CSTEP: a HPC Platform for Scenario Reduction Research on Efficient Stochastic Modeling

--Representative Scenario Approach

Yvonne C. Chueh and Paul H. Johnson, Jr.

Abstract

The CSTEP (Cluster Sampling for Tail Estimation of Probability) system is a desktop-based application to assist actuaries in efficient stochastic modeling through the implementation of Representative Scenario approaches [1]. Its goal is to provide actuaries with an open-source HPC (High Performance Computation) research platform to select a sample set of representative (pivot) risk scenarios through user-specified or user-researched distance definitions. The sample is used to represent the entire scenario population (universe of up to one million scenarios each with up to 4,500 projected time periods) and provides a more accurate representation of the tail distributions of model outcome measures than simple random sampling. The traditional distance definitions used to select the representative scenarios were built into the program and are extended to allow modifications of distance definitions through user-inputted distance formulas with insight from the general pattern of model-specific asset and liability cash flows. The effectiveness of the representative (pivot) scenario approach, once achieved, can help actuaries and executives make and analyze decisions associated with Cash Flow Testing, Risk Based Capital C-3, Economic Capital, Principle-Based Reserving, Capital Budgeting, Solvency Analysis, Pricing, . . . , etc., for integrated large lines of business within the permitted run time. Moreover, the CSTEP formula-building platform allows for continual tests and tail enhancement of the Representative Scenarios with real-world assets and liabilities models tested by any stochastic scenario generators the modelers choose.

Introduction

Evaluating the adequacy of asset cash flows, risk-based capital, principle-based capital and reserves, and company solvency are critical tasks for many businesses, particularly insurance companies, as well as regulators and rating agencies. Valuation actuaries need to calculate the conditional tail expectation or probability distribution of relevant financial outcomes while refining their techniques whenever possible. It is especially difficult to truly capture the tail distributions without modeling a sufficiently large number of risk scenario simulations or adopting advanced model efficiency techniques. Through collaboration between the Central Washington University (CWU) Actuarial Science program and the Computer Science department, a HPC (High Performance Computation) platform named CSTEP (Cluster Sampling for Tail Estimation of Probability) was designed and engineered to deliver an open-source software for actuaries and researchers to apply pivot/representative scenario technique for modeling efficiency (through scenario reduction). The output of the CSTEP program is a set of pivot scenarios representing their distance-linked scenario clusters existing in the universe. The cluster size, which is the

number of scenarios grouped in that cluster, defines the probability measure of their pivot. The scenario universe is optionally nested in the pivoting process to help save run time or allocate computing time for advanced model complexity and dynamics. The pivot-scenario set calculated from CSTEP are the scenarios to be investigated and their model-projected financial outcomes to be analyzed. These simulated empirical model outcomes can then fit the optimal parametric probability distribution function to enhance probability tail accuracy [3],[6]. Such a probability distribution can help analyze conditional events and their financial or risk consequences given informed or assumed prior events/conditions.

CSTEP incorporates specialized, flexible distance formulas, to a quasi "cluster–distance" sampling method formally defined as the pivoting process and tested by [Chueh][1]. Earlier forms of representative scenario approaches such as cherry-picking and New York 7-associated were initiated and published earlier by [Christiansen][1] and [Longley-Cook][1] with the goal of attaining the best-fit tail probability estimation of the stochastic model outcome under a practical run time constraint.

Currently, actuaries implement pivoting processes using in-house spreadsheet programs or the SALMS[2] (Stochastic Asset Liability Model Sampling) software that either limit the size of the scenario universe and the speed of the pivoting process or do not allow for new distance definitions. CSTEP upgrades the current SALMS to a HPC platform that allows the user to edit the distance formula based on asset and liability attributes, or other considerations that may closely connect scenario paths to asset/liability model cash flow projections. The study of the relevance between the new distance metrics and the assets and liabilities attributes, independent of the scenario generator, should contribute to the knowledge of such scenario reduction strategies.

