Answers to the Comments of the Associated Editor, Reviewers #1, #2 and #3

We would like to thank three reviewers and the Associated Editor for their time and their valuable comments which have significantly enhanced the paper.
Below, we provide each comment made by the reviewers (in italics), followed by our response.

Associated Editor:

Three reviewers were consulted to assess the quality of this submission, and its suitability for Operations Research. Two of the reviewers recommend revision, with different levels of excitement, while the other reviewer recommends a rejection. After reading the manuscript, I agree with the majority and believe this paper could be published in OPRE after a suitable, well-executed revision along the lines proposed by the reviewers. Please note that all three reviews are outstanding, and they should be taken with extreme care.
On this note, let add that one of the positive reviewers contacted me after submitting her/his review with additional concerns about this paper. I reproduce her/his additional comments here:

FROM REVIEWER ("This is an interesting paper. I have only two main comments."):
"I filed my report yesterday perhaps a little too quickly and wanted to add this additional concern about the paper. While I am still in favor of ultimate publishing it if you and the other referees agree too, I think there are two main concerns about their approach that the authors need to discuss more clearly.
1. The POMDP approach essentially converts the problem into a higher dimensional problem where the decision maker carries a "posterior" distribution over the unobserved states in the model. This is a standard approach but the curse of dimensionality renders this approach practically infeasible if the unobserved state takes on more than a few values since then the posterior distribution is a point in an N dimensional simplex that makes it a continuous state variable requiring numerical approximation. In the Zurcher application the authors assumed that the unobserved latent state of the bus was either good or bad so there are only two possible values (the smallest possible) and so the simplex is the 1 dimensional simplex, the continuous unit interval. This is feasible and perhaps feasibility extends to maybe dimension 10 or 15, but for higher dimensional problems the authors need to discuss how the problem can be solved or alternative estimation methods that bypass the nested numerical solution of the model.
2. The authors seem to treat the posterior beliefs about the unobserved state, which they denote by x_t in their paper, as observed by the econometrician. Their choice probability (27) depends on x_t and so does their likelihood (30). But the econometrician does not "observe" these beliefs: they are also latent. So it is not clear how the maximum likelihood is actually implemented unless the unobserved x_t random variables are margined out and depend only on the observable history $\{(z_t, a_t)\}$. The authors would have to clarify the discussion of this and explain how they get observations of x_t or else integrate out the x_t if unobserved in forming their likelihood.
The latter is a crucial point and so I would be willing to review a revision to make sure that the authors have satisfactorily addressed this concern before I would be willing to recommend this paper for publication. "

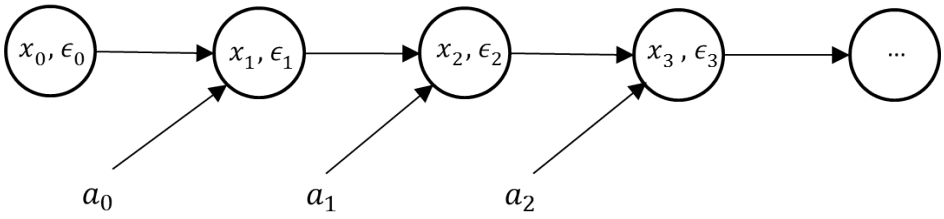
The above comments are also very important and should be addressed in full. Likewise, the attached referee reports include a wealth of useful suggestions and critical assessments for the manuscript under review. At

this point, it would be unproductive for me to offer further comments on the manuscript, given the detailed and excellent reports received.

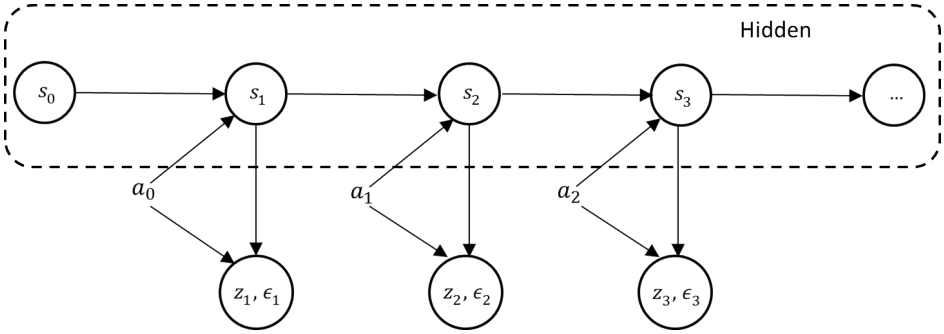
Consequently, my AE recommendation at the moment is for the authors to address the referee reports (including the additional comments above) in full. Once a revision along those lines is received, I will consult the three reviewers again. If they agree with the revisions and see a path forward for publication, I will offer final comments myself, if any, which I will try to keep minimal.

Response: We thank the AE for reading our paper, encouraging remarks, and proper guidance. It seems that the writing of our previous manuscript may not be clear enough to present our research and caused some serious confusion and unnecessary misunderstanding which we now want to clarify at front.

