

Evolutionary Optimization of ZIP60: A Controlled Explosion in Hyperspace

Dave Cliff¹

Foreign Exchange Complex Risk Group, Deutsche Bank
1 Great Winchester Street, London EC2N 2DB.
dave.cliff@bcs.org

Abstract: *The “ZIP” adaptive trading algorithm has been demonstrated to outperform human traders in experimental studies of continuous double auction (CDA) markets. The original ZIP algorithm requires the values of eight control parameters to be set correctly. A new extension of the ZIP algorithm, called ZIP60, requires the values of 60 parameters to be set correctly. ZIP60 is shown here to produce significantly better results than the original ZIP (called “ZIP8” hereafter), for negligible additional computational costs. A genetic algorithm (GA) is used to search the 60-dimensional ZIP60 parameter space, and it finds parameter vectors that yield ZIP60 traders with mean scores significantly better than those of ZIP8s. This paper shows that the optimizing evolutionary search works best when the GA itself controls the dimensionality of the search-space, so that the search commences in an 8-d space and thereafter the dimensionality of the search-space is gradually increased by the GA until it is exploring a 60-d space. Furthermore, the results from ZIP60 cast some doubt on prior ZIP8 results concerning the evolution of new ‘hybrid’ auction mechanisms that appeared to be better than the CDA.*

1. Introduction

The Zero-Intelligence Plus (ZIP) adaptive automated trading algorithm [6] has been demonstrated to outperform human traders in experimental studies of continuous double auction (CDA) markets populated by mixtures of human and “robot” traders [15]. To successfully populate a market with ZIP traders, the values of eight real-valued control parameters need to be set correctly. While these eight values can of course be set manually, previous papers have demonstrated that this 8-d parameter-value vector can be automatically optimized using a simple genetic algorithm (GA) search to tailor ZIP traders to particular markets, thereby producing results superior to those from ZIP traders with manually-set parameter values [7, 8]. Furthermore, a simple extension of the GA-ZIP approach (i.e., adding a single additional real-valued numeric parameter, its value set by the GA) allows for automated market-mechanism design, and has been demonstrated as a possible way of automatically discovering novel “hybrid” forms of auction mechanism that appear to be more efficient than the CDA [10, 11, 12]. This paper introduces a more sophisticated version of the ZIP algorithm, which is shown to produce significantly better results. The extended variant is known as “ZIP60”, because it requires 60 real-valued control parameters to be set

¹ Author’s current address: School of Electronics & Computer Science, University of Southampton, Southampton SO17 1BJ, England, UK. dc@ecs.soton.ac.uk. Copyright © D.Cliff, 2006.

correctly, and thus the original algorithm is now re-named as “ZIP8”. Manually identifying the correct values for 60 control parameters could be a very laborious task, but it is demonstrated here that an appropriate automatic search or optimization process (such as a GA) can reliably discover good sets of values for the parameters, so long as some care is exercised in controlling a gradual expansion of the dimensionality of the search-space. The GA operating in the 60-dimensional parameter space is shown to produce markets populated by ZIP60 traders with mean scores significantly better than those of ZIP8s. Moreover, the ZIP60 results presented in this paper, while better than ZIP8, show a markedly reduced incidence of cases where the GA also discovers novel hybrid auction mechanisms within which the ZIP traders perform significantly better than when they interact within the fixed CDA mechanism. A plausible conclusion drawn from this is that it indicates that the earlier ZIP8 results (where apparent “improvements” on the CDA were common) were actually consequences of the relative lack of sophistication in the ZIP8 algorithm, rather than consequences of previously-undiscovered weaknesses in the CDA mechanism that the ZIP8 traders were operating within.

In the interests of scientific openness and ease of replicability, the C source-code that was used to generate the ZIP60 results in this paper has been published in a technical report freely available on the web [13].

This paper reports on an ongoing line of research, and there are several open avenues of research that could be pursued to extend or further explore the ideas presented here. In particular, it is important to note that the results in this paper are certainly not intended as an absolute and conclusive demonstration that ZIP60 is superior to all other CDA bidding algorithms, or that the solutions discovered by the GA are optimal in the sense of the GA routinely discovering Nash equilibria in the experimental markets that ZIP60 is studied within here. This paper studies the equilibrating performance of markets that are homogeneously populated with one type of trader-agent, in the style of frequently-cited prior work such as that by Gode & Sunder [20], Cliff [6, 9, 12], Preist & van Tol [30], and Gjerstad & Dickhaut [19]; rather than studying strategic interactions within markets heterogeneously populated by two or more different types of trading algorithms or market mechanisms, such as is exemplified by [38, 39, & 29]. Although the original paper [6] that introduced the ZIP8 algorithm also studied ZIP8’s performance *only* in homogeneously populated markets, nevertheless ZIP8 was subsequently used as a benchmark trading algorithm in numerous studies of strategic interactions between heterogeneous mixes of trading algorithms, performed by several independent groups of researchers. The number of such papers in which ZIP8 (or close derivatives of ZIP8) have been used is fairly large, and the list includes: [15, 38, 39, 23, 29, 40, & 1]. Thus, given that so much prior work exploring strategic interactions and heterogeneous populations has been based on ZIP8, it seems reasonable at least to presume that researchers with an interest in studying heterogeneous marketplaces might find ZIP60 a useful new benchmark, even though this current paper reports only on ZIP60 in homogeneous settings. While the study of ZIP60’s strategic interactions with other CDA bidding algorithms is certainly an important topic of further research, it is beyond the scope of this current paper.

Furthermore, it is worth noting that in pretty much all of the above-cited papers studying strategic interactions between heterogeneous mixtures of bidding algorithms, the results come from experiments in which the nature of the market supply and de-

mand curves are essentially fixed for the duration of each experiment. That is, studies exploring the effects that significant changes to the supply or demand (or both) curves can have on the trading-agent market’s internal dynamics seem pretty rare. Most often, the supply and demand curves in any one trader-agent experiment remain largely the same for the entire duration of that experiment. This seems very curious, given that one commonly-claimed motivation for studying market systems is that mechanisms such as the CDA are interesting because of their ability to quickly and robustly adapt to dynamic and unexpected changes in supply and/or demand; that studies of shock-changes in *human* CDA markets date back as far as Vernon Smith’s seminal 1962 paper [36]; and that such changes are known to occur in real-world markets.² If CDA markets are interesting because they exhibit attractive adaptation to dynamic changes in supply and demand, why is there this affection in the trading-agent literature for studying CDA systems where such changes are largely absent? In contrast, the results reported in this paper all come from experiments in which the marketplaces periodically undergo sudden “shock” changes to the supply and/or demand curves, and where the ZIP60 traders are optimized on the basis of their ability to rapidly and stably adapt to the new market conditions prevailing after each shock-change.

