Market Efficiency, Sales Competition, and the Bullwhip Effect in the TAC SCM Tournaments

Patrick R. Jordan, Christopher Kiekintveld, Jason Miller, and Michael P. Wellman

University of Michigan
Computer Science & Engineering
Ann Arbor, MI 48109-2121, USA
{prjordan,ckiekint,mijason,wellman}@umich.edu

Abstract. The TAC SCM tournament is moving into its fourth year. In an effort to track agent progress, we present a benchmark market efficiency comparison for the tournament, in addition to prior measures of agent competency through customer bidding. Using these benchmarks we find statistically significant increases in intratournament market efficiency, whereas agents are generally decreasing in manufacturer market power. We find that agent market share and bid efficiency have increased while the variance of average sales prices has been significantly reduced. Additionally, we test for a statistical relationship between agent profits and the bullwhip effect.

1 Introduction

The supply chain management (SCM) game of the Trading Agent Competition (TAC) has provided three years of rich competition among a diverse pool of participants. We seek to evaluate progress and changes in the field of agents, and employ a variety of measures for this evaluation. These include measures of both social welfare and individual performance. We also raise the issue of bullwhip effects in the TAC SCM game, since this is a commonly discussed phenomenon in other supply chain settings. Our analysis is complicated by the strategic interactions that played out in the component procurement markets on day 0 during the early years of competition, and the subsequent changes to the game specification. We will discuss the possible effects of these changes at relevant points in the discussion.

In Section 3 we discuss a method for calculating market efficiency and the division of surplus in the SCM game. Using this measure of efficiency as a benchmark, we compare agent performance between rounds in each of the three tournaments and note interesting trends between tournaments. Section 4 gives results comparing several measures of sales performance in the customer market. Since the specification of the customer market has changed less drastically than the supply market, comparisons across tournaments on these metrics hold more weight. In Section 5 we apply a basic measure of the bullwhip effect to the SCM market and consider the relationship of this measure with market efficiency and the division of surplus. We conclude with a summary of the results and discussion of the usefulness of the various measures.

2 TAC SCM Game

The TAC supply chain management game ([1], [2], [3]) consists of six manufacturing agents competing simultaneously in two separate markets over a period of 220 simulated days to assemble and sell PCs to customers. The manufacturing agents attempt to procure processors, motherboards, memory, and hard drives from eight suppliers at low cost, while assembling and selling 16 different types of PCs to customers at a high price. Each agent is assigned an identical factory for production. Factories have a limited number of cycles each day for production. PC types vary in the number of cycles necessary for production.

Supplier prices are determined by available capacity and inventory levels. The available capacity is driven by a mean reverting random walk, while inventory is determined by past excess capacity and orders already on the books. In 2005, a reputation component was added to the supplier pricing equation. An agent's reputation determines the order in which the agent's request for quotes will be considered. Lower reputations result in higher price offers. The agent's reputation is assigned by considering the ratio of requested amounts to actual amounts purchased. If that ratio is within some acceptable range (acceptable purchase ratio) the agent is designated as having a perfect reputation.

Daily customer demand is Poisson distributed about a mean for each market segment. The mean evolves according to a random walk with an evolving trend parameter. Each customer request for quote contains the requested PC type, quantity, due date, reserve price, and penalty amount. An agent receives an order if the agent issues the lowest bid that meets the reserve price. Agents pay the assigned daily penalty if they faily to deliver by the due date.

3 Market Efficiency and Power

In order to gauge the aggregate agent behavior in the TAC SCM market, we present market power and efficiency benchmarks for evaluating and analyzing the entire economy. We generally expect market efficiency to increase during each subsequent tournament. However, as noted in [4], market efficiency is not a direct measure of agent progress and competent agents may not necessarily maximize market efficiency. We also show the results of analyzing market power (distribution of surplus). The market power held by manufacturers tends to decrease in subsequent tournaments and across rounds.

