

20-Oct-2022

Dear Author:

The review process for your Marketing Science manuscript, MKSC-22-0247, entitled "Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve," is now complete. I have received comments from two reviewers and an Associate Editor, which are combined in the attached file. The paper was submitted on July 6th, and I would like to apologize on behalf of the journal for the rather long turnaround time.

The review team sees some merit in the ideas you have presented, but there are major revisions that would need to be done if the paper is going to be accepted. However, given the major extent of these needed revisions, there is considerable uncertainty as to the final outcome. Therefore, I invite you to respond to the review team's comments and revise your manuscript.

The quality of both reviews and the AE report is quite exceptional. The AE's report will serve as our revision contract. As you can see, there is a lot of work ahead.

In particular, you will need to benchmark your proposed approach against several extant papers compiled in AE.1 (mostly outside of marketing), both theoretically and empirically.

You will also need to justify various building blocks of your approach (e.g., use of GP, see R1.2.2, AE.2).

In terms of the comparison with Misra et al. (2019), it calls the question of the impact of assuming a differentiable demand function. It might be interested to add a scenario to the simulations that would be inspired by the 99c ending phenomenon (see R2.1), to explore how robust your approach is to situations like these. More generally on the topic of varying scenarios and contexts, AE.3 wonders about contexts in which the proposed approach "will still provide gains even in the longer-run"?

The exposition of the paper will also need a bit of work (e.g., I agree with R1.5 that it might be helpful to get to your proposed approach earlier in the paper).

As the AE notes, once AE.1-3 are addressed, it is not clear whether the proposed approach will still look novel and useful enough to justify publication in Marketing Science. Accordingly, this revision should be considered particularly risky. The AE concludes with a summary of ideas for further improving the paper's contribution (e.g., a field application, theoretical analysis, multiple products). These would all be wonderful additions, but it is hard to comment on these ideas at this stage, as it is unclear whether such additions will indeed be needed in order to improve the paper's contribution, once the points summarized in AE.1-3 are addressed.

Our deadline for revisions is nine months from the date of decision. However, I hope you will be able to submit the revision sooner, preferably in five-six months. When submitting your revised manuscript, please make sure that you respond to the above comments and the comments made by the review team in a separate document. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewers.

*Starting in May 2022, Marketing Science has implemented a competing interests policy ([https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fpubsonline.informs.org%2Fpage%2Fmksc%2Fcompeting-interests-](https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fpubsonline.informs.org%2Fpage%2Fmksc%2Fcompeting-interests-policy&data=05%7C01%7Cvineet.kumar%40yale.edu%7C7dc2a60178854454580208dab2aa5c9a%7Cdd8cbebb21394df8b4114e3e87abeb5c%7C0%7C0%7C638018742148297057%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6IjEhaWwiLCJXVCi6Mn0%3D%7C3000%7C%7C%7C&sdata=xi4d3VsY8z%2BZ0k76NdYuaaX0BzBILa%2FS2q7O%2FAC%2BYL4%3D&reserved=0)

[policy&data=05%7C01%7Cvineet.kumar%40yale.edu%7C7dc2a60178854454580208dab2aa5c9a%7Cdd8cbebb21394df8b4114e3e87abeb5c%7C0%7C0%7C638018742148297057%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6IjEhaWwiLCJXVCi6Mn0%3D%7C3000%7C%7C%7C&sdata=xi4d3VsY8z%2BZ0k76NdYuaaX0BzBILa%2FS2q7O%2FAC%2BYL4%3D&reserved=0](https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fpubsonline.informs.org%2Fpage%2Fmksc%2Fcompeting-interests-policy&data=05%7C01%7Cvineet.kumar%40yale.edu%7C7dc2a60178854454580208dab2aa5c9a%7Cdd8cbebb21394df8b4114e3e87abeb5c%7C0%7C0%7C638018742148297057%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6IjEhaWwiLCJXVCi6Mn0%3D%7C3000%7C%7C%7C&sdata=xi4d3VsY8z%2BZ0k76NdYuaaX0BzBILa%2FS2q7O%2FAC%2BYL4%3D&reserved=0)).

Please add a "Funding and Competing Interests" section above the references section and include all relevant declarations here. Please do not include author names as the peer review process is double blind.

Once again, thank you for submitting your manuscript to Marketing Science and I look forward to receiving your revision.

