

Article

Minimax Estimation in Regression under Sample Conformity Constraints

Andrey Borisov ^{1,2,3,4} 

- ¹ Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44/2 Vavilova Str., 119333 Moscow, Russia; A.Borisov@frccsc.ru
- ² Moscow Aviation Institute, 4, Volokolamskoe Shosse, 125993 Moscow, Russia
- ³ Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, GSP-1, 1-52 Leninskiye Gory, 119991 Moscow, Russia
- ⁴ Moscow Center for Fundamental and Applied Mathematics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, 119991 Moscow, Russia

Abstract: The paper is devoted to the guaranteeing estimation of parameters in the uncertain stochastic nonlinear regression. The loss function is the conditional mean square of the estimation error given the available observations. The distribution of regression parameters is partially unknown, and the uncertainty is described by a subset of probability distributions with a known compact domain. The essential feature is the usage of some additional constraints describing the conformity of the uncertain distribution to the realized observation sample. The paper contains various examples of the conformity indices. The estimation task is formulated as the minimax optimization problem, which, in turn, is solved in terms of saddle points. The paper presents the characterization of both the optimal estimator and the set of least favorable distributions. The saddle points are found via the solution to a dual finite-dimensional optimization problem, which is simpler than the initial minimax problem. The paper proposes a numerical mesh procedure of the solution to the dual optimization problem. The interconnection between the least favorable distributions under the conformity constraint, and their Pareto efficiency in the sense of a vector criterion is also indicated. The influence of various conformity constraints on the estimation performance is demonstrated by the illustrative numerical examples.

Keywords: mathematical modeling; estimation; minimax techniques; pareto optimization; regression analysis; statistical uncertainty



Citation: Borisov, A. Minimax Estimation in Regression under Sample Conformity Constraints. *Mathematics* **2021**, *9*, 1080. <https://doi.org/10.3390/math9101080>

Academic Editors: Mikhail Posypkin, Andrey Gorshenin and Vladimir Titarev

Received: 6 April 2021
Accepted: 6 May 2021
Published: 11 May 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The problems of the heterogeneous parameter estimation in the regression under the model uncertainty are considered intensively from the various points of view. The guaranteeing (or minimax) approach gives one of the most prospective tools to solve these problems. For the proper formulation of an estimation problem in minimax terms one usually needs:

- A description of the uncertainty set in the observation model;
- A class of the admissible estimators;
- An optimality criterion (a loss function) as a function of the argument pair “estimator–uncertain parameter value”.

The problem is to find the estimator that minimizes the maximal losses over the whole uncertainty set.

In the related literature, the parametric uncertainty set is specified either by geometric [1–7], or by statistical [8–15] constraints. In the former case, the uncertain parameters are treated as non-random but unknown ones lying within the fixed uncertainty set. In the latter case, the parameters are supposed to be random with unknown distribution, and the uncertainty set is formed by all

the admissible distributions. In both cases, the guaranteeing estimation presumes a solution to a two-person game problem: the first player is “a statistician”, and the performer of the second, “external” player role is dictated by the problem statement—it might be nature, another human or device. Nevertheless, the guaranteeing approach suggests the unified prescription: finding the best estimator under the worst behavior of the uncertainty. In practice, such a universality leads to a loss of some prior information.

Let us explain this point by an example: the statistician knows that the source of the uncertainty is nature. This means he/she “should bear in mind that nature, as a player, is not aiming for a maximal win (that is, does not want us to suffer a maximal loss), and in this sense, it is ‘impartial’ in the choice of strategies” [12]. Hence, in this case, the minimax approach is too pessimistic and leads to cautious and coarse estimates. Even if we know the second player is a human, this does not imply his/her “bad will” towards the statistician. Hopefully, the second player has goal other than maximizing the loss of the statistician. If the goal of the second player is known, one can change the estimation criterion and transform the initial problem into a non-antagonistic game [16]. Otherwise, the statistician can identify the goal indirectly, relying on the available observations. Hence, in the latter case, it seems natural to introduce additional constraints to the uncertainty set, depending on the realized observations.

The paper aims to present a solution to the minimax estimation problem under additional constraints, which are determined by a conformity index of the uncertain parameters to the available observations.

The paper is organized as follows. Section 2 contains the formal problem statement with the conformity index based on the likelihood function. The section presents the assumptions concerning the observation model, which guarantee the correctness of the stated estimation problem and the existence of its solution. It also contains the comparison of the problem with the recent investigations.

Section 3 provides the main result: the initial estimation problem is reformulated as a game problem, which has a saddle point, defining the minimax estimator completely. Moreover, the point is a solution to a dual finite-dimensional constrained optimization problem, which is simpler than the initial minimax problem. The form of the minimax estimator and properties of the least favorable distributions (LFD) is also included in the section.

Section 4 is devoted to the analysis of the obtained results. First, a numerical algorithm for the dual optimization problem solution is presented along with its accuracy characterization. Second, some other conformity indices based on the empirical distribution function (EDF) and sample mean are also introduced. Third, a new concept of the uncertain distribution choice under a vector criterion is considered. The first criterion component, being the loss function introduced in Section 2, describes the influence of the uncertainty on the estimation quality. The second component is the conformity index, which characterizes the accordance of the unknown distribution of γ and the realized observations $Y = y$. We present an assertion that the LFD in the minimax estimation problem is Pareto-efficient in the sense of the introduced vector criterion.

Section 5 presents the numerical examples, which illustrate the influence of various conformity constraints on the estimation performance. Section 6 contains concluding remarks.

The following notations are used in this manuscript:

- $\mathcal{B}(S)$ is the Borel σ -algebra of the topological space S (is S is the whole space) or its contraction to the set S (if S is a set of the topological space);
- $\text{col}(A_1, \dots, A_n)$ is a column vector formed by the ordinary or block components A_1, \dots, A_n ;
- $\text{row}(A_1, \dots, A_n)$ is a row vector formed by the ordinary or block components A_1, \dots, A_n ;
- $\langle a, b \rangle$ is a scalar product of two finite-dimensional vectors;
- $\mathcal{C}(\mathcal{X})$ is a set of all continuous real-valued functions with the domain \mathbb{X} ;
- $\|x\|$ is the Euclidean norm of the vector x ;
- $P_F\{A\}$ is the probability of the event A corresponding to the distribution F ;

- $E_F\{X\}$ is a mathematical expectation of the random vector X with the distribution F ;
- $\text{conv}(\mathbb{S})$ is a convex hull of the set \mathbb{S} .

2. Statement of Problem

2.1. Formulation

Let us consider the following observation model:

$$Y = A(X, \gamma) + B(X, \gamma)V. \tag{1}$$

Here:

- $\gamma \in \mathcal{C} \in \mathcal{B}(\mathbb{R}^m)$ is an unobservable random vector, having an unknown cumulative distribution function (cdf) F ;
- $X \in \mathbb{R}^n$ is a random unobservable vector with a known cdf $\Psi(dx|\gamma)$ dependent on the value of γ ;
- $Y \in \mathbb{R}^k$ is a vector of observations;
- $V \in \mathbb{R}^k$ is a random vector of observation errors with the known probability density function (pdf) $\phi_V(v)$;
- $A(\cdot) : \mathcal{C} \times \mathbb{R}^n \rightarrow \mathbb{R}^k$ is a nonrandom function characterizing the observation plant;
- $B(\cdot) : \mathcal{C} \times \mathbb{R}^n \rightarrow \mathbb{R}^{k \times k}$ is a nonrandom function characterizing the observation error intensity.

The observation model is defined on the family of the probability triplets $\{(\Omega, \mathcal{F}, P_F)\}_{F \in \mathbb{F}}$, where:

- The outcome space $\Omega \triangleq \mathcal{C} \times \mathbb{R}^m \times \mathbb{R}^k$ contains all admissible values of the compound vector $\text{col}(\gamma, X, V)$;
- σ -algebra is determined as $\mathcal{F} \triangleq \mathcal{B}(\mathcal{C} \times \mathbb{R}^m \times \mathbb{R}^k)$;
- The probability measures P_F are determined as:

$$P_F\{\gamma \in dq, X \in dx, V \in dv\} \triangleq \Psi(dx|q)F(dq)\phi_V(v)dv. \tag{2}$$

Using the generalized Bayes rule [17], it is easy to verify that the function:

$$\mathcal{L}(y|q) \triangleq \int_{\mathbb{R}^n} |\det(B(q, x))|^{-1} \phi_V(B^{-1}(q, x)(y - A(q, x))) \Psi(dx|q) \tag{3}$$

is the conditional pdf of the observation Y given γ : $P_F\{Y \in dy|\gamma = q\} = \mathcal{L}(y|q)dy$. Furthermore, the function:

$$\mathfrak{L}(y, F) \triangleq \int_{\mathcal{C}} \mathcal{L}(y|q)F(dq) \tag{4}$$

defines the pdf of the observation Y under the assumption that the distribution law of γ equals F :

$$\mathfrak{L}(y, F) = \frac{P_F\{Y \in dy\}}{dy} = \int_{\mathcal{C}} \mathcal{L}(y|q)F(dq). \tag{5}$$

Below in the paper we refer to the function $\mathfrak{L}(y, F)$ as *the sample conformity index based on the likelihood function*.

Our aim is to estimate the function $h(\gamma, X) : \mathcal{C} \times \mathbb{R}^m \rightarrow \mathbb{R}^l$ of (γ, X) , and the admissible estimators are the functions $\bar{h}(Y) : \mathbb{R}^k \rightarrow \mathbb{R}^l$ of the available observations.

The loss function is a conditional mean square of estimate error given the available observations:

$$J(\bar{h}, F|y) \triangleq E_F\{\|h(\gamma, X) - \bar{h}(Y)\|^2|Y = y\}, \tag{6}$$

and the corresponding estimation criterion:

$$J^*(\bar{h}|y) \triangleq \sup_{F \in \mathbb{F}_L} J(\bar{h}, F|y) \tag{7}$$

characterizes the maximal loss for a fixed estimator \bar{h} within the class \mathbb{F}_L of the uncertain distributions of γ , for which $\mathcal{L}(y, F) \geq L$.

The minimax estimation problem for the vector h is to find an estimator $\hat{h}(\cdot)$, such that:

$$\hat{h}(y) \in \underset{\bar{h} \in \mathbb{H}}{\text{Argmin}} J^*(\bar{h}|y), \tag{8}$$

where \mathbb{H} is a class of admissible estimators.

2.2. Necessary Assumptions Concerning Observation Model

To state the minimax estimation problem (8) properly and guarantee the existence of its solution we have to make additional assumptions concerning the uncertainty of γ , the observation model (1) and the estimated vector h :

- (i) The set \mathcal{C} is compact.
- (ii) Let \mathbb{F} be a family of all probability distributions with a support lying within the set \mathcal{C} . The set \mathbb{F}_L is itself a convex $*$ -weakly compact [18] subset of \mathbb{F} .
- (iii) The constraint

$$\mathcal{L}(y, F) \geq L \tag{9}$$

holds for all $F \in \mathbb{F}_L$. The inequality (9) is called *the conformity constraint of the level L based on the likelihood function* (or, shortly, likelihood constraint).