(i) The existing dynamic discrete choice (DDC) models assume that the decision-making agent makes decision based on a Markov decision process (MDP), where the agent can completely observe the system state. However, it is well-known that MDPs do not consider situations where systems are only partially observed to the agent due to measurement noise, lack of access to information, etc. In these partially observed scenarios, partially observable Markov decision processes (POMDPs) provide a more realistic framework for dynamic decision making. This fundamental POMDP assumption on the *decision-making agent* side set our work apart from the existing DDC literature.



(a) Existing MDP-based DDC model. At each stage, the decision-making agent can completely observe the system state $s_t = (x_t, \epsilon_t)$ and makes decision.



(b) Our POMDP-based DDC model where the system state s_t can not be directly observed. At each stage, the decision-making agent can only collect noisy observations (z_t, ϵ_t) to infer the underlying system state and makes decision on its belief about the system state.

Figure 1: the existing DDC models vs. our POMDP based DDC model.

In this regard, please also see Shi et al. (2020) and *Comment 1* of *Reviewer 3* which also provides a very nice insight and our response. In other words, we develop a POMDP-based DDC model to consider the situation where the data is generated via a POMDP process, rather than a MDP (or high order MDP) process.

Shi, C., Wan, R., Song, R., Lu, W, and Leng, L. (2020) "Does the Markov decision process fit the data: Testing for Markov property in sequential decision making", *Proceedings of the 37 th International Conference on Machine Learning*, Vienna, Austria, PMLR 119, 2020.

(ii) Consequently, the "unobservables" or "hidden states" has different meanings in our paper than in the existing DDC literature. DDC models with unobserved state variables have been extensively studied (as reviewers have pointed out, such as Kasahara and Shimotsu 2009, Hu and Shum 2012, Connault 2016), where "unobserved state variables" refer to the state component observed by the agent, but unobserved to the econometrician/researcher (e.g., ϵ_t in Figure 1(a)). If the econometrician/researcher could access these "unobservables", he/she would know the system state exactly as the decision making agent.

On the contrary, the "hidden state" in our model refers to that the system state s_t is hidden to the decision-making agent, and of course, to the econometrician/researcher. That is, (a) the agent does not know exactly the system state s_t and only can receive noisy observation (z_t, ϵ_t) on the system state, and (b) even if the econometrician/researcher could access the whole information data that the agent has, the econometrician/researcher would still not be able to figure out the system state.

The following table summarizes how our proposed model is different from the existing DDC models.

	Existing DDC Models (including DDCs with unobserved states)	Our Model
Decision-making Agent	MDP	POMDP
System State	fully observable to the agent but partially observable to the researcher	hidden to both the agent and the researcher
Agent makes decision on	system state at time t	*entire* information history up to time t to form a *belief* on system state
"unobservables"	state component hidden to the researcher	system state

Table 1: The difference between our model and the existing DDC models.

Reviewer #1:

This is an interesting paper. I have only two main comments.
1. The authors should cite and discuss the differences from their paper and the paper by Benjamin Connault "Hidden Rust Models" which also put forth the idea of modeling dynamic discrete choice models via POMDP and has identification results, though I think there are significant differences between the two papers. See `https://urldefense.com/v3/__https://www.sas.upenn.edu/*connault/hidden-rust-models.pdf__;fg!!KwNVnqRv!QIgaL1aJqi1JfXcja31rRl26Z8LaxFaECU9PVq1Xs-2WjTW6llXrRIyw_u3eg1RRpjqb$`

Response: We thank the reviewer for pointing out this paper. The reviewer is right that our paper is fundamentally different from the "Hidden Rust Models" of Connault. In the work of Connault, it is assumed that the decision-making agent makes decision according to a Markov decision process (MDP), namely, the decision-making agent can fully observe the system state. It is the econometrician/researcher who can only observe part of the system state. Please see the last paragraph of page 2, "agents make decisions exactly as in classical Rust models but the econometrician observes only part of the state variable" and Fig. 1 on page 7.
On the contrary, our POMDP-based hidden state model assumes that the decision-making agent cannot fully observe the system state, and hence the econometrician/researcher can not know the exact system state either. This is the crucial difference from the classical Rust model and its derivatives. We have updated our Fig. 1 to make this clear.
This difference also explains why we do *not* need to marginalize out the unobserved state as what Connault has done. In the latter, the econometrician/researcher needs to marginalize out the unobserved state because the agent is making decision on the true system state; however, in our model, the agent is making decision based on its *belief* x_t on the true system state, where the belief is a sufficient statistic for decision making. *While x_t is not directly observable to the econometrician/researcher, x_t can be calculated for each trajectory using Eqs. (14)-(15) for any given structural estimate θ.* The researcher can then update the estimates θ to best rationalize the data. Hence, marginalization is not required in our model.

