Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

DESIGNING SIMULTANEOUS SIMULATION EXPERIMENTS

Paul Hyden

School of Operations Research
and Industrial Engineering
Cornell University
206 Rhodes Hall
Ithaca, NY 14853, U.S.A.

Lee Schruben

Department of Industrial Engineering
and Operations Research
University of California at Berkeley
4135 Etcheverry Hall
Berkeley, CA 94720-1777, U.S.A.

ABSTRACT

Simulation experiments are often designed assuming that a fixed, and known, computing budget is to be allocated sequentially among different alternatives. However, in actual simulation experiments, there may be budget uncertainty or at least flexibility - for example, when there is a soft deadline for obtaining the study results. In such situations, it may be beneficial to allocate resources simultaneously in dynamically changing proportions. In this paper, we will examine optimal resource allocation paths. These paths climb the contour curves of the probability of selecting the best of several alternatives in a manner that insures that the highest probability of correct selection P(CS) is obtained when the study is halted. To gain insight into the complexity of optimal resource allocation paths, simple models exhibiting serial correlation, cross correlation, and trends are studied.

1 INTRODUCTION AND BACKGROUND

Simulation study design often focuses on the serial allocation of a predetermined computer budget, such as the sequential efforts of Chick and Inoue (1998) and Chen, Yücesan, and Dai (1998). In addition, these approaches focus on output that is independent and identically distributed. As in Schruben (1997), we want to exploit some of the advantages of simultaneously replicating different models with continuously variable allocations of effort to each model. We focus on an graphical representation of the problem in cases where the output series have serial correlation, cross correlation, and trends.

Simulation experiment budgets can include more than computing resources. For example, a study deadline or constraints on the analysts' time may be more critical than computer time. Furthermore, the exact amount of computing resources available until the study deadline is reached may not be known - particularly for important studies were extending deadlines may be preferred to making the wrong decision. Automation of experimental design decisions may be critical to efficiently using analysts' time or maximizing the use of available computing resources before a deadline is reached.

This is further complicated by the fact that when an experiment is halted may depend on the results obtained - budgets may not be set a priori but are negotiable as the study proceeds. A study may end sooner than anticipated when the early results suggest a clear answer, allowing attention to be directed toward other studies or other aspects of the same project. Conversely, more information may be desired when the performances of competing alternatives are close and the project will require further resources. An analyst may realize that a marginal increase in effort will result in a significantly improved answer. Not uncommonly, a reformulated study objective may be dictated to support a failing "pet" option or to invalidate the current best choice, requiring the study to continue. For these, and other, reasons it makes sense that simulation studies be designed to give the best answers available at any time during the study.

2 P(CS) CURVES AND OPTIMAL ALLOCATION PATHS

The simplest example that illustrates these ideas, is the problem of determining which of two random processes has the greater asymptotic mean. The experimental design problem is simply to determine how long to observe each process. We will consider more realistic situations later; however, for the time being, we will assume that, unknown to us, the underlying random variables for the two series are i.i.d. N(1,16) and N(0,16), respectively. The probability that the first sample mean is larger than the second sample mean can be given in a contour plot for different sampling combinations. The vertical axis represents the number of points sampled from the first series and the horizontal axis represents the number of samples from the second series. Since the first series has a larger asymptotic mean, the probability of correct selection (P(CS)) is the probability that the first sample mean is greater than the second sample mean. This contour plot is given in Figure 1. These P(CS) curves will help us determine the optimal design for an experiment.

