Nonparametric Adaptive Robust Control Under Model Uncertainty

Erhan Bayraktar* Tao Chen †

This Version: March 21, 2022

ABSTRACT: We consider a discrete time stochastic Markovian control problem under model uncertainty. Such uncertainty not only comes from the fact that the true probability law of the underlying stochastic process is unknown, but the parametric family of probability distributions which the true law belongs to is also unknown. We propose a nonparametric adaptive robust control methodology to deal with such problem. Our approach hinges on the following building concepts: first, using the adaptive robust paradigm to incorporate online learning and uncertainty reduction into the robust control problem; second, learning the unknown probability law through the empirical distribution, and representing uncertainty reduction in terms of a sequence of Wasserstein balls around the empirical distribution; third, using Lagrangian duality to convert the optimization over Wasserstein balls to a scalar optimization problem, and adopting a machine learning technique to achieve efficient computation of the optimal control. We illustrate our methodology by considering a utility maximization problem. Numerical comparisons show that the nonparametric adaptive robust control approach is preferable to the traditional robust frameworks.

KEYWORDS: nonparametric adaptive robust control, model uncertainty, stochastic control, adaptive robust dynamic programming, Wasserstein distance, Markovian control problem, utility maximization.

MSC2010: 49L20, 49J55, 93E20, 93E35, 60G15, 65K05, 90C39, 90C40, 91G10, 91G60, 62G05

Contents

*E. Bayraktar is partially supported by the National Science Foundation under grant DMS-2106556 and by the Susan M. Smith chair.

Department of Mathematics, University of Michigan, Ann Arbor
530 Church Street, Ann Arbor, MI 48109, USA
Email: erhan@umich.edu, URL: https://sites.lsa.umich.edu/erhan/

†
Email: chenta@umich.edu, URL: http://taochen.im

1 Introduction

In this paper we propose a new methodology for solving a stochastic Markovian control problem in discrete time under model uncertainty. Unlike many works in this area that assume the unknown probability law of the underlying stochastic process belongs to some parametric family of distributions, we avoid making such postulation to prevent model misspecification. When it comes to handling model uncertainty, there are different approaches, parametric and nonparametric, developed in the past decades to incorporate learning into solving control problems with unknown system models (cf. [KV15], [CG91], [Rie75], [CM20]). However, earlier studies show that a pure learning approach without awareness of the model risk is prone to risk caused by estimation error and often leads to overly aggressive controls and system outcomes with high variances. On the other hand, the central idea of robust control goes back to [GS89]. A large body of research have been devoted to this area since then, and produced fruitful resutls which are briefly summarized in Section 2. Robust techniques are extremely successful in dealing with model risk but if the learning phase is lacking in the framework, corresponding controls can be overly conservative and even trivial. Our work aims to address all the issues mentioned above when handling a Markovian control problem by proposing a nonparametric adaptive robust methodology and develop an efficient numerical scheme for implementing such method.

A robust control problem can be viewed as a game between the controller and the nature. In the traditional setup, the nature chooses the worst case model against the controller at the beginning of the game. To respond, the controller adopts a control law which determines the game strategies at all time steps through the timeline. In a sense, both counterparties' strategies are pre-committed. Mathematically, the controller takes a set of considered models, solves the optimization problem for every model in such set, and chooses the strategy corresponding to the worst model against the controller. We refer to [HSTW06], [HS08], and [BB95], for more information regarding this setup. More recent works consider a robust control problem as a sequential game: from a fixed set of models, at each time step the nature chooses one that is the worst for the controller, and the controller will apply an optimal control in response (cf. [Sir14], [BCP16]). The main difference between the two approaches mentioned so far is that the worst case model is time independent in the former case and time dependent in the latter. In [Nut16], the author presented a robust framework where the nature chooses models from a time dependent set. In other words, the nature can pick strategies from different sets of available actions at different stages of the game. In [BCC+19], the authors specified the dynamics of such sets via recursive confidence regions of the unknown model parameters. We refer to [BCC17] for the detailed discussion of recursive construction of confidence regions. Such idea is also utilized in this work. The advantage of using confidence regions are twofold. On one hand, as new realization of the random noise in the system is observed between the decision-making time points, the confidence region updates itself and naturally represents the learning of the unknown system model. On the other hand, such sequence of sets is asymptotically shrinking in size which leads to reduction of the model uncertainty. To the best of our knowledge, [BCC+19] is the first work that incorporates the idea of online learning into the robust control paradigm. A follow-up work in [BC21] is an attempt to extend the adaptive robust control to the continuous time setup.

Note that the methods in [BCC+19] and [BC21] are parametric and the practical usage of such methods relies on the assumption that the family of the unknown probability law of the underlying stochastic process is known to the controller. Some researchers have realized this drawback and adopts nonparametric statistical methods by assuming uncertainty for the family of parametric models. To formulate a robust setup, one will define a set of probability distributions that includes the estimated distribution. For example, in [KENA19] and [OW21], the authors take a Wasserstein

ball around the empirical distribution and use the ball as the set of considered models. However, such setup has only been implemented in one-period control problems so far, and the feasibility of this approach in multi-period setup remains to be investigated. To overcome this obstacle, we develop a nonparametric adaptive robust control methodology in this work to handle multi-period stochastic control problems where the family of distributions which the true law of the system model belongs to is unknown. Naturally, we use the empirical distribution as the estimate of the distribution of the underlying stochastic process. Another candidate for this purpose is the perturbed empirical distribution when such distribution is known to be continuous. For construction of confidence regions in this setup, we utilize the Wasserstein ball around the empirical distribution. There are several works on the concentration results regarding the empirical distribution and the Wasserstein distance (cf. [DBGM99], [FG15]). Backed by these papers, one obtains a CLT-type of result for the empirical distribution that leads to construction of confidence regions through Wasserstein distance under rather mild assumptions. Practically, numerical search of the worst case model in a set of probability distributions is extremely difficult. Another advantage of using the Wasserstein ball as the confidecnce region is that the aforementioned task of searching for the worst case model in a Wasserstein ball can be converted to a scalar optimization problem. Last but not the least, we implement a machine learning technique via the Gaussian process surrogates [RW06] to build regression models for the relevent value function and the optimal control. The former surrogate enables us to proceed the backward recursion according to the dynamic programming principle, and the latter allows us for fast computation of the optimal control when applying our framework.

The rest of the paper is organized as follows. We begin Section 2 with setting up the model and in Section 2.1 we discuss the contruction of confidence region for the unknown true probability law in terms of the Wasserstein ball. Such sets of distributions represent the uncertainty of the system model. Section 2.2 is dedicated to the formulation of the nonparametric adaptive robust control framework. We investigate the solution of the nonparametric adaptive robust control problem and derive the associated Bellman equations in Section 2.3. Also in this section, we prove the Bellman principle of optimality for the problem and show the existence of measurable worst-case model selector as well as the existence of measurable optimal control. In Section 2.4, we discuss the convergence and deviation of the adaptive robust value function to the true value function. Finally, in Section 3 we consider an illustrative example. Namely, the uncertain utility maximization problem where the investor needs to allocate the wealth between the money market account and the risky asset without knowing the true distribution of the risky asset's return process. We apply the nonparametric adaptive robust control approach to such problem and provide a numerical solver by using machine learning techniques. Numerical results presented in this section show the favorable aspects of the proposed methodology to the traditional robust control framework and the case of knowing the true model.

2 Nonparametric Stochastic Control Problem Subject to Model Uncertainty

Let $(\Omega, \mathscr{F})$ be a measurable space, and $T \in \mathbb{N}$ be a fixed time horizon. Let $\mathcal{T} = \{0, 1, 2, \ldots, T\}$, $\mathcal{T}' = \{0, 1, 2, \ldots, T-1\}$, and $\mathcal{T}'' = \{1, 2, \ldots, T\}$. On the space $(\Omega, \mathscr{F})$ we consider a controlled random process $X = \{X_t,\ t \in \mathcal{T}\}$ taking values in $\mathbb{R}^n$ with dynamics

$$X_{t+1} = S(X_t, \varphi_t, Z_{t+1}), \quad t \in \mathcal{T}', \quad X_0 = x_0 \in \mathbb{R}^n. \tag{2.1}$$

The above $Z = \{Z_t, t \in \mathcal{T}\}$ is an i.i.d. real valued random sequence of which the natural filtration is denoted by $\mathbb{F} = (\mathcal{F}_t, t \in \mathcal{T})$. The process $\varphi = \{\varphi_t, t \in \mathcal{T}'\}$ is $\mathbb{F}$-adapted and takes values in a compact set A. The function $S : \mathbb{R}^n \times A \times \mathbb{R} \to \mathbb{R}^n$ is deterministic and continuous. For every $t \in \mathcal{T}'$, we denote by $\mathcal{A}_t$ the set of all processes that take values in A and are adapted to the filtration $\mathbb{F}_t := (\mathcal{F}_s, t \leq s \leq T-1)$. Each element in $\mathcal{A}_t$ is called an admissible control starting at time t, and we use the convention $\mathcal{A} = \mathcal{A}_0$. In this work, we assume that the process Z is observable but the distribution F^* of each Z_t is unknown. We write $\mathcal{P}(\mathbb{R})$ as the set of all distributions on $\mathbb{R}$ and $\mathbb{P}_F$ as the probability measure on $(\Omega, \mathscr{F})$ corresponds to $F \in \mathcal{P}(\mathbb{R})$. The expectation associated to $\mathbb{P}_F$ is $\mathbb{E}_F$, and $\ell : \mathbb{R}^n \to \mathbb{R}$ is the loss function which is continuous. In this work we will formulate and solve a robust optimization problem aiming to minimize the expected loss when taking into consideration that the true distribution F^* of Z is unknown. In order to avoid model misspecification caused by assuming a wrong parametric family of distrubtions, we will conduct online learning of the underlying system in a nonparametric manner via empirical distribution. In the spirit of [BCC+19], we define the sets of model candidates as approximated confidence regions around the empirical distribution. Such sets are Wasserstein balls and their sizes decreases as time goes on in general. Therefore, in our robust framework, the uncertainty is dynamically reduced through online learning and shrinkage of the Wasserstein balls.

2.1 Empirical Distribution and Uncertainty Set

We make a standing postulation that F^* satisfies that

$$\int_{-\infty}^{\infty} \sqrt{F^*(z)(1-F^*(z))}dz < \infty. \tag{2.2}$$

We note that any distribution that has finite moments with order higher than 2 will satisfy the above assumption. Next, denote by $\widehat{F}_t$, $t \in \mathcal{T}'$, the empirical distrubtion of Z_{t+1} given the initial guess $\widehat{F}_0$ of F^* and the observations $Z_{1:t} := \{Z_i, i = 1, \ldots, t\}$, where $\widehat{F}_0$ is the empirical distribution of Z_1 based on historical data of Z with sample size t_0. In other words, $\widehat{F}_t$ is the contructed based $Z_{-t_0+1:t}$. Defined as an average of indicator functions, $\widehat{F}_t$ satisfies the following recursion similarly to any estimated mean:

$$\widehat{F}_{t+1}(z) = \frac{(t_0+t)\widehat{F}_t + \mathbb{1}_{\{Z_{t+1}<z\}}}{t_0+t+1} := R(t, \widehat{F}_t, Z_{t+1}), \quad z \in \mathbb{R},\ t \in \mathcal{T}'. \tag{2.3}$$

