Does it matter to play Nash: the case of adaptive sellers ?

Eric DARMON [a], Roger WALDECK [b] ♣

[a] GREDEG-CNRS and Université de Nice Sophia-Antipolis; 250, avenue Albert Einstein, F-06560 Valbonne, France
[b] GET ENST-Bretagne, Department LUSSI; Technopole Brest Iroise CS 83818, F-29238 Brest Cedex 3, France

Preliminary Version - 14th February 2005

Abstract: Usual models of search markets describe the equilibrium pricing strategies of sellers confronted to a population of buyers searching for the best deal: in equilibrium, a seller perfectly anticipates both the pricing strategy of his rivals and the search behaviour of buyers, and optimally responds to it. We show in this paper that "less rational" sellers are not able to learn this Nash-equilibrium pricing rule through adaptation and repeated interactions. The strategy of sellers can be summed by the distribution of posted prices. Using an analytical model characterized by consumers using a fixed-sample size searching rule, we first define the variance and mean of this price distribution. We then compare the characteristics of this Nash-theoretical distribution with the one obtained through a reinforcement learning process. Interestingly, we show that the qualitative properties of the Nash search equilibrium (NSE) of changes in the market structure are still valid for the RL distribution. That is, mean price and variance exhibit similar variations to a change in the degree of information of consumers.

Key words: search market equilibrium – reinforcement learning process – fixed sample size search strategy – numerical computation

JEL Codes: D43 - D83 - C63

1. Introduction

This paper analyses market pricing by adaptive sellers in the presence of imperfectly informed consumers: some consumers are uniformed (then captive to one seller) while others shop around and visit a fixed number of sellers (fixed sample size search). In that case, it has been shown (see Varian [1980]) that the unique symmetric equilibrium is in mixed strategy. Indeed, sellers' profit-maximizing strategy can be broadly comprehended as a mix between a high sales – low price strategy (thus targeting the informed population of buyers) on the one hand and a low sales – high price strategy (thus targeting the uninformed population of

♣ E-mail adresses: darmon@idefi.cnrs.fr (corresponding author) ; Tel : 06-61-57-97-66 ; Fax : 04 93 65 37 98 ; homepage : http://idefi.cnrs.fr/hp/ed/

buyers) on the other hand. In such a context, it has been conjectured that higher information transparency (for example on electronic markets) would lead to lower prices and to a reduced price dispersion. However, price dispersion has remained a commonly observed phenomena on the Internet for seemingly homogeneous goods (Brynjolfsson et Smith [2000]). Waldeck [2005] shows that the link between information and pricing is often misleading in the context of search theory models. Notably, better information may lead to higher price dispersion and more intensive search by shoppers may lead to higher prices in a symmetric mixed strategy equilibrium of the game. Besides, although Varian [1980] suggests a dynamic interpretation of the equilibirium concept, this equilibrium concept remains essentially static and these models miss for a convincing story about how the choices made by adaptive-learning sellers may converge to the Nash-equilibrium price distribution.This paper questions the relevance of this equilibrium concept to explain real observed phenomena. It shows that even if convergence to the mixed strategy equilibrium is not observed, the Nash prediction with respect to a change in the market structure are still valid. More specifically, we test for a change in the degree of information on the price distribution. A second, but related contribution of the paper may be on the market outcome itself, notably on price dispersion analysis.

We consider in this paper less rational sellers using a simple reinforcement rule. One question is whether the equilibrium converges to the Nash Search Equilibrium (further NSE) predicted by standard game theorical models. Reinforcement Learning (further RL) rules, may be at a first view justified, by their ability to fit experimental data: Erev et Roth [1998] explore the robustness of mixed strategy equilibrium prediction in normal form games with a unique mixed strategy equilibrium. For a large number of experiments in games that involve repeated plays of 100 periods, and that are characterized by different information structures, they show that a simple one parameter model outperforms largely the equilibrium price prediction of the game. Moreover, not only are data better explained but also the predictive power of a simple reinforcement learning model performs far better than the mixed strategy equilibrium. Therefore, RL seems to have find some support in the literature of experimental economics for explaining observed data. However, Sarin et Vahid [2001] use a model with optimising myopic players and show that this model does at least as good as the RL model used by Erev and Roth. Despite the previous argument, we chose in this article to use a RL-based learning rule. This choice can be justified by three arguments. First, Salmon [2001] shows that econometric models used to retrieve agents' behaviour may be inadequate. He begins simulating experimental data and then then estimates the econometric models on the simulated data to determine if they can correctly identify the rule that was intially used to generate the data. The results show that all the models examined possess difficulties in accurately distinguishing between the generating process, justifying the fact that assuming a specific learning model only by introspection of experimental data may be difficult and misleading. Second, Hopkins [2002] shows that in normal form games, a *perturbed* reinforcement rule and a stochastic fictitious play may exhibit the same dynamical properties. Especially, fixed points and stability properties are the same as those of a continuous version of the replicator dynamics. This result suggests that although a reinforcement rule is used in our model, the derived properties may be more general than supposed with the addition of some noise in the choice process. Third, Simon [1977] argues that agents confronted to a complex situation, may revert to simple heuristics to avoid computational overload. In addition, as noted by Simon, the point must be on the way human choose, *i.e.* procedural rationality and not on the choice made i.e substantive rationality. Or in complex situations, agents may find a procedure to simplify the complexity of the world and then take good or optimal decision within this simplified framework. An example of such procedural rationality is provided by A.P. Kirman et Brousseau [1991]. They show that, although agents may use some complicated econometric models, misspecification in the models used by agents may lead to a very different outcome would the agents know the true underlying model. Therefore, from the modeller point of view, the modelling of a market is highly stylised. Not only does a modeller not know the exact market structure both on the supply and demand side but also one can hardly imagine what kind of rationality players have and what kind of learning rules they can use. Consequently, we can only compare two types of stylised markets: one in which sellers are Nash and another where sellers are adaptive. These three arguments show that the choice of a specific learning rule may, once considered our objectives in this paper, is only justified on an *ad hoc* basis.

In a search context, using Reinforcement Learning has then the advantage of introducing a dynamical choice model. One key issue is how sellers may learn the demand curve and buyers' search strategy. Some papers have already tackled with this issue: Alan P. Kirman et Vriend [2001] propose an agent-based (ACE) model which tries to reproduce two stylised facts observed on Marseille fish market: high loyalty and high price dispersion. In their model, sellers are capacity constrained, prices are not posted, and buyers make repeated purchases. They show how buyers' behaviours become heterogeneous since buyers can be discriminated according to their loyalty[1]. Using the same kind of behavioural model but in a capacity-unconstrained framework, Brenner [2002] shows how buyers and sellers may learn to bargain. As bargaining is costly, he determines in which cases, such market may converge to a posted price market. Cason *et al.* [2003] try to fit simulated data with experimental data. An hybrid of gradient dynamics and logit dynamics seems to best reproduce the observed data. Hopkins et Seymour [2002] show that Nash Search Equilibrium (NSE hereafter) is dynamically unstable under a wide class of learning dynamics, including reinforcement learning. However, they assume a continuum number of small firms.