The Purpose of CSTEP Platform

One of the goals in developing CSTEP is to build a HPC platform for testing and developing in-house distance metrics for the challenging task of stochastically modeling blocks of contracts with volatile or sensitive financial risks triggered by policy benefit guarantees. A good distance metric is one that can force the hidden extreme stochastic model outcomes to appear provided any small change in the scenario path that gets picked by the pivoting process. Assuming that the stochastic model outcome measure is a continuous functional of high-dimensional scenarios and other inter-related drivers (such as the policy factors, contract liabilities, investment policies, etc), there exists one or more distance metrics that will stabilize the outcome-functional values with respect to the small scenario change (when the scenario is shifted to any direction in the high-dimensional distance metric space). This stabilization, if achieved, will grant uniform continuity and differentiability of the functional that ideally highly relate the model outcome to the high-dimensional scenarios. Due to the stochastic dynamics of the model outcome, there is so far no analytical way to derive the distance metric without simplifying

the assets and liabilities and other stochastic factors. Thus, empirical studies based on real asset and liability models that replicate not only the real assets and liabilities but their interactions and dynamics involving investment and consumer behavior are essential to the understanding of model efficiency and accuracy.

It is notable that CSTEP is not only designed for scenario reduction through pivot processes for stochastic model efficiency, but also for stochastic scenario research involving interest rates, equity returns, and stock indices. With its potential 8,388,608 scenario capacity with 4,500 projection periods each, and flexible rate formats (negative rates), CSTEP is a valuable contribution to research on equity risk scenarios, stock market indices, and stochastic nested scenarios for the main purpose of representative sampling or improving modeling efficiency for timely and realistic financial reporting.

CSTEP is a sampling platform that enhances efficient stochastic modeling through scenario reduction techniques and studies. It provides actuaries with a HPC research platform to select a sample set of pivot risk scenarios so that tedious programming work can be avoided and empirical studies, supported by real assets liabilities modeling practice, can be feasibly conducted. This is valuable since theoretical or replicated models have been the main approaches that may not capture the real complications and interactions among assets and liabilities, as well as real-world risk factors. Such empirical findings will be available to facilitate a more accurate, time-efficient, and reliable stochastic modeling practice. Practicing actuaries and researchers can now analyze the probability distributions of model outcomes for large blocks of integrated business with a drastically reduced run time (typically up to 1/10 of the current run time) through a substantially more effective representative scenario approach of their choice through the editable distance formula.

The option of changing the distance definition is intended to take into account the distance between scenarios measured in the light of perceived associations between the distance metric space and the model outcome variables. For example, the user can incorporate asset runoff in asset liability models by defining economic value parameters in the distance measure. This feature can be used to seek other meaningful distance measures to improve product-driven or block-driven modeling efficiency in tail distributions, which has historically been challenging for the products with scenario-sensitive guarantees.

In the end, CSTEP will provide several scenario sampling results for practicing actuaries and researchers to compare and analyze the probability distributions of stochastic assets/liabilities model outcome based on these scenario samples within a short amount of run time. Through the editable distance definitions, the stochastic model outcome distribution from small sample runs can be compared and analyzed to obtain the best estimates of tail probabilities and conditional tail expectations (CTE) supported by the empirical evidences.

Research Related to CSTEP

CSTEP was engineered according to the techniques from a published paper on modeling efficiency [1] and its associated published open-source software SALMS [2]. It is designed to increase the functionality and capacity of SALMS and provide the user with the opportunity to research their own customized input distance formula. It has been difficult for actuaries to program these sampling algorithms without training or experience in standard computer software engineering. CSTEP replaces spreadsheet based programs that require days of computation to sample from a moderately sized universe (such as 10,000) when the advanced pivoting process is performed. With CSTEP's user friendly interface and automatic file management, valuable research time can be dedicated to studying model techniques instead of tool building and/or manually processing data using a transparent spreadsheet format.