The rest of this paper is structured as follows. Section 2 gives an overview of ZIP traders and of the experimental methods used, including a description of the continuously variable space of auction types. This description is largely identical to the account given in previous papers (e.g., [10, 12]), albeit extended to describe how the new experiments whose results are presented here differ from the previous work. The new ZIP60 results are then presented, analyzed, and discussed in Section 3.

2. Methods

2.1 The original eight-parameter ZIP

The original eight-parameter ZIP trading algorithm was first described fully in a lengthy report [6], which included source-code (in ANSI C) of an example implementation. For the purposes of this paper, a high-level description of the algorithm and its eight key parameters is sufficient. Illustrative C source-code for ZIP60 has been published in [13]. As will be seen in Section 3, there are in fact a family of ZIP algorithms between ZIP8 and ZIP60, and so hereinafter the acronym “ZIP” with no numeric suffix is intended to mean “all ZIP n for $8 \leq n \leq 60$ and beyond”.

ZIP traders deal in arbitrary abstract commodities. Each ZIP trader i is given a private (i.e., secret) limit-price, λ_i , which for a seller is the price below which it must not sell and for a buyer is the price above which it must not buy. If a ZIP trader completes a transaction at its λ_i price then it generates zero utility (“profit” for the sellers or “saving” for the buyers). For this reason, each ZIP trader i maintains a time-varying utility margin $\mu_i(t)$ and generates quote-prices $p_i(t)$ at time t using $p_i(t) = \lambda_i(1 + \mu_i(t))$ for sellers and $p_i(t) = \lambda_i(1 - \mu_i(t))$ for buyers. The “aim” of traders is to maximize their utility over all trades, where utility is the difference between the accepted quote-price and the trader’s λ_i value. Trader i is given an initial value $\mu_i(0)$ (i.e., $\mu_i(t)$ for $t=0$) which is subsequently adapted over time using a simple machine learning technique known as *the Widrow-Hoff rule* which is also used in back-propagation neural networks and in learning classifier systems. This rule has a “learning rate” parameter β_i that governs

² E.g., in high-frequency foreign-exchange price time series, “gap” step-changes in price are not unusual.

the speed of convergence between trader i 's quoted price $p_i(t)$ and the trader's idealized "target" price $\tau_i(t)$. When calculating $\tau_i(t)$, ZIP traders introduce a small random *absolute* perturbation generated from³ $U[0, c_a]$ (this perturbation is positive when increasing $\tau_i(t)$, negative when decreasing) and also a small random *relative* perturbation generated from $U[1-c_r, 1]$ when decreasing $\tau_i(t)$, or from $U[1, 1+c_r]$ when increasing $\tau_i(t)$, where c_a and c_r are global system constants. To smooth over noise in the learning system, there is an additional "momentum" parameter γ_i for each trader (such momentum terms are also common in back-propagation neural networks).

So, adaptation in each ZIP trader i has the following parameters: initial margin $\mu_i(0)$; learning rate β_i ; and momentum term γ_i . In an entire market populated by ZIP traders, values for these three parameters are randomly assigned to each trader via $\mu_i(0)=f_a(\mu_{min}, \mu_{\Delta})$, $\beta_i=f_a(\beta_{min}, \beta_{\Delta})$, and $\gamma_i=f_a(\gamma_{min}, \gamma_{\Delta})$; for $f_a(\alpha, \kappa)=U[\alpha, \alpha+\kappa]$. Hence, to initialize an entire ZIP-trader market, it is necessary to specify values for the six market-initialization parameters μ_{min} , μ_{Δ} , β_{min} , β_{Δ} , γ_{min} , and γ_{Δ} , and for the two system constants c_a and c_r . Thus any set of initialization parameters for a ZIP-trader market exists within an eight-dimensional real space – hence "ZIP8".

Vectors in this 8-space can be considered as "genotypes" in a genetic algorithm (GA), and from an initial population of randomly generated genotypes it is possible to allow a GA to find new genotype vectors that best satisfy an appropriate evaluation function. This is exactly the process that was first introduced in [7, 8]. For the purposes of this paper, we will consider the GA optimizer as a "black box" and leave it largely un-discussed: full details accompany the source-code in [13].

In addition to using the GA to optimize the control parameters for the trader-agents, one more real-valued numeric parameter was introduced in [10–12] to give the GA automated control over the auction mechanism. This market-mechanism parameter is called Q_s and it governs the *exogenously imposed* probability that the next quote in the marketplace will be taken from a seller, so $Q_s=0.0$ is a pure one-sided auction where only buyers can quote (and hence is similar to an English auction); $Q_s=1.0$ is pure one-sided with only sellers quoting (as in a Dutch Flower auction); and $Q_s=0.5$ makes quotes from buyers or sellers equi-probable (as in a CDA). The surprising result reported in [10–12] is that "hybrid" auction mechanisms (such as $Q_s=0.25$) were found by the GA to give the best evaluation scores when the value of Q_s was evolved alongside the values of the eight ZIP control parameters. Experiments where the value of Q_s was under control of the GA are referred to here as "EM" (for "evolving mechanism") experiments, and experiments where the value of Q_s was fixed, typically at the CDA value of 0.5, are referred to as "FM" experiments (for "fixed mechanism").

The fitness of genotypes was evaluated here using the methods described previously [7, 8, 10–12]: one *trial* of a particular genome was performed by initializing a ZIP-trader market from the genome, and then allowing the ZIP traders to operate within the market for a fixed number of trading periods (often colloquially referred to as "days"), with allocations of stock and currency being replenished between each trading period. During each trading period, Smith's [36] α measure (root mean square deviation of transaction prices from the market's theoretical competitive equilibrium price) was monitored, and a weighted average of α was calculated across the days in

³ Here $v=U[x,y]$ denotes a random real value v generated from a uniform distribution over the range $[x,y]$.

the trial, using a method described in more detail in the next section. As the outcome of any one such trial is influenced by stochasticity in the system, the final evaluation score for an individual was calculated as the arithmetic mean of 100 such trials. Note that as *minimal* deviation of transaction prices from the theoretical equilibrium price is desirable, lower scores are better: we aim here to *minimize* the evaluation scores. That is, individuals with lower scores have greater reproductive fitness.