We address the same issue of convergence to a NSE when the number of firms is finite. A related issue is the property of the stationary price distribution when the degree of information varies. In a previous paper, Waldeck [2003] calculated the mean price and variance of the price distribution at the NSE as consumers are endowed with varying degrees of information. He considered two types of consumers: informed consumers who sample k sellers (fixed sample search, in proportion a of the whole population) and uninformed consumers who only visit one seller. He showed that the mean price increases with k and that in most cases, price dispersion does not reduce with an increase in the proportion of informed consumers or with an increase in k[2]. Taking this model as a benchmark, we will compare the results of the NSE model (mean price, variance) to those obtained with adaptive sellers using reinforcement learning.

Section 2 presents the simulation model. Section 3 compares the main findings of the NSE to those obtained with RL sellers. Section 4 discusses the generality of the results stated in Section 3. Section 5 concludes.

2. The Model

We model the market as a posted-price mechanism where S sellers (indexed by s) and B buyers (indexed by b) respectively produce and consume an indivisible and homogeneous

---

[1] In most cases, sellers reward loyalty positively by giving to loyal buyers are given a higher chance of being served.

[2] k has to be sufficiently low (around k = 8 depending on the proportion of informed consumers). However, it corresponds well to real world considerations and moreover the variance remains significant for higher k.

item. Each market period t ($t = 1,...T$) is divided into three sub-periods (see *Appendix A* for the timeline): first, sellers independently choose the price they will post during the period; second, buyers visit sellers and transact; third, sellers evaluate the profit generated by the pricing rule currently used and reward this rule. Using the updated reward, they set the price posted at period *(t+1)*, *etc*.

21. Sellers

There is no capacity constraint and sellers incur a constant *per* unit cost (c). As costs are identical for every seller, we further assume w.l.g. that the production cost is zero. At the beginning of each period, sellers independently set the price they will post for this period. Sellers cannot change their posted prices for the current period (no bargaining) but can freely change price from one period to the next period (no menu cost). They set their posted price (p_t) by using a simple RL process[3] (see Alan P. Kirman et Vriend [2001]): assuming that prices are discrete variables ranging from c to v with an exogenous step ϕ (with v being buyers' reservation price), sellers then have $\left(1+\left(v-c\right)/\phi\right)$ price rules namely {Price c, Price $c+\phi$, Price $c+2\phi$, …, Price v}. Using an RL-system, each price corresponds to a rule which is defined by a triplet {Condition , Action , Fitness}. In our framework, rules have no condition part i.e. they *can* be applied whatever sellers' environment. The action part is simply the price sellers decide to post during the period. The fitness of a rule (F) refers to its ability to satisfy sellers' objective (i.e. profit). At the first period, sellers anticipate that all posted prices yield the same expected profit. Hence, all pricing rules are initially endowed with the same fitness noted F_0. Coefficient F_0 can also be interpreted as sellers' expected belief about market-profitability.

$$F^i_{t=0} = F_0 \quad \left(\forall i\right) \tag{1}$$

During the subsequent periods, sellers revise the expected profit of each rule according to the actual instantaneous profits π_t generated by that rule. Let n_t denote the number of actual transactions implemented at period *t*. For any pricing rule *i* (used during period *t*), the fitness of Rule *i* (F^i_t) is thus updated as follows[4]:

$$F^i_{t+1} = F^i_t + \alpha\left(\pi_t - F^i_t\right) \text{ with } \pi_t := (p_t - c)(n_t) \tag{2}$$

Coefficient α $\left(\alpha \in \left]0,1\right]\right)$ measures the relative weight attached to current experiences. This updating rule dynamics mimics an adaptive trial-and-error process: if, during the current period, Rule *i* generates a higher profit than what it did in the past, its reward increases and hence, the probability of selecting this rule is reinforced for the next periods (see Equation (3) below).

At each period, each seller sets a price according to a trembling-hand process. In other terms, sellers' behaviour is characterized by a trade off between exploration and exploitation: at each

[3] See Alan P. Kirman et Vriend [2001] for more details about the learning process; see also Sutton [1991] for a general description of RL processes.

[4] Appendix *A* provides further details about the implementation of the learning process.

period, each rule is selected with a probability drawn from a Boltzmann distribution described by Equation (3):

$$prob\{Select\ Rule\ i\} = \frac{e^{\frac{F_t^i}{\tau}}}{\sum_j e^{\frac{F_t^j}{\tau}}} \quad \text{with } \tau > 0 \tag{3}$$

Parameter τ sets the trade off between exploration and exploitation: as $\tau \to 0$, sellers tend to select only the rules which have generated the highest payoffs in the past (choice of the "greedy action"). As τ increases, sellers tend to explore alternative rules more frequently.

We can notice that thanks to this formulation of the reinforcement learning process, sellers do not set a single price but may set different prices sampled from a learned distribution. Consequently, a stable strategy is here defined by a stable price *distribution*. Besides, it should be noted that the reinforcement scheme is formulated so that *i)* there is no strategic interaction between two different sellers and *ii)* the reward associated to each rule are strictly individual (individual learning).

2.2. Buyers

We are dealing with non repeated purchases. For that reason, we assume that a constant flow of B "new" buyers come to the market and that buyers cannot recall the identity of sellers. Each of them needs to purchase one unit of an indivisible good. The accepted price at which the transaction is sealed should not exceed their reservation price (v) which we assume identical for all buyers. We distinguish two types of buyers: *informed* buyers (proportion a of the total population of buyers with $a \in [0,1]$) and *uniformed* buyers. Informed buyers visit k sellers per period (fixed sample search strategy) and buy at the lowest proposed price; *uniformed* buyers have higher search costs and hence visit only one seller during each period and buy at that price (since this price is lower than their reservation price). As we deal with non-repeated purchases, new buyers permanently come to the market, and consequently cannot remember the identity of sellers from one period to another. For that sake, the sample of selected sellers is randomized for each period and each buyer, so that each seller has the same probability, $1/S$ (resp. C_S^k) of being sampled by an uninformed (resp. informed) buyer.

2.3. Derivation of the NSE

We can derive the Nash pricing equilibrium in such setting. We suppose first that each seller gets an equal share $\left((1-a)B\right)/n$ of captive consumers. The probability that a seller (say i) has the lowest price among k prices randomly chosen by an informed buyer is $(1-F(p_i))^{k-1}$. There are C_n^k ways of choosing k sellers among n. The probability, that a seller i gets a sale by a shopper, is then equal to $\left(C_{n-1}^{k-1}/C_n^k\right)\left(1-F\left(p_i\right)\right)^{k-1} = (k/n)\left(1-F\left(p_i\right)\right)^{k-1}$. Consequently, the expected demand of a seller fixing price p is thus $D(p) = (B/n)\left(1-a+ak\left(1-F(p)\right)^{k-1}\right)$. Given a null marginal production cost $(c=0)$, the expected profit of a seller fixing price p is

$\Pi(p) = p\ D(p) = p(B/n)\left(1 - a + ak(1 - F(p))^{k-1}\right)$ and the profit of the monopoly price v is $\Pi(v) = (B/n)(1-a)v$.