2. The identification result in Theorem 6 strikes me as obviously wrong as stated. Adding a constant to all reward functions cannot change the optimal decision rule and as such there is an equivalence class of observationally equivalent reward functions that differ from the two reward function only by an additive constant. So at best the identification can only hold "up to an additive constant". The authors should also more rigorously define "identification" rather than just stating "there is only one reward function rationalizing the data". The term "rationalize" has not been defined and in the econometric literature "identification" is not defined relative to a given data set (which would have stochastic elements) but rather with respect to the probability limit of estimators so the result is essentially a result on the invertibility of the "reduce form" objects of the model to recover the underlying "structure" which in this case is the reward function r(s,a). The fact that the unobserved state s is not observed by either econometrician or agent controlling the system and the reward function could still be identified is surprising, but it does

rely on the very strong assumption that the cardinality of the unobserved state is known apriori. In many practical applications this will be considered to be an overly strong assumption.

Response: the reward under the reference action is fixed in our model. If the number of states is unknown, we could try different possible values and select the best one in practice. In many applications, it may be common knowledge. such as possible phases of a disease or a cancer. agree we need to define "identification". need to discuss.

3. However I am willing to recommend the paper for publication after a revision if the statement and proof of Theorem 6 can be fixed since the empirical application is interesting and plausible that the bus manager allocates buses that are more likely to be in good operating condition to high mileage bus routes. The authors do obtain a significant improvement in fit of the model from their approach, which outperforms the approach of Reich (Operations Research 2018) that allowed for serial correlation in the unobserved shocks affecting engine replacement decisions, and who could not find strong evidence in favor of serial correlation. The POMDP approach seems to have paid off in revealing an aspect of the data that I would agree has "also revealed economically meaningful features of Mr. Zucker's behavior ignored by the Rust's model."

Response: We sincerely appreciate the reviewer's time and the encouraging comments. Indeed, we worked diligently to accommodate all suggestions and we are thrilled with his/her assessment. Please accept our sincere gratitude for helping us improve our manuscript!

4. Note however the bus manager's name is Zurcher not Zucker.
Response: Done. Thank you for the good catch!

I filed my report yesterday perhaps a little too quickly and wanted to add this additional concern about the paper. While I am still in favor of ultimate publishing it if you and the other referees agree too, I think there are two main concerns about their approach that the authors need to discuss more clearly.
5. The POMDP approach essentially converts the problem into a higher dimensional problem where the decision maker carries a "posterior" distribution over the unobserved states in the model. This is a standard approach but the curse of dimensionality renders this approach practically infeasible if the unobserved state takes on more than a few values since then the posterior distribution is a point in an N dimensional simplex that makes it a continuous state variable requiring numerical approximation. In the Zurcher application the authors assumed that the unobserved latent state of the bus was either good or bad so there are only two possible values (the smallest possible) and so the simplex is the 1 dimensional simplex, the continuous unit interval. This is feasible and perhaps feasibility extends to maybe dimension 10 or 15, but for higher dimensional problems the authors need to discuss how the problem can be solved or alternative estimation methods that bypass the nested numerical solution of the model.

Response: The reviewer is right that POMDPs are hard: computing an optimal policy is PSPACE-complete (Papadimitriou and Tsitsiklis 1987) and finding an ϵ-optimal policy is NP-hard (Lusena et al. 2001). Nevertheless, many exact and approximate algorithms for POMDPs have been developed over decades, such as Sondik's one-pass algorithm (Smallwood and Sondik 1973), Witness algorithm (Littman 1994), and various point-based POMDP algorithms (see review in Shani et

al. 2013). As a result, POMDP based models have been successfully applied in numerous realistic problems (indicatively, Foka and Trahanias 2007, Byon et al. 2010, Bhatt and Krishnamurthy 2019).

As the main focus of the paper is to present a new dynamic discrete choice model using the POMDP framework, we thus will examine in detail the computational issues in our follow-up research. We expect that the existing efficient POMDP algorithms can be leveraged and extended to the dynamic discrete choice models. In addition, the recent development in computer science community also shows that artificial neural networks (e.g., Generative Adversarial Imitation Learning) can be very effective in address the high dimensional issue.

Papadimitriou, C. H., and Tsitsiklis, J. N. (1987). The complexity of Markov decision processes. *Mathematics of Operations Research*, 12(3), 441-450.
Lusena, C., Goldsmith, J., and Mundhenk, M. (2001). Nonapproximability results for partially observable Markov decision processes. *Journal of Artificial Intelligence Research*, 14, 83-103.
Smallwood, R., and Sondik, E. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21, 1071-1088.
Littman, M. L. (1994). The Witness algorithm for solving partially observable Markov decision processes. *Technical Report*. Department of Computer Science, Brown University.
Shani, G., Pineau, J., and Kaplow, R. (2013). A survey of point-based POMDP solvers. *Autonomous Agents and Multi-Agent Systems*, 27, 1-51.
Foka, A., and Trahanias, P. (2007). Real-time hierarchical POMDPs for autonomous robot navigation. *Robotics and Autonomous Systems*, 55(7), 561-571.
Byon, E., Ntaimo, L., and Ding, Y. (2010). Optimal maintenance strategies for wind turbine systems under stochastic weather conditions. *IEEE Transactions on Reliability*, 59(2), 393-404.
Bhatt, S., and Krishnamurthy, V. (2019). Adaptive polling in hierarchical social networks using Blackwell dominance. *IEEE Transactions on Signal and Information Processing over Networks*, 5(3), 538-553.