2.2. Previous ZIP8 Results

In [12], results from 32 sets of experiments were published, where each experiment involved sequences built from one or more of four specific market supply and demand schedules. These four schedules are referred to as markets M1, M2, M3, and M4, and are illustrated in [12, 13]. In all four schedules there are 11 buyers and 11 sellers, each empowered to buy/sell one unit of commodity. Market M1 is taken from Smith's seminal 1962 paper [36] on his early experimental economics work, and the remaining three markets are variations on M1. In M2 the slope of the demand curve has been greatly reduced while the slope of the supply curve has been increased only slightly; and in M4 the slope of the supply curve has been greatly reduced while the slope of the demand curve has been increased only slightly. In M3 the slopes of both the supply and demand curves are only slightly steeper than the slopes in M1, yet these minor differences between the supply and demand curves in M1 and M3 can still lead to significant differences in the final best evolved solutions.

The experiments reported in this paper use a method first explored in ZIP8 experiments, involving "shock changes" being inflicted on the market by swapping from one schedule to another partway through the evaluation process. Here, two shocks occurred during each evaluation process (i.e., switching between three schedules). For instance, in one experiment referred to here as M121, the evaluation involved six trading periods ("days") with supply and demand determined by M1, then a sudden change to M2, then six periods/days later a reversion to M1 for a final six periods. The other sets of experiments are similarly named M212, M123, M321, and so on. Each of the three market schedules was used for six "days", so the two-shock trials last for 18 days. As in the previous GA-ZIP work, the evaluation function was a weighted average of Smith's [36] " α " measure of root mean square deviation of transaction prices from the underlying theoretical equilibrium price at the start of the experiment, measured across the six periods for each schedule used: in each trading period p the value α_p was calculated, and the evaluation score was computed as $(1/\sum w_p) \cdot \sum (\alpha_p \cdot w_p)$ for $p=1 \dots 18$ with weights $w_1=1.75$, $w_2=1.5$, $w_3=1.25$, $w_{3 < p < 7}=1.0$, $w_{p > 6}=w_{p-6}$, and $w_{p > 12}=w_{p-12}$.

The process used to compare the EM and FM cases is as follows. In any one experiment, here involving a population of 30 genotypes over 500 generations, in each generation the elite (best-scoring) individual is of most interest, and so the time-series of the elite fitness score for the population is monitored across the 500 generations. These results are non-deterministic: different runs of the GA (with different seed values for its random number generator) will yield different elite trajectories. Examining the results from 50 repetitions of an experiment (varying only the random seed between repetitions) often gives multimodal results, and in all experiments we are interested only in the best elite mode (i.e. the mode with lowest scores), which can be summarized by the mean and standard deviation (s.d.) of the scores within that mode

at each generation: these two values will be referred to here as the *best elite-mode* fitness mean and s.d.. For comparison purposes, in the ZIP8 work reported in [12], similar trajectories of best elite-mode fitness values were recorded from 50 repetitions of the each experiment in fixed-mechanism (FM) conditions, where the value of Q_s was *not* evolved but instead was fixed at the CDA value of $Q_s=0.5$.

The results from 18 dual-shock (triple-schedule) experiments were presented in four separate data-tables in [12], grouped by the nature of the shocks (i.e., the “treatment regime”). Table 3 showed results from experiments where only the demand curve undergoes a major change on each shock (i.e.: M121, M212, M232, M323, M123, and M321). Table 4 showed results from experiments where only the supply curve undergoes a major change on each shock (i.e.: M141, M414, M434, M343, M143, and M341). In Table 5, one of the two shocks involves a major change only to the demand curve while the other shock involves a major change only to the supply curve (i.e.: M432, M234, M412, and M214); and in Table 6 each shock involved a major change to both the supply curve and the demand curve (i.e.: M242 and M424). In this paper, all 18 dual-shock results are shown together in a single graph, but the results appear in table order, as was just listed.

Analysis of the ZIP8 results showed that the GA never failed to discover EM genotypes that were at least as good (i.e. had elite evaluation scores at least as low) as the corresponding FM genotypes, and in several cases the EM result was significantly better (lower) than the FM result, at the 1% confidence level, using appropriate non-parametric significance tests such as the Wilcoxon-Mann-Whitney (see, e.g., [35]), or latterly the Robust Rank Order test [16].

The histogram in Figure 1 shows the results for GA-optimized ZIP8 in FM and EM conditions. Fig.1 also shows the results from various styles of ZIP60 EM experiments, discussed further in Section 3 of this paper. The ZIP8 statistics in Fig.1 are the results of conducting a more rigorous and careful analysis (discussed in [13]) of the data than was originally summarized and tabulated in [12]. The final evaluation score recorded as the outcome of any one experiment is now taken as an average of the final few elite scores (over generations 490 to 500) to smooth over noise in the evaluation process; and the summary statistics for each type of experiment are here always calculated from the top 10% (i.e., the upper decile) of the 50 repetitions of each type of experiment, regardless of how many repetitions converged on solutions with final elite scores in the best elite mode. So, the data in Fig.1 show the mean and s.d. of the final outcome elite scores from the best (lowest-scoring) five experiments in each study.

2.3 Related Work

These previous GA-ZIP results have subsequently been replicated, adapted, and extended in a number of independent studies. Robinson [32] explored the use of evolved market-mechanisms in the context of market-based control (e.g. [4]) of scarce resources in utility-scale corporate data centers. Walia [41] explored the use of the same evolving-mechanism techniques but with markets populated by Gode & Sunder’s [20] ZI trader-agents rather than ZIP traders, again finding evidence that non-standard hybrid mechanisms were discovered as good/best solutions by the GA; and Byde [2] demonstrated that the same techniques could lead to the evolution of hybrid sealed-bid auction mechanisms, regardless of the type of trader operating in the market. Shipp [34] investigated how the nature of the evolved solutions changed as the num-

ber of “market shocks” used in the evaluation process increased; and Wichett [43] explored a system in which multiple reproductively separate “gene-pools” of ZIP traders competed, co-adapted, and co-evolved along with the market mechanism. Other recent uses of ZIP include modifying it for bargaining in sealed-bid auctions [1]; using ZIP traders to study speculative trading in business-to-business exchanges [25]; and using ZIP traders to explore issues of reputation and information quality in a variety of market configurations [24].