In addition to the need for such a HPC platform for pivot selection [1], to achieve higher model efficiency, various efficient modeling issues call for a new tool to help further investigate these areas. CSTEP will serve as a HPC platform that conveniently assists researchers and practitioners to analyze:

1. The size of the original scenario universe that can adequately estimate probabilities and moments of the tail distribution of stochastic model outcomes;
2. The minimum sample size (reduced sample run) required, justified with research, given the size of original universe and the type of policies (business blocks) and risk characteristics;
3. A comparison of various scenario reduction techniques, and a refinement of their technique efficiency and effective implementations;
4. Composite/nested risk scenarios of path dependent scenario simulations for more truthful financial reporting and well-grounded enterprise risk management;
5. Equity risk with a tail distribution that is difficult to effectively model compared to traditionally less volatile or severe interest rate risk;
6. In conjunction with the AMOOF (Actuarial Model Outcome Optimal Fit) technique [3], the model efficiency technique incorporating optimal parametric probability model fitting;
7. In conjunction with BMLE (Bias-Corrected Maximum Likelihood Estimation) [4], the model efficiency technique using bias-corrected MLE probability models;
8. In conjunction with BMLE [4],[5], the model efficiency technique using bias-corrected MLE to mixed probability models.

The stochastic model outcome distribution resulting from the reduced sample run can be analyzed to obtain estimates of probabilities and conditional expectations from the tail of the distribution. Such a scenario reduced empirical model distribution can test the effectiveness of pivot scenario approaches through the open-entry distance formulas—a new functionality for academic and company research and implementation. When the distance metric is more closely connected to the mathematical relation between the scenario and its stochastic model projection, the extreme scenarios are more likely to be detected, and the pattern will be more likely to be immediately identified and evaluated. The effectiveness of the pivot scenario approach, to be enhanced by an editable and researchable distance metric, is that it can help actuaries and executives analyze or make decisions on Cash Flow Testing, Risk

Based Capital, C-3, Economic Capital, Principle-based Reserving, Capital Budgeting, Solvency, Pricing, . . . , etc., for integrated one-time-prohibitive models for large lines of business with a feasible run time supported by a company's computing technology. Moreover, CSTEP has an open distance formula function to allow continual tests and improvement of these distance metrics. In the future, when companies have the capability to run tens, hundreds, thousands, or even millions of nested risk scenarios, CSTEP will still be able to search out extreme pivot scenarios from the nested scenario universe of at least 8,388,608 risk scenarios.

Open Source Computational Platform

For readers who are interested in using CSTEP, please visit www.cwu.edu/~chueh for the installation link. To assist CSTEP users, the CSTEP installation package contains a Documentation page (Figure 1) with a linked document describing the pivot sampling methods, article references, the user manual, the development team's website, and other relevant websites. Appendix A contains a list of definitions for the terms used in the CSTEP. More program screenshots can be found in Appendix B containing Figures 2-7.