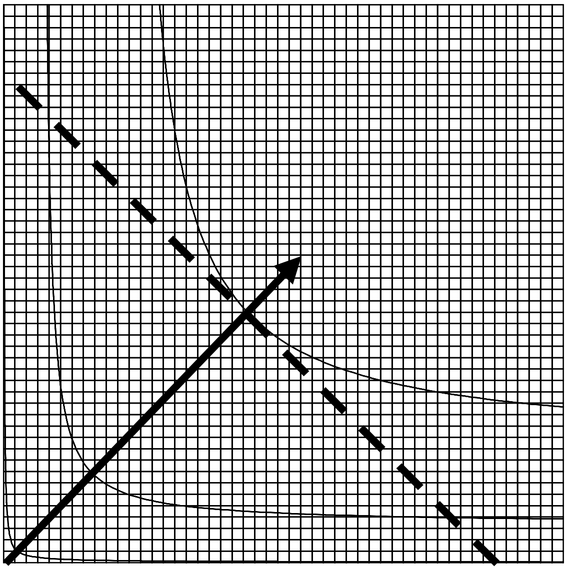

Figure 1: P(CS) Contours – Optimal Allocation Path

To decide on the optimal allocation, we also need to know the relative sampling costs for the two series. If the relative sampling costs are equal, then any given sampling budget will give an allocation falling on a line parallel to the dashed line in Figure 1. The optimal allocation for any budget is the tangent point of that budget line to the highest P(CS) contour. Varying the budget gives a series of optimal allocations. The optimal allocation path is given as the solid arrow in Figure 1. Different sampling costs will yield different optimal allocation paths. Figure 2 gives a representative budget line and optimal allocation path for the same problem given in Figure 1, when samples from series 2 are three times as costly as samples from series 1.

We considered but do not present the optimal sampling paths when the relative sampling costs vary over time. In this case, the budget lines are no longer be mutually parallel.

The optimal allocation path will also be a function of the output series. By changing the variance of the second series to 36, the optimal allocation path shifts toward sampling series 2 more. This case is given in Figure 3.

To contrast with the optimal allocation path, consider sampling each series equally. This choice of paths could cause the allocation of resources to one of the models too early and may be wasteful. Since we can't take back our

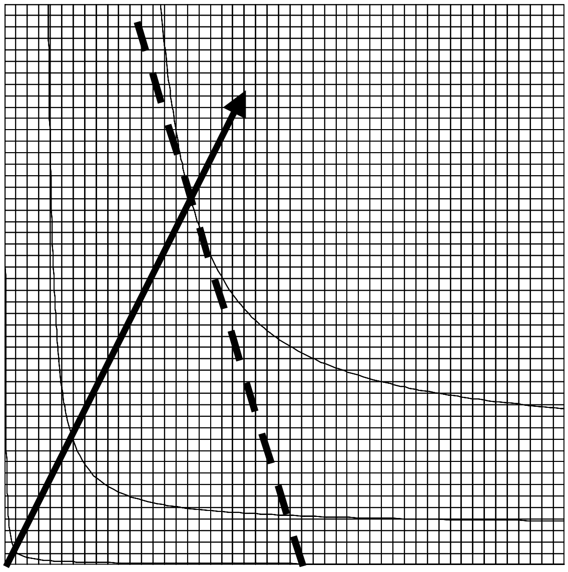

Figure 2: Different Relative Sampling Costs

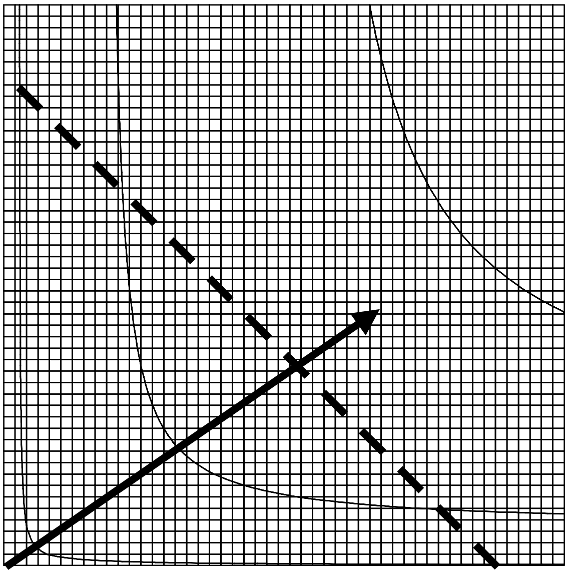

Figure 3: Unequal Variances

samples, we may not be able to recover quickly from a poor allocation early in our study. In addition, this path is highly dependent on a fixed budget. If, in the worst case, the budget for this study were cut off after sampling from only one model, the solution quality will be no better than if the study were never done.