The results in [10] were the first demonstration that radically new market mechanisms for artificial traders may be designed by automatic means. But, at much the same time as they were being generated, Steve Phelps and his colleagues were independently working on a conceptually very similar (but algorithmically rather different) theme of using artificial evolution to develop and study new auction-market mechanisms [29]. In addition to the contemporaneous work of Phelps *et al.*, a number of other authors have more recently reported on the results of using artificial evolution and other forms of automated search, learning, or optimization for exploring spaces of possible trader-agent strategies, and possible new auction mechanisms, generally with positive results [39, 18, 26, 28, 21, 27, 31, & 42]. Of course, the paper introducing ZIP [6] was not the first-ever study of artificial trading agents in double-auction markets; notable prior work includes [44], [17], and [33]. Also, [19] was developed independently at much the same time. For additional discussion of earlier work, see [6].

3. ZIP60

3.1 From 8 to 60 in five paragraphs

The results from using a GA to fine-tune the ZIP8 trader were sufficiently encouraging that they provoke the question of whether new variants of ZIP can be developed to take advantage of the fact that we can now (generally, at least) rely on automated optimizers like the GA to set appropriate values for the numeric parameters affecting the traders. If we commit to using an optimizer to set the parameter values, we don’t need to keep the number of parameters small enough for them all to be manageable or comprehensible by humans. That’s the rationale for ZIP60.

To this end, observe that in ZIP8 the genome specifies the same vector of eight real values $\{\mu_{min}, \mu_{\Delta}, \beta_{min}, \beta_{\Delta}, \gamma_{min}, \gamma_{\Delta}, c_a, c_r\}$ whether the trader is a buyer or a seller. But in some situations it’s plausible that the market dynamics might be better if the parameter-values used by the buyers were different to those used by the sellers, so we could in principle have a GA-ZIP system dealing with these two cases (i.e. where *Case 1* is that the trader is a buyer; *Case 2* is that the trader is a seller) and hence optimizing sixteen real parameters (i.e., “ZIP16”), with the first vector of eight values being used to initialize the buyers and the second being used to initialize the sellers.

Next, note that in some situations a ZIP trader (whether it is a buyer or a seller) has to increase its margin, and in others it has to decrease its margin, and that it may be useful to have different parameter-values depending on which of the four cases we are in, i.e. whether the trader is a buyer raising its margin, a buyer lowering its margin, a seller raising, or a seller lowering. That’s 4 cases, each with 8 values, and so “ZIP32”. But we can then additionally note that, in the original specification of the ZIP algorithm, both for buyers and for sellers, there are actually *three* different cases or circumstances in which the trader alters its margin (see [6] pp.42-43 for the details of and rationale for this design). For example, a seller’s margin is *raised* if *one* condition

holds true (i.e., if the last quote was accepted and the seller's current price is less than the price of the current quote); but a seller's margin is *lowered* if either of *two* other possible conditions are true (i.e.: if the last quote was an accepted bid and the seller is active and the seller's price is greater than the price of the last quote; *or* if the last quote was an offer that was accepted and the seller is active and its price is greater than the price of the last quote). So we could have the genome specify *three* corresponding parameter-value vectors for the buyers and also *three* such vectors for the sellers, i.e. a total of six different vectors for six different cases, which at eight values per vector gives us "ZIP48".

And in a final flourish of parameter-count inflation, let's abandon the use of a mere pair of system-wide global constants c_a and c_r and in place initialize each trader i with its own corresponding "personal" values $c_{a,i}$ and $c_{r,i}$ generated at initialization from the uniform distributions $U[c_{a,min}, c_{a,min}+c_{a,\Delta}]$ & $U[c_{r,min}, c_{r,min}+c_{r,\Delta}]$. This addition of extra parameters still allows solutions involving the old system-wide constant c_a and c_r values to be "discovered" by the GA – that will happen if better evaluation scores are associated with (near-)zero values of $c_{a,\Delta}$ and $c_{r,\Delta}$. So, the parameter-value vectors for each case needs now to specify not only the six previous system parameters (μ_{min} , μ_A , β_{min} , β_A , γ_{min} , and γ_A) but also the values for the four newly-introduced system parameters $c_{a,min}$, $c_{a,\Delta}$, $c_{r,min}$, and $c_{r,\Delta}$ – i.e., ten values per vector. For six cases, each with ten values per vector, we get to sixty values: "ZIP60".

It is worth noting that this final increase from eight to ten parameter-values per case could also be applied to any of the other ZIP n versions mentioned in the preceding paragraphs. That is, by the expansion of the specification of c_a and c_r , ZIP8 becomes ZIP10; ZIP16 becomes ZIP20; and ZIP32 becomes ZIP40.

We need also to introduce some terminology that will ease the analysis and discussion that come later. While a ZIP8 trader has one genetically-specified value for each parameter (so, for example, it has only one β_{min} value), a ZIP60 genome specifies six related parameter values – one for each case – which we will refer to by adding case-numbers to the subscript (e.g.: $\beta_{min:1}$, $\beta_{min:2}$, ..., $\beta_{min:6}$). For ZIP60, the entire set of sixty parameters can be generated from the pattern $P_{t:n}$ where P is one of $\{\mu, \beta, \gamma, c_a, c_r\}$; t is one of $\{min, \Delta\}$; and n is an integer in $\{1, \dots, 6\}$. We'll refer to the set of six values for any one parameter-type (i.e., $\{P_{t:1}, P_{t:2}, \dots, P_{t:6}\}$ for some given P and t) as the *homologous set* of P_t parameter values.

Finally, note that the additional computational costs of using ZIP60 as a replacement for ZIP8 are virtually zero. The space costs are those incurred in storing the additional 52 real-valued parameters: this is a large percentage increase, but in absolute terms it is still a very small amount of storage when expressed as actual additional bytes required. The additional time costs are also very low indeed: a tiny amount of extra processing is needed in initialising the ZIP60 trader (i.e., populating its look-up table of 60 real values) and then in doing table look-up while the trader is operating (i.e. choosing the values to use that are appropriate to the current "case"), but that's it.