6. The authors seem to treat the posterior beliefs about the unobserved state, which they denote by x_t in their paper, as observed by the econometrician. Their choice probability (27) depends on x_t and so does their likelihood (30). But the econometrician does not "observe" these beliefs: they are also latent. So it is not clear how the maximum likelihood is actually implemented unless the unobserved x_t random variables are margined out and depend only on the observable history $\{(z_t, a_t)\}$. The authors would have to clarify the discussion of this and explain how they get observations of x_t or else integrate out the x_t if unobserved in forming their likelihood.

Response: We thank the reviewer for the careful read and the great question.
As we mentioned above (the difference between our model and the Connault's work), we do not need to marginalize x_t. The objective is to estimate the structural parameter θ. For each proposed θ, the belief x_t, although latent, can be calculated for each sample path via Eqs. (14)-(15). Thus, the whole likelihood is a function of θ, hence, can be maximized. This is another difference between

the classical Rust's model, where s_t can be directly observed. In our case, x_t can not be directly observed but can be calculated as a function of θ.

The latter is a crucial point and so I would be willing to review a revision to make sure that the authors have satisfactorily addressed this concern before I would be willing to recommend this paper for publication.
Response: a thank you note.

Reviewer #2:

This paper generalizes the single-agent dynamic discrete choice model of Rust (1987) to the case where the latent payoff-relevant states s_t are partially or imperfectly observed as z_t. The authors argue that the main elements of the model, including the static return function in each period and the state transition itself, can be identified under the classical conditional independence assumption on state transition (equation (8) on page 9), without further knowledge about how z_t and s_t are related. The authors also investigates analytically the consequences of using z_t instead of s_t. They then apply the model to Rust (1987) data, and conclude that route assignment can be further improved by assigning buses in worse conditions to routes with lower mileages.

1. Some earlier works (such as Kasahara and Shimotsu (2009), Hu and Shum (2012)) had dealt with dynamic discrete choice models with unobserved states or hidden Markov chains. The model in KS2009 and HS2012 can be reviewed as a special case where z_t is a subvector of s_t. Identification in that case requires additional restrictions on how the distribution of unobserved states is related to the observed ones. Could the model considered in this manuscript be observationally equivalent to those considered in these other papers? That is, will the current model in this manuscript deliver the same distribution of observable states and actions as those other papers, for certain specification of model elements? To motivate the generalization in this paper, it is necessary to discuss such differences.

Response: We thank the reviewer for pointing out the existing elegant work on MDP-based DDC models with unobserved states. We also sincerely apologize for the confusion caused by our first draft of the manuscript. In order to clear up the confusion, we summarize the main differences between MDP-based DDC models (including unobserved states) with our POMDP-based DDC model in Figure 1 and Table 1, which can be found on page 2-3 of this document, in the response to the AE's report.
Essentially, in the existing DDC models, the "unobserved states" refers to the state component "which are observed by the agent, but unobserved to the econometrician" (in the first paragraph of the introduction of HS2012). However, in our POMDP-based model, the "unobservable state" refers to the system state which is unobserved to the agent, and hence, is also unobserved to the econometrician/researcher. That is, the decision-agent makes decision according to a POMDP process, instead of a MDP process. Hence, our work is not equivalent to the mentioned ones.

2. The use of $x_t(s_t) = Pr(s_t|\zeta_t)$, with ζ_t being the history of observed action and states up to time t, is counter-intuitive and confusing. The author claims this avoids the issue of the increasing dimension of observed history ζ_t, and builds the identification results around a "sufficient statistics" statement in Theorem

1. But it appears to me that this is just a change of notation. After all, when it comes to identification, one still needs to condition on the full history of ζ_t. So how does the introduction of x_t helps to circumvent the issue with increasing dimension of the conditioning set as t increases? Why not use the time-homogenous assumption $Pr(s_t|\zeta_t) = Pr(s_t|z_t, a_t)$ for simplicity?

Response: We thank the reviewer for carefully reading and thinking our paper and for the opportunity of elaborating the implication of using belief $x_t(s_t) = P(s_t|\zeta_t)$.
The use of belief x_t is inherited from the POMDP literature and it is not a simple change of notation. In fact, the existence and the property of belief state x_t is a critical basis for the existing POMDP literature which we now summarize below:

(i) In POMDP, the agent has no direct access to the current system state and makes decisions based on his/her *entire* information history, including all actions implemented and all observations seen. Thus, we do not use the time-homogenous assumption $Pr(s_t|\zeta_t) = Pr(s_t|z_t, a_t)$ for simplicity.

(ii) The existing POMDP literature has shown that the belief, which is a probability distribution over all possible system states, provides the agent the same information for decision making as if the agent maintained the entire information history, and the belief process (also called information process) is Markovian. That is, the next belief x_{t+1}, only depends on the current belief x_t, the current action a_t, and new observation z_{t+1}. See Eqs. (14)-(15) in our paper and page 218 of White 1991.