The map R defined above will be viewed as the dynamics of the process $\widehat{F}$. Regarding other properties of $\widehat{F}$, it is well known that $\widehat{F}$ is a consistent estimator of F^*:

$$\lim_{t\to\infty} \widehat{F}_t(z) = F^*(z), \quad \text{a.s..}$$

Moreover, by the assumption (2.2) and using the results in [DBGM99], we have that

$$\sqrt{t_0+t}\, d_{W,1}(\widehat{F}_t, F^*) = \sqrt{t_0+t} \int_{-\infty}^{\infty} |\widehat{F}_t(z) - F^*(z)|dz \to \int_0^1 |B(s)|dQ^*(s), \tag{2.4}$$

where $d_{W,1}$ is the Wasserstein distance of order 1, $B(s)$, $0 \leq s \leq 1$, and Q^* are the Brownian bridge and the quantile function of F^*, respectively, and the convergence is in distribution. We will construct an approximated confidence region for F^* based on (2.4). Since Q^* is unknown, we

will approximate it by using $\widehat{Q}_t$ which is the quantile function corresponding to $\widehat{F}_t$. At time t, the integral $\int_0^1 |B(s)|dQ^*(s)$ is then approximated as

$$\int_0^1 |B(s)|dQ^*(s) \approx \sum_{i=1}^{t_0+t-1} \left| B\left(\frac{i}{t_0+t}\right) \right| \left(z_{(i+1)} - z_{(i)}\right) =: H_t(\widehat{F}_t),$$

where $z_{(1):(t_0+t)}$ is the order statistics of $z_{-t_0+1:t}$. We define the α-uncertainty set $\mathcal{C}_t^\alpha$, $0 < \alpha < 1$, which is an approximated confidence region for F^* as

$$\mathcal{C}_t^\alpha(\widehat{F}_t) = \left\{ F \in \mathcal{P}_1(\mathbb{R}) : d_{W,1}(\widehat{F}_t, F) \le \frac{Q_t^H(1-\alpha)}{\sqrt{t_0+t}} \right\}, \tag{2.5}$$

where $\mathcal{P}_1(\mathbb{R})$ is the set of all distributions with finite first moment, and Q_t^H is the quantile function of $H_t(\widehat{F}_t)$. Due to the discussion above, the rational behind (2.5) is that the probability that $C_t^\alpha(\widehat{F}_t)$ contains F^* is approximately $1-\alpha$. Note that theoretically we can derive the distribution of $\sum_{i=1}^{t_0+t-1} |B(\frac{i}{t_0+t})|(z_{(i+1)} - z_{(i)})$. But since $B(\frac{i}{t_0+t})$, $i = 1, \ldots, t_0+t-1$, are not independent, then such computation will be too tedious. Hence, we will estimate $Q_t^H(1-\alpha)$ via simulation instead. As another way to justify that the radius of the Wasstertein ball being a multiple of $\frac{1}{\sqrt{t_0+t}}$ is the calculation from [FG15] where they show, under the stronger assumption

$$\int_{\mathbb{R}} e^{c_1|z|^{c_2}} F^*(dz) < \infty, \tag{2.6}$$

for some $c_1 > 0$, and $c_2 > 1$, that for any fixed $0 < \alpha < 1$, there exists some constants C and c such that

$$\mathbb{P}\left(d_{W,1}(\widehat{F}_t, F^*) \le \sqrt{\frac{\log(C/\alpha)}{c(t+t_0)}} \right) \ge 1-\alpha. \tag{2.7}$$

A different formulation of the uncertainty sets can be obtained from (2.7). However, radius of the resulting Wasserstein ball has the same order, namely, $\frac{1}{\sqrt{t_0+t}}$ as of $\mathcal{C}_t^\alpha$.

Next, by using a different representation of the Wasserstein distance between probability distributions, we have the following technical result for the map defined in (2.3).

Lemma 2.1. *For fixed $t \in \mathcal{T}'$, the mapping $R(t,\cdot,\cdot) : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \to \mathcal{P}_1(\mathbb{R})$ is continuous.*

Proof. Assume that $(F_n, z_n) \to (F, z)$ where $F_n,\ F \in \mathcal{P}_1(\mathbb{R})$, $z_n,\ z \in \mathbb{R}$, $n = 1, 2, \ldots$. Then, $d_{W,1}(F_n, F) \to 0$ and $z_n \to z$. Denote $\mu_{F_n,z_n} = R(t, F_n, z_n)$ and $\mu_{F,z} = R(t, F, z)$. For $\mathcal{M} := \{f : |f(x) - f(y)| \le |x-y|\}$, we have that

$$\begin{aligned}
d_{W,1}(\mu_{F_n,z_n}, \mu_{F,z}) &= \sup\left\{ \int_{\mathbb{R}} f d\mu_{F_n,z_n} - \int_{\mathbb{R}} f d\mu_{F,z} :\ f \in \mathcal{M} \right\} \\
&= \sup\left\{ \frac{t_0+t}{t_0+t+1} \left(\int_{\mathbb{R}} f dF_n - \int_{\mathbb{R}} f dF \right) + \frac{1}{t_0+t+1}(f(z_n) - f(z)) :\ f \in \mathcal{M} \right\} \\
&\le \frac{t_0+t}{t_0+t+1} \sup\left\{ \int_{\mathbb{R}} f dF_n - \int_{\mathbb{R}} f dF :\ f \in \mathcal{M} \right\} + \frac{1}{t_0+t+1}|z_n - z| \\
&= \frac{t_0+t}{t_0+t+1} d_{W,1}(F_n, F) + \frac{1}{t_0+t+1}|z_n - z|.
\end{aligned}$$

Therefore, we get that $d_{W,1}(\mu_{F_n,z_n}, \mu_{F,z}) \to 0$ and the mapping $R(t,\cdot,\cdot)$ is continuous. □

One property that the set valued function $\mathcal{C}_t^\alpha$ satisfies is upper hemicontinuity (u.h.c.). That is for any for any $F \in \mathcal{P}_1(\mathbb{R})$ and any open set E such that $\mathcal{C}_t^\alpha(F) \subset E \subset \mathcal{P}_1(\mathbb{R})$, there exists a neighbourhood D of F such that for all $F' \in D$, $\mathcal{C}_t^\alpha(F') \subset E$ (cf. [Bor85, Definition 11.3]). To see that $\mathcal{C}_t^\alpha$ is u.h.c., let $\varepsilon = \text{dist}(F, \partial E) - Q_t^H(1-\alpha)/\sqrt{t_0 + t}$ where $\text{dist}(F, \partial E)$ is the shortest distance from F to the boundary of E. Then, take D as the ball centered at F with radius ε. It is not hard to see that for any $F' \in D$, we have $\mathcal{C}_t^\alpha(F') \subset E$. We summarize the result as follows

Lemma 2.2. *For every $t \in \mathcal{T}'$, the set valued function $\mathcal{C}_t^\alpha$ is upper hemicontinuous.*

As per our discussion above, the proof is straightforward and we omit it here.

2.2 Nonparametric Adaptive Robust Control Problem

Now we proceed to formulate the nonparametric adaptive robust control problem. For the rest of the paper, we will consider $\mathcal{P}_1(\mathbb{R})$ with the metric $d_{W,1}$. Since $\mathbb{R}$ is separable and complete, then $(\mathcal{P}_1(\mathbb{R}), d_{W,1})$ is also separable and complete. Hence, $\mathcal{P}_1(\mathbb{R})$ is a Polish space and thus a Borel space. Define the augumented state process $Y = \{Y_t = (X_t, \widehat{F}_t),\ t \in \mathcal{T}\}$, and the augumented state space $E_Y = \mathbb{R}^n \times \mathcal{P}_1(\mathbb{R})$. For E_Y we equip the product topology, it is then a Borel space and the Borel σ-algebra $\mathcal{E}_Y$ conicides with the product σ-algebra. The process Y has the following dynamics

$$Y_{t+1} = \mathbf{G}(t, Y_t, \varphi_t, Z_{t+1}) := (S(X_t, \varphi_t, Z_{t+1}), R(t, \widehat{F}_t, Z_{t+1})), \quad t \in \mathcal{T}'. \tag{2.8}$$

According to the assumption that S is continuous and Lemma 2.1, we get that $\mathbf{G}$ is continuous and therefore Borel measurable. Next, given our setup, the process Y is $\mathbb{F}$-adapted and Markovian. The transition probability for the state process Y is defined as follows. For any $t \in \mathcal{T}'$, $(y, a) \in E_Y \times A$, and $F \in \mathcal{P}_1(\mathbb{R})$, Q_t is a probability measure on $\mathcal{E}_Y$ such that

$$Q_t(D|y, a, F) = \mathbb{P}_F(\mathbf{G}(t, y, a, Z_{t+1}) \in D), \quad D \in \mathcal{E}_Y.$$

One important property of the stochastic kernel Q_t is that it is in fact Borel measurable which will be proved below. Such property is crucial for showing the existence of measurable optimal controls.

Proposition 2.3. *For each $t \in \mathcal{T}'$, the probability $Q_t(\ \cdot\ |y, a, F)$ is a Borel measurable stochastic kernel on E_Y given $E_Y \times A \times \mathcal{P}_1(\mathbb{R})$.*

Proof. According to [BS78], it is enough to show that for any $b_1, b_2 \in \mathbb{R}^n$ and closed ball $D \subseteq \mathcal{P}(\mathbb{R})$ with finite radius, $Q_t([b_1, b_2] \times D|y, a, F)$ is a measurable function in (y, a, F), where

$$[b_1, b_2] := \mathsf{X}_{i=1}^n [b_1^{(i)}, b_2^{(i)}], \quad b_j = (b_j^{(1)}, \ldots, b_j^{(n)}),\ j = 1, 2.$$

We will prove that $Q_t([b_1, b_2] \times D|y, a, F)$ is upper semi-continuous, and then it will be Borel measurable.

Fix any $(y_0, a_0, F_0) \in E_Y \times A \times \mathcal{P}_1(\mathbb{R})$, and let $\{(y_n, a_n, F_n), n > 0\}$ be a sequence that converges to (y_0, a_0, F_0). Note that the set $C_0 := \{z : \mathbf{G}(t, y_0, a_0, z) \in [b_1, b_2] \times D\}$ is a closed set since the map $\mathbf{G}$ is continuous. We similary define $C_n = \{z : \mathbf{G}(t, y_n, a_n, z) \in [b_1, b_2] \times D\}$, $n > 0$, and they satisfy the same properties.

We first prove that $\bigcup_{i=0}^{\infty} C_i$ is bounded. Assume the union contains at least two points. If the two points $z_1 < z_2$ belong to the same C_n, denote by $\hat{f}_n$ the second component of y_n, we have that

$$\begin{aligned} d_{W,1}(\mu_{\hat{f}_n,z_1}, \mu_{\hat{f}_n,z_2}) &= \sup\left\{\int_{\mathbb{R}} g d\mu_{\hat{f}_n,z_1} - \int_{\mathbb{R}} g d\mu_{\hat{f}_n,z_2} : g \in \mathcal{M}\right\} \\ &= \sup\left\{\frac{g(z_1) - g(z_2)}{t_0 + t + 1} : g \in \mathcal{M}\right\} \\ &\geq \frac{z_2 - z_1}{t_0 + t + 1}. \end{aligned}$$

Since D is a bounded set, then z_2 must be within a bounded range of z_1. Next assume that there are $z_k \in C_k$ and $z_l \in C_l$, and $z_l > z_k$. Again, we have

$$\begin{aligned} d_{W,1}(\mu_{\hat{f}_l,z_l}, \mu_{\hat{f}_k,z_k}) &= \sup\left\{\int_{\mathbb{R}} g d\mu_{\hat{f}_l,z_l} - \int_{\mathbb{R}} g d\mu_{\hat{f}_k,z_k} : g \in \mathcal{M}\right\} \\ &\geq \frac{t_0 + t}{t_0 + t + 1}(\mathbb{E}_{\hat{f}_l}[Z_l] - \mathbb{E}_{\hat{f}_k}[Z_k]) + \frac{z_l - z_k}{t_0 + t + 1}. \end{aligned}$$

Since $\{y_n, n > 0\}$ is a convergent sequence, then the value of the first term on the right hand side of the above inequality is bounded for any k and l. Moreover, the value $z_l - z_k$ should be bounded as well. Now we see that $\bigcup_{i=0}^{\infty} C_i$ is bounded and every single C_n is compact. Next, we show that if $C_0 = \emptyset$ then for large enough n the set C_n is also empty. In particular, if the preimage $R^{-1}(t, \hat{f}_0, D) = \emptyset$, then for large enough n, $R^{-1}(t, \hat{f}_n, D) = \emptyset$. Otherwise, we can find a subsequence n_k, $k > 0$, such that there exist $z_{n_k} \in R^{-1}(t, \hat{f}_0, D)$ for all k. Without loss of generality, we assume that z_{n_k} is convergent due to the fact that $\bigcup_{i=0}^{\infty} C_i$ is compact. Then, $R(t, \hat{f}_{n_k}, z_{n_k})$, $k > 0$, is a convergent sequence and $R(t, \hat{f}_{n_k}, z_{n_k}) \in D$, $k > 0$. Because D is closed, the limit $R(t, \hat{f}_0, z_0) \in D$ which implies that $z_0 \in C_1$. This contradicts to the assumption that $R^{-1}(t, \hat{f}_0, D) = \emptyset$, hence for large enough n, $R^{-1}(t, \hat{f}_n, D) = \emptyset$. On the other hand, by using the continuity argument, it is also easy to see that if $S(x_0, a_0, z) \notin [b_1, b_2]$ for all $z \in \mathbb{R}$, then for large enough n, $S(x_n, a_n, z) \notin [b_1, b_2]$ for all $z \in \mathbb{R}$. We have proved that if $C_0 = \emptyset$ then for large enough n the set C_n is also empty. In this case,

$$\lim_{n\to\infty} Q_t([b_1, b_2] \times D | y_n, a_n, F_n) = Q([b_1, b_2] \times D | t, y_0, a_0, F_0) = 0,$$

and the function $Q_t([b_1, b_2] \times D | y, a, F)$ is continuous and therefore upper semi-continuous at such (y_0, a_0, F_0).

