Preliminary Analysis of Distributional Equivalence in a Model Docking Study*

Brian F. Tivnan

btivnan@mitre.org

Command and Control Center

The MITRE Corporation, McLean, VA 22102 USA

August 15, 2006

*This research was conducted in conjunction with attendance at the 2006 Santa Fe Institute Complex Systems Summer School. As such, the author gratefully acknowledges financial support from the Santa Fe Institute and The MITRE Corporation. Additionally, the author would also like to acknowledge technical support from Stephen Upton. The remaining shortcomings, of course, remain the sole responsibility of the author as do the opinions and views expressed in this paper, which do not necessarily reflect those of the aforementioned organizations or person.

This paper briefly describes an analytical extension to Axtell et al. (1996) in their research on model alignment, or “docking.” A replication of March’s (1991) Organizational Code Model (OCM) and the original OCM are the two models to be docked. Tivnan (2006b) provides a detailed formulation of the original OCM as well as a discussion of the initial replication study which established relational equivalence between the original OCM and the replicated OCM – termed OCM(REP). This paper describes the analysis to determine the distributional equivalence between the OCM and OCM(REP).

Because attempts to locate an executable form of March’s OCM were unsuccessful (e.g., repeated searches of the literature and the Internet did not uncover any instances of the model), the previous phase of this study involved replicating March’s results using the model formulation in the published paper. However, preliminary results from the replication did not consistently demonstrate relational equivalence between March’s original results and those from this replication. Not until the personal correspondence with Riolo (2004) and his gracious sharing of his research notes from the Cohen, Axelrod Riolo (CAR) Project (1997) was a robust, relational equivalence confirmed between the original results and those from this replication. Relational equivalence differs from distributional equivalence; the former being that the same internal, qualitative patterns are reflected in the results from the two models in a docking exercise, the latter being the difference in the results from the two are not statisitically significant (1996).

Following the presentation to the 2006 Organization Science Winter Conference (Tivnan, 2006a)of the results from the previous docking exercise to establish relational equivalence, the author had the distinct honor of March's detailed analysis of the replicated model, dynamics and graphics – after which he confirmed that indeed the original rules, activation sequence and dynamics had been successfully replicated. March provided the original computer program for the OCM, and in order to preserve the independence of this study, the original OCM computer program was provided at that time to an independent and expert software engineer, Stephen Upton.

Upton was able to run March's original code using QuickBasic, therefore producing 50 MarchExperiments with the baseline model. Because the actual data was not available from March's (1991) original research, results from Upton's 50 MarchExperiments provide the raw data for this docking exercise and the subsequent analysis of the distributional equivalence between March's OCM and this replication.

Although the author is extraordinarily grateful to James March for his willingness to personally engage in the analysis of this replication and provide his feedback, successful replications should not hinge on opportunities for subsequent researchers to obtain direct access to the original researchers, their models and their data. Since a return to providing complete computer programs in published works (e.g., both Cyert and March (1963) as well as Cohen, March and Olsen (1972) each contained complete versions of the original computer programs) seems impractical given the complexity of today's models, the inclusion of a comprehensive set of pseudo-code (e.g., both Epstein and Axtell (1996) as well as Axtell (1999) provide more than a sufficient model formulation in addition to an abundance of pseudo-code) should become compulsory for publication.

Kolmogorov-Smirnov Tests Between OCM and OCM(REP)

As with Axtell et al. (1996), this comparison applies the Kolmogorov-Smirnov goodness-of-fit (K-S) test to initially establish the distributional equivialence between OCM and OCM(REP). The K-S test is appropriate for determining whether two underlying probability distributions statistically differ from each other, or whether an underlying probability distribution statistically differs from a hypothesized distribution. In both cases, the K-S test requires finite samples. This analysis relies on the former case, two samples and whether their underlying empirical cumulative distribution functions statistically differ. The usefulness of the two-sample KS test as a general nonparametric method for comparing two samples arises from the test's sensitivity to differences in both location and shape of the empirical cumulative distribution functions of the two samples.[1] The null hypothesis in the two-sample case is that the two samples share the same cumulative distribution function.

To confirm the distributional equivalence between OCM and OCM(REP), three sets of K-S tests were conducted. The first set consisted of three independent K-S tests, each test comparing one random March Experiment from OCM with one random March Experiment from OCM(REP). Recall from Chapter 4 that a March Experiment consists of 80 replicates for each of 27 design points (i.e., a complete parametric sweep of p_1 = 0.1, 0.2,…, 0.9 and p_2 = 0.1, 0.5 and 0.9) for a total of 2160 runs.

The second set also consisted of three independent K-S tests, with each test comparing one random March Experiment from OCM with another random March

---

[1] http://en.wikipedia.org/wiki/K-S_test

Experiment also from OCM. The third set once again consisted of three independent K-S tests, here each test comparing one random March Experiment from OCM(REP) with another random March Experiment also from OCM(REP). Notice that the first set of tests is intended to confirm distributional equivalence *between* OCM and OCM(REP), whereas the second and third sets of tests are intended to confirm similar internal consistencies *within* OCM and OCM(REP), respectively. All tests compare the data as if generated from one design point (i.e., one test with 2160 observations from each sample) and for 27 design points (i.e., 27 tests, each with 80 observations from each sample).

