USING EMPIRICAL DEMAND DATA AND COMMON RANDOM NUMBERS IN AN AGENT-BASED SIMULATION OF A DISTRIBUTION NETWORK

William J. Sawaya III

School of Civil and Environmental Engineering
220 Hollister Hall
Cornell University
Ithaca, NY 18853, USA

ABSTRACT

Agent-based simulation provides a methodology to investigate complex systems behavior, such as supply chains, while incorporating many empirical elements relative to both systems structure and agent behavior. While there is a significant amount of simulation and analytical research investigating the impact of information sharing in supply chains, few studies have used empirical demand for the model. This research utilizes empirical distributions in order to determine the demand process faced by distribution centers in a distribution network. Therefore, the distribution centers face independent and heterogeneous demand that is not normal, and exhibits a much larger coefficient of variation than is generally utilized in similar research. With so much complexity and variability, contrasting different inter-organizational information sharing configurations provides an ideal setting for utilizing common random numbers for variance reduction. Comparisons made using this methodology show clear differences between the different information sharing schemes.

1 INTRODUCTION

There is already a considerable amount of research that explores the impact of shared information in supply chain settings (see Huang 2003). This setting is unique because of the variability and distribution of the empirically-informed demand data, the variability of the transportation lead-times, and their incorporation into an agent-based representation of the network. While the findings are not dissimilar to those of other related research, they extend the applicability to a more realistic, empirically-motivated setting. Using empirical distributions for demand has received very little attention as demand processes are generally assumed to be either Poisson or normal, generally with a very small coefficient of variation. The empirical data exhibited demand distributions that were neither normal nor Poisson, and had relatively large

coefficients of variation: ranging from .42 to 3.4 and with a mean of just over one for all the demands that were examined. With this high level of demand variability, and with variable transportation lead-times and production capacities, it seemed an ideal setting for utilizing common random numbers for variance reduction. This is the only study of this type, of which the author is aware, that utilizes common random numbers for variance reduction.

1.1 Inter-organizational Information Sharing

There is a wealth of available research on the effects of information sharing in supply network contexts. Many researchers examine the effect of passing point of sale (POS) demand data upstream (see Lee et al. 1997, Chen et al. 2000, Croson and Donohue 2003, Li 2003, Steckel et al. 2004) to mitigate the bullwhip effect. Others have also found that information sharing provides significant benefits to suppliers (see Bourland et al. 1996, Gavirneni et al. 1999, and Cachon and Fisher 2000). Haung et al. (2003) provide a detailed literature review of inter-organizational information sharing. This research extends that previous analytical and simulation research by exploring the questions of information sharing in contexts where it has not been examined. Specifically where there is considerably more variability, as indicative of the empirical data.

1.2 Conceptualizing Supply Networks as a Complex System

It is a natural step to identify supply chains as complex adaptive systems. Choi et al. (2001) argue aptly that supply chain networks should be recognized as complex adaptive systems. One of their key points is that in a complex adaptive system, organization often emerges rather than being prescribed. Thus, organizations can only manage supply chains directly to a certain extent, as some of the organization is likely to emerge over time. They further outline complex adaptive systems in a simi-

lar fashion to Anderson (1999), and provide a detailed mapping of each element of their description of complex adaptive systems to a supply network. But this conceptualization also recognizes that the complex relationships and interactions between organizations cannot always be abstracted away with impunity. Thus the behavior of one organization can have significant impact on both their immediate trading partners as well as organizations that are only peripherally related to organization taking action.

1.3 Agent-based Simulation Models of Complex Supply Networks

Agent-based simulation modeling of supply networks has received increasing attention over the last decade (see Lin 1996, Swaminathan et al. 1998, Tan 1999, Chang and Harrington 2000, Chatfield 2001, Chatfield et al. 2004, Pathak 2005, and Sawaya 2006). Zeggelink et al. (1996) observe that object oriented programming methods provide a natural way of modeling agents, their schemata, and their interconnections. One of the major benefits of this modeling methodology is the ability to model things from the ground up, without abstracting away much of the real systems complexity and interactions (see Anderson 1999).

1.4 Empirical Distribution Network Setting

A portion of the data for this study came from that used in Dong et al. (2005), the remainder was collected independently. An electronics component distributor provided the point of sale data for many of their products in many of the warehouses which they supply. As part of the data, 20 products at 5 different distribution centers (DCs) were analyzed from one organization. Input analysis was conducted for each of the approximately 90 demands since not every product was carried at every DC. This structure of 5 distribution centers was then implemented in the model with the empirical demands as well as the empirical average lead-times. Additional POS demand data from other sources and industries was also considered as well.