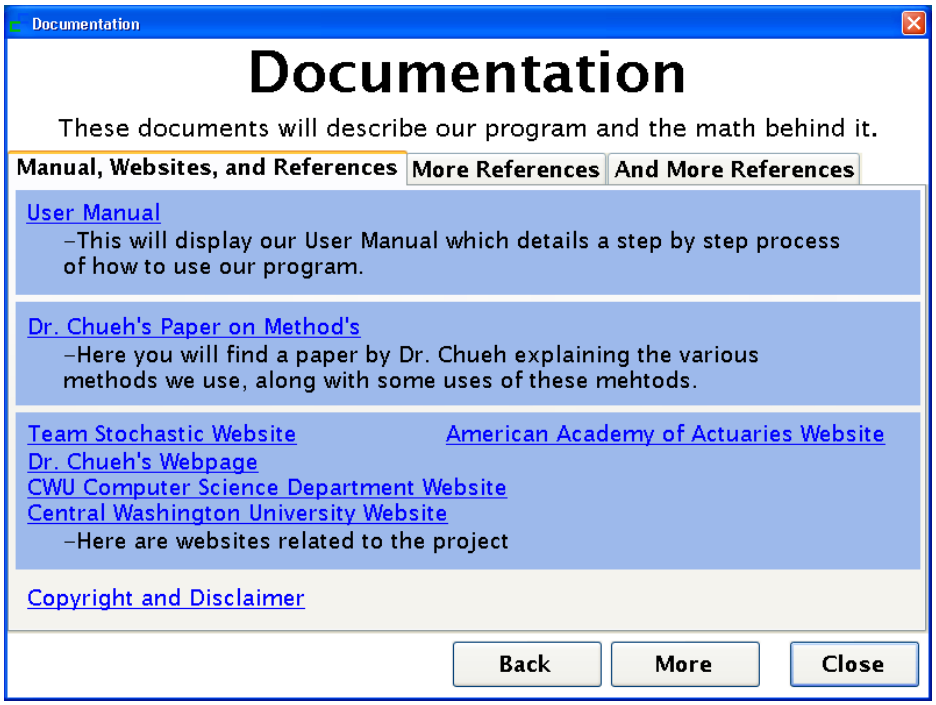


Figure 1 CSTEP Documentation Page

Acknowledgements

The authors would like to acknowledge the CWU Computer Science department and its chair Dr. James Schwing for their software engineering expertise and long-term generous support to the authors' research program. These former senior students have directly contributed to three published software projects (CSTEP, AMOOF, SALMS) in Actuarial Science studies: Alan Chandler, Eric Brown, Nathan Wood, Temourshah Ahmady, Ryan Green, Daniel Hanson, Matthew Miller, Edward Badgley, Jon Swenson, Paul Reed, Doug Love, Keith Lambert, and Mitsuharu Yasuda. Dr. Chueh would also like to acknowledge the University of Connecticut for the privileged educational and research opportunities she received, her former professors and advisors Dr. Charles Vinsonhaler, ASA and Dr. Jeyaraj Vadiveloo, FSA, MAAA, CFA, for their professional dedication and inspiration; Dr. Jeyaraj Vadiveloo and Mr. Alastair Longley-Cook's vision and guidance had motivated the later efficient sampling tools. Special thanks to Edward F. Cowman, FSA, MAAA, for his extraordinary real model data contribution and the valuable testing assistance.

References [in the order of relevance]

[1] Chueh, Yvonne C., "*Efficient Stochastic Modeling for Large and Consolidated Insurance Business: Interest Sampling Algorithms*" ***North American Actuarial Journal*** **vol.3,** 88-103**,** the **Society of Actuaries,** 2002.

[2] Chueh, Yvonne C., "*Insurance Modeling and Stochastic Cash Flow Scenario Testing: Effective Sampling Algorithms to Reduce Number of Run* and *SALMS (Stochastic Asset Liability Modeling Sampling)*", ***Contingencies***, on-line link to full paper, the major publication of the **American Academy of Actuaries**, 2003.

[3] Chueh, Yvonne C., "*Efficient Stochastic Modeling: Scenario Sampling Enhanced by Parametric Model Outcome Fitting*", ***Contingencies***, the major publication of the **American Academy of Actuaries**, 2005.

[4] Johnson, Paul H. Jr; Qi, Yongxue; Chueh, Yvonne, "Bias-Corrected Maximum Likelihood Estimation in Actuarial Science", working paper, 2011.

[5] Chueh, Yvonne C., "*Efficient Stochastic Modeling: From Scenario Sampling to Parametric Model Fitting Utilizing ASEM as an Example*", CD Rom Audio File, Proceedings of ***Symposium of Stochastic Modeling***, 1-40, an International Professional Development Symposium Co-sponsored by **Canadian Institute of Actuaries**, **Actuarial Foundation,** and **Society of Actuaries,** Toronto, Canada, 2003.

[6] Chueh, Yvonne C. and Curtis, Dan, "*Optimal PDF (Probability Density Function) Models for Stochastic Model Outcomes: Parametric Model Fitting on Tail Distributions*", 1-17, New Ideas in Symbolic Computation: Proceedings of the 6th ***International Mathematica Symposium***, 2004.

Appendix A: Glossary of Terms for CSTEP

Scenario:

A vector in h dimensional space which defines each value h on:

- Interest rates
- Stock return
- Risk parameters.

Universe:

A user created and imported .csv file which lists the risk scenarios (such as interest rates, equity returns, . . ., etc) and their ID numbers.

Population:

A subset of the imported Universe from which the CSTEP program runs calculations and draws samples given the sampling method and distance metric selected.

ID:

A unique identifier for each scenario in the scenario universe file, which if not provided by the user, will be generated by the program using the natural numbers.

Sampling Process:

It is a statistical process that finds the most extreme and representative scenarios for a population of risk scenarios.

Horizon:

The time horizon (time-period) of each scenario, the number of time periods to be modeled and projected by the stochastic modeler(s).