3.2 ZIP60 Results: Control of Search-Space Dimensionality Required

In testing the performance of ZIP60, all effort thus far has been devoted to exploring the performance of ZIP60 on dual-shock tests: if markets populated by ZIP traders cannot cope with sudden shock-changes in supply and demand, then they are of little

interest. Moreover, it seems highly likely (but has not yet actually been empirically verified) that if ZIP60 does better than ZIP8 on these multi-shock tests, then it will also do better in those cases where there are fewer or no market shocks.

Experience with GA optimisation of ZIP60 indicates that significant care is needed in managing the dimensionality of the search-space: simply applying the old methods that worked well with ZIP8 does *not* give best results when working with ZIP60. This is a lesson learnt from experience: for the very first attempts at evolutionary optimization of ZIP60 traders, the same experiment methods as described in Section 2 were used, except that the initial population was composed entirely of randomly generated ZIP60 individuals, rather than ZIP8s. The results from these attempts were somewhat mixed. Although the scores of the elite evolved ZIP60 traders were on average significantly better than the elite ZIP8 scores in the same experiments, the standard deviation on that average improvement was almost identical to the mean improvement itself. This large standard deviation was a reflection of the fact that, in a few cases, the evolved elite ZIP60 results were actually significantly *worse* than the corresponding ZIP8 results. Now there is nothing preventing the ZIP60 GA system from evolving genotypes that correspond to ZIP8 solutions, so it seems peculiar that the ZIP60s perform worse than the ZIP8s in some cases. There are certainly points within the ZIP60 genome-space that correspond perfectly to ZIP8 solutions: if for each of the ten homologous sets the within-set variance of the parameter values for the set is (near) zero, then that ZIP60 genome is functionally equivalent to the corresponding single-case ZIP10 genome; and furthermore if the values of the $c_{a,\Delta}$ and $c_{r,\Delta}$ homologous sets are all zero, then the ZIP60 is functioning as a ZIP8. So, how come the ZIP60 results are sometimes worse than ZIP8? The fact that the GA failed to find ZIP8 solutions within the ZIP60 genome space is a strong indication that the 60-dimensional search space has characteristics (such as local maxima, sharp ridges, and plateaus in the fitness landscape) which make the search for good genomes a nontrivial process.

To address this, the ZIP genetic encoding was extended, allowing the number of cases (1, 2, 4, or 6, as discussed in Section 3.1) to be specified *on the genome itself*. The rest of the genome is still a set of ten homologous-set vectors (each made of six real numbers). If an individual's gene specifying the number of cases is set to one, then all six parameter-values are set to be identical within each homologous set, by copying the values from the first element of the set into the remaining five. If the number of cases is set to two, then the three buyer-case parameter values within each set are forced to be identical copies of each other, as are the three seller values; and if the number of cases is set to be six, then the three buyer and the three seller parameters can all be different numeric values. Thus, the ZIP60 genomes are always 60 parameter-values long, but over-writing duplication of values within the genome can reduce the *effective* dimensionality of the parameter-vectors encoded on a particular genome so that it codes for *any* of the family of ZIP algorithms between ZIP60 and ZIP8.

The motivating hypothesis for placing the dimensionality of the search-space under evolutionary control was the belief that the GA's evolutionary search would be more successful if it could start by first simply optimizing the 1-case genome, and then (only once all the values are approximately correct) could successive multi-case refinements be progressively introduced by the GA as necessary. So, for example, if a 1-case individual mutated to become a high-case individual, thereby decoupling its genome-values across the different cases, such a mutant would only be retained in the

population if the mutation that increases the number of cases is also associated with higher fitness. Strictly speaking, the initial case-increasing mutation is selectively neutral: the genome values for the different cases start out as identical copies of each other, but the case-increasing mutation allows *subsequent* mutations to introduce differences across cases, and it is those mutations that will be retained if they are correlated with higher fitness. Handing evolutionary control of the dimensionality of a search-space to the GA that is searching that space is an idea that was first explored in depth in Harvey's [22] thesis, where he developed the "species adaptation genetic algorithm", which was first successfully applied in evolving neural-network controllers for autonomous mobile physical robots [5].⁴

Two new sets of ZIP60 experiments were performed to test the effects of GA-controlled dimensionality. In the first set, the population was initialised with individuals that had a randomly-assigned value for the number of cases on their genome, with the values 1, 2, 4, and 6 being equally probable. This is the initialization we refer to here as ZIP60(1:6) (for "from 1 case to 6 cases"). In the second set, every individual in the initial population was set to have a 1-case genome; this is referred to here as the ZIP60(1:1) initialization. And so the first set of experiments, where all individuals in the initial populations were 6-case individuals, are referred to as ZIP60(6:6). Results from the EM experiments with ZIP60 with the (1:1), (1:6), and (6:6) initializations are shown in Figure 1, with ZIP8 EM and FM scores shown alongside, for comparison.

The histogram in Figure 1 shows the mean elite ZIP60 EM scores alongside the ZIP8 EM and FM scores: ZIP60 consistently out-performs ZIP8, and the error bars showing the s.d. values make it clear that these differences are significant. On the average, the ZIP60(1:1) scores are 14.0% better (lower) than the ZIP8 scores (and the s.d. on that mean improvement is 5.7%). In comparison, the ZIP60(6:6) scores are on average 12.91% better than the ZIP8, but the s.d. on that improvement is 12.88%; and for ZIP60(1:6), the average improvement is 12.32% with s.d.=7.03%. So, ZIP60(1:1) has the highest mean increase in performance and the lowest s.d. on its mean increase.

Results from significance analysis of the differences between the ZIP60(1:1) and ZIP60(1:6) upper-decile elite scores for the 18 dual-shock experiment schedules are tabulated in [13], and they offer weakly supportive evidence for the claim that ZIP60(1:1) is a better initialization than ZIP60(1:6). Using the Robust Rank Order test [16] at the 1% significance level reveals that, over the 18 types of experiment, only for schedule M242 does ZIP60(1:1) lead to significantly better results than ZIP60(1:6). In all other cases, no statistically significant difference in the scores is detected. So, ZIP60(1:1) is certainly no worse than ZIP60(1:6), and the evidence thus far is that it is actually significantly better in one of the 18 cases studied. The absence of a huge difference is perhaps no surprise given that a ZIP60(1:1) system will, after sufficiently many generations, be pretty much indistinguishable from a ZIP60(1:6) as mutants with case-values greater than unity are progressively retained in the (1:1)-seeded population.