(iii) Thus, a famous POMDP result is that the value function depends on ζ_t only through x_t and POMDP can be viewed as a belief-based MDP by treating the belief state as the system state. Moreover, if the state space S is finite, then the belief state lives in a finite-dimensional simplex which is also time invariant (on the contrary, the cardinality of $\{\zeta_t\}$ will explode as t increases).

The relationship between MDP and POMDP is illustrated in Figure 2 below. Please also see Smallwood and Sondik (1973), White (1991), and `https://cs.brown.edu/research/ai/pomdp/tutorial/pomdp-background.html` for more detailed discussion.

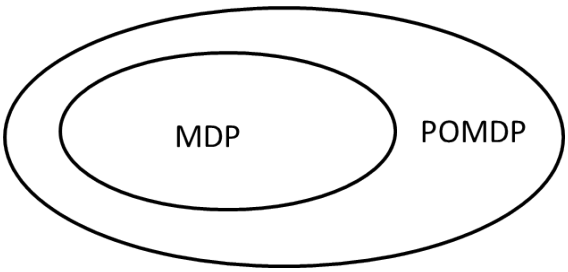


Figure 2: MDP vs. POMDP (Please also see Figure 2 of Steimle et al. 2019)

Steimle, L. N., Kaufman, D.L., and Denton B.T. Multi-model Markov Decision Processes. *Optimization-online*, Updated on August 10, 2019.

Smallwood R. and Sondik, E. (1973) The Optimal Control of Partially Observable Markov Processes over a Finite Horizon, *Operations Research*, 21(5), 1071-1088.
White, C.C. (1991) A survey of solution techniques for the partially observed Markov decision process. *Annals of Operations Research*, 32, 215–230.

3. Theorem 5 delivers one of the main identification results. In this proof, σ_0 is the observed conditional distribution, while both x and $P(z,a)$ are unknown parameters to recover. So I do not follow claims in the proof, where the sequence of "if and only if" statements lead to identification immediately. It should also be helpful to relate this proof to those in the literature (e.g. identification of models with unobserved states in KS2009, HS2012). In particular, it should be made clear in the text that how the current manuscript manages to identify a more general model with less restrictive assumptions? For readability, I also recommend the authors to write out $P(z,a)$,$x(s)$ explicitly as vectors or matrices with specified dimension in the proof.

Response: I am not sure i understand the comment. What does it mean "a more general model with less restrictive assumptions"? What assumptions?

4. Theorem 6 is another main identification result in the paper. It builds on earlier work in Magnac and Thesmar (2002). MT2002 essentially had a negative identification result (Proposition 2-(ii) and Corollary 3), and identification requires further restrictions such as the exogenous variation in state transition. The approach taken in the current manuscript is to "normalize" the static return r(x,a=0) to zero. This is known to be problematic for counterfactual analysis, see Kalouptsidi, Scott, and Eduardo Souza-Rodrigues (2019).

Response: this one can be serious.

5. The exercise in Section 5 is not surprising ex ante. Essentially, with the cardinality of the support being the same $|S| = |Z|$, one can alternatively frame the issue of unobserved state as a measurement error issue. In this sense the consequences mentioned in Section 5 is not surprising. A more useful insight would be how the consequences of mismeasurement varies with some analytical relation between s_t and z_t. I'd recommend the authors modify Section 5 to try to give more insights long these lines.

Response: What does analytical relation mean?

6. In general, the manuscript needs to cite more earlier work on dynamic discrete choice models from the econometrics literature. A partial list is provided in the references below. It helps to discuss the relation with these papers fully in the text.
References:
Hiroyuki Kasahara and Katsumi Shimotsu. Econometrica. 2009. Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices
Yingyao Hu and Matthew Shum. Journal of Econometrics. 2012. Nonparametric Identification of Dynamic Models with Unobserved State Variables.
Magnac and Thesmar. Econometrica. 2002. Identification of Dynamic Discrete Choice Process.
Myrto Kalouptsidi, Paul Scott and Eduardo Souza-Rodrigues. 2019. Quantitative Economics.

Response: doable

Reviewer #3:

This paper studies the identification and estimation of dynamic discrete choice structural models when some states are hidden to both the econometrician and the decision maker. The authors present this class of models, denoted Partially Observable Markov Decision Process, POMDP. The paper presents four main results. First, the authors characterize the solution and properties of the model, establishing a clear connection with similar properties in the standard Rust's model. Second, they show the identification of the structural parameters of a POMDP model under similar conditions as for the identification of Rust's model. Third, they characterize the misspecification bias of ignoring the partial observability of state variables. Finally, the authors illustrate these biases using simulated data from numerical experiments and using actual data from Rust (1987)'s application to bus engine data.

The paper is well written and I have enjoyed reading it. In my opinion, it contains original and interesting contributions. I have some comments and suggestions on the context of the paper, the relationship of the POMDP model to other dynamic discrete choice structural models, and on the empirical application in the paper.

Response: We sincerely thank the reviewer for all outstanding review comments and suggestions and we are very pleased with and encouraged by his/her assessment.