A property of a mixed strategy equilibrium is that all prices charged with positive density must yield the same expected profit: if a price yields a greater expected profit than some other prices, then it would lead to increase the frequency with which the more profitable price was charged. Using the fact that *v* is always fixed in equilibrium and $\Pi(p) = \Pi(v)$ for all prices in the support of the distribution, the cumulative equilibrium price distribution $F(p)$ is equal to: $F(p;a,k) = 1 - \left(\frac{(1-a)(v-p)}{kap}\right)^{\frac{1}{k-1}}$. Since $F(p,a,k)$ is atomless, the lower bound $b(a,k)$ of the support of the distribution is such that $F(b,a,k) = 0$ then $b(a,k) = \frac{(1-a)v}{(k-1)a+1}$. The upper bound of the equilibrium price distribution is *v*.

3. Results of the analytical model

To compare NSE- with RL- outcomes, we will run {100} simulations for each parameters configuration (a,k). We define each simulation as a session. For each session, we will compute the aggregate mean and the variance of the price distribution over the last 100 periods [5]. The process has been simulated in Java: the appendix reports an abridged pseudo code of the model. Source code, classes and an executabe version of the program are available on request.

We assume w.l.g. that $c = 0$ and $v = 100$. We present the results of simulations on (a,k) using a model where τ = 0.05, α = 0.8 and F_0 = 0.2. We want to simulate a situation where individuals' initial expectations are not too optimistic (F_0 = 0.2), price volatility is not too high or random (as it might be with higher temperature) and memory of individuals is not to long (typically with α = 0.8 , only 3 to 4 periods of time remain significant in individual decisions). The influence of parameters τ ,α and F_0 on the different results will be detailed in a latter section.

Result 0. *The stationary distribution with reinforcement learners does not converge to the Nash distribution for parameters "a" different from 1.*

By performing a Kolmogorov-Smirnoff adequation test, we test the hypothesis that the RL- and NSE- distributions are equal. With a 5% significance level, these tests systematically reject the equality assumption for all parameter (a,k) with $0 < a < 1$. We could have expected such a result. Indeed, Posch [1997] has already shown that in normal form games with a unique mixed strategy equilibrium, Reinforcement Learning, of the type used in Erev et Roth [1998][6], does not converge to this type of equilibrium. Moreover, the expected motion of

[5] We have chosen the length of time in each session (the number of rounds (T)) so that the process is ensured to converge to a stationary position.

[6] Although Erev and Roth use a choice rule of the form $\text{proba (Select rule i)} = F_t^i \Big/ \sum_j F_t^j$

reinforcement learning is given by the evolutionary replicator dynamics and so inherits from the same stability properties.

3.1.Mean Price

Corollary, the NSE model leads to the following propositions[7]:

Proposition 1. The NSE expected posted price decreases from the monopoly price v to the marginal cost 0 as a varies from 0 to 1. The NSE expected posted price increases with k : $k_1 < k_2 \Rightarrow E(p_1, a, k_1) < E(p_2, a, k_2)$. For a = 0 (resp. a= 1) the Nash equilibrium is the monopoly price (resp. the Bertrand equilibrium at p=1)

Table 1 shows the average price obtained with RL and confronts it to the NSE predictions.

Result 1. *RL mean price is a decreasing function of the proportion of informed consumers in the market. Moreover, NSE mean price is in general higher than the RL mean price except when all consumers are informed (a=1) or for value of k less or equal to 4 and a large proportion of informed consumers.*

This conforms to the prediction that an increase in the proportion of informed consumers leads to lower prices.

Result 2. *The RL mean price is an increasing function of k except for k=2 or when a= 0 or a= 1.*

Result 2 is in line with the NSE result that an increase in k leads to a higher mean price.

Remark 1. We will show in the discussion to follow that for a=1, RL mean price should be equal to marginal cost pricing for all k. However, the larger k, the lower the number of market periods (T) required for marginal cost pricing. For T=1000 market periods, the process did not stabilized for “small” values of k.

Table 1: Theoritical Mean Market Price versus RL Mean price for v=100

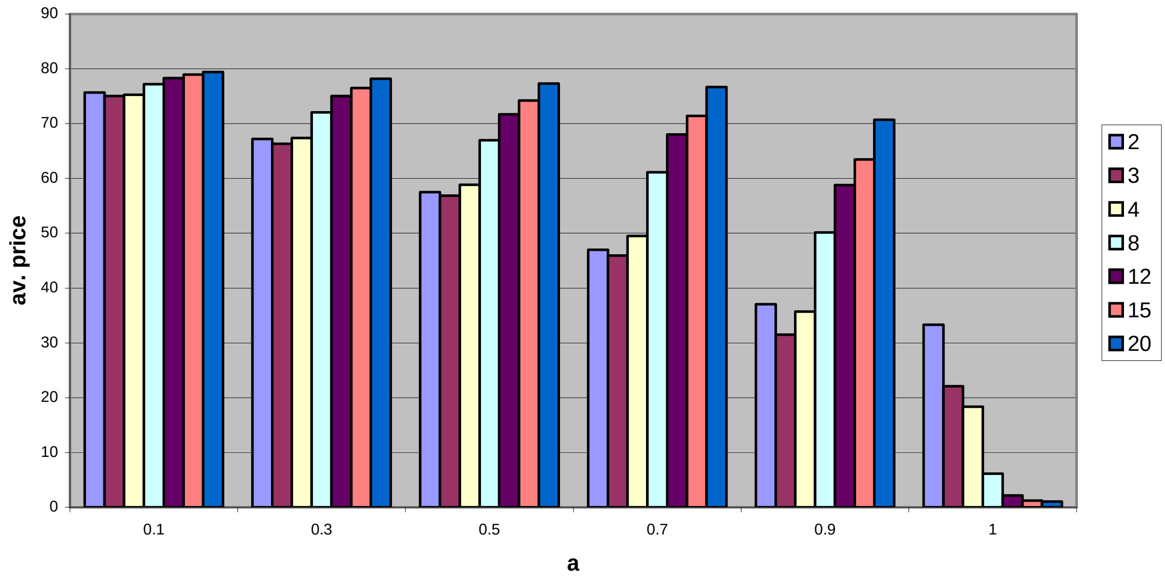
a \ k	2		3		4		5		8		10		15	
0	100	**79.8**	100	**79.8**	100	**79.8**	100	**79.8**	100	**79.8**	100	**79.9**	100	**79.9**
0.1	90	**75.7**	91	**75.0**	91	**75.3**	91	**75.9**	92	**77.2**	93	**77.8**	94	**79.0**
0.2	81	**71.5**	82	**70.7**	84	**71.4**	85	**72.2**	87	**74.7**	88	**75.9**	90	**77.6**
0.3	72	**67.2**	75	**66.3**	77	**67.4**	79	**68.8**	82	**72.1**	84	**73.7**	87	**76.5**
0.4	64	**62.4**	68	**61.7**	71	**63.2**	73	**65.0**	79	**69.5**	81	**71.6**	85	**75.3**
0.5	55	57.5	60	**56.9**	65	**58.9**	68	**61.3**	75	**67.0**	78	**69.5**	83	**74.2**
0.6	46	52.3	53	**51.7**	59	**54.4**	63	**57.3**	71	**64.1**	75	**67.4**	80	**72.9**
0.7	37	47.0	46	**45.9**	52	**49.5**	57	**52.9**	67	**61.1**	71	**64.9**	78	**71.4**
0.8	27	41.7	37	39.5	45	**43.5**	51	**47.6**	63	**56.9**	67	**61.1**	75	**68.9**
0.9	16	37.0	27	31.5	35	35.7	42	**40.0**	56	**50.2**	62	**55.0**	71	**63.5**
1.0	1	33.3	1	22.1	1	18.3	1	9.2	1	6.2	1	3.6	1	1.2

Comments: the dashed cells are the theoretical prediction of NSE. The bold numbers indicate that the RL mean price is lower than the NSE mean price. Note that for a = 0, there are no informed consumers so that the variation in the mean price as a function of k is only a matter of statistical randomness .

