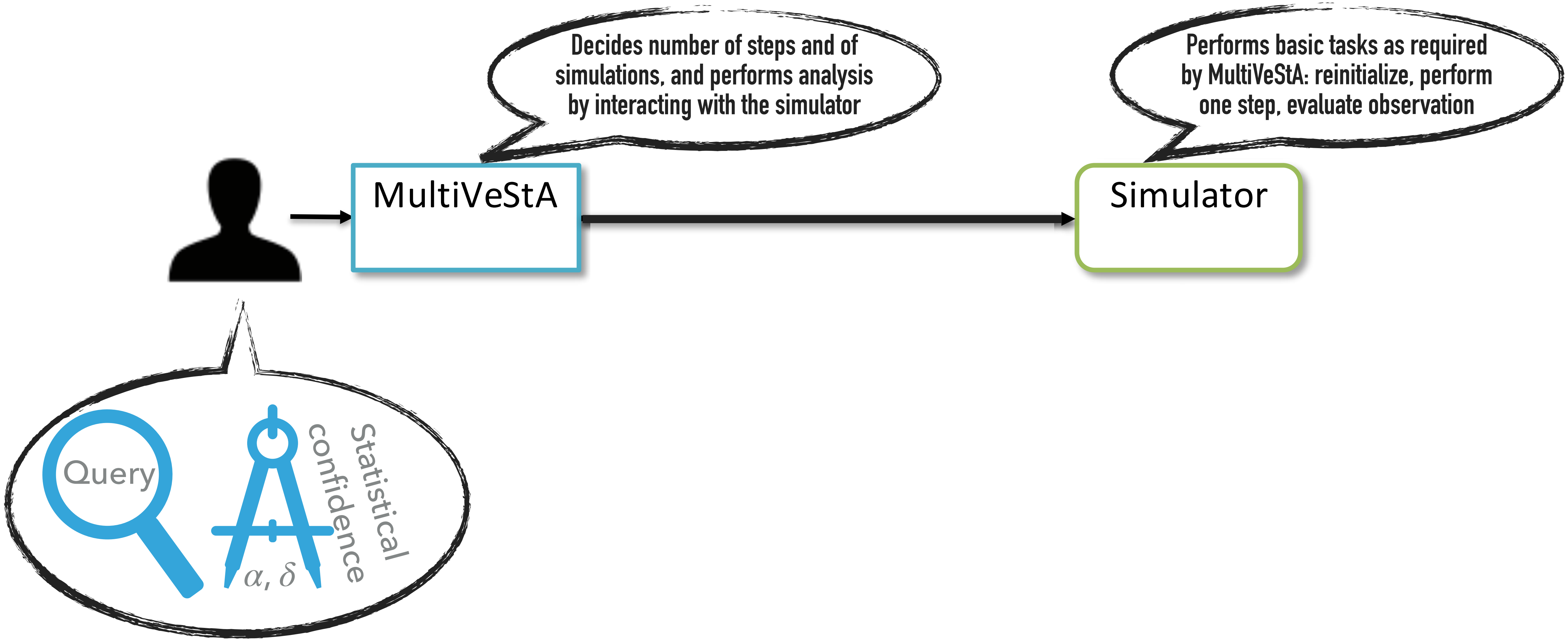
MultiVeStA: SMC For Discrete-Event Simulators



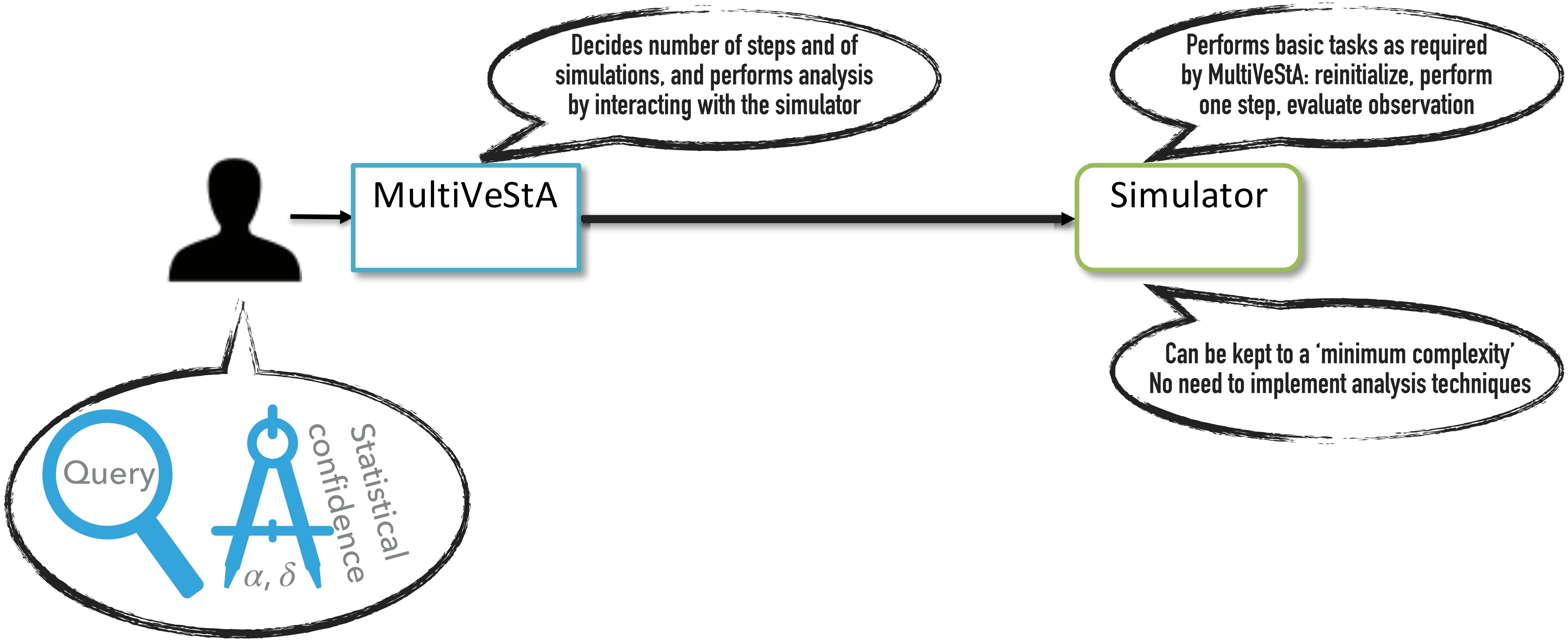
MultiVeStA: SMC For Discrete-Event Simulators



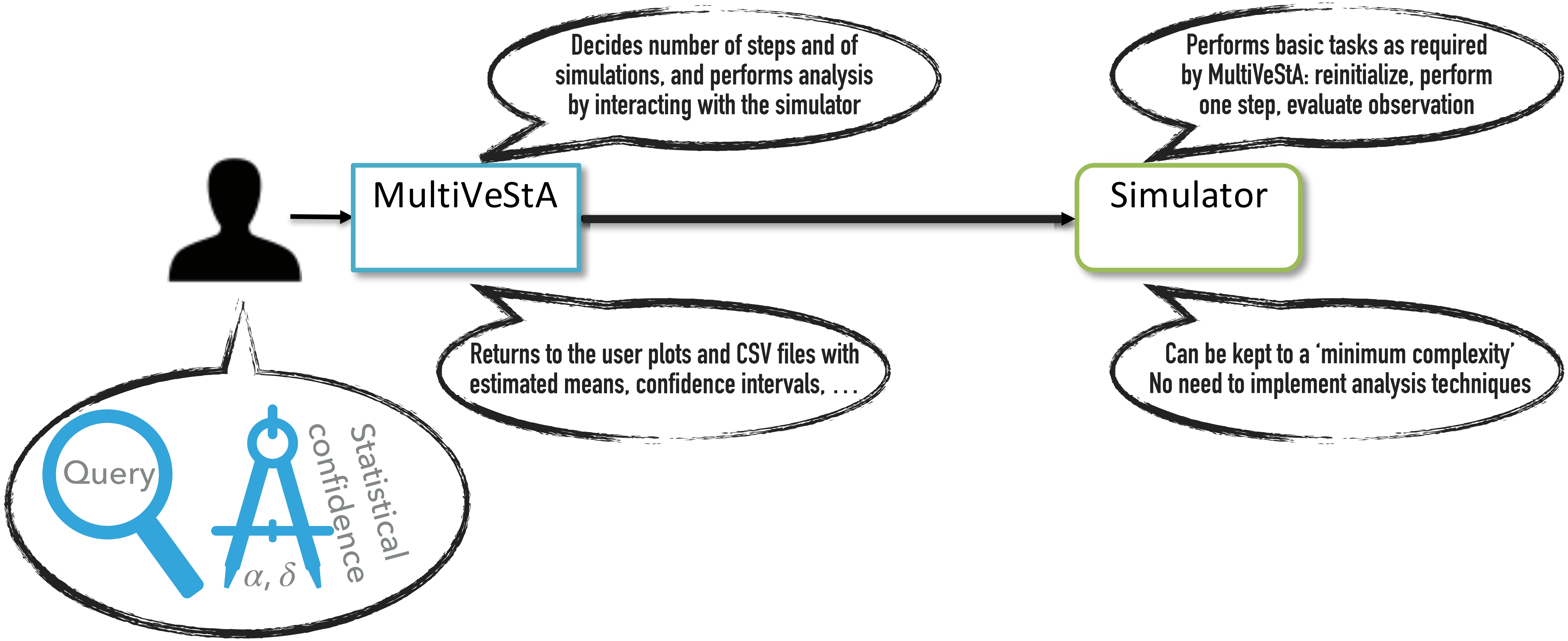
MultiVeStA: SMC For Discrete-Event Simulators



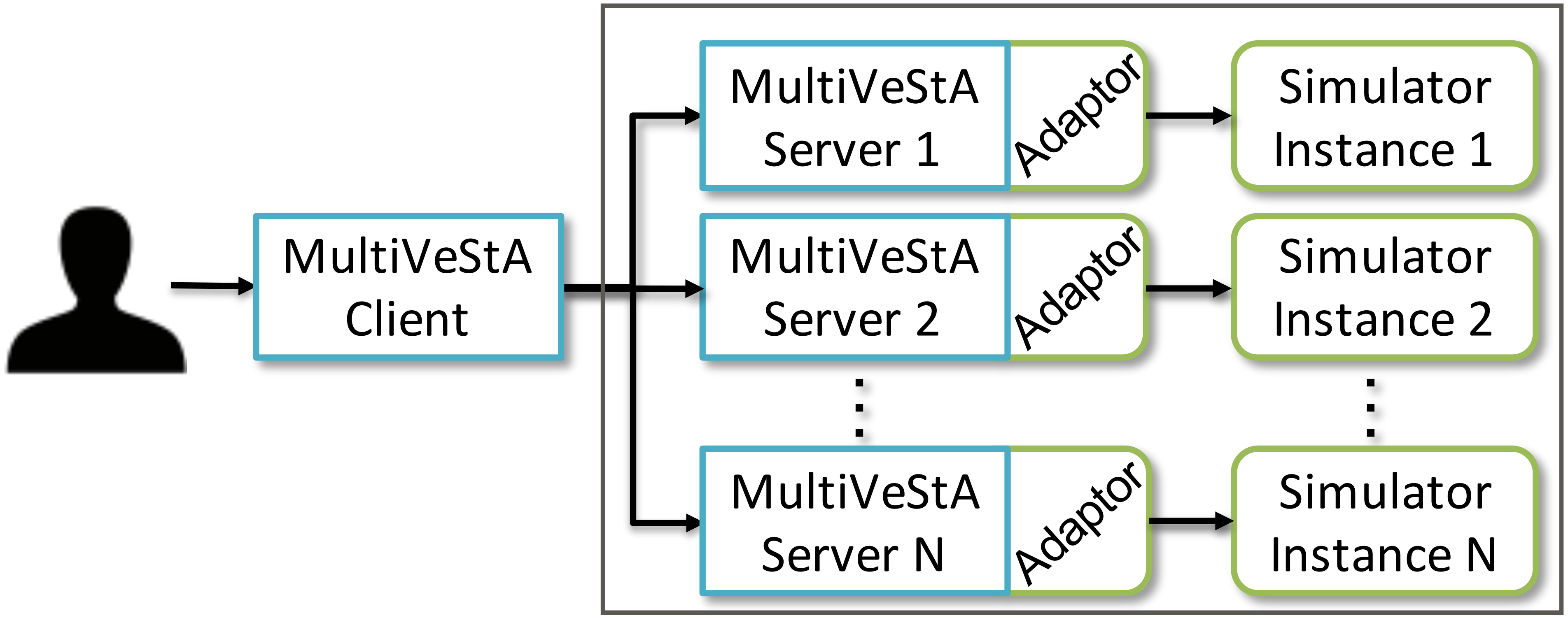
MultiVeStA: SMC For Discrete-Event Simulators