3.1 Calculating Market Efficiency

We calculate market efficiency by comparing TAC SCM market revenue to the revenue generated by a central planner using complete knowledge of supplier capacity and customer demand. We assume that the supplier value of components is the discounted base price, $P_c^{base}(1-\delta)^1$, and that the customer value of a PC in an order is the reserve price of the request for quote. Therefore the revenue extracted from the market for a single PC is the customer value minus the suppliers' cost of components, penalties, and storage costs. On any day d, a PC given in a customer RFQ may be produced from the available supplier components of days $2 \ldots d-3$, denoted C_d, and the available factory capacity on days $3 \ldots d-2$, denoted F_d. Therefore the constraints on a day d accepted RFQ set A_d are

i. The set of components $Components(A_d)$ required by A_d must be contained in the available component set C_d:

$$Components(A_d) \subseteq C_d(A_4, \ldots, A_{d-1}).$$

ii. The factory cycles $Factory(A_d)$ required to produce A_d must be contained in the available factory cycles set F_d:

$$Factory(A_d) \subseteq F_d(A_4, \ldots, A_{d-1}).$$

Therefore, we would maximize *social welfare* by finding the RFQ acceptance set A which solves the following problem

$$\begin{aligned} \max_{\{A_d \;:\; A_d \subseteq RFQ_d\}} \quad & \sum_{d \in Days} \mathcal{V}(A_d) - \mathcal{C}(A_d) \\ s.t. \quad \forall d \in Days \quad & Components(A_d) \subseteq C_d \\ & Factory(A_d) \subseteq F_d \end{aligned} \tag{1}$$

where $\mathcal{V}(\cdot)$ is the customer value and $\mathcal{C}(\cdot)$ is the supplier component cost plus any storage costs and penalties. This amounts to a large mixed integer linear programming problem. For any given day, the *central planner* may use the current day's component availability and factory capacity as well as the remaining aggregate available components and factory capacity. In doing this we actually allow component availability and factory capacity to be scheduled for early production in order to fill orders that otherwise would not have been obtainable. It is typical behavior for agents in the tournament to procure components in the expectation of customer sales. In our implementation we make a simplification that allows for the possibly infeasible case where aggregate factory capacity is used to build PCs from aggregate component availability that is not yet technically available. For example, consider an acceptance set decision for day 30. An infeasible production state would occur if day 10 had available factory capacity that was used to manufacture PCs with components available on day 20 due for delivery on day 30. As mentioned later in the results regarding the *naive central planner*, shifting component availability and factor capacity accounts for less

[1] $P_c^{base}(1-\delta)$ is the lower bound on component price given in [3]

than 9% of market surplus. Given the current (2005) settings, factory capacity is usually the binding constraint and therefore we feel justified in considering the possibility and additionally the effect of generating infeasible production states negligible. A typical example of a game in the TAC 2005 finals contains over 35,000 values (potential orders) and 2,500 constraints (aggregate component availability and factor capacity), which is not computationally feasible for a standard solver. In order to approximate an optimal central planner, we create a *greedy central planner* that seeks to maximize Equation 1 given a subset of the game's potential orders subject to the constraints. For instance, suppose that we make a greedy choice for every simulation day, the central planner solves 220 instances of the 0-1 multiple knapsack problem in approximately 100 variables and 11 constraints.

3.2 0-1 Multiple Knapsack Solver

We test two different types of solvers for the 0-1 multiple knapsack problem (MKP). Both use a genetic algorithm to search for optimal solutions. The first solver, which we denote the direct search genetic algorithm (DSGA), searches the variable domain directly. The second is a variant of the hybrid genetic algorithm developed in [5], which we denote the meta-search genetic algorithm (MSGA). The MSGA uses a construction heuristic to search over perturbed problem spaces. Both genetic algorithms are tested using a population size of 100 with mutation rate $1/n$ where n is the length of the chromosome, two point crossover, and tournament selection with tournament size 4.