[7] Due to lack of space, we could not report the proofs in this paper. All demonstrations are available in Waldeck [2003].

Figure 1 shows the average posted price for different value of a and k. For $a = 1$, the convergence to the Bertrand equilibrium is better achieved for larger values of k.

Figure 1: Average posted price



One interesting question is whether firms learn to fix the same average price. To test this assumption we conducted a Fisher Snedecor test over the last 100 periods. Let H0 be that all firms set an equal mean price. Let H1 be that at least one firm set a different mean price. The table below shows that in general firms learn different prices. This is especially true when a is greater than 0.2 and k>3 for which in general at least one firm fixes as a different mean price from the rivals.

Number of time H0 is true at 5% level

	k																		
a	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0.0	66	67	57	52	62	59	57	57	53	64	55	62	60	53	53	61	65	59	63
0.1	69	51	51	41	47	43	38	37	51	43	45	42	42	44	50	40	54	45	43
0.2	64	38	33	28	17	21	18	26	24	22	29	27	21	19	25	22	23	18	19
0.3	51	28	10	11	7	10	7	8	7	6	5	8	4	13	6	8	9	11	8
0.4	45	15	4	3	8	1	4	4	0	3	5	1	1	4	3	1	2	1	2
0.5	33	13	3	2	0	0	0	0	1	0	0	0	1	0	2	0	0	0	0
0.6	37	8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.7	33	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.8	38	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.9	47	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.0	45	34	21	6	4	2	0	0	0	10	10	24	14	17	10	13	11	11	3

Proposition 3. The NSE mean price paid by informed buyers decreases with k and a (proportion of informed buyers).

Result 3. *The RL mean price paid by informed consumers decreases with k and with* a.

Table 2: Average Accepted Price (informed buyers)

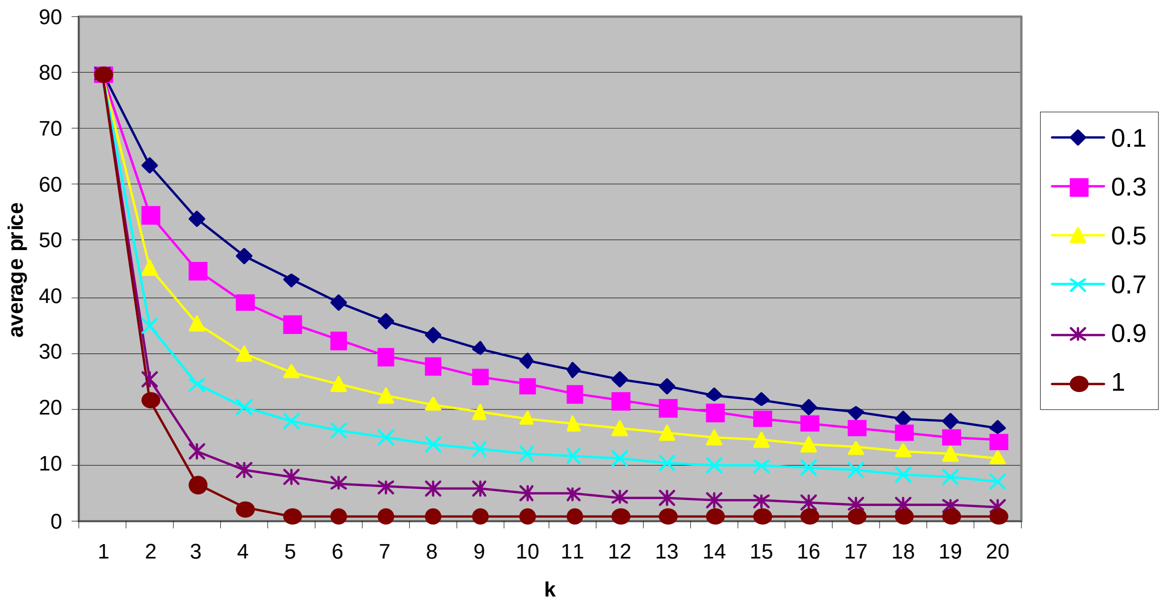
	k											
a	2		3		5		8		10		15	
0.1	87	63.4	84	54.0	78	43.0	70	33.3	66	28.7	58	21.8
0.2	76	59.1	70	49.5	62	39.1	53	30.8	48	27.0	40	20.2
0.3	65	54.8	59	44.9	50	35.3	41	27.7	37	24.3	30	18.5
0.4	55	50.0	49	40.1	40	31.1	32	24.4	29	21.3	23	16.5
0.5	45	45.1	40	35.2	32	26.8	25	21.0	22	18.5	17	14.6
0.6	36	40.1	31	29.9	25	22.3	19	17.4	17	15.3	13	12.2
0.7	27	34.9	23	24.3	18	17.8	14	13.8	12	12.2	9.4	9.8
0.8	18	29.7	16	18.5	12	13.0	9.4	9.8	8.1	8.7	6.2	7.1
0.9	9.3	25.3	8.2	12.3	6.4	7.9	4.9	5.7	4.2	5.0	3.2	3.6
1	1	21.7	1	6.7	1	1.0	1	1.0	1	1.0	1	1.0

The dashed cells are for the NSE average price and the white cells for the RL average price.

We can also note that the average price paid by informed consumers in NSE is larger than the average mean price in RL for low values of (a).

Figure 2 shows the average price paid by informed consumers as a function of k.

Figure 2: Average accepted price informed consumers



3.2. Standard deviation

Proposition 4. The variance of the NSE posted price (excluding the standard cases of a=0 or a =1) is an inverse U-shaped function of a. However, the variance is in general increasing with a (depending on the value of k): for $k = 2$, the variance starts to decrease at $a = 70\%$; this proportion increases steadily with k since for $k = 3$, we have a maximum at $a = 80\%$, with $k = 4$ we have $a = 90\%$ and for $k \geq 8$, a must be greater or equal to 99%. Keeping a constant, the variance increases steadily with k whenever $k < 8$ and remains significant even for large k.[8]

[8] The variation of the variance with respect to k has not been proven analytically in Waldeck (2005) and is only a numerical result.