Samples:

The final outputs of CSTEP from the sampling process provided the imported universe scenarios and user specified methods.

Euclidean Distance Method:

$$\sqrt{\sum_{t=1}^{h}(i_t - i_t^P)^2 \bullet V^t}$$

This method defines the distance between each pair of risk scenarios using a single weight value, defined by V, to allow for a decreasing weight over the time period. (See Paper [1] for more Details).

Significance Method:

$$\sqrt{\sum_{t=1}^{h}\left(\prod_{k=1}^{t}\frac{C_k}{1+i_k}\right)^2}$$

This method is the simplest. It calculates a significance measure for each scenario in the universe, sorts the scenarios based on their significance values, and then uniformly chooses a sample set of scenarios from the sorted scenario population list. Thus, the scenarios corresponding to the evenly marked percentiles will be chosen. Also this method uses a constant, 1, for C_k (See Paper [1] for more Details).

Present Value Distance Method:

$$\sqrt{\sum_{t=1}^{h}\left(\prod_{k=1}^{t}\frac{C_k}{1+i_k}-\prod_{k=1}^{t}\frac{C_k}{1+i_k^P}\right)^2}$$

This method defines distance, like the Euclidean distance method, and then selects the next pivot scenarios based on the above distance measure. This method only uses 1 for C_k as a special case of the Economic Value Distance Method.

Economic Value Distance Method:

$$\sqrt{\sum_{t=1}^{h}\left(\prod_{k=1}^{t}\frac{C_k}{1+i_k}-\prod_{k=1}^{t}\frac{C_k}{1+i_k^P}\right)^2}$$

This method defines a more general distance between each pair of risk scenarios. The constant C_k allows users to incorporate the asset runoff speed and pattern of the business block into the distance measure. By studying the asset runoff to assign proper values to C_k, it is more likely to closely associate the extreme scenarios with the extreme model outcomes.

Economic Significance Method:

$$\sqrt{\sum_{t=1}^{h}\left(\prod_{k=1}^{t}\frac{C_k}{1+i_k}\right)^2}$$

This method is the simplest, and calculates a significance measure for each scenario in the universe, sorts them by the significance values, and then uniformly chooses from the sorted scenario population list. Like the Economic Value Distance formula, this economic significance measure allows users to incorporate the asset runoff speed and pattern of the business block into the significance of each scenario. By studying the asset runoff to assign proper values to C_k, it is more likely to closely associate the extreme scenarios with the extreme significance measures.

Pivot:

Pivot scenarios form the sample and are produced by the Present Value Distance Method, the Economic Value Distance Method, or the Euclidean Distance Method, that rely on the pivot selection process. Pivot scenarios are also named Representative Scenarios that are meant to represent the entire scenario population without under or over sampling from the two extreme tails.

Scenario Probability:

The probability of a scenario occurring. CSTEP calculates the probability of each pivot scenario and summarizes the scenario on the output page.

Pivoting Process:

To choose the pivots, start with a seed pivot scenario and then select the second pivot that maximizes the distance to the seed pivot. The third pivot is the one that is the farthest from the two pivots selected. Continue this process until the entire sample of pivots is selected.