In DSGA, the objective function we use allows infeasible assignments. When a constraint is violated a set of *active* objects that contribute to the constraint are ordered by value. The *active* object with the minimum profit is deactivated and its value subtracted from the objective function. This process continues until all constraints are satisfied. The fitness of the remaining object set is then computed according to Equation 1. This treatment of infeasible strings is similar to [6] but differs in that the penalty term is added per constraint violation, which amounts to a smaller jump in fitness. We initialize the population by generating object sets such that each set is either infeasible or exactly satisfies some constraint. Each chromosome is created in the initial population by starting with all objects *inactive* and then sequentially activating until the chromosome is infeasible, exactly satisfies at some constraint, or is fully active.

We chose to compare DSGA and MSGA on a subset of the problems used by [5]. This also served as a validation of our implementation of MSGA. While DSGA was competitive with MSGA, MSGA did have a higher mean score in every problem instance. We chose to use our implementation of DSGA for the greedy central planner due to its significant superiority in execution speed[2].

[2] On both the Java VM implementations for OS X and FreeBSD DSGA ran an order of magnitude faster than MSGA.

3.3 Market Efficiency Comparison

We consider changes in market efficiency both between years and rounds within the same tournament. These results are shown in Table 1, while the p-values for intratournament comparison are shown in Table 2. The comparisons between tournaments are presented with the caveat that specification changes prevent us from making definative judgements.

In 2005 there was a statistically significant improvement from semifinal to finals play while the improvement in agent efficiency from quarterfinals to semifinals was statistically insignificant. This is consistent with the understanding that teams are improving agents across rounds, and weaker agents are eliminated. This was quite different from the scenario in 2004. While agents exhibited a significant improvement in the quarterfinals to semifinals, this trend did not continue into the finals. Analysis by [7] identified the blocking strategy employed by *FreeAgent*, which barred agents from procuring non-CPU components for a substantial part of the game. This behavior, while competitive, seems to have resulted in an overall loss in market efficiency.

Table 1. Intratournament Market Efficiency

	QF	SF	F
2003	57.0%	56.3%	60.1%
2004	64.6%	73.1%	54.3%
2005	84.1%	84.8%	87.7%

QF - Quarterfinals, SF - Semifinals, F - Finals

Table 2. p-value of Intratournament Market Efficiency Comparisons

	2003			2004			2005		
	QF	SF	F	QF	SF	F	QF	SF	F
QF	1	0.42	0.13	1	2.8e-3	9.6e-4	1	0.21	2.8e-5
SF		1	0.12		1	1.8e-9		1	1.9e-3
F			1			1			1

In order to better understand the measure of market efficiency we tested two benchmarks which give a relative comparison for the TAC SCM 2005 tournament. The first benchmark used a set of six *Dummy* agents provided in the SCM server software. Given the 2005 configuration, this set of agents had a mean market efficiency of 54.4%. The second benchmark considers only the factory and component capacity for a given day, instead of the aggregate prior respective values, as the constraints on production. Using this approach the *naive central planner* achieves an average market efficiency of 91.2%.

3.4 Market Power Comparison

Given the supplier and customer utility defined above we analyze the distribution of surplus within the TAC SCM market over the prior years' tournaments. The market power distribution is given in Table 3 with the corresponding significances given in Table 4. We expect that a more competent field of agents will have lower aggregate market power. The 2004 tournament has a statistically significant decrease in market power from the semifinal to final rounds, while the 2005 tournament shows a decrease in the transition from quarterfinals to semifinals.

Table 3. Tournament Market Power Distribution

	QF			SF			F		
Year	S	M	C	S	M	C	S	M	C
2003	14%	39%	47%	22%	23%	55%	28%	11%	61%
2004	21%	32%	47%	16%	33%	51%	38%	9%	53%
2005	36%	13%	51%	48%	2%	50%	46%	4%	50%

S - Supplier, M - Manufacturer, C - Customer

Table 4. *p*-value of Intratournament Manufacturer Market Power Comparisons

	2003			2004			2005		
	QF	SF	F	QF	SF	F	QF	SF	F
QF	1	0.26	0.04	1	0.34	1.5e-4	1	1.4e-4	2.2e-3
SF		1	0.31		1	1.5e-5		1	0.19
F			1			1			1

