RATIONALITY & COMPETITION
CRC TRR 190

Robust Decision-Making under Risk and Ambiguity

Maximilian Blesch (HU Berlin, DIW Berlin)
Philipp Eisenhauer (Amazon)

Discussion Paper No. 463

November 22, 2023

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Georg Weizsäcker, Ph.D., Humboldt University Berlin, 10117 Berlin, Germany
info@rationality-and-competition.de

Robust decision-making under risk and ambiguity*

Maximilian Blesch[1, 2] and Philipp Eisenhauer[3]

[1]Humboldt Universität Berlin
[2]German Institute of Economic Research
[3]Amazon

November 2023

Economists often estimate economic models on data and use the point estimates as a stand-in for the truth when studying the model's implications for optimal decision-making. This practice ignores model ambiguity, exposes the decision problem to misspecification, and ultimately leads to post-decision disappointment. Using statistical decision theory, we develop a framework to explore, evaluate, and optimize robust decision rules that explicitly account for estimation uncertainty. We show how to operationalize our analysis by studying robust decisions in a stochastic dynamic investment model in which a decision-maker directly accounts for uncertainty in the model's transition dynamics.

JEL Codes D81, C44, D25

Keywords decision-making under uncertainty, robust Markov decision process

*Corresponding author: Maximilian Blesch, maximilian.blesch@hu-berlin.de. Support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged. We thank Anton Bovier, Annica Gehlen, Lena Janys, Ken Judd, John Kennan, Simon Merkt, Sven Rady, Gregor Reich, John Rust, Jörg Stoye, and Rafael Suchy for numerous helpful discussions. We thank Annica Gehlen for her outstanding research assistance. We are grateful to the Social Sciences Computing Service (SSCS) at the University of Chicago for the permission to use their computational resources.

Contents

1. Introduction **3**

2. Statistical decision theory **7**

2.1. Decision problem 7

2.2. Urn example 9

3. Data-driven robust Markov decision problem **12**

3.1. Setting 13

3.2. Solution 15

3.3. Evaluation 17

4. Bus replacement problem **18**

4.1. Setting 19

4.2. Ex-post analysis 22

4.3. Ex-ante analysis 26

5. Conclusion **28**

A. The robust contraction mapping **A-1**

1. Introduction

Decision-makers often confront uncertainties when determining their course of action. For example, individuals save to cover uncertain medical expenses in old age (French and Song, 2014). Firms set prices in an uncertain competitive environment (Ilut, 2020), and policy-makers face uncertainties about future costs and benefits when voting on climate change mitigation efforts (Barnett et al., 2020). We consider the situation in which a decision-maker posits a collection of economic models to inform his decision-making process. Each model formalizes the relevant objectives and trade-offs involved and provides an implicit rule for optimal decisions. Uncertainty is limited to risk for a given model, as the model induces a unique probability distribution over possible future outcomes. However, a decision-maker also faces model ambiguity as the true model within the collection remains uncertain (Arrow, 1951; Knight, 1921).

It is the standard practice in economics to estimate models on data and use the point estimates as a stand-in for the truth when studying the model's implications and optimal decision-making.[1] This approach ignores model ambiguity, resulting from the remaining parametric uncertainty after the estimation, and opens the door for the misspecification of the decision problem. As-if decisions, decisions that are optimal if the point estimates are accurate (Manski, 2021), often turn out to be very sensitive to misspecification (Smith and Winkler, 2006). This creates the need for robust decisions that explicitly account for the parametric uncertainty. Robust optimization offers tools to construct robust decisions, by using a worst-case realization of the parameters, instead of the point estimates when selecting the optimal decisions (Ben-Tal et al., 2013). Increasing the level of robustness comes at the price of reduced performance in all other scenarios when the true parameters are not as extreme as the worst-case parameters. The level of robustness used to construct the robust decision is a choice by the decision-maker and finding the optimal level of robustness is crucial for the performance of the decision (Gotoh et al., 2021).

Our contribution is to outline a rigorous framework to determine the optimal level of robustness. We do so by applying statistical decision theory to evaluate the performance of robust decisions and show how to operationalize our methodology in a seminal model of dynamic decision-making. In particular, we translate the problem of choosing the optimal level of robustness into finding the optimal statistical decision function (SDF), a task that is at the center of statistical decision theory. An SDF maps each potential data realization into feasible decisions (Manski, 2021) and statistical decision theory provides a systematic, transparent and

[1] See examples in labor economics (Adda et al., 2017; Blundell and Shephard, 2012), industrial organization (Hortaçsu et al., 2019; Igami, 2017), and international trade (Bagwell et al., 2021; Eaton et al., 2011).

well-researched framework for assessing the performance of different SDFs as well as selecting the optimal SDF from any class of SDFs (Berger, 2010). At the core of our contribution lies the interpretation, that a parametrized model of decision-making is a single SDF. Data-driven robust optimization (Bertsimas et al., 2018) directly generates a class of SDFs with different levels of robustness, allowing us to employ the analysis framework of statistical decision theory to pick the optimal level of robustness. In doing so, we bring together and extend research in economics and operations research by using econometric models in complex decision problems (Bertsimas and Thiele, 2006; Manski, 2021).

In our application, we revisit Rust's (1987) seminal bus replacement problem. Model ambiguity is particularly consequential in dynamic models where the impact of erroneous decisions accumulates over time (Mannor et al., 2007). In the model, the manager Harold Zurcher implements a maintenance plan for a fleet of buses that maximizes his expected discounted utility. He faces uncertainty about the future mileage utilization of the buses but has data on past utilization available to inform his decisions. While Rust's (1987) original goal was to describe the investment behavior of Harold Zurcher, our analysis is a prescriptive approach. Prescriptive decision analysis seeks to develop reasonable, even if not optimal, ways to make decisions under uncertainty (Manski, 2021). We choose Rust's (1987) seminal model, as it serves in a variety of research as an illustrative example (Christensen and Connault, 2023; Iskhakov et al., 2016; Reich, 2018; Su and Judd, 2012).

The bus replacement problem is typically modeled as a standard Markov decision problem (MDP), and the point estimates for the mileage utilization are treated as-if they correspond to the true parameters. The solution of the MDP is an as-if decision rule that is optimal given the estimates. This approach ignores model ambiguity. From the perspective of statistical decision theory, an MDP is just one particular example of an SDF suitable for analyzing the bus replacement problem. We, on the other hand, consider a whole class of SDFs called robust Markov decision problems (RMDP) (Ben-Tal et al., 2009). RMDPs generalize the standard MDP, as they consider a whole set of distributions for the transition dynamics collected in an ambiguity set. The solution of an RMDP is a robust decision rule that is optimal under a worst-case scenario for all mileage utilization distributions inside the ambiguity set. We follow the literature and construct the ambiguity set so that it contains all distributions we cannot reject with a certain level of confidence $\omega \in [0, 1]$ around the point estimates under any possible realization of the data (Ben-Tal et al., 2013). The size of the ambiguity set is a choice by the decision-maker and determines the level of robustness. Given the realization of the data, the robust decision rule based on the solution of an RMDP is always conditional on the specified level of robustness. Each choice of ω defines a different RMDP, and applying the toolkit of statistical decision theory allows us to determine the optimal level of robustness ω^* within the whole class of SDFs.

To do so, we compare the performance of RMDPs with varying levels of robustness under different decision-theoretic criteria. We consider the situation before any data on mileage utilization is available and implement an ex-ante decision-theoretic analysis. We explore the performance of robust decision rules over the whole probability simplex and are thus able to determine the optimal level of robustness. Throughout, we compare robust and as-if decision rules, as the standard Markov decision problem remains one SDF within the broader class we consider.

Figure 1 stresses the point that each RMDP is a different SDF that characterizes robust decisions for any realization of the data. Here, for example, we consider two RMPDs with different levels of robustness – ω_1 and ω_2 – that, once data is realized, lead to different decision rules. This situation creates the need to compare their performance under alternative decision-theoretic criteria.

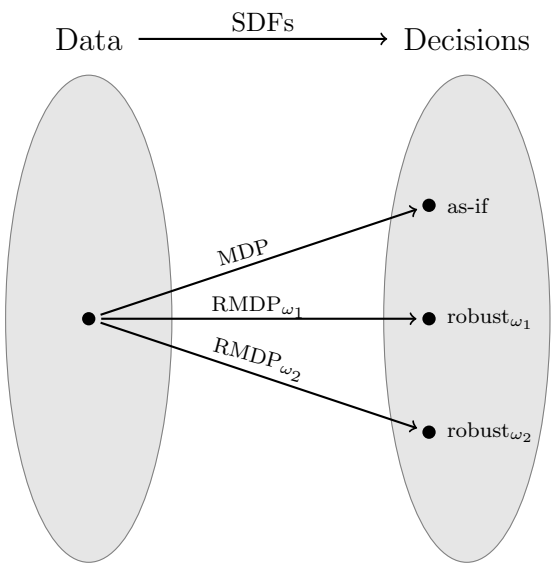


Figure 1: RMDPs as statistical decision functions

Our insight to evaluate robust decisions using statistical decision theory applies to the whole literature on data-driven robust optimization. There exists a growing number of empirical applications of data-driven robust decision-making in a variety of settings, including portfolio decisions (Jin et al., 2020; Zymler et al., 2013), elective admission to hospitals (He et al., 2019; Meng et al., 2015), the timing of medical interventions (Goh et al., 2018; Kaufman et al., 2017), and managing the production of renewable energy (Alismail et al., 2018; Samuelson and Yang, 2017).[2] Additionally, robust optimization has become a prominent tool to advance research in machine and in particular reinforcement learning (Blanchet et al., 2019; Bruns-Smith and

[2]See the recent surveys by Bertsimas et al. (2018), Keith and Ahner (2021), and Rahimian and Mehrotra (2019) for numerous additional examples.

Zhou, 2023; Blanchet et al., 2023).

Despite its broad field of application, the existing literature only offers limited guidance on choosing the optimal level of robustness. At the most basic level, the recommendations range from simply advocating a high level of robustness (Ben-Tal et al., 2013; Delage and Ye, 2010; Duchi et al., 2021; Lam, 2016) to choosing a level of robustness that ensures a pre-specified worst-case performance (Brown et al., 2012). These approaches ignore the trade-off between a performance guarantee under a worst-case scenario and reduced performance in all other cases. Most recently, and much closer to our approach, Gotoh et al. (2021) put the robustness trade-off front and center. Adopting ideas from the machine learning literature, they calibrate the level of robustness by trading off the mean and variance in the out-of-sample performance of a robust decision rule. However, their approach restricts attention to the neighborhood of a realized point estimate. Thus, their analysis is ex-post and does not aggregate performance over all possible true parameterizations.

At the same time, in econometrics, there is a burgeoning interest in assessing the sensitivity of findings to model or moment misspecification.[3] Our work is related to Jørgensen (2023), who develops a measure to assess the sensitivity of results by fixing a subset of parameters of a model before the estimation of the remaining parameters. Our approach differs as we directly incorporate model ambiguity in the design of the decision-making process and assess the performance of a decision rule under misspecification of the decision environment. As such, our focus on ambiguity faced by decision-makers about the model draws inspiration from the research program summarized in Hansen and Sargent (2016) that tackles similar concerns with a theoretical focus. We complement recent work by Saghafian (2018), who works in a setting similar to ours, but does not use statistical decision theory to determine the optimal robust decision rule. In ongoing work, Eisenhauer et al. (2021) use statistical decision theory to structure policy decisions in light of uncertainty about counterfactual policy predictions due to the remaining model ambiguity after the estimation of a model. While they conduct an ex-post evaluation of alternative policy proposals using decision-theoretic criteria, we perform a proper ex-ante analysis of competing decision rules. We evaluate each rule's performance under all possible parameterizations of the model and directly account for the model ambiguity in their construction. In addition, we contribute to the work on optimal treatment allocation started in Manski (2004) and Manski (2009), which characterizes the structure of optimal statistical decision functions and provides (asymptotic) bounds on their performance (Hirano and Porter, 2009; Stoye, 2009; Tetenov, 2012; Stoye, 2012; Kitigawa et al., 2018).