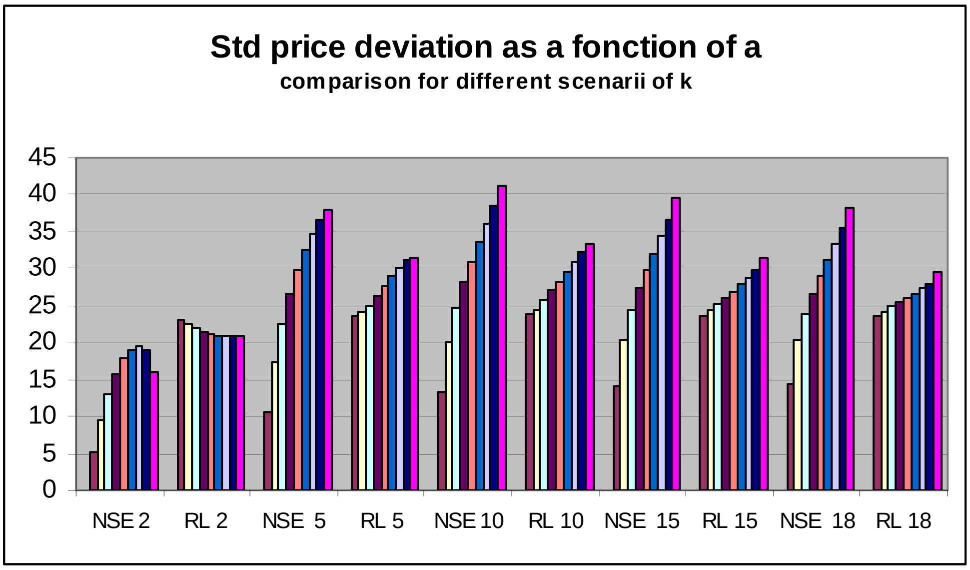
Result 4: *Variation with respect to the proportion "a" of informed consumers: standard deviation is decreasing for k= 2. For other values of k standard deviation is an inverse U-shaped function of a with a maximum at 80% of informed consumers for k=3 and 4 and a maximum at or above 90% for larger values of k. (Figure 3a)*

Table 3: RL average standard deviation of posted price

	k												
a	2	3	4	5	6	7	8	9	10	12	15	18	20
0	23.5	23.4	23.4	23.4	23.4	23.4	23.3	23.3	23.4	23.5	23.3	23.3	23.3
0.1	23.2	23.4	23.7	23.7	23.7	23.8	23.7	23.7	23.8	23.7	23.5	23.5	23.5
0.2	22.6	23.3	23.9	24.2	24.5	24.6	24.5	24.6	24.5	24.5	24.3	24.2	24.2
0.3	21.9	23.4	24.4	25.0	25.4	25.7	25.6	25.7	25.6	25.5	25.2	24.9	24.7
0.4	21.4	23.7	25.3	26.2	26.8	27.0	27.1	27.0	27.0	26.7	26.1	25.6	25.2
0.5	21.1	24.3	26.5	27.6	28.1	28.4	28.4	28.4	28.3	27.9	27.0	26.1	25.5
0.6	20.9	25.1	27.6	28.9	29.5	29.9	29.8	29.8	29.6	29.0	27.9	26.7	25.7
0.7	20.8	25.7	28.6	30.1	30.8	31.2	31.2	31.1	30.9	30.1	28.8	27.3	25.9
0.8	20.8	**26.0**	**29.3**	31.1	32.0	32.5	32.6	32.5	32.2	31.5	29.8	27.9	25.9
0.9	20.8	25.3	29.3	**31.3**	**32.7**	**33.4**	**33.7**	**33.4**	**33.4**	**32.8**	**31.3**	**29.6**	**27.5**
1	21.0	22.5	22.4	18.7	14.7	15.2	13.5	10.9	7.2	3.1	0.8	0.5	0.3

The bold numbers represent the maximum std deviation for a given *k*. The dashed cells represent the maximum std deviation for a given a.

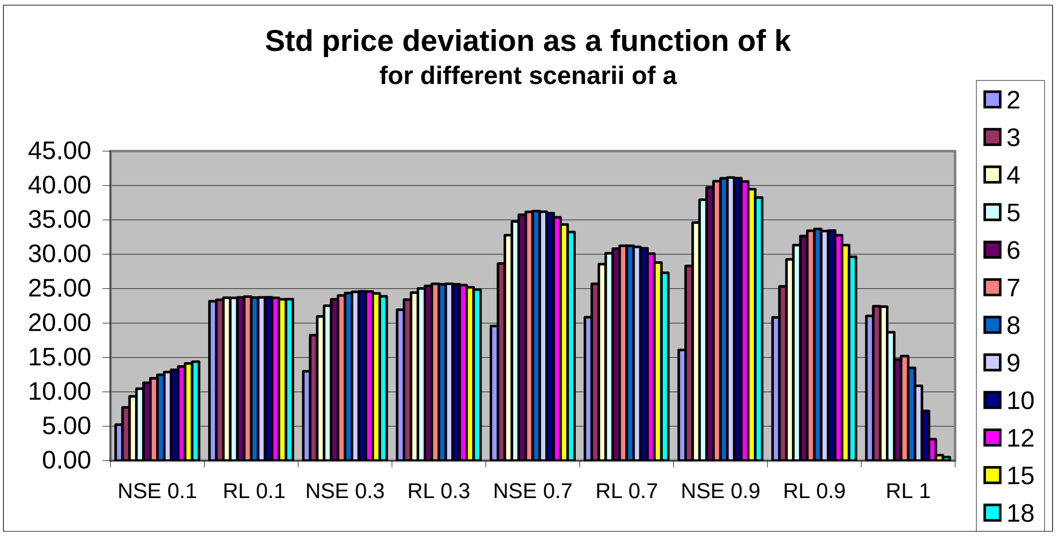
Figure 3a



We can note that although price dispersion increases as a fonction of *a* (except when *k*=2), the variation are less pronounced for RL. The same remark holds for variations with respect to k as pictured in figure 3a.

Result 5 (*variations with respect to k): for a given value of a the variance is an inversed U-shaped function of k with maximum value around k=7 or 8. Variance is steadily increasing for low values of k whereas it decreases at a lower pace for larger values of k.*

Figure 3b



To sum up, also the RL algorithm does not converge to the NSE equilibrium, we see that the qualitative properties of the NSE equilibrium with respect to variation in the market structure are still valid.