Similar paths are produced when the budget is broken up into finite stages. Alternatively, we may simply assign a fixed allocation between the two samples at the beginning of the study. These paths will give much better answers if the budget is modified during the course of the study. However, the quality of the answer will be very dependent on the initial allocation choice. A multistage version of this path modifies the allocation during the study. In the limit, a combination of these paths approach the optimal allocation path. Figure 4 illustrates this variety of alternative paths in the context of Figure 3. For reference, the dashed arrow indicates the optimal allocation path.

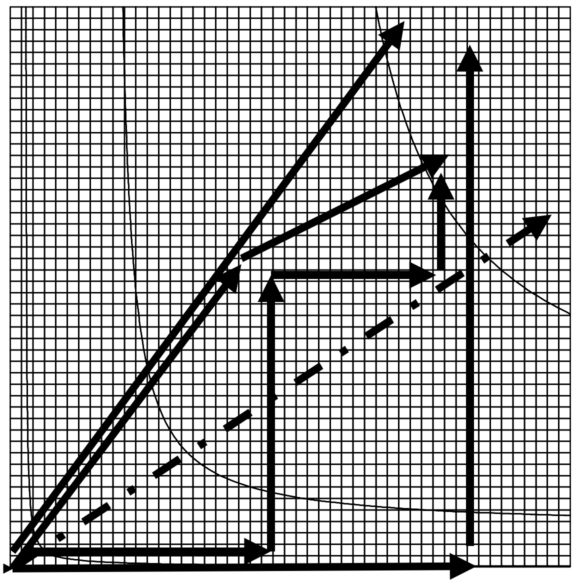

Figure 4: Alternatives to the Optimal Allocation Path

These P(CS) curves will be used to examine the effects of serial correlation, cross correlation, and trends in three simple pairs of models. In all cases, we take the sample mean to be the performance measure. Our decision rule chooses the model with the current highest sample mean. In addition, for simplicity, we will assume for the remainder of the paper that sampling costs are the same for both models.

3 SERIAL CORRELATION

The first model focuses on serial correlation. Here we use an AR(1) process for each of the two models i=1,2.

$$X_{i,t} = \phi_i X_{i,t-1} + Z_{i,t}$$

Where $\{Z_{i,t}\}$ are independent and identically distributed normal random variables and $\phi_i \in (0,1)$. The models are initialized at $X_{i,0}=0$, but this point is not included in the sample mean. Note that the asymptotic first order serial correlation of model i is ϕ_i.

By choosing parameters for the models such that $\lim_{r\to\infty} E(X_{1,t})=1/3$ and $\lim_{r\to\infty} E(X_{2,t})=0$, $\phi_i=0.7$ for i=1,2 and $var(Z_{i,t})= 1$ for all i and t, we generate the contours in Figure 5.

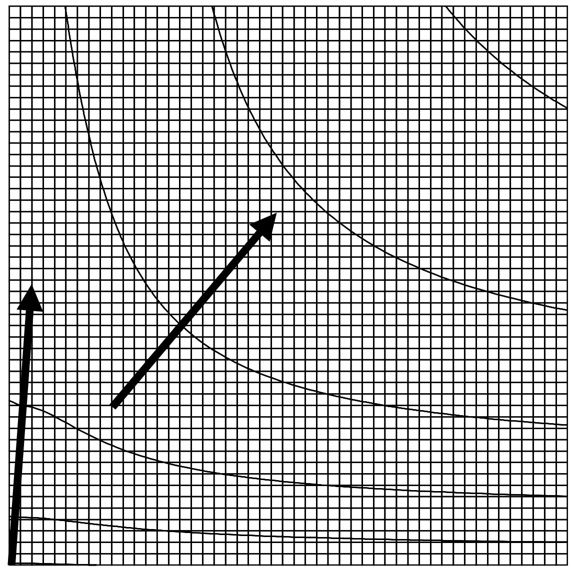

Figure 5: Positive Serial Correlation

The solid arrow indicates the optimal allocation path for equal sampling costs. The discontinuity occurs because the contours in the range of the discontinuity closely match the slope of the budget line. In the space of adding an additional sample to the budget, the contours shift just enough to dramatically change the optimal allocation. Figure 6 shows a close up of the contours in the region of the discontinuity. For reference, the budget line is shown as the dotted line.