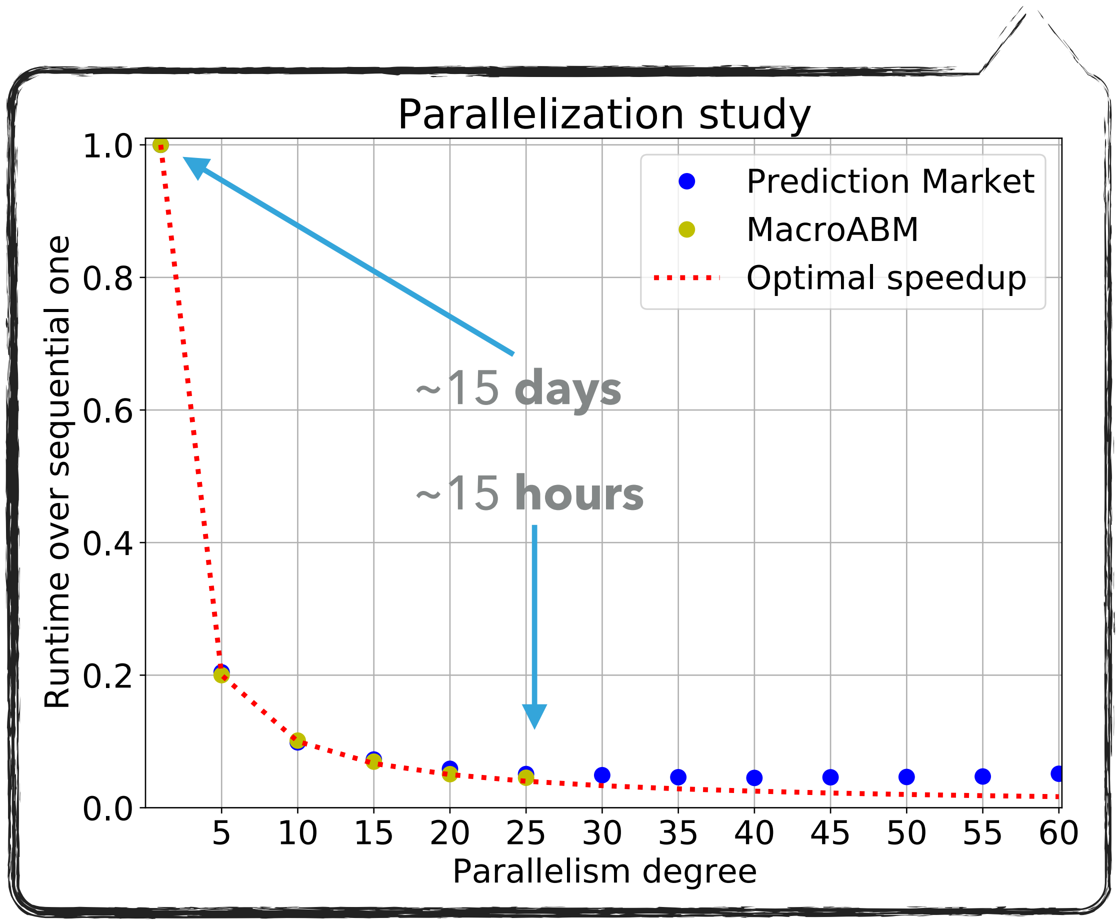


MultiVeStA: SMC For Discrete-Event Simulators

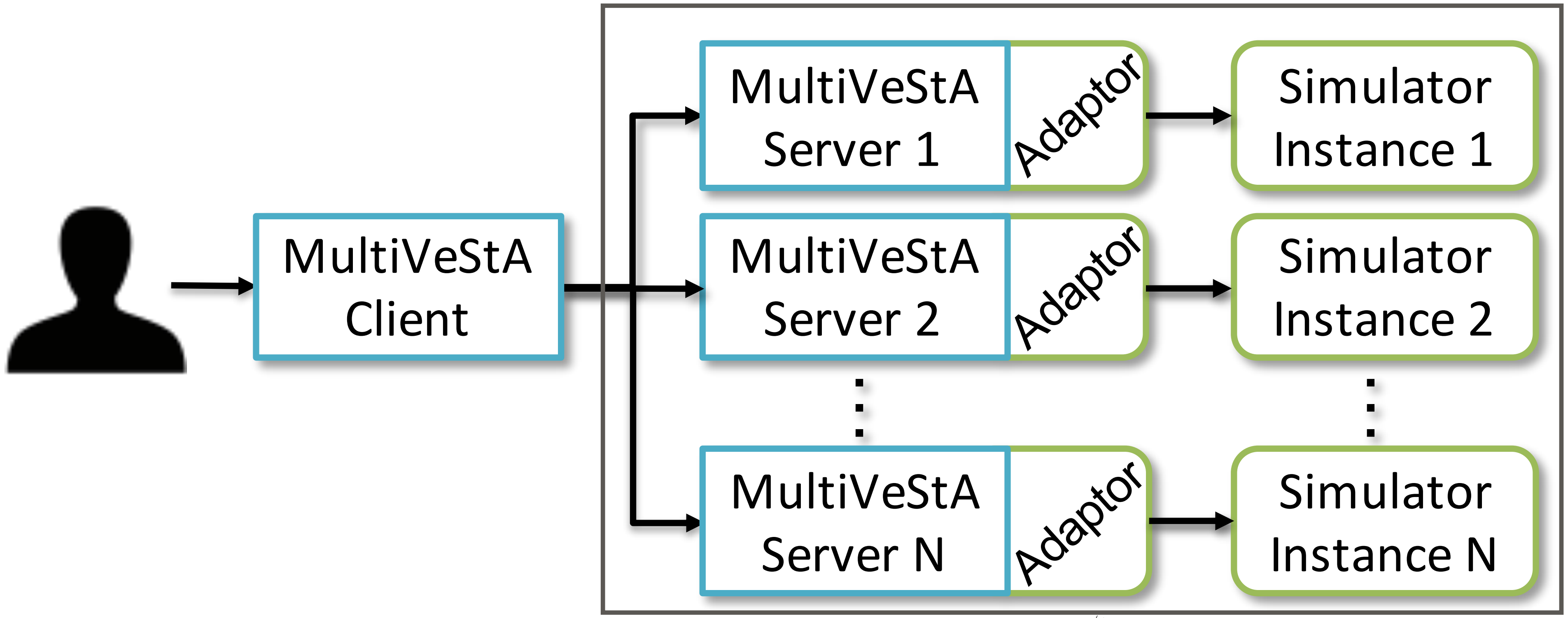


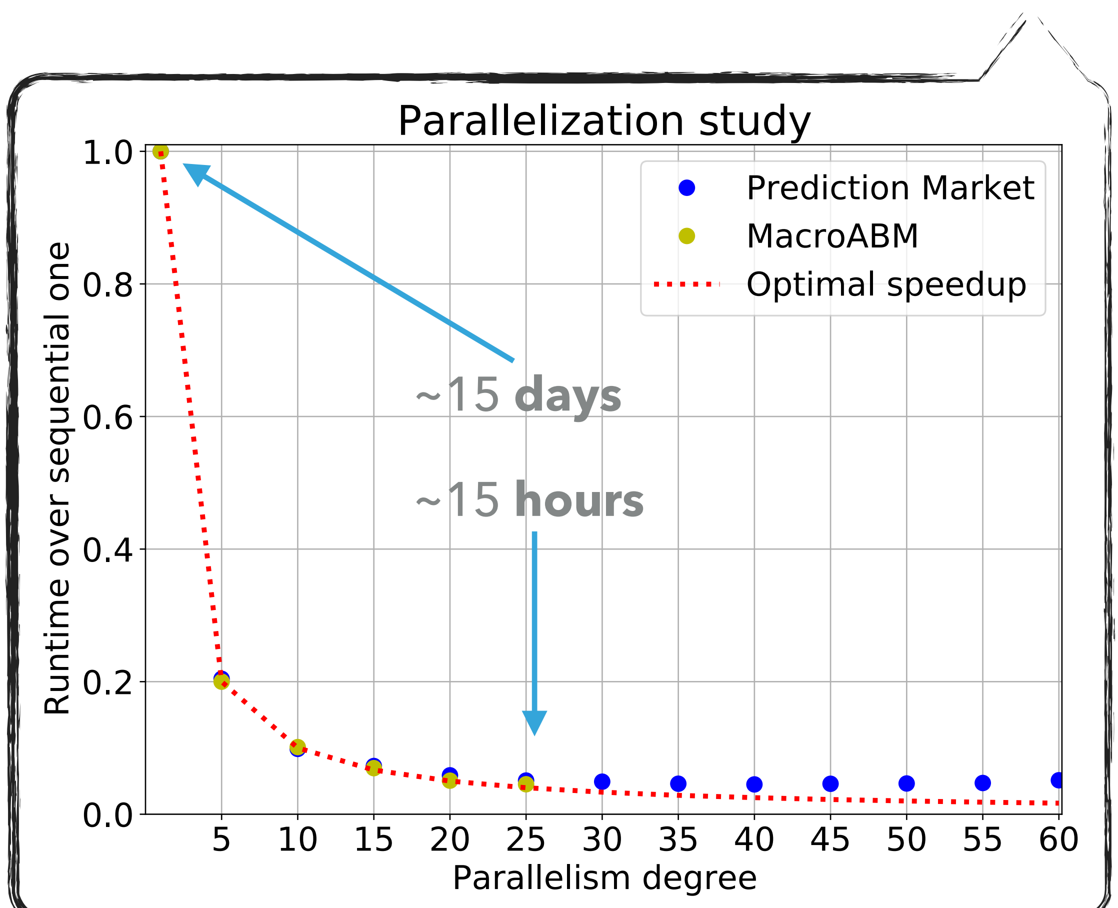
MultiVeStA: SMC For Discrete-Event Simulators





MultiVeStA: SMC For Discrete-Event Simulators





OUTLINE

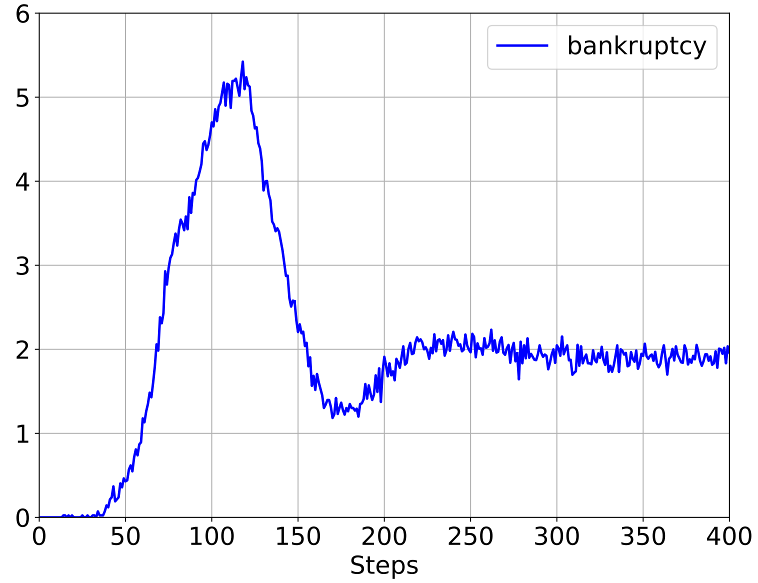
1. Motivation, vision, and proposal
    1. Automated analysis with statistical guarantees for ABMs
    2. The MultiVeStA Statistical Model Checker
2. **Transient Analysis of a large-scale financial macro ABM**
    1. Estimation of expected outcome and Confidence Interval
    2. Counterfactual analysis for different model configurations
3. Steady-state analysis of a prediction market model
    1. Steady-state analysis by Replication and Deletion (RD)
    2. Warmup estimation
    3. Steady-state analysis by Batch Means (BM)
    4. A methodology for ergodicity analysis based on RD and BM
4. Conclusions & Future works

Transient Analysis by autoIR: Large Macro ABM

Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

Evolution of bankruptcies



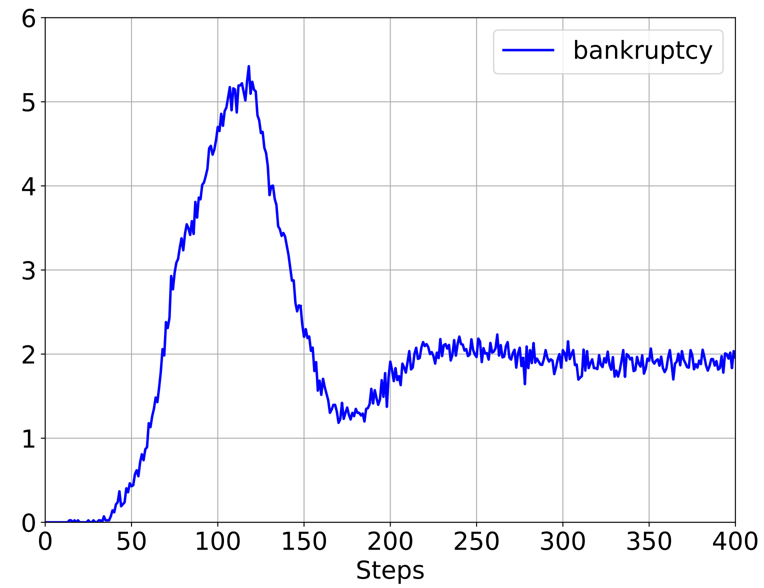
Transient Analysis by autoIR: Large Macro ABM

Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

$$y_1 \quad y_2 \quad \cdots \quad y_{400}$$

Evolution of bankruptcies

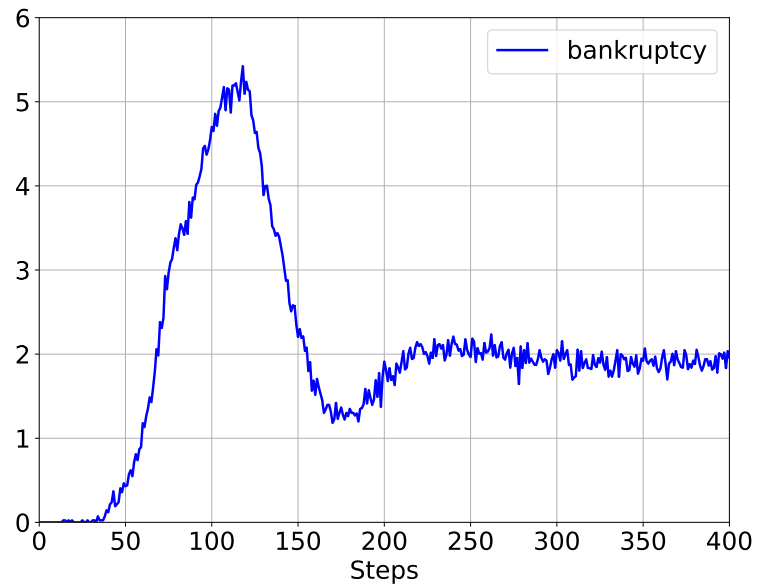


Transient Analysis by autoIR: Large Macro ABM

Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

Evolution of bankruptcies



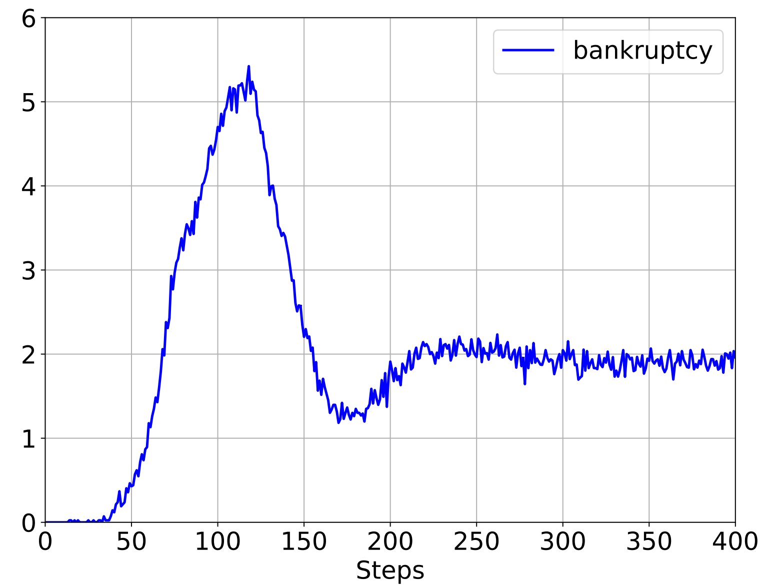
$y_{1,1}$	$y_{1,2}$	$\cdots$	$y_{1,400}$
$y_{2,1}$	$y_{2,2}$	$\cdots$	$y_{2,400}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{n,1}$	$y_{n,2}$	$\cdots$	$y_{n,400}$

Transient Analysis by autoIR: Large Macro ABM

Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

Evolution of bankruptcies



$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,400} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,400} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,400} \end{array}$$

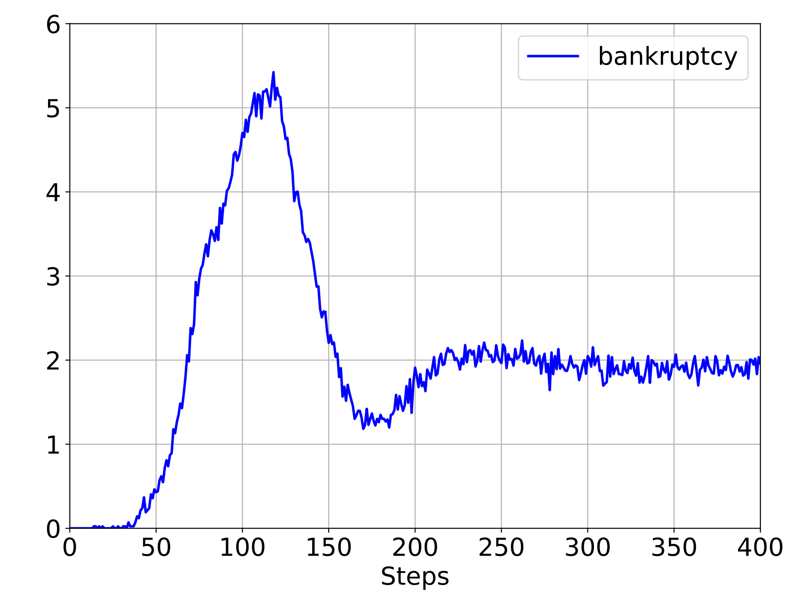
$$\sum_{i=1}^{n} \frac{y_{i,2}}{n} = \overline{Y}_2 \approx E[Y_2]$$

Transient Analysis by autoIR: Large Macro ABM

Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

Evolution of bankruptcies



$$\begin{matrix} y_{1,1} & y_{1,2} & \cdots & y_{1,400} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,400} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,400} \end{matrix}$$

$$\sum_{i=1}^{n} \frac{y_{i,2}}{n} = \overline{Y}_2 \approx E[Y_2]$$

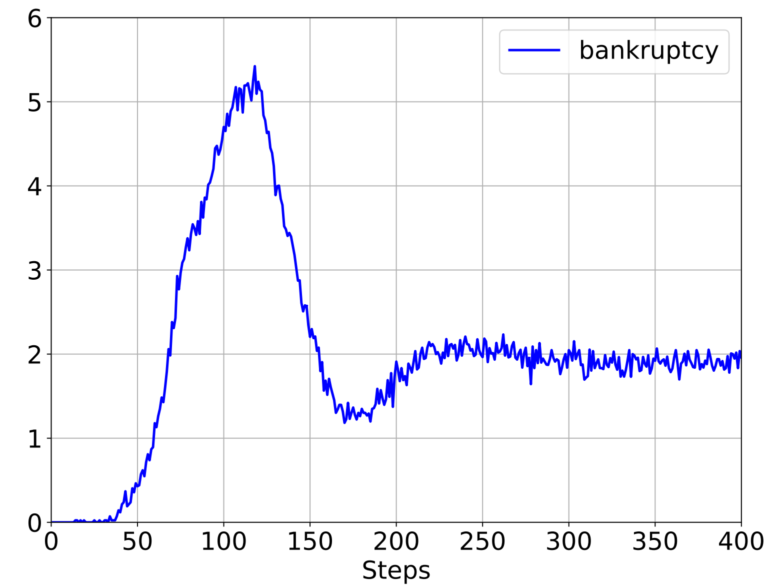
What is the correct value of n?

Transient Analysis by autoIR: Large Macro ABM

Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

Evolution of bankruptcies



$$\begin{matrix} y_{1,1} & y_{1,2} & \cdots & y_{1,400} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,400} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,400} \end{matrix}$$

$$\sum_{i=1}^{n} \frac{y_{i,2}}{n} = \overline{Y}_2 \approx E[Y_2]$$

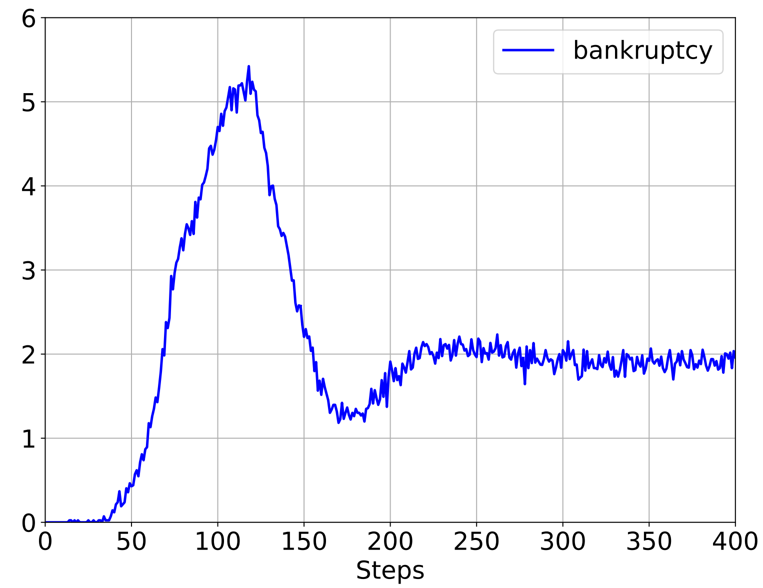
What is the correct value of n?
Typical answer: 100

Transient Analysis by autoIR: Large Macro ABM

Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

Evolution of bankruptcies



$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,400} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,400} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,400} \end{array}$$

$$\sum_{i=1}^{n} \frac{y_{i,2}}{n} = \overline{Y}_2 \approx E[Y_2]$$

What is the correct value of n?
Typical answer: 100

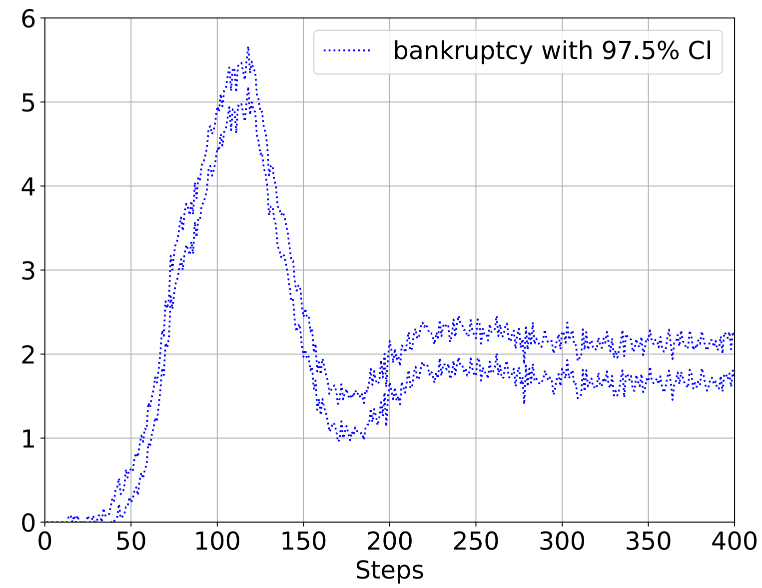
Our answer:
The question itself is ill-posed
Each property and time step might require a different n

Transient Analysis by autoIR: Large Macro ABM

Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

Evolution of bankruptcies



$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,400} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,400} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,400} \end{array}$$

$$\sum_{i=1}^{n} \frac{y_{i,2}}{n} = \overline{Y}_2 \approx E[Y_2]$$

What is the correct value of n?
Typical answer: 100

Our answer:
The question itself is ill-posed
Each property and time step might require a different n

Transient Analysis by autoIR: Large Macro ABM

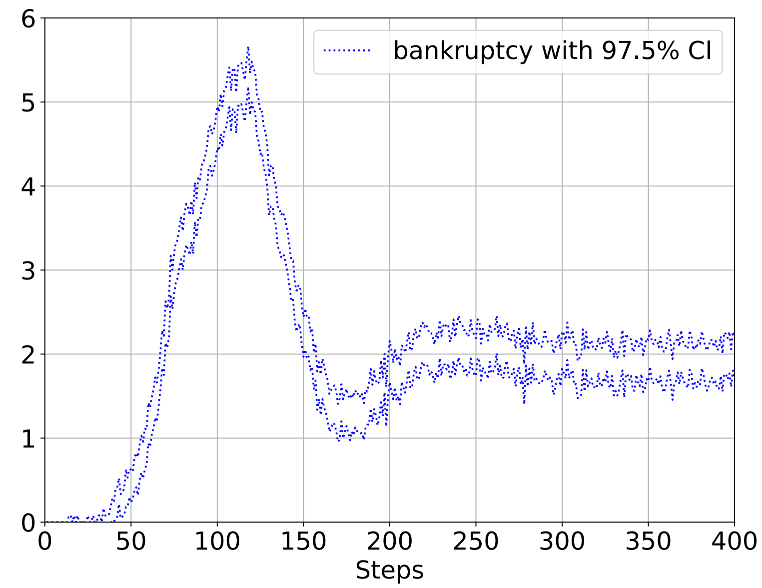
Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,400} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,400} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,400} \end{array}$$