Appendix B: Screenshots of CSTEP

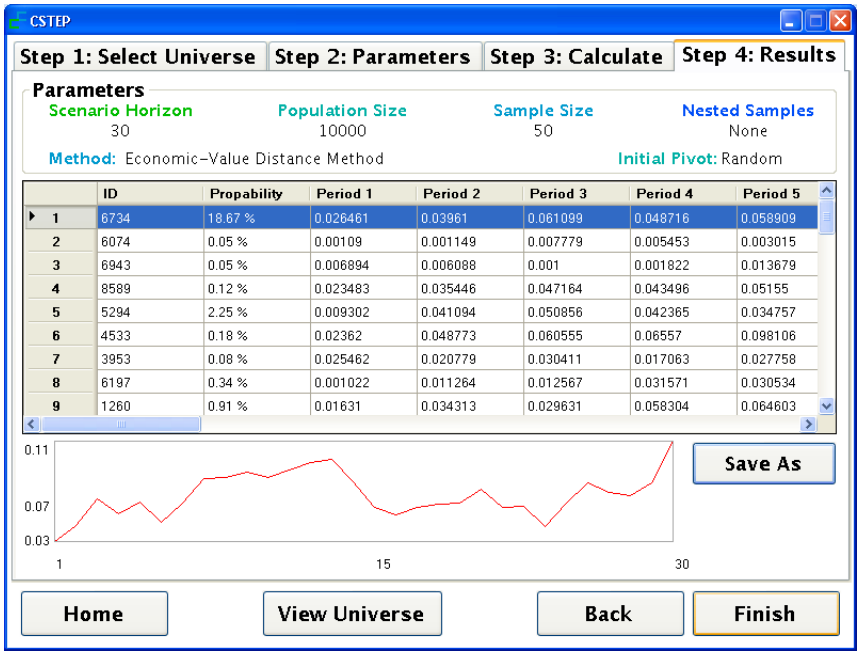


Figure 2 CSTEP Main Page Results Tab

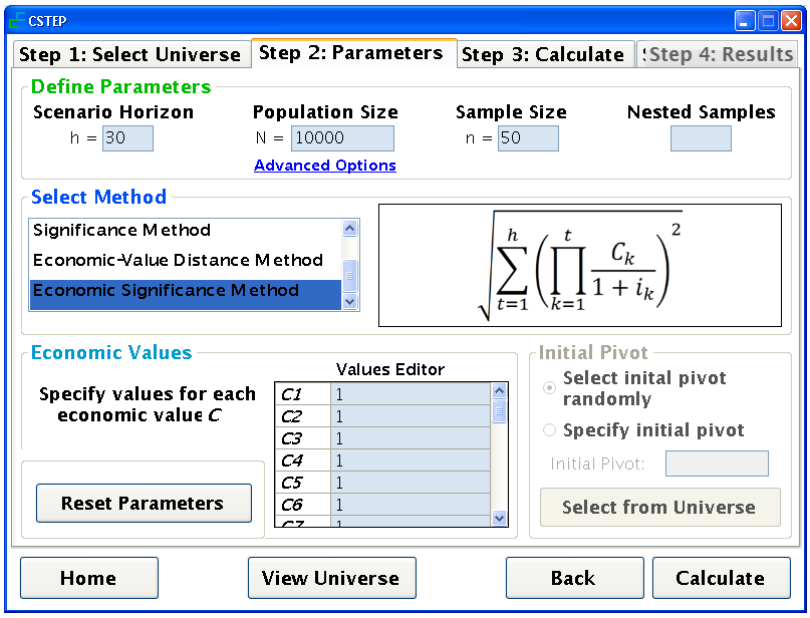


Figure 3 CSTEP Main Page Parameters Tab

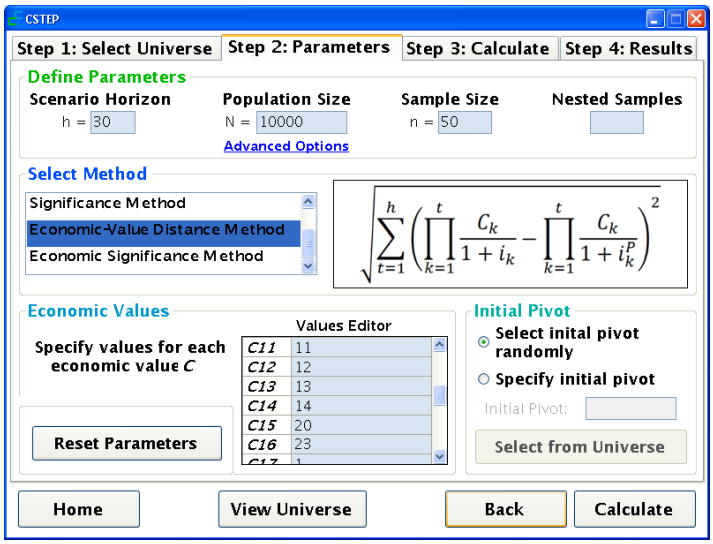