For the rest of the proof, we assume that $C_0 \neq \emptyset$. Let $\varepsilon_m > 0$, $m > 0$, be a strictly decreasing sequence that converges to 0. For any ε_m and $z \in \mathbb{R}$, let $\mathcal{B}_m(z)$ be the open ball centered at z with radius ε_m. The collection $\{\mathcal{B}_m(z) : z \in C_0\}$ is an open cover of the compact C_0, and there exists a finite subcover $\mathcal{B}_m(z_{(1)}), \ldots, \mathcal{B}_m(z_{(k_m)})$. Define the set $C_0^m = \bigcup_{i=1}^{k_m} \overline{\mathcal{B}_m(z_{(i)})}$, and we argue that for any $m > 0$, there exists $N_m > 0$ such that for any $n > N_m$, we have $C_n \subseteq C_0^m$.

We prove the above statement by contradiction. Assume that it is not true. Then for any $N > 0$, there exits $n > N$ such that $C_n \not\subseteq C_0^m$. Consequently, there exists a sub-sequence n_j, $j > 0$, such that $z_{n_j} \in C_{n_j}$ but $z_{n_j} \notin C_0^m$. From previous discussions we know the sequence z_{n_j} is bounded, and moreover there exists a z^* that is a limiting point of z_{n_j}. It is safe to assume

$$z^* \notin C_0^m \tag{2.9}$$

for the reason that if z^* is on the boundary of C_0^m, we can replace ε_m with a number in the interval $(\varepsilon_{m+1}, \varepsilon_m)$. Let us consider the sequence $\{\mathbf{G}(t, y_{n_j}, a_{n_j}, z_{n_j}), j > 0\}$. Recall that $z_{n_j} \in C_{n_j}$, hence $\mathbf{G}(t, y_{n_j}, a_{n_j}, z_{n_j}) \in [b_1, b_2] \times D$ for all $j > 0$. Due to Lemma 2.1, $\mathbf{G}(t, y_0, a_0, z^*)$ is a limiting point

of such sequence. In addition, since $[b_1, b_2] \times D$ is a closed set, then $z^* \in C_0 \subseteq C_0^m$ which contradicts to (2.9). Now we conclude that any $m > 0$, there exists $N_m > 0$ such that for any $n > N_m$, we have $C_n \subseteq C_0^m$.

Next, we obtain that

$$Q_t([b_1, b_2] \times D | y_n, a_n, F_n) = \mathbb{P}_{F_n}(C_n) \leq \mathbb{P}_{F_n}(C_0^m). \tag{2.10}$$

Since C_0^m is a closed set, and F_n converges weakly to F_0, then (2.10) implies that

$$\limsup_n Q_t([b_1, b_2] \times D | y_n, a_n, F_n) = \limsup_n \mathbb{P}_{F_n}(C_n) \leq \limsup_n \mathbb{P}_{F_n}(C_0^m) \leq \mathbb{P}_{F_0}(C_0^m).$$

Finally, note that one can construct $\{C_0^m, m > 0\}$ such that the sequence of sets is decreasing and $\bigcap_m C_0^m = C_0$. We have

$$\lim_{m \to \infty} \mathbb{P}_{F_0}(C_0^m) = \mathbb{P}_{F_0}(C_0).$$

It follows immediately that

$$\limsup_n Q_t(b_1, b_2] \times D | y_n, a_n, F_n) \leq Q_t([b_1, b_2] \times D | y_0, a_0, F_0).$$

To summarize, we obtain that $Q_t([b_1, b_2] \times D | \cdot, \cdot, \cdot)$ is upper semi-continuous. Therefore, it is a Borel measurable function. □

In this work, we are dealing with a closed loop feedback control problem. To this end, a control process φ is called Markovian if for every $t \in \mathcal{T}'$ (with a slight abuse of notation)

$$\varphi_t = \varphi_t(Y_t)$$

where on the right hand side $\varphi_t : E_Y \to A$ is a measurable mapping. Similarly, A process ψ is called a Markovian model selector if

$$\psi_t = \psi_t(Y_t)$$

where $\psi_t : E_Y \to \mathcal{P}_1(\mathbb{R})$ is measurable. In the adaptive robust framework, we consider the Markovian control processes and Markovian model selectors such that $\psi_t(y) \in \mathcal{C}_t^\alpha(y)$ for any $y \in E_Y$. For every $t \in \mathcal{T}'$, any time t state $y_t \in E_Y$, and control process $\varphi \in \mathcal{A}_t$, we denote

$$\mathbf{\Psi}_{y_t,t}^{\varphi} = \{\psi_{t:T-1}, \psi_s(y_s) \in \mathcal{C}_s^\alpha(y_s), \exists z \in \mathbb{R}, \text{s.t. } y_{s+1} = \mathbf{G}(s, y_s, \varphi_s(y_s), z), t \leq s < T-1\}.$$

and

$$\mathbf{\Psi}_{y_t,t} = \{\psi_{t:T-1}, \psi_s(y_s) \in \mathcal{C}_s^\alpha(y_s), \exists a \in A, z \in \mathbb{R}, \text{s.t. } y_{s+1} = \mathbf{G}(s, y_s, a, z), t \leq s < T-1\}.$$

Next, for every $t \in \mathcal{T}'$, any $y_t \in E_Y$, $\varphi \in \mathcal{A}_t$, and $\psi \in \mathbf{\Psi}_{y_t,t}$, we define the probability measure $\mathbb{Q}_{y_t,t}^{\varphi,\psi}$ on the concatenated canonical space $\mathsf{X}_{s=t+1}^T E_Y$ as

$$\mathbb{Q}_{y_t,t}^{\varphi,\psi}(B_{t+1} \times \cdots \times B_T) = \int_{B_{t+1}} \cdots \int_{B_T} \prod_{u=t}^{T-1} Q_u(dy_{u+1} | y_u, \varphi_u(y_u), \psi_u(y_u)).$$

Correspondingly, we define the family of probability measures $\mathcal{Q}_{y_t,t}^{\varphi} = \{\mathbb{Q}_{y_t,t}^{\varphi,\psi}, \psi \in \mathbf{\Psi}_{y_t,t}\}$. In particular, we let $\mathcal{Q}_{y_0}^{\varphi} = \mathcal{Q}_{y_0,0}^{\varphi}$. Then, for given $y_0 \in E_Y$, the nonparametric adaptive robust control problem is formulated as

$$\inf_{\varphi \in \mathcal{A}} \sup_{\mathbb{Q} \in \mathcal{Q}_{y_0}^{\varphi}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)]. \tag{2.11}$$

In a traditional robust setup, one would choose a fixed set $\mathcal{P}_0$ in place of $\mathcal{C}_t^\alpha$. Due to such reason, we will call it the static robust framework throughout. In comparison, the advantage of (2.11) is that such framework integrates robust control with learning and reducing uncertainty. The learning of the unknown model is carried through via the evolution of the process Y, and reduction of uncertainty is embedded in the construction of $\mathcal{Q}_{y_0}^\varphi$ since for any $y_t \in E_Y$ instead of finding the worst case model in the fixed set $\mathcal{P}_0$, the selectors take values in the uncertainty sets $\mathcal{C}_t^\alpha(y_t)$ which is a sequence of random sets that shrink in size.

2.3 Solution of Nonparametric Adaptive Robust Control Problem

We will show that solution of the nonparametric adaptive robust control problem is given by solving the following adaptive robust Bellman equations

$$\begin{aligned} V_T(y) &= \ell(x), \quad y \in E_Y, \\ V_t(y) &= \inf_{a\in A} \sup_{F\in\mathcal{C}_t^\alpha(y)} \int_{E_Y} V_{t+1}(y_{t+1}) Q_t(dy_{t+1}|y,a,F), \quad y \in E_Y, t \in \mathcal{T}'. \end{aligned} \tag{2.12}$$

Before we prove the main theorem in this section, let us first provide the following technical result.

Lemma 2.4. *Fix $t \in \mathcal{T}'$, for any $\widehat{F} \in \mathcal{P}_1(\mathbb{R})$, let*

$$\widetilde{\mathcal{C}}_t^\alpha(\widehat{F}) = \left\{ F \in \mathcal{P}_1(\mathbb{R}) : d_{W,1}(F,\widehat{F}) < \frac{Q_t^H(1-\alpha)}{\sqrt{t_0+t}} \right\}.$$

Then,

$$\mathcal{O}_t^\alpha := \bigcup_{y\in E_Y} \widetilde{\mathcal{C}}_t^\alpha(y)$$

is an open set in $E_Y \times \mathcal{P}_1(\mathbb{R})$.

Proof. We prove the statement by contradiction. Assume there exists $(y_0, F_0) \in \mathcal{O}_t^\alpha$, and there exists a sequence $(y_n, F_n) \to (y_0, F_0)$, such that for any $n > 0$, $(y_n, F_n) \notin \mathcal{O}_t^\alpha$. Note that $F_0 \in \widetilde{\mathcal{C}}_t^\alpha(y_0)$, hence $d_{W,1}(F_0, \hat{f}_0) < Q_t^H(1-\alpha)/\sqrt{t_0+t} - \varepsilon$ for some $\varepsilon > 0$, where $\hat{f}_0$ is the second component of y_0. We have

$$d_{W,1}(F_n, \hat{f}_n) \le d_{W,1}(F_n, F_0) + d_{W,1}(F_0, \hat{f}_0) + d_{W,1}(\hat{f}_0, \hat{f}_n).$$

For large enough n, we have $d_{W,1}(F_n, F_0) < \varepsilon/4$, and $d_{W,1}(\hat{f}_0, \hat{f}_n) < \varepsilon/4$. Then, for such n, the following equality holds true

$$d_{W,1}(F_n, \hat{f}_n) \le \frac{Q_t^H(1-\alpha)}{\sqrt{t_0+t}} - \frac{\varepsilon}{2},$$

which implies that $F_n \in \widetilde{\mathcal{C}}_t^\alpha(y_n)$ and $(y_n, F_n) \in \mathcal{O}_t^\alpha$. We get the contradiction so the set $\mathcal{O}_t^\alpha$ is open. □

Next, we have the main result of this section which shows that the optimal control φ and model selector ψ exist, and they are sequences of measurable functions.