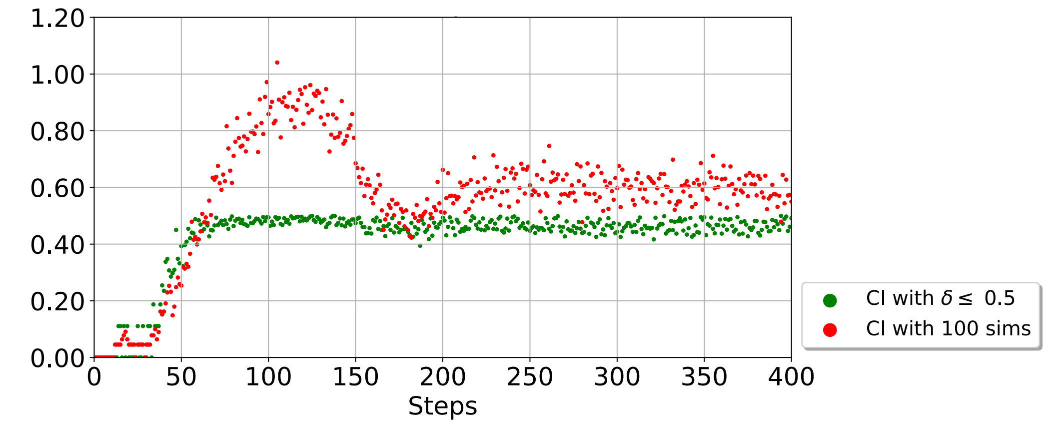
$$\sum_{i=1}^{n} \frac{y_{i,2}}{n} = \overline{Y}_2 \approx E[Y_2]$$

Evolution of bankruptcies



Confidence Intervals width

MultiVeStA VS 100 sims



What is the correct value of n?
Typical answer: 100

Our answer:
The question itself is ill-posed
Each property and time step might require a different n

Transient Analysis by autoIR: Large Macro ABM

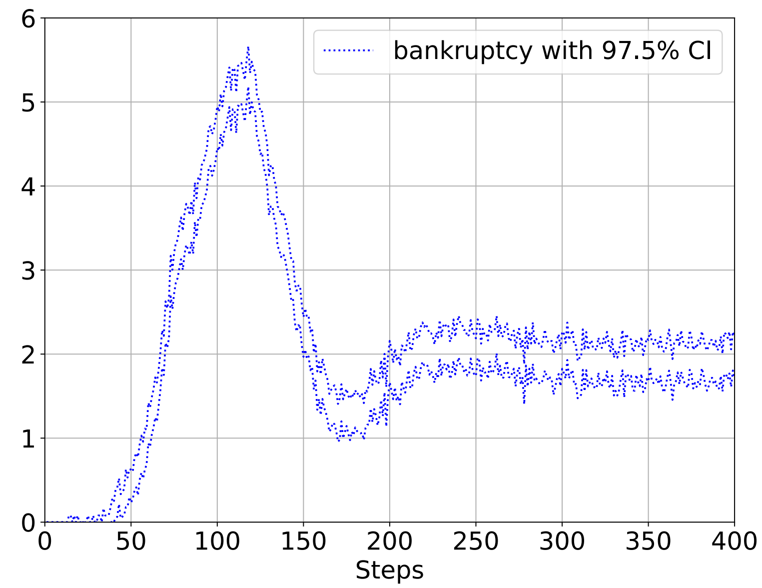
Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

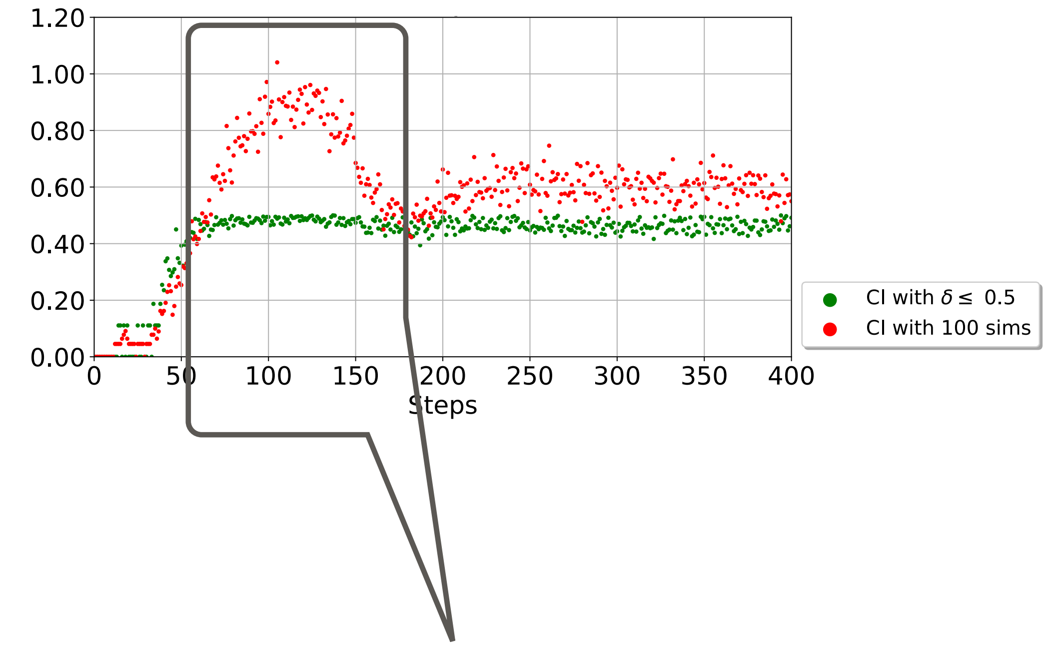
$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,400} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,400} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,400} \end{array}$$

$$\sum_{i=1}^{n} \frac{y_{i,2}}{n} = \overline{Y}_2 \approx E[Y_2]$$

Evolution of bankruptcies



Confidence Intervals width
MultiVeStA VS 100 sims



More simulations required here
We save a lot of time halting the
latest simulations to 150 steps

What is the correct value of n?
Typical answer: 100

Our answer:
The question itself is ill-posed
Each property and time step might require a different n

Transient Analysis by autoIR: Large Macro ABM

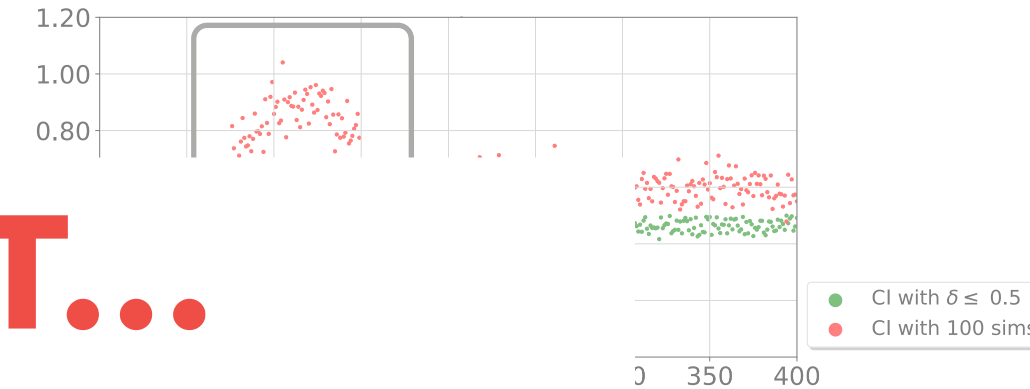
Large-scale macro financial ABM from Caiani et al, JEDC, 2016[Caiani et al2016]

- An economy with households, consumption/capital firms, commercial banks, government, central bank
- Thousands of agents
- Implemented in JMAB: Java framework for macro stock-flow consistent ABM models.
  - Side product: any model implemented in JMAB is now natively integrated with MultiVeStA

Evolution of bankruptcies

Confidence Intervals width

MultiVeStA VS 100 sims



OK, BUT...

IS THIS *IMPORTANT*?

$$\begin{matrix} y_{1,1} & y_{1,2} & \cdots & y_{1,400} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,400} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,400} \end{matrix}$$

$$\sum_{i=1}^{n} \frac{y_{i,2}}{n} = \overline{Y}_2 \approx E[Y_2]$$

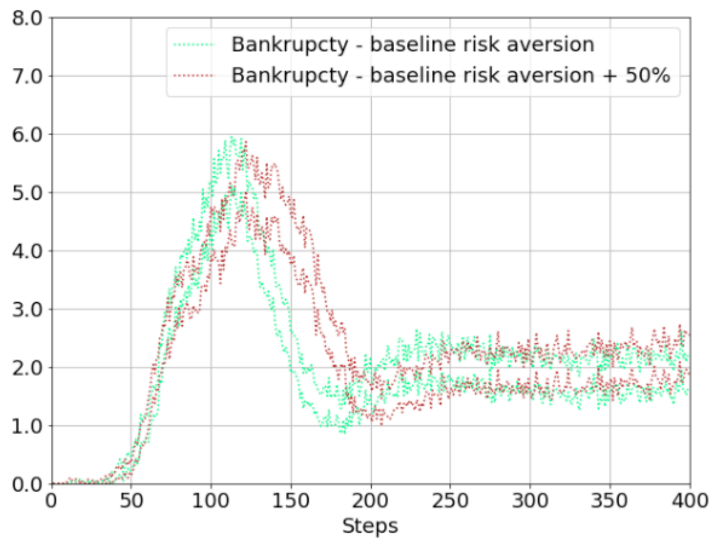
More simulations required here
We save a lot of time halting the
latest simulations to 150 steps

What is the correct value of n?
Typical answer: 100

Our answer:
The question itself is ill-posed
Each property and time step might require a different n

Statistically Meaningful Counterfactual Analysis

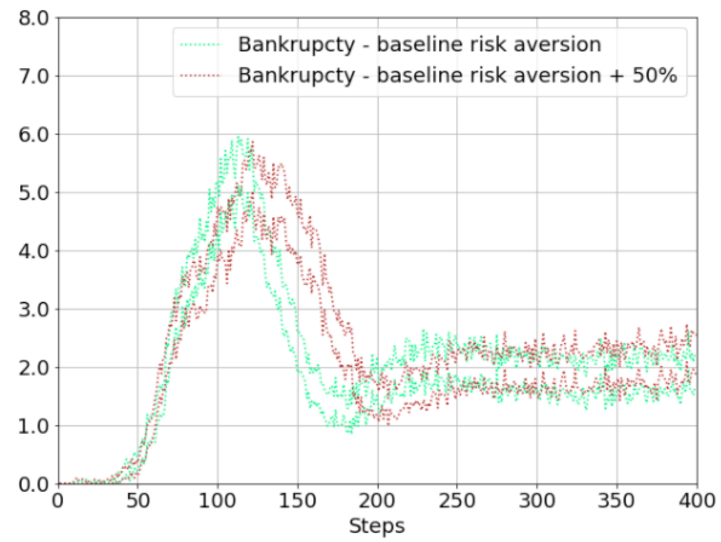
97.5% CI
100 Simulations



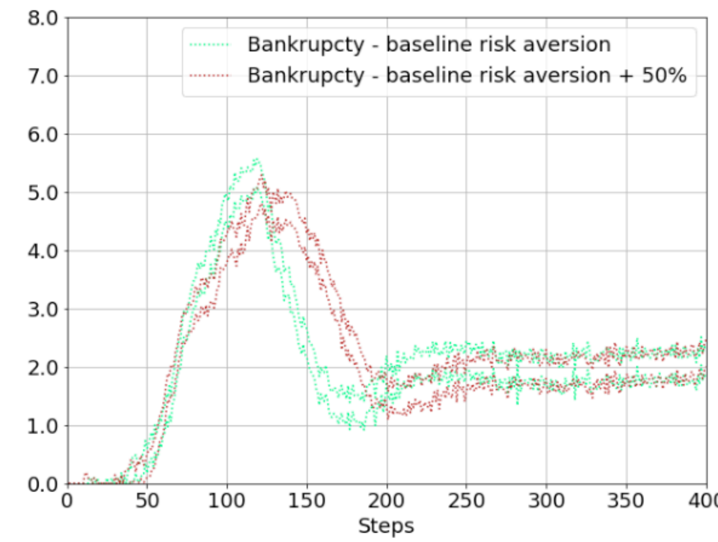
(a) CIs width for $\alpha = 0.025$ and $N = 100$ simulations