- (iv) The set \mathbb{F}_L is nonempty.
- (v) $A(\cdot, \cdot), B(\cdot, \cdot), h(\cdot, \cdot) \in \mathbf{C}(\mathcal{C} \times \mathbb{R}^n)$.
- (vi) pdf $\phi_V(v) > 0$ for $\forall v \in \mathbb{R}^k$; $\phi_V(v) \in \mathbf{C}(\mathbb{R}^k)$; the function $\Psi(dx|q)$ is a regular version of the conditional distribution for $\forall q \in \mathcal{C}$.
- (vii) The observation noise is uniformly non-degenerate, i.e.,

$$\min_{(q,x) \in \mathcal{C} \times \mathbb{R}^n} B(q, x)B^T(q, x) \geq \lambda_0 I > 0.$$

- (viii) The inequalities

$$\begin{aligned} & \int_{\mathbb{R}^k} \|v\|^2 \phi_V(v) dv < \infty, \\ \sup_{q \in \mathcal{C}} \int_{\mathbb{R}^n} \|A(q, x)\|^2 \Psi(dx|q) & \triangleq K_A < \infty, \\ \sup_{q \in \mathcal{C}} \int_{\mathbb{R}^n} \|h(q, x)\|^2 \Psi(dx|q) & \triangleq K_h < \infty \end{aligned}$$

are true.

- (ix) The set of admissible estimators \mathbb{H} contains only the functions $\bar{h}(\cdot) : \mathbb{R}^k \rightarrow \mathbb{R}^l$, for which:

$$\sup_{q \in \mathcal{C}} \int_{\mathbb{R}^k} \|\bar{h}(y)\|^2 \mathcal{L}(y|q) dy < \infty. \tag{10}$$

2.3. Argumentation

First, we discuss the sense of the assumptions in the subsection above.

Conditions (i)–(iv), describing the set \mathbb{F}_L , have the following interpretation.

The requirement for \mathcal{C} to be compact (i.e., fulfillment of condition (i)) is standard for the minimax estimation problems (see, e.g., [2,3]). In the case the prior information about the vector γ limited by the knowledge of its domain \mathcal{C} only, it is rather natural to treat γ as a random vector with an unknown distribution $F \in \mathbb{F}$. In practice we often have some additional prior information concerning the moment characteristics of γ , hence the entire uncertainty set \mathbb{F} can be significantly reduced. If, for example, $\mu(q) = \text{col}(\mu_1(q), \dots, \mu_N(q)) : \mathcal{C} \rightarrow \mathbb{R}^N$ is a vector of convex moment functions, and we know the vector $\bar{\mu} \triangleq \text{col}(\bar{\mu}_1, \dots, \bar{\mu}_N) \in \mathbb{R}^N$ of their upper bounds, then the set of ad-

missible distributions takes the form $\{F \in \mathbb{F} : \int_{\mathcal{C}} \mu_j(q)F(dq) \leq \bar{\mu}_j, j = \overline{1, N}\}$. The $*$ -weak compactness and convexity can be easily verified for this subset. Further in the presentation, we do not stress the explicit form of the “total” constraints other than (9) forming the subset \mathbb{F}_L : they should just guarantee the closure and convexity for \mathbb{F}_L . That is the sense of condition (ii).

The conditional pdf $\mathcal{L}(y|q)$ (3) can also be treated as the likelihood function of the parameter γ , calculated at the point q given the observed sample $Y = y$. This likelihood value reflects the relevance of the parameter value q to the realized observation y . By analogy, the function $\mathfrak{L}(y, F)$ can be considered as some generalization of the likelihood function that evaluates the correspondence between the uncertain distribution F and the realized observation y . The following lower and upper bounds for this value are obvious:

$$0 < \underline{\mathcal{L}}(y) \triangleq \min_{q \in \mathcal{C}} \mathcal{L}(y|q) \leq \mathfrak{L}(y, F) \leq \max_{q \in \mathcal{C}} \mathcal{L}(y|q) \triangleq \overline{\mathcal{L}}(y).$$

Below in the paper we suppose that the likelihood level L lies in $[\underline{\mathcal{L}}(y), \overline{\mathcal{L}}(y)]$. The subset formed by constraint $\{F \in \mathbb{F} : \int_{\mathcal{C}} \mathcal{L}(y|q)F(dq) \geq L\}$ is called *a distribution subset satisfying the likelihood conformity constraint of the level L* . It is nonempty because it contains at least all distributions with the support lying within the set $\{q \in \mathcal{C} : \mathcal{L}(y|q) \geq L\}$.

Adjusting the level L , we can vary the uncertainty set \mathbb{F}_L , choosing the distributions F , which are more or less relevant to the realized observations $Y = y$. That is an essence of condition (iii). Condition (iv) is obvious: all the constraints, defining the set \mathbb{F}_L , should be feasible.

Condition (v) is technical: it provides correctness of a subsequent change of measure. The condition is non-restricting because the broad class of the functions A, B and h can be approximated by the continuous functions. Conditions (vi) and (vii) guarantee correct utilization of the Fubini theorem and an abstract variant of the Bayes formula [19]. In practice these conditions are usually valid. Condition (viii) guarantees finite variance for both the observations and estimated vector independently of the distribution F .

Condition (ix) guarantees a finite variance of the estimate $\bar{h}(Y)$ independently of $F \in \mathbb{F}_L$.

The solution to (8) is obvious in the case of the one-point set $\mathbb{F}_L = \{F\}$. This means the distribution F of the parameter γ is known, and the initial problem is reduced to the traditional optimal in the mean square sense (MS-optimal) estimation problem. The case of the one-point set $\mathcal{C} = \{q\}$ is quite similar. In both cases the optimal estimator is completely defined by the conditional expectation (CE): $\hat{h}(y) = E_F\{h(\gamma, X)|Y = y\}$ in the case of a known distribution F , and $\hat{h}(y) = E_{\{q\}}\{h(q, X)|Y = y\}$ in the “one point” case.

In the general case of \mathbb{F}_L this result is inapplicable, because the CE $E_F\{h(\gamma, X)|Y = y\}$ is a functional of the unknown distribution F .

The stated estimation problem has a transparent interpretation. First, under prior uncertainty of the distribution F the replacement of the loss function (6) by guaranteeing analog looks natural. Second, utilization of the CE in the criterion means that the desired estimate should be calculated optimally for each observed sample. The criteria in the form of the CE appear often in estimation and control problems [11,17,20–22]. Mostly, the estimation is the preliminary stage in the solution to the optimization and/or control problem under incomplete information. The random disturbances/noises in such observation systems represent:

- A result of natural (non-human) impacts;
- A randomized (or generalized) control [23,24], used in the dynamic system;
- A result of some uncontrollable (parasitic) input signals of “the external players”.

The impact of the two latter types is not necessarily the nonrandom functions of available observations, but some “extra generated” random processes with distributions dependent on the observations. This type of control is used in the areas of telecommunications [25,26], cellular networks [27], technical systems [28], etc. The proposed minimax

criterion allows inhibiting the negative effect of the “additional randomness” in the external signals (the third type of disturbances mentioned above) to the estimation quality.

The additional comprehension of the natural gaps, which are inherent to the minimax estimation paradigm, and the ways of their partial coverage can be revealed by the following interpretation. It is well-known that in the case a minimax estimation problem can be reduced to a two-person game with a saddle point, the minimax estimator is the best one calculated for the LFD. The form of the LFD can be very strange and artificial. Moreover, the conformity degree of the LFD to the realized observations can be too low. Thus, the utilization of various sample conformity indices (particularly the ones based on the likelihood function) admits all to describe this degree, restrict it from below, implicitly reduce the distribution uncertainty set and exclude “exotic” variants of the LFDs.

Minimax estimation of the regression parameters is an investigation object in the various settings. Mostly, the observation model is a linear function of the estimated parameters corrupted by an additive Gaussian noise. The optimality criterion is a mathematical expectation of some loss function. In [29], the problem is solved by engaging the framework of the fuzzy sets. The authors of [30,31] used the criterion other than the traditional mean square one, and the estimated vector was random with the uncertain discrete distribution. In [32], the Gaussian noises have an uncertain but bounded covariance matrix. The papers [33–35] are also devoted to the minimax Bayesian estimation in the regression under various geometric and moment constraints of the estimated parameters. The criterion functions are ℓ_p norms of the estimation errors.

The optimality criterion in the form of CE and the admissibility of nonlinear estimates distinguish the proposed estimation problem from the recently considered ones [2,3,5–7,9]. A closely related problem considered in [11] has an essential difference. The uncertain parameter in [11] was treated as unknown and nonrandom, and hence the initial minimax problem could not be solved in terms of the saddle points. Moreover, the statistic uncertainty in [11] gave no possibility to take into account any additional prior and posterior information about the moment characteristics, conformity indices, etc. The paper [14] was devoted to the particular case of the likelihood constraints only. An idea to use confidence sets, calculated by the available statistical data, as the uncertainty sets of the distribution moments was used in [36] for the conditionally-minimax prediction.

3. The Main Result

As is known, the CE is determined in a non-unique way, hence we should specify a version of the CE so as to use it in further inferences. If the distribution F of the vector γ is known, then the CE of an integrable random value $h(\gamma, X) : \mathcal{C} \times \mathbb{R}^m \rightarrow \mathbb{R}$ can be calculated by the abstract variant of the Bayes formula:

$$\widehat{h}^F(y) = \frac{\int_{\mathcal{C} \times \mathbb{R}^n} h(q, x) |\det(B(q, x))|^{-1} \phi_V(B^{-1}(q, x)(y - A(q, x))) \Psi(dx|q) F(dq)}{\int_{\mathcal{C} \times \mathbb{R}^n} |\det(B(q', x'))|^{-1} \phi_V(B^{-1}(q', x')(y - A(q', x')))) \Psi(dx'|q') F(dq')}, \quad (11)$$

i.e., $E_F\{h(\gamma, X)|Y = y\} = \widehat{h}^F(y)$ (11) P_F – a.s. Below in the presentation we use the CE version, defined by (11). It should also be noted that if $\widehat{h}(\cdot)$ is the desired minimax estimator, then the inclusion (8) must be satisfied point-wise for any sample $y \in \mathbb{R}^k$.

Further in the paper the function:

$$J_*(F|y) \triangleq \min_{\widehat{h} \in \mathbb{H}} J(\widehat{h}, F|y) = J(\widehat{h}^F, F|Y) = \|\widehat{h}\|^2(y) - \|\widehat{h}^F(y)\|^2 \quad (12)$$

is called *the dual criterion* for J^* (7). All CEs in (12) are calculated by (11).

Using (3) for the calculation of \mathcal{L} , the notation:

$$v(q, x|y) \triangleq |\det B(q, x)|^{-1} \phi_V(B(q, x)^{-1}(y - A(q, x))), \quad (13)$$

and the CE version (11), the loss function (6) can be rewritten in the form:

$$J(\bar{h}, F|y) = \frac{\int_{\mathcal{C} \times \mathbb{R}^n} \|h(q, x) - \bar{h}(y)\|^2 \nu(q, x|y) \Psi(dx|y) F(dq)}{\int_{\mathcal{C}} \mathcal{L}(y|q') F(dq')} \tag{14}$$

As can be seen from (14), the function $J(\bar{h}, F|y)$ is neither convex nor concave in F , which complicates the solution to the estimation problem (8). Moreover, the argument F lies in the abstract infinite-dimensional space of the probability measures. Nevertheless, the problem can be reduced to a standard finite-dimensional minimax problem with a convex–concave criterion.

First, we introduce a new reference measure F' and verify that the loss function (14) represents a functional, which is linear in F' .

