A Normative Probabilistic Design of a Fair Government Decision Strategy

Miroslav Kárný, Jan Kracík

ÚTIA AV ČR, POBox 18, 18208 Prague, Czech Republic

tel. +420-266062274, fax. +420-266052068

http://utia.cas.cz, E-mail: school@utia.cas.cz

Abstract

A government (a global decision maker) is supposed to search for a fair strategy generating its decisions influencing a population of citizens (local decision makers). The strategy should respect the fact that each citizen has their personal preferences as well as observation and decision spaces. A non-standard problem formulation and its solution are proposed.

Specifically, each citizen (male, female, possibly child) is supposed to express his wishes and restrictions in the way that can be translated into an ideal distribution of data he observes and influences. He is recommended to select his decision strategy so that the real distribution of these data is close to his ideal distribution. This approach is called fully probabilistic design. The government is assumed to be able to influence a few data entries that have an impact on each citizen. The government optimal decision strategy is also formulated in the fully probabilistic sense with its ideal defined as a mixture of the ideal distributions of citizens. Portions of different types of citizens are taken as weights of components forming the mixture.

The paper characterizes individual problem elements and information flow, provides an approximate feasible solution and specializes it to normal government model and normal ideal distributions of citizens. Qualitative consequences with respect to rational governing are drawn.

Keywords: Fully probabilistic design, Bayesian decision making, Multi-participant decision making, E-democracy

1 Introduction

Searching for a compromise between different goals is a well-developed art, dominated by Pareto and Nash equilibria. A range of variants have been published and applied in the past decades. Decision making with multiple decision makers represents an important class of tasks belonging to this category. A commonly accepted and sufficiently general approach to practical problems is, however, missing (cf. a dearth of references on the Web of Science). Such an approach should respect the limited and uncertain information of individual decision makers as well as different constraints on their decisions and acceptable consequences.

Here, a step is proposed towards a desirable general solution. It is undertaken for a specific scenario involving a global decision maker (government) whose decisions influence local decision makers (citizens). Each citizen is supposed to express his wishes and restrictions in a way that can be translated into an ideal distribution of data he deals with. The government is to select the strategy which searches for a compromise among these disparate wishes and restrictions using the data available at government level.

The paper recommends to adopt a fully probabilistic design (Kárný, 1996) for the design of the government strategy. A rational government has to create a model relating its data and decisions. Given its complex uncertain environment, the model should be probabilistic in nature. Here, the government is, moreover, recommended to take a convex combination of appropriately extended citizen ideal probability density functions (pdfs) as the ideal distribution to be used in the design of its strategy. The proposed approach

- provides a normative solution of complex decision making under uncertainty with many independent decision makers and a single global one influencing all of them,
- respects not only differences in aims and restrictions but also differences among observation and decision spaces of citizens,
- uses a realistic description of practically met circumstances,
- describes a normative way for decision making of citizens but does not rely on its implementation: just uses their data,
- brings a qualitative and potentially quantitative insight into the structure of the problem solved.

We close this introduction with a few comments on our approach. It is based on the following simple engineering principles.

- Everything around us is uncertain and approximate so that a soft description is a good model of reality. A formalization in terms of probabilistic beliefs is better than a fuzzy-set-based one at least because it allows us to estimate citizen preferences from data.
- The World is "optimized" by optimizing optional elements of the soft model that is to be kept as close as possible to the real World behavior. The best achievable compromise under given circumstances is to be constructed and then analyzed. This avoids danger that a priori unrealistic solutions are inspected.
- It is waste of the technical power of the Internet to use it as a voting mechanism only. It can and should serve as an intelligent information channel connecting

stratified society.

The proposed approach has proved itself invaluable in number of engineering problems Kárný et al. (2003), nevertheless, the field of its applications is, in our opinion, much wider. We hope, this paper can contribute a new point of view to multi criteria decision making. We also believe we can avoid the traditional traps of the area, like Arrow's impossibility theory (Arrow, 1995), by introducing soft preferences in the vein (Nurmi, 2001).

The layout of the paper is as follows.

Section 2 introduces basic notation and recalls fully probabilistic design of decision strategies (Kárný, 1996). It allows the problem to be formulated and the articulation of its component elements.

Section 3 provides an approximation to the general solution of the optimum design of government strategy.

Section 4 elaborates algorithmic details of this design for government model and citizen ideals in the form of multivariate autoregressive models with exogenous variable (ARX). Practical importance of this special case makes us to present a complete solution in a numerically safe algorithmic forms. Readers who are not interested in manipulations with factorized kernels of quadratic form may skip this section if they are willing to take applicability of the proposed approach to large scale problems for granted.