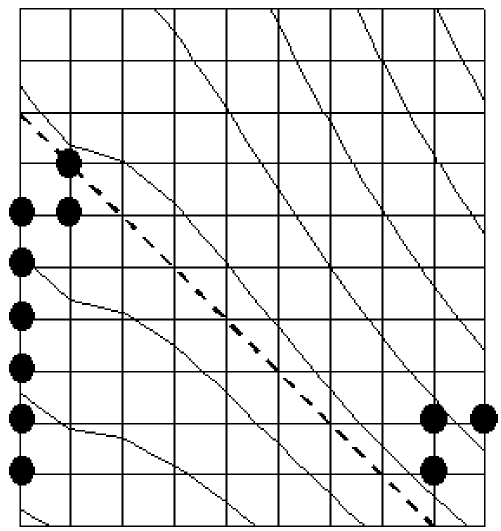

Figure 6: Close-up of Discontinuity

In Figure 7 we choose the same parameters as in Figure 5 except ϕ_i =-0.7 for i=1,2. Using the same scale on the contours as in Figure 4, we see that the same solution quality comes at a much lower cost. Empirically, we observe that the same budget will yield a poorer (better) quality solution in the presence of positive (negative) serial correlation. We also observe a different optimal allocation path. Note that the general character of the contours greatly resembles the case found in Figure 1.

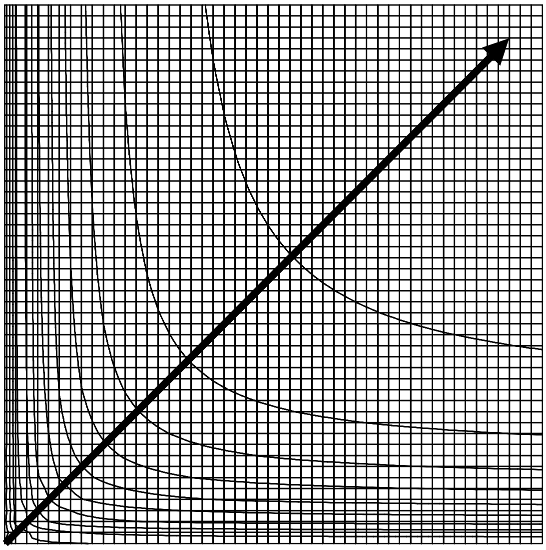

Figure 7: Negative Serial Correlation

To contrast positive and negative serial correlation, Figure 8 shows the contour plot for the same parameters in Figure 5 except ϕ_1 =0.7 and ϕ_2 =-0.7. Not surprisingly, the optimal allocation path forces more samples on model 1 and provides solution quality between that of the two previous cases.

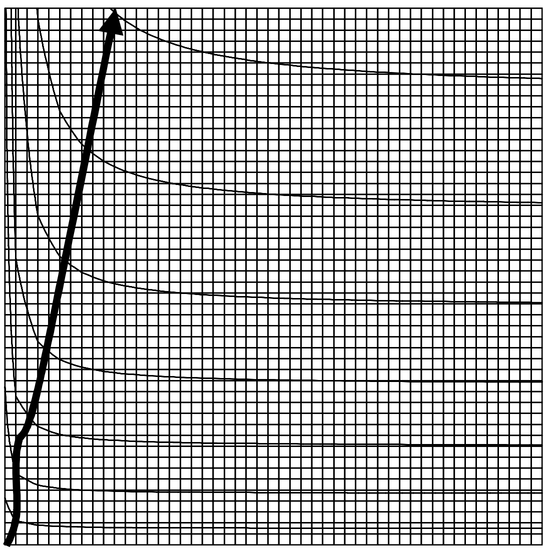

Figure 8: Both Positive and Negative Serial Correlation

4 CROSS CORRELATION

The next pair of models illustrates the impact of cross correlation. The two models are defined as follows.

$$\text{Model 1: } X_t = Z_t + A_t, \; t>0$$
$$\text{Model 2: } Y_t = W_t + A_t, \; t>0$$

$\{A_t\}$, $\{Z_t\}$, and $\{W_t\}$ are independent. $\{A_t\}$, $\{Z_t\}$, and $\{W_t\}$ are normal random variables. Both series are initialized at X_0=0 and Y_0=0 but the sample mean does not include this value. The variance of the shared series $\{A_t\}$ generates the cross correlation between these two output series.