Let:

$$F'(F, dq|y) \triangleq \frac{\mathcal{L}(y|q) F(dq)}{\int_{\mathcal{C}} \mathcal{L}(q'|y) F(dq')} \tag{15}$$

Lemma 1. *If conditions (i)–(ix) are satisfied, then the following assertions are true.*

1. $F'(F, dq|y)$ is a probability measure for $\forall y \in \mathbb{R}^k$, and $F'(F, \cdot|y) \sim F(\cdot)$. The transformation (15) is a bijection of \mathbb{F} into itself, and its inversion F'' has the form:

$$F''(F', dq|y) \triangleq \frac{\mathcal{L}^{-1}(y|q) F'(dq)}{\int_{\mathcal{C}} \mathcal{L}^{-1}(q'|y) F'(dq')} \tag{16}$$

2. The set \mathbb{F}'_L of all distributions obtained from \mathbb{F}_L by the transformation (15):

$$\mathbb{F}'_L \triangleq \{\bar{F}(\cdot) : \exists F \in \mathbb{F}_L, \bar{F}(\cdot) = F'(F, \cdot|y)\} \tag{17}$$

is convex and $*$ -weakly closed.

The proof of Lemma 1 is given in Appendix A.

Applying the Fubini theorem and keeping in mind (11) and (15), we can rewrite the loss function (14) in the form:

$$\begin{aligned} J(\bar{h}, F|y) &= \frac{\int_{\mathcal{C}} \frac{\int_{\mathbb{R}^n} \|h(q, x) - \bar{h}(y)\|^2 \nu(q, x|y) \Psi(dx|y)}{\mathcal{L}(y|q)} \mathcal{L}(y|q) F(dq)}{\int_{\mathcal{C}} \mathcal{L}(q'|y) F(dq')} \\ &= \int_{\mathcal{C}} \frac{\int_{\mathbb{R}^n} \|h(q, x) - \bar{h}(y)\|^2 \nu(q, x|y) \Psi(dx|y)}{\mathcal{L}(y|q)} F'(F, dq|y) = J(\bar{h}, F'|y). \end{aligned} \tag{18}$$

To reduce the initial problem to some finite-dimensional equivalent, we also introduce the vectors:

$$\begin{aligned} w(y|q) &\triangleq \text{col}(w_1(y|q), w_2(y|q)) \in \mathbb{R}^{\ell+1} : \\ w_1(y|q) &\triangleq E_F \left\{ \|h(\gamma, X)\|^2 | Y = y, \gamma = q \right\} = \frac{\int_{\mathbb{R}^n} \|h(q, x)\|^2 \nu(q, x) \Psi(dx|y)}{\mathcal{L}(y|q)}, \\ w_2(y|q) &\triangleq E_F \{ h(\gamma, X) | Y = y, \gamma = q \} = \frac{\int_{\mathbb{R}^n} h(q, x) \nu(q, x) \Psi(dx|y)}{\mathcal{L}(y|q)}; \end{aligned} \tag{19}$$

$$\begin{aligned} w(F|y) &\triangleq \text{col}(w_1(F|y), w_2(F|y)) \in \mathbb{R}^{\ell+1} : \\ w_1(F|y) &\triangleq E_F \left\{ \|h(\gamma, X)\|^2 | Y = y \right\} = \int_{\mathcal{C}} w_1(y|q) F'(F, dq|y), \\ w_2(F|y) &\triangleq E_F \{ h(\gamma, X) | Y = y \} = \int_{\mathcal{C}} w_2(y|q) F'(F, dq|y), \end{aligned} \tag{20}$$

and their collections generated by the subsets \mathcal{C} and \mathbb{F}_L :

$$\begin{aligned} \mathbb{W}(\mathcal{C}|y) &\triangleq \{w(y|q) : q \in \mathcal{C}\}, \\ \mathbb{W}(\mathbb{F}_L|y) &\triangleq \{w(F|y) : F \in \mathbb{F}_L\}. \end{aligned} \tag{21}$$

Here and below the notation $\mathbb{H}(y)$ also stands for the whole set of the estimate values $\bar{h} \in \mathbb{H}$ calculated for the fixed argument y .

The set $\mathbb{W}(\mathbb{F}_L|y) \in \mathcal{B}(\mathbb{R}^{\ell+1})$ is compact; moreover (see [37]), the inclusion $\mathbb{W}(\mathbb{F}_L|y) \subseteq \text{conv}(\mathbb{W}(\mathcal{C}|y))$ holds.

On the set $\mathbb{R}^\ell \times \mathbb{R}^{\ell+1}$ we prepare the new loss function:

$$\mathbf{J}(\eta, w) \triangleq w_1 - 2\langle \eta, w_2 \rangle + \|\eta\|^2 = w_1 - \|w_2\|^2 + \|\eta - w_2\|^2. \tag{22}$$

It is easy to verify that the loss function (18) can be expressed via (22):

$$J(\bar{h}, F|y) = \int_{\mathcal{C}} \mathbf{J}(\bar{h}(y), w(y|q)) F'(F, dq|y) = \mathbf{J}(\bar{h}(y), w(F|y)).$$

The corresponding guaranteeing criterion takes the form:

$$\mathbf{J}^*(\eta|y) \triangleq \sup_{w \in \mathbb{W}(\mathbb{F}_L|y)} \mathbf{J}(\eta, w), \tag{23}$$

and its dual can be determined as:

$$\mathbf{J}_*(w) \triangleq \min_{\eta \in \mathbb{H}(y)} \mathbf{J}(\eta, w) = J(w_2, w) = w_1 - \|w_2\|^2. \tag{24}$$

The finite-dimensional minimax problem is to find:

$$\hat{\mathbf{h}}(y) \in \underset{\eta \in \mathbb{H}(y)}{\text{Argmin}} \mathbf{J}^*(\eta|y). \tag{25}$$

From the definitions of $\mathbb{W}(\mathbb{F}_L|y)$, $\mathbb{H}(y)$ and criterion (23) it follows that the problem (25) is equivalent to the initial minimax estimation problem (8):

$$\min_{\bar{h} \in \mathbb{H}} J^*(\bar{h}|y) = \min_{\eta \in \mathbb{H}(y)} \mathbf{J}^*(\eta|y) \triangleq \mathcal{J}(y), \tag{26}$$

$$\underset{\bar{h} \in \mathbb{H}}{\text{Argmin}} J^*(\bar{h}|y) \Big|_y \triangleq \{\hat{\mathbf{h}}(y) : J^*(\hat{\mathbf{h}}|y) = \mathcal{J}(y)\} = \underset{\eta \in \mathbb{H}(y)}{\text{Argmin}} \mathbf{J}^*(\eta|y) \tag{27}$$

for $\forall y \in \mathbb{R}^k$.

The following theorem characterizes the solution to the finite-dimensional minimax problem in terms of a saddle point of the loss function \mathbf{J} .

Theorem 1. For $\forall y \in \mathbb{R}^k$, the loss function $\mathbf{J}(\eta, w)$ (22) has the unique saddle point $(\hat{\mathbf{h}}(y), \hat{\mathbf{w}}(y))$ on the set $\mathbb{H}(y) \times \mathbb{W}(\mathbb{F}_L|y)$. The second block subvector $\hat{\mathbf{w}}(y) = \text{col}(\hat{\mathbf{w}}_1(y), \hat{\mathbf{w}}_2(y)) \in \mathbb{W}(\mathbb{F}_L|y)$ of the saddle point is the unique solution to the finite-dimensional dual problem:

$$\{\hat{\mathbf{w}}(y)\} = \underset{w \in \mathbb{W}(\mathbb{F}_L|y)}{\text{Argmax}} \mathbf{J}_*(w), \tag{28}$$

and $\hat{\mathbf{h}}(y) = \hat{\mathbf{w}}_2(y)$ is the second sub-vector of this optimum $\hat{\mathbf{w}}(y)$.

The proof of Theorem 1 is given in Appendix B.

By the definition of $\mathbb{W}(\mathbb{F}_L|y)$, for any vector $\widehat{w}(y)$ there exists at least one distribution \widehat{F} such that:

$$\widehat{w}_1(y) = E_{\widehat{F}}\{\|h(\gamma, X)\|^2|Y = y\}, \quad \widehat{w}_2(y) = E_{\widehat{F}}\{h(\gamma, X)|Y = y\}. \tag{29}$$

\widehat{F} is an LFD, and the whole set of the distributions satisfying (29) is denoted by $\widehat{\mathbb{F}}_L$.

Theorem 1 allows to obtain a solution to the initial minimax estimation problem. The result is formulated as:

Corollary 1. *The estimator $\widehat{w}_2(y)$ introduced in Theorem 1 is a solution to the minimax estimation problem (8), i.e., $\widehat{h}(y) = \widehat{w}_2(y)$ point-wise. The set $\{(\widehat{h}, \widehat{F})\}_{\widehat{F} \in \widehat{\mathbb{F}}_L}$ presents the saddle points of the loss function J (6) on the set $\mathbb{H} \times \mathbb{F}_L$. The estimator $\widehat{h}(y)$ is invariant to the LFD choice: if \widehat{F}' and \widehat{F}'' are different LFDs then $E_{\widehat{F}'}\{h(\gamma, X)|Y = y\} = E_{\widehat{F}''}\{h(\gamma, X)|Y = y\} = \widehat{w}_2(y)$.*

The following assertion characterizes the key property of the LFD set $\widehat{\mathbb{F}}_L$.

Corollary 2. *There exists a variant of the LFD $\widehat{F}_L \in \widehat{\mathbb{F}}$ concentrated at most at $\dim(\mathbb{W}(\mathbb{F}_L|y)) + 1$ points of the set \mathcal{C} .*

The proof of Corollary 2 is given in Appendix C.

The presence of the discrete version of LFD is a remarkable fact. Let us remind the reader that initially, we suppose that the uncertain vector γ lies in the set \mathcal{C} . The deterministic hypothesis concerning γ hopelessly obstructed the solution to the minimax estimation problem. To overcome this obstacle we provide the randomness of γ : the vector keeps constant during the individual observation experiment, and the stochastic nature of γ appears from experiment to experiment only. The existence of a discrete LFD returns us partially to the primordial situation. The point is that there exists a set of γ values that are the most difficult for estimation. Tuning to these parameters we can obtain estimates of γ with the guaranteed quality.

Theorem 1 and Corollary 1 simplify the solution to the initial problem (8), reducing it to the maximization of the finite-dimensional quadratic function (28) over the convex compact set.

4. Analysis and Extensions

4.1. Dual Problem: A Numerical Solution

To simplify presentation of the numerical algorithm of problem (28)'s solution, we suppose that the uncertainty set \mathbb{F}_L takes the form $\mathbb{F}_L = \{F \in \mathbb{F} : \mathcal{L} \geq L\}$, i.e., it is restricted by the conformity constraint only.

Let us consider the case $\mathcal{C} \triangleq \{q_j\}_{j=1, \overline{M}} \subset \mathbb{R}^m$, which corresponds to the practical problem of Bayesian classification [10,38]. Here the dual problem (28) has the form $\widehat{w}(y) = \underset{w \in \text{conv}(\mathbb{W}(\mathcal{C}|y))}{\text{Argmax}} \mathbf{J}_*(w)$. Its solution can be represented as $\widehat{w}(y) = \sum_{j=1}^M \widehat{P}_j(y)w(q_j|y)$,

where $\widehat{P}(y) \triangleq \text{row}(\widehat{P}_1(y), \dots, \widehat{P}_M(y))$ is a solution to the standard quadratic programming problem (QP problem):

$$\widehat{P}(y) \in \underset{\substack{p_1, \dots, p_M \geq 0: \\ \sum_{j=1}^M p_j = 1}}{\text{Argmin}} \left(\sum_{j=1}^M p_j w_1(q_j|y) - \sum_{j,j'=1}^M p_j p_{j'} \langle w_2(q_j|y), w_2(q_{j'}|y) \rangle \right). \tag{30}$$

Consequently, in the case of finite \mathcal{C} the minimax estimation problem can be reduced to the standard QP problem with well-investigated properties and advanced numerical procedures.