Section 5 outlines how to get the involved elements through the standard procedures of descriptive statistics complemented by Internet-based extraction of citizens wishes. Qualitative consequences of the formal design are in Section 6. For a better readability, proofs and two auxiliary Propositions are shifted into Appendix.

2 Preliminaries and problem formulation

The following notation is used throughout the text.

Symbol	**Meaning**
$\equiv$	equality by definition
x^*	a set of x-values
$\mathring{x}$	length of a vector x or cardinality of a finite set x^*
$f(\cdot\|\cdot)$	probability density functions (pdf)
$F(\cdot\|\cdot)$	ideal pdf describing desired behavior
$\mathcal{E}$	expectation
t	discrete-time, always the last subscript separated by semicolon
$d = (\Delta, a)$	data record = (observable consequences, decisions)
$d(t)$	sequence $(d_1, \ldots, d_t)$
$'$	transposition
I, (I_k)	unit $((k,k))$ matrix
L	lower triangular matrix with ones on diagonal
D	diagonal matrix with non-negative diagonal
${}^{\llcorner a}B$	a non-numerical superscript a of a variable B
$\mathcal{N}_x(\mu, r)$	normal pdf of x with mean μ and variance r

The pdfs are distinguished by the identifiers in their arguments. No formal distinction is made between random variable, its realization and an argument of a pdf.

The correct meaning follows from the context. The elementary properties of pdfs (normalization, marginalization, chain and Bayes rules) are used (Peterka, 1981). The adopted probabilistic modelling operates on the joint pdf of all uncertain variables encountered. It composes this joint pdf through the chain rule and derives its particular marginal or conditional versions. It inserts the measured realization of any variable at disposal into the treated pdf. Here, a sequence $d(\mathring{t}) = (d_1, \ldots, d_{\mathring{t}})$, over discrete time $t \in t^* \equiv \{1, \ldots, \mathring{t}\}$, of multi-variate data records d_t is considered. Each data record d_t consists of optional decisions a_t and of *observable consequences* Δ_t. An optimized *causal strategy*, mapping randomly $d^*(t-1) \to a_t^*$, $t \in t^*$, generates the decisions. Causality is meant in information sense, i.e. the data $d(t-1)$ can be at most used for selecting the decision a_t. The optimization of the mappings forming the strategy is performed up to the *design horizon* $\mathring{t}$.

2.1 Ideal pdf

Joint pdf $f(d(\mathring{t}))$ of data sequences $d(\mathring{t})$ is a widely accepted tool for describing *where the data are expected* to be. It models joint realizations of (possibly random) decisions and their observable uncertain consequences. An ideal pdf $F(d(\mathring{t}))$, discussed here, is its counterpart for describing *where the data are required* to be. It expresses wishes and constraints of decision maker by prescribing where joint realizations of (possibly random) decisions and their observable uncertain consequences should ideally be. The higher value of the ideal pdf, the more preferred data. A zero value of the ideal pdf means that the corresponding data are absolutely undesirable. Using ideal pdfs for preferences descriptions allows greater discrimination than simple "yes" or "no" statements used in classical voting system. It offers a detailed quantitative characterization of preferences.

For example, let us consider an ideal pdf to be a one-dimensional Gaussian distribution. In such a case, the mean corresponds to the most preferred value, and the variance characterizes strictness of the preferences. Two extreme cases, in this sense, are distributions with the variance close to zero and to infinity. The first one describes preferences where only values in a very narrow range, close to the mean, are acceptable; in the second one, all data have the same preferences.

Adopting an ideal pdf does not mean that preferences are of random nature; it is only a tool for describing them. A merit of this approach is that ideal pdfs provide a unified, intuitively plausible way to articulate preferences on commodities of a very different physical nature. This applies also to decisions whose domain is "naturally" restricted to support of the ideal pdf. Moreover, its introduction makes it possible to define a "universal" design of optimal decision strategies with an explicit and unique optimizing solution. These statements are supported by results presented in the next subsection.