[3]See for example Andrews et al. (2020), Andrews et al. (2017), Armstrong and Kolesár (2021), Bonhomme and Weidner (2021), Chernozhukov et al. (2022), Christensen and Connault (2023), and Honoré et al. (2020).

The structure of the remaining analysis is as follows. In Section 2, we present statistical decision theory as our framework to compare decision rules. We then set up a canonical model of a data-driven robust Markov decision problem in Section 3 and outline the decision-theoretic determination of the optimal level of robustness. Section 4 presents our analysis of the robust bus replacement problem. Section 5 concludes.

2. Statistical decision theory

We now show how to compare as-if decision-making to its robust alternatives using statistical decision theory. We first review the basic setting and then turn to a classic urn example to illustrate some key points.

2.1. Decision problem

We study a decision problem in which the consequence $c \in \mathcal{C}$ of various alternative actions $a \in \mathcal{A}$ depend on the parameterization $\theta \in \Theta$ of an economic model. A consequence function $\rho : \mathcal{A} \times \Theta \mapsto \mathcal{C}$ details the consequence of action a under parameters θ:

$$c = \rho(a, \theta).$$

A decision-maker ranks consequences according to a utility function $u : \mathcal{C} \mapsto \mathbb{R}$, where higher values are more desirable. The structure of the decision problem $(\mathcal{A}, \Theta, \mathcal{C}, \rho, u)$ is known, but the true parameterization θ_0 is uncertain. As a result, the consequences of a particular action are ambiguous. An observed sample of data $\psi \in \Psi$, however, provides a signal about the true parameters, as P_θ – the sampling distribution of ψ – differs by θ. A statistical decision function (SDF) $\delta : \Psi \mapsto \mathcal{A}$ is a procedure that determines an action for each possible realization of the sample.

In our application, we study the bus replacement problem with unknown future mileage utilizations. The decision problem is dynamic, so a decision-maker acts by committing to a plan that specifies whether to maintain or replace a bus in any possible future scenario. The consequences of executing a particular plan are a stream of maintenance costs, aggregated by its discounted sum of utilities. A plan's total utility is determined by the true distribution of the bus mileage utilization. The optimal decision rule based on an RMDP depends on the observed sample of past mileage transitions as the sample informs the construction of the ambiguity set. So, each RMDP is one example of an SDF for the bus replacement problem. We consider many RMDPs with varying levels of robustness and thus analyze a whole class of SDFs.

Statistical decision theory provides the framework to compare the performance of alternative decision functions $\delta \in \Gamma$. The utility achieved by any δ is a random variable before realizing ψ. Thus, Wald (1950) suggests measuring the performance of δ at all possible parametrizations θ by computing the expected utility with respect to its induced sampling distribution P_θ:

$$\mathrm{E}_\theta \left[u \left(\rho(\delta(\psi), \theta) \right) \right] = \int_\Psi u \left(\rho(\delta(\psi), \theta) \right) dP_\theta(\psi).$$

In general, no single decision function yields the highest expected utility for all possible parameterizations. In this case, determining the best decision function δ^* is not straightforward. Still, decision theory proposes various criteria (Gilboa, 2009; Marinacci, 2015) to aggregate the performance of a decision function at all possible parameterizations. At the most fundamental level, any decision function is admissible if another function does not exist, whose expected utility is always at least as high. In most cases, several decision functions are admissible, and thus additional optimality criteria are needed. Our analysis explores three of the most common decision criteria: (1) maximin, (2) minimax regret, and (3) subjective Bayes.

Following the maximin decision criterion (Wald, 1950), we determine the optimal decision function by computing the minimum expected performance for each decision function over all points in the parameter space. We then choose the one with the highest minimum performance. Stated concisely,

$$\delta^* = \arg\max_{\delta \in \Gamma} \min_{\theta \in \Theta} \mathrm{E}_\theta \left[u(\rho \left(\delta(\psi), \theta \right)) \right].$$

For the minimax regret criterion (Niehans, 1948), we compute the maximum regret for each decision function over all points in the parameter space. The regret of choosing a decision function for any realization of θ is the difference between the maximum possible performance, where the true parameterization informs the decision, and its actual performance. We then select the decision function with the lowest maximum regret. Thus, the minimax regret criterion solves:

$$\delta^* = \arg\min_{\delta \in \Gamma} \max_{\theta \in \Theta} \underbrace{\left[\max_{a \in \mathcal{A}} u(\rho \left(a, \theta \right)) - \mathrm{E}_\theta \left[u(\rho \left(\delta(\psi), \theta \right)) \right] \right]}_{\text{regret}}.$$

Subjective Bayes (Savage, 1954) requires a subjective probability distribution f_θ over the parameter space. Then, we select the decision function with the highest expected subjective utility:

$$\delta^* = \arg\max_{\delta \in \Gamma} \int_\Theta \mathrm{E}_\theta \left[u(\rho \left(\delta(\psi), \theta \right)) \right] df_\theta.$$

2.2. Urn example

We now illustrate the key ideas that allow us to compare as-if and robust decision-making using statistical decision theory with an urn example. As in our empirical application, we study a whole class of statistical decision functions. We first compare the performance of two distinct alternatives and then determine the optimal function within the class.

We consider an urn with black b and white w balls where the true share of black balls θ_0 is unknown. In this example, the action constitutes a guess $\tilde{\theta}$ about θ_0 after drawing a fixed number of n balls at random with replacement. The parameter and action space both correspond to the unit interval $\Theta = \mathcal{A} = [0, 1]$.

If the guess matches the true share, we receive a payment of one. On the other hand, the payment is reduced by the squared error in case of a discrepancy. Thus, the consequence function takes the following form:

$$\rho(\tilde{\theta}, \theta_0) = 1 - (\tilde{\theta} - \theta_0)^2.$$

Going forward, we assume a linear utility function and directly refer to the monetary consequences of a guess as its utility. The sample space is $\Psi = \{b, w\}^n$ where a sequence $(b, w, b, \ldots, b)$ of length n is a typical realization of ψ. The observed number of black balls r among the n draws in a sample ψ provides a signal about θ_0. The sampling distribution for the possible number of black balls R takes the form of a probability mass function (PMF):

$$\Pr(R = r) = \begin{pmatrix} n \\ r \end{pmatrix} (\theta_0)^r \, (1 - \theta_0)^{n-r}.$$

Any function $\delta : \{b, w\}^n \mapsto [0, 1]$ is a possible statistical decision function.

We focus on the following class of decision functions $\delta \in \Gamma$, where each λ indexes a particular decision function:

$$\delta_\lambda(\psi) = \lambda \left(\frac{r}{n}\right) + (1 - \lambda) \left(\frac{1}{2}\right), \qquad \text{for some} \quad 0 \leq \lambda \leq 1.$$

The empirical share of black balls r/n in the sample ψ provides the point estimate $\hat{\theta}$. The decision functions in Γ specify the guess as a weighted average between the point estimate and the midpoint of the parameter space. The larger λ is, the more weight is put on the point estimate. At the extremes, the guess is either the point estimate ($\lambda = 1$) itself or fixed at 0.5 ($\lambda = 0$).

We begin by comparing the performance of the two decision functions with $\lambda = 1$ and $\lambda = 0.9$. We refer to the former as the as-if decision function (ADF), as it announces the point estimate as if it is the true parameter. For reasons that will later become clear, we identify $\lambda = 0.9$ as the robust decision function (RDF). Following Wald (1950), we evaluate their relative performance by aggregating the vector of expected payoffs over the unit interval using the different decision-theoretic criteria. We set the number of draws n to 50.

Figure 2 shows the sampling distribution of the number of black balls R and the associated payoff of following the two decision functions for each possible draw. The true, but unknown, share in this example is 40%, i.e. $\theta_0 = 0.4$. The RDF outperforms the ADF for realizations of the point estimates smaller than its true value due to the shift towards 0.5. At the same time, the ADF leads to a higher payoff at the center of the distribution.

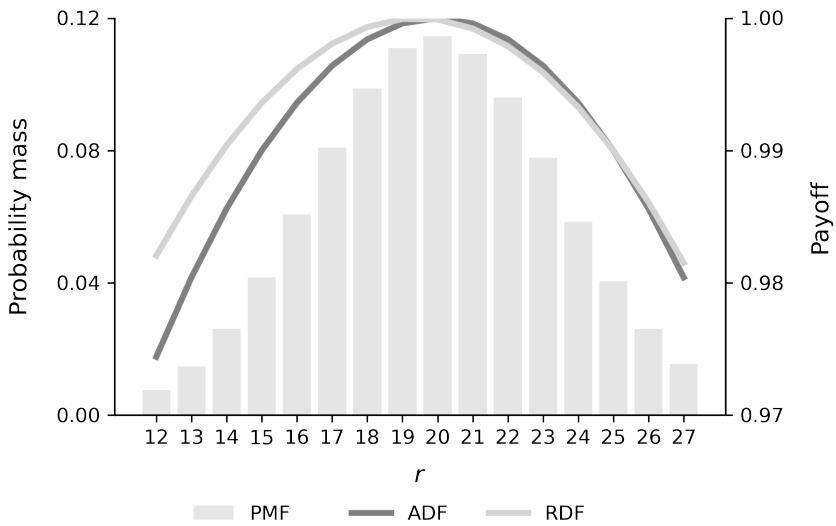
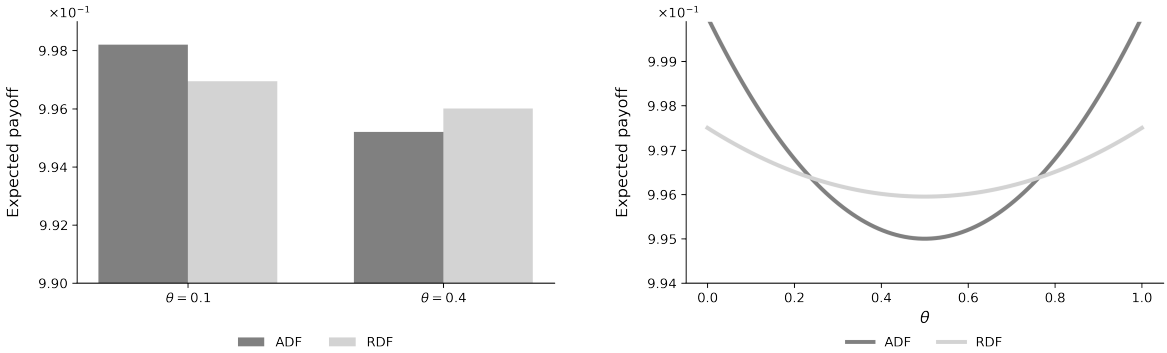


Figure 2: Calculating the expected payoff

Figure 3 shows the expected payoff for varying shares θ of black balls in the urn. On the left, we show the expected payoff at two selected points. While the ADF performs better than the RDF at $\theta = 0.1$, the opposite is true at $\theta = 0.4$. Thus, both decision functions are admissible, as neither outperforms the other for all possible true shares. On the right, we trace the expected payoff of both functions over the whole parameter space. Although the RDF outperforms the ADF for shares in the center of the parameter space, it performs worse at the boundaries. Overall, the performance of the RDF is more balanced across the whole parameter space, which motivates its name.