Examination of the elite genomes across the course of the 500 generations, discussed and illustrated in [13], shows that although the ZIP60(1:1) population starts out composed entirely of 1-case genomes, after a while the number of 2-case, 4-case, and

⁴ A recent paper by Stanley & Miikkulainen [37] re-discovers some of Harvey's [22] ideas of evolutionarily controlled dimensionality increase, which Stanley & Miikkulainen rename as "complexification".

6-case mutant genomes starts to increase, and by the end of each experiment the elite individual is almost always a 6-case genome. ZIP60(1:6)-seeded experiments also virtually always end with 6-case elite genomes.

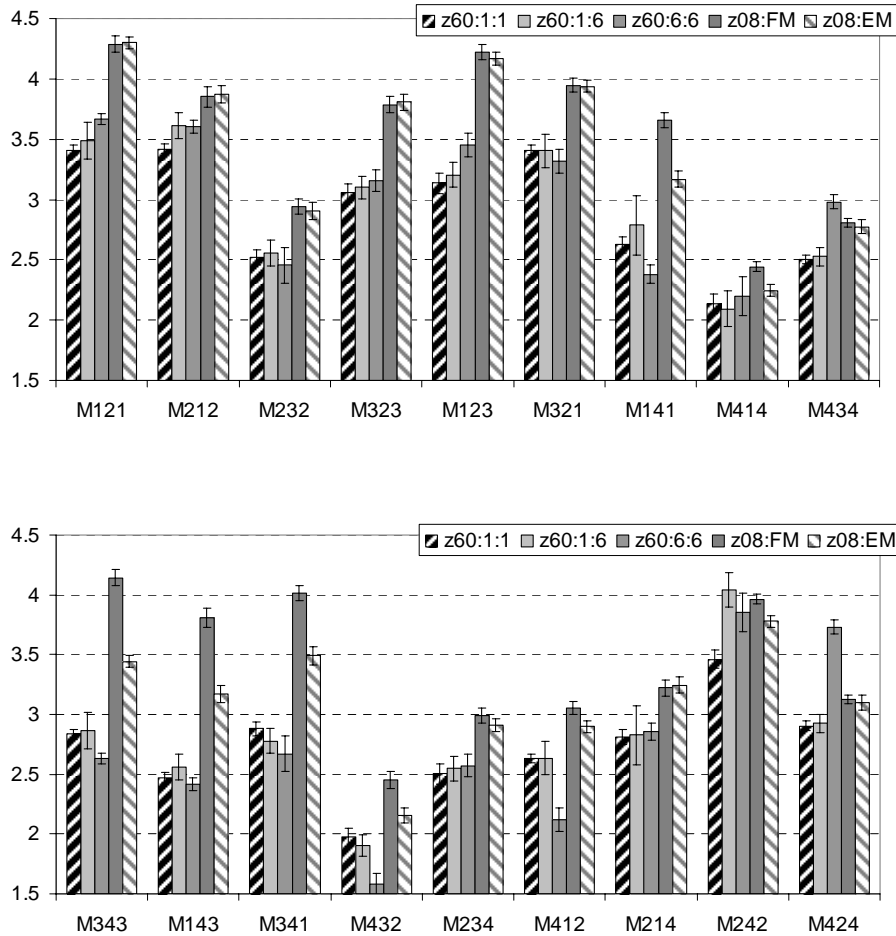


Figure 1: Mean elite outcome scores from the best 10% ($n=5$) of the 50 repetitions of each of the 18 “dual-shock” experiments involving two sudden changes in market supply and demand function, as described in the text. Labels on the horizontal axis indicate the specific shock sequence. Vertical axis is evaluation score: a weighted average of root mean square deviation of transaction prices from the theoretical competitive equilibrium price, expressed as a percentage of the equilibrium price; a metric labeled “ α ” by Smith [36]. Lower scores are better. Each bar in the graph shows a mean score, with error bars at plus and minus one s.d. For each shock-sequence, the cluster of 5 bars shows the results for (from left to right): EM ZIP60(1:1); EM ZIP60(1:6); EM ZIP60(6:6); FM ZIP8, & EM ZIP8. See text for further discussion.

3.3 Principal Component Analysis

The results just presented demonstrate that ZIP60(1:1) most consistently out-performs ZIP8, which strongly suggests that the larger number of additional parameters are indeed useful. However, as was noted above, it is possible for a ZIP60 genome to be functionally equivalent to a lower-dimensioned ZIP n genome. In the most extreme case, if all the values in each homologous set of parameters are equal for any one genome (so, e.g., $\beta_{\Delta 1}=\beta_{\Delta 2}=\beta_{\Delta 3}=\beta_{\Delta 4}=\beta_{\Delta 5}=\beta_{\Delta 6}$), or if the differences between them are all sufficiently small to be ignored as mutational noise, then that ZIP60 genome is functionally equivalent to a ZIP10 genome. And if it has zero values for its $c_{a\Delta}$ and $c_{r\Delta}$ parameters, then it is effectively a ZIP8.

So, to confirm that ZIP60 is indeed an advance on ZIP8 (or ZIP10), some analysis of the final evolved parameter-sets is necessary, to see whether they contain any low-dimensional solutions embedded in higher-dimensional spaces. To this end, principal component analysis (PCA) was used on the parameter-values from the top-decile ZIP60 genomes. PCA is explained in most textbooks on multivariate analysis, e.g. [3]. Each six-dimensional homologous set of final evolved parameter values from all 18 sets of experiments was individually subjected to PCA, and the percentage of the variance in the parameter values accounted for by each principal component (PC) was calculated. If all the values in any one homologous set were equal or approximately equal, the first PC would account for very nearly 100% of the variance. However, the first PC would also account for close to 100% of the variance if the values in the homologous set were positioned along/around *any* line in the 6-D space, e.g. one where $P_{i:t} \neq P_{i:j}$ (for some P in $\{\mu, \beta, \gamma, c_a, c_r\}$; some t in $\{min, \Delta\}$; and for i, j integers in $\{1, \dots, 6\}$ with $i \neq j$). So, to identify a ZIP8/ZIP10 embedded in a ZIP60 genome, we'd need to see the first PC for each homologous set accounting for close to 100% of the variance, *and* see the angle θ between the first PC and the line $P_{t:1}=P_{t:2}=\dots=P_{t:6}$ being very close to zero. That is for $\theta = \pi - |\cos^{-1}((\mathbf{p}_1 \cdot \mathbf{u}) / |\mathbf{p}_1|) - \pi|$ with \mathbf{p}_1 being the first PC (a 6-D vector) and \mathbf{u} being a 6-D unit vector with elements $u_{i:i=1\dots 6}$ such that $u_1=u_2=u_3=u_4=u_5=u_6 > 0$.