1. MDP with serially correlated unobservables. The authors compare their POMDP model to a standard Rust's model without serially correlated unobservables. However, there is a substantial literature that has extended Rust's model to incorporate serially correlated unobservables, including Norets (2009), Arcidiacono and Miller (2011), or Aguirregabiria, Gu, and Luo (2020), among others.

In opinion, this literature is very relevant to understand the contribution of this paper. The POMDP model and Rust's models with serially correlated unobservables are trying to explain the same features of the data: the fact that the observable state at period t is not a sufficient statistic in the CCPs or in the transition probability of the state variable such that lagged values of the observable state and of the agent's decision have also explanatory power in these probabilities. Interestingly, the two models provide very different explanations for this empirical finding. In a Rust's model with serially correlated unobservables, the researcher considers that the decision maker has more information than the researcher, and that this additional information is serially correlated. In contrast, in a POMDP model, the researcher considers that (excluding the i.i.d. logit shocks) she(he) has the same information as the decision maker, but there are variables affecting the agent's reward function that are unobservable both to the researcher and the decision maker. The two explanations are plausible, and definitively, they are not mutually exclusive. Therefore, a relevant question is whether we can distinguish between the two, and if not, under which conditions one modelling approach can be better than the other.

I understand that studying the separate identification of serially correlated unobservables and the POMDP model is beyond the scope of this paper. It is also a model-specific question because it depends on the specification of the stochastic process of the unobservables in the two models. However, the comparison of the two models can be very relevant in empirical applications. The applied researcher who finds evidence against the Markov structure in the standard Rust's model can choose between these two alternative models. I have

the impression that serially correlated unobserved state variables (observed by the agent) will be a reasonable competing hypothesis to POMDP in basically any empirical application. Therefore, the applied researcher using a POMDP needs to consider serially correlated unobservables as a serious alternative hypothesis.
I will come back to this point below in the context of the empirical application in the paper. However, I think that it also deserves some comments in the Introduction.

Response: We thank the reviewer for his/her deep thinking which is also very intriguing and inspiring to us.
We totally agree with the reviewer that both the POMDP-based model and the Rust's models with serially correlated unobservables can be possible in an empirical application. Comparing these two models is definitely an interesting and very important future research task. In a recent paper of Shi et al. (2020), the authors developed a model selection procedure to test the Markov assumption to decide which models among MDP, high-order MDP, and POMDP fits best for a given dataset (see Figure 3 below).
Shi, C., Wan, R., Song, R., Lu, W, and Leng, L. (2020) "Does the Markov decision process fit the data: Testing for Markov property in sequential decision making", Proceedings of the 37 th International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.

2. MDP with Bayesian learning. There is a literature on dynamic discrete choice structural models where agents have uncertainty about some primitives and they use some form of learning (typically Bayesian learning). Ching, Erdem, and Keane (2017) provide a recent survey of this literature with particular attention to dynamic demand models where consumers learn over time about some product characteristics. The structure of this type of models is very similar to the structure of the POMDP model presented in this paper. In fact, under Bayesian learning, I think that they are basically the same model, where the object of learning is s_t and the sequence of signals that the agent receives is $\{z_t\}$. In any case, the two models are closely related. I think that the authors need to mention this literature and comment on its relationship to the POMDP model.

Response:

3. POMDP and the standard Rust's model. I think that it would be worthwhile to clarify an aspect of the relationship between the proposed POMDP and the standard Rust' MDP model. It could be argued that POMDP is a particular case of the Rust's MDP. This is because the primitives in Rust's MDP - the reward function $r(z_t, a_t)$ and the transition probability function $P(z_{t+1}|z_t, a_t)$ - can be always interpreted as expectations conditional on (z_t, a_t) of other functions that depend on variables which are unobservable to the decision maker: that is, $r(z_t, a_t) = \int r(s_t, a_t)dp(s_t|z_t, a_t)$ and $P(z_{t+1}|z_t, a_t) = \int P(z_{t+1}|s_t, a_t)dp(s_t|z_t, a_t)$. In Rust's model we are interested in the functions $r(z_t, a_t)$ and $P(z_{t+1}|z_t, a_t)$ instead of $r(s_t, a_t)$, $P(z_{t+1}|s_t, a_t)$, and $p(s_t|z_t, a_t)$.
I can see two reasons why this argument is not exactly valid. First, in some applications, the functions that depend on the latent state s_t are the actual structural objects, and the answers to some empirical questions may require the estimation of these functions and not only of the "integrated" versions $r(z_t, a_t)$ and