(a) Expected payoff

(b) Performance

Figure 3: Evaluating decision functions

Figure 4 ranks the two functions according to different decision-theoretic criteria. Both decision functions have their lowest expected payoff at $\theta = 0.5$. As the RDF outperforms its ADF alternative at that point, the RDF is preferred based on the maximin and minimax regret criteria. The maximin and minimax regret criteria are identical in this setting, as the payoff at the true share is constant across the parameter space. Using the subjective Bayes criterion with a uniform prior, we select the ADF, as its better performance at the boundaries of the parameter space is enough to offset its worse performance in the center.

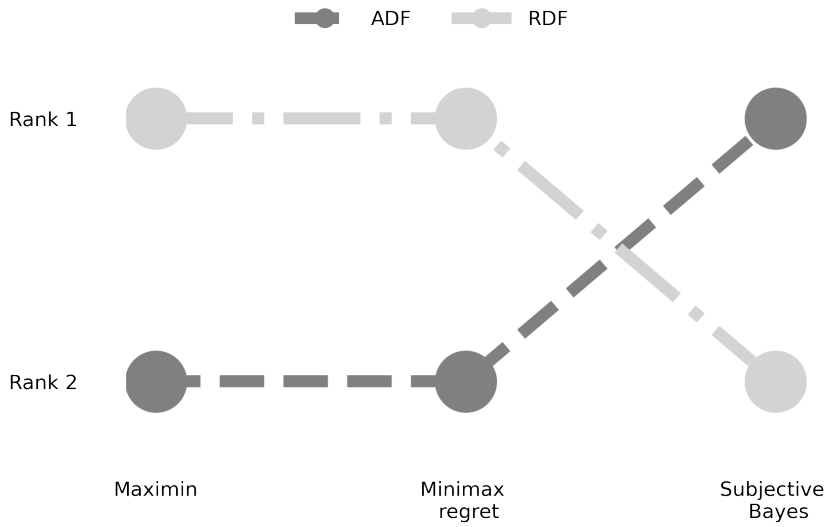
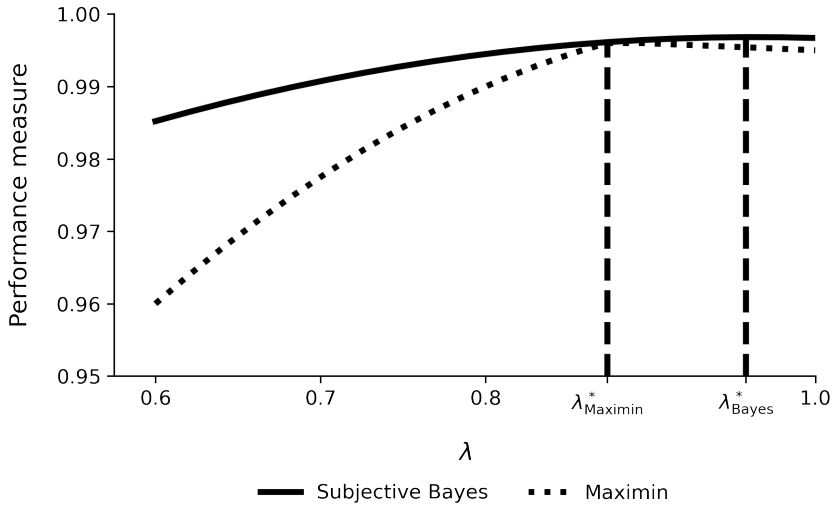


Figure 4: Ranking of decision functions

Returning to the whole set of decision functions, we can construct the optimal statistical decision function in Γ for the alternative criteria by varying λ to maximize the relevant performance measure. For example, Figure 5 shows the minimum and the uniformly weighted performance for varying λ.



Notes: We omit the performance measure for the minimax regret criterion, as $\lambda^*_{\text{Maximin}} = \lambda^*_{\text{Regret}}$ in this setting as noted earlier.

Figure 5: Optimality of decision functions

Neither of our two decision functions analyzed earlier turns out to be optimal, as $\lambda^*_{\text{Bayes}} \approx 0.96$ and $\lambda^*_{\text{Maximin}} \approx 0.87$. Overall, the performance measure is more sensitive to the choice of λ under the maximin criterion than under subjective Bayes.

In summary, the urn example illustrates the performance comparison of alternative decision functions over the whole parameter space. It shows how to construct an optimal decision function within a class for alternative decision-theoretic criteria. Next, we move to the more involved setting of a sequential dynamic decision problem with ambiguous transitions that we analyze in our application.

3. Data-driven robust Markov decision problem

We now outline the framework of an RMDP for the analysis of sequential decision-making in light of model ambiguity. From the perspective of statistical decision theory, any RMDP with a fixed level of robustness is a statistical decision function. Once a sample of transitions is available, we construct the ambiguity set of a given size and solve the RMDP for a robust decision rule.

We first present the general setup of an RMDP and discuss the construction of the ambiguity set. We then turn to the solution approach and describe our decision-theoretic analysis to determine the optimal level of robustness. Throughout, we address the new challenges of analyzing an RMPD as opposed to a standard MDP. In line with our application, we discuss an infinite horizon model in discrete time, stationary utility and transition probabilities, and

discrete states and actions.[4]

We focus our exposition on ambiguity in the transition dynamics of the Markov decision process. We do not address uncertainty about the parameters of the reward functions. Our central insight to use statistical decision theory to determine the optimal level of robustness is also relevant for the parameters of the reward functions. However, we do not address the uncertainty about these parameters, as each setting introduces its unique computational challenges (Mannor and Xu, 2019).

3.1. Setting

We consider the following decision problem. At time $t = 0, 1, 2, \ldots$ a decision-maker observes the state of their environment $s_t \in \mathcal{S}$ and chooses an action a_t from the set of admissible actions $\mathcal{A}$. The decision has two consequences. It creates an immediate utility $u(s_t, a_t)$, and the environment evolves to a new state s_{t+1}. The transition from s_t to s_{t+1} is affected by the action, and governed by a transition probability distribution $p(s_t, a_t)$.

Decision-makers take the future consequences of the current action into account. While a decision rule d_t specifies the planned action for all possible states within period t, a policy $\pi = \{d_0, d_1, d_2, \ldots\}$ is a collection of decision rules and specifies all planned actions for all time periods.

Figure 6 depicts the timing of events in the decision problem. At the beginning of period t, a decision-maker learns about the utility of each alternative, chooses one according to the decision rule d_t, and receives its immediate utility. Then, the state evolves from s_t to s_{t+1}, and the process repeats itself in $t + 1$.

In a standard Markov decision process (MDP), a single transition probability distribution $p(s_t, a_t)$ is associated with each state-action pair. This distribution is assumed to be known, and thus the MDP incorporates risk only. In an RMDP, there is a whole set of distributions associated with each state-action pair collected in an ambiguity set $p(s_t, a_t) \in \mathcal{P}(s_t, a_t)$. For a particular RMDP, the ambiguity set is assumed to be known, and thus the RMDP incorporates risk for a given distribution and ambiguity about the true distribution.

[4]See Puterman (1994) for a textbook introduction to the standard MDP and Rust (1994) for a review of MDPs in economics and structural estimation.

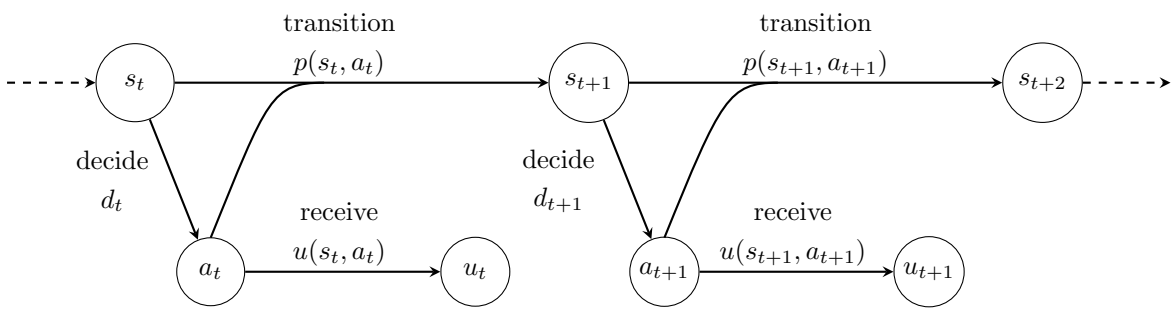


Figure 6: Timing of events

In a standard MDP, the objective of a decision-maker in state s_t at time t is to choose the optimal policy π^* from the set of all possible policies Π that maximizes their expected total discounted utility $\tilde{v}_t^{\pi^*}(s_t)$ as formalized in Equation (3.1):

$$\tilde{v}_t^{\pi^*}(s_t) \equiv \max_{\pi \in \Pi} \mathrm{E}^{\pi} \left[\sum_{r=0}^{\infty} \delta^{t+r} u(s_{t+r}, d_{t+r}(s_{t+r})) \right]. \tag{3.1}$$

The exponential discount factor δ parameterizes a taste for immediate over future utilities. The superscript of the expectation emphasizes that each policy induces a different probability distribution over sequences of possible futures. As long as transition probabilities used to construct the policy are in fact correct, the standard value function $\tilde{v}_t^{\pi^*}(s_t)$ measures the performance of the optimal policy.

In an RMDP, the goal is to implement an optimal policy that maximizes the expected total discounted utility under a worst-case scenario. Given the ambiguity about the transition dynamics, a policy induces a whole set of probabilities over sequences of possible future utilities $\mathcal{F}^{\pi}$, and the worst-case realization determines its ranking. The formal representation of the decision-maker's objective is Equation (3.2):

$$v_t^{\pi^*}(s_t) \equiv \max_{\pi \in \Pi} \left\{ \min_{\mathbf{P} \in \mathcal{F}^{\pi}} \mathrm{E}^{\mathbf{P}} \left[\sum_{r=0}^{\infty} \delta^{t+r} u(s_{t+r}, d_{t+r}(s_{t+r})) \right] \right\}. \tag{3.2}$$

We consider a setting where historical data provides information about the transition dynamics. In the data-driven standard MDP, the empirical probabilities $\hat{p}(s_t, a_t)$ serve as a plug-in for the truth, and the solution of the MDP provides an as-if decision rule. In a data-driven RMDP, the empirical probabilities are used to construct the ambiguity sets for the transitions, and the solution of the RMDP provides a robust decision rule.

We follow Ben-Tal et al. (2013) and create the ambiguity sets using statistical hypothesis testing. We restrict attention to distributions we cannot reject with a certain level of confidence

$\omega \in [0,1]$ around the empirical probabilities and collect them in an estimated ambiguity set $\hat{\mathcal{P}}(s_t, a_t; \omega)$. Different values of ω result in different RMDPs, each with its own statistical decision function. Two special cases stand out. First, if $\omega = 0$, then a decision-maker treats the empirical probabilities as if they are correct. This case captures the notion of as-if decision-making. Second, for $\omega = 1$, a robust decision-maker considers the worst-case scenario over the whole probability simplex at each state-action pair when constructing the optimal policy.