Utilization of the finite subsets $\mathcal{C}(\cdot)$ instead of the original domain \mathcal{C} allows to calculate the “mesh” approximations for the solution to (8).

Let:

- $\epsilon_n \downarrow 0$ be a decreasing nonrandom sequence characterizing the approximation accuracy;
- $\{\mathcal{C}(\epsilon_n)\}_{n \in \mathbb{N}}: \mathcal{C}(\epsilon_1) \subseteq \mathcal{C}(\epsilon_2) \subseteq \mathcal{C}(\epsilon_3) \subseteq \dots \subseteq \mathcal{C}$ be a sequence of embedded subdivisions;
-

$$\begin{aligned} \omega_1^\epsilon(y) &\triangleq \max_{\substack{q_1, q_2 \in \mathcal{C}: \\ \|q_1 - q_2\| < \epsilon}} |w_1(q_1|y) - w_1(q_2|y)|, \\ \omega_2^\epsilon(y) &\triangleq \max_{\substack{q_1, q_2 \in \mathcal{C}: \\ \|q_1 - q_2\| < \epsilon}} \|w_2(q_1|y) - w_2(q_2|y)\| \end{aligned} \tag{31}$$

be modulus of continuity for $w_1(y|q)$ and $w_2(y|q)$.

The assertion below characterizes the divergence rate of the approximating solutions to the initial minimax estimate.

Lemma 2. *If $\{\widehat{w}(n|y)\}_{n \in \mathbb{N}}$ are corresponding solutions to the problems:*

$$\widehat{w}^n(y) = \underset{w \in \text{conv}(\mathbb{W}(\mathcal{C}(\epsilon_n)|y))}{\text{Argmax}} \mathbf{J}_*(w)$$

then the following sequences are convergent as $\epsilon_n \downarrow 0$:

$$\begin{aligned} \widehat{w}^n(y) &\rightarrow \widehat{w}(y), \\ \mathcal{J}(y) - \max_{w \in \text{conv}(\mathbb{W}(\mathcal{C}(\epsilon_n)|y))} \mathbf{J}_*(w) &\leq \omega_1^{\epsilon_n}(y) + K[\omega_2^{\epsilon_n}(y)]^2 \downarrow 0 \end{aligned} \tag{32}$$

with some constant $0 < K < \infty$.

The proof of Lemma 2 is given in Appendix D.

4.2. The Least Favorable Distribution in the Light of the Pareto Efficiency

The minimax estimation problem under the conformity constraints is tightly interconnected with the choice of the distribution \widehat{F} that is optimal in the sense of a vector-valued criterion. On the one hand, the solution to the considered estimation problem is grounded on the evaluation of the distribution \widehat{F} , maximizing the dual criterion (12): $\mathbf{I}_1(F|y) \triangleq J_*(F|y) \rightarrow \max_F$. On the other hand, the distribution F should conform to the realized sample $Y = y$, and the maximization of the conformity index leads to the following optimization problem: $\mathbf{I}_2(F|y) \triangleq \mathcal{L}(y, F) \rightarrow \max_F$.

Obviously, the criteria \mathbf{I}_1 and \mathbf{I}_2 are conflicting; hence the proper choice of F requires the application of the vector optimization techniques.

Let:

- $\widehat{\mathbb{F}}_0$ be a set of the LFDs in the estimation problem (8) without conformity constraints (i.e., as $L = 0$);
- $\widetilde{\mathcal{L}}(y) \triangleq \max_{F \in \widehat{\mathbb{F}}_0} \mathcal{L}(y, F)$;
- $M \in [\widetilde{\mathcal{L}}(y), \overline{\mathcal{L}}(y)]$ be an arbitrary fixed conformity level;
- $\widehat{w}(y) = \underset{w \in \mathbb{W}(\mathbb{F}_M|y)}{\text{Argmax}} \mathbf{J}_*(w)$ be a solution to the finite-dimensional dual problem;
- $\widehat{\mathbb{F}}_M$ be the set of corresponding LFDs.

Lemma 3. *Any least favorable distribution $\widehat{F}_M \in \widehat{\mathbb{F}}_M$ is Pareto-efficient with respect to the vector-valued criterion $(\mathbf{I}_1, \mathbf{I}_2)$.*

The proof of Lemma 3 follows directly from the Germeyer theorem [16].

Consideration of the constrained minimax estimation problem in light of the optimization by the vector criterion is somehow close to the one investigated in [31], where the

estimation quality is characterized by the ℓ_2 norm of the error, and the Shannon entropy is characterized as a measure of the statistical uncertainty of the estimated vector.

4.3. Other Conformity Indices

First, we consider the conformity constraint (9) thoroughly. It admits the following treatment. Let $\tilde{F} \in \mathbb{F}$ be some reference distribution. The constraint $\mathfrak{L}(y, F) \geq \mathfrak{L}(y, \tilde{F})$ is a specific case of (9); the feasible distributions F should be relevant to the available observations $Y = y$ no less than the reference distribution \tilde{F} is. One more treatment is also acceptable. Let $\tilde{q} \in \mathcal{C}$ be some “guess” value of the uncertain parameter γ , and $\alpha > 0$ be a fixed value. The constraint:

$$\frac{\mathfrak{L}(y, F)}{\mathfrak{L}(y|\tilde{q})} \geq \alpha \tag{33}$$

is a specific case of (9): it means that the likelihood ratio of any feasible distribution F to the one-point distribution at \tilde{q} should be no less than the level α . Obviously, the guess value \tilde{q} could be chosen from the maxima of the function \mathcal{L} , i.e., $\tilde{q} \in \text{Argmax}_{q \in \mathcal{C}} \mathcal{L}(y|q)$, but calculation of these maxima is itself a nontrivial problem of likelihood function maximization. In Section 5 we use some modification of (33):

$$\frac{\mathfrak{L}(y, F) - \min_{q \in \mathcal{C}_n} \mathcal{L}(y|q)}{\max_{q \in \mathcal{C}_n} \mathcal{L}(y|q) - \min_{q \in \mathcal{C}_n} \mathcal{L}(y|q)} \geq r \tag{34}$$

where $\mathcal{C}_n \subseteq \mathcal{C}$ is a known subset, and $r \in (0, 1)$ is a fixed parameter. This form is important, because in the case of $\mathcal{C} = \mathcal{C}_n$ it guarantees for the constraint (34) to be active in the considered minimax optimization problem for each $r \in (0, 1)$.

Furthermore, the proposed conformity index $\mathfrak{L}(y, F)$ (9) is a non-unique numerical characteristic that describes the interconnection between F and Y . For example, an alternative conformity index can be defined as $\int_{\mathcal{C}} f(\mathcal{L}(y|q))F(dq)$, where $f(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is some continuous nondecreasing function. Another way to introduce this index is to set it as $\int_{\mathcal{S}(y)} \mathfrak{L}(y', F)dy' = P_F\{Y \in \mathcal{S}(y)\}$, i.e., as a probability that the observation Y lies in the confidence set $\mathcal{S}(y) \in \mathcal{B}(\mathbb{R}^k)$.

For a particular case of the observation model (1) we can propose one more conformity index that is based on the EDF. Let us consider the observation model with the “pure uncertain” estimated parameter γ :

$$Y_t = A(\gamma) + B(\gamma)V_t, \quad t = \overline{1, T}. \tag{35}$$

Here:

- $\mathbf{Y}_T \triangleq \text{col}(Y_1, \dots, Y_T)$ are available observations;
- $\gamma \in \mathcal{C} \in \mathbb{R}^m$ is a random vector with unknown distribution F ;
- $\mathbf{V}_T \triangleq \text{col}(V_1, \dots, V_T)$ are the observation errors that are i.i.d. centered normalized random values with the pdf $\phi_V(v)$.

If the value γ is known, the observations $\{Y_t\}_{t=\overline{1, T}}$ can be considered as i.i.d. random values, whose pdf is equal to $\phi_V(v)$ after some shifting and scaling. The EDF of the sample $\{Y_t\}_{t=\overline{1, T}}$ has the form:

$$F_T^*(y) \triangleq \frac{1}{T} \sum_{t=1}^T \mathbf{I}(y - Y_t). \tag{36}$$

On the other hand, the cdf $F^Y(y)$ of any observation Y_t for a fixed distribution F can be calculated as:

$$F^Y(y) \triangleq \int_{-\infty}^y \int_{\mathcal{C}} \phi_V\left(\frac{u - A(q)}{B(q)}\right) F(dq) du. \tag{37}$$

The sample conformity index based on the EDF is the following value:

$$\mathfrak{M}(\mathbf{Y}_T, F) \triangleq \|F_T^* - F^Y\|_\infty = \sup_{y \in \mathbb{R}} |F_T^*(y) - F^Y(y)|. \tag{38}$$

The new uncertainty set \mathbb{F}_M describing all admissible distributions F satisfies conditions (i), (ii) and (iv) above, but condition (iii) is replaced by the following one:

(x) the constraint

$$\mathfrak{M}(\mathbf{Y}_T, F) \leq M \tag{39}$$

This holds for all $F \in \mathbb{F}_M$ and some fixed level $M > 0$. It is called *the constraint based on the EDF*.

The proposed conformity index represents the well known Kolmogorov distance used in the goodness-of-fit test. One also knows the asymptotic characterization of $\mathfrak{M}(\mathbf{Y}_T, F)$:

$$\lim_{T \rightarrow \infty} P \left\{ \mathfrak{M}(\mathbf{Y}_T, F) < \frac{x}{\sqrt{T}} \right\} = \sum_{-\infty}^{+\infty} (-1)^j e^{-2j^2 x^2}.$$

Furthermore, the value $\mathfrak{M}(\mathbf{Y}_T, F)$ can be easily calculated, because the function F_T^* is piece-wise constant while F^Y is continuous:

$$\mathfrak{M}(\mathbf{Y}_T, F) = \max_{1 \leq t \leq T} \max(|F_T^*(Y_t) - F^Y(Y_{t-})|, |F_T^*(Y_t) - F^Y(Y_t)|),$$

and the cdf F^Y is calculated by (37).

The distribution set determined by (39) takes the form:

$$\left\{ F \in \mathbb{F} : -M + F_T^*(Y_t) \leq \int_{-\infty}^{Y_t} \int_{\mathcal{C}} \phi_V \left(\frac{u - A(q)}{B(q)} \right) F(dq) du \leq M + F_T^*(Y_{t-}), \quad t = \overline{1, T} \right\}. \tag{40}$$

Using the variational series $\mathbf{Y}_{(T)} \triangleq \text{col}(Y_{(1)}, \dots, Y_{(T)})$ of the sample \mathbf{Y}_T , and recalling $F_T^*(Y_{(t)}) = \frac{t}{T}, F_T^*(Y_{(t)-}) = \frac{t-1}{T}$, (40) can be rewritten in the form:

$$\left\{ F \in \mathbb{F} : -M + \frac{t}{T} \leq \int_{-\infty}^{Y_{(t)}} \int_{\mathcal{C}} \phi_V \left(\frac{u - A(q)}{B(q)} \right) F(dq) du \leq M + \frac{t-1}{T}, \quad t = \overline{1, T} \right\}. \tag{41}$$

It can be seen that this set is a convex closed polyhedron, lying in \mathbb{F} , with at most $2T$ facets. All assertions formulated in Section 3 are valid after replacing the uncertainty set \mathbb{F}_L , generated by the likelihood function, by the set \mathbb{F}_M , generated by the EDF. Moreover, the mesh algorithm for the dual optimization problem solution, presented above in Section 4.1, can also be applied to this case.

