The paper seeks to show, using a simple agent-based model of a goods market, that the behaviour of rational agents and that of agents using a reinforcement learning mechanism to form expectations can be sufficiently similar that it may be hard to discern whether observed outcomes are the result of fully rational behaviour or bounded-rational learning behaviour. I like the simplicity of the framework used and the paper is quite well-written and compact. However, there are two overarching issues I have with the paper in its current state.

The first is that it is not entirely clear what the specific contribution of the paper is and whether its results are sufficiently strong to merit publication. The main result appears to be a demonstration that rational and non-rational behaviour/expectations can produce qualitatively similar outcomes. I do not believe that this is a claim any economist, whether `mainstream' or unconventional would dispute. The crucial question in this context, also discussed by some of the literature cited by the author, is whether it is possible to statistically distinguish between rational and non-rational behavior.

How do the statistical tests used in the present paper differ from those typically applied in the literature? What would be the result of applying a test such as the one discussed in Ilek (2017), if this is possible? Would it, just as in the examples given by Ilek, be unable to reject the rationality hypothesis on the data produced by the learning agents in some cases? Linked to this, what is the precise innovation of this paper w.r.t. papers such as Ilek (2017)? The author states that the main difference is the use of agents using reinforcement learning. Why is this more appropriate to answer the question than e.g. the use of OLS-learning or naive expectations as in Ilek (2017)?

In the paper the criteria used to judge whether or not rational and non-rational behavior are similar appear to be whether certain coefficients have the same signs under rationality and learning and whether the IRFs to particular innovations look similar. I am not sure whether this really provides for a very strong result. It would be more interesting to see whether particular tests are unable to statistically distinguish between data produced by rational and non-rational agents (in the sense that there is no statistically significant difference between the outcomes under rationality and learning). More broadly, it would also be interesting to see how rational and learning agents perform in terms of forecast errors. Do rational agents significantly outperform learning ones in terms of forecast errors? Overall I think the paper needs to be linked more closely to the existing literature and the author would need to find some way to strengthen the results.

The purpose of using an agent-based model with reinforcement learning was to provide an example of a set-up that is dramatically different from the standard model. Note that Ilek (2017) and other papers still model an aggregate observable measure of expectations (albeit using a somewhat differently specified law of motion). In my model (where the agents simply make one binary choice) the concept of expected value does not even exist. I believe that this (together with the fact that only limited information is available to the agents) shows how minimal are the requirements for the "correct" correlations to emerge.

Note that that since there is no observable measure of the forecasted developments in this model the test applied by Ilek (2017) is not applicable. I will consider the options but do not have any (good) ideas so far. Also, personally, I believe that the IRFs comparison may in fact be more relevant. There is only limited number of empirical studies that test rationality using the aforementioned test, but abundant number of papers use theoretical models to provide an interpretation for IRFs or IRFs to validate theoretical models.

For the same reason I cannot compare the forecasts' errors from different models. Nevertheless, the efficiency of the agents may be implicitly compared by observing the profitability measures reported in Table 1. The learning agents appear to be less efficient. But arguably this result is model-dependent and can be easily altered by e.g. introducing endogenous state partitioning. In fact, I purposefully wanted the outcomes to be different but still indistinguishable in terms of GMM-regressions and VARs.

I will emphasize all these considerations in the revision. Also I intend to add the VARs' forecast performance to show that using information on aggregate demand and trend costs improves the forecasts in all cases.

My second concern is that I found some aspects of the model description confusing. On page 3, it is stated that "There are n agents. Each of them may produce q_n goods". Is q_n a parameter or a choice variable? Below, the choice of agents is presented as a binary one, i.e. to either participate or not to participate in the market. Do agents decide whether to participate in the market and if so, produce a fixed quantity q_n? Or do they also choose their quantity? In the latter case, it should be made more explicit how this quantity is chosen. In the former case, the value of q_n is not given anywhere. Table 1 column 5 reports that under random strategies, 1000 agents on average produce a quantity of 1000. Random strategies here means that agents participate with probability 0.5. This would indicate that q_n = 2 or some distribution centred around 2. In either case, the model description should be updated to make this aspect more easily understandable.

Indeed, the agents only have a binary choice: produce q_n or zero. These are agent-specific parameters determined as $q_n = s_n Q^{max}$ where Q^{max} = 2000 (representing the maximum output) and the shares s_n are determined randomly (first drawn from the uniform distribution $s'_n \in (0,1)$ and then normalized $s_n = s'_n / \sum s'_n$ so that the sum equals unity). Erroneously, this explanation is missing from the paper and will be added in the revision.