Result 5 bis: *The RL coefficient of variation (CoV) is higher than the NSE CoV for proportion of informed consumers lower than 0.5. For a greater than 0.5 the reverse is true with however less difference in variability between NSE and RL. Moreover for small values of a the RL CoV show small variability with respect to k.*

Table 3 (Bis): Coefficient of Variation

	k											
a	2		3		5		8		10		15	
0	0	29.4	0	29.3	0	29.3	0	29.2	0	29.3	0	29.2
0.1	5.8	30.6	8.5	31.2	11.5	31.2	13.6	30.7	14.2	30.6	15.0	29.7
0.2	11.8	31.6	16.6	32.9	20.5	33.5	22.4	32.8	22.8	32.2	22.6	31.3
0.3	18.1	32.7	24.3	35.3	28.5	36.4	29.7	35.6	29.3	34.8	27.9	32.9
0.4	24.6	34.3	32.3	38.5	36.3	40.4	35.5	38.9	34.7	37.7	32.2	34.6
0.5	32.3	36.7	41.4	42.7	43.7	45.0	41.5	42.5	39.8	40.7	36.0	36.3
0.6	41.5	39.9	51.1	48.6	51.4	50.4	47.6	46.6	44.8	43.9	40.2	38.2
0.7	52.9	44.3	62.3	56.0	61.0	57.0	54.2	51.1	50.6	47.6	44.0	40.3
0.8	70.0	50.0	79.3	65.8	72.0	65.4	61.3	57.3	57.2	52.8	48.9	43.3
0.9	100.6	56.2	104.8	80.4	90.4	78.3	73.4	67.1	66.3	60.8	55.6	49.4
1	0	63.1	0	101.6	0	202.4	0	218.9	0	198.5	0	65.7

Comments: dashed cells = NSE CoV; white cells : RL CoV

For small values of *a,* NSE prices are more concentrated (*i.e.* with less variability) around the monopoly price than RL prices. For small values of *a,* the RL mean price is lower than the NSE mean price whereas standard deviation is higher for RL than for NSE (figure 3b). For large values of *a*, RL and NSE mean prices have grossly the same levels whereas standard deviation is higher for NSE than RL (figure 3b) meaning that prices are more concentrated in RL around the competitive price.

3.3. Profit

A property of the NSE equilibrium is that within the support of the price distribution each price leads to an equal profit. If not, price strategies that lead to an inferior profit would be systematically withdrawed. Moreover, the supremum of the support is equal to the monopoly price v. (Varian [1980]).

Proposition 5. In NSE equilibrium with $a < 1$, each firm earns an expected profit equal to $v(1-a)B/S = 5000(1-a)$. As a=1, the equilibrium strategy is a Bertand equilibrium $(p = 1)$ and each firm earns a expected profit of 50.

An implication of proposition 1 is that expected profit is decreasing with "a" as long as a < 1. Moreover, the NSE expected profit is independent of k.

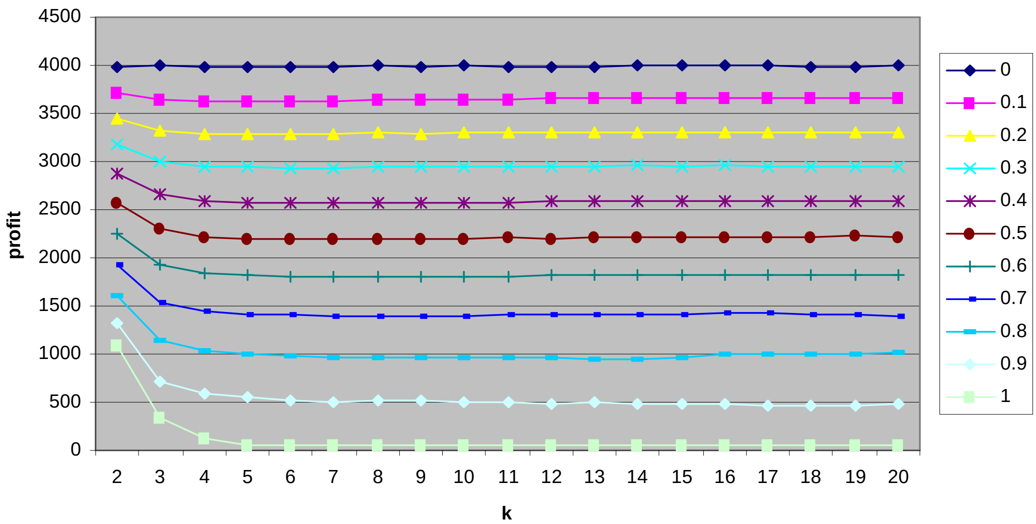
Result 6: *Profits are a decreasing function of a. However profits are independent of k for value of k larger than 4. In general profit are independent of k. (see table 4 and figure 4).*

Table 4: mean profit (left cell) and coefficient of variation (CV %) (right cell)

	k																		
	2		3		4		5		8		10		12		15		20		
a	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	NSE
0	3988	33	3991	33	3989	33	3988	33	3991	33	3993	33	3986	33	3995	33	3991	33	**5000**
0.1	3721	31	3645	30	3626	30	3632	29	3641	29	3644	29	3653	29	3663	29	3656	31	**4500**
0.2	3450	30	3324	28	3288	27	3280	27	3296	26	3307	27	3303	28	3308	30	3301	34	**4000**
0.3	3173	28	**2994**	26	2945	25	2938	25	2938	26	2946	28	2950	30	2955	33	2952	40	**3500**
0.4	2872	26	**2652**	24	2590	24	2571	25	2573	28	2574	31	2586	33	2590	38	2586	48	**3000**
0.5	2566	25	**2304**	24	2221	25	2201	26	2199	31	2197	35	2205	38	2219	44	2217	57	**2500**
0.6	2249	25	**1930**	25	1846	26	1817	29	1802	35	1806	40	1814	44	1824	51	1817	65	**2000**
0.7	1927	26	**1541**	27	**1453**	30	1415	33	1400	41	1400	46	1409	52	1415	61	1395	75	**1500**
0.8	1607	28	**1136**	31	**1035**	35	996	39	962	49	959	55	956	60	972	71	1017	109	**1000**
0.9	1325	32	**712**	38	**596**	44	**556**	51	509	64	498	72	487	78	480	88	475	107	**500**
1	1084	40	**334**	59	**117**	76	**50**	50	49	32	49	23	49	18	49	15	49	14	**50**

Comments: the dashed cells are the value of (a,k) for which RL profits > NSE profits. The cells in bold indicates that average profit are decreasing by more than 5% with an increase in k. The coefficient of variation is equal to the standard deviation per mean profit unit.

figure 4: average profit



The cases where profit decreases with k are due to the fact that the process need a longer convergence time for low k (see remark 2, in the discussion to follow).

4. Discussion

An open question is whether these results are robust to a change in the initial strength F_0 and in the reward updating coefficient α. To test the effect of F_0 , and temperature = 0.05, we run 25 simulations for each of the following parameter configurations (a,k).