4 Customer Market

One of our goals is to develop useful metrics for evaluating agent play in addition to overall profits. In this section we focus specifically on the customer market, presenting three different benchmarks that provide information about agent interactions in this market. These are *market share*, *average selling price* (ASP), and *bid efficiency*. One motivation for specific analysis of the customer market is that the structure of this market has remained relatively constant, allowing for somewhat better intertournament comparisons. There were some changes after 2003, with modifications to the demand evolution process and separation of the market into three distinct segments. In addition, we note that changes in the supplier market can have an effect on customer market measures.

4.1 Market Share

Market share numbers are presented in Table 5. The most obvious feature is that unbid market share dropped dramatically in 2005. Figure 1 shows a breakdown of the total satisfied fraction of the market by simulation day, averaged over all finals games in the three tournaments. In 2003 and 2004 the fraction of the market satisfied early in the game is very low compared to 2005. The difference is attributable to changes in the day-0 purchasing behavior. In 2003 and 2004 the large orders placed on day 0 prevented agents from getting a full complement of components necessary for production until many days into the game in most cases. They were willing to make this tradeoff because of the very low prices for these components. Another explanation for the higher fraction of the market satisfied in 2005 is the increase in nominal supplier capacity from 500 to 550 components per line per day. This reduces prices and instances where production is not feasible due to supplier capacity constraints. Both of these are instances where changes to the supplier market had a clear effect on behaviors in the customer market.

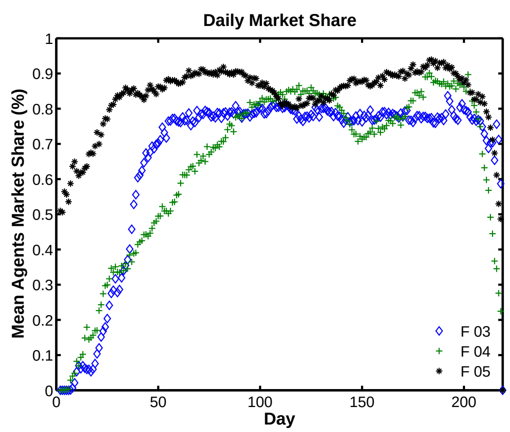

Fig. 1. Mean Daily Market Share
The figure shows the mean aggregate agent market share over the 2003, 2004, & 2005 finals

Another interesting feature of the data may be less influenced by the changes in the supplier market. The variability in agent's market share decreased noticable in 2005, and particularly in the final round. This is an indicator of increased parity between the agents and possibly greater consistency in dealing with a variety of market conditions.

Table 5. Market Share per Agent

Year		Mean	Std. Dev.	Unbid
	QF	8.9%	2.8%	47.3%
2003	SF	8.3%	2.4%	50.2%
	F	9.6%	2.2%	42.2%
	QF	10.8%	1.2%	35.0%
2004	SF	10.1%	3.1%	39.4%
	F	10.1%	0.9%	39.7%
	QF	13.4%	1.1%	19.6%
2005	SF	13.6%	1.1%	18.4%
	F	13.5%	0.8%	19.0%

4.2 Average Sales Price

We now consider the average selling price (ASP) for PCs. All prices are normalized by the base price. We look particularly at the standard deviation of these values. In earlier tournaments, agents were able to corner markets for significant periods of time because of the strategic day-0 procurement issues. This resulted in a decreasing trend in customer ASPs over the course of the game, identified in [7]. While there is still a benefit to early production, this effect has been attenuated by the specification changes. This is noticable in the significant drops in the standard deviation of ASP in subsequent tournaments.

Table 6. TAC SCM Tournament ASP

	2003		2004		2005	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
QF	77.1%	30.1%	82.3%	21.2%	77.4%	15.4%
SF	76.2%	32.0%	80.2%	21.0%	78.3%	15.3%
F	78.2%	26.8%	84.0%	20.6%	78.5%	15.2%

4.3 Bid Efficiency

We define *bid efficiency* to be the ratio of actual revenue achieved by bids and the maximum possible revenue that could have been achieved by bidding on the same set of requests, but with perfect information about opponents' bids. In other words, this is the winning price divided by either the first losing bid or the reserve price if there was no other bid. This is an interesting measure because we can directly compare at least one aspect of agent behavior against the optimal behavior.