Sincerely,

Prof. Olivier Toubia
Senior Editor, Marketing Science

Associate Editor

Comments to the Author:

The authors propose a nonparametric multi-armed bandit reinforcement learning algorithm in the context of a pricing problem. Their approach builds on recent work that combines Gaussian process with Thompson sampling (GPTS) for MAB applications. Two expert reviewers have provided their comments. Both reviewers appreciate the topic and the theory-based approach to improving exploration-exploitation performance in this context. Both reviewers raise significant and distinct concerns that are ultimately related to the overall contribution over prior work. Major concerns in this regard are:

1. The authors have sought to position their work relative to UCB-based approach for MAB in a pricing context in marketing by Misra et al. 2019 (MSA19). Both reviewers question whether this is indeed the relevant benchmark to evaluate the contribution of the current paper. Reviewer 1 notes that GPTS for pricing applications has been previously introduced in Ringbeck & Huchzermeier 2017 (RH17). In particular, RH17 also note the advantage GPTS provides over UCB-based approaches with regards to what is referred to in the current paper as the informational externality across nearby price points. Reviewer 2 asks the authors to consider other recent work on GPTS such as Urteaga and Wiggins 2022 as a possible benchmark. Another relevant example seems to be Chowdhury and Gopalan 2017 (On Kernelized Multi-armed Bandits) who find that a GP-UCB variant (referred to as IGP-UCB) outperforms GPTS.

While MSA19 is a useful benchmark (being an early paper in marketing on this topic), given the backdrop of the current literature on GP-based MAB, I would be hard-pressed in viewing it as the sole benchmark. The authors should provide a better discussion of the current literature on GPTS and other GP-based MAB algorithms and frame their contribution in this context and identify further suitable benchmarks (Reviewer 1 incidentally notes the sparse discussion of RH17 and the incorrect reference to Srinivas et al 2009 in the context of GPTS).

2. Like Reviewer 1, I currently see the main contribution of the paper is in incorporating the monotonicity restriction in a GP-based framework. Reviewer 1 questions whether / why GP is crucial in this context (as opposed to a polynomial-based approach). Along similar lines, I would question whether / why TS is crucial. If indeed one or both elements are crucial, this needs to be justified appropriately. If either is not crucial then the value of incorporating monotonicity needs to be established in a broader context (e.g., also with a GP-UCB variant).

3. The authors show that the gain from incorporating monotonicity restriction within GPTS is mainly in the case where consumer WTP is left-skewed (referring to Table 4). Since the price support is scaled, the advantage from the monotonicity restriction is mainly when the price support is chosen too wide on the higher end of prices – while this is plausible the very first time, I suspect this is unlikely to happen frequently past the first learning cycle unless there are frequent drastic changes in consumer WTP. So I am not seeing the advantage in the longer-run performance of ongoing price adaptations

(e.g., weekly or monthly adjustments to track demand changes). Are there contexts where this approach will still provide gains even in the longer-run?

In light of the above concerns, both reviewers have offered some great suggestions that can help strengthen the contribution in light of the above concerns. Reviewer 2 suggests examining application with field data. Reviewer 1 suggests a theoretical analysis of the algorithm's performance. Reviewer 1 also suggests an application to a multi-product context. To be clear, the main concern here is that addressing points 1 to 3 above may not be sufficient to clear the bar for contribution and the authors need to expand the scope of their work further.

Reviewer: 1

This paper proposes a multiarmed bandit algorithm for a firm to learn its demand curve and converge to optimal prices. The algorithm leverages a Gaussian process (GP) to model the dependence of demand at nearby price points and uses Thompson Sampling to balance exploration and exploitation. The algorithm also imposes weak monotonicity of demand by sampling the derivatives of (an approximated) GP, conditional on that the derivatives must be non-positive. The paper presents simulation results and shows that the proposed algorithm outperforms several existing algorithms.

I enjoyed reading the paper, and I have learned quite a bit about the topic. Below are my questions and comments:

Major comments

1. Addressing Ringbech and Huchzermeier

The paper cites Ringbech and Huchzermeier (2019) as a closely related paper. I read through this paper as well, and I found quite a few similarities. Ringbech and Huchzermeier (2019) propose to combine Gaussian Processes with Thompson Sampling to estimate demand. Their motivation for doing so is to leverage GP's ability to characterize correlated information between the focal test point and nearby price points. I believe their proposed algorithm is the GPTS algorithm in the authors' paper. However, I am a bit confused by the fact that the authors' paper makes little mention of Ringbech and Huchzermeier (2019), except for one paragraph in the literature review. I think the paper needs to be addressed up-front, and my take is that the authors' main contribution is to add monotonicity restrictions (adapting a method by Maatouk and Bay, 2017) to the GPTS algorithm. If that understanding is right, then I think the paper's writing should be significantly reformulated to highlight Ringbech and Huchzermeier's results and the authors' contribution on top of their paper.

2. Is the Gaussian Process important? Or is the method more generalizable as a polynomial-based approach?

2.1 From my understanding, the authors follow Maatouk and Bay (2017) and approximate the GP using a polynomial, so that they can impose shape restrictions on the demand function. It took me a while to understand that the method is an approximation rather than exact, so the writing might need some improvements.

2.2 The bigger question: can any polynomial be used together with Thompson Sampling? Once one introduces a global or local function, the function naturally generates information externalities between price points. And given that polynomials are easy to impose shape restrictions (both in first derivatives and in higher derivatives), is the main added value coming from the use of polynomials rather than GPs?