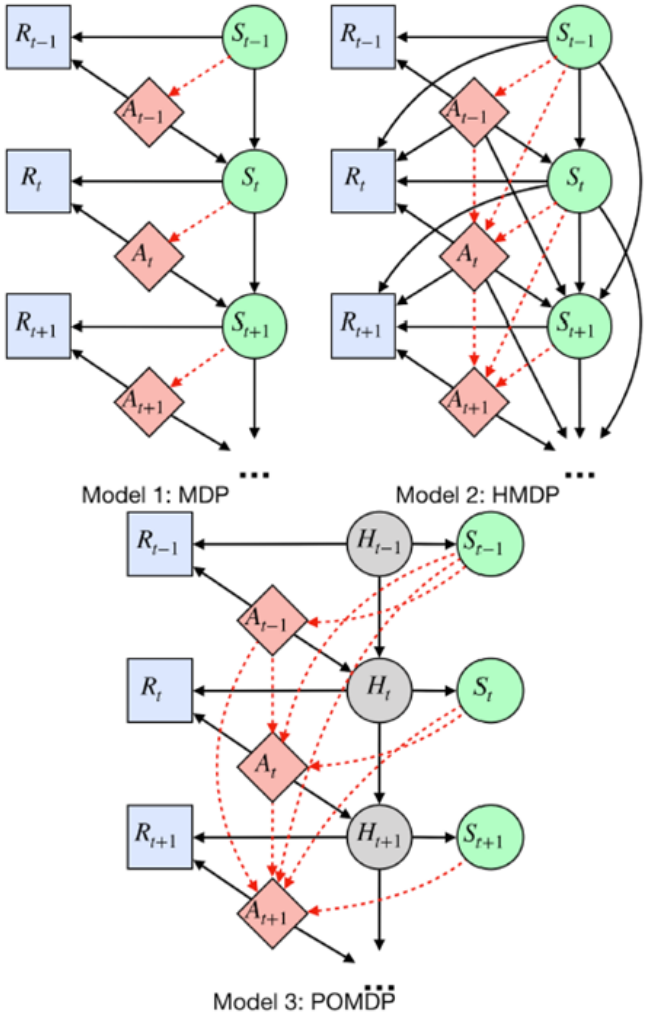


Figure 3: Diagrams for MDP, High-order MDP (HMDP), and POMDP (from Shi et al. 2020)

$P(z_{t+1}|z_t, a_t)$. *Second, in the POMDP proposed in this paper the observable state* z_t *is not restricted to be Markovian.*

Related to this second point, in Rust's model one can define the observable state z_t *in such a way that it can accommodate non-Markovian state variables. For instance, if the state vector* z_t *follows a Markov process with n-periods memory, one can always redefine the state variable as* $x_t = (z_t, z_{t-1}, ..., z_{t-n})$ *such that* x_t *follows a first order Markov process. Based on this observation, one could argue that the second point in the paragraph above is not a substantial difference between POMDP and Rust's MDP.*

I think that these points on the relationship between POMDP and Rust's model deserve some discussion in this paper.

Response: We are really excited by the reviewer's deep knowledge and insightful comments. The reviewer is exactly right that our model and the Rust's MDP model is not equivalent due to the two

points that the reviewer have mentioned. We just want to point out that the observation process of a POMDP can be non-Markovian no matter how many past observations are included. Indeed, if there exists a n such that $x_t = (z_t, z_{t-1}, ..., z_{t-n})$ and x_t follows a first order Markov process, then the process is a High-order MDP whose redefined state space could be very large; See Figure 3 above. Such n may not exist in a data generated by a POMDP process (see Shi et al. 2020).

4. The description of the POMDP version of Rust's model – in section 6.1 – never mentions the specification assumptions for the stochastic process of the latent variable $s_t \in \{0, 1\}$ that represents whether the state of a bus engine is good $(s_t = 0)$ or bad $(s_t = 1)$. The authors specify the probability distribution of the (observed) incremental mileage conditional on the (unobserved) state of the bus engine, i.e., $P(z_{t+1} - z_t|s_t, a_t)$. However, the paper seems silent about the stochastic process of the latent variable s_t, or more generally, about $P(s_{t+1}|s_t, z_t, a_t)$. This probability seems to me fundamental for the estimation of the parameters of the POMDP presented in Table 1. Of course, assuming that s_t follows an i.i.d. Bernoulli distribution would be extremely unrealistic. A simple model would be one where s_t does not depend on cumulative mileage and follows a Markov process with transition probabilities $P(s_{t+1} = 0|s_t = 0, a_t = 0) = p_{00} < 1$, $P(s_{t+1} = 1|s_t = 1, a_t = 0) = 1$,and $P(s_{t+1} = 1|s_t, a_t = 1) = 0$. However, the assumption that the state of a bus engine does not depend on its cumulative mileage seems also very unrealistic and it can have important implications on the estimation results. The authors need to make explicit and transparent their assumptions about the stochastic process of the latent variable s_t.

Response: Thank you for the great catch. The hidden system dynamics is now included in the revised manuscript. Note that in the POMDP framework, the internal system state follows exactly a MDP; the relationship between system state s_t and mileage z_t is related through the observation probabilities, i.e., $P(z_{t+1} - z_t|s_t, a_t)$. The Bayes' rule will update the distribution over system state x_t conditional on mileage z_t ("switch the order"), hence, the belief on good state will tend to decrease as mileage increases.