Theorem 2.5. *For every $t \in \mathcal{T}$, the function V_t is lower semicontinuous (l.s.c.) and upper semianalytic (u.s.a.). There exists Borel measurable optimal control φ_t^*, $t \in \mathcal{T}'$, and universally measurable model selector ψ_t^*, $t \in \mathcal{T}'$.*

Proof. Since $V_T(y) = \ell(x)$ which is continuous by assumption, then V_T is l.s.c. and u.s.a.. Next, denote

$$v_{T-1}(y,a,F) = \int_{E_Y} V_T(y_T) Q_{T-1}(dy_T|y,a,F).$$

By using Proposition 2.3, we have that $v_{T-1}(y,a,F)$ is u.s.a.. Let $D = \bigcup_{(y,a)\in E_Y\times A} \mathcal{C}^\alpha_{T-1}(y)$. Note that $\mathcal{C}^\alpha_{T-1}$ is u.h.c. from Lemma 2.2 and closed valued, by adopting the proof of [BCC21] in our setup, we obtain that the graph of $\mathcal{C}^\alpha_{t-1}$, which is D, is closed. Hence, the set D is analytic. The (y,a) section of D is $\mathcal{C}^\alpha_{T-1}(y)$, and according to [BS78], we get that

$$\widetilde{v}_{T-1}(y,a) := \sup_{F\in\mathcal{C}^\alpha_{T-1}(y)} v_{T-1}(y,a,F)$$

is u.s.c.. Moreover, for any $\varepsilon > 0$, there exists an analytically measurable function $\widetilde{\psi} : E_Y \times A$ such that for any (y,a),

$$v_{T-1}(y,a,\widetilde{\psi}(y,a)) \geq \begin{cases} \widetilde{v}_{T-1}(y,a) - \varepsilon, & \text{if } \widetilde{v}_{T-1}(y,a) < \infty, \\ 1/\varepsilon, & \text{if } \widetilde{v}_{T-1}(y,a) = \infty. \end{cases} \tag{2.13}$$

Define the set

$$I = \left\{(y,a) \in E_Y \times A : \text{for some } F^* \in \mathcal{C}^\alpha_{T-1}(y), v_{T-1}(y,a,F^*) = \widetilde{v}_{T-1}(y,a)\right\},$$

and we claim that $I = E_Y \times A$. That is for any $(y,a) \in E_Y \times A$ there exists an F^* such that $v_{T-1}(y,a,F^*) = \widetilde{v}_{T-1}(y,a)$. To see why this is true, by taking in (2.13) $\varepsilon = 1/n$, we obtain a sequence $\widetilde{\psi}_n$ such that

$$\lim_{n\to\infty} v_{T-1}(y,a,\widetilde{\psi}_n(y,a)) = \widetilde{v}_{T-1}(y,a).$$

Next, note that $\mathcal{C}^\alpha_{T-1}(y)$ is weakly compact, so there exists $\widetilde{\psi}^*(y,a)$ as a limiting point of $\widetilde{\psi}_n(y,a)$, $n > 0$, such that $v_{T-1}(y,a,\widetilde{\psi}^*(y,a)) = \widetilde{v}_{T-1}(y,a)$, and indeed $I = E_Y \times A$. Therefore, by [BS78], there exists a universally measurable function $\psi^*_{T-1} : E_Y \times A \to \mathcal{C}^\alpha_{T-1}(y)$ which satisfies

$$v_{T-1}(y,a,\psi^*_{T-1}(y,a)) = \widetilde{v}_{T-1}(y,a).$$

Now we prove that the function $\widetilde{v}_{T-1}(y,a)$ is l.s.c.. To this end, we write

$$v_{T-1}(y,a,F) = \int_{\mathbb{R}} V_T(\mathbf{G}(T-1,y,a,z)) dF(z).$$

Since V_T is l.s.c. and $\mathbf{G}(T-1,y,a,z)$ is continuous in (y,a,z), then $V_T(\mathbf{G}(T-1,y,a,z))$ is l.s.c.. On the other hand, F is clearly a continuous stochastic kernel on $\mathbb{R}$ given $\mathcal{P}_1(\mathbb{R})$. In view of the assumption that ℓ is bounded below, we know the function $v_{T-1}(y,a,F)$ is l.s.c.. Let us consider the optimization problem

$$\widehat{v}_{T-1}(y,a) = \sup_{F\in\widetilde{\mathcal{C}}^\alpha_t(y)} v_{T-1}(y,a,F).$$

Lemma 2.4 shows that the set $\mathcal{O}^\alpha_t$ is open in $E_Y \times \mathcal{P}_1(\mathbb{R})$, so it is also open in $D' := E_Y \times A \times \mathcal{P}_1(\mathbb{R})$. The $\widetilde{\mathcal{C}}^\alpha_t(y)$ is the (y,a) section of D'. By [BS78], we obtain that $\widehat{v}_{T-1}(y,a)$ is l.s.c.. Note for any $y \in E_Y$, the uncertainty set $\mathcal{C}^\alpha_t(y)$ is the closure of $\widetilde{\mathcal{C}}^\alpha_t(y)$, it follows immediately that $\widetilde{v}_{T-1}(y,a) = \widehat{v}_{T-1}(y,a)$ and the former is therefore l.s.c..

It remains to show that

$$V_{T-1}(y) = \inf_{a\in A} \widetilde{v}_{T-1}(y,a)$$

is l.s.c., and there exists a Borel measruable function $\varphi^* : E_Y \to A$ such that

$$V_{T-1}(y) = \widetilde{v}_{T-1}(y, \varphi^*(y)).$$

Towards this end, we note that $D'' = E_Y \times A$ is closed, and A by assumption is compact. The y section of D'' is A for any $y \in E_Y$. Thus, by [BS78], the function V_{T-1} is l.s.c., and the Borel measurable optimal control φ^* exists.

We shall prove the statement of all $t = T-2, \dots, 0$ by backward induction. Recall from Proposition 2.3, the stochastic kernel $Q_{T-2}(\,\cdot\,|y,a,F)$ is Borel measurable. Also, the function V_{T-1} is u.s.a.. Therefore,

$$v_{T-2}(y,a,F) = \int_{E_Y} V_{T-1}(y_{T-1}) Q_{T-2}(dy_{T-1}|y,a,F)$$

is u.s.a.. By using a similar argument as above, the function

$$v_{T-2}(y,a,F) = \int_{E_Y} V_{T-1}(\mathbf{G}(T-2,y,a,z)) dF(z)$$

is l.s.c.. The rest of the proof follows analogously. □

Finally, we show that the problem (2.11) will be solved by the adaptive robust Bellman equations (2.12). To this end, we introduce the set $\mathcal{A}_t = \{\varphi_{t:T-1}, t \in \mathcal{T}'\}$, and provide the following technical results for preparation.

Lemma 2.6. *For every $t \in \mathcal{T}'$, and any $\varphi \in \mathcal{A}_t$, the function*

$$\sup_{\mathbb{Q}\in\mathcal{Q}^{\varphi}_{y_t,t}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)]$$

is upper semianalytic in y_t.

The proof for this lemma is a direct modification of Theorem 2.5 and hence we omit it here. Such result ensures that the mentioned function is measurable and can be integrated. Now we are ready to present the solution of the adaptive robust control problem.

Theorem 2.7. *For every $t \in \mathcal{T}'$, and any $y_t \in E_Y$, we have*

$$V_t(y_t) = \inf_{\varphi\in\mathcal{A}_t} \sup_{\mathbb{Q}\in\mathcal{Q}^{\varphi}_{y_t,t}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)].$$

Moreover, with φ^*_t and ψ^*_t, $t \in \mathcal{T}'$, in Theorem 2.5, we get

$$\inf_{\varphi\in\mathcal{A}_t} \sup_{\mathbb{Q}\in\mathcal{A}^{\varphi}_{y_t,t}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] = \mathbb{E}_{\mathbb{Q}^{\varphi^*_{t:T-1},\psi^*_{t:T-1}}_{y_t,t}}[\ell(X_T)].$$

Proof. We prove the result via backward induction in $t = T-1, \dots, 1, 0$.

First, for $t = T-1$ and $y_{T-1} \in E_Y$, we have

$$\inf_{\varphi \in \mathcal{A}_{T-1}} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi}_{y_{T-1}, T-1}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] = \inf_{a \in A} \sup_{F \in \mathcal{C}^{\alpha}_{T-1}(y_{T-1})} \int_{E_Y} V_T(y_T) Q_{T-1}(y_T | y_{T-1}, a, F) = V_{T-1}(y_{T-1}).$$

Next, for $t = T-2, \dots, 0$ and $y_t \in E_Y$, by induction

$$\begin{aligned}
\inf_{\varphi \in \mathcal{A}_t} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi}_{y_t, t}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] &= \inf_{(\varphi_t, \varphi_{t+1:T-1}) \in \mathcal{A}_t} \sup_{F \in \mathcal{C}^{\alpha}_t(y_t)} \int_{E_Y} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi_{t+1:T-1}}_{y_{t+1}, t+1}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] Q_t(dy_{t+1} | y_t, \varphi_t(y_t), F) \\
&\geq \inf_{(\varphi_t, \varphi_{t+1:T-1}) \in \mathcal{A}_t} \sup_{F \in \mathcal{C}^{\alpha}_t(y_t)} \int_{E_Y} V_{t+1}(y_{t+1}) Q_t(dy_{t+1} | y_t, \varphi_t(y_t), F) \\
&= \inf_{a \in A} \sup_{F \in \mathcal{C}^{\alpha}_t(y_t)} \int_{E_Y} V_{t+1}(y_{t+1}) Q_t(dy_{t+1} | y_t, a, F) = V_t(y_t),
\end{aligned}$$

where the inequality is due to that

$$\sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi_{t+1:T-1}}_{y_{t+1}, t+1}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] \geq \inf_{\varphi \in \mathcal{A}_{t+1}} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi}_{y_{t+1}, t+1}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] = V_{t+1}(y_{t+1}).$$

On the other hand, for any $\varepsilon > 0$, let $\varphi^{\varepsilon}_{t+1:T-1} \in \mathcal{A}_{t+1}$ be an ε-optimal control starting at time $t+1$. We get

$$\sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi^{\varepsilon}_{t+1:T-1}}_{y_{t+1}, t+1}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] \leq \inf_{\varphi \in \mathcal{A}_{t+1}} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi}_{y_{t+1}, t+1}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] = V_{t+1}(y_{t+1}) + \varepsilon.$$

It is followed by

$$\begin{aligned}
\inf_{\varphi \in \mathcal{A}_t} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi}_{y_t, t}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] &= \inf_{(\varphi_t, \varphi_{t+1:T-1}) \in \mathcal{A}_t} \sup_{F \in \mathcal{C}^{\alpha}_t(y_t)} \int_{E_Y} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi_{t+1:T-1}}_{y_{t+1}, t+1}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] Q_t(dy_{t+1} | y_t, \varphi_t(y_t), F) \\
&\leq \inf_{(\varphi_t, \varphi_{t+1:T-1}) \in \mathcal{A}_t} \sup_{F \in \mathcal{C}^{\alpha}_t(y_t)} \int_{E_Y} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi^{\varepsilon}_{t+1:T-1}}_{y_{t+1}, t+1}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] Q_t(dy_{t+1} | y_t, \varphi_t(y_t), F) \\
&\leq \inf_{a \in A} \sup_{F \in \mathcal{C}^{\alpha}_t(y_t)} \int_{E_Y} V_{t+1}(y_{t+1}) Q_t(dy_{t+1} | y_t, a, F) + \varepsilon \\
&= V_t(y_t) + \varepsilon.
\end{aligned}$$

Since ε is arbitrary, we obtain

$$\inf_{\varphi \in \mathcal{A}_t} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi}_{y_t, t}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] \leq V_t(y_t).$$

Hence, we have

$$\inf_{\varphi \in \mathcal{A}_t} \sup_{\mathbb{Q} \in \mathcal{Q}^{\varphi}_{y_t, t}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)] = V_t(y_t).$$

To see that φ^* and ψ^* in Theorem 2.5 solve the adaptive robust control problem, we just need to note that for every $t \in \mathcal{T}'$

$$\begin{aligned}
V_t(y_t) &= \int_{E_Y} V_{t+1}(y_{t+1}) Q_t(dy_{t+1}|y_t, \varphi_t^*(y_t), \psi_t^*(y_t)) \\
&= \int_{E_Y} \int_{E_Y} V_{t+2}(y_{t+2}) Q_{t+1}(dy_{t+2}|y_t, \varphi_t^*(y_{t+1}), \psi_{t+1}^*(y_{t+1})) Q_t(dy_{t+1}|y, \varphi_t^*(y_t), \psi_t^*(y_t)) \\
&= \int_{E_Y} \cdots \int_{E_Y} V_T(y_T) \prod_{s=t}^{T-1} Q_{s+1}(dy_s|y_s, \varphi_s^*(y_s), \psi_s^*(y_s)) \\
&= \mathbb{E}_{\mathbb{Q}_{y_t,t}^{\varphi_{t:T-1}^*, \psi_{t:T-1}^*}}[\ell(X_T)],
\end{aligned}$$

where the above $\psi_s^*(y_s) = \psi_s^*(y_s, \varphi_s^*(y_s))$, $t \in \mathcal{T}'$, can be viewed as a composition of universally measurable functions and therefore universally measurable. □