2.3 However, one thing that GP seems to be able to achieve (at least in other contexts, e.g. dynamic programming using GP to approximate the value function) is that they approximate high-dimensional objects well. On the other hand, polynomials are often terrible in high-dimensional space. The authors' examples all focus on one product so this advantage (or disadvantage, if the method is essentially a polynomial) is unclear. To be fair, existing MAB-based pricing papers seem to be low-dimensional, and the authors

are sticking to the literature's convention. But I wonder whether there's anything interesting if one pushes the dimensionality to be much higher.

3. Theoretical discussions about the algorithm's behavior.

I have not seen theoretical discussions about the algorithm's properties. Related to point 2, if the algorithm is essentially a polynomial approach, some of the theoretical properties might have changed. But regardless of whether my understanding above is right or wrong, I think discussing the theoretical properties of the algorithm might help the reader understand its behavior in simulations.

4. More simulation exercises to understand the algorithm's behavior.

There is one case where GPTS gets stuck at \$1 (way too high compared to the optimum) and GPTS-Mono gets out of that point due to the monotonicity restrictions. I appreciate the authors' discussions here, but presenting only a few cases leaves many questions unanswered. For example, how often, and under what empirical conditions, does GPTS get stuck? Does monotonicity always help? What if the price upper bound is higher so that it is less binding? I apologize if my questions are not very clear. But broadly I'd suggest looking at many many different scenarios and present systematic empirical evidence of when/why shape restrictions help.

Broader suggestions and minor comments:

5. Writing.

5.1 The algorithm section is very unclear. I suggest the authors refer to equations/symbols to be precise about what each step is doing. And I never figured out what is "perform Bayesian update" as the last step of several algorithms.

5.2 Much of the paper is dedicated to broad overviews of existing algorithms. While I appreciate the overviews, I have been looking forward to the authors' exact approach, and I only find it more than halfway into the paper.

6. Deeper discussion about the two forms of information externalities

A suggestion: One realization of (quantity, price) under a GP generates information to nearby points in a symmetric way (at least in a one-dimensional case), because of the way the covariance function is set up. On the other hand, monotonicity generates information spillover effects in different ways to all prices above the test point compared to below the test point. I find that these two spillover effects are different, and how they work together is worth thinking about more. The authors have touched upon this point here and there in the paper, but the discussions are so far unclear. It might be interesting to explore a bit more on this angle.

7. On page 14, just into section 4, the authors cite Srinivas et al (2009) for GPTS. I believe they should cite Ringbech and Huchzermeier. I looked at Srinivas et al very briefly, and that paper looks like a GP-UCB algorithm.

Reviewer: 2

In this paper the authors consider a new approach to consider demand learning for optimal price experiments. In their approach they consider GP bandit as opposed to the prior literature that consider an index-based bandit and then adds downward sloping demand as an imposed assumption.

While I am supportive of the approach highlighted here, I do have some concerns:

1. The authors suggest the approach is micro-founded. A key difference between this approach a partially identified approach (ala Mankiw) is that rather than consider a directional result as in Misra et al. they consider a differentiable equivalent. This introduces an additional restriction to the demand system that it must be differentiable at all price points. In contrast, the Misra et al model the underlying demand model can be discontinuous as is built on the weak axiom of revealed preference (as an aside this necessitates the use of segments in their model). This has a few implications:

a. Assuming additional ex-ante information about the demand curve could be in form of a researcher's prior information. However, I do think it is not consistent with the Partial Identification literature. To be clear, I do not mean to suggest that this assumption is incorrect - simply not consistent with the partial identification literature they are adding to (based on axioms of choice).

b. Adding an assumption implies that the model is not directly comparable to the UCB-PI model the authors use as the baseline as clearly additional information (if consistent with the underlying DGP) will make the bandit converge faster. With the addition of the new demand assumption – the authors could consider a different set of models to compare their proposed solution.

c. Implications for feasible demand systems. In the Misra et al paper the authors consider a demand system with discontinuous preference – this is not consistent with the proposed model and has not been considered in the demand implications. In terms of empirical demand systems – consider the literature on 99c ending these impose non differentiable constraint (as the derivative from below is different from the derivative from above) in the demand model at a particular price point (see figure 2 in https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3413019).

2. If we consider the additional assumption added about the demand system as continuity or differentiability (see point 1a) then the question arises why this is the most appropriate model is the one chosen for this paper. With imposing more restrictions the authors can consider more specific choice model frameworks (e.g a choice model with a flexible heterogeneity distribution) to Bayesian NP gaussian frameworks to impose these beliefs. See for example, <https://arxiv.org/pdf/1808.02932.pdf> for a recent paper that shows lower regret bounds compared to the Srinivas et al method (labelled MultitaskGP) implemented in the current paper, with the advantage of direct interpretability as the authors suggest as an advantage on page 17.

3. Application: It would be great to see a real application of this method to show potential advantages in field data.

Overall I think it is great to see work in the area of pricing algorithms combining theory and CS methods, but hope the authors will push this research further.