For these models, setting $E(Z_t)=2$, $E(W_t)=1$, $E(A_t)=0$, $Corr(X_t,Y_t)=10/11$ and $Var(Z_t)=Var(W_t)=1$, we get the optimal sampling path given in Figure 9. Notice that the optimal allocation path closely follows an even split between the two models, because the variance due to $\{A_t\}$ can be completely eliminated when the two sample sizes are equal. For this reason, the optimal allocation path does not change when either $Var(Z_t)$ or $Var(W_t)$ is increased to 20 while the other variance remains at 1. With $Var(Z_t)=25$, $Var(W_t)=1$ and $Corr(X_t,Y_t)$ now at 0.51, the optimal allocation path finally begins to favor model 1, as shown in Figure 10. Also note that, due to the shape of the contours, the optimal allocation path is not affected by the relative sampling costs.

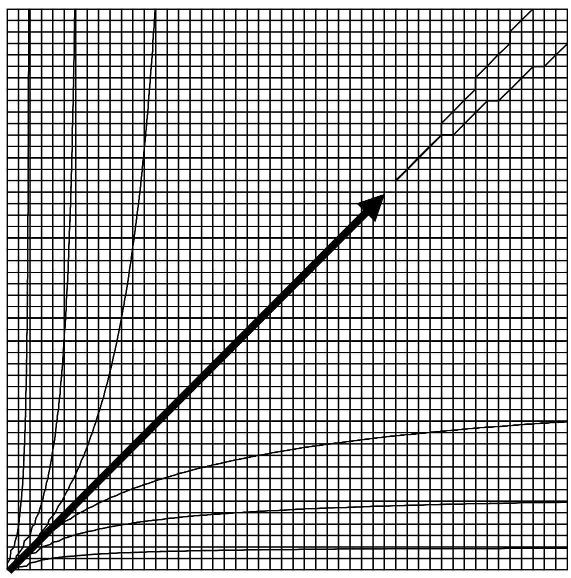

Figure 9: Positive Cross Correlation

For the case of negative cross correlation, we use the same model as above except for a modification to model 2.

$$\text{Model 2: } Y_t = W_t - A_t$$

Setting $\lim_{r\to\infty}E(X_t)=2$ and $\lim_{r\to\infty}E(Y_t)=1$, and $Var(Z_t)=Var(W_t)=1$ as before with $Corr(X_t,Y_t)=-10/11$, we get Figure 11.

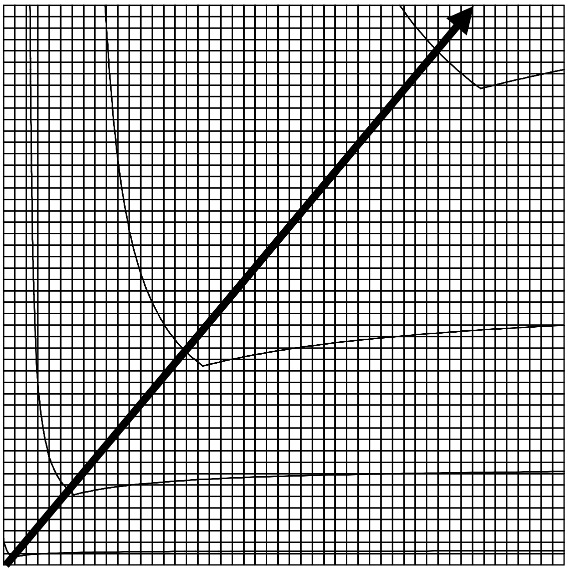

Figure 10: Positive Cross Correlation

For this special case, there are two optimal allocation paths, but a slight change in any of the parameters will favor one path. Now the relative sampling costs will again affect the optimal allocation path. Also, the optimal sampling path is much more sensitive to differences in variance for the two models. In addition, since the scale is the same as that in Figure 11, we see that the same solution quality comes at a lower cost. Empirically, we observe that the same budget will yield a better (poorer) quality solution in the presence of negative (positive) cross correlation.