2.4 Convergence Analysis

A nice property of the combination of Wasserstein metric and adaptive robust control is that convergence analysis can be done in such framework very easily. As shown in Theorem 2.5 and 2.7, to deal with (2.11) one employs the dynamic programming principle and solves the following Bellman equation

$$V_t(y) = \inf_{a \in A} \sup_{F \in \mathcal{C}_t^\alpha(y)} \mathbb{E}_F[V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))], \quad t \in \mathcal{T}'.$$

According to [BDOW21, Theorem 2], by assuming V and S to be differentiable w.r.t. x, and denoting

$$V_t^a(y) = \sup_{F \in \mathcal{C}_t^\alpha(y)} \mathbb{E}_F[V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))],$$

we get that

$$V_t^a(y) = \mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))] + \frac{Q_t^H(1-\alpha)}{\sqrt{t_0+t}} \mathbb{E}_{\widehat{F}_t}\left[\left|\frac{\partial}{\partial x} V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))\right|\right] + o\left(\frac{1}{\sqrt{t_0+t}}\right). \tag{2.14}$$

For any given state $y = (x, \hat{f}) \in E_Y$, denote by $z_{-t_0+1:0}$ the historical sample points that generate $\widehat{F}_0$, and let $z_{1:t}$ be the observations of Z such that $\hat{f}(z) = \frac{\sum_{i=-t_0+1}^{t} \mathbb{1}_{\{z_i < z\}}}{t_0+t}$. The following expectation is computed as

$$\mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))] = \frac{1}{t_0+t} \sum_{i=-t_0+1}^{t} V_{t+1}(\mathbf{G}(t, y, a, z_i)),$$

which is the sample mean of the random variable $V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))$ given sample $z_{-t_0+1:t}$. By central limit theorem, we obtain that the convergence speed of $\mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))]$ to the expectation $\mathbb{E}_{F^*}[V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))]$ is asymptotically of order $\frac{1}{\sqrt{t_0+t}}$. Thus, as t increases the adaptive robust control problem converges to the control problem without uncertainty and the convergence speed is of order $\frac{1}{\sqrt{t_0+t}}$. Moreover, we get by using the Chebyshev inequality that

$$\mathbb{P}\left(\left|\mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))] - \overline{\mu}_V^*\right| > \varepsilon\right) \leq \frac{\mathrm{Var}\left(\mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t, y, a, Z_{t+1}))]\right)}{(t_0+t)\varepsilon^2}, \tag{2.15}$$

where $\overline{\mu}_V^* = \mathbb{E}_{F^*}[V_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))]$. Inequality (2.15) implies that the first term on the right hand side in (2.14) has a high probability of being close to $\mathbb{E}_{F^*}[V_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))]$. For example, taking $\varepsilon = \frac{1}{\sqrt{t_0+t}}$, then (2.15) implies

$$\mathbb{P}\left(\left|\mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))] - \overline{\mu}_V^*\right| > \frac{1}{\sqrt{t_0+t}}\right) \leq \mathrm{Var}\left(\mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))]\right).$$

Clearly, the above probability will continue to decrease as t increases. Note that with a further assumption given in the Cramer's Theorem:

$$\int_{\mathbb{R}} e^{\theta z} F^*(dz) < \infty, \quad \forall \theta \in \mathbb{R}, \tag{2.16}$$

we have

$$\lim_{t\to\infty} \frac{1}{t+t_0} \log \mathbb{P}\left(\left|\mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))] - \overline{\mu}_V^*\right| > \varepsilon\right) = -\sup_{\theta\in\mathbb{R}}((\overline{\mu}_V^* + \varepsilon)\theta - \log \mathbb{E}_{F^*}[e^{\theta Z_{t+1}}]).$$

As a result, the probability that $\mathbb{E}_{\widehat{F}_t}[V_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))]$ deviates from the true value function for more than ε has an exponential decay in time with speed $-\sup_{\theta\in\mathbb{R}}((\overline{\mu}_V^* + \varepsilon)\theta - \log \mathbb{E}_{F^*}[e^{\theta Z_{t+1}}])$ which is an obvious improvement over (2.15). In summary, if assuming (2.16) and using $\mathcal{C}_t^\alpha(y)$ as the uncertainty set, even though the overall convergence speed of V_t is still of order $\frac{1}{\sqrt{t_0+t}}$, we obtain a more accurate value function compared to the true one.

Based on (2.14), we can compare the adaptive robust framework to the static robust setup in a qualitative manner. For the latter, the uncertainty set is fixed for all $t \in \mathcal{T}'$, and we denote it by $\mathcal{P}_0$. To solve the static robust control problem, one also utilizes the dynamic programming principle and solves

$$\widetilde{V}_t(y) = \inf_{a\in A} \sup_{F\in\mathcal{P}_0} \mathbb{E}_F[\widetilde{V}_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))], \quad t \in \mathcal{T}',$$

where $\widetilde{V}$ is the corresponding value function. We consider the set $\mathcal{P}_0$ defined as $\mathcal{B}_\delta(\widehat{F}_0)$ which is a Wasserstein ball arond $\widehat{F}_0$ with radius δ. We also define a preference relation between value functions via

$$V_t(y) \succeq \widetilde{V}_t(y) \iff \sup_{F\in\mathcal{C}_t^\alpha(y)} \mathbb{E}_F[V_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))] \leq \sup_{F\in\mathcal{B}_\delta(\widehat{F}_0)} \mathbb{E}_F[\widetilde{V}_{t+1}(\mathbf{G}(t,y,a,Z_{t+1}))],$$

for any $a \in A$. Next, suppose that $F^* \in \mathcal{P}_0^o$ which is the interior of the set $\mathcal{P}_0$. For large T and t, we have $d_{W,1}(\widehat{F}_t, F^*) < d_{W,1}(F^*, \partial\mathcal{P}_0)$ with high probability, where $d_{W,1}(F^*, \partial\mathcal{P}_0)$ is the Wasserstein distance from between F^* and the closest point on the boundary of $\mathcal{P}_0$. Consequently, $\mathcal{C}_t^\alpha(Y_t) \subset \mathcal{P}_0$ with high probability, and loosely speaking we get

$$V_t(y) \succeq \widetilde{V}_t(y), \tag{2.17}$$

asymptotically. Note that such discussion is rather qualitative since it is not easy to compute $\mathbb{P}(\mathcal{C}_t^\alpha(Y_t) \subset \mathcal{P}_0)$ and prove (2.17) rigorously. Nevertheless, we argue that adaptive robust framework is more preferrable than static robust.

For a more quantitative analysis, we assume that $\mathcal{P}_0 = \mathcal{C}_0^\alpha(y_0)$. Similarly to (2.14), we have

$$\widetilde{V}_t^a(x) = \mathbb{E}_{\widehat{F}_0}[\widetilde{V}_{t+1}(S(x,a,Z_{t+1}))] + \frac{Q_0^H(1-\alpha)}{\sqrt{t_0}} \mathbb{E}_{\widehat{F}_0}\left[\left|\frac{\partial}{\partial x}\widetilde{V}_{t+1}(S(x,a,Z_{t+1}))\right|\right] + o(1).$$

It is obvious that the right hand side of the above equality does not converge with respect to t. As a result, the static robust framework will produce strategies that in general distant from the optimal strategies without uncertainty. Such strategies behave very conservatively while adaptive robust has a better balance between being aggressive and conservative due to the embedded learning feature. In view of such, the adaptive robust methodology is more favorable compared to the static robust framework which offers no convergence to the true optimization problem.

Note that discussions in this section are possible since we are using the Wasserstein metric to define the uncertainty sets. Similar analysis could be done when utilizing the Kullback-Leibler divergence but stronger assumptions on the considered probability distributions are required.

3 Nonparametric Adaptive Robust Utility Maximization

In this section, we consider a utility maximization problem under model uncertainty and we will solve it under the nonparametric adaptive robust framework. To this end, we take X to be the investor's wealth process. Any portfolio includes two assets: a banking account with 1-period return $1+r$, where r is the interest rate and fixed throughout, and a stock with i.i.d. log-return Z_t, $t \in \mathcal{T}''$, of which the distribution F^* is unknown. For each $t \in \mathcal{T}'$, denote by φ_t the ratio of the wealth invested in the stock. We rule out leverage and short selling, so φ_t takes values in $A = [0,1]$. Imposing the self-financing strategy, and given $X_0 = x_0 > 0$, the dynamics of X is given by

$$X_{t+1} = X_t((1-\varphi_t)(1+r) + \varphi_t e^{Z_{t+1}}), \quad t \in \mathcal{T}'.$$

Take $n = 1$, and the function S is defined on $\mathbb{R} \times A \times \mathbb{R}$. The prices of the risky asset are observable and thus the return process Z of the risky asset is also observable. We will use the observations of Z to construct the empirical distribution iteratively as in (2.3). Then, we build the α-uncertainty sets for the distribution F of Z according to (2.5). Next, by taking $\ell(x) = \frac{e^{-\eta x}-1}{\eta}$ for some $\eta > 0$, we formulate the nonparametric adaptive robust utility maximization problem as

$$\inf_{\varphi \in \mathcal{A}} \sup_{\mathbb{Q} \in \mathcal{Q}^{\alpha}_{y_0}} \mathbb{E}_{\mathbb{Q}}[\ell(X_T)],$$

where $y_0 = (x_0, \widehat{F}_0)$ such that $\widehat{F}_0$ is the initial guess of F^*. Note that the funtion ℓ is bounded and we are equivalently dealing with

$$\sup_{\varphi \in \mathcal{A}} \inf_{\mathbb{Q} \in \mathcal{Q}^{\alpha}_{y_0}} \mathbb{E}_{\mathbb{Q}} \left[\frac{1 - e^{-\eta X_T}}{\eta} \right] \tag{3.1}$$

which is a maximization problem of the exponential utility function. Due to Theorem 2.7, we will solve the following Bellman equations to get the solution of (3.1).

$$\begin{aligned} V_T(y) &= \frac{1-e^{-\eta x}}{\eta}, \\ V_t(y) &= \sup_{a \in A} \inf_{F \in \mathcal{C}^{\alpha}_t(y)} \int_{E_Y} V_{t+1}(y_{t+1}) Q_t(dy_{t+1}|y,a,F), \quad t \in \mathcal{T}'. \end{aligned} \tag{3.2}$$

Moreover, by applying Theorem 2.5, we get that the optimal trading strategies and worst case models exist which are optimizers of (3.2).

Remark 3.1. Several types of utility functions satisfy the assumptions in Theorem 2.5 so that the corresponding optimal trading strategies and worst case models exist, and the adaptive robust control problem can be solved by utilizing the dynamic programming principle. Another example of such utility functions is the power utility $\frac{x^{1-\eta}-1}{1-\eta}$ where $\eta > 1$.

Note that the loss function $\ell(x) = \frac{e^{-\eta x}-1}{\eta}$ is not only bounded from below but actually bounded. Here we provide the following technical result regarding the corresponding value functions.

Proposition 3.2. *The value function $V_t(y)$ as in (2.12) is lower semicontinuous for every $t \in \mathcal{T}'$.*

Proof. The function $V_T(y) = \frac{1-e^{-\eta x}}{\eta}$ is clearly continuous and hence u.s.c.. Because $\mathbf{G}(T-1, y, a, z)$ is continuous in (y, a, z), $V_T(\mathbf{G}(T-1, y, a, z))$ is l.s.c. in (y, a, z). Moreover,

$$v_{T-1}(y, a, F) = \int_{\mathbb{R}} V_T(\mathbf{G}(T-1, y, a, z))dF(z)$$

is l.s.c. due to that W_T is bounded.