Statistically Meaningful Counterfactual Analysis

97.5% CI
100 Simulations



(a) CIs width for $\alpha = 0.025$ and $N = 100$ simulations

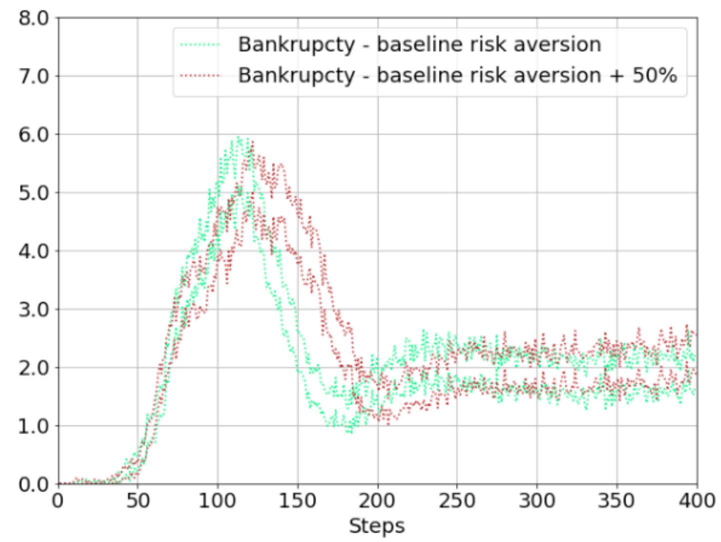


(b) CIs width for $\alpha = 0.025$ and $\delta = 0.5$

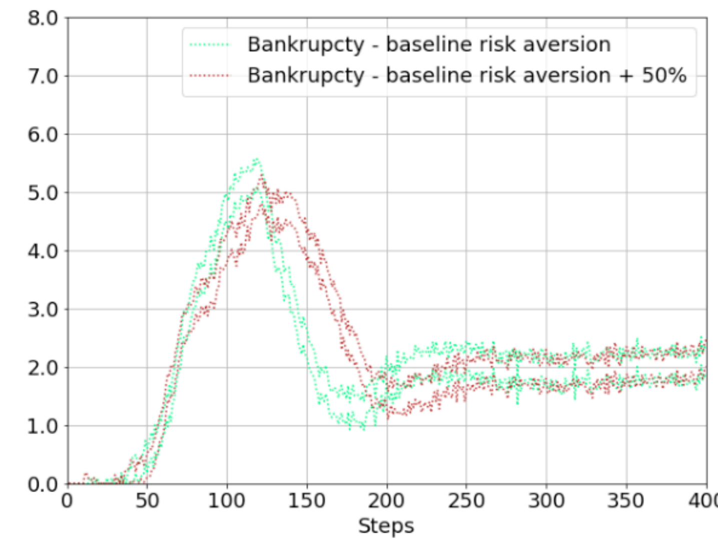
97.5% CI
MultiVeStA
$|\text{CI}| \leq \delta = 0.5$

Statistically Meaningful Counterfactual Analysis

97.5% CI
100 Simulations



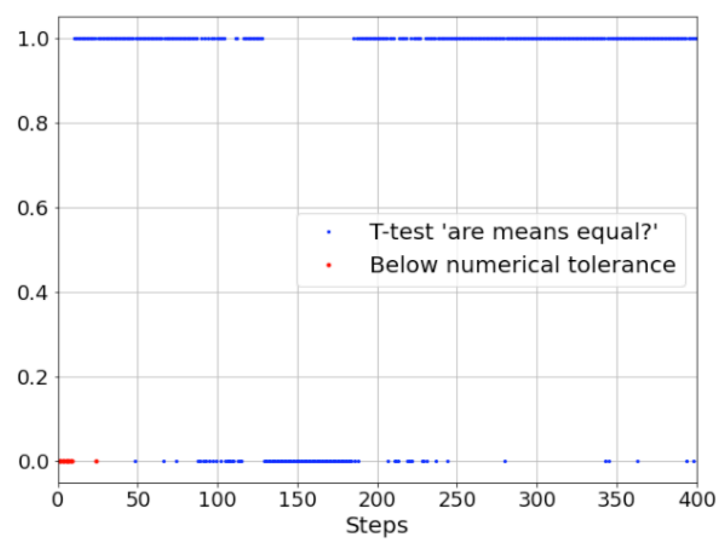
(a) CIs width for $\alpha = 0.025$ and $N = 100$ simulations



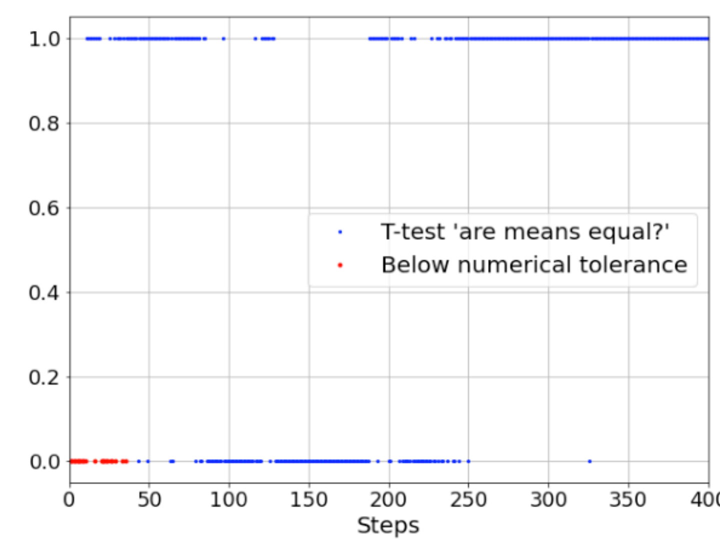
(b) CIs width for $\alpha = 0.025$ and $\delta = 0.5$

97.5% CI
MultiVeStA
$|\mathrm{CI}| \leq \delta = 0.5$

Welch's t-test with
significance α=2.5%
[Welch1947]



(c) T-test *are means in (a) point-wise equal* for significance $\alpha = 0.025$

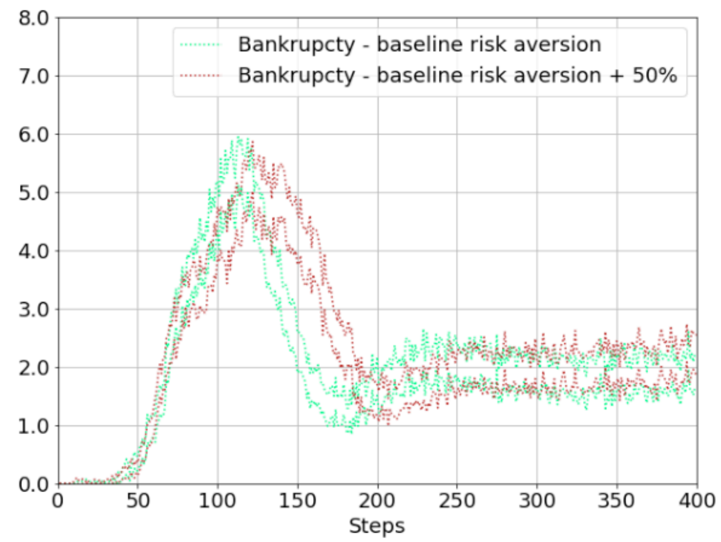


(d) T-test *are means in (b) point-wise equal* for significance $\alpha = 0.025$

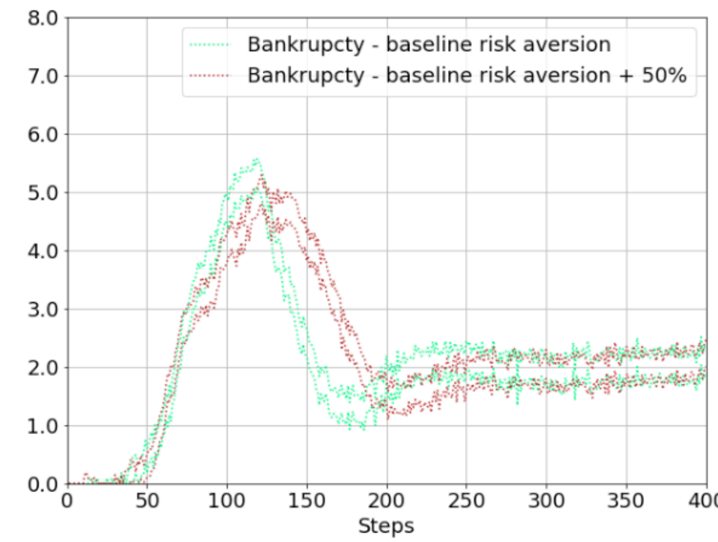
Welch's t-test with
significance α=2.5%
[Welch1947]

Statistically Meaningful Counterfactual Analysis

97.5% CI
100 Simulations



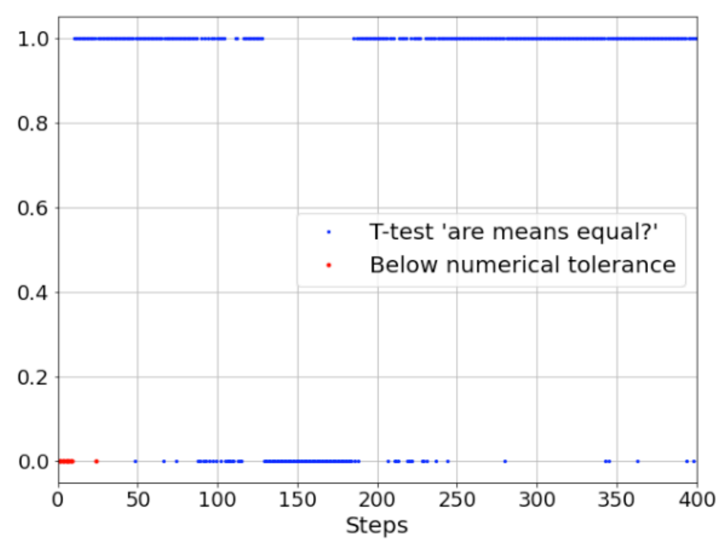
(a) CIs width for α = 0.025 and N = 100 simulations



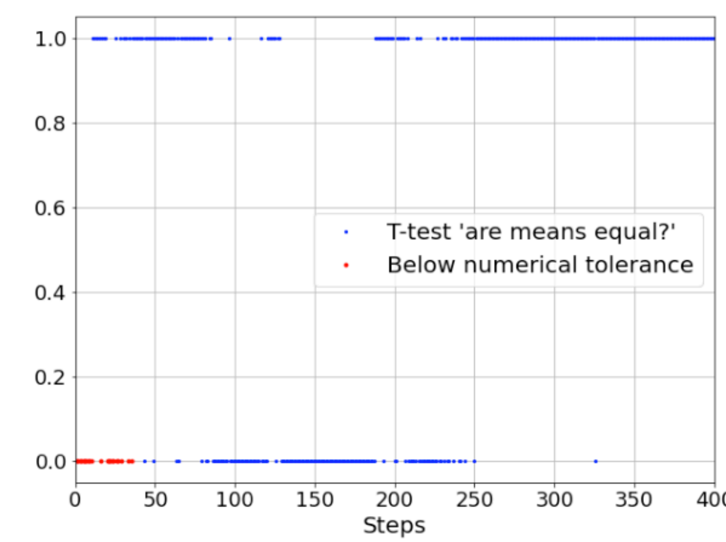
(b) CIs width for α = 0.025 and δ = 0.5

97.5% CI
MultiVeStA
$|\mathrm{CI}| \leq \delta = 0.5$

Welch's t-test with
significance α=2.5%
[Welch1947]



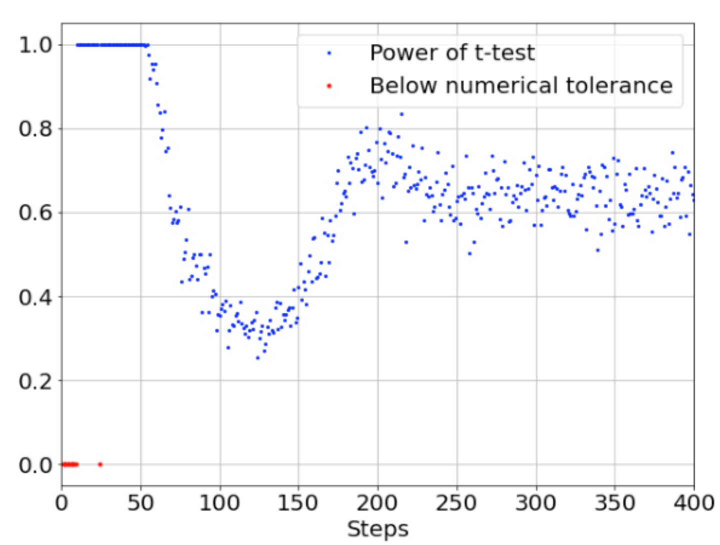
(c) T-test *are means in (a) point-wise equal* for significance α = 0.025