Table : Parameter tested (25 runs for each (a,k))			
Sessions	temperature	Reward Updating Parameter	Initial strenght
2500	0.05	0.3	0.6
1000	0.1	0.8	0.2
1000	0.05	0.5	0.6
1000	0.05	0.6	1
1000	0.05	0.8	0.2
1000	0.05	0.8	0.6
1000	0.05	0.8	1
1000	0.05	1	0.2

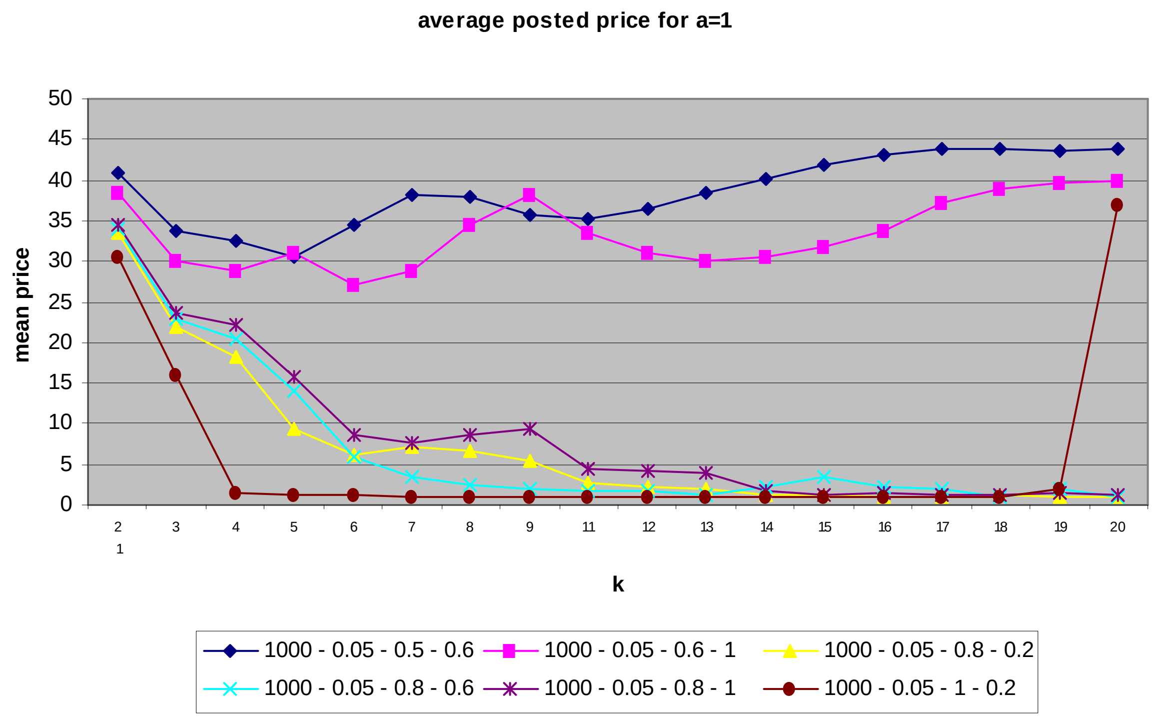
Result 7. *mean posted price: (for all parameter configurations)*

- *RL mean price is a decreasing function of $0 < a < 1$*
- *For $0 < a < 1$, RL mean price is a increasing function of k except for k=2.*
- *The qualitative properties of the mean price are maintained for all parameter configurations.*

Remark 2 for the case a= 1:

- In the case of 1000 market sessions, RL mean price seems decreasing with k. However, this is principally due to the fact that the process has not converged. When the number of sessions rises to 3000, the average posted price is nearer from marginal cost pricing for all k (typically for k=3 mean price = 1.9 and for k>3 the mean price was below 1.2). For k =2, convergence is slower and it requires more than 20000 rounds for an average price of 18. Moreover (for a=1) a higher α leads to lower mean price for all k (figure below).
- Moreover the smaller α, the more rounds are required for convergence to a stationary state. The difference in shape for low values of α (below 0.6) is due to the fact that the process has not converged within 1000 sessions. It requires more that 1500 sessions to converge in that case. In addition, the number of sessions needed are lower the larger k. This may explain why profit is decreasing when a=1 (or a=0.9) from k=2 to k=3, since in this case the process has not converged. For example for a= 0.9 average posted price is equal to 34.3 for 2500 rounds and to 37.6 for 1000 rounds.[9]

[9] A even more curious observation is when RUP =1 and k=20. In this case, the average posted price and standard deviation shows a sudden increase . In addition , if a slight memory effect is added to he individual learning process (for example RUP =0.99) then this sudden increase disappear. We did not find a clear explanation.



Result 8. *Accepted price paid by informed buyers are decreasing with a and with k.*

Result 9.
- *the shape of standard deviation of posted price as a function of a (for a given k) is the same across all reward updating parameters* (α)*, Initial strength specifications tested. It corresponds grossly to the observation of section 3.2.*
- *The shape of standard deviation of posted price as a function of k (for a given a) is the same across all temperature, RUP and Initial strength specifications except for a=1 where the same remark as in result 7 concerning convergence time applies.*