Consider the set $D = \bigcup_{(y,a)\in E_Y \times A} \mathcal{C}^{\alpha}_{T-1}(y)$, and define the function

$$\check{v}_{T-1}(y, a, F) = \begin{cases} v_{T-1}(y, a, F) & \text{if } (y, a, F) \in D, \\ \infty & \text{otherwise.} \end{cases}$$

For any $c \in \mathbb{R}$, we have

$$\begin{aligned} &\{(y, a, F) \in E_Y \times A \times \mathcal{P}_1(\mathbb{R}) \mid \check{v}_{T-1}(y, a, F) \le c\} \\ =&\{(y, a, F) \in E_Y \times A \times \mathcal{P}_1(\mathbb{R}) \mid v_{T-1}(y, a, F) \le c\} \cap D. \end{aligned}$$

Since v_{T-1} is l.s.c., and D is closed, then the set $\{(y, a, F) \in E_Y \times A \times \mathcal{P}_1(\mathbb{R}) \mid \check{v}_{T-1}(y, a, F) \le c\}$ is closed and $\check{v}_{T-1}(y, a, F)$ is l.s.c.. Next, for any $c \in \mathbb{R}$,

$$\begin{aligned} &\left\{(y, a) \in E_Y \times A \mid \inf_{F\in\mathcal{C}^{\alpha}_{T-1}(y)} v_{T-1}(y, a, F) \le c\right\} \\ =&\left\{(y, a) \in E_Y \times A \mid \inf_{F\in\mathcal{P}_1(\mathbb{R})} \check{v}_{T-1}(y, a, F) \le c\right\}. \end{aligned}$$

Fix (y, a) and let $\{F_n, n > 0\} \subset \mathcal{P}_1(\mathbb{R})$ be such that

$$\check{v}(y, a, F_n) \downarrow \widetilde{v}_{T-1}(y, a) := \inf_{F\in\mathcal{P}_1(\mathbb{R})} \check{v}_{T-1}(y, a, F).$$

By definition of $\check{v}_{T-1}$, we know for large enough n, $F_n \in \mathcal{C}^{\alpha}_{T-1}(y)$ which is a weakly compact set. Then, there exists F^* such that $\check{v}_{T-1}(y, a, F^*) = \widetilde{v}_{T-1}(y, a)$. Let $\{(y_n, a_n), n > 0\}$ be a sequence that converges to some (y_0, a_0). We choose a sequence $\{F_n, n > 0\} \subset \mathcal{P}(\mathbb{R})$ such that $\check{v}_{T-1}(y_n, a_n, F_n) = \widetilde{v}_{T-1}(y_n, a_n)$. Obviously, for each $n > 0$, $F_n \in \mathcal{C}^{\alpha}_{T-1}(y_n)$. Due to the fact that $\{y_n, n > 0\}$ converges to y_0, the set $\widetilde{D} = \bigcup_n \mathcal{C}^{\alpha}_{T-1}(y_n)$ is bounded. Hence, there exists $F' \in \widetilde{D}$ and $\delta > 0$ such that $\widetilde{D} \subset \mathcal{B}_\delta(F')$ where the latter is a Wasserstein ball around F' with radius δ.

Now we consider the topology consistent with the weak convergence for the argument F in the function $\check{v}_{T-1}(y, a, F)$. In such case, $\check{v}_{T-1}$ is still l.s.c.. There exists a subsequence $(y_{n_k}, a_{n_k}, F_{n_k})$, $k > 0$, such that

$$\liminf_{n\to\infty} \check{v}_{T-1}(y_n, a_n, F_n) = \lim_{k\to\infty} \check{v}_{T-1}(y_{n_k}, a_{n_k}, F_{n_k}).$$

As $\mathcal{B}_\delta(F')$ is compact under the Prokhorov metric, there exists F_0 that is a limit point of $\{F_{n_k}, n > 0\}$. We obtain

$$\begin{aligned} \liminf_{n\to\infty} \widetilde{v}_{T-1}(y_n, a_n) = \liminf_{n\to\infty} \check{v}_{T-1}(y_n, a_n, F_n) &= \lim_{k\to\infty} \check{v}_{T-1}(y_{n_k}, a_{n_k}, F_{n_k}) \\ &\ge \check{v}_{T-1}(y_0, a_0, F_0) \ge \widetilde{v}_{T-1}(y_0, a_0). \end{aligned}$$

This shows that $\widetilde{v}_{T-1}(y,a)$ is l.s.c.. Next, take set $\mathcal{O} = E_Y \times (0,1)$ and such set is open. The y section of $\mathcal{O}$ is the interval $(0,1)$. By [BS78]

$$\widehat{V}_{T-1}(y) = \sup_{a\in(0,1)} \widetilde{v}_{T-1}(y,a)$$

is l.s.c.. Note that $A = [0,1]$ is the closure of $(0,1)$, thus

$$V_{T-1}(y) = \sup_{a\in A} \widetilde{v}_{T-1}(y,a) = \sup_{a\in(0,1)} \widetilde{v}_{T-1}(y,a) = \widehat{V}_{T-1}(y),$$

and $V_{T-1}(y)$ is l.s.c.. Following the backward induction for $t = T-2,\dots,0$, the proof is complete. □

Proposition 3.2 is of great importance for numerical computation of Bellman equations (3.2). As in [KENA19], when V_{t+1} is l.s.c., for any fixed $(y,a) \in E_Y \times A$, the inner optimization problem can be solved as follows

$$\begin{aligned}\inf_{F\in\mathcal{C}_t^\alpha(y)} \int_{E_Y} V_{t+1}(y_{t+1}) Q_t(dy_{t+1}|y,a,F) &= \inf_{F\in\mathcal{C}_t^\alpha(y)} \int_{\mathbb{R}} V_{t+1}(\mathbf{G}(t,y,a,z)) dF(z) \\ &= \sup_{\gamma\geq 0} \left\{ \mathbb{E}_{\widehat{F}} \left[V_{t+1}^\gamma(\mathbf{G}(t,y,a,Z_{t+1})) \right] - \frac{\gamma Q_t^H(1-\alpha)}{\sqrt{t_0+t}} \right\},\end{aligned}$$

where $V_{t+1}^\gamma(\mathbf{G}(t,y,a,Z_{t+1})) = \inf_{z\in\mathbb{R}} \{V_{t+1}(\mathbf{G}(t,y,a,z)) + \gamma|z - Z_{t+1}|\}$, and $y = (x, \widehat{F})$. With such results in hand, the Bellman equation (3.2) becomes

$$V_T(y) = \frac{1-e^{-\eta x}}{\eta},$$

$$V_t(y) = \sup_{a\in A,\gamma\geq 0} \left\{ \mathbb{E}_{\widehat{F}}[V_{t+1}^\gamma(\mathbf{G}(t,y,a,Z_{t+1}))] - \frac{\gamma Q_t^H(1-\alpha)}{\sqrt{t_0+t}} \right\}, \tag{3.3}$$

$$V_{t+1}^\gamma(\mathbf{G}(t,y,a,Z_{t+1})) = \inf_{z\in\mathbb{R}} \{V_{t+1}(\mathbf{G}(t,y,a,z)) + \gamma|z - Z_{t+1}|\}. \tag{3.4}$$

In the sequel, we will dicuss the challenges in the numerical computation of (3.3) and (3.4) and explain our algorithm for dealing with such problem.

3.1 Algorithm

In this practice, we will mainly follow the idea represented in [CL21] and propose a similar numerical scheme that uses regression Monte Carlo and GP surrogates to solve the Bellman equations (3.3) and (3.4). Then, we analyze the performance of the obtained optimal control on out-of-sample paths by simulating the realized terminal utility and estimating the expected utility.

Towards this end, we begin with discretizing the state space by choosing $y_t^i = (x_t^i, \widehat{F}_t^i) \in E_Y$, $i = 1,\dots,N$, $t \in \mathcal{T}$. These y_t^i's are called design points. Then, we solve the equation (3.3) for the design points $y = y_t^i$, $i = 1,\dots,N$, $t = T, T-1\dots,0$. One of the main tasks in the numerical algorithm is computing $\mathbb{E}_{\widehat{F}}[V^\gamma(\mathbf{G}(t,y_t^i,a,Z_{t+1})]$ for $i = 1,\dots,N$, $t \in \mathcal{T}$. In view of $\widehat{F}_t^i$ being an empirical distribution and assuming that

$$\widehat{F}_t^i(z) = \frac{1}{t_0+t} \sum_{j=-t_0+1}^{t} \mathbb{1}_{z_j^i \leq z},$$

we have

$$\mathbb{E}_{\widehat{F}_t^i}[V_{t+1}^\gamma(\mathbf{G}(t, y_t^i, a, Z_{t+1}))] = \frac{1}{t+t_0} \sum_{j=-t_0+1}^{t} V_{t+1}^\gamma(\mathbf{G}(t, y_t^i, a, z_j^i)). \tag{3.5}$$

Remark 3.3. In our current setup, $\widehat{F}_0$ is defined to be an empirical distribution constructed from historical data prior to the beginning of the investment, but it does not have to be. For example, there are estimation techniques that produce continuous prior distribution $\widehat{F}_0$ (cf. perturbed empirical distribution), and in such case Monte Carlo method will be needed to compute $\mathbb{E}_{\widehat{F}}[V_{t+1}^\gamma(\mathbf{G}(t, y, a, Z_{t+1}))]$ due to the fact that $\widehat{F}$ is no longer a discrete distrubtion anymore.

Since the value function V_{t+1}, and in turn V_{t+1}^γ, cannot be computed analytically, we will need a regression model for V_{t+1} so that we can estimate the right hand side of (3.5). The general strategy is then, for every $t \in \mathcal{T}'$, we use $(y_{t+1}^i, V_{t+1}(y_{t+1}^i))$, $i = 1, \ldots, N$, called training points to build a regression model for V_{t+1}, and use it to evaluate V_{t+1}^γ. Thus, we have an optimize–train–optimize loop in our algorithm. The state component $\widehat{F}_t^i$ is a probability distribution which is infinitely dimensional, or can be equivalently replaced by the vector $z_{-t_0+1:t}^i$. In both cases, we are dealing with a high dimensional problem and facing the challenge of "curse of dimensionality". Due to such reason, the traditional grid-based method for choosing the design points y_t^i, $i = 1, \ldots, N$, $t \in \mathcal{T}$, will be inefficient. To overcome this difficulty, we use the idea of randomized control so that we can focus on the points in the state space that are likely to be visited by the state process Y. In particular, for $t \in \mathcal{T}'$, given the design points $y_t^i, \ldots, y_t^N$, we will uniformly generate $a^1, \ldots, a^N$ from A and use them to update $y_t^1, \ldots, y_t^N$ to $y_{t+1}^1, \ldots, y_{t+1}^N$, respectively, according to

$$y_{t+1}^i = \mathbf{G}(t, y_t^i, a^i, Z_{t+1}^i), \quad i = 1, \ldots, N,$$

where Z_{t+1}^i is the simulated random noise.

Next, we discuss the choise of regression model for the value function V_{t+1} in detail. From above we see that for each $t \in \mathcal{T}'$, V_{t+1} can be viewed as a function of $(x_{t+1}, z_{-t_0+1:t+1})$ where $z_{-t_0+1:t+1}$ yields the empirical distribution $\widehat{F}_{t+1}$. Therefore, it is natural to regress V_{t+1} against $(x_{t+1}, z_{-t_0+1:t+1})$ instead of $(x_{t+1}, \widehat{F}_{t+1})$. Such treatment will reduce an infinite dimensional problem to a finite one. However, note that $(x_{t+1}, z_{-t_0+1:t+1})$ has a dimension of $t_0 + t + 2$ and to regress V_{t+1} against such high dimensional input requires an enormous amount of training points $(x_{t+1}^i, z_{-t_0+1:t+1}^i)$, $i = 1, \ldots, N$, so that we can obtain an accurate regression model for V_{t+1}. Hence, solely for the regression purpose, we will approximate $\widehat{F}_{t+1}$ with its first d moments denote by $m_{t+1}^1, \ldots, m_{t+1}^d$, and regress V_{t+1} against $(x_{t+1}, m_{t+1}^1, \ldots, m_{t+1}^d)$. By doing so, we effectively approximate a $t_0 + t + 2$-dimensional function with a $d+1$-dimensional regression model. Since the moments of a distrubtion capture the features of the distribution quite well, our strategy is a sound way to reduce the dimension of the problem that we are facing. To this end, we propose to use the GP surrogate to build regression models for V_{t+1}, $t \in \mathcal{T}'$. Gaussian process is a popular tool in machine learning that is suitable for dealing with regression problem with mid-range dimensions. It produces nonparametric functional approximations of functions by utilizing the location information of the function input. Namely, for some "usual" function g, if $\|u_1 - u_2\|$ is small, then a GP user assumes that $\|g(u_1) - g(u_2)\|$ should be relatively small as well. Recall that from Theorem 2.5 and Proposition 3.2, we immediately get the following result.