Let us consider the observation model (35) again. We can use the sample mean $\bar{Y} \triangleq \frac{1}{T} Y_t$ as one more conformity index. Let us remind the reader that due to the model property, the random parameter $\gamma(\omega)$ is constant for each sample \mathbf{Y}_T . For rather large T values, the central limit theorem allows to treat the normalized value $\frac{\sqrt{T}(\bar{Y} - A(\gamma))}{|B(\gamma)|}$ as a standard Gaussian random one. We then fix a standard Gaussian quantile c_α of the confidence level α and excise the subset:

$$\mathcal{C}_\alpha \triangleq \left\{ q \in \mathcal{C} : \bar{Y} - \frac{c_\alpha |B(\gamma)|}{\sqrt{T}} \leq A(q) \leq \bar{Y} + \frac{c_\alpha |B(\gamma)|}{\sqrt{T}} \right\} \subseteq \mathcal{C}.$$

If \mathcal{C}_α is compact then the set \mathbb{F}_α of all probability distributions with the domain lying in \mathcal{C}_α is called *the set of admissible distributions satisfying the sample mean conformity constraint of the level α* .

The comparison of the minimax estimates, calculated under various types of the conformity constraints, is presented in the next section.

5. Numerical Examples

5.1. Parameter Estimation in the Kalman Observation System

Let us consider the linear Gaussian discrete-time (Kalman) observation system:

$$\begin{cases} X_t = aX_{t-1} + bV_t, & t = \overline{1, T}, \quad x_0 \sim \mathcal{N}(0, P_0), \\ Y_t = cX_t + fW_t, & t = \overline{0, T}, \end{cases} \tag{42}$$

where:

- $\mathbf{X}_T \triangleq \text{col}(X_0, \dots, X_T)$ is an unobservable state trajectory (the autoregression X_t is supposed to be stable);
- $\mathbf{Y}_T \triangleq \text{col}(Y_0, \dots, Y_T)$ are available observations;
- $\mathbf{V}_T \triangleq \text{col}(V_1, \dots, V_T)$ and $\mathbf{W}_T \triangleq \text{col}(W_0, \dots, W_T)$ are vectorizations of independent standard Gaussian discrete-time white noises;
- P_0, c and f are known parameters;
- $\gamma \triangleq \text{col}(a, b)$ is an uncertain vector lying in the fixed rectangle $\mathcal{C} \triangleq [a, \bar{a}] \times [b, \bar{b}]$.

Our goal is to calculate the proposed minimax estimates of the uncertain vector γ and analyze their performance depending on the specific form of the loss function (6). To vary the loss function we can either specify the estimated test signal $h(\cdot)$ or determine different Euclidean weighted norms. We choose the second approach and define the following norm $\|\cdot\|_{\xi_X, \xi_\gamma}$ for the compound vector: $Z \triangleq \text{col}(\mathbf{X}_T, \gamma)$:

$$\|Z\|_{\xi_X, \xi_\gamma} \triangleq \sqrt{\xi_X^2 \sum_{t=1}^T X_t^2 + \xi_\gamma^2 (a^2 + b^2)},$$

and the corresponding loss function takes the form:

$$J_{\xi_X, \xi_\gamma}(\bar{Z}, F|\mathbf{Y}_T) \triangleq E_F \left\{ \|Z - \bar{Z}(\mathbf{Y}_T)\|_{\xi_X, \xi_\gamma}^2 | \mathbf{Y}_T \right\}. \tag{43}$$

In the case $\xi_\gamma = 1$ and $\xi_X = 0$ we obtain “the traditional” case of the mean-square loss conditional function $J_{0,1}(\bar{Z}, F|\mathbf{Y}_T) = E_F \{ \|\gamma - \bar{\gamma}(\mathbf{Y}_T)\|^2 | \mathbf{Y}_T \}$, and the estimation quality of $\bar{\gamma}(\cdot)$ is determined directly through the loss function. Using $\xi_\gamma = 0$ and $\xi_X = 1$ we transform the loss function into $J_{1,0}(\bar{Z}, F|\mathbf{Y}_T) = E_F \{ \|X - \bar{X}(\mathbf{Y}_T)\|^2 | \mathbf{Y}_T \}$, and the estimation of γ appears indirectly via the estimation of the state trajectory \mathbf{X}_T .

The minimax estimation is calculated by the numerical procedure introduced in Section 4.1 with the uniform mesh \mathcal{C}_{h_a, h_b} of the uncertainty set \mathcal{C} ; h_a and h_b are corresponding mesh steps along each coordinate.

We calculate the minimax estimate with the likelihood conformity constraint of the form:

$$\frac{\mathcal{L}(\mathbf{Y}_T, F) - \min_{(a,b) \in \mathcal{C}_{h_a, h_b}} \mathcal{L}(\mathbf{Y}_T | (a, b))}{\max_{(a,b) \in \mathcal{C}_{h_a, h_b}} \mathcal{L}(\mathbf{Y}_T | (a, b)) - \min_{(a,b) \in \mathcal{C}_{h_a, h_b}} \mathcal{L}(\mathbf{Y}_T | (a, b))} \geq r,$$

where $r \in (0, 1)$ is a confidence ratio.

We compare the proposed minimax estimate with some known alternatives.

The calculations have been executed with the following parameter values: $\mathcal{C} = [-0.1; 0.1] \times [0.1; 1]$, $a = -0.1$, $b = 0.1$, $P_0 = 0.5$, $c = 1$, $f = 0.5$, $T = 1000$, $h_a = 0.01$, $h_b = 0.045$. The choice of the parameters can be explained by the following facts. First, the point $(-0.1; 0.1)$ of actual parameter values belongs to the domain of the LFD for both loss functions $J_{0,1}$ and $J_{1,0}$. This means the appearance of just the LFD for both cases. Second, in spite of sufficient observation length, the signal-to-noise ratio is rather small, which prevents high performance of the asymptotic estimation methods.

Figure 1 presents the evolution of the minimax estimates $\hat{a}_{0,1}(r)$ and $\hat{a}_{1,0}(r)$ of a drift coefficient depending on the confidence ratio $r \in (0, 1)$. The minimax estimates are compared with;

- The estimate $\bar{a}^{MS}(\mathbf{Y}_T)$ calculated by the moment/substitution method [12]:

$$\bar{a}^{MS} = \sum_{t=1}^T y_{t-1}y_t / \left(\sum_{t=1}^T y_t^2 - T f^2 \right), \quad \bar{b}^{MS} = \sqrt{\frac{1}{c^2} \left(1 - (\bar{a}^{MS})^2 \right) \left(\sum_{t=1}^T y_t^2 - T f^2 \right)}$$

- The Bayesian estimate $\hat{a}^{F_1}(\mathbf{Y}_T)$ (11) calculated under the assumption that prior distribution F_1 of γ is uniform over the whole uncertainty set \mathcal{C} ;
- The Bayesian estimate $\hat{a}^{F_2}(\mathbf{Y}_T)$ (11) calculated under the assumption that the prior distribution F_2 of γ is uniform over the vertices of \mathcal{C} ;
- The estimate $\bar{a}^{EKF}(\mathbf{Y}_T)$ calculated by the extended Kalman filter (EKF) algorithm [39] and subsequent residual processing;
- The maximum likelihood estimate (MLE) $\bar{a}^{MLE}(\mathbf{Y}_T)$ calculated by the expectation/maximization algorithm (EM algorithm) [17].

Figure 2 contains a similar comparison of the diffusion coefficient estimates $\hat{b}_{0,1}(r)$ and $\hat{b}_{1,0}(r)$.

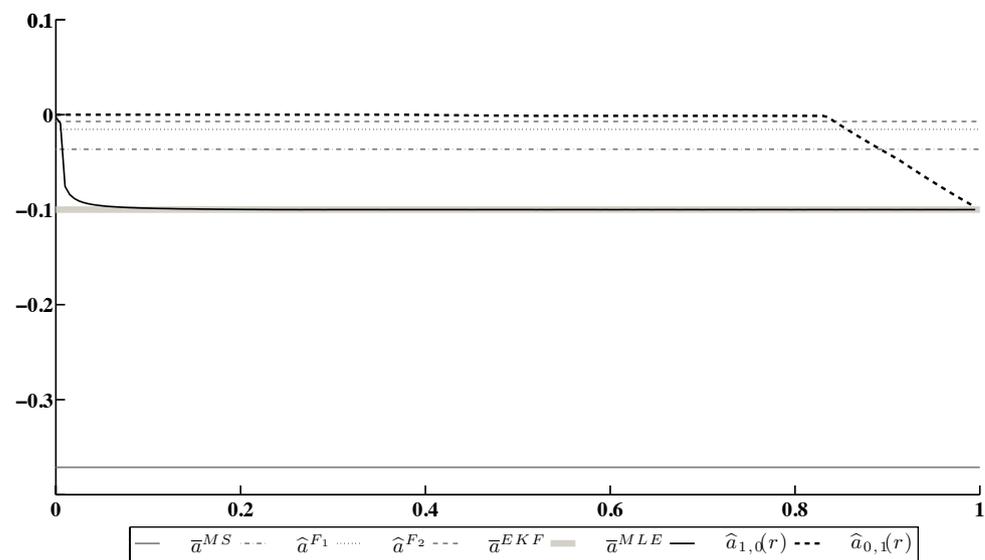


Figure 1. Estimation of the drift coefficient a .

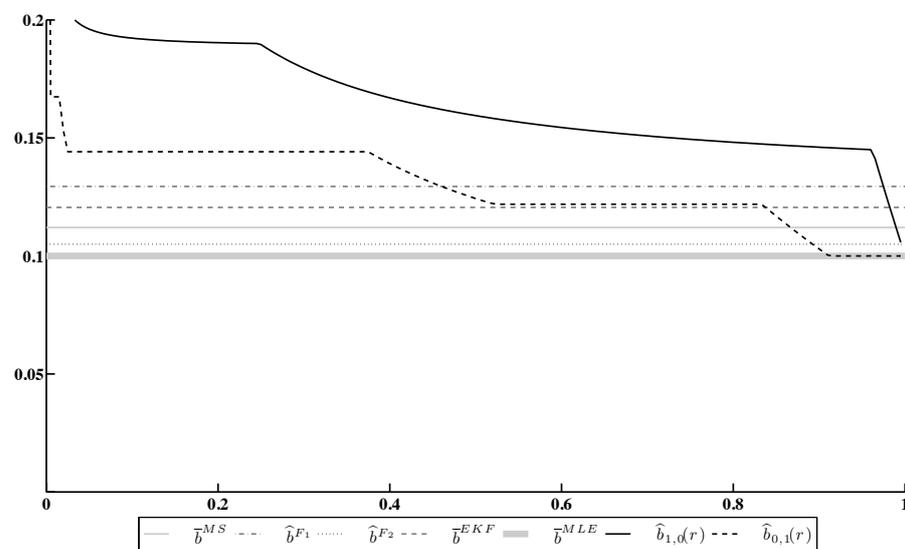


Figure 2. Estimation of the diffusion coefficient b .