Beyond these two major comments, I have some minor ones which are listed below.

1. Introduction

 Page 2, second paragraph: The author mentions some references on the possible `rationality' of heuristics/rules of thumb in complex environments. Such considerations have a long tradition in non-conventional economic thinking dating back to the work of Herbert Simon, and much important work on the concept of `ecological rationality' has been done by Gerd Gigerenzer. The author may wish to include references to some of these works to better ground the paper in the existing literature. Moreover, Dosi et al (2017) (https://www.ofce.sciences-po.fr/pdf/dtravail/WP2017-32.pdf) present a paper on the fitness of simple heuristics in a complex macro-ABM to which the author may wish to refer.

 There is also some literature on the idea that identical aggregate patterns can be obtained from diverse behavioural assumptions at the micro level. One example of this is A. Shaikh's `Capitalism' (2016, Oxford University Press, Ch.

3). Perhaps some discussion of such literature should be incorporated

 Page 2, fourth paragraph: Here, too, a bit more discussion of existing literature would be helpful. There is a range of existing papers by authors such as Herbert Dawid, Jasmina Arifovic, Giovanni Dosi, the Santa Fe group etc. on the use of various different learning algorithms in relatively simple models.

How does the learning algorithm used in the present paper differ from others which have been proposed in the literature? Why is this one in particular used?

Thank you for the references. I will use them in the revision.

2. The Model

 Some equations are numbered while others are not. This should be fixed.

I will fix this in the revision.

 p. 3: Why is it necessary to have both an aggregate cost shock (v_t) and an agent-specific one ($e_{n,t}$)?

Note that agents have information on their current costs $c_{n,t}$ when making the market participation choice. If there was a direct link between trend and individual costs (i.e no $\varepsilon_{n,t}$) it would mean that they also possess the information on contemporaneous trend costs. An alternative set-up (without introducing $\varepsilon_{n,t}$) would be to restrict the available information on costs to the previous period's C_{t-1} which seemed less intuitive.

 p. 4, equation (2): It would be useful to include a derivation of the supply curve to aid understanding of the model.

In the revision I intend to use an arguably more intuitive version of the supply curve (expressing Q instead of P) with the following explanation:

Note that the individual agents' costs are uniformly distribute from $C_t^e\lambda^{min}$ to $C_t^e\lambda^{max}$. Accordingly if $P_t^e<C_t^e\lambda^{min}$ zero goods will be produced. If $P_t^e > C_t^e\lambda^{max}$ maximum (Q^{max}) goods will be produced. In other cases the share of goods supplied (out of Q^{max}) will approximately be proportional to the ratio of a current margin ($P_t^e - C_t^e\lambda^{min}$) to a maximum margin ($C_t^e\lambda^{max} - C_t^e\lambda^{min}$):

$$Q_t^e = \frac{P_t^e - C_t^e\lambda^{min}}{C_t^e\lambda^{max} - C_t^e\lambda^{min}} Q^{max}$$

3. The Experiments

 Mixed strategies: Earlier it is claimed that when the model is populated only by rational agents, each agent is aware that the other agents, too, are rational. In this mixed case, are the rational agents aware that 2/3 of the other agents are non-rational? If not, why not? Shouldn't they be if the goal is to have model-consistent expectations for the rational agents?

No, they are not. In this respect the agents are only pseudo-rational. I will acknowledge this in the revision.

page 7; fourth paragraph: It is claimed that \even the model estimated for the limited information agents indicates that output `reacts' to fluctuations of demand, although we know that formally this is not the case". It appears clear from the model description that information about the price level indirectly provides information about the level of demand; indeed, this is explicitly stated later on in the paper (page 8, last paragraph). The result hence does not seem surprising and perhaps this should be stated here.

I will emphasize this in the revision. Notably, the point of the paper is to show how different mechanisms inevitably lead to the same set of correlations.

page 8; top paragraph: If a low autocorrelation of residuals is indeed an indicator of rationality in the present model, why do residuals for rational agents display relatively high autocorrelation (table 2)?

The agents are only pseudo-rational. They only know the distributions of λ_n and q_n (but not the agent-specific values and their interplay) and do not attempt to correct for that. Accordingly, the errors that are smaller compared to the learning agents (as follows from the profitability indicators reported in table 1) but more systematic emerge . I will acknowledge this in the revision.

4. General

 It would be helpful to have some indication of the extent to which the reported results are sensitive to the chosen parameter values. Given that this is a relatively simple model, a basic sensitivity analysis could be conducted at low computational cost.

I will add such analysis in the revision.

Thank you for your comments!