3.2. Solution

In a standard MDP, the objective is to maximize the expected total discounted utility as formalized in Equation (3.1). This requires evaluating the performance of all policies based on all possible sequences of utilities and the probability that each occurs. Fortunately, the stationary Markovian structure of the problem implies that the future looks the same whether the decision-maker is in state s at time t or any other point in time. The only variable that determines the value to the decision-maker is the current state s. Thus, the optimal policy is stationary as well (Blackwell, 1965), and the same decision rule is used in every period. The value function is independent of time and of the solution to the following Bellman equation:

$$\tilde{v}(s) = \max_{a \in \mathcal{A}} \left[u(s,a) + \delta \int \tilde{v}(s')\, \hat{p}(ds'|s,a) \right]. \tag{3.3}$$

The as-if decision rule is recovered from Equation (3.3) by finding the value $a \in \mathcal{A}$ that attains a maximum for each $s \in \mathcal{S}$.

Let $\mathbb{V}$ denote the set of all bounded real value functions on $\mathcal{S}$. Then, the Bellman operator $\tilde{\Lambda} : \mathbb{V} \to \mathbb{V}$ is defined as follows: For all $w \in \mathbb{V}$

$$\tilde{\Lambda}(w)(s) = \max_{a \in \mathcal{A}} \left[u(s,a) + \delta \int w(s')\, \hat{p}(ds'|s,a) \right], \quad s \in \mathcal{S}. \tag{3.4}$$

Under mild conditions, $\tilde{\Lambda}$ is a contraction mapping and allows to compute the value function $\tilde{v}(\cdot)$ as its unique fixed point (Denardo, 1967).

For an RMDP, where transition probabilities are ambiguous, the contraction mapping property of the Bellman operator and the optimality of a stationary deterministic Markovian decision rule both require the assumption of rectangularity of $\mathcal{F}^\pi$ (Iyengar, 2005; Nilim and El Ghaoui, 2005). As the realization of any particular distribution in a state-action pair does not affect future realizations, rectangularity is a form of an independence assumption. The uncertainty is uncoupled across states and actions. This approach rules out any kind of learning about future ambiguity from past experiences due to, for example, a common source of uncertainty across states. While restrictive, most applications rely on the rectangularity assumption, as general

notions of coupled uncertainties are intractable (Wiesemann et al., 2013).[5]

We now develop the formal definition of rectangularity. Let $\mathcal{M}(\mathcal{S})$ denote the set of all probability distributions on $\mathcal{S}$. Then, the set of all conditional transition probability distributions associated with any decision rule d is given by:

$$\mathcal{F}^d = \{p : \mathcal{S} \to \mathcal{M}(\mathcal{S}) \,|\, \forall s \in \mathcal{S},\, p(s) \in \hat{\mathcal{P}}(s, d(s); \omega)\}.$$

For every state $s \in \mathcal{S}$, the next state can be determined by any $p \in \hat{\mathcal{P}}(s, d(s); \omega)$.

A policy π now induces a set of probability distributions $\mathcal{F}^\pi$ on the set of all possible histories $\mathcal{H}$. Any particular history $h = (s_0, a_0, s_1, a_1, \ldots)$ can be the result of many possible combinations of transition probabilities. Rectangularity imposes a structure on the combination possibilities.

Assumption 1. ***Rectangularity*** *The set $\mathcal{F}^\pi$ of probability distributions associated with a policy π is given by*

$$\begin{aligned}\mathcal{F}^\pi &= \left\{\mathbf{P} \mid \forall\, h \in \mathcal{H} : \mathbf{P}(h) = \prod_{t=0}^{\infty} p(s_{t+1}|s_t, a_t),\ \text{with } p(s_t, a_t) \in \hat{\mathcal{P}}(s_t, d_t(s_t); \omega) \text{ for } t = 0, 1, \ldots\right\} \\ &= \mathcal{F}^{d_0} \times \mathcal{F}^{d_1} \times \mathcal{F}^{d_2} \times \ldots = \prod_{t=0}^{\infty} F^{d_t},\end{aligned}$$

where the notation simply denotes that each element in $\mathcal{F}^\pi$ is a product of $p \in \mathcal{F}^{d_t}$, and vice versa (Iyengar, 2005).

Assumption 1 formalizes the idea that ambiguity about the transition probability distribution is uncoupled across states and time. All elements of the ambiguity sets can be freely combined to generate a particular history.

The objective when facing ambiguity is to implement a policy π^* that maximizes the expected total discounted utility under a worst-case scenario as presented in Equation (3.2). Under the rectangularity assumption, the decision-maker faces the same uncertainty, whether he is in state s at time t or any other point in time. Thus, the value function is independent of time and solely depends on the current state s. It is the solution to the robust Bellman equation (3.5), where the future value is evaluated using the worst-case element in the ambiguity set (Iyengar,

[5]See Mannor et al. (2016) and Goyal and Grand-Clement (2023) for recent attempts to introduce milder rectangularity conditions.

2005):

$$v(s) = \max_{a \in \mathcal{A}} \left[u(s,a) + \delta \min_{p \in \hat{\mathcal{P}}(s,a;\omega)} \int v(s')\, p(ds'|s,a) \right]. \tag{3.5}$$

The robust decision rule is recovered from Equation (3.5) by finding the value $a \in \mathcal{A}$ that attains a maximum for each $s \in \mathcal{S}$ under the worst-case scenario for all distributions in the ambiguity set.

The robust Bellman operator on $\mathbb{V}$ follows directly: For all $w \in \mathbb{V}$

$$\Lambda(w)(s) = \max_{a \in \mathcal{A}} \left[u(s,a) + \delta \min_{p \in \hat{\mathcal{P}}(s,a;\omega)} \int w(s')\, p(ds'|s,a) \right] \quad s \in \mathcal{S}. \tag{3.6}$$

Algorithm 1 allows solving the RMDP by a robust version of the value iteration algorithm where κ denotes a convergence threshold. The calculation of future values under the worst-case scenario is the key difference to the standard approach.

Algorithm 1 Robust Value Iteration Algorithm

Input: $v \in \mathbb{V}, \kappa > 0$

For each $s \in \mathcal{S}$, set $\hat{v}(s) = \max_{a \in \mathcal{A}} \left\{ u(s,a) + \min_{p \in \hat{\mathcal{P}}(s,a;\omega)} \int v(s')\, p(ds'|s,a)] \right\}$.

while $\| v - \hat{v} \|_\infty > \kappa$ **do**

$v = \hat{v}$

$\forall\, s \in \mathcal{S}$, set $\hat{v}(s) = \max_{a \in \mathcal{A}} \left\{ u(s,a) + \min_{p \in \hat{\mathcal{P}}(s,a;\omega)} \int v(s')\, p(ds'|s,a)] \right\}$

end while

3.3. Evaluation

The solution of an RMDP is tailored to the simultaneous worst-case realization of all distributions in all ambiguity sets. Although this conservative approach ensures a minimum performance over all distributions in the set, the performance of the robust decision rule in all other cases is disregarded. This indifference introduces a trade-off when determining the size of the ambiguity set (Delage and Mannor, 2010). The larger the set, the more scenarios for which a minimum performance is ensured. However, the robust rule's general performance suffers. This trade-off is particularly pronounced when the actual structure of the decision problem exhibits coupled uncertainties that are ignored in the construction of the robust rule to ensure

its computational tractability.

Statistical decision theory allows us to navigate the trade-off and determine the optimal level of robustness. Each RMDP is a different statistical decision function, and we consider the whole class of statistical decision functions each indexed by $\omega \in [0, 1]$. Adapting our urn example from earlier to accommodate setting of a data-driven RMDP, the parameter space corresponds to the set of transition probability distributions $\mathcal{L}(\mathcal{S}, \mathcal{A}) = \{p : \mathcal{S} \times \mathcal{A} \to \mathcal{M}(\mathcal{S})\}$. We observe data on the transition probabilities and measure the actual performance of a robust decision rule $\eta(\hat{p}; p_0, \omega)$ as the discounted sum of utilities, which depends on the estimate of the transition probabilities $\hat{p}$, the true underlying probabilities p_0, and the confidence level ω used to construct the robust decision function. The standard decision-theoretic criteria translate to this setting as follows:

Maximin	$\omega^* = \arg\max_{\omega \in [0,1]} \min_{p \in \mathcal{L}(\mathcal{S},\mathcal{A})} \mathrm{E}_p \left[\eta(\hat{p}; p, \omega)\right]$
Minimax regret	$\omega^* = \arg\min_{\omega \in [0,1]} \max_{p \in \mathcal{L}(\mathcal{S},\mathcal{A})} \left[\max_{\tilde{\omega} \in [0,1]} \eta(p; p, \tilde{\omega}) - \mathrm{E}_p \left[\eta(\hat{p}; p, \omega)\right]\right]$
Subjective Bayes	$\omega^* = \arg\max_{\omega \in [0,1]} \int_{\mathcal{L}(\mathcal{S},\mathcal{A})} \mathrm{E}_p \left[\eta(\hat{p}; p, \omega)\right] \, df_p$

Note that even for genuinely uncoupled uncertainties, the maximin criterion does not automatically select the most robust statistical decision function ($\omega = 1$). This particular decision function is based on the worst-case scenario over the full probability simplex at each state-action pair. In fact, the worst-case decision function might not be admissible in particular settings where it is weakly dominated by the as-if (or some other) decision function. Suppose, for example, the true distribution corresponds to the worst-case distributions. In this case, the distribution of sampled transitions is degenerate, as the worst-case scenario at each state-action pair is the certain transition to the state with the lowest future value (Nilim and El Ghaoui, 2005). Thus, the as-if and worst-case decision functions share the same performance. For all other true distributions, the as-if decision function may very well outperform the worst-case decision function if the sampled data is sufficiently informative.

4. Bus replacement problem

We now study robust decision-making in the seminal bus replacement problem. First, we discuss the general setting and the details of the computational implementation. Second, we conduct an ex-post analysis of robust decision rules constructed for the observed sample of mileage transitions analyzed in Rust (1987). Third, considering the situation before any data is realized, we conduct an ex-ante analysis of robust decision functions with varying levels of robustness over the whole probability simplex, which allows us to determine the optimal level of robustness using statistical decision theory.

4.1. Setting

The bus replacement model is set up as a regenerative optimal stopping problem (Chow et al., 1971). It is motivated by the sequential decision problem of a maintenance manager, Harold Zurcher, for a fleet of buses. He makes repeated decisions about their maintenance to maximize the expected total discounted utility under a worst-case scenario. Each month t, a bus arrives at the bus depot in state $s_t = (x_t, \epsilon_t)$ described by its mileage since the last engine replacement x_t and other signs of wear and tear ϵ_t. He faces the decision to either conduct a complete engine replacement ($a_t = 1$) or perform basic maintenance work ($a_t = 0$). The cost of maintenance $c(x_t)$ increases with the mileage state, while the cost of replacement RC remains constant. In the case of an engine replacement, the mileage state is reset to zero. Note that we do not attempt to describe Harold Zurcher's decision-making process. Instead, we are interested in how a generic decision-maker should make decisions in this setting.