Result 10. *Profits are as described in Result 6 (with Remark 2 holding).*

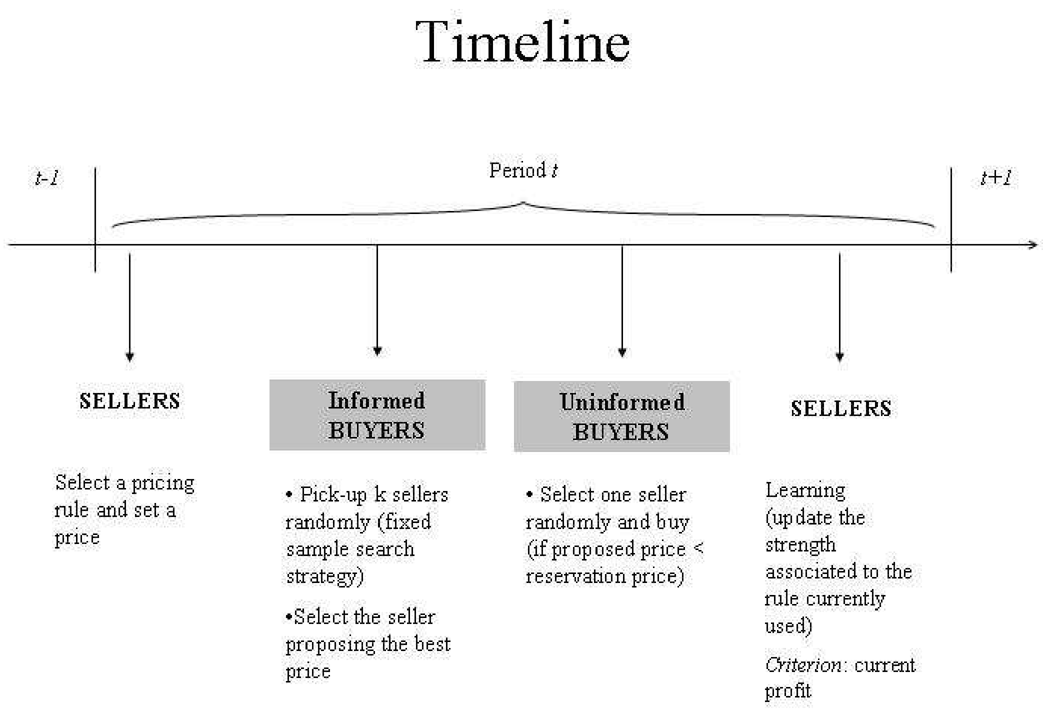
5. Conclusion

The principal contribution of this paper is to show that Nash predictions are performing well with respect to a change in the degree of information and this even if the dynamics does not convergence to Nash. Especially, we studied the impact of varying degrees of information on prices and profits. We showed that Nash predictions are still respected with respect to variations of these parameters. Especially, the mean price is decreasing with a and increasing with k (except for k=2) and the profits are decreasing with a (and constant with k except for k=2 ,3). However, although our result holds on an aggregate level, they are four limitations. First, we studied the stationary distribution of prices , *i.e.* once the process has converged. In general, it takes less than 500 periods for the process to converge. One question is whether 1000 sessions are a short enough period of time to represent real pricing behaviour of sellers. One could argue that real sellers may not start without any prior knowledge about the price distribution of a product. This knowledge may be inherited from experience on similar good markets. However, such initial beliefs are difficult to model. One escape way is to consider that the initial sessions in our simulations stands for this outside knowledge. Second, sellers do not have the same pricing behaviour. However, we did not observe individual pricing behaviour. Third, one may wonder if the result obtained in this paper holds for more general learning behaviours. What if Experience Weighted Average learning (Camerer et Ho [1999]) or imitation is introduced? Fourthly, consumers search was exogenously given. We could proceed with the same analysis with adaptive searching consumers. Points two, three and four may be interesting extensions of this work and should be treated in a near future with the main objective to question whether the qualitative result of NSE are still valid.

Finally, the empirical literature on e-commerce (Brynjolfsson et Smith [2000], Pan *et al.* [2002]) leads to contrasted results about whether electronic markets reduce the level of price and price dispersion as compared to conventional markets. This paper adds comprehension to this contrasting result. More performant search instruments may not have the desired result with respect to prices and price dispersion. Moreover, factors like confidence may lead to a less elastic demand in electronic markets than conventional market (Waldeck 2005). This may impact positively on price dispersion in electronic markets.

Appendix A- Technical specifications

A1- Timeline



A2- Parameters

Numerical simulations have been performed with the following set of parameters:

	Parameters	Value
	T (total number of periods) Γ (number of periods on which results are averaged	1000 20
BUYERS	B (number of buyers) v (reservation price) a (proportion of informed buyers) k (number of visits of informed buyers	1000 100 Variable Variable
SELLERS	S (number of sellers) c (production cost) ϕ (price step) α (Agents 'memory') τ (exploitation/exploration parameter) F_0 (initial strengths)	20 0 1 0.8 (Q-learning) 0.1

Appendix A-3 Abridged Pseudo Code of the model

Main procedure {
- For all Sellers, do {{ Initialize Seller}}
- For all market sessions, do{{
  - For all sellers, do {Seller posts prices}
  - For all informed buyers, do { Informed Buyer makes transaction }
  - For all uninformed buyers, do { Uninformed Buyer makes transaction }
  - Sellers update reward }} }

Proc. *Initialize Seller* {
- Create $\left(1+\left(v-c\right)/\phi\right)$ price rules (Price c, Price $c+\phi$,…, Price v)
- Initialize the strength of each rule to F_0 }

Proc. *Sellers post prices* {
- Normalize all fitnesses
- while { all rule have been tested or one rule is selected }, do {
  - Choose one rule at random
  - Select this rule with probability $\left(e^{\frac{F_i^t}{\tau}} \Big/ \sum_j e^{\frac{F_j^t}{\tau}} \right)$
- if no rule has been selected, choose one rule randomly
- Set the price corresponding to that rule }

Proc. *Informed Buyers make transaction* {
- Sample k sellers randomly
- Buy one unit to the best-price seller, if (proposed price) < v }

Proc. *Uninformed Buyers make transaction* {
- Pick up one seller randomly
- Buy one unit to this seller, if (proposed price) < v }

Proc. *Sellers update reward* {

- Compute current profit
- Update the reward of the rule used: new reward = old reward + α (current profit – old reward) }

Acknowledgements

We are grateful to the participants of the 1st *Bonzenfreies Conference on Market Dynamics and Quantitative Economics* (Alessandria, Sept. 2004, 9-10th) and of the 2nd ELICCIR Workshop (Grenoble, Sept. 2004, 13-14th) for their remarks and comments. This paper has benefited from a support of the Program "Complex systems in Social Sciences" (ELICCIR project) of the French National Center for Scientific Research (CNRS).

References

BRENNER T. [2002], "A Behavioural Learning Approach to the Dynamics of Prices", *Computational Economics*, 19, p. 67-94.

BRYNJOLFSSON E., SMITH M. [2000], "Frictionless Commerce? A Comparison of Internet and Conventional Retailers", *Management Science*, 46(4), p. 563-585.

CAMERER C., HO T.-H. [1999], "Experience-Weighted Attraction Learning in Normal Form Games", *Econometrica*, 67(4), p. 827-874.

CASON T., FRIEDMAN D., WAGENER F. [2003], "The dynamics of Price Dispersion or Edgeworth Variations", *Journal of Economic Dynamic and Control*, (forthcoming)

EREV I., ROTH A. E. [1998], "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria", *American Economic Review*, 88(4), p. 848-881.

HOPKINS E. [2002], "Two Competing Models of How People Learn in Games", *Econometrica*, 70(6), p. 2141-2166.

HOPKINS E., SEYMOUR R. [2002], "The stability of price dispersion under seller and consumer learning", *International Economic Review*, 19(1), p. 46-76.

KIRMAN A. P., BROUSSEAU V. [1991], "Apparent convergence of learning processes in mis-specified games", dans AL., D.B.E. (eds) [1991], *Game Theory and Economic Applications*. Berlin,Heildelberg, Springer Verlag.

KIRMAN A. P., VRIEND N. J. [2001], "Evolving Market Structure: An ACE Model of Price Dispersion and Loyalty", *Journal of Economic Dynamics and Control*, 25(3-4), p. 459-502.

PAN X., RATCHFORD B. T., SHANKAR V. [2002]. "Can Price Dispersion in Online Markets be explained by Differences in e-Tailer Service Quality?" *Maryland University Working Paper*.

POSCH M. [1997], "Cycling in a Stochastic Learning Algorithm for Normal Form Games", *Journal of Evolutionary Economics*, 7(2), p. 193-207.

SALMON T. C. [2001], "An Evaluation of Econometric Models of Adaptive Learning", *Econometrica*, 69(6), p. 1597-1628.

SARIN R., VAHID F. [2001], "Predicting How People Play Games: A Simple Dynamic Model of Choice", *Games and Economic Behavior*, 34(1), p. 104-122.

SIMON H. [1977], "From Substantive to Procedural rationality", dans LATSIS, S.J. (eds) [1977], *Method and Appraisal in Economics*, Cambridge University Press.

SUTTON R. G. [1991], "Reinforcement Learning": Cambridge University Press.

VARIAN H. [1980], "A Model of Sales", *American Economic Review*, 70, p. 651-659.

WALDECK R. [2003]. "Search and Price Competition." *ENST-Bretagne Working Paper downloadable at http://www-eco.enst-bretagne.fr/Recherche/Biblio/Roger/search%20competition.pdf*.

WALDECK R. [2005], "Prices in a Shopbot Market", *To appear in Electronic Commerce Research*,