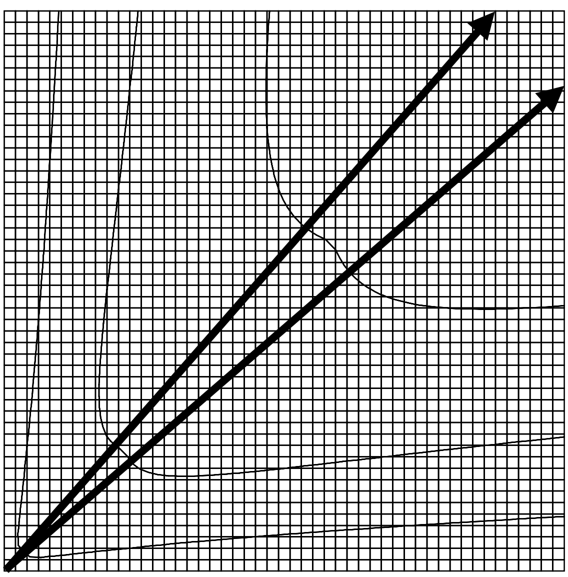

Figure 11: Negative Cross Correlation

5 TRENDS

The final pair of models focuses on the impact of a trend. In this model, the trend disappears asymptotically, in a similar way to warm-up.

$$Y_{i,t}=X_{i,j} + (\alpha_i)^{t-1},\ i\in\{1,2\},\ t>0$$

$X_{1,t}\sim$ i.i.d. $N(\mu_1,\sigma_1^2)$,$X_{2,t}\sim$ i.i.d. $N(\mu_2,\sigma_2^2)$ and $\alpha_i \in (0,1)$. The series is initialized at $Y_{i,0}=1$ for both models, but this value is not included in the sample mean. We have a steady state preference for model 1 if $\mu_1>\mu_2$ and model 2 otherwise.

The optimal path for this situation is shown in Figure 12. Here, we have chosen $(\mu_1,\sigma_1^2)=(0.001,1)$, $(\mu_2,\sigma_2^2)=(0,1)$, and $\alpha_1=\alpha_2=0.99$, so the steady state preference is for model 1. The solid line shows the optimal allocation path. The dotted line shows the worst possible path. The probability of correct selection actually decreases along this path.

This model pair shows an interesting characteristic. The optimal designs exaggerate the true difference between the two models. For example, if we observe model one m times and model two n-m times, then the expected value of the difference in sample means is

$$\mu_1 - \mu_2 + (1-\alpha_1^{m})/(m(1-\alpha_1)) - (1-\alpha_2^{(n-m)})/((n-m)(1-\alpha_2)),$$

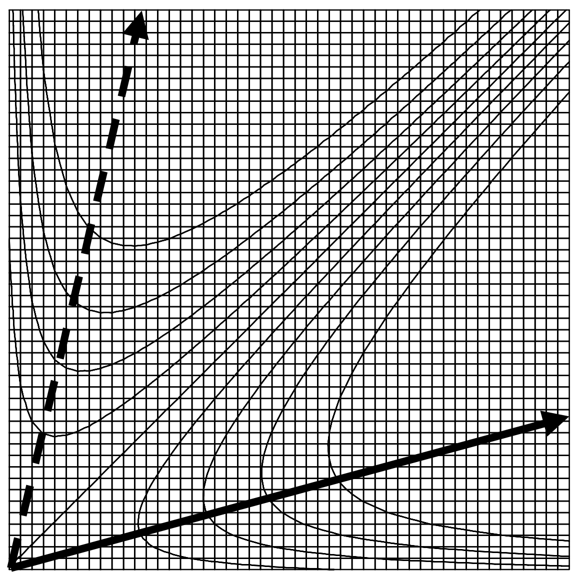

Figure 12: Trend

which is greater than $\mu_1 - \mu_2$ along the optimal allocation path. This also explains why the probability of correct selection can decrease, because the design of the experiment can favor the wrong conclusion as well. Of course, if this is the case, perhaps we should be asking which model will be favorable for the horizon of our study.