PCA analysis was performed on the entire data-set of top-decile elite genomes; and the results are presented graphically and discussed further in [14]. Although \mathbf{p}_1 accounts for more than 50% of the variance in all homologous sets, the highest value is 90.29% for \mathbf{p}_1 of the β_{min} set, which is not high enough to cause alarm. The mean variance accounted for by \mathbf{p}_1 across all homologous sets was 68% (s.d.=10%), and the minimum value was 58%. Also, the angle θ is safely high in all cases (mean=24°; s.d.=17°; min=5°; max=50°). So, the evolved ZIP60s are not ZIP8s in disguise.

3.4 Discussion: Fewer Hybrids?

Comparing the ZIP8 and ZIP60 results presented here reveals that for ZIP60 the GA much less frequently discovers hybrid values of Q_s yielding overall market dynamics that are better than those of the corresponding fixed-market CDA $Q_s=0.5$ experiments. That is, despite the final ZIP60 EM evolved Q_s values varying quite widely, few of them give results that are statistically significantly better than the corresponding FM results. Data tables available in [13] show that in two thirds (12 out of 18) of the original ZIP8 experiments, the EM experiment found a “hybrid” Q_s value that improved on the corresponding FM score; yet in the ZIP60 experiments, the occurrence of superior EM results fell by 67%, i.e. from 12/18 down to 4/18. This could be an

indication that the previously-published results showing evolved hybrid auction mechanisms are to some extent artifacts of the lack of sophistication in the ZIP8 traders that were used in those studies. A counterargument to this is that Bye [2] presented results from applying similar GA-search for designs for hybrid sealed-bid auctions, where the GA found hybrid solutions to be preferable to the traditional first-price and second-price sealed bid auctions and those results were *independent* of the sophistication of the traders in the market. Clearly this is another issue that should be explored in more depth in future research.

4. Conclusions

From the data summarized and analyzed in this paper, it is clear that the ZIP60 variant of ZIP is a genuine improvement on the original ZIP8, and that ZIP60 parameter-vectors that outperform ZIP8 by over 10% can be found by a search/optimization process such as the simple GA used here, provided that care is taken in the progressive expansion of the dimensionality of the search-space explored by that GA. Principal component analysis of the elite evolved parameter-sets from multiple runs under differently-changing sequences of supply and demand schedules revealed that the evolved parameter-vectors make active use of considerably more values than the eight available in ZIP8. The fact that (in comparison to previous experiments using ZIP8 traders) the experiments with ZIP60 traders reported here show a reduced incidence of the discovery of “hybrid” auction mechanisms is possibly an indication that the hybrid auctions reported on in the *E-Commerce Research and Applications* journal paper [12] actually evolved as a consequence of the lack of sophistication in the behavior of ZIP8 traders: with the comparatively finer-grained responses of ZIP60 traders, hybrid mechanisms evolve much less frequently, and so it is tempting to conjecture that if the same type of auction-design experiments were repeated with even more sophisticated trader agents, hybrid mechanisms would not occur at all. Exploring that question remains one of several topics for further research.

References

- [1] Bagnall, A. & Toft, I. (2005) “Autonomous Adaptive Agents for Single-Seller Sealed Bid Auctions”, *Autonomous Agents & Multi Agent Systems*. In press.
- [2] Bye, A. (2003), “Applying Evolutionary Game Theory to Auction Mechanism Design”. *Proc. 2003 ACM Conf. on E-Commerce*. Also available as Hewlett-Packard Laboratories Technical Report HPL-2002-321.
- [3] Chatfield, C. & Collins, A. (1980), *An Introduction to Multivariate Analysis*. Chapman and Hall.
- [4] Clearwater, S., ed. (1995), *Market-Based Control*. World Scientific Press.
- [5] Cliff, D., Harvey, I., & Husbands, P. (1993), “Explorations in Evolutionary Robotics”, *Adaptive Behavior*, 2:73–110.
- [6] Cliff, D. (1997), “Minimal-intelligence agents for bargaining behaviours in market environments”. Hewlett-Packard Laboratories Technical Report HPL-97-91.
- [7] Cliff, D. (1998a), “Genetic optimization of adaptive trading agents for double-auction markets” in *Proceedings of Computational Intelligence in Financial Engineering (CIFEr98), New York, IEEE/ IAFE/ Informs* (preprint proceedings), pp.252–258, Apr 1998.
- [8] Cliff, D. (1998b), “Evolutionary optimization of parameter sets for adaptive software-agent traders in continuous double-auction markets”. Presented at the *Artificial Societies & Computational Markets (ASCMA98)* workshop at 2nd *International Conference on Autonomous Agents*, May 1998. Also available as HP Labs Technical Report HPL-2001-99.
- [9] Cliff, D. & Bruten, J. (1999), “Animat Market-Trading as Collective Social Adaptive Behavior”. *Adaptive Behavior* 7(3&4):385–414.
- [10] Cliff, D. (2002a), “Evolution of market mechanism through a continuous space of auction-types”. Presented at *Computational Intelligence in Financial Engineering (CIFEr)*, Hawaii, May 2002.