The results of this experiment allow us to make the following conclusions.

1. Both minimax estimates $(\widehat{a}_{0,1}(r), \widehat{b}_{0,1}(r))$ and $(\widehat{a}_{1,0}(r), \widehat{b}_{1,0}(r))$ converge to the MLE $(\bar{a}^{MLE}, \bar{b}^{MLE})$ as $r \rightarrow 1$. Nevertheless, the rate of convergence depends on the specific choice of the loss function ($J_{0,1}$ or $J_{1,0}$ in the considered case).
2. Both minimax estimates are more conservative than the MLE, because they take into account a chance for other points of the LFD domain to be realized.
3. Under an appropriate choice of the confidence ratio r , both minimax estimates become more accurate than other candidates, except for the MLE.

5.2. Parameter Estimation under Additive and Multiplicative Observation Noises

We consider the observations $\mathbf{Y}_T \triangleq \text{col}(Y_1, \dots, Y_T)$:

$$Y_t = aX_t + V_t, \quad t = \overline{1, T}. \tag{44}$$

Here:

- a is an estimated value;
- $\mathbf{X}_T \triangleq \text{col}(X_1, \dots, X_T)$ is a vector of the i.i.d. unobservable random values (multiplicative noise): $X_1 \sim \mathcal{R}[0, 1]$;
- $\mathbf{V}_T \triangleq \text{col}(V_1, \dots, V_T)$ is a vector of the i.i.d. unobservable random values (additive noise): $V_1 \sim \mathcal{N}(0, \sigma)$.

We assume that the parameter a is random with unknown distribution, whose support set lies within the known set $\mathcal{C} \triangleq [c^1, c^2]$. The loss function has the form:

$$J(\bar{a}, F | \mathbf{Y}_T) = \mathbb{E}_F \left\{ \|a - \bar{a}(\mathbf{Y}_T)\|^2 | \mathbf{Y}_T \right\}.$$

In this example our goal is to compare the minimax estimates of the parameter a under conformity constraint based either on the likelihood function or on the EDF.

The minimax estimations are calculated for the following parameter values: $a = 2$, $T = 20$, $\mathcal{C} = [2, 3]$, $\sigma = 0.1$. We use the proposed numerical procedure under a uniform mesh \mathcal{C}_h of the set \mathcal{C} with the step $h = 0.005$. The example has some features. First, the observation model contains both the additive (\mathbf{V}_T) and multiplicative (\mathbf{X}_T) heterogeneous noises. Second, the available observed sample is not too long to provide the high quality for the consistent estimates. Third, the exact value of a is equal 2; meanwhile under the constraint absence there exists a discrete variant of the LFD with the finite support set $\{2, 3\}$. This means that the LFD is realized only in the considered observation model.

The likelihood conformity constraint looks similar to the one from the previous subsection:

$$\frac{\mathcal{L}(\mathbf{Y}_T, F) - \min_{q \in \mathcal{C}_h} \mathcal{L}(\mathbf{Y}_T | q)}{\max_{q \in \mathcal{C}_h} \mathcal{L}(\mathbf{Y}_T | q) - \min_{q \in \mathcal{C}_h} \mathcal{L}(\mathbf{Y}_T | q)} \geq r, \tag{45}$$

where $r \in (0, 1)$ is a confidence ratio.

Figure 3 contains comparison of the minimax estimate $\widehat{a}(r)$ with its actual value a , the (consistent asymptotically Gaussian) M-estimate $\bar{a}^{sub} \triangleq \frac{2}{T} \sum_{t=1}^T Y_t$, obtained by the moment/substitution method [12] and the MLE \bar{a}^{MLE} .

Next, we investigate minimax posterior estimates under the conformity constraint based on the EDF. The constraint is of the form:

$$\frac{\max_{F \in \mathbb{F}_{\mathcal{C}_h}} \mathfrak{M}(\mathbf{Y}_T, F) - \mathfrak{M}(\mathbf{Y}_T, F)}{\max_{F \in \mathbb{F}_{\mathcal{C}_h}} \mathfrak{M}(\mathbf{Y}_T, F) - \min_{F \in \mathbb{F}_{\mathcal{C}_h}} \mathfrak{M}(\mathbf{Y}_T, F)} \geq r, \tag{46}$$

where $r \in (0, 1)$ is some fixed confidence ratio, and $\mathbb{F}_{\mathcal{C}_h}$ is a “mesh” approximation of the set $\mathbb{F}_{\mathcal{C}}$ corresponding to the uniform “mesh” \mathcal{C}_h . The form (46) of the conformity constraint provides that it is active in the minimax optimization problem for any $r \in (0, 1)$.

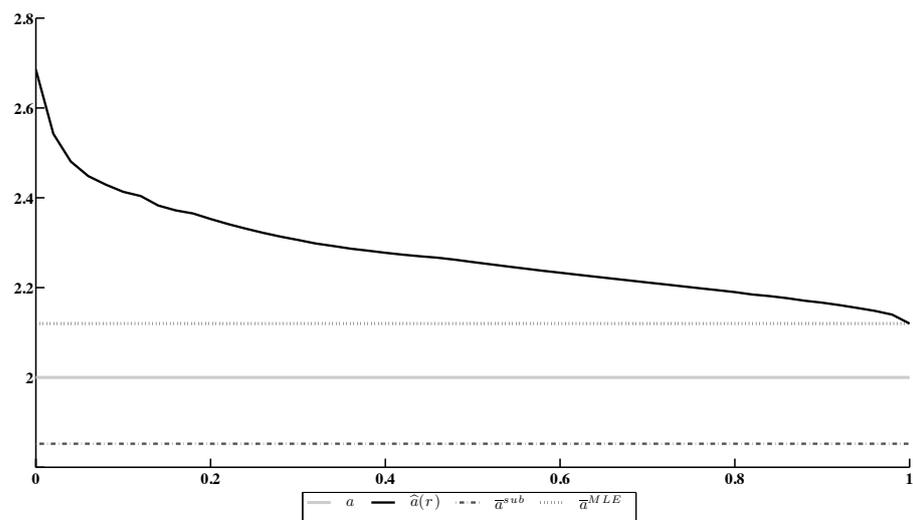


Figure 3. Estimation of the coefficient a under conformity constraint based on the likelihood function.

Figure 4 contains:

- The EDF $F_Y^*(y)$ calculated by the sample Y_T ;
- The cdf's $F_Y^q(y) = \int_{-\infty}^y \phi_V\left(\frac{u-A(q)}{B(q)}\right) du$ of Y , corresponding to the one-point distribution concentrated at the point q ($q = 2, 3$);
- The cdf $\bar{F}_Y(y) \bar{F}_Y(y) \in \text{Argmin}_{F \in \mathbb{F}_{C_h}} \mathfrak{M}(Y_T, F)$, closest to the EDF $F_Y^*(y)$ within the set \mathbb{F}_{C_h} .

Note that $F_Y^2(y)$ is a cdf of Y corresponding to the actual value of a .

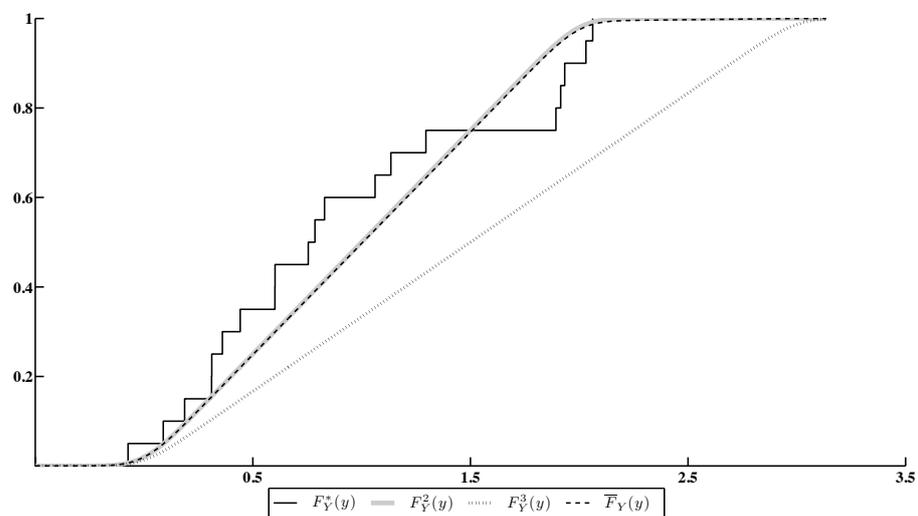


Figure 4. The EDF of Y and different cdf's of Y under various choices of a .

Figure 5 contains a comparison of the minimax estimate $\hat{a}(r)$ under the conformity constraint, based on the EDF, with its actual value a , the moment/substitution estimate \bar{a}^{sub} and the MLE \bar{a}^{MLE} .

The results of this experiment allow us to make the following conclusions.

1. The minimax estimate $\hat{a}(r)$ under the conformity constraint, based on the EDF, does not converge to the MLE \bar{a}^{MLE} as $r \rightarrow 1$.
2. Under an appropriate choice of the confidence ratio r , the minimax estimate under the EDF constraint becomes more accurate than other candidates, including the MLE.

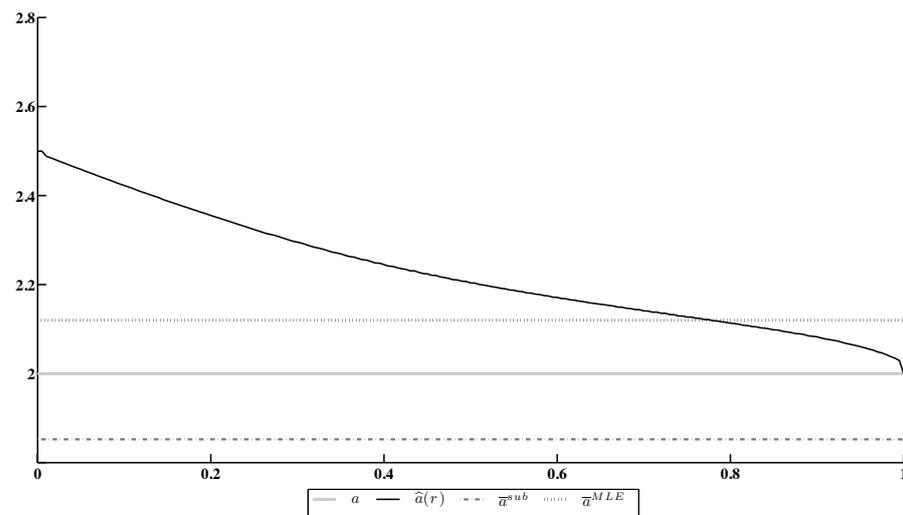


Figure 5. Estimation of the coefficient a under conformity constraint based on the EDF.

6. Conclusions

The paper contains the statement and solution to a new minimax estimation problem for the uncertain stochastic regression parameters. The optimality criterion is the conditional mean square of the estimation error given the realized observations. The class of admissible estimators contains all (linear and nonlinear) statistics with finite variance. The a priori information concerning the estimated vector is incomplete: the vector is random and the part of its components lies in the known compact. The key feature of the considered problem is the presence of the additional constraints for the statistical uncertainty, restricting from below the correspondence degree between the uncertainty and realized observations. The paper presents various indices, characterizing this conformity via the likelihood function, the EDF and the sample mean.