The immediate utility of each action is given by:

$$u(a_t, x_t) + \epsilon_t(a_t) \quad \text{with} \quad u(a_t, x_t) = \begin{cases} -RC & a_t = 1 \\ -c(x_t) & a_t = 0. \end{cases}$$

Decisions are made in light of uncertainty about next month's state variables captured by their conditional distribution $p(x_t, \epsilon_t, a_t)$.

Although in this framework, the utility and consequently the value function is finite in each state, they are not uniformly bounded. This property, however, is a crucial assumption for the results of Blackwell (1965) and Denardo (1967) on the contraction property of the Bellman operator and the stationarity of the optimal policy in the standard MDP setting. For the original as-if analysis, Rust (1988) circumvents this problem by imposing conditional independence between the observable and unobservable state variables, i.e. $p(x_{t+1}, \epsilon_{t+1}|x_t, \epsilon_t, a_t) = p(x_{t+1}|x_t, a_t)\, q(\epsilon_{t+1}|x_{t+1})$, and assuming that the unobservables $\epsilon_t(a_t)$ are independent and identically distributed according to an extreme value distribution with mean zero and scale parameter one. These two assumptions, together with the additive separability between the observed and unobserved state variables in the immediate utilities, ensure that the expectation of the next period's value function is independent of the time. The regenerative structure of the process implies that the transition probabilities in case of replacement in any mileage state correspond to the probabilities of maintenance in the zero mileage state. Therefore, the expected value function is the unique fixed point of a contraction mapping on the reduced space of mileage states only. In addition, the conditional choice probabilities $P(a_t|x_t)$ have a closed-form solution (McFadden, 1973). We build on these results and extend them to our robust setting with ambiguous transition dynamics. The proof is available in Appendix A.

In the analysis of the original bus replacement problem, the distribution of the monthly mileage transitions are estimated in a first step and used as plug-in components for the subsequent analysis. We extend the original setup and explicitly account for the ambiguity in the estimation. Following the arguments on the regenerative structure of the process above, we incorporate ambiguity in the RMDP with ambiguity sets conditional on the mileage states x only. We construct ambiguity sets $\hat{\mathcal{P}}(x;\omega)$ based on the Kullback-Leibler divergence D_{KL} (Kullback and Leibler, 1951) that are statistically meaningful, computationally tractable, and anchored in empirical estimates $\hat{p}(x)$.

Our ambiguity set takes the following form for each mileage state x:

$$\hat{\mathcal{P}}(x;\omega) = \left\{ q \in \mathring{\Delta}_{|J_x|} : D_{KL}(q \parallel \hat{p}(x)) = \sum_{i=1}^{|J_x|} q_i \ln\left(\frac{q_i}{\hat{p}(j_i|x)}\right) \leq \rho_x(\omega) \right\},$$

where $J_x = \{j_1, \ldots, j_{|J_x|}\}$ denotes the set of all states that have an estimated non-zero probability to be reached from x, $\mathring{\Delta}_{|J_x|} = \{p \in \mathbb{R}^{|J_x|} \,|\, p_i > 0 \text{ for all } i = 1, \ldots, |J_x| \text{ and } \sum_{i=1}^{|J_x|} p_i = 1\}$ is the interior of the $(|J_x| - 1)$ - dimensional probability simplex, and $\rho_x(\omega)$ captures the size of the set for the state x with a given level of confidence ω.

Iyengar (2002) and Ben-Tal et al. (2013) provide the statistical foundation to calibrate $\rho_x(\omega)$ such that the true (but unknown) distribution p_0 is contained within the ambiguity set for a given level of confidence ω. Let χ^2_{df} denote a chi-squared random variable with df degrees of freedom, and let $F_{df}(\cdot)$ denote its cumulative distribution function with inverse $F^{-1}_{df}(\cdot)$. Then, the following approximate relationship holds as the number of observations N_x for state x tends to infinity (Pardo, 2005):

$$\begin{aligned}
\omega &= \Pr[\, p_0 \in \hat{\mathcal{P}}(x;\omega)\,] \\
&\approx \Pr[\, \chi^2_{|J_x|-1} \leq 2N_x\rho_x(\omega)\,] \\
&= F_{|J_x|-1}(2N_x\rho_x(\omega)).
\end{aligned}$$

We can therefore calibrate the size of the ambiguity set based on the following relationship:

$$\rho_x(\omega) = \tfrac{1}{2N_x} F^{-1}_{|J_x|-1}(\omega). \tag{4.1}$$

We use Rust's (1987) original data to inform our computational experiments. His data consists of monthly odometer readings x_t and engine replacement decisions a_t for 162 buses. The fleet consists of eight groups that differ in their manufacturer and model. We focus on the fourth group of 37 buses with a total of 4,292 monthly observations. We discretize mileage into 78

equally spaced bins of length $5,000$ and set the discount factor to $\delta = 0.9999$.

Figure 7 highlights the limited information about the true distribution of mileage utilization. It shows the number of observations available to estimate next month's utilization for different levels of accumulated mileage. While there are more than 1,150 observations on buses with less than 50,000 miles, there are only about 220 with more than 300,000.

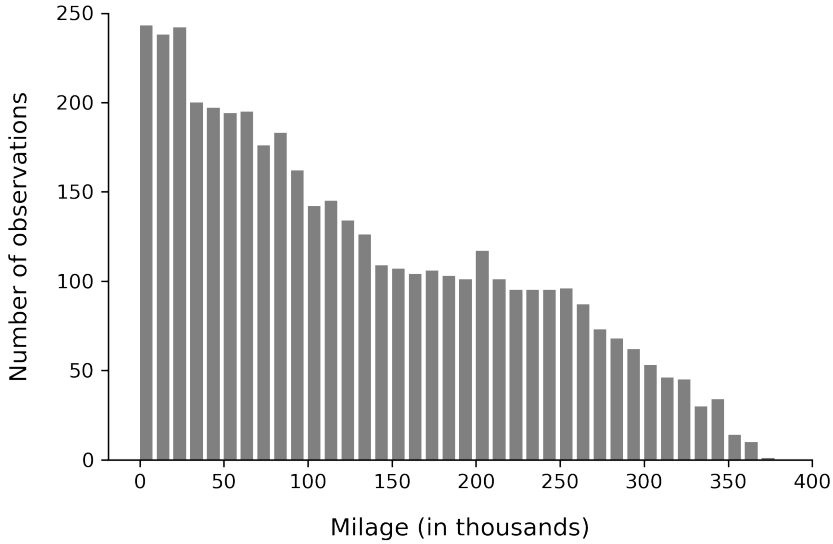


Figure 7: Distributions of observations

We analyze a specific example of Rust's (1987) bus replacement problem. We do not use his reported estimates of the maintenance and replacement costs. Given these estimates, decisions are mainly driven by the unobserved state variable ϵ_t, and so ambiguity about the evolution of the observed state variable x_t does not have a substantial effect on decisions. We ensure that a bus's accumulated mileage has a considerable impact on the timing of engine replacements by increasing the maintenance and replacement costs compared to their empirical estimates. Thus, we specify the following cost function $c(x_t) = 0.4\, x_t$ and set the replacement costs RC to 50.

We solve the model using a modified version of the original nested fixed point algorithm (NFXP) (Rust, 1988), and we determine the worst-case transition probabilities in each successive approximation of the fixed point. Given the size of the ambiguity set, we can determine the worst-case probabilities as the solution to a one-dimensional convex optimization problem (Iyengar, 2005; Nilim and El Ghaoui, 2005).[6]

[6]The core routines are implemented in our group's ruspy (2020) and robupy (2020) software packages and are publicly available.

4.2. Ex-post analysis

We first study as-if and robust decision rules for Rust's (1987) observed sample of mileage transitions. We present the estimated transition probabilities and the corresponding worst-case distributions. We then explore alternative decision rules based on several RMDPs, outline the resulting differences in maintenance decisions, and evaluate their relative performance under different scenarios.

Figure 8 shows the point estimates $\hat{p}$ for the transition probabilities of monthly mileage usage. We pool all 4,292 observations to estimate this distribution by maximum likelihood, and thus the probability of the next period's mileage utilization is the same for each state x_t. We only observe increases of at most $J = 3$ grid points per month. For about 60% of the sample, monthly bus utilization is between 5,000 and 10,000 miles. Very high usage of more than 10,000 miles amounts to only 1.2%.

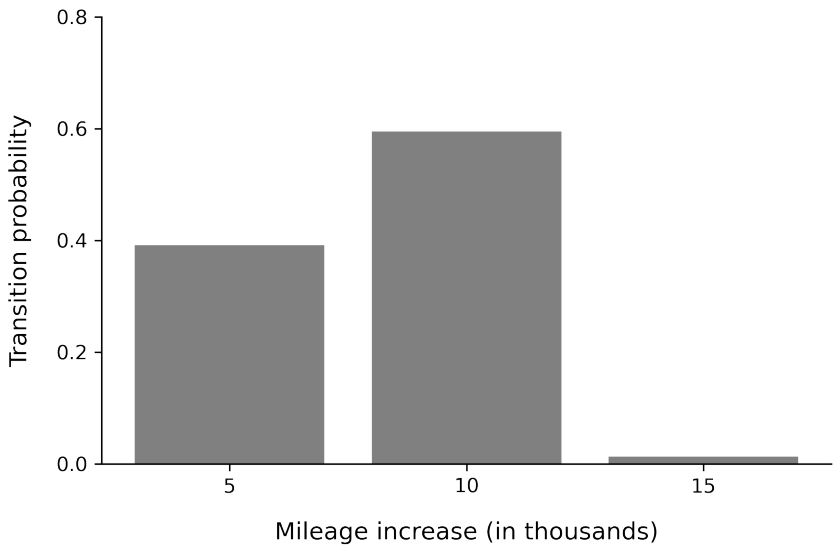
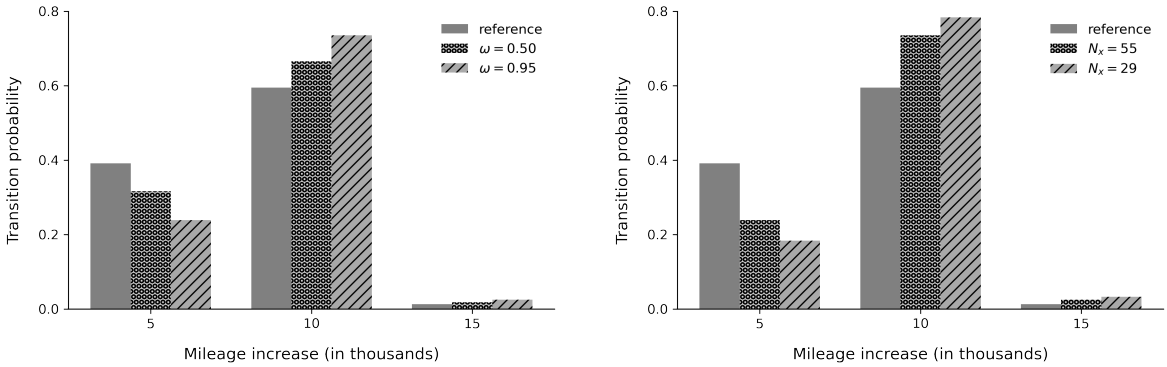


Figure 8: Estimated transition probabilities

The confidence level ω and the available number of observations N_x determine the size of the ambiguity set as outlined in Equation (4.1). From now on, we mimic state-specific ambiguity sets by constructing them based on the average number of 55 observations per state. Note that while the estimated distribution is the same for all mileage levels, its worst-case realization is not. However, there are only minor differences across mileage levels, so we focus our following discussion on a bus with an odometer reading of 75,000.