2 EMPIRICAL DEMAND PROCESS IN THE DISTRIBUTION NETWORK

2.1 Empirical Demand Input Analysis

Regarding input analysis, Vincent (1998) notes a paradoxical position regarding the accuracy and the representativeness of real data. Although the collection of empirical data is the best argument for specifying and evaluating a specific model; it should not be paramount to the model

and research question. That is, there should be inherent mistrust of models that are over-reliant on a particular "clean" data sample. The goal is robust strategies for modeling. His idea is part of the motivation in the use of empirical data to begin with. Most other models rely on normalized data with relatively small coefficients of variation. By using empirical data, this research relaxes the assumptions made in previous work regarding the shape and variance of the demand distributions.

Input analysis of these empirical demand streams using the ARENA input analyzer led to some interesting observations. First, none of the data was significantly fit to a normal distribution. Second, about one in five could be significantly fit to any distribution at all. But most fit best to beta distributions (lowest MSE) whether they were significant or not. Third, they all exhibited relatively high variance. The smallest coefficient of variation in any of the demand strings was 0.42 and the highest was 3.33, with a mean of 1.30 and a median of 1.13. Typical demand coefficient of variations ranges in previous research of this type range only as high as 0.4 for a high level of variance, largely due the lack of empirical motivation for most demand and use of normally distributed demand. Thus, the variation in demand observed here were significantly higher that those typically assumed. One product was chosen for extended analysis where there was demand at all five locations and the following summary statistics in Table 1 below. It should also be noted that there was no seasonality detected in any of the demand processes.

Table 1: Summary Data for Empirical Demand

	DC1	DC2	DC3	DC4	DC5
Daily Average Demand	5.61	32.17	6.96	12.27	10.68
Daily Standard Deviation	7.24	32.26	9.00	11.36	16.50
Coefficient of Variation of Demand	1.29	1.00	1.29	0.93	1.54
Best Distributional Fit (MSE)	Beta	Exponential	Beta	Beta	Beta
Significant Fit?	Yes	No	Yes	No	No

2.2 Demand Data Generation for Simulation

Additional data were generated for conducting simulations of length, since some demand streams consisted of only one years data. From the best fit distribution for each distribution centers' demand an additional 30,000 data points were generated using the ARENA input analyzer for each DC and then the data was split into 10 sets of 3,000. This resulted in a set of demand suitable for 10 replications of 3000 days length for each of the DCs for a total of 100 demand sets. Then an additional set of 10

replications of 3000 days was produced for each DC from a normal distribution with the low coefficients of variation for the sets with homogeneous and heterogeneous demand across the distribution centers. And the last set had empirically based distribution shapes, but which had reduced coefficients of variation and were shifted such that the means were the same as the empirically based set. These demand sets were used to explore the effect of the distributional shape on the network. In this model there is no negative demand, so in any distribution where the value was negative it is truncated to zero. This also helps establish that the observed results are not merely an artifact of the distributional shape of the empirical data.

3 SIMULATION MODEL CONSTRUCTION, VERIFICATION AND VALIDATION

3.1 Model Construction, Verification and Validation via Extreme Programming

In order to develop a model with high face validity at the early stages of model development, this research draws extensively on the existing theory and models of supply chains and networks. In addition, actual system observation is possible, in a limited sense, through the extensive literature and case studies that examine this type of behavior. Additionally, interviews provide "experts" to help the model maintain face validity. Ongoing discussions (inquiries) with academic researchers and business practitioners maintain this grounding. Further all of the principles suggested by Law and Kelton (2000) for developing a model with high face validity were employed in the construction and verification of this model. The model was constructed and run in the Eclipse programming environment (http://www.eclipse.org/). It was coded with the JAVA (http://java.sun.com/) programming language. To aid in the model verification process many of the elements of "extreme programming" (see http://www.extremeprogramming.org and http://www.xprogramming.com) were used. These have been developed by computer programmers in order to write code quickly, correctly, and they utilize best programming practice. They aid in ensuring that the code is written correctly and performs correctly and involved programming in modules and performing constant tests and checks to make sure that everything performs properly.