We propose a reduction of the initial optimization problem in the abstract infinite-dimensional spaces to the standard finite-dimensional QP problem with convex constraints along with an algorithm of its numerical realization and precision analysis.

The minimax estimation problem is solved in terms of the saddle points, i.e., besides the estimators with the guaranteed quality, we have a description of the LFDs. First, the investigation of the LFDs' domains allowed us the detection of the uncertain parameter values, which are the worst for the estimation. Second, the consideration of the performance index pair "conformity index–guaranteed estimation quality" uncovered rather a new conception of the parameter estimation under a vector optimality criterion. The paper contains an assertion, which states that the LFDs are Pareto-optimal for the vector-valued criterion above.

The paper focuses mostly on the conformity indices related to the likelihood function; thus, it is obvious that the performance of the minimax estimate is compared with the one of the MLE. In general, the MLE has several remarkable properties, in particular the asymptotic minimaxity under some additional restrictions [12]. However, the estimate is non-robust to the prior statistical uncertainty. The proposed minimax estimate can be considered as a robustified version of the MLE, which is ready for application in the cases of the short non-asymptotic samples or the violation of the conditions for the MLE asymptotic minimaxity.

The conformity constraints are not exhausted by the likelihood function. In the paper, we present other conformity indices based on the EDF and sample mean. We demonstrate that the minimax estimates with the EDF conformity constraint are better than the MLE. One of the points of the paper is that the flexible choice of the conformity indices and design of the additional conformity constraints for each individual applied estimation problem allows obtaining a tradeoff between the prior uncertainty and available observations.

The reason to choose one or another conformity index depends not only on the conditions of the specific practical estimation problem solved under the minimax settings. One of the essential conditions is the possibility of its quick computation for the subsequent verification of the conformity constraint. For example, calculation of the likelihood conformity constraint (33) with the guess value $\mathcal{L}(y|\bar{q}) = \max_q \mathcal{L}(y|q)$ tends to necessarily solve the auxiliary maximization problem for the likelihood function, which is nontrivial itself. Thus, the conformity indices based on the EDF or sample moments look more prospective from the computational point of view.

The applicability of the proposed minimax estimate also depends on the presence of the analytical formula of the estimates $w(y|q)$, or the fast numerical algorithms of its calculation. In turn, this possibility is a base for the subsequent effective solution to the QP problem and specification of the LFD.

Finally, the key indicator affecting the estimate calculation process and its precision is the number of the mesh nodes in the approximation $\mathcal{C}(\epsilon_n)$ of the uncertainty set \mathcal{C} . It is a function of “the size of \mathcal{C} / the mesh step ϵ_n ” ratio and dimensionality m of \mathcal{C} .

All of the factors above characterize the limits of possible applicability of the proposed minimax estimation method for the solution to one or another practical problem.

Funding: The research was supported by the Ministry of Science and Higher Education of the Russian Federation, project No. 075-15-2020-799.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

cdf	cumulative distribution function
CE	conditional expectation
EDF	empirical distribution function
EKF	extended Kalman filter
EM algorithm	expectation/maximization algorithm
LFD	least favorable distribution
MLE	maximum likelihood estimate
MS-optimal	optimal in the mean square sense
pdf	probability density function
QP problem	quadratic programming problem

Appendix A

Proof of Lemma 1. Conditions (v)—(viii) imply fulfillment of the inequalities:

$$\mathcal{L}(y|q) \leq \sup_{x \in \mathbb{R}^n} \nu(q, x|y) \leq \frac{1}{\lambda_0^{k/2}} \max_{x \in \mathbb{R}^n} \phi_V(x) \triangleq M < \infty.$$

Furthermore, for $\forall \epsilon$ ($0 < \epsilon < 1$) there exists a compact set $S(\epsilon) \in \mathcal{B}(\mathbb{R}^n)$, such that $\int_{S(\epsilon)} \Psi(dx|q) \geq 1 - \epsilon$, and by the Weierstrass theorem $m(y) \triangleq \min_{(q,x) \in \mathcal{C} \times S(\epsilon)} \nu(q, x|y) > 0$. Each measure $F \in \mathbb{F}$ can be associated with the measure $\mu_F(dq|y) \triangleq \mathcal{L}(y|q)F(dq)$. Obviously, $\mu_F \ll F$, and μ_F is finite, i.e., $0 < m(y) \leq \int_{\mathcal{C}} \mu_F(dq|y) \leq M < \infty$. Hence, $\forall y \in \mathbb{R}^k$ and $\forall F \in \mathbb{F}$. The measure $F'(F, dq|y)$ (15) is probabilistic; moreover $F' \ll F$. The measure $F''(F', dq|y)$ (16) is also a probabilistic one defined on $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, $F'' \ll F$, and the denominator in (16) has the following lower and upper bounds:

$$0 < \frac{1}{M} \leq \int_{\mathcal{C}} \mathcal{L}^{-1}(y|q)F(dq) \leq \frac{1}{m(y)} < \infty.$$

From (15) and (16) it follows that $F \sim F'$, and the corresponding measure transformations are mutually inverse, i.e., $\forall F \in \mathbb{F}$ the identity $F''(F'(F)) \equiv F'(F''(F)) \equiv F$ holds, and, moreover, $\{F'(F) : F \in \mathbb{F}\} = \{F''(F) : F \in \mathbb{F}\} = \mathbb{F}$. Assertion (1) of Lemma 1 is proven.

The set \mathbb{F}'_L is $*$ -weakly closed, because the set \mathbb{F}_L is, and the function $\mathcal{L}(y|q)$ is non-negative, continuous and bounded in $q \in \mathcal{C}$.

Let $F'_1, F'_2 \in \mathbb{F}'_L$ be two arbitrary distributions from \mathbb{F}'_L , and $F'_\alpha \triangleq \alpha F'_1 + (1 - \alpha)F'_2$ be its convex linear combination with a fixed parameter $\alpha \in [0, 1]$. We should prove that $F'_\alpha \in \mathbb{F}'_L$. By the definition of \mathbb{F}'_L there exist distributions $F_1, F_2 \in \mathbb{F}_L$ such that $F'_1 = F'(F_1)$ and $F'_2 = F'(F_2)$. Furthermore, for the convex combination $F_\beta = \beta F_1 + (1 - \beta)F_2$ with

$$\beta \triangleq \frac{\alpha \mathcal{L}(F_2|y)}{\alpha \mathcal{L}(F_2|y) + (1 - \alpha)\mathcal{L}(F_1|y)} \in [0, 1],$$

we can verify easily that $F'_\alpha = F'(F_\beta)$, i.e., $F'_\alpha \in \mathbb{F}'_L$. Assertion (2) of Lemma 1 is proven. \square

Appendix B

Proof of Theorem 1. The set $\mathbb{H}(y) = \mathbb{R}^\ell$ by condition (ix); thus it is convex and closed. The set \mathbb{F}'_L is convex and $*$ -weakly closed due to Lemma 1. From this fact and (20) it follows that $\mathbb{W}(\mathbb{F}_L|y)$ is also a convex closed set. Moreover, it is bounded due to condition (viii). The function \mathbf{J} (22) is strictly convex in η and concave (affine) in w . These conditions are sufficient for the existence of a saddle point [40]. It should be noted that both the set $\mathbb{H}(y) \times \mathbb{W}(\mathbb{F}_L|y)$ and the saddle point $(\hat{\mathbf{h}}(y), \hat{\mathbf{w}}(y))$ depend on the observed sample y . For the saddle point the following equalities are true:

$$\mathbf{J}(\hat{\mathbf{h}}(y), \hat{\mathbf{w}}(y)) = \min_{\eta \in \mathbb{H}(y)} \max_{w \in \mathbb{W}(\mathbb{F}_L|y)} \mathbf{J}(\eta, w) = \max_{w \in \mathbb{W}(\mathbb{F}_L|y)} \min_{\eta \in \mathbb{H}(y)} \mathbf{J}(\eta, w) = \max_{w \in \mathbb{W}(\mathbb{F}_L|y)} \mathbf{J}_*(w),$$

$$\text{i.e., } \hat{\mathbf{w}}(y) \in \underset{w \in \mathbb{W}(\mathbb{F}_L|y)}{\text{Argmax}} \mathbf{J}_*(w).$$

Now we prove the uniqueness of the saddle point $\hat{\mathbf{w}}(y)$. Let $w'(y) = \text{col}(w'_1(y), w'_2(y))$ and $w''(y) = \text{col}(w''_1(y), w''_2(y))$ be two different saddle points, and $\mathcal{J}(y) \triangleq \mathbf{J}_*(w'(y)) = \mathbf{J}_*(w''(y))$ and $w'''(y) \triangleq \alpha w'(y) + (1 - \alpha)w''(y)$ be arbitrary convex combinations of the chosen points ($0 < \alpha < 1$). After elementary algebraic transformations we have:

$$\mathbf{J}_*(w'''(y)) = \mathcal{J}(y) + \alpha(1 - \alpha)\|w'(y)_2 - w''(y)\|^2 > \mathcal{J}(y),$$

which contradicts our assumption that $w'(y)$ and $w''(y)$ are two different solutions to the finite-dimensional dual problem. Theorem 1 is proven. \square

Appendix C

Proof of Corollary 2. The set $\mathbb{W}(\mathbb{F}_L|y) \in \mathcal{B}(\mathbb{R}^{\ell+1})$ is compact, and $\mathbb{W}(\mathbb{F}_L|y) \subseteq \text{conv}(\mathbb{W}(\mathcal{C}|y))$. By the Krein–Milman theorem [37], each point of the set $\mathbb{W}(\mathbb{F}_L|y)$ can be represented as a convex combination at most of $\dim(\mathbb{W}(\mathbb{F}_L|y)) + 1$ extreme points of the set $\mathbb{W}(\mathbb{F}_L|y)$.