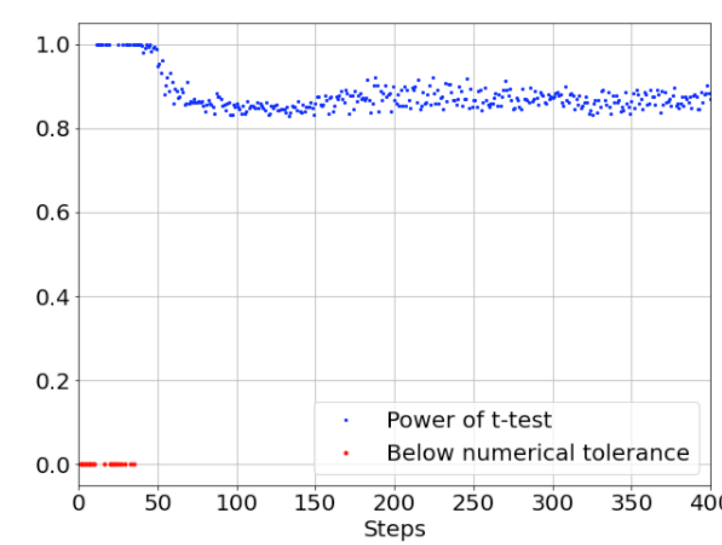
(d) T-test *are means in (b) point-wise equal* for significance α = 0.025

Welch's t-test with
significance α=2.5%
[Welch1947]

Power of the test
P(Test=0 | Real=0)
1 - P(Type II error)
[Chow2002]



(e) Power of t-test in (c) for difference ε = 0.5



(f) Power of t-test in (d) for difference ε = 0.5

Power of the test
P(Test=0 | Real=0)
1 - P(Type II error)
[Chow2002]

OUTLINE

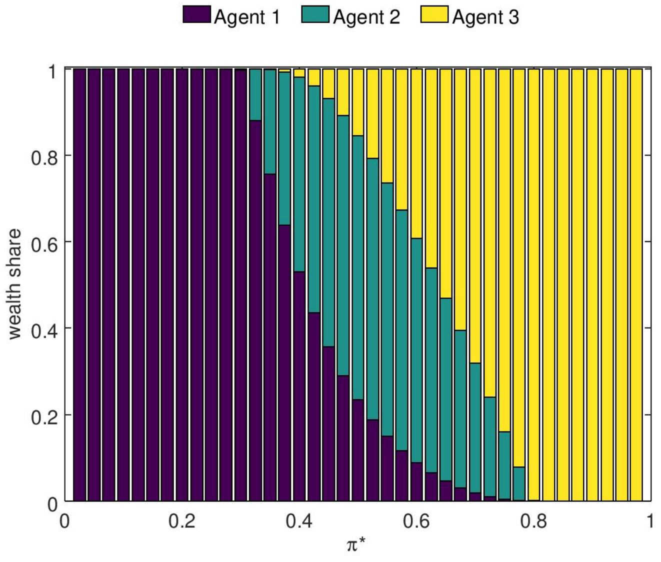
1. Motivation, vision, and proposal
    1. Automated analysis with statistical guarantees for ABMs
    2. The MultiVeStA Statistical Model Checker
2. Transient Analysis of a large-scale financial macro ABM
    1. Estimation of expected outcome and Confidence Interval
    2. Counterfactual analysis for different model configurations
3. **Steady-state analysis of a prediction market model**
    1. **Steady-state analysis by Replication and Deletion (RD)**
    2. Warmup estimation
    3. Steady-state analysis by Batch Means (BM)
    4. A methodology for ergodicity analysis based on RD and BM
4. Conclusions & Future works

Steady-State Analysis by autoRD: Market Selection

Simple repeated betting market from Kets et al, AAAI 2014

- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

Agents wealth at steady state



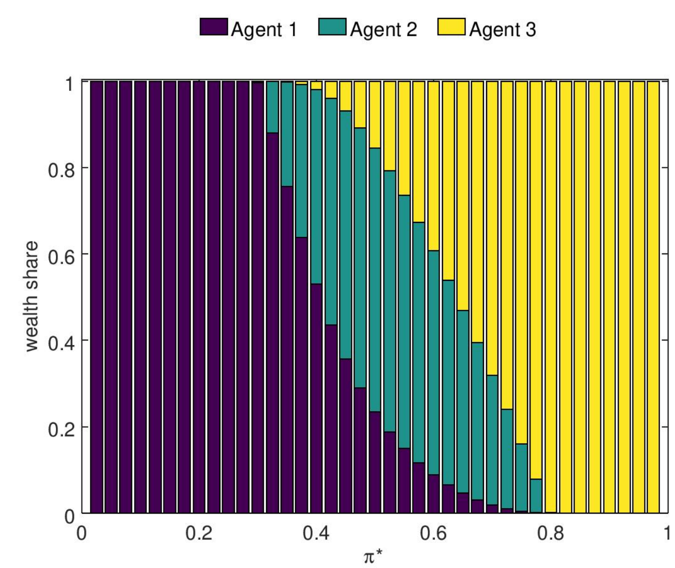
Steady-State Analysis by autoRD: Market Selection

Simple repeated betting market from Kets et al, AAAI 2014

- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,m} \end{array}$$

Agents wealth at steady state



Steady-State Analysis by autoRD: Market Selection

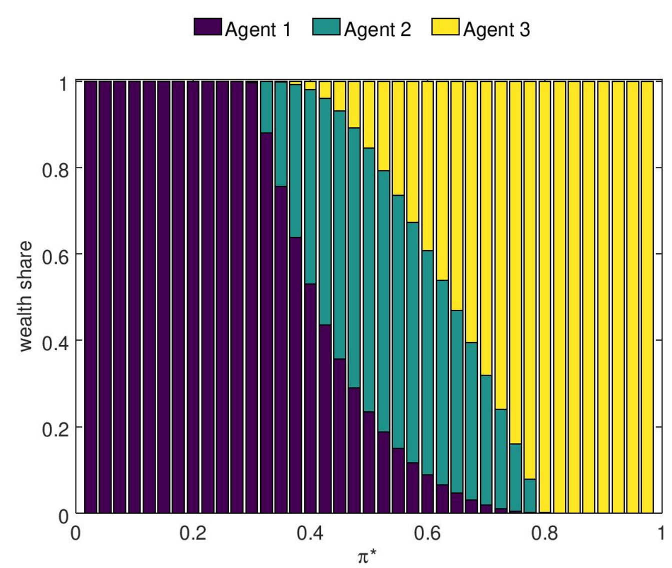
Simple repeated betting market from Kets et al, AAAI 2014

- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

Warmup
w steps

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,m} \end{array}$$

Agents wealth at steady state



Steady-State Analysis by autoRD: Market Selection

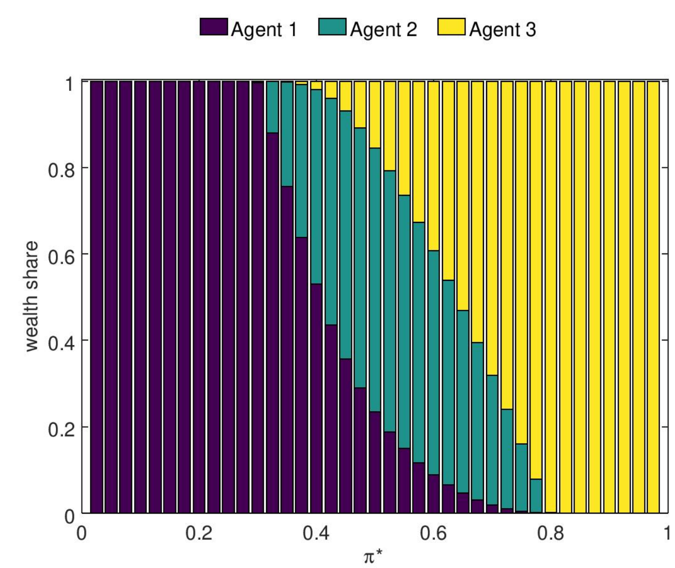
Simple repeated betting market from Kets et al, AAAI 2014

- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

Warmup
w steps

$$
\begin{array}{cccc}
y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\
y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\
\vdots & \vdots & \vdots & \vdots \\
y_{n,1} & y_{n,2} & \cdots & y_{n,m}
\end{array}
\qquad
\begin{array}{l}
\sum_{t=w+1}^{m} \frac{y_{1,t}}{m-w} = \overline{Y}_1(w) \\
\vdots \qquad = \overline{Y}_2(w) \\
\vdots \qquad \vdots \\
\vdots \qquad = \overline{Y}_n(w)
\end{array}
$$

Agents wealth at steady state



Steady-State Analysis by autoRD: Market Selection

Simple repeated betting market from Kets et al, AAAI 2014

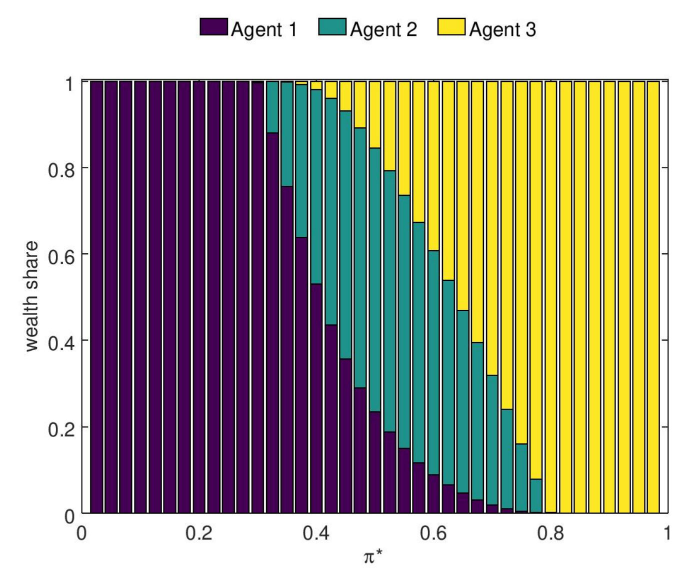
- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

Warmup w steps

$$
\begin{array}{ccccl}
y_{1,1} & y_{1,2} & \cdots & y_{1,m} & \sum_{t=w+1}^{m} \frac{y_{1,t}}{m-w} = \bar{Y}_1(w) \\
y_{2,1} & y_{2,2} & \cdots & y_{2,m} & \vdots \quad = \bar{Y}_2(w) \\
\vdots & \vdots & \vdots & \vdots & \vdots \qquad \vdots \\
y_{n,1} & y_{n,2} & \cdots & y_{n,m} & \vdots \quad = \bar{Y}_n(w)
\end{array}
$$

$$\sum_{i=1}^{n} \frac{\bar{Y}_i(w)}{n} = \bar{Y}(w) \approx E[Y] = \lim_{t \to \infty} E[Y_t]$$

Agents wealth at steady state



Steady-State Analysis by autoRD: Market Selection

Simple repeated betting market from Kets et al, AAAI 2014
- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

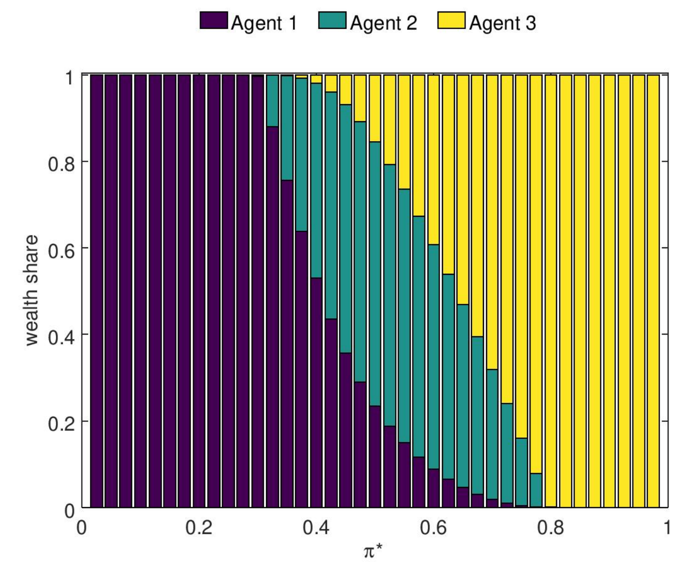
Warmup
w steps

$$\begin{array}{cccc|c|c}
y_{1,1} & y_{1,2} & \cdots & y_{1,m} & \sum_{t=w+1}^{m} \frac{y_{1,t}}{m-w} & = \overline{Y}_1(w) \\
y_{2,1} & y_{2,2} & \cdots & y_{2,m} & \vdots & = \overline{Y}_2(w) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
y_{n,1} & y_{n,2} & \cdots & y_{n,m} & \vdots & = \overline{Y}_n(w)
\end{array}$$

$$\sum_{i=1}^{n} \frac{\overline{Y}_i(w)}{n} = \overline{Y}(w) \approx E[Y] = \lim_{t \to \infty} E[Y_t]$$

Replication and Deletion (RD) [Law,Kelton2015]

Agents wealth at steady state



Steady-State Analysis by autoRD: Market Selection

Simple repeated betting market from Kets et al, AAAI 2014

- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

Warmup
w steps

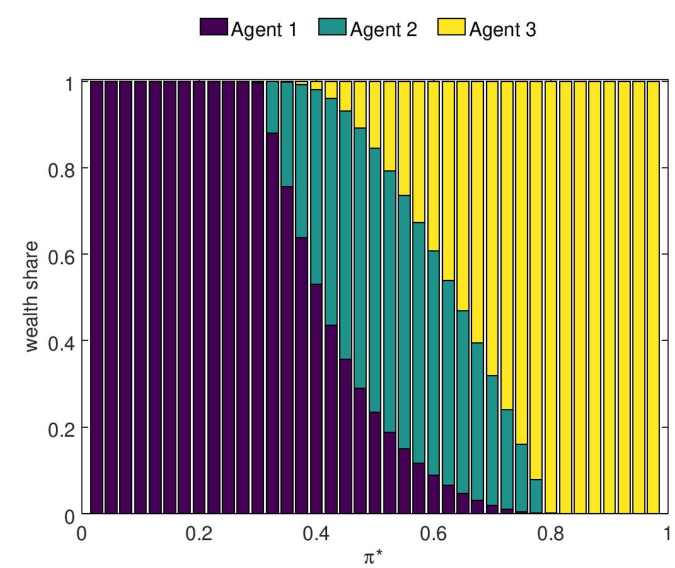
$$
\begin{array}{cccc}
y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\
y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\
\vdots & \vdots & \vdots & \vdots \\
y_{n,1} & y_{n,2} & \cdots & y_{n,m}
\end{array}
\qquad
\begin{array}{l}
\sum_{t=w+1}^{m} \frac{y_{1,t}}{m-w} = \overline{Y}_1(w) \\
\vdots \qquad\qquad = \overline{Y}_2(w) \\
\vdots \qquad\qquad \vdots \\
\vdots \qquad\qquad = \overline{Y}_n(w)
\end{array}
$$

$$\sum_{i=1}^{n} \frac{\overline{Y}_i(w)}{n} = \overline{Y}(w) \approx E[Y] = \lim_{t\to\infty} E[Y_t]$$

Replication and Deletion (RD) [Law,Kelton2015]

What are the correct values for w, m, n?