Table 7 shows the evolution of agents' bidding efficiency over the competition's three year history. All comparisons are statistically significant. Note that the standard deviation of agent bid efficiency in the 2005 tournament finals is less than 5.2% in comparison to 14.2% and 11.8% in the 2003 and 2004 tournament finals, respectively. One possible explanation for this is increased stability in the customer market prices, which would make it somewhat easier for agents to make bids close to the winning price levels. In general, increases in bidding efficiency reflect improvements in an agent's ability to predict the winning price levels. Bidding efficiency is very high in the 2005 finals, so it is likely to be difficult to gain much advantage from improving this aspect of bidding performance in future agent designs.

Table 7. TAC SCM Tournament Bid Efficiency

	2003		2004		2005	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
QF	88.7%	15.6%	92.6%	11.0%	95.7%	6.7%
SF	87.9%	16.2%	93.8%	9.3%	97.1%	4.5%
F	89.0%	14.2%	94.0%	11.8%	97.1%	5.2%

5 Bullwhip Effect

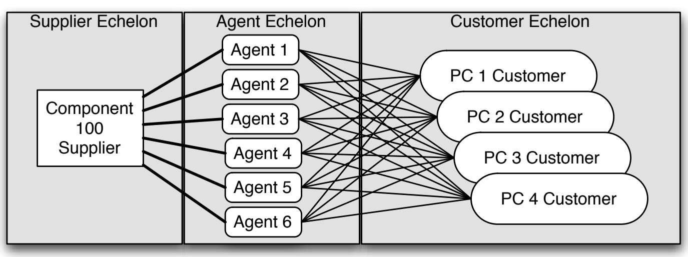

Fig. 2. Component 100 Processor Supply Web

The *bullwhip effect* is a widely known and commonly studied issue in supply chain scenarios [8]. The effect was made famous by the MIT beer game, which is often used to demonstrate the idea in business schools. The basic problem is that volatility in demand (orders) is amplified as it moves up the supply chain, making planning more difficult for entities further from end demand. The SCM scenario is a supply chain with three echelons: customers, manufacturers, and suppliers. Since it is a supply chain, we might expect that it would exhibit the bullwhip effect. We looked for evidence of this by comparing the relative demand variability in the Supplier-Manufacturer echelon and the Manufacturer-Customer echelon.

For a specific supply web, we define the component demand C and the customer demand Q. An example of the supply web formed by component 100 and PCs 1, 2, 3, and 4 is shown in Figure 2. We can then quantify the bullwhip effect as the ratio of the component demand standard deviation to customer demand standard deviation[9]:

$$\omega = \frac{\sigma[C]}{\sigma[Q]} \tag{2}$$

The customer demand Q is Poisson distributed with an evolving trend parameter[3]. TAC agents must plan their component procurement according to this evolution in order to effectively manage the supply chain. Figure 3 shows these signals for the component 100 supply web of game 3717 in the 2005 finals. The bullwhip measure ω for this supply web is 1.37. Any measure greater than 1 indicates a possible instance of the bullwhip effect.

We measured the bullwhip effect for all of the 2005 finals games. The mean and standard deviation of ω for the finals are 2.35 and 0.36, respectively. In addition, we measured the correlation between Q and C, denoted $\rho_{Q,C}$. For the finals the mean and standard deviation for this correlation are 0.19 and 0.09, respectively. We performed linear regressions of the base bullwhip metric ω and the correlation measure against market efficiency and the division of surplus. The results are shown in Table 8. The bullwhip measure had very little predictive power on any of the market measures, though there seems to be some relationship with the correlation measure.