Figure 9 shows the transition probabilities for different sizes of the ambiguity set. We vary the confidence level for the whole number of observations ($N_x = 55$) on the left, while on the right, the level of confidence remains fixed ($\omega = 0.95$), and we cut the number of observations roughly

in half. The larger the ambiguity set, the more probability is attached to higher mileage utilization, resulting in higher costs overall. For example, while the probability of mileage increases of 10,000 or more is an infrequent occurrence in the data, its probability increases first to 1.7%. It then doubles to 2.5% as we increase the confidence level. When only about half the data is available, this probability increases even further to 3.2%.



(a) Variation in ω, ($N_x = 55$) **(b)** Variation in N_x, ($\omega = 0.95$)

Figure 9: Worst-case transition probabilities

The decision-maker chooses whether to perform regular maintenance work on a bus or replace its complete engine each month. The assumed transition probabilities correspond to their worst-case transitions within the ambiguity set. As a result, any differences between the as-if and worst-case distributions translate into different maintenance decisions.

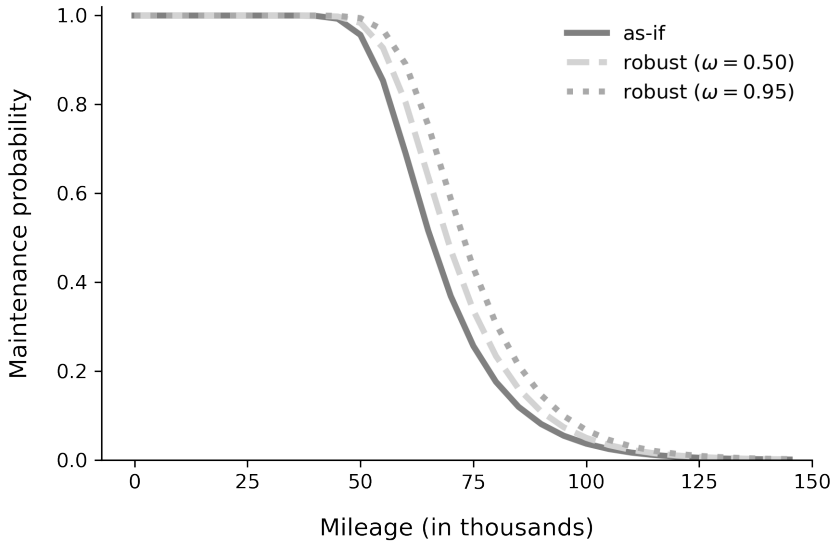


Figure 10: Maintenance probabilities

Figure 10 shows the maintenance probabilities for different levels of accumulated mileage and alternative rules. Overall, the maintenance probability decreases with accumulated mileage,

as maintenance becomes more costly than an engine replacement. Robust rules result in a higher probability of maintenance compared to the as-if decision rule. Under the worst-case transitions, a bus is more likely to experience higher usage during the period. As the cost of maintenance is determined by the mileage level at the beginning of the period, maintenance becomes more attractive. For example, again considering a bus with 75,000 miles, the as-if maintenance probability is 25%, while it is 33% ($\omega = 0.50$) and 43% ($\omega = 0.95$) following the robust rule.

To gain further insights into the differences between the as-if and robust decisions, we simulate a fleet of 1,000 buses for 100,000 months under the alternative decision rules.

Figure 11 shows the level of accumulated mileage over time for a single bus under different decision rules. It clarifies our simulation setup, where we apply different decision rules to the same bus. The realizations of observed transitions and unobserved signs of wear and tear remain the same. The bus accumulates more and more mileage until Harold Zurcher replaces the complete engine and the odometer is reset to zero. The first replacement happens after 20 months at 60,000 miles following the as-if decision rule, while it is delayed for another four months under the robust alternative ($\omega = 0.95$). As its timing differs, the odometer readings will start to diverge after 20 months, even though monthly utilization remains the same.

We now evaluate the as-if and robust decisions at the boundary of the ambiguity set. We measure the performance of the alternative decision rules based on their total discounted utility under different assumed and actual mileage transitions.

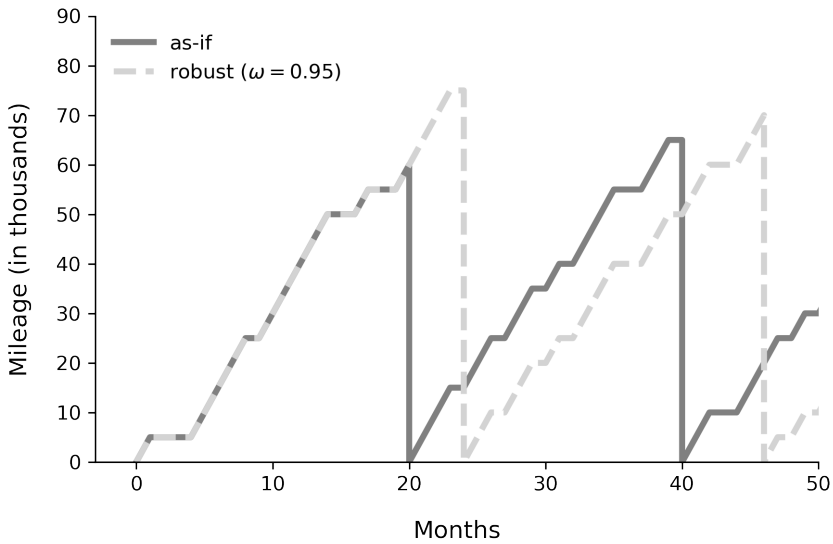


Figure 11: Single bus under alternative decision rules

Figure 12 shows the performance of the as-if decision rule over time when the worst-case distribution for a confidence level of 0.95 governs the actual transitions. It illustrates the sensitivity of the as-if decision rule to perturbations in the transition probabilities. The solid line corresponds to its expected long-run performance without misspecification of the decision problem, while the dashed line indicates its observed performance. After about 20,000 months, it accumulates the expected long-run average cost and performs about 14% worse overall.

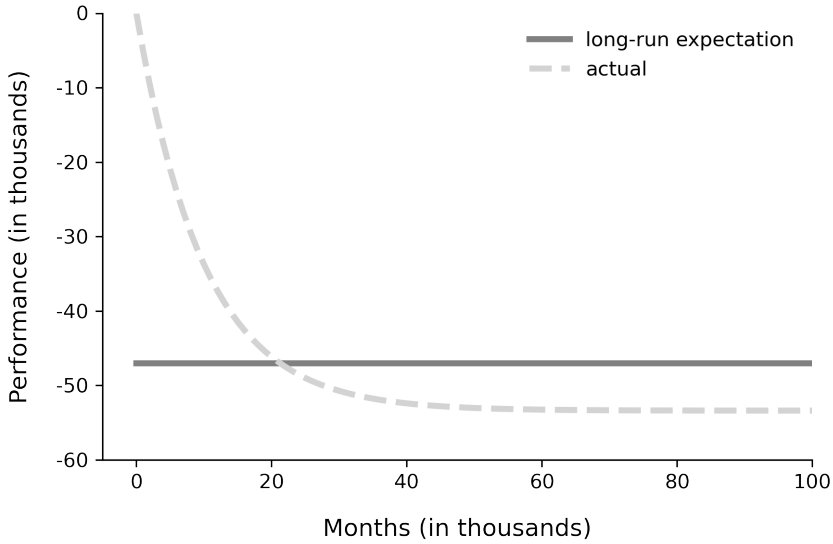
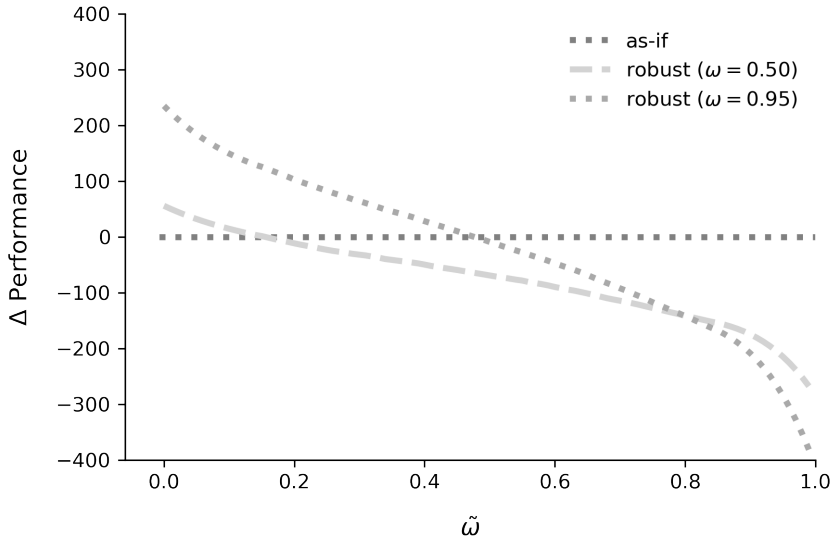


Figure 12: Performance of as-if decision rule

Figure 13 shows the average difference in performance between the as-if and two robust decision rules with confidence levels of 0.50 and 0.95, respectively. The actual transitions follow the worst-case distribution with varying ω. A positive value indicates that the robust decision rule outperforms the as-if decision rule. In the absence of any misspecification, the as-if decision rule must defeat any other decision rule. The same is true for the robust decision rule when the actual transitions are governed by the same ω used for their construction. Nevertheless, the as-if decision rule continues to outperform both robust decisions for moderate levels of ω. For worst-case distributions with ω larger than 0.2, the first robust decision rule ($\omega = 0.5$) starts to beat the as-if decision rule. For the other robust decision rule ($\omega = 0.95$), the same is true for worst-case transitions of ω equal to 0.5.



Notes: We apply a Savitzky-Golay filter (Savitzky and Golay, 1964) to smooth the simulation results.

Figure 13: Performance and misspecification

4.3. Ex-ante analysis

We now turn to the situation before any data are realized. We evaluate the ex-ante performance of as-if and robust decision functions over the whole probability simplex and determine the optimal level of robustness.

We operationalize our analysis as follows. In line with Rust's (1987) assumption on the distribution of the mileage utilization, we specify a uniform grid with 0.1 increments over the interior of the two-dimensional probability simplex $\mathring{\Delta}_3$. At each grid point, we draw 100 samples of 55 random mileage utilizations. For each sample, we solve several robust decision functions for a grid of $\omega = \{0.0, 0.1, \ldots, 1.0\}$ using the estimated transition probabilities. Note that the uncertainties are coupled across states, as the same underlying probability creates the sample of bus utilizations. Thus, the rectangularity assumption does not reflect the economic environment. However, we still impose it when constructing the robust decision functions to ensure tractability. We then simulate the implied decision rules' actual performance and compute their expected performance by averaging across the 100 runs for each grid point. Using this information, we measure the performance of the different decisions based on the maximin criterion, the minimax regret rule, and the subjective Bayes approach using a uniform prior.

In Figure 14 we illustrate the differences in expected performance between a robust decision function ($\omega = 0.1$) and the as-if alternative over the probability simplex.