Agents wealth at steady state



Steady-State Analysis by autoRD: Market Selection

Simple repeated betting market from Kets et al, AAAI 2014

- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

Warmup
w steps

$$\begin{array}{ccccc}
y_{1,1} & y_{1,2} & \cdots & y_{1,m} & \sum_{t=w+1}^{m} \frac{y_{1,t}}{m-w} = \bar{Y}_1(w) \\
y_{2,1} & y_{2,2} & \cdots & y_{2,m} & \vdots \quad = \bar{Y}_2(w) \\
\vdots & \vdots & \vdots & \vdots & \vdots \quad \vdots \\
y_{n,1} & y_{n,2} & \cdots & y_{n,m} & \vdots \quad = \bar{Y}_n(w)
\end{array}$$

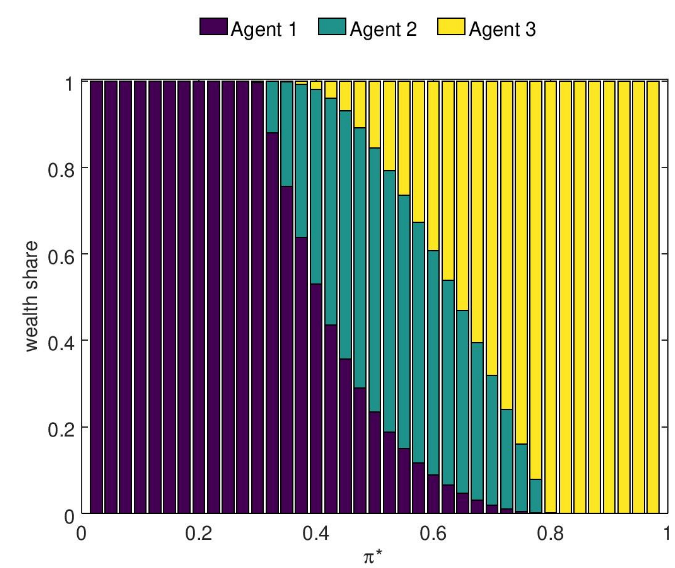
$$\sum_{i=1}^{n} \frac{\bar{Y}_i(w)}{n} = \bar{Y}(w) \approx E[Y] = \lim_{t \to \infty} E[Y_t]$$

Replication and Deletion (RD) [Law,Kelton2015]

What are the correct values for w, m, n?

THESE ARE DIFFICULT QUESTIONS
ARE THEY CRUCIAL?

Agents wealth at steady state



Steady-State Analysis by autoRD: Market Selection

Simple repeated betting market from Kets et al, AAAI 2014

- 1 event realises at every step with a fixed probability π^*
- 3 Fractional Kelly bettors with a belief on π^* and place bets accordingly

Warmup
w steps

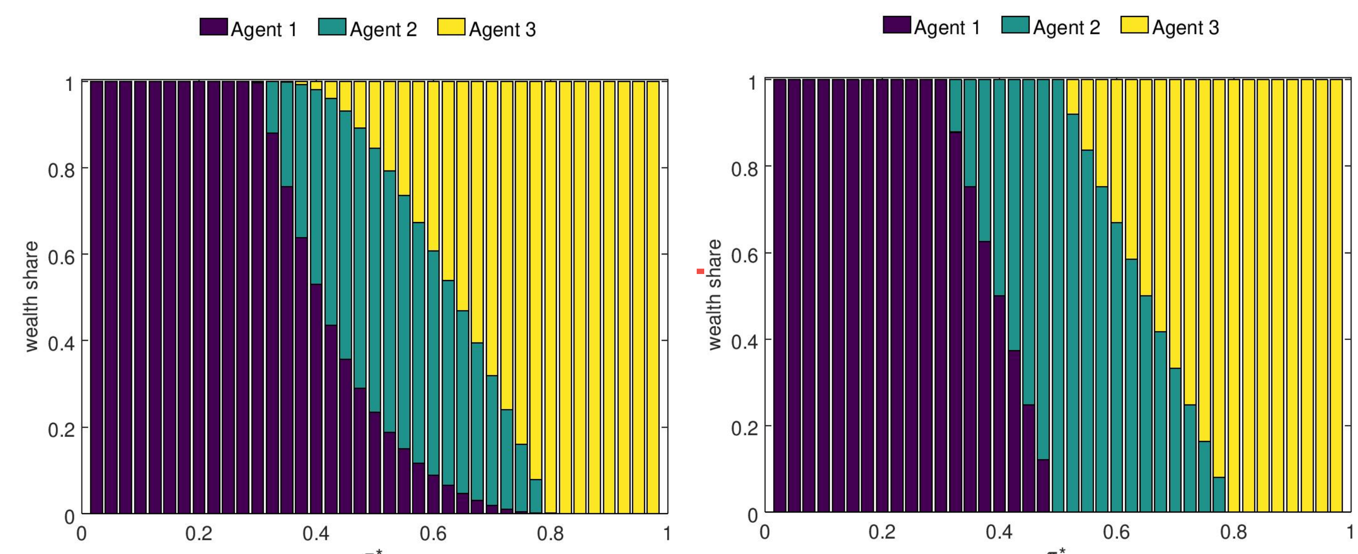
$$
\begin{array}{ccccc}
y_{1,1} & y_{1,2} & \cdots & y_{1,m} & \displaystyle\sum_{t=w+1}^{m} \frac{y_{1,t}}{m-w} = \overline{Y}_1(w) \\
y_{2,1} & y_{2,2} & \cdots & y_{2,m} & \vdots \quad = \overline{Y}_2(w) \\
\vdots & \vdots & \vdots & \vdots & \vdots \quad \vdots \\
y_{n,1} & y_{n,2} & \cdots & y_{n,m} & \vdots \quad = \overline{Y}_n(w)
\end{array}
$$

$$\sum_{i=1}^{n} \frac{\overline{Y}_i(w)}{n} = \overline{Y}(w) \approx E[Y] = \lim_{t \to \infty} E[Y_t]$$

Replication and Deletion (RD) [Law,Kelton2015]

What are the correct values for w, m, n?

Agents wealth at steady state



[Kets et al2014]

MultiVeStA

Same as analytical solution
from [Bottazzi,Giachini2019]

THESE ARE DIFFICULT QUESTIONS
ARE THEY CRUCIAL?

YES

OUTLINE

1. Motivation, vision, and proposal
    1. Automated analysis with statistical guarantees for ABMs
    2. The MultiVeStA Statistical Model Checker
2. Transient Analysis of a large-scale financial macro ABM
    1. Estimation of expected outcome and Confidence Interval
    2. Counterfactual analysis for different model configurations
3. **Steady-state analysis of a prediction market model**
    1. Steady-state analysis by Replication and Deletion (RD)
    2. **Warmup estimation**
    3. Steady-state analysis by Batch Means (BM)
    4. A methodology for ergodicity analysis based on RD and BM
4. Conclusions & Future works

Warmup Estimation by autoWarmup: our Automated Proposal

1. Do *1 long simulation* of length *m=B*b*
2. Divide it in *B batches* of *b* consecutive steps

$$\boxed{y_{1,1} \cdots y_{1,b}} \; \boxed{\cdots \; y_{1,2b}} \quad \cdots \quad \boxed{\cdots \; y_{1,Bb}}$$

Warmup Estimation by autoWarmup: our Automated Proposal

0. Set *m*= *B***b*,
   B=128 is the number of batches of simulation
   b=16 is the number of steps in each batch

16 16 16

$y_{1,1} \cdots y_{1,b}$ $\cdots\ y_{1,2b}$ $\cdots$ $\cdots\ y_{1,Bb}$

1. Do *1 long simulation* of length *m*=*B***b*
2. Divide it in *B batches* of *b* consecutive steps

Warmup Estimation by autoWarmup: our Automated Proposal

0. Set $m = B*b$,
   B=128 is the number of batches of simulation
   b=16 is the number of steps in each batch

16 16 16

$$\boxed{y_{1,1} \cdots y_{1,b}} \quad \boxed{\cdots\ y_{1,2b}} \quad \cdots \quad \boxed{\cdots\ y_{1,Bb}}$$

$$\sum_{t=1}^{b} \frac{y_{1,t}}{b} = \overline{B}_1 \qquad \overline{B}_2 \qquad \cdots \qquad \overline{B}_B$$

1. Do *1 long simulation* of length $m=B*b$
2. Divide it in *B batches* of *b* consecutive steps
3. Compute the mean $\overline{B}_i$ within each batch

Warmup Estimation by autoWarmup: our Automated Proposal

0. Set *m*= *B***b*,
   B=128 is the number of batches of simulation
   b=16 is the number of steps in each batch

16 16 16

$$y_{1,1} \cdots y_{1,b} \quad \cdots\ y_{1,2b} \quad \cdots \quad \cdots\ y_{1,Bb}$$

$$\sum_{t=1}^{b} \frac{y_{1,t}}{b} = \overline{B}_1 \qquad \overline{B}_2 \qquad \cdots \qquad \overline{B}_B$$

1. Do *1 long simulation* of length *m*=*B***b*
2. Divide it in *B batches* of *b* consecutive steps
3. Compute the mean $\overline{B}_i$ within each batch
4. Perform statistical tests to check if *m* is large enough
   Perform a normality test on the computed means
   Check for low lag-1 autocorrelation on the means

Warmup Estimation by autoWarmup: our Automated Proposal

0. Set *m*= *B***b*,
   B=128 is the number of batches of simulation
   b=16 is the number of steps in each batch

$$\underbrace{y_{1,1} \cdots y_{1,b}}^{16} \quad \underbrace{\cdots\ y_{1,2b}}^{16} \quad \cdots \quad \underbrace{\cdots\ y_{1,Bb}}^{16}$$

$$\sum_{t=1}^{b} \frac{y_{1,t}}{b} = \overline{B}_1 \qquad \overline{B}_2 \qquad \cdots \qquad \overline{B}_B$$

1. Do *1 long simulation* of length *m*=*B***b*
2. Divide it in *B batches* of *b* consecutive steps
3. Compute the mean $\overline{B}_i$ within each batch
4. Perform statistical tests to check if *m* is large enough
   Perform a normality test on the computed means
   Check for low lag-1 autocorrelation on the means

5.1 If all tests pass, we conclude that the warmup has ended

Warmup Estimation by autoWarmup: our Automated Proposal

0. Set *m*= *B***b*,
   B=128 is the number of batches of simulation
   b=16 is the number of steps in each batch

16 16 16

$$y_{1,1} \cdots y_{1,b} \quad \cdots\ y_{1,2b} \quad \cdots \quad \cdots\ y_{1,Bb}$$

$$\sum_{t=1}^{b} \frac{y_{1,t}}{b} = \overline{B}_1 \quad \overline{B}_2 \quad \cdots \quad \overline{B}_B$$

1. Do *1 long simulation* of length *m*=*B***b*
2. Divide it in *B batches* of *b* consecutive steps
3. Compute the mean $\overline{B}_i$ within each batch
4. Perform statistical tests to check if *m* is large enough
   Perform a normality test on the computed means
   Check for low lag-1 autocorrelation on the means