We also considered bullwhip effects for individual agents. Figure 4 gives an example of the evolution of the C and Q signals for the six agents during a game. We regressed each agent's demand adjusted profits (DAP, calculated similarly to [10]) against ω and $\rho_{Q,C}$. The R^2 value for the ω regression is 3.8% and the R^2 value for the $\rho_{Q,C}$ regression is 17.1%, both with positive coefficients. The ω measure has very little explanatory power for agent scores, but well correlated procurement behavior may have positive benefits.

Table 8. Bullwhip Regression R^2

	ω Coefficient	ω R^2	$\rho_{Q,C}$ Coefficient	$\rho_{Q,C}$ R^2
Market Efficiency	0.086	1.7%	-0.089	13%
Supplier Power	0.014	1.47%	-0.199	21%
Manufacturer Power	0.017	0.8%	-0.215	8.4%
Customer Power	-0.03	2.2%	0.414	27%

In general, our current measures of the bullwhip effect do not seem to yield useful insights in the SCM domain. This does not mean that the bullwhip effect is not relevant. It may well be that our simple measures are not adequate to capture the complex dynamics of the SCM environment. Another issue is that the upstream agents (suppliers agents) in the SCM game are not very adaptive

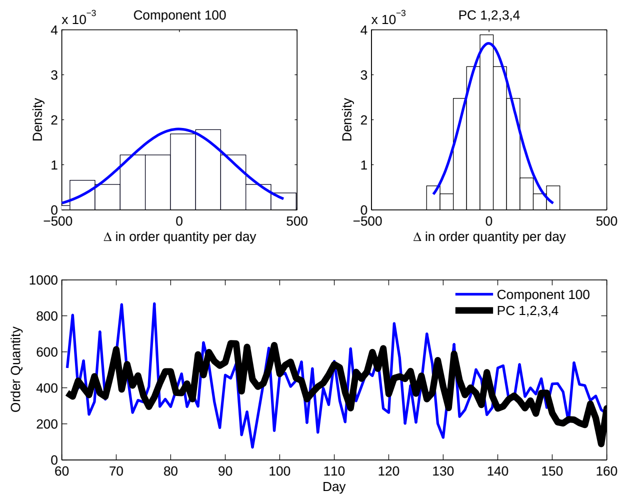

Fig. 3. Supply web signal for component 100 and PCs 1,2,3, and 4 in game 3717. The top-left plot shows the distribution of the change in order quantity per day of component 100. Similarly, the top-right plot shows the distribution of the change in order quantity per day of the PCs corresponding to component 100. These distributions are formed from the component and PC signals shown in the lower plot. Notice that the component, which is in a higher echelon, has a distribution with a significantly larger variance, signifying that there is a *bullwhip effect* in this supply web.

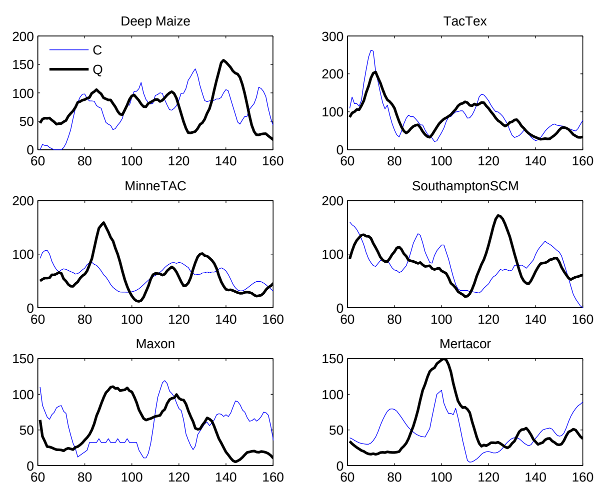

Fig. 4. Agent breakdown of supply web signal

Table 9. Finals' Agent Bullwhip

Agent	Mean	Std. Dev.
Deep Maize	2.69	0.24
TacTex	2.88	0.53
MinneTAC	1.41	0.16
Southampton	2.30	0.32
Maxon	2.52	0.77
Mertacor	1.23	0.15

to changing demand conditions. This may be one reason that bullwhip effects are muted, particularly in how they effect overall measures of market efficiency. However, suppliers do have some behavior modifications based on demand, notably the behavior of producing ahead for future committments with unused capacity. This should be sufficient to see effects from better demand visibility under certain circumstances. We believe that the bullwhip effect is an interesting area for further investigation in the TAC/SCM game, and may shed new light on existing supply chain literature.