Agents were generically programmed as organizations which can purchase, sell, and manufacture objects. The make decisions relative to purchasing, production, and inventory levels based on information available to them. These generic agents can then be specialized to function as suppliers, manufacturers, distribution centers or customers. Multiple layers of organizations and any

set of network relationships can be established given the fully configurable nature of the coding. Ordering, stocking, and production decisions may be individually specified for each agent. In addition, transportation times, and information sharing between organizations can also be individually specified for each trading relationship within the supply network. The network structure used in this research consisted of five distribution centers, with unique transportation times, based on empirical data as seen below in Figure 1.
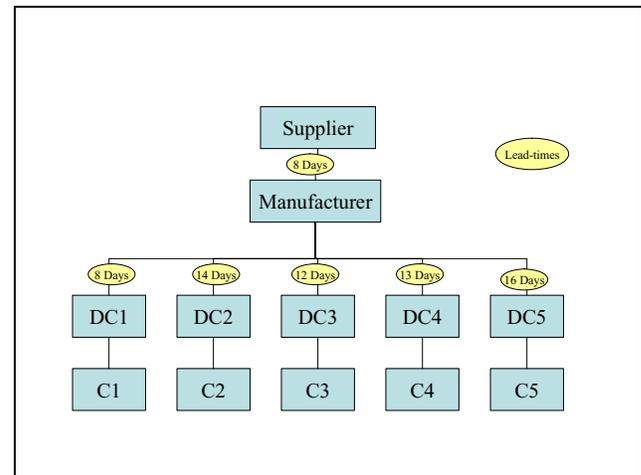

Figure 1: Supply Network Configuration

3.2 Turing Test for Validation of Simulation Output

Additional empirical opportunities for validation were possible due to the cooperation of a system "expert" and primary decision maker, as well as in the availability of real data relative to both system inputs and a limited set of output for one of the distribution networks. By collecting actual data on the network structure, decision rules, and performance data; the model can be validated via the second and third methods suggested by Naylor and Finger (1967), namely testing the assumptions of the model empirically, and determining how representative the simulation output data are of the actual system data. These are both discussed in greater detail in Law and Kelton (2000).

The additional data available provided the opportunity for a, so called, Turing test (see Schruben 1980 and Turing 1950), which further established the validity of the model. Michael Ness, Manager and Owner of Ness Electronics, in Saint Paul Minnesota provides detailed information about the operations of three separate distribution centers and well as empirical data of actual orders, sales and inventory levels. This data was used to construct agents similar to the actual distribution centers. The actual demand data was then used as an input and the resulting patterns of inventory levels for each distribution cen-

ter was then graphically compared to the actual inventory patterns by Michael Ness. Who confirmed that the modeled agents where behaving in a similar fashion to the actual organizations. By applying this procedure and the principles of extreme programming, validation is achieved to an extent that is not common in this type of research. It also provides an example of validation in an agent-based simulation setting of a supply network. Thus the model is credible for use in exploring research questions that have been laid out relative to information sharing, and their impact on performance of the system in a setting that is informed by extensive empirical data.

4 UTILIZING COMMON RANDOM NUMBERS METHOD OF VARIANCE REDUCTION TO COMPARE DIFFERENT INFORMATION SHARING CONFIGURATIONS

The use of common random numbers (CRN) is an important variance reduction technique (Law and Kelton 2000) specifically for simulation experiments where alternate system configurations are considered (Goldsman and Nelson 1998). See Law and Kelton (2000) for a detailed mathematical explanation of CRN. The method of CRN essentially means putting exactly the same inputs into different model system configurations. Therefore, if the stream of random numbers has peculiarities, then both of the configurations must deal with the same elements of randomness. This is analogous to blocking on the random behavior of the input in a designed experiment. In this case we are blocking for the randomness in the random numbers and in a complex systems setting if a small deviation has a large impact, both models will see it at the same time. Therefore great care and effort must be taken when using CRN for all of the random model inputs such that they are properly synchronized. Thus, every single combination of demand strings, whether empirical, empirically-based generated strings, or purely fabricated are stored and tracked so that they can be run through every model. The result is a large number of demand string files that must be generated, tracked and reused. Likewise, when the capacity is variable, the variation is blocked using CRN.