5. Empirical application to Rust data (i). As the authors point out, there is strong evidence of serial correlation in the incremental mileage in Rust's bus engine data. Other papers have pointed out this before. Also, other papers have shown evidence of serial correlation in bus engine replacement decisions, after controlling for cumulative mileage. Most previous studies have interpreted this empirical finding as evidence of serially correlated unobserved state variables. For instance, Aguirregabiria, Gu, and Luo (2020) consider that there are different types of bus engines according to their maintenance cost (or probability of failure). The decision maker knows the type of each bus engine, but this is unobservable to the researcher. They find evidence for this type of unobserved heterogeneity using Rust's bus engine data. The POMDP model is a very interesting alternative explanation. I think that it would be very interesting to see the relative performance of the two alternative explanations. I suggest that the authors include also estimates of a model with permanent unobserved heterogeneity – e.g., two types of buses – and compare its predictive power to the POMDP.

Response: We thank the reviewer for the great suggestion. POMDP-based DDC with permanent unobserved heterogeneity is definitely an interesting research topic. In fact, this is exactly one of our ongoing research under exploration (including identifiability). We believe this topic deserves a

separate comprehensive study (also due to the recommended 32-page limit); hence, we will report our modeling and analysis approach, structural results, and empirical findings on heterogeneity in our next paper. We thank the reviewer again for pointing out the relevant paper and for the help in shaping our research direction.

6. Empirical application to Rust data (ii). The authors compare the goodness of fit of the POMDP model relative to Rust's model and specification in his 1987 Econometrica paper. This specification, imposes the restriction that the probability distribution of incremental mileage $z_{t+1} - z_t$ is independent of cumulative mileage z_t. As the authors point out, this restriction is clearly (and trivially) rejected by the data. Of course, this restriction is not a characteristic of Rust's MDP models and this type of model can be estimated using a flexible specification of the transition probability function $P(z_{t+1} - z_t|z_t, a_t)$ that allows for dependence with respect to z_t.
I wonder if most of the 17.7% improvement in the goodness of fit of the POMDP comes from this restriction in Rust's bus engine model. I suggest that the authors include also the estimation of a version of Rust's MDP model that allows for a flexible specification of the transition probability $P(z_{t+1} - z_t|z_t, a_t)$.

Response: This is a great suggestion. The results of using MDP-based DDC where unobservables follow an AR(1) process has been nicely reported and summarized in Reich (2018). The author stated that he/she did not observe improvement in the goodness of fit. Under the extreme value type I assumption, there are some changes in parameter estimates but the the ratio of engine replacement cost to the regular maintenance cost parameter is relatively the same as in the Rust's (Please also see *Comment 3* from Reviewer 1).
As these results have been published and are readily available, we decide to refer potential readers who are also interested in this comparison to the paper of Reich. Furthermore, we believe it may be more reasonable to have a detailed comparison study among Rust's MDP model, MDP-based DDC models with serially correlated unobservables, POMDP-based DDC model, and POMDP-based DDC model with serially correlated unobservables (to be developed) to be crystal clear on the root causes of these estimation differences. This is a clearly important research question and should be pursued once relevant models are established.
Reich, G. (2018). Divide and Conquer: Recursive Likelihood Function Integration for Hidden Markov Models with Continuous Latent Variables. *Operations Research*, 66(6):1457-1759.

7. Empirical application to Rust data (iii). As shown in equations (6) and (30) in the paper, the log-likelihood functions of the Rust's MDP model and of the POMPD model can be written as the sum the log-likelihood from choice data and the log-likelihood from the transition of the observable state variables. It would be interesting to decompose the improvement in the goodness of fit of the POMDP model relative to Rust's MDP into the improvements in the log-likelihood of choice data and in the log-likelihood of transition data. This is relevant because – for this class of models and applications – we are typically more interested in explaining choice data. It is quite intuitive that the POMDP should do better that Rust's MDP in terms of explaining the transition data, but it is not so obvious that it necessarily improves the goodness of fit for the choice data.

Response: agreed and doable but $P(z, a)$ is also meaningful. our $P(z, a)$ shows optimal route assignment behavior, in addition to CCP.

8. Empirical application to Rust data (iv). Please, include standard errors of the parameter estimates in Tables 1 and 2.

Response: Can we do this? Not sure how to do it without further distribution assumptions.

References

Aguirregabiria, V, J. Gu, and Y. Luo (2020): "Sufficient statistics for unobserved heterogeneity in dynamic structural logit models," Journal of Econometrics. Forthcoming.

Arcidiacono P., and R. Miller (2011): "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," Econometrica, 79(6), 1823-1867.

Ching, A., T. Erdem, and M. Keane (2017): "Empirical models of learning dynamics: A survey of recent developments," Handbook of marketing decision models (pp. 223–257). Cham: Springer.

Norets, A. (2009): "Inference in dynamic discrete choice models with serially correlated unobserved state variables," Econometrica, 77(5), 1665-1682.

Response: It is our honor and privilege to communicate such excellent reviewers. These insightful and constructive comments not only significantly help us in improving the current manuscript, they also help us promote our thinking and shape our research agenda. We sincerely thank these outstanding reviewers.