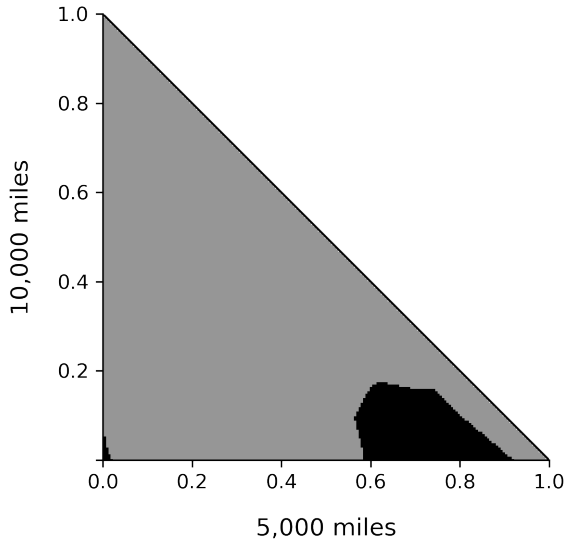


Figure 14: Relative performance of decision rules

In the gray areas, the as-if decisions outperform the robust alternative based on their expected performance. The opposite is true for the black areas: robust decisions perform very well when the true probability of mileage increases of $5,000$ per month is high and when the true probability of increases amounting to $10,000$ is low. Otherwise, the as-if decisions outperform the robust alternative. Thus, no rule dominates the other, and it is essential to aggregate the performance over the whole probability simplex using decision theory before settling on a decision rule.

Figure 15 ranks the as-if decisions against selected robust alternatives for the different performance criteria.

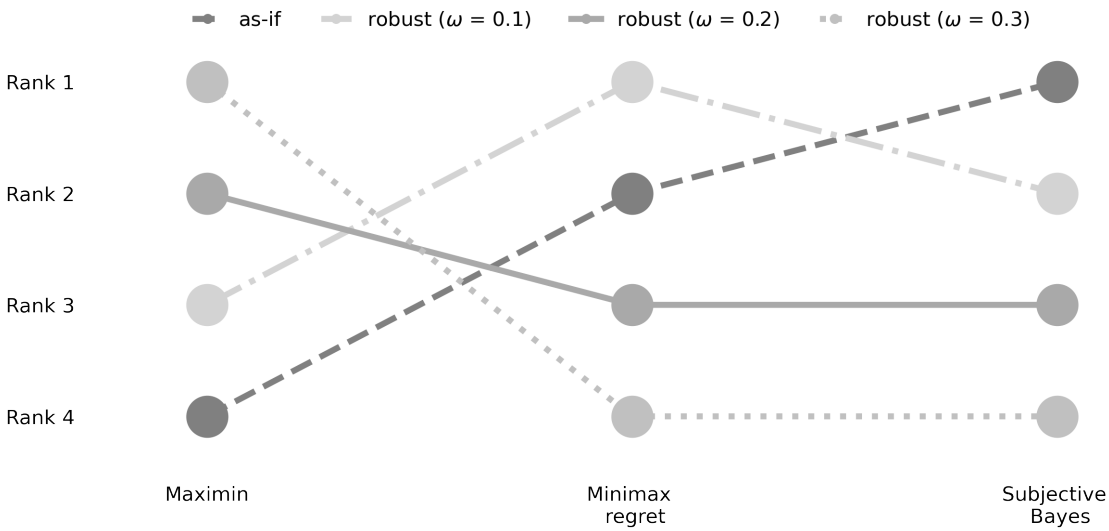


Figure 15: Ranking of decision rules

Based on a maximin criterion, decision functions rank higher when the confidence level ω used to construct them is greater. The decision function with $\omega = 0.3$ comes in first, while as-if decisions rank last. Thus, decision-makers can improve their worst-case outcomes by adopting a robust decision function. However, this comes at a cost, as indicated by the improved rankings for the as-if decision function as we move to different criteria. As-if decisions move to second place for minimax regret. The as-if decision rule comes in first when we aggregate performance across all states using a subjective Bayes approach with a uniform prior. Thus, our approach clarifies the trade-offs involved when choosing a particular decision function for decision-making.

We now determine the optimal size of the ambiguity set ω^* for each decision-theoretic criterium. Figure 16 shows the minimum performance of the decision functions for varying levels of ω normalized between zero and one. Among all decision functions, robust decisions with $\omega = 0.36$ have the highest minimum performance. They thus strike a balance between the conservatism of the worst-case approach and the protection against unfavorable transition probabilities. Based on the maximin criterion, the as-if decision function performs worst.

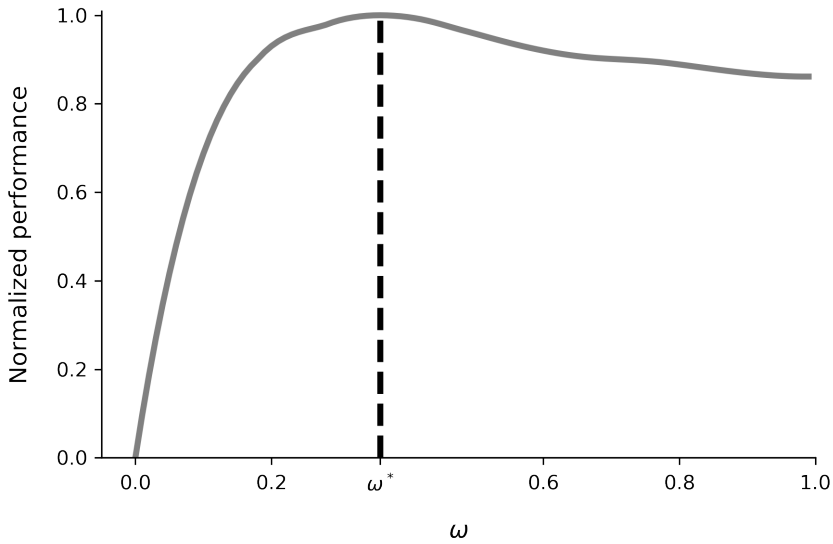


Figure 16: Optimality of decision rules

The minimax regret criterion leads to a slightly reduced level of $\omega^* = 0.1$. As-if decisions are optimal based on the subjective Bayes criterion with a uniform prior.

5. Conclusion

Economists often estimate economic models on data and use the point estimates as a stand-in for the truth when studying the model's implications for optimal decision-making. This practice ignores model ambiguity, exposes the decision problem to misspecification, and ultimately leads to post-decision disappointment. We develop a framework to explore, evaluate, and op-

timize robust decision rules that explicitly account for the uncertainty in the estimation using statistical decision theory. We show how to operationalize our analysis by studying robust decisions in a stochastic dynamic investment model in which a decision-maker directly accounts for uncertainty in the model's transition dynamics.

As our core contribution, we combine ideas from data-driven robustness optimization (Bertsimas et al., 2018), robust Markov decision processes (Ben-Tal et al., 2009), and statistical decision theory (Berger, 2010) to optimize robustness in decision-making. This insight transfers directly to many other settings. For example, the COVID-19 pandemic provides a timely example of economists informing policy-making by using highly parameterized models in light of ubiquitous uncertainties (Avery et al., 2020). When analyzing these models, economists treat many of their parameters as if they are known. However, their actual values are uncertain, as they are often estimated based on external data sources. Using statistical decision theory, our research illustrates how to conduct robust policy-making and to evaluate its relative performance against policies that ignore uncertainty. Such an approach promotes a sound decision-making process, as it provides decision-makers with the tools to systematically navigate the uncertainties they face (Berger et al., 2021).

References

Adda, J., Dustmann, C., and Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, 125(2):293–337.

Alismail, F., Xiong, P., and Singh, C. (2018). Optimal wind farm allocation in multi-area power systems using distributionally robust optimization approach. *IEEE Transactions on Power Systems*, 33(1):536–544.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2020). On the informativeness of descriptive statistics for structural estimates. *Econometrica*, 88(6):2231–2258.

Armstrong, T. B. and Kolesár, M. (2021). Sensitivity analysis using approximate moment condition models. *Quantitative Economics*, 12(1):77–108.

Arrow, K. J. (1951). Alternative approaches to the theory of choice in risk-taking situations. *Econometrica*, 19(4):404–437.

Avery, C., Bossert, W., Clark, A., Ellison, G., and Ellison, S. F. (2020). Policy implications of models of the spread of coronavirus: Perspectives and opportunities for economists. *NBER Working Paper*.

Bagwell, K., Staiger, R. W., and Yurukoglu, A. (2021). Quantitative analysis of multiparty tariff negotiations. *Econometrica*, 89(4):1595–1631.

Barnett, M., Brock, W., and Hansen, L. P. (2020). Pricing uncertainty induced by climate change. *The Review of Financial Studies*, 33(3):1024–1066.

Ben-Tal, A., den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

Ben-Tal, A., El Ghaoui, L., and Nemirowski, A. (2009). *Robust optimization*. Princeton University Press, Princeton, NJ.

Berger, J. O. (2010). *Statistical decision theory and Bayesian analysis*. Springer, New York City, NY.

Berger, L., Berger, N., Bosetti, V., Gilboa, I., Hansen, L. P., Jarvis, C., Marinacci, M., and Smith, R. D. (2021). Rational policymaking during a pandemic. *Proceedings of the National Academy of Sciences*, 118(4).

Bertsimas, D., Gupta, V., and Kallus, N. (2018). Data-driven robust optimization. *Mathematical Programming*, 167(2):235–292.

Bertsimas, D. and Thiele, A. (2006). Robust and data-driven optimization: Modern decision making under uncertainty. In Johnson, M. P., Norman, B., and Secomandi, N., editors, *Models, Methods, and Applications for Innovative Decision Making*, pages 95–122. INFORMS.

Blackwell, D. (1965). Discounted dynamic programming. *The Annals of Mathematical Statistics*, 36(1):226–235.

Blanchet, J., Kang, Y., and Murthy, K. (2019). Robust wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, 56(3):830–857.

Blanchet, J., Kang, Y., Olea, J. L. M., Nguyen, V. A., and Zhang, X. (2023). Dropout training is distributionally robust optimal. *Journal of Machine Learning Research*, 24(180):1–60.

Blundell, R. and Shephard, A. (2012). Employment, hours of work and the optimal taxation of low-income families. *The Review of Economic Studies*, 79(2):481–510.

Bonhomme, S. and Weidner, M. (2021). Minimizing sensitivity to model misspecification. *Quantitative Economics*, 13(3):907–954.

Brown, D. B., Giorgi, E. D., and Sim, M. (2012). Aspirational preferences and their representation by risk measures. *Management Science*, 58(11):2095–2113.

Bruns-Smith, D. and Zhou, A. (2023). Robust fitted-q-evaluation and iteration under sequentially exogenous unobserved confounders. *arXiv preprint arXiv:2302.00662*.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022). Locally robust semiparametric estimation. *Econometrica*, 90(4):1501–1535.

Chow, Y. S., Robbins, H., and Siegmund, D. (1971). *Great expectations: The theory of optimal stopping*. Houghton Mifflin, Boston, MA.

Christensen, T. and Connault, B. (2023). Counterfactual sensitivity and robustness. *Econometrica*, 91(1):263–298.

Delage, E. and Mannor, S. (2010). Percentile optimization for Markov decision processes with parameter uncertainty. *Operations Research*, 58(1):203–213.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612.

Denardo, E. V. (1967). Contraction mappings in the theory underlying dynamic programming. *SIAM Review*, 9(2):165–177.

Duchi, J. C., Glynn, P. W., and Namkoong, H. (2021). Statistics of robust optimization: A generalized empirical likelihood approach. *Mathematics of Operations Research*, 46(3):946–969.