4.1 Benefits of Common Random Numbers in This Study

Testing the efficacy of using CRN can be done via a simple test for two sets of data simulated using CRN (Law & Kelton 2000). Let $S_1^2(n)$ and $S_2^2(n)$ be the sample variance of the X_{1j}'s and the X_{2j}'s respectively where X_{1j} is the parameter output for replication j from system 1 and X_{2j} is the parameter output for replication j from system 1. Simply by defining $Z_j = X_{1j} - X_{2j}$; since the runs were

made with common random numbers $S_Z^2(n)$ is an unbiased estimator of the variance of Z_j. Regardless of whether CRN are used or not $S_1^2(n)$ is an unbiased estimator of the Var(X_{1j}) and $S_2^2(n)$ is an unbiased estimator for the Var(X_{2j}), so $S_1^2(n) + S_2^2(n)$ is an unbiased estimator of the variance of Z_j. Thus, if the CRN method is working we would expect that $S_Z^2(n) < S_1^2(n) + S_2^2(n)$, and the difference estimates how much the use of CRN is reducing the variance of Z_j. There are a few examples of the use of common random numbers backfiring. Therefore the previous test is performed on the per unit cost output for some of the simulation runs. In no case was the induced correlation estimated to be negative, and it appears to have a increasing effect as the randomness in the models increased. For instance, the variance reduction seemed to be much greater in the case of variable capacity. In many cases the variance reduction appears to cut the variance in half or reduce it even further which indicates that this was an appropriate variance reduction technique to employ as can be seen in Tables 2 and 3 below.

Table 2: Sample Variances for Different Information Sharing Configurations with Constant Capacity

	$S_1^2(n)$	$S_2^2(n)$	$S_1^2(n) +$ $S_2^2(n)$		$S_Z^2(n)$
DC1	0.0528	0.0236	0.07634	>	0.03325
DC2	0.0776	0.0416	0.11922	>	0.01959
DC3	0.0168	0.0119	0.0287	>	0.0121
DC4	0.0281	0.012	0.04014	>	0.00864
DC5	0.0353	0.0152	0.05048	>	0.00685
MFG	2E-05	3E-05	4.4E-05	>	2.5E-05

Table 3: Sample Variances for Different Information Sharing Configurations with Variable Capacity

	$S_1^2(n)$	$S_2^2(n)$	$S_1^2(n) +$ $S_2^2(n)$		$S_Z^2(n)$
DC1	0.1357	0.1538	0.2895	>	0.00389
DC2	0.3128	0.2419	0.55468	>	0.01171
DC3	0.2038	0.0561	0.25994	>	0.02732
DC4	0.2232	0.1128	0.33597	>	0.04514
DC5	0.2306	0.1616	0.39222	>	0.00809
MFG	0.0084	0.0057	0.01406	>	2.3E-05

Tables 2 and 3 each show the sample variance for 10 replications. The first configuration is with no information being shared beyond orders placed from the DCs to the manufacturer. This is a typical configuration of 'no' information sharing. The second information sharing configuration passes daily point-of-sale data from each of the 5 DCs to the manufacturer. As increasing variability

is induced in the system, the benefits of the common random numbers increases as greater covariance is induced between the simulations of different system configurations. In addition to the sharing of point-of-sale information upstream from the distribution center to the manufacturer, ultimately other configurations can be examined including the sharing of lead-time estimates downstream from the manufacturer to the distribution centers.

5. CONCLUSION AND DISCUSSION

The results comparing the different information sharing configurations are generally significant. The sharing of point-of-sale information is beneficial to the manufacturer and the distribution network system as a whole. Further, the individual distribution centers are generally no worse off for the sharing of their point-of-sale information with the manufacturer. Many similar studies find a decreasing significance for sharing demand information up the supply chain as the variance of the demand increases. In this experimental setting there are much larger demand variances than those used in other similar studies. In addition the use of random lead-times ensures that this model has greater variability than other models of this type. Nonetheless, here the findings indicate extremely significant benefits to the sharing of this type of information; in both practical significance as well as in statistical significance. If other similar studies used this methodology it is likely that they would have found much more significant results; particularly considering that they are generally using constant lead-times and much lower coefficients of variation in their demands. One might further speculate that a reason it is not done is the effort involved in keeping track of so many different demand strings. Additionally the induced covariance can complicate statistical analysis. But the benefits of the variance reduction and the ability to examine complex systems, such as supply networks, make the effort well worth it. The increasing use of CRN in agent-based models will undoubtedly aid in study of all types of complex systems.

REFERENCES

Anderson, P. 1999. Complexity theory and organization science. *Organization Science*, Vol. 10, No. 3, 216-232.

Bourland, K. E., S. G. Powell, and D. F. Pike. 1996. Exploiting timely demand information to reduce inventories. *European Journal of Operational Research*, Vol. 92, 239-253.

Cachon, G. P., and M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Science*, Vol. 46, No. 8, 1032-1048.