6 Conclusion

We have examined a variety of measures of agent performance and overall market efficiency in the TAC SCM tournaments. These measures provide indirect evidence that agent competency is increasing, but it is difficult to separate out the effects of the changes in the game specification. The effects of these changes are pronounced in some of the measures. Overall market efficiency has increased in successive tournaments, while the market power held by manufacturing agents has decreased markedly. This is consistent with increases in the competitiveness of these agents. It is also an indication that the specification changes have increased competitive behavior. In the customer market, agents have increased market share while simultaneously increasing their bid efficiency. The variance on both bid efficiency and mean average selling prices has also decreased, effectively compressing the spread of relevant prices in the customer market. The addition of a repository of binary agents and stability in the specification will allow for more direct and extensive comparisons of agent performance in future tournaments. We also believe that there is a great deal to be learned by exploring widely studied phenomena like the bullwhip effect in the context of TAC SCM, and we look forward to further refinements of this analysis.

Acknowledgments

We would like to thank the TAC SCM participants for providing a rich, complex field of agents on which to perform analysis. We would also like to thank Kevin O'Malley for his involvement in design of our agent **Deep Maize**. This research was supported in part by the U.S. National Science Foundation under IGERT grant 0114368 and grant IIS-0205435.

References

1. Arunachalam, R., Sadeh, N.M.: The supply chain trading agent competition. Electronic Commerce Research and Applications **4** (2005) 63–81
2. Arunachalam, R., Eriksson, J., Finne, N., Janson, S., Sadeh, N.: The Supply Chain Management Game for the Trading Agent Competition 2004. Technical Report T2004-09, Swedish Institute of Computer Science (2004)
3. Collins, J., Arunachalam, R., Sadeh, N., Eriksson, J., Finne, N., Janson, S.: The Supply Chain Management Game for the 2005 Trading Agent Competition. CMU-ISRI-04-139 (2004)
4. Wellman, M.P., Chen, S.F., Reeves, D.M., Lochner, K.M.: Trading Agents Competing: Performance, Progress, and Market Effectiveness. IEEE Intelligent Systems **18**(6) (2003) 48–53
5. Cotta, C., Troya, J.: A Hybrid Genetic Algorithm for the 0-1 Multiple Knapsack Problem. In Smith, G., Steele, N., Albrecht, R., eds.: Artificial Neural Nets and Genetic Algorithms 3, Wien New York, Springer-Verlag (1998) 251–255
6. Khuri, S., Bäck, T., Heitkötter, J.: The Zero/One Multiple Knapsack Problem and Genetic Algorithms. In Deaton, E., Oppenheim, D., Urban, J., Berghel, H., eds.: Proc. of the 1994 ACM Symposium of Applied Computation proceedings, ACM Press (1994) 188–193
7. Kiekintveld, C., Vorobeychik, Y., Wellman, M.P.: An Analysis of the 2004 Supply Chain Management Trading Agent Competition. In: IJCAI-05 Workshop on Trading Agent Design and Analysis. (2005)
8. Lee, H.L., Padmanabhan, V., Whang, S.: The Bullwhip Effect in Supply Chains. Sloan Management Review **38**(3) (1997) 93–102
9. Fransoo, J.C., Wouter, M.J.: Measuring the bullwhip effect in the supply chain. Supply Chain Management **5**(2) (2000) 78–89
10. Wellman, M.P., Estelle, J., Singh, S., Vorobeychik, Y., Kiekintveld, C., Soni, V.: Strategic Interactions in a Supply Chain Game. Computational Intelligence **21**(1) (2005) 1–26