Eaton, J., Kortum, S., and Kramarz, F. (2011). An anatomy of international trade: Evidence from french firms. *Econometrica*, 79(5):1453–1498.

Eisenhauer, P., Gabler, J., and Janys, L. (2021). Structural models for policy-making: Coping with parametric uncertainty. *arXiv preprint arXiv:2103.01115*, submitted.

French, E. and Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, 6(2):291–337.

Gilboa, I. (2009). *Theory of decision under uncertainty*. Cambridge University Press, New York City, NY.

Goh, J., Bayati, M., Zenios, S. A., Singh, S., and Moore, D. (2018). Data uncertainty in Markov chains: Application to cost-effectiveness analyses of medical innovations. *Operations Research*, 66(3):697–715.

Gotoh, J.-y., Kim, M. J., and Lim, A. E. B. (2021). Calibration of distributionally robust empirical optimization models. *Operations Research*.

Goyal, V. and Grand-Clement, J. (2023). Robust Markov decision process: Beyond rectangularity. *Mathematics of Operations Research*, 48(1):203–226.

Hansen, L. P. and Sargent, T. J. (2016). *Robustness*. Princeton University Press, Princeton, NJ.

He, S., Sim, M., and Zhang, M. (2019). Data-driven patient scheduling in emergency departments: A hybrid robust-stochastic approach. *Management Science*, 65(9):4123–4140.

Hirano, K. and Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, 77(5):1683–1701.

Honoré, B., Jørgensen, T., and de Paula, Á. (2020). The informativeness of estimation moments. *Journal of Applied Econometrics*, 35(7):797–813.

Hortaçsu, A., Luco, F., Puller, S. L., and Zhu, D. (2019). Does strategic ability affect efficiency? Evidence from electricity markets. *American Economic Review*, 109(12):4302–42.

Igami, M. (2017). Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998. *Journal of Political Economy*, 125(3):798–847.

Ilut, C. (2020). Paralyzed by fear: Rigid and discrete pricing under demand uncertainty. *Econometrica*, 88(5):1899–1938.

Iskhakov, F., Lee, J., Rust, J., Schjerning, B., and Seo, K. (2016). Comment on "Constrained optimization approaches to estimation of structural models". *Econometrica*, 84(1):365–370.

Iyengar, G. N. (2002). Robust dynamic programming. *CORC Tech Report*.

Iyengar, G. N. (2005). Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280.

Jin, X., Luo, D., and Zeng, X. (2020). Tail risk and robust portfolio decisions. *Management Science*, 67(5):3254–3275.

Jørgensen, T. H. (2023). Sensitivity to calibrated parameters. *The Review of Economics and Statistics*, 105(2):474–481.

Kaufman, D. L., Schaefer, A. J., and Roberts, M. S. (2017). Living-donor liver transplantation timing under ambiguous health state transition probabilities. *SSRN Working Paper*.

Keith, A. J. and Ahner, D. K. (2021). A survey of decision making and optimization under uncertainty. *Annals of Operations Research*, 300(2):319–353.

Kitigawa, Kitagawa, T., and Tetenov, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrics*, 86(2):591–616.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Houghton Mifflin Harcourt, Boston, MA.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Lam, H. (2016). Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, 41(4):1248–1275.

Mannor, S., Mebel, O., and Xu, H. (2016). Robust MDPs with k-rectangular uncertainty. *Mathematics of Operations Research*, 41(4):1484–1509.

Mannor, S., Simester, D., Sun, P., and Tsitsiklis, J. N. (2007). Bias and variance approximation in value function estimates. *Management Science*, 53(2):308–322.

Mannor, S. and Xu, H. (2019). Data-driven methods for Markov decision problems with parameter uncertainty. In Netessine, S., editor, *Operations Research & Management Science in the Age of Analytics*, pages 101–129. INFORMS.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246.

Manski, C. F. (2009). The 2009 Lawrence R. Klein Lecture: Diversified treatment choice under ambiguity. *International Economic Review*, 50(4):1013–1041.

Manski, C. F. (2021). Econometrics for decision making: Building foundations sketched by Haavelmo and Wald. *Econometrica*, 89(6):2827–2853.

Marinacci, M. (2015). Model uncertainty. *Journal of the European Economic Association*, 13(6):1022–1100.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York City, NY.

Meng, F., Qi, J., Zhang, M., Ang, J., Chu, S., and Sim, M. (2015). A robust optimization model for managing elective admission in a public hospital. *Operations Research*, 63(6):1452–1467.

Niehans (1948). Zur Preisbildung bei ungewissen Erwartungen. *Swiss Journal of Economics and Statistics*, 84(5):433–456.

Nilim, A. and El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798.

Pardo, L. (2005). *Statistical inference based on divergence measures*. Chapman & Hall, London, UK.

Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, New York City, NY.

Rahimian, H. and Mehrotra, S. (2019). Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*.

Reich, G. (2018). Divide and conquer: Recursive likelihood function integration for hidden Markov models with continuous latent variables. *Operations Research*, 66(6):1457–1470.

robupy (2020). A Python package for robust optimization.

ruspy (2020). An open-source package for the simulation and estimation of a prototypical infinite-horizon dynamic discrete choice model based on Rust (1987).

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Rust, J. (1988). Maximum likelihood estimation of discrete control processes. *SIAM Journal on Control and Optimization*, 26(5):1006–1024.

Rust, J. (1994). Structural estimation of Markov decision processes. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, pages 3081–3143. North-Holland Publishing Company, Amsterdam, Netherlands.

Saghafian, S. (2018). Ambiguous partially observable Markov decision processes: Structural results and applications. *Journal of Economic Theory*, 178:1–35.

Samuelson, S. and Yang, I. (2017). Data-driven distributionally robust control of energy storage to manage wind power fluctuations. In *2017 IEEE Conference on Control Technology and Applications (CCTA)*, pages 199–204.

Savage, L. J. (1954). *The foundations of statistics*. John Wiley & Sons, New York City, NY.

Savitzky, A. and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639.

Smith, J. E. and Winkler, R. L. (2006). The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322.

Stoye (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics*, 151(1):70–81.

Stoye (2012). Minimax regret treatment choice with covariates or with limited validity of experiments. *Journal of Econometrics*, 166(1):138–156.

Su, C.-L. and Judd, K. L. (2012). Constrained optimization approaches to estimation of structural models. *Econometrica*, 80(5):2213–2230.

Tetenov (2012). Statistical treatment choice based on asymmetric minimax regret choice. *Journal of Econometrics*, 166(1):157–165.

Wald, A. (1950). *Statistical decision functions*. John Wiley & Sons, New York City, NY.

Wiesemann, W., Kuhn, D., and Rustem, B. (2013). Robust Markov decision processes. *Mathematics of Operations Research*, 38(1):153–183.

Zymler, S., Kuhn, D., and Rustem, B. (2013). Worst-case value at risk of nonlinear portfolios. *Management Science*, 59(1):172–188.

A. The robust contraction mapping

Rust (1987) shows that the expectation of the next period's value function is a fixed point on the mileage states x only. He uses the regenerative property of the mileage process and introduces a separate notion $\tilde{EV}(x)$ for the expected value function of maintenance. $\tilde{EV}(x)$ is the fixed point of the contraction mapping defined as follows: For all $v \in \mathbb{V}$

$$\tilde{\Lambda}(v)(x) = \sum_{x' \in X} p(x'|x) \log \sum_{a \in \{0,1\}} \exp\big(u(x', a) + \delta\, v((1-a)\, x')\big),\ x \in X. \tag{A.1}$$

Following Iyengar (2005), we adopt a similar approach and show:

Theorem 1. *Let the robust Bellman operator $\Lambda : \mathbb{V} \to \mathbb{V}$ be defined as follows: For all $v \in \mathbb{V}$*

$$\begin{aligned}\Lambda(v)(x) &= \int \max_{a \in \{0,1\}} \Big[u(x, a) + \epsilon(a) + \delta \min_{p \in \mathcal{P}((1-a)\, x, \omega)} \sum_{x' \in X} p(x') v(x')\Big] q(d\epsilon) \\ &= \log \sum_{a \in \{0,1\}} \exp\Big[u(x, a) + \delta \min_{p \in \mathcal{P}((1-a)\, x, \omega)} \sum_{x' \in X} p(x') v(x')\Big]. \end{aligned} \tag{A.2}$$

Then $\Lambda(\cdot)$ is a contraction mapping on $\big(\mathbb{V}, \|\cdot\|_\infty\big)$ with unique fixed point EV.

Proof. Let $v, w \in \mathbb{V}$ be arbitrary. Fix $x \in X$ and assume without loss of generality that $\Lambda(w)(x) \geq \Lambda(v)(x)$. Let $\nu > 0$ be arbitrary. Then choose $p^a \in \mathcal{P}((1-a)x, \omega)$, such that

$$\begin{aligned}&\max_{a \in \{0,1\}} [u(x, a) + \epsilon(a) + \delta \min_{p \in \mathcal{P}((1-a)\, x, \omega)} \sum_{x' \in X} p(x') v(x')] \geq \\ &\max_{a \in \{0,1\}} [u(x, a) + \epsilon(a) + \delta \sum_{x' \in X} p^a(x') v(x')] - \nu.\end{aligned}$$

By construction:

$$\begin{aligned}&\max_{a \in \{0,1\}} [u(x, a) + \epsilon(a) + \delta \min_{p \in \mathcal{P}((1-a)\, x, \omega)} \sum_{x' \in X} p(x') w(x')] \leq \\ &\max_{a \in \{0,1\}} [u(x, a) + \epsilon(a) + \delta \sum_{x' \in X} p^a(x') w(x')].\end{aligned}$$

Rust (1988) shows for any conditional distribution measure p and mileage state $x \in X$:

$$\begin{aligned}&\max_{a \in \{0,1\}} [u(x, a) + \epsilon(a) + \delta \sum_{x' \in X} p(x') w(x')] - \max_{a \in \{0,1\}} [u(x, a) + \epsilon(a) + \delta \sum_{x' \in X} p(x') v(x')] \\ &\leq \delta \max_{a \in \{0,1\}} |\sum_{x' \in X} p(x')(w(x') - v(x')| \leq \delta\, \|w - v\|_\infty .\end{aligned}$$

This holds in particular for p^a, which yields:

$$
\begin{aligned}
0 &\leq \Lambda(w)(x) - \Lambda(v)(x) \\
&\leq \int \Bigg(\max_{a \in \{0,1\}} [u(x,a) + \epsilon(a) + \delta \sum_{x' \in X} p^a(x') w(x')] \\
&\qquad - \max_{a \in \{0,1\}} [u(x,a) + \epsilon(a) + \delta \sum_{x' \in X} p^a(x') v(x')] + \nu \Bigg) q(d\epsilon) \\
&\leq \int (\delta \, \|w - v\|_\infty + \nu) \, q(\epsilon) \\
&= \delta \, \|w - v\|_\infty + \nu.
\end{aligned}
$$

Arguing vice versa for $\Lambda(w)(x) \leq \Lambda(v)(x)$, this implies that

$$
\|\Lambda(w) - \Lambda(v)\|_\infty \leq \delta \, \|w - v\|_\infty + \nu.
$$

With ν arbitrary and $\delta \in [0, 1)$ this shows that Λ is a contraction mapping on $\mathbb{V}$ with respect to $\|\cdot\|_\infty$. As $\left(\mathbb{V}, \|\cdot\|_\infty\right)$ is a Banach space, the result is established. $\square$