Chang, M.-H., and J. E. Harrington Jr. 2000. Centralization vs. decentralization in a multi-unit organization: A computational model of a retail chain as a multi-agent adaptive system. *Management Science*, Vol. 46, No. 11, 1427-1440.

Chatfield, D. C. 2001. SISCO and SCML – Software tools for supply chain simulation modeling and information sharing. Ph.D. Dissertation, The Pennsylvania State University.

Chatfield, D. C., J. G. Kim, T. P. Harrison, and J. C. Hayya. 2004. The bullwhip effect – impact of stochastic lead time, information quality, and information sharing: A simulation study. *Production and Operations Management*, Vol. 13, No. 4, 340-353.

Choi, T. Y., K. J. Dooley, and M. Rungtusanatham. 2001. Supply networks as complex adaptive systems: Control versus emergence. *Journal of Operations Management*, Vol. 19, 351-366.

Croson, R., and K. Donohue. 2003. Impact of POS data sharing on supply chain management: An experimental study. *Production and Operations Management*, Vol. 12, No. 1, 1-11.

Chen, F., Z. Drezner, J. K. Ryan, and D. Simchi-Levi. 2000. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, Vol. 46 No. 3, 436-443.

Dong , Y., M. Dresner, and Y. Yao. 2005. Beyond information sharing: The value of vendor managed inventory to downstream firms. Working Paper.

Gavirneni, S., R. Kapuscinski, and S. Tayur. 1999. Value of information is capacitated supply chains. *Management Science*, Vol. 45, No. 1, 16-24.

Goldsman, D., and B. L. Nelson. 1998. Comparing systems via simulation. Chapter 8 *in Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, J. Banks ed., John Wiley and Sons.

Huang, G. Q., J. S. K. Lau, and K. L. Mark. 2003. The impacts of sharing production information on supply chain dynamics: a review of the literature. *International Journal of Production Research*, Vol. 41, No. 7, 1483-1517.

Lin, F.-r. 1996. Reengineering their order fulfillment process: A multi-agent information systems approach. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling & analysis.* 3rd ed. New York: McGraw-Hill, Inc.

Lee, H. L., V. Padmanabhan, and S. Whang, 1997. Information distortion in a supply chain: the bullwhip effect. *Management Science*, Vol. 43, No. 4, 546-558.

Naylor, T. H. and J. M. Finger. 1967. "Verification of Computer Simulation Models." *Management Science, Vol.* 14, No. 2, B92-B106.

Pathak, S. D. 2005. An investigative framework for studying the growth and evolution dynamics of supply networks. Ph.D. Dissertation, Vanderbilt University.

Sawaya III, W. J. 2006. An investigation of the performance impact of the extent of inter-organizational information sharing: Using a complex adaptive system paradigm and agent-based simulation. Ph.D. Dissertation, University of Minnesota.

Schruben, L. W. 1980. Establishing the Credibility of Simulations. *Simulation*, Vol. 34, No. 3, 101-105

Steckel, J. H., S. Gumpta, and A Banerji. 2004. Supply chain decision making: Will shorter cycle times and shared point-of-sale data help? *Management Science*, Vol. 50, No. 4, 458-464.

Swaminathan, J. M., S. F. Smith, and N. M. Sadeh. 1998. Modeling supply chain dynamics: A multiagent approach. *Decision Sciences*, Vol. 27, No. 3, 607-632.

Tan, G. W. 1999. "The impact of demand information sharing on supply chain network," Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

Turing, A.M. 1950. "Computing machinery and intelligence," *Mind*, Vol. 59, 433-460.

Van Horn, L. 1971. Validation of simulation results. *Management Science*, Vol. 17, No. 17, No. 5, 247-258.

Vincent, S. 1998. "Input data analysis," Chapter 3 *in Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, J. Banks ed., John Wiley and Sons.

Zeggelink, E. P. H., R. Van Oosten, and F. N. Stokman. 1996. "Object oriented modeling of social networks," *Computational & Mathematical Organization Theory*, Vol. 2, No. 2, 115-138.

AUTHOR BIOGRAPHIES

WILLIAM J. SAWAYA III is a post doctoral associate in the department of Civil and Environmental Engineering at Cornell University. He completed in Ph.D. at the in Operations and Management Science from the Carlson School of Management at the University of Minnesota. His research interests include the impact of inter-organizational information sharing on organizational performance, the simulation of complex systems, and the impact of supply disruptions on organizations' operations. E-mail: <wjs32@cornell.edu>