Corollary 3.4. *For every $t \in \mathcal{T}$, and V_t defined in (3.2), V_t is a continuous function on E_Y.*

Hence, GP is the ideal tool for us to build the statistical surrogates for each V_t, $t \in \mathcal{T}''$, so that we can proceed with the backward iteration and solve the Bellman equations. To be more specific,

we approximate each of the design points y_{t+1}^i, $i = 1, \ldots, N$, by $\check{y}_{t+1}^i := (x_{t+1}^i, m_{t+1}^{i,1}, \ldots, m_{t+1}^{i,d})$, and denote by $\check{V}_{t+1}$ the GP surrogate of V_{t+1}. Then, in the context of GP regression, the values $\check{V}_{t+1}(\check{y}_{t+1}^i)$, $i = 1, \ldots, N$, are jointly normal distributed. For any $y \in E_Y$, the predicted value $\check{V}_{t+1}(y)$ that approximates $V_{t+1}(y)$ is then computed as

$$\check{V}_{t+1}(y) = (k(y, \check{y}_{t+1}^1), \ldots, k(y, \check{y}_{t+1}^N))[\mathbf{K} + \epsilon^2 \mathbf{I}]^{-1}(V_{t+1}(\check{y}_{t+1}^1), \ldots, V_{t+1}(\check{y}_{t+1}^N))^\top,$$

where $\mathbf{I}$ is the $N \times N$ identity matrix and entries of $\mathbf{K}$ has the form $\mathbf{K}_{ij} = k(\check{y}_{t+1}^i, \check{y}_{t+1}^j)$, $i, j = 1, \ldots, N$. The function $k(\cdot, \cdot)$ is called the kernel function of the GP surrogate and in this project, we choose it from the Matern-5/2 family (cf. [Gen02]). We fit $\check{V}_{t+1}$ to the training points $\{(\check{y}_{t+1}^i, V_{t+1}(\check{y}_{t+1}^i)), i = 1, \ldots, N\}$ and during this process the hyperparameters inside of $k(\cdot, \cdot)$ will be estimated. For a comprehensive discussion of the Gaussian process surrogates, we refer to the book [RW06].

We summarize our algorithm for solving (3.3) and (3.4) as follows:

1. (Assume that $V_{t+1}(\cdot)$ and $\varphi_{t+1}^*(\cdot)$ are computed (estimated) at design points $y_{t+1}^1, \ldots, y_{t+1}^N$, $t \in \mathcal{T}''$, and the GP surrogates $\check{V}_{t+1}$ and $\check{\varphi}_{t+1}^*$ [1] are fitted.)

2. For time t, any $a \in A$, $\gamma > 0$, $z \in \mathbb{R}$, and each of the design points $\{y_t^i, i = 1, \ldots, N\} \subset E_Y$, use the GP surrogate $\check{V}_{t+1}$ and command `scipy.optimize.minimize_scalar` in the scipy package for `Python` to compute

$$\check{V}_{t+1}^\gamma(\mathbf{G}(t, y_t^i, a, z)) := \inf_{z' \in \mathbb{R}} \{\check{V}_{t+1}(\mathbf{G}(t, y_t^i, a, z)) + \gamma |z' - z|)\}, \quad i = 1, \ldots, N,$$

   and $\check{V}_{t+1}^\gamma$ is an approximation of V_{t+1}^γ.

3. For time t, any $a \in A$, and each of the design points $\{y_t^i, i = 1, \ldots, N\} \subset E_Y$, approximate $\mathbb{E}_{\widehat{F}}[V_{t+1}^\gamma(\mathbf{G}(t, y_t^i, a, Z_{t+1}))]$ as

$$\mathbb{E}_{\widehat{F}}[V_{t+1}^\gamma(\mathbf{G}(t, y_t^i, a, Z_{t+1}))] \approx \frac{1}{t + t_0} \sum_{j=-t_0+1}^{t} \check{V}_{t+1}^\gamma(\mathbf{G}(t, y_t^i, a, z_j^i)).$$

4. Use the command `scipy.optimize.minimize_scalar` to compute

$$V^{(1)}(y_t^i, a) = -\inf_{\gamma \geq 0} \left\{ -\frac{1}{t + t_0} \sum_{j=-t_0+1}^{t} \check{V}_{t+1}^\gamma(\mathbf{G}(t, y_t^i, a, z_j^i)) + \frac{\gamma Q_t^H(1 - \alpha)}{\sqrt{t_0 + t}} \right\}, \quad i = 1, \ldots, N,$$

   and

$$V_t(y_t^i) = -\inf_{a \in A}(-V^{(1)}(y_t^i, a)),$$

   where we also obtain the optimizer $\varphi_t^*(y_t^i)$, $i = 1, \ldots, N$.

5. Fit the GP surrogate $\check{V}_t$ by using $(\check{y}_t^i, V_t(y_t^i))$, $i = 1, \ldots, N$, as the training points. Similarly, fit $\check{\varphi}_t^*$ by using $(\check{y}_t^i, \varphi_t(y_t^i))$, $i = 1, \ldots, N$.

6. Goto 1.: start the next recursion for $t - 1$.

[1] The GP surrogate $\check{\varphi}_{t+1}^*$ is the Gaussian process regression model constructed by using the training data $\{(\check{y}^i, \varphi_{t+1}^*(y^i)), i = 1, \ldots, N\}$.

To analyze the performance of the optimal control we obtain from solving the Bellman equations, we generate N' forward simulated paths by starting with the initial state $y_0 = (x_0, \widehat{F}_0)$ and applying the control $\check{\varphi}_t^*(\check{y}_t^i)$, $i = 1, \ldots, N'$, to obtain the next-step state y_{t+1}^i for $t \in \mathcal{T}'$ according to

$$y_{t+1}^i = \mathbf{G}(t, y_t^i, \check{\varphi}_t^*(\check{y}_t^i), Z_{t+1}^i).$$

The corresponding forward Monte Carlo algorithm is summarized as

1. Take $y_0^i \equiv (x_0, \widehat{F}_0)$, $i = 1, \ldots, N'$.
2. For $t = 1, \ldots, T$, generate Z_t^i, $i = 1, \ldots, N'$.
3. Approximate y_t^i as $\check{y}_t^i$ and use the GP surrogates to compute the control $a_t^i = \check{\varphi}_t(\check{y}_t^i)$, $i = 1, \ldots, N'$, $t \in \mathcal{T}'$.
4. Update the states $y_{t+1}^i = \mathbf{G}(t, y_t^i, a_t^i, Z_{t+1}^i)$, $i = 1, \ldots, N'$, $t = 0, \ldots, T-1$.
5. Compute $\widehat{V}_0(y_0) = \frac{1}{N'} \sum_{i=1}^{N'} \frac{1 - e^{-\eta x_T^i}}{\eta}$.

The average $\widehat{V}_0(y_0)$ is then the Monte Carlo estimator of the expected utility. In addition, we are interested in the distribution of the utility

$$\widehat{U}(y_0) = \left(\frac{1 - e^{-\eta x_T^1}}{\eta}, \ldots, \frac{1 - e^{-\eta x_T^{N'}}}{\eta} \right)$$

and the numerical results will be reported in the sequel.

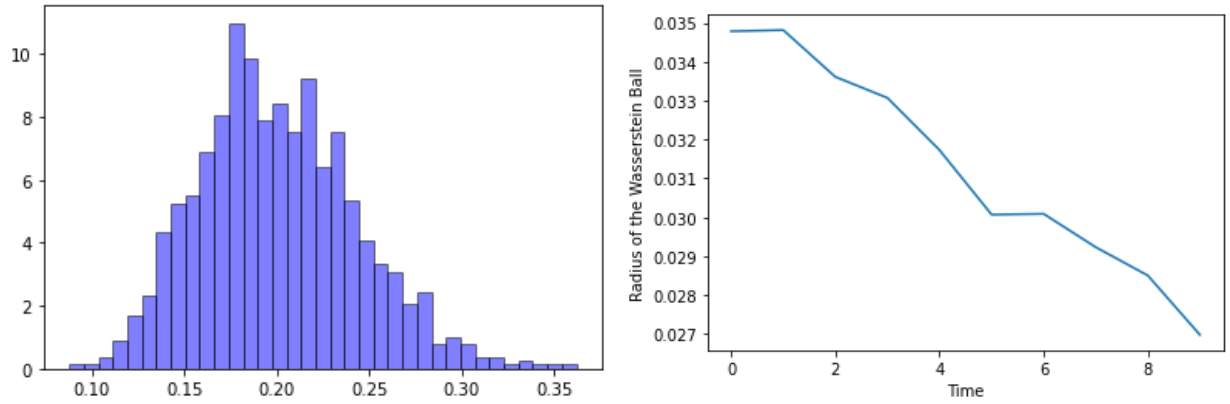

Figure 1: *Left panel: Histogram of simulated* $Q_0^H(1-\alpha)$ *for* $t_0 = 20$ *and* $\alpha = 0.1$. *Right panel: simulated path of the radius of* $\mathcal{C}_t^\alpha$.

3.2 Numerical Results

In this section, we apply the machine learning algorithm described above to some specific sets of parameters. We will compare the performance of nonparametric adaptive robust method to that of some other frameworks used to deal with model uncertainty. Theoretically, the optimal control is attained when there is no model uncertainty. We will also analyze the difference in performance between the cases of knowing the true model and having to estimate and learning the dynamics

of the underlying stochastic process. To this end, we consider three types of investors: the one who knows the true model with terminal utility $\widehat{U}^{\mathrm{TR}}(y_0)$ and expected utility $\widehat{V}_0^{\mathrm{TR}}(y_0)$; the one that applies the nonparametric adaptive robust with terminal utility $\widehat{U}^{\mathrm{AR}}(y_0)$ and expected utility $\widehat{V}_0^{\mathrm{AR}}(y_0)$; finally, the one uses the static robust methods, meaning the corresponding uncertainty sets do not change with respect to the state and time. In particular, the static robust investor utilizes the nonparametric setup and builds the uncertainty set as a Wasserstein ball around the empirical distribution generated by historical data with sample size t_0. The terminal utility and expected terminal utility of the nonparametic static robust investor are $\widehat{U}^{\mathrm{SR}}(y_0)$ and $\widehat{V}_0^{\mathrm{SR}}(y_0)$, respectively.

	AR	TR	SR
$\widehat{V}_0$	65.425570	66.805075	63.947066
$\mathrm{var}(\widehat{U})$	36.679199	108.601415	$6.451175 \cdot 10^{-9}$
$q_{0.20}(\widehat{U})$	59.682528	58.740896	63.947003
$q_{0.90}(\widehat{U})$	72.937869	78.899913	63.947173
$\max(\widehat{U})$	82.953448	90.773115	63.947302
$\min(\widehat{U})$	46.192910	26.307049	63.946811

Table 1: *Mean, variance, 20%-quantile, 90%-quantile, maximum, and minimum of the out-of-sample terminal utility for the AR, TR and SR methods;* $Q_0^H(1-\alpha) = 0.199165$.

Note that we can easily modify the above algorithm to compute $\widehat{U}^{\mathrm{TR}}$, $\widehat{V}_0^{\mathrm{TR}}$, $\widehat{U}^{\mathrm{SR}}$, and $\widehat{V}_0^{\mathrm{SR}}$. In fact, by taking $\mathcal{C}_t^\alpha(y) \equiv \mathcal{B}_0(F^*)$ which is the Wasserstein ball around F^* with 0 radius, we are able to compute $\widehat{U}^{\mathrm{TR}}$ and $\widehat{V}_0^{\mathrm{TR}}$. For $\widehat{U}^{\mathrm{SR}}$ and $\widehat{V}_0^{\mathrm{SR}}$, we take $\mathcal{C}_t^\alpha(y) \equiv \mathcal{C}_0^\alpha(y_0)$.