2.2 Fully probabilistic design of decision strategy

Both the government and citizens are recommended to use, so called, fully probabilistic design (Kárný, 1996) to optimize their decision strategies. The design, recalled here, uses the *Kullback-Leibler (KL) distance* (Kullback and Leibler, 1951) as a measure of a proximity of a pair of pdfs f, g acting on a common set x^*. Their KL distance $\mathcal{D}(f||g)$ is defined by the formula

$$\mathcal{D}(f||g) \equiv \int f(x) \ln \left(\frac{f(x)}{g(x)} \right) dx. \tag{1}$$

We shall need the following properties of the KL distance

$$\mathcal{D}(f||g) \geq 0,\ \mathcal{D}(f||g) = 0 \text{ iff } f = g \text{ almost everywhere.} \tag{2}$$

The fully probabilistic design problem is formulated and solved as follows. The joint pdf $f(d(\mathring{t})) \equiv f(\Delta(\mathring{t}), a(\mathring{t}))$ describing globally observable data sequences up to the considered horizon $\mathring{t}$ can be factorized by a repetitive use of the chain rule

$$f(\Delta(\mathring{t}), a(\mathring{t})) = \prod_{t \in t^*} f(\Delta_t | a_t, d(t-1)) f(a_t | d(t-1)), \tag{3}$$

$d(0)$ is assumed to be fixed known prior information.

The decomposition (3) contains two types of factors under the product sign.

The first factors $\{f(\Delta_t | a_t, d(t-1))\}_{t \in t^*}$ describe immediate observable consequences of the decision a_t under the available experience reflected in the data $d(t-1)$. These pdfs form the outer model of the system in question. The model should reflect objective relationships between decisions and their consequences as exactly as possible. It is constructed a priori through a combined modelling and parameter or state estimation Peterka (1981); Ljung (1987) and represents a given part of the deign.

Similarly, the factors $\{f(a_t | d(t-1))\}_{t \in t^*}$ represent an outer model of the used randomized decision strategy. It should be selected within the design, it is optimized optional part of the overall probabilistic description of interactions of decisions and observable consequences.

The recognition of the optional factors in the global probabilistic suggests designing the decision strategy as a minimizer of the KL distance of the joint pdf (3) to a pre-specified ideal joint pdf $F(d(\mathring{t}))$ that assigns preferred occurrences among possible data $d(\mathring{t})$. The design is formalized as follows.

The ideal joint pdf $F(d(\mathring{t}))$ is assumed to be specified in the way mimic to (3)

$$F(d(\mathring{t})) = \prod_{t \in t^*} F(\Delta_t | a_t, d(t-1)) F(a_t | d(t-1)). \tag{4}$$

The optimal strategy is selected among causal, possibly randomized, decision strategies $\{f(a_t|d(t-1)\}_{t\in t^*}$ as a minimizer of the KL distance (1)

$$\mathcal{D}\left(f||F\right) \equiv \int f(\Delta(\mathring{t}), a(\mathring{t})) \ln \left(\frac{f(\Delta(\mathring{t}), a(\mathring{t}))}{F(\Delta(\mathring{t}), a(\mathring{t}))} \right) d(\Delta(\mathring{t}), a(\mathring{t})) \tag{5}$$

of the pdf $f(\Delta(\mathring{t}), a(\mathring{t}))$ (3) to the ideal pdf $F(\Delta(\mathring{t}), a(\mathring{t}))$ (4).

Proposition 1 (Fully probabilistic optimal strategy) *The optimal strategy minimizing the KL distance (5) of the pdf (3) to the pdf (4) is (almost surely) unique and has the form*

$$\begin{aligned}
f^o(a_t|d(t-1)) &= F(a_t|d(t-1)) \frac{\exp[-\omega_\gamma(a_t, d(t-1))]}{\gamma(d(t-1))}, \quad \textit{where} \\
\gamma(d(t-1)) &\equiv \int F(a_t|d(t-1)) \exp[-\omega_\gamma(a_t, d(t-1))]\, da_t \\
\omega_\gamma(a_t, d(t-1)) &\equiv \int f(\Delta_t|a_t, d(t-1)) \times \\
&\times \ln \left(\frac{f(\Delta_t|a_t, d(t-1))}{\gamma(d(t)) F(\Delta_t|a_t, d(t-1))} \right) d\Delta_t \\
\gamma(d(\mathring{t})) &= 1.
\end{aligned}$$

The solution of these functional equations is performed against the time course, starting at $t = \mathring{t}$.

For proof see Appendix.

The result of the fully probabilistic design, i.e., the optimal strategy $f^o(a_t|d(t-1))$, can be taken in as a recommended distribution of the preferences of the decisions a_t. Again, it does not mean that optimal decisions are of a random nature. While the ideal pdf $F(a_t|d(t-1))$ describes preferences of the quantities

a_t without taking into account their impact on observable consequences Δ_t, the optimal pdf $f^o(a_t|d(t-1))$ describes preferences of decisions a_t so that all quantities (a_t and Δ_t) are as close (in the sense of KL distance) to their ideal pdfs as possible.