Figure 4 CSTEP Main Page Parameters Tab

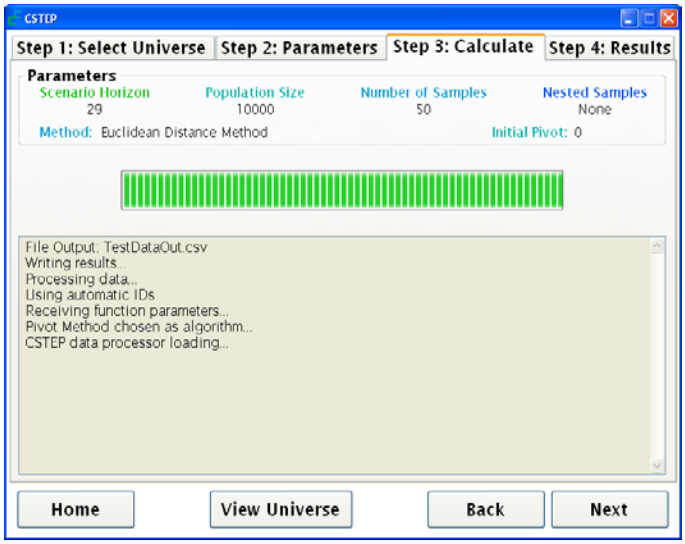


Figure 5 CSTEP Main Page Calculate Tab

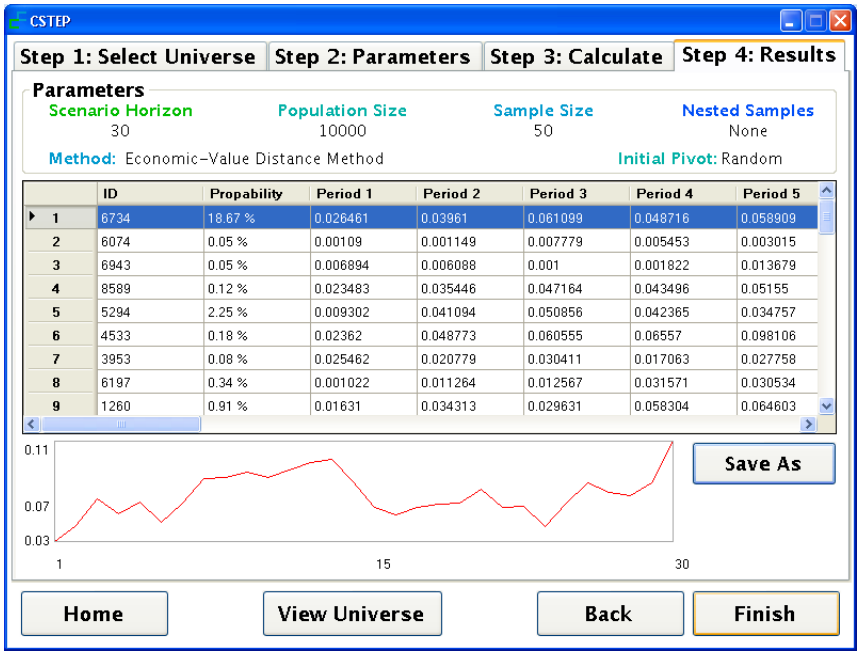
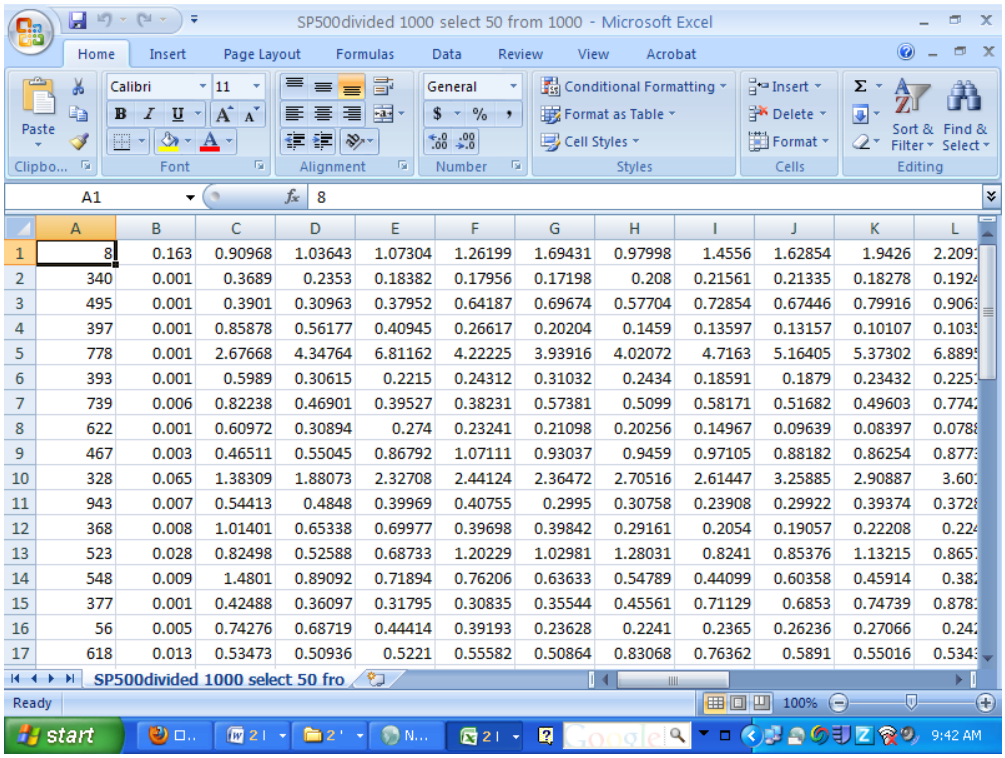