Obviously, all extreme points of $\mathbb{W}(\mathbb{F}_L|y)$ belong to the set $\mathbb{W}(\mathcal{C}|y)$. Hence, for the point $\hat{\mathbf{w}}(y)$ which is a solution to the finite-dimensional dual problem (28), there exists a finite set $\{\mathbf{q}_s(y)\}_{s=1, \overline{S}} \subseteq \mathcal{C}$, $1 \leq S \leq \dim(\mathbb{W}(\mathbb{F}_L|y)) + 1$ of parameters, and weights $\{P_s(y)\}_{s=1, \overline{S}}$ ($P_s(y) \geq 0, \sum_{s=1}^S P_s(y) = 1$) such that:

$$\hat{\mathbf{w}}(y) = \sum_{s=1}^S P_s(y)w(\mathbf{q}_s(y)|y). \tag{A47}$$

The parameters and weights define the reference measure (15) on the space $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$:

$$\hat{F}'(dq|y) \triangleq \sum_{s=1}^S P_s(y) \delta_{\mathbf{q}_s(y)}(dq).$$

We can establish the initial measure by (16):

$$\hat{F}(dq|y) = \frac{\sum_{s=1}^S \mathcal{L}^{-1}(\mathbf{q}_s(y)|y) P_s(y) \delta_{\mathbf{q}_s(y)}(dq)}{\sum_{s'=1}^S \mathcal{L}^{-1}(\mathbf{q}_{s'}(y)|y) P_{s'}(y)}.$$

It is easy to verify that $E_{\hat{F}}\{\|h(\gamma, X)\|^2 | Y = y\} = \mathbf{w}_1(y)$ and $E_{\hat{F}}\{h(\gamma, X) | Y = y\} = \mathbf{w}_2(y)$, i.e., \hat{F} is the required LFD. Corollary 2 is proven. \square

Appendix D

Proof of Lemma 2. Without loss of generality we suppose each ϵ_n -mesh contains at least $\dim(\mathbb{W}(\mathbb{F}|y)) + 2$ points. By Corollary 2 the solution to problem (28) can be represented in form (A1). By the condition of Lemma 2 there exists a set $\{\hat{q}_s(\epsilon_n|y)\}_{s=\overline{1,S}} \subseteq \mathcal{C}(\epsilon_n)$, such that $\max_{1 \leq s \leq S} \|\hat{\mathbf{q}}_s(y) - q_s(\epsilon_n|y)\| \leq \epsilon_n$. For the vector $w(\epsilon_n|y) \triangleq \sum_{s=1}^S \hat{P}_s(y) w(q_s(\epsilon_n|y)|y)$ the inequalities

$$|\hat{\mathbf{w}}_1(y) - w_1(\epsilon_n|y)| \leq \sum_{s=1}^S \hat{P}_s(y) |w_1(\hat{\mathbf{q}}_s(y)|y) - w_1(q_s(\epsilon_n|y)|y)| \leq \omega_1(\epsilon_n|y),$$

$$\|\hat{\mathbf{w}}_2(y) - w_2(\epsilon_n|y)\| \leq \sum_{s=1}^S \hat{P}_s(y) \|w_1(\hat{\mathbf{q}}_s(y)|y) - w_1(q_s(\epsilon_n|y)|y)\| \leq \omega_2(\epsilon_n|y)$$

hold. Furthermore, the sequence of inequalities

$$\begin{aligned} \max_{w \in \text{conv}(\mathbb{W}(\mathcal{C}(\epsilon_n)|y))} \mathbf{J}_*(w) &= \mathcal{J}(y) - \min_{w \in \text{conv}(\mathbb{W}(\mathcal{C}(\epsilon_n)|y))} (\hat{\mathbf{w}}_1(y) - w_1 + \|w_2\|^2 - \|\hat{\mathbf{w}}_2(y)\|^2) \geq \\ &\geq \mathcal{J}(y) - \left[|\hat{\mathbf{w}}_1(y) - w_1(\epsilon_n|y)| + \|\hat{\mathbf{w}}_2(y) - w_2(\epsilon_n|y)\|^2 - 2\langle \hat{\mathbf{w}}_2(y), \hat{\mathbf{w}}_2(y) - w_2(\epsilon_n|y) \rangle \right] \geq \\ &\geq \mathcal{J}(y) - \left[\omega_1(\epsilon_n|y) + \left(\omega_2(\epsilon_n|y) + 2 \frac{M}{m(y)} K_h \right) \omega_2(\epsilon_n|y) \right] \end{aligned}$$

proves the convergence $\max_{w \in \text{conv}(\mathbb{W}(\mathcal{C}(\epsilon_n)|y))} \mathbf{J}_*(w) \uparrow \mathcal{J}(y)$ as $\epsilon_n \downarrow 0$.

Let $\hat{\mathbf{w}}(n|y) \rightarrow \hat{\mathbf{w}}(y)$ as $\epsilon_n \downarrow 0$. Then there exists a subsequence $\{\epsilon_{n_k}\}_{n_k \in \mathbb{N}}$, such that $\hat{\mathbf{w}}(n_k|y) \rightarrow \bar{\mathbf{w}}(y) \neq \hat{\mathbf{w}}(y)$. This means that $\mathbf{J}_*(\hat{\mathbf{w}}(y)) = \mathbf{J}_*(\bar{\mathbf{w}}(y))$, which contradicts the uniqueness of the solution to the finite-dimensional dual problem (28). Lemma 2 is proven. \square

References

1. Calafiore, G.; El Ghaoui, L. Robust maximum likelihood estimation in the linear model. *Automatica* **2001**, *37*, 573–580. [\[CrossRef\]](#)
2. Kurzhanski, A.B.; Varaiya, P. *Dynamics and Control of Trajectory Tubes*; Birkhäuser: Basel, Switzerland, 2014.
3. Matasov, A. *Estimators for Uncertain Dynamic Systems*; Kluwer: Dordrecht, The Netherlands, 1998.
4. Borisov, A.V.; Pankov, A.R. Optimal filtering in stochastic discrete-time systems with unknown inputs. *IEEE Trans. Autom. Control* **1994**, *39*, 2461–2464. [\[CrossRef\]](#)
5. Pankov, A.R.; Semenikhin, K.V. Minimax identification of a generalized uncertain stochastic linear model. *Autom. Remote Control* **1998**, *59*, 1632–1643.
6. Poor, V.; Looze, D. Minimax state estimation for linear stochastic systems with noise uncertainty. *IEEE Trans. Autom. Control* **1981**, *26*, 902–906. [\[CrossRef\]](#)
7. Soloviev, V. Towards the Theory of Minimax-Bayesian Estimation. *Theory Probab. Its Appl.* **2000**, *44*, 739–754. [\[CrossRef\]](#)
8. Blackwell, D.; Girshick, M. *Theory of Games and Statistical Decisions*; Wiley: New York, NY, USA, 1954.

9. Martin, C.; Mintz, M. Robust filtering and prediction for linear systems with uncertain dynamics: A game-theoretic approach. *IEEE Trans. Autom. Control* **1983**, *28*, 888–896. [[CrossRef](#)]
10. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*; Springer: Berlin/Heidelberg, Germany, 1985.
11. Anan'ev, B. Minimax Estimation of Statistically Uncertain Systems Under the Choice of a Feedback Parameter. *J. Math. Syst. Estim. Control* **1995**, *5*, 1–17.
12. Borovkov, A. *Mathematical Statistics*; Australia Gordon & Breach: Blackburn, Australia, 1998.
13. Epstein, L.; Ji, S. Ambiguous volatility, possibility and utility in continuous time. *J. Math. Econ.* **2014**, *50*, 269–282. [[CrossRef](#)]
14. Borisov, A.V. A posteriori minimax estimation with likelihood constraints. *Autom. Remote Control* **2012**, *73*, 1481–1497. [[CrossRef](#)]
15. Arkhipov, A.; Semenikhin, K. Minimax Linear Estimation with the Probability Criterion under Unimodal Noise and Bounded Parameters. *Autom. Remote Control* **2020**, *81*, 1176–1191. [[CrossRef](#)]
16. Germeier, Y. *Non-Antagonistic Games*; Springer: New York, NY, USA, 1986.
17. Elliott, R.J.; Moore, J.B.; Aggoun, L. *Hidden Markov Models: Estimation and Control*; Springer: New York, NY, USA, 1995.
18. Yosida, K. *Functional Analysis*; Grundlehren der Mathematischen Wissenschaften; Springer: Berlin/Heidelberg, Germany, 2013.
19. Liptser, R.; Shiryaev, A. *Statistics of Random Processes: I. General Theory*; Springer: Berlin/Heidelberg, Germany, 2001.
20. Kats, I.; Kurzhanskii, A. Estimation in Multistep Systems. *Proc. USSR Acad. Sci.* **1975**, *221*, 535–538.
21. Petersen, I.R.; James, M.R.; Dupuis, P. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Trans. Autom. Control* **2000**, *45*, 398–412. [[CrossRef](#)]
22. Xie, L.; Ugrinovskii, V.A.; Petersen, I.R. Finite horizon robust state estimation for uncertain finite-alphabet hidden Markov models with conditional relative entropy constraints. In Proceedings of the 2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601), Nassau, Bahamas, 14–17 December 2004; Volume 4, pp. 4497–4502. [[CrossRef](#)]
23. El Karoui, N.; Jeanblanc Picque, M. Contrôle de processus de Markov. *Séminaire Probab. Strasbg.* **1988**, *22*, 508–541.
24. Lee, E.; Markus, L. *Foundations of Optimal Control Theory*; SIAM Series in Applied Mathematics; Wiley: Hoboken, NJ, USA, 1967.
25. Floyd, S.; Jacobson, V. Random early detection gateways for congestion avoidance. *IEEE/ACM Trans. Netw.* **1993**, *1*, 397–413. [[CrossRef](#)]
26. Low, S.H.; Paganini, F.; Doyle, J.C. Internet congestion control. *IEEE Control Syst. Mag.* **2002**, *22*, 28–43. [[CrossRef](#)]
27. Altman, E.; Avrachenkov, K.; Menache, I.; Miller, G.; Prabhu, B.J.; Shwartz, A. Dynamic Discrete Power Control in Cellular Networks. *IEEE Trans. Autom. Control* **2009**, *54*, 2328–2340. [[CrossRef](#)]
28. Perruquetti, W.; Barbot, J.P. *Sliding Mode Control in Engineering*; Marcel Dekker, Inc.: New York, NY, USA, 2002.
29. Arnold, B.F.; Stahlecker, P. Fuzzy prior information and minimax estimation in the linear regression model. *Stat. Pap.* **1997**, *38*, 377–391. [[CrossRef](#)]
30. Donoho, D.; Johnstone, I.; Stern, A.; Hoch, J. Does the maximum entropy method improve sensitivity? *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 5066–5068. [[CrossRef](#)] [[PubMed](#)]
31. Donoho, D.L.; Johnstone, I.M.; Hoch, J.C.; Stern, A.S. Maximum Entropy and the Nearly Black Object. *J. R. Stat. Society. Ser. B* **1992**, *54*, 41–81. [[CrossRef](#)]
32. Pham, D.S.; Bui, H.H.; Venkatesh, S. Bayesian Minimax Estimation of the Normal Model with Incomplete Prior Covariance Matrix Specification. *IEEE Trans. Inf. Theory* **2010**, *56*, 6433–6449. [[CrossRef](#)]
33. Donoho, D.L.; Johnstone, I.M. Minimax risk over l_p -balls for l_q -error. *Probab. Theory Relat. Fields* **1994**, *99*, 277–303. [[CrossRef](#)]
34. Donoho, D.L.; Johnstone, I.M. Minimax estimation via wavelet shrinkage. *Ann. Stat.* **1998**, *26*, 879–921. [[CrossRef](#)]
35. Donoho, D.L.; Johnstone, I.; Montanari, A. Accurate Prediction of Phase Transitions in Compressed Sensing via a Connection to Minimax Denoising. *IEEE Trans. Inf. Theory* **2013**, *59*, 3396–3433. [[CrossRef](#)]
36. Bosov, A.; Borisov, A.; Semenikhin, K. Conditionally Minimax Prediction in Nonlinear Stochastic Systems. *IFAC-PapersOnLine* **2015**, *48*, 802–807. [[CrossRef](#)]
37. Kadets, V. *A Course in Functional Analysis and Measure Theory*; Springer: Berlin/Heidelberg, Germany, 2018.
38. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian Data Analysis*, 2nd ed.; Chapman and Hall/CRC: London, UK, 2004.
39. Anderson, B.; Moore, J. *Optimal Filtering*; Prentice-Hall: Upper Saddle River, NJ, USA, 1979.
40. Grabiner, J.; Balakrishnan, A. *Applications of Mathematics: Applied Functional Analysis*; Applications of Mathematics; Springer: New York, NY, USA, 1981.