Table 1 contains the results from these multiple comparisons of distributional equivalence using the K-S test. In all comparisons as if the data was generated from one design point, no statistically significant differences existed at the conventional standard of 0.05 level of significance. Therefore, the "Notes" column reflects which of the 27 design points in that test and its corresponding p-value where a statistically significant difference between the two samples occurred. Or, if no statistically significant difference existed between the two samples, the "Notes" column simply reports which of the 27 design points in that test had the lowest corresponding p-value.

KS Test of Distributional Equivalence between OCM and OCM(REP)– heading 2

In each of the three independent comparisons, considering the data as if generated from one design point, no statistically significant differences existed between the two samples. In two of the three independent comparisons where it was assumed the data was generated from 27 independent design points, no statistically significant differences existed between the two samples. In one comparison, assuming 27 design points, one statistically significant difference existed for one of the 27 design point. Therefore, even ignoring the results from the tests assuming one design point, only one statistically significant difference between OCM and OCM(REP) occurred in 81 (i.e., 3 comparisons, each consisting of 27 independent tests) independent, K-S tests.

Summary and Initial Conclusions

A brief discussion of inferential statistics follows. An unavoidable aspect of all inferential statistics results from the inherent likelihood of drawing incorrect inferences from statistical analysis. One such categorization of these incorrect inferences occurs from a Type I error – the incorrect rejection of the null hypothesis when the null hypothesis actually reflects the true state of nature (Larsen & Marx, 1986). By definition, the likelihood of committing a Type I error equates to the level of significance for the statistical analysis under consideration. In this discussion, the level of significance follows the conventional standard of 0.05 (Axtell, Axelrod, Epstein, & Cohen, 1996). That is, the researcher accepts the probability of committing a Type I error for one in every twenty tests. Therefore, for the purposes of this analysis, an average of approximately one Type I error would occur when testing the 27 design points of one OCM March Experiment with those from one OCM(REP) March Experiment (i.e., 27 independent, hypothesis tests).

In this analysis of distributional equivalence, fewer rejections of the null hypothesis resulted than the theoretical occurrence of Type I errors. Additionally, the OCM and the OCM(REP) demonstrated similar statistical signatures in internal comparisons of their respective distributional equivalence between March Experiments. These global outcomes enhance the robustness of this analysis and provide additional confirmative support to the distributional equivalence between OCM and OCM(REP).

REFERENCES

Axtell, R. L. (1999). The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions. In *Center on Social and Economic Dynamics* (Vol. Working Paper No. 3). Washington: Brookings Institution.

Axtell, R. L., Axelrod, R., Epstein, J. M., & Cohen, M. D. (1996). Aligning Simulation Models: A Case Study and Results. *Computational and Mathematical Organization Theory, 1*(2), 123-141.

Cohen, M. D., Axelrod, R., & Riolo, R. L. (1997). CAR Project: Replication of Eight "Social Science" Simulation Models. Retrieved August 5, 2006, from http://www.cscs.umich.edu/Software/CAR-replications.html

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A Garbage Can Theory of Organizational Choice. *Administrative Science Quarterly, 17*(1), 1-25.

Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

Epstein, J. M., & Axtell, R. L. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Washington: Brookings Institution Press.

Larsen, R. J., & Marx, M. L. (1986). *An Introduction to Mathematical Statistics and its Applications*. Englewood Cliffs, NJ: Prentice-Hall.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science, 2*(1), 71-87.

Riolo, R. (2004). RE: Code for March's (1991) Model.

Tivnan, B. F. (2006a). *March-ing Forward by Leaps and Boundary Spanning.* Paper presented at the 2006 Organization Science Winter Conference, Steamboat Springs, CO.

Tivnan, B. F. (2006b). *March-ing Forward by Leaps and Boundary Spanning: Coevolutionary Dynamics of the Apdative Tension between Exploration and Exploitation.* George Washington University, Washington, DC.

Table 1. Rejections of the Null Hypothesis in KS Tests

Batch Test #	Test #	OCM Batch #	OCM (REP) Batch #	Rejections 1 Design Point	Rejections 27 Design Points	Notes
1	1	42	47	0	1*	for $p_2 = 0.5$; $p_1 = 0.5$: p-value = 0.035
2	2	38	27	0	0	for $p_2 = 0.9$; $p_1 = 0.5$: p-value = 0.082
3	3	35	49	0	0	for $p_2 = 0.1$; $p_1 = 0.3$: p-value = 0.120
Batch Test #	**Test #**	**1st OCM Batch #**	**2nd OCM Batch #**	**Rejections 1 Design Point**	**Rejections 27 Design Points**	**Notes**
20	1	3	11	0	0	for $p_2 = 0.1$; $p_1 = 0.9$: p-value = 0.120
21	2	20	21	0	0	for $p_2 = 0.5$; $p_1 = 0.9$: p-value = 0.054
22	3	25	43	0	0	for $p_2 = 0.1$; $p_1 = 0.4$: p-value = 0.172
Batch Test #	**Test #**	**1st OCM (REP) Batch #**	**2nd OCM (REP) Batch #**	**1 Design Point**	**27 Design Points**	**Notes**
35	1	11	8	0	0	for $p_2 = 0.9$; $p_1 = 0.3$: p-value = 0.120
36	2	8	30	0	0	for $p_2 = 0.9$; $p_1 = 0.8$: p-value = 0.120
37	3	28	38	0	1*	for $p_2 = 0.1$; $p_1 = 0.3$: p-value = 0.035

* Level of significance = 0.05.