Figure 6 CSTEP Main Page Results Tab



	A	B	C	D	E	F	G	H	I	J	K	L
1	8	0.163	0.90968	1.03643	1.07304	1.26199	1.69431	0.97998	1.4556	1.62854	1.9426	2.209
2	340	0.001	0.3689	0.2353	0.18382	0.17956	0.17198	0.208	0.21561	0.21335	0.18278	0.192
3	495	0.001	0.3901	0.30963	0.37952	0.64187	0.69674	0.57704	0.72854	0.67446	0.79916	0.906
4	397	0.001	0.85878	0.56177	0.40945	0.26617	0.20204	0.1459	0.13597	0.13157	0.10107	0.103
5	778	0.001	2.67668	4.34764	6.81162	4.22225	3.93916	4.02072	4.7163	5.16405	5.37302	6.889
6	393	0.001	0.5989	0.30615	0.2215	0.24312	0.31032	0.2434	0.18591	0.1879	0.23432	0.225
7	739	0.006	0.82238	0.46901	0.39527	0.38231	0.57381	0.5099	0.58171	0.51682	0.49603	0.774
8	622	0.001	0.60972	0.30894	0.274	0.23241	0.21098	0.20256	0.14967	0.09639	0.08397	0.078
9	467	0.003	0.46511	0.55045	0.86792	1.07111	0.93037	0.9459	0.97105	0.88182	0.86254	0.877
10	328	0.065	1.38309	1.88073	2.32708	2.44124	2.36472	2.70516	2.61447	3.25885	2.90887	3.60
11	943	0.007	0.54413	0.4848	0.39969	0.40755	0.2995	0.30758	0.23908	0.29922	0.39374	0.372
12	368	0.008	1.01401	0.65338	0.69977	0.39698	0.39842	0.29161	0.2054	0.19057	0.22208	0.22
13	523	0.028	0.82498	0.52588	0.68733	1.20229	1.02981	1.28031	0.8241	0.85376	1.13215	0.865
14	548	0.009	1.4801	0.89092	0.71894	0.76206	0.63633	0.54789	0.44099	0.60358	0.45914	0.38
15	377	0.001	0.42488	0.36097	0.31795	0.30835	0.35544	0.45561	0.71129	0.6853	0.74739	0.878
16	56	0.005	0.74276	0.68719	0.44414	0.39193	0.23628	0.2241	0.2365	0.26236	0.27066	0.24
17	618	0.013	0.53473	0.50936	0.5221	0.55582	0.50864	0.83068	0.76362	0.5891	0.55016	0.534

Figure 7 CSTEP Result File Saved: A- ID, B- Probability, C and after - Scenario Columns