- [11] Cliff, D. (2002b), "Evolution of market mechanism through a continuous space of auction-types II: Two-sided auction mechanisms evolve in response to market shocks". In: H. Arabnia, & Y. Mun. (eds) *Proc. Int. Conf. Internet Computing IC02*, Vol. III, CSREA Press, pp.682–688.
- [12] Cliff, D. (2003), Explorations in evolutionary design of online auction market mechanisms. *Electronic Commerce Research & Applications Journal*, 2(2):162–175, 2003.
- [13] Cliff, D. (2005), *ZIP60: Further Explorations in the Evolutionary Design of Online Auction Market Mechanisms*. Hewlett-Packard Laboratories Technical Report HPL-2005-85.
- [14] Cliff, D. (2006), *ZIP60: an enhanced extension of the ZIP trading algorithm*. Submitted for review.
- [15] Das, R., Hanson, J., Kephart, J., & Tesauro, G. (2001), "Agent-human interactions in the continuous double auction" *Proc. of the Int. Joint Conference on Artificial Intelligence (IJCAI-01)*.
- [16] Feltovich, N. (2003), "Nonparametric tests of differences in medians: comparison of the Wilcoxon-Mann-Whitney and robust rank-order tests", *Experimental Economics*, 6:273–297.
- [17] Friedman, D. (1991), "A Simple Testable Model of Price Formation in the Double Auction Market", *Journal of Economic Behavior & Organization* 15:47-70.
- [18] Gerding, E., Somefun, K., & La Poutré, H. (2004), "Multi-Attribute Bilateral Bargaining in a One-to-Many Setting", In *Proc. Workshop on Agent Mediated Electronic Commerce VI (AMEC-04)*.
- [19] Gjerstad, S. & Dickhaut, J. (1998), "Price Formation in Double Auctions", *Games & Economic Behavior*, 22:1–29.
- [20] Gode, D. & Sunder, S. (1993), "Allocative efficiency of markets with zero-intelligence traders", *Journal of Political Economy*, 101:119–137.
- [21] Greenwald, A., Guillemette, B., Naroditskiy, V., & Tschantz, M., (2005), "Scaling Up the Sample Average Approximation Method for Stochastic Optimization with Applications to Trading Agents" in Jansen, S. (ed) *Notes IJCAI-05 Workshop on Trading Agent Design & Analysis (TADA05)*.pp.21–27.
- [22] Harvey, I. (1994), *The Artificial Evolution of Adaptive Behavior*. PhD Thesis, School of Cognitive and Computing Sciences, University of Sussex, U.K.
- [23] He, M., Leung, H., & Jennings, N. (2003), "A fuzzy logic based bidding strategy for autonomous agents in continuous double auctions" *IEEE Transactions on Knowledge and Data Engineering* 15(6):1345–1363.
- [24] Ladley, D. & Bullock, S. (2005), "Who to listen to: Exploiting information quality in a ZIP-agent market" in Jansen, S. (ed) *Notes IJCAI-05 Workshop on Trading Agent Design & Analysis (TADA05)*. pp.28-34.
- [25] Li, L. & Smith, S. (2004), "Speculation Agents for Dynamic, Multi-period Continuous Double Auctions in B2B Exchanges", *Proc. 37th Hawaii International Conference on System Science*, Hawaii, January 2004.
- [26] Lochner, K. & Wellman, M. (2004), "Rule-Based Specification of Auction Mechanisms." *Proc. Third International Joint Conference on Autonomous Agents and Multiagent Systems*. pp.818-825.
- [27] Pardoe, D. & Stone, P. (2005), "Developing Adaptive Auction Mechanisms", *ACM SIGecom Exch.*, 5(3):1-10.
- [28] Park, S., Durfee, E., & Birmingham, W. (2004), "Use of Markov Chains to Design an Agent Bidding Strategy for Continuous Double Auctions", *Journal of Artificial Intelligence Research (JAIR)*, 22: 175–214.
- [29] Phelps, S., Parsons, S., & McBurney, P. (2004), "An Evolutionary Game-theoretic Comparison of two Double Auction Market Designs". *Proceedings 4th International Workshop on Agent-Mediated E-Commerce (AMEC-IV)*.
- [30] Preist, C. & van Tol, M. (1998), "Adaptive Agents in a Persistent-Shout Double Auction". *Proceedings of ICE-98*, ACM.
- [31] Reeves, D., Wellman, M., Mackie-Mason, J., & Osepayshvili, A. (2005), "Exploring Bidding Strategies for Market-Based Scheduling", *Decision Support Systems*, 39:67–85.
- [32] Robinson, N. (2002), *Evolutionary Optimization of Market-Based Control Systems for Resource Allocation in Compute Farms*. MSc Thesis, School of Cognitive & Computing Sciences, University of Sussex, U.K.
- [33] Rust, J., Miller, J., & Palmer, R. (1992), "Behavior of trading automata in a computerized double auction market" in Friedman, D., & Rust, J. (eds) *The Double Auction Market: Institutions, Theories, and Evidence*, pp.155-198. Addison-Wesley;
- [34] Shipp, D. (2004), *The effects of changes to supply and demand on trader agents and marketplaces*. MSc Thesis, School of Computing, University of Leeds, UK.
- [35] Siegel, S., & Castellan, N. (1988), *Nonparametric Statistics for the Behavioral Sciences*, McGraw Hill.
- [36] Smith, V. (1962), "Experimental study of competitive market behavior" *Journal of Political Economy*, 70:111–137.

- [37] Stanley, K. & Miikkulainen, R. (2004), "Competitive Coevolution through Evolutionary Complexification", *Journal of Artificial Intelligence Research (JAIR)*, **21**:63–100.
- [38] Tesauro, G. & Das, R. (2001), "High-Performance Bidding Agents for the Continuous Double Auction", *Economic Agents, Models, & Mechanisms Workshop, International Joint Conference on Artificial Intelligence (IJCAI-01)*.
- [39] Tesauro, G. & Bredin, J. (2002), "Strategic Sequential Bidding in Auctions using Dynamic Programming", *Proc First Int. Conf. Autonomous Agents & Multiagent Systems (AAMAS-02)*.
- [40] Vytelingum, P., Dash, R., David, E., & Jennings, N. (2004), "A risk-based bidding strategy for continuous double auctions" *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI04)*, Valencia, Spain, pp.79-83.
- [41] Walia, V. (2002), *Evolving Market Design*, MSc Thesis, Department of Computer Science, University of Birmingham, UK.
- [42] Wellman, M., Reeves, D., Lochner, K., & Suri, R. (2005), "Searching for Walverine 2005" in Jansen, S. (ed.) *Notes IJCAI-05 Workshop on Trading Agent Design & Analysis (TADA05)* pp.1–6.
- [43] Wichett, D. (2004), *Coadaptive Dynamics of Minimal Intelligence Trading Agents*. MSc Thesis, Department of Computer Science., University of Birmingham, UK.
- [44] Wilson, R. (1987), "On Equilibria of Bid-Ask Markets" in Feiwel, G. (ed.) *Arrow & the Ascent of Modern Economic Theory*, pp.375-414. New York University Press.