Finally, Figure 13 shows an interesting optimal allocation path where the allocations vary radically over the course of the study. The parameter settings are $(\mu_1,\sigma_1^2)=(10,16)$, $(\mu_2,\sigma_2^2)=(9.5,1)$, $\alpha_1=0.98$, and $\alpha_2=0.995$.

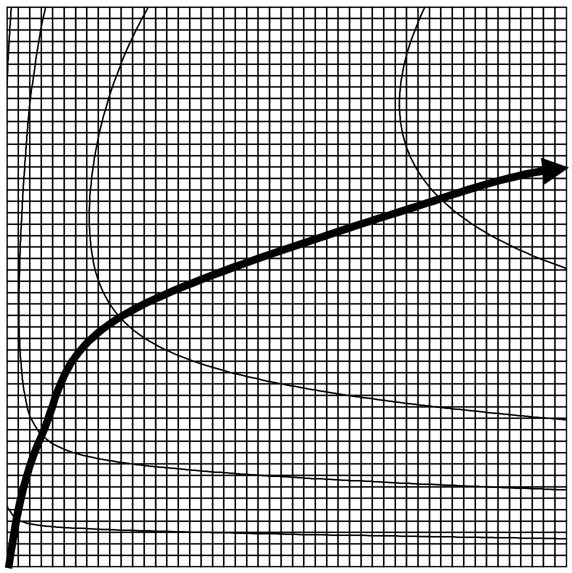

Figure 13: Trend with Highly Variable Path

6 CONCLUSION

P(CS) curves and optimal allocation paths provide perspective on the great untapped diversity of simulation experiment designs. Stepping away from the i.i.d. case to consider three fundamental attributes of time series output (serial correlation, cross correlation, and trends), we see a great motivation to move toward highly adaptive and simultaneous experimental designs. Further enrichments that examine initial conditions, time varying parameters, and variable sampling costs, as well as the interactions between all of these characteristics are worthy of further investigation.

ACKNOWLEDGEMENTS

The research reported here was partially supported by a joint SRC (FJ-490) and NSF (DMI-9713549) research project in semiconductor operations modeling, and a NSF Graduate Research Fellowship.

REFERENCES

Chen, C.H., E. Yücesan, and L. Dai. 1998. Computing Budget Allocation for Simulation Experiments with Different System Structures. In *Proc. of the 1998 Winter Simulation Conference*, ed. D.J. Medeiros, E.F. Watson, J.S. Carson, and M.S. Manivannan, 735-741.

Chick, S. and K. Inoue. 1998. Sequential Allocations That Reduce Risk for Multiple Comparisons. In *Proc. of the 1998 Winter Simulation Conference*, ed. D.J. Medeiros, E.F. Watson, J.S. Carson, and M.S. Manivannan, 669-676.

Schruben, L.W. 1997. Simulation Optimization Using Simultaneous Replication and Event Time Dilation. In *Proc. of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K.J. Healy, D.H. Withers, and B.L. Nelson, 177-180.

AUTHOR BIOGRAPHIES

PAUL HYDEN is a Ph.D. candidate at Cornell University's School of Operations Research and Industrial Engineering. His dissertation work focuses on designing efficient and practical simultaneous simulation experiments. He is also interested in semiconductor manufacturing, service operations, and finance.

LEE SCHRUBEN is in the Department of Industrial Engineering and Operations at the University of California at Berkeley. He was recently awarded the Andrew Schultz .Professorship at Cornell University from where he is currently on a leave of absence. His research interests are in statistical design and analysis of simulation experiments and in graphical simulation modeling methods. His simulation application experiences and interests include semiconductor manufacturing, dairy and food science, health care, banking, and the hospitality industry.