We choose the terminal time to be 1 year with $T = 10$ time steps which means one unit of time is 0.1 year. The annual insterest rate is 0.02 so that $r = 0.02/10 = 0.002$. Initial endowment is $x_0 = 100$. Some other parameters are $\alpha = 0.1$, $\eta = 0.01$, and $m = 4$. The number of paths is $N = 1000$ for nonparametric adaptive robust and 200 for other methods. The reason for such choice is that the state space of adaptive robust has dimension $m + 1$ while the others have dimension 1. For the sampling measure and test measure, we consider a Gaussian mixture model: with 40% probability, $Z_t \sim N(0.06/10, 0.4^2/10)$, and with 60% probability, $Z_t \sim N(0.16/10, 0.25^2/10)$. Recall that the parametric static robust investor assumes that $Z_t \sim N(\mu, \sigma^2)$ and constructs the confidence region for μ and σ^2. We will compute and compare the distributions of utlities among the mentioned four frameworks with the above choice of parameters for $t_0 = 20$. We also want to point out that the behavior of the optimal strategies would depend on the simulated $Q_0^H(1-\alpha)$. In this exercise, we present two cases with $Q_0^H(1-\alpha) \approx 0.199165$ and $Q_0^H(1-\alpha) \approx 0.092942$. Note that among 1000 simulated paths, 0.199165 sits very closely to the average value of $Q_0^H(1-\alpha)$ which is 0.200395, and 0.092942 is below the 1% quantile which is 0.115721. We refer to the left panel of Figure 1 for the histogram of simulated $Q_0^H(1-\alpha)$.

For $Q_0^H(1-\alpha) \approx 0.199165$, comparison among AR, TR, and SR are reported in Table 1. Since TR knows the true model of the risky asset return, the corresponding strategy will be optimal and $\widehat{V}_0^{TR}$ will outperform any other optimal control provided by investors who do not know the true model. Nevertheless, AR does better in three indices of risky management: AR has lower variance, higher 20% quantile, and minimum value of the simulated terminal utilities than TR. AR also beats SR quite significantly in regard to the mean, 90% quantile and maximum value of the

	AR	TR	SR
$\widehat{V}_0$	65.440839	66.805075	63.947067
$\mathrm{var}(\widehat{U})$	41.907675	108.601415	$6.776359 \cdot 10^{-9}$
$q_{0.20}(\widehat{U})$	59.575356	58.740896	63.946997
$q_{0.90}(\widehat{U})$	73.363772	78.899913	63.947175
$\max(\widehat{U})$	85.322523	90.773115	63.947367
$\min(\widehat{U})$	45.40310	26.307049	63.946763

Table 2: *Mean, variance, 20%-quantile, 90%-quantile, maximum, and minimum of the out-of-sample terminal utility for the AR, TR and SR methods;* $Q_0^H(1-\alpha) = 0.092942$.

simulated terminal utility. In addition, by viewing the Figure 2, we argue that AR produces wealth paths with more favorable distribution than TR. On the other hand, SR generates trivial optimal strategies similarly to the observations made in some earlier work (cf. [BCC+19], [CM20]). By ignoring the numerical instability, the terminal wealth produced by SR is a constant 102.018 which means all the money is invested in the banking account. With no surprises, as such a conservative control method, SR performs well in the department of risk management: it has apparent minimal variance, higher 20% quantile and minimum value of the terminal utility compared to AR and TR.

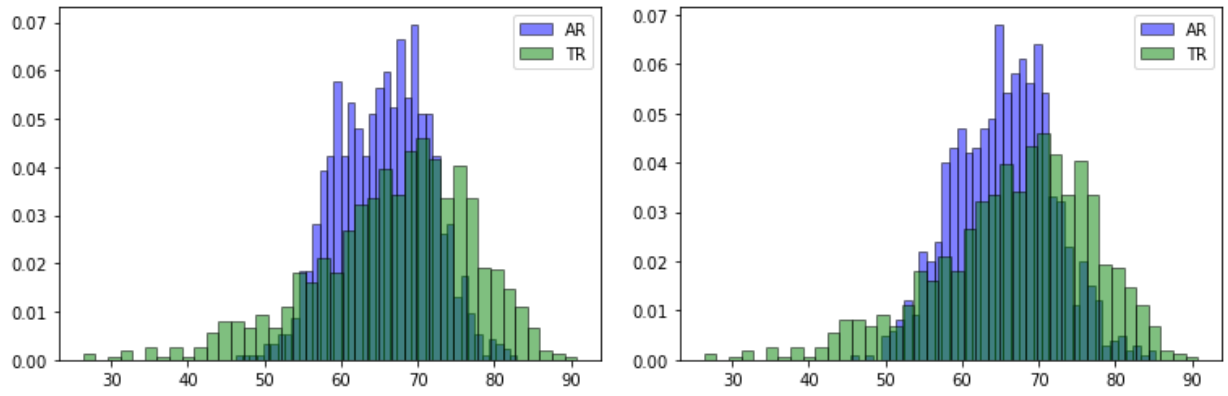

Figure 2: *Histogram of the out-of-sample terminal utility* U*: AR vs TR. Left panel:* $Q_0^H(1-\alpha) = 0.199165$*; right panel:* $Q_0^H(1-\alpha) = 0.092942$.

For $Q_0^H(1-\alpha) \approx 0.092942$, comparison of the performance of AR, TR, and SR on the same out-of-sample paths as in the previous case are reported in Table 2. Since that $Q_0^H(1-\alpha)$ is smaller, the size of $\mathcal{C}_t^\alpha$ along the simulated paths is in general smaller as a consequence. Hence, we expect more aggressive strategies given by the robust approaches. One needs to be aware that $Q_0^H(1-\alpha) \approx 0.092942$ has an extremely low probability. Thus, we expect the value of $Q_0^H(1-\alpha)$, and in turn the radius of $\mathcal{C}_t^\alpha$ to be oscillating after $t = 0$. Nevertheless, we see from Table 2 that there is an improvement of AR in this case. Estimated expected utility $\widehat{V}_0^{\mathrm{AR}}$ and the 90% quantile of $\widehat{U}^{\mathrm{AR}}$ are marginally larger than in the case of $Q_0^H(1-\alpha) = 0.199165$. Increase in the maximum value of $\widehat{U}^{\mathrm{AR}}$ on the other hand is somewhat significant. An unavoidable trade-off is that, even though only slightly, the strategy becomes more risky as the variance increases and 20% quantile, as well as the minimum value, of $\widehat{U}^{\mathrm{AR}}$ both decrease. In line with our discussion, we also observe

in Figure 2 that the distribution of $\widehat{U}^{\mathrm{AR}}$ in the right panel has moderately larger tails on both left and right sides compared to that in the left panel. Such change is expected to be more significant if the computation is done for larger t_0 and T. To conclude, AR is more aggressive when the size of $\mathcal{C}_t^\alpha$ is smaller but it is in general stable for our choice of parameters in the computation. Regarding SR, we observe changes following a similar pattern as for AR. However, such changes are so tiny and almost negligible. Consequently, the computed SR strategies are considered as trivial and one needs to further reduce $Q_0^H(1-\alpha)$ in order to obtain a non-trivial SR optimal control.

The main argument for why SR being so conservative is that for relatively small historical data size t_0, the corresponding confidence region is usually too large. On top of that, there is no shrinkage of the confidence region in static robust. Hence, no matter at which time step, the worst case model in such a large set is strongly against the controller which implies that, in the context of optimal portfolio, the money should only be invested in the banking account. Dynamic reduction of uncertainty is thereby an apparent advantage maintained by AR over SR. In practice, static robust control should only be used when there is sufficient historical data. One still needs be cautious of potential estimation error as, for uncertainty set with small size, the SR optimal control will heavily depend on the initial guess of the unknown distribution. Due to the lack of dynamic learning, SR optimal control in such case will be biased if the initial guess has large distance to the true model. On the contrary, learning is incorporated in adaptive robust and thus the corresponding control will be almost optimal for time steps close to T, and this feature will be carried out to earlier time steps following the dynamic programming principle.

References

[BB95] T. Başar and P. Bernhard. H^∞-optimal control and related minimax design problems. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, second edition, 1995. A dynamic game approach.

[BC21] T. Bhudisaksang and A. Cartea. Adaptive robust control in continuous-time. SIAM Journal on Control and Optimization, 59(5):3912–3945, 2021.

[BCC17] T. Bielecki, T. Chen, and I. Cialenco. Recursive construction of confidence regions. Electron. J. Statist., 11(2):4674–4700, 2017.

[BCC+19] T. Bielecki, T. Chen, I. Cialenco, A. Cousin, and Jeanblanc M. Adaptive robust control under model uncertainty. SIAM J. Control Optim., 57(2), 2019.

[BCC21] T. Bielecki, T. Chen, and I. Cialenco. Risk-sensitive markov decision problems under model uncertainty: finite time horizon case. arXiv, 2021.

[BCP16] E. Bayraktar, A. Cosso, and H. Pham. Robust feedback switching control: Dynamic programming and viscosity solutions. SIAM Journal on Control and Optimization, 54(5):2594–2628, 2016.

[BDOW21] D. Bartl, S. Drapeau, J. Obloj, and J. Wiesel. Sensitivity analysis of wasserstein distributionally robust optimization problems. Proc. R. Soc. A, 477: 20210176, 2021.

[Bor85] K. Border. Fixed Point Theorems with Applications to Economics and Game Theory. Cambridge University Press, 9 edition, 1985.

[BS78] D. P. Bertsekas and S. Shreve. Stochastic Optimal Control: The Discrete-Time Case. Academic Press, 1978.

[CG91] H. F. Chen and L. Guo. Identification and stochastic adaptive control. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., 1991.

[CL21] T. Chen and M. Ludkovski. A machine learning approach to adaptive robust utility maximization and hedging. SIAM Journal on Financial Mathematics, 3(12):1226–1256, 2021.

[CM20] T. Chen and J. Myung. Nonparametric adaptive bayesian stochastic control under model uncertainty. Preprint, 2020.

[DBGM99] E. Del Barrio, E. Giné, and C. Matrán. Central limit theorems for the wasserstein distance between the empirical andthe true distributions. The Annals of Probability, 27(2):1009–1071, 1999.

[FG15] N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. Probability Theory and Related Fields, 162:707–738, 2015.

[Gen02] M. G. Genton. Classes of kernels for machine learning: a statistics perspective. The Journal of Machine Learning Research, 2:299–312, 2002.

[GS89] I. Gilboa and D. Schmeidler. Maxmin expected utility with nonunique prior. J. Math. Econom., 18(2):141–153, 1989.

[HS08] P. L. Hansen and T. J. Sargent. Robustness. Princeton University Press, 2008.

[HSTW06] L. P. Hansen, G. Sargent, G. Turmuhambetova, and N. Williams. Robust control and model misspecification. J. Econom. Theory, 128(1):45–90, 2006.

[KENA19] D. Kuhn, P. Esfahani, V. Nguyen, and S. Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. INFORMS Tutorials in Operations Research, pages 130–166, 2019.

[KV15] P. R. Kumar and P. Varaiya. Stochastic Systems: Estimation, Identification, and Adaptive Control. Prentice Hall, Inc., 2015.

[Nut16] M. Nutz. Utility maximization under model uncertainty in discrete time. Mathematical Finance, 26(2):252–268, 2016.

[OW21] J. Oblój and J. Wiesel. Distributionally robust portfolio maximisation and marginal utility pricing in one period financial markets. Mathematical Finance Special Issue in Memory of Professor Mark H. A. Davis, pages 1454–1493, 2021.

[Rie75] U. Rieder. Bayesian dynamic programming. Adv. Appl. Prob., 7:330–348, 1975.

[RW06] C. E. Rasmussen and C. K. I. Williams. Gaussian Processes for Machine Learning. The MIT Press, 2006.

[Sir14] M. Sirbu. A note on the strong formulation of stochastic control problems with model uncertainty. Electronic Communications in Probability, 19, 2014.