5.1 If all tests pass, we conclude that the warmup has ended

5.2 If one test fails
   Double *b* squeezing the batches in the first *B/2* ones
   Double *m* by performing *m* new simulation steps
   Go back to step 3

Warmup Estimation by autoWarmup: our Automated Proposal

0. Set *m*= *B***b*,
   - *B*=128 is the number of batches of simulation
   - *b*=16 is the number of steps in each batch

16 16 16

$$y_{1,1} \cdots y_{1,b} \quad \cdots\ y_{1,2b} \quad \cdots \quad \cdots\ y_{1,Bb}$$

$$\sum_{t=1}^{b} \frac{y_{1,t}}{b} = \overline{B}_1 \qquad \overline{B}_2 \qquad \cdots \qquad \overline{B}_B$$

32 32

$$y_{1,1} \cdots \qquad y_{1,b} \cdots \qquad \cdots\ y_{1,\frac{B}{2}b}$$

1. Do *1 long simulation* of length *m*=*B***b*
2. Divide it in *B batches* of *b* consecutive steps
3. Compute the mean $\overline{B}_i$ within each batch
4. Perform statistical tests to check if *m* is large enough
   - Perform a normality test on the computed means
   - Check for low lag-1 autocorrelation on the means

5.1 If all tests pass, we conclude that the warmup has ended

5.2 If one test fails
   - Double *b* squeezing the batches in the first *B/2* ones
   - Double *m* by performing *m* new simulation steps
   - Go back to step 3

Warmup Estimation by autoWarmup: our Automated Proposal

0. Set *m*= *B***b*,
   B=128 is the number of batches of simulation
   b=16 is the number of steps in each batch

16 16 16

$$y_{1,1} \cdots y_{1,b} \qquad \cdots\ y_{1,2b} \qquad \cdots \qquad \cdots\ y_{1,Bb}$$

$$\sum_{t=1}^{b} \frac{y_{1,t}}{b} = \overline{B}_1 \qquad \overline{B}_2 \qquad \cdots \qquad \overline{B}_B$$

32 32

$$y_{1,1} \cdots \qquad y_{1,b} \cdots \qquad \cdots\ y_{1,\frac{B}{2}b}$$

32 32

$$y_{1,\frac{B}{2}b} \cdots \qquad y_{1,(\frac{B}{2}+1)b} \cdots \qquad \cdots\ y_{1,Bb}$$

1. Do *1 long simulation* of length *m*=*B***b*
2. Divide it in *B batches* of *b* consecutive steps
3. Compute the mean $\overline{B}_i$ within each batch
4. Perform statistical tests to check if *m* is large enough
   Perform a normality test on the computed means
   Check for low lag-1 autocorrelation on the means

5.1 If all tests pass, we conclude that the warmup has ended

5.2 If one test fails
   Double *b* squeezing the batches in the first *B/2* ones
   Double *m* by performing *m* new simulation steps
   Go back to step 3

Warmup Estimation by autoWarmup: our Automated Proposal

0. Set *m*= *B**b*,
   - *B*=128 is the number of batches of simulation
   - *b*=16 is the number of steps in each batch

$$y_{1,1} \cdots y_{1,b} \quad \cdots\ y_{1,2b} \quad \cdots \quad \cdots\ y_{1,Bb}$$

$$\sum_{t=1}^{b} \frac{y_{1,t}}{b} = \overline{B}_1 \qquad \overline{B}_2 \qquad \cdots \qquad \overline{B}_B$$

1. Do *1 long simulation* of length *m*=*B**b*
2. Divide it in *B batches* of *b* consecutive steps
3. Compute the mean $\overline{B}_i$ within each batch
4. Perform statistical tests to check if *m* is large enough
   - Perform a normality test on the computed means
   - Check for low lag-1 autocorrelation on the means

5.1 If all tests pass, we conclude that the warmup has ended

5.2 If one test fails
   - Double *b* squeezing the batches in the first *B/2* ones
   - Double *m* by performing *m* new simulation steps
   - Go back to step 3

Based on the **Batch Means** (**BM**) method
- First proposal in [Conway1963], First automatic version in [Law,Carson1979]
- Approach for steady state analysis
  - Alternative to Replication and Deletion based on 1 long simulation

Steady State Analysis by autoBM: our BM-Based Proposal

0. Set $m = B*b$,
   - B=128 is the number of simulation batches
   - b=16 is the number of steps in each batch

$\boxed{y_{1,1} \cdots y_{1,b}}$ $\boxed{\cdots\ y_{1,2b}}$ $\cdots$ $\boxed{\cdots\ y_{1,Bb}}$

$$\sum_{t=1}^{b} \frac{y_{1,t}}{b} = \overline{B}_1 \qquad \overline{B}_2 \qquad \cdots \qquad \overline{B}_B$$

$$\sum_{j=l+1}^{n} \frac{\overline{B}_j}{n-l} = \overline{B}(l) \approx E[Y] = \lim_{t\to\infty} E[Y_t]$$

1. Do *1 long simulation* of length $m=B*b$
2. Divide it in *B batches* of *b* consecutive steps
3. Compute the mean $\overline{B}_i$ within each batch
4. Perform statistical tests to check if *m* is large enough
   - Perform a normality test on the computed means
   - Check for low lag-1 autocorrelation on the means

5.1 If all tests pass, we conclude that the warmup has ended
   - Compute the grand mean (mean of the means)
   - Compute the width *d* of the CI of grand mean
   - Adjust *d* according to the residual correlation in the means

5.2 If one test fails
   - Double *b* squeezing the batches in the first *B/2* ones
   - Double *m* by performing *m* new simulation steps
   - Go back to step 3

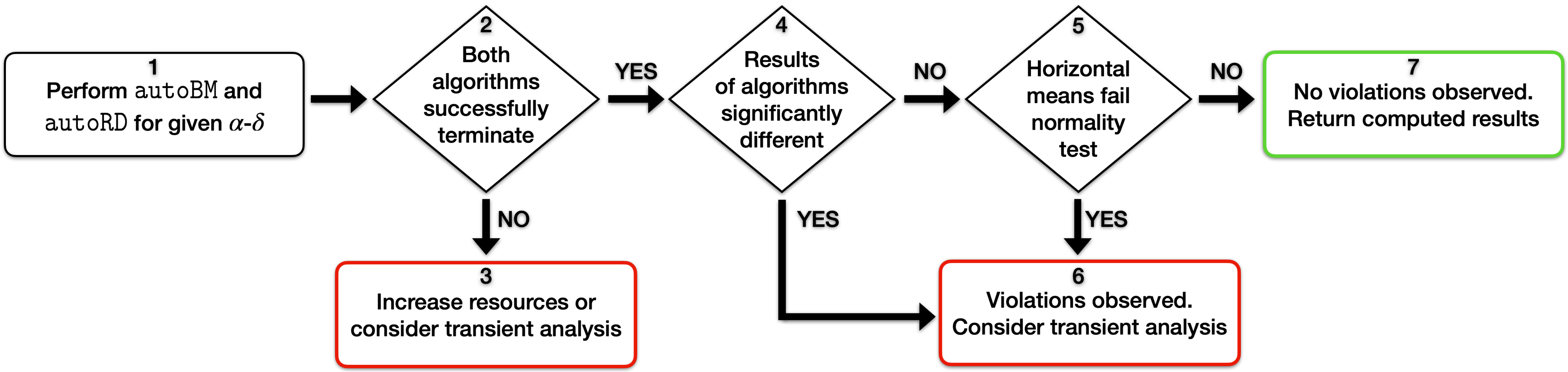
Based on the **Batch Means** (**BM**) method
- First proposal in [Conway1963], First automatic version in [Law,Carson1979]
- Approach for steady state analysis
  - Alternative to Replication and Deletion based on 1 long simulation

We also propose a simple novel version of BM for steady-state analysis
- Based on [Law,Carson1979] [Steiger et al 2005] [Tafazzoli et al 2011] [Gilmore et al 2017]

OUTLINE

1. Motivation, vision, and proposal
   1. Automated analysis with statistical guarantees for ABMs
   2. The MultiVeStA Statistical Model Checker
2. Transient Analysis of a large-scale financial macro ABM
   1. Estimation of expected outcome and Confidence Interval
   2. Counterfactual analysis for different model configurations
3. **Steady-state analysis of a prediction market model**
   1. Steady-state analysis by Replication and Deletion (RD)
   2. Warmup estimation
   3. Steady-state analysis by Batch Means (BM)
   4. **A methodology for ergodicity analysis based on RD and BM**
4. Conclusions & Future works

A Methodology for Ergodicity Diagnostics



OUTLINE

1. Motivation, vision, and proposal
   1. Automated analysis with statistical guarantees for ABMs
   2. The MultiVeStA Statistical Model Checker
2. Transient Analysis of a large-scale financial macro ABM
   1. Estimation of expected outcome and Confidence Interval
   2. Statistical comparison of different model configurations
3. Steady-state analysis of a prediction market model
   1. Steady-state analysis by Replication and Deletion (RD)
   2. Warmup estimation
   3. Steady-state analysis by Batch Means (BM)
   4. A methodology for ergodicity analysis based on RD and BM
4. **Conclusions & Future works**

CONCLUSIONS

- Fully automated framework for statistical analysis of ABMs
  - Transient analysis with statistical tests to compare model configurations
  - Warmup estimation, Steady-state analysis by RD and BM
  - Ergodicity diagnostics
- Tool-supported one-click analysis:
  - Less manual error-prone tasks => more reproducibility & reliability
  - Automatically parallelise simulations: 15 days => 15 hours
  - Support for simulators written in Java, Python, R, C++, JMAB
- Validated on two models from the literature:
  - Large-scale macro financial ABM, Small-scale prediction market model
  - We obtained new insights on the considered models
  - We avoid analysis issues from previous publications
- Our approach is rooted in results from:
  - Communities of Simulation, Computer Science, Operations Research
  - We aim at strengthening the cross-fertilisation with the ABM one

FUTURE WORK

- Add more techniques
  - Detection of multiple stationary points
  - Advanced SMC techniques to
    - Handle rare events, Reduce number of simulations required
  - More!? Model calibration, Sensitivity analysis, …
- Apply the approach to further models and domains
  - Any JMAB model is now natively supported
    - We wish to natively support further frameworks for ABM modelling
  - Would you like to **use MultiVeStA** to analyse your models?
    - Just contact us

REFERENCES

[Secchi,Seri2017] Secchi, D. and R. Seri (2017). Controlling for false negatives in agent-based models: a review of power analysis in organizational research. Computational and Mathematical Organization Theory 23(1), 94–121.

[Caiani et al2016] Caiani, A., A. Godin, E. Caverzasi, M. Gallegati, S. Kinsella, and J. E. Stiglitz (2016). Agent based-stock flow consistent macroeconomics: Towards a benchmark model. Journal of Economic Dynamics and Control 69, 375–408

[Welch1947] Welch, B. L. (1947). The generalization ofstudent's' problem when several different population variances are involved. Biometrika 34(1/2), 28–35.

[Chow2002] Chow, S.-C., J. Shao, and H. Wang (2002). A note on sample size calculation for mean comparisons based on noncentral t-statistics. Journal of biopharmaceutical statistics 12(4), 441–456.

[Law,Kelton2015] Law, A. M. and D. M. Kelton (2015). Simulation Modeling and Analysis (5th ed.). McGraw-Hill Higher Education, http://www.averill-law.com/simulation-book/

[Kets et al2014] Kets, W., D. M. Pennock, R. Sethi, and N. Shah (2014). Betting strategies, market selection, and the wisdom of crowds. In Twenty-Eighth AAAI Conference on Artificial Intelligence.

[Bottazzi,Giachini2019] Bottazzi, G. and D. Giachini (2019b). Far from the madding crowd: Collective wisdom in prediction markets. Quantitative Finance 19(9), 1461–1471

[Conway1963] Conway, R. W. (1963). Some tactical problems in digital simulation. Management Science 10(1), 47–61.

[Law,Carson1979] Law, A. M. and J. S. Carson (1979). A sequential procedure for determining the length of a steady-state simulation. Operations Research 27(5), 1011–1025.

[Steiger et al 2005] Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, and D. Goldsman (2005). Asap3: A batch means procedure for steady-state simulation analysis. ACM Transactions on Modeling and Computer Simulation (TOMACS) 15(1), 39–73.

[Tafazzoli et al 2011] Tafazzoli, A., J. R. Wilson, E. K. Lada, and N. M. Steiger (2011). Performance of skart: A skewness-and autoregression-adjusted batch means procedure for simulation analysis. INFORMS Journal on Computing 23(2), 297–314.

[Gilmore et al 2017] Gilmore, S., D. Reijsbergen, and A. Vandin (2017). Transient and steady-state statistical analysis for discrete event simulators. In Integrated Formal Methods - 13th International Conference, IFM 2017, Proceedings, pp. 145–160.

THANK YOU FOR YOUR ATTENTION!

QUESTIONS?

FEEDBACK?

LEM Working Paper available at:
www.lem.sssup.it/WPLem/2020-31.html

Tool and models available at:
https://bit.ly/MultiVeStATool