The following example illustrates using of the fully probabilistic design and an impact of variances of ideal pdfs on optimal pdfs.

Example Let us consider a static system with a one-dimensional decision a and a one-dimensional observable consequence Δ. The outer model of the system is

$$f(\Delta|a) = \mathcal{N}_\Delta(\mu + \theta a, r),$$

where $\mathcal{N}_x(\mu, r)$ denotes normal pdf of the quantity x with mean μ and variance r. Let the ideal pdfs $F(a)$ and $F(\Delta|a)$ be given by independent normal distributions, i.e.,

$$F(a) = \mathcal{N}_a(M_a, R_a), \quad F(\Delta|a) = \mathcal{N}_\Delta(M_\Delta, R_\Delta).$$

In order to illustrate results of the fully probabilistic design in a graphical form, let us consider the following values of the parameters: $\mu = 2$, $r = 1$, $\theta = 0.5$, $M_\Delta = 1$, $R_\Delta = 1$, $M_a = 2$,and R_a being either 1, or 5, or 25.

The results of the fully probabilistic design, for the individual values of R_a, are in Figures 1a, 1b, and 1c. The figures in the upper row represents the ideal and optimal pdfs of decisions a, i.e., $F(a)$ and $f^o(a)$; the figures in the bottom row demonstrates the corresponding pdfs of observable consequences Δ - the ideal ones and marginal ones for optimal distributions of decisions, i.e., $F(\Delta)$ and $f^o(\Delta) = \int f(\Delta|a)f^o(a)da$. The ideal pdfs are plotted by dotted lines, and the optimal ones by solid lines.

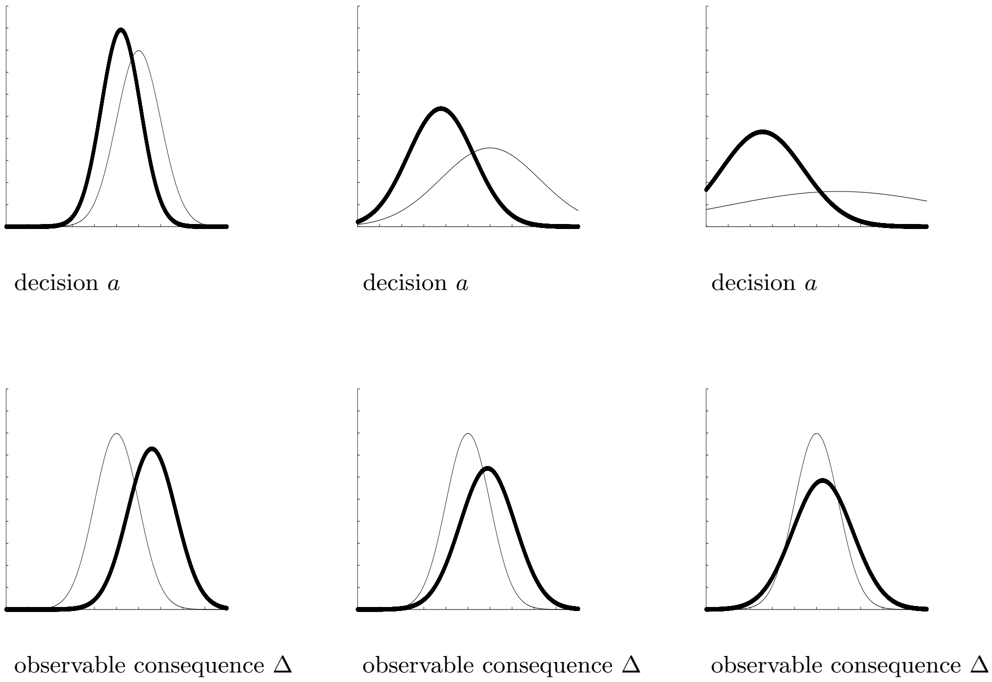

Fig. 1a: $R_a = 1$ Fig. 1b: $R_a = 5$ Fig. 1c: $R_a = 25$

The example clearly demonstrates an impact of a variance of ideal pdfs on optimal pdfs. The narrow ideal distribution $F(a)$, (Fig. 1a) causes the mean of the optimal pdf $f^o(a)$ to be close to the ideal pdf, and consequently, the mean of the corresponding optimal pdf $f^o(\Delta)$ to be relatively far from the ideal pdf $F(\Delta)$. The consequences of the opposite case – the wide pdf $F(a)$ – are shown in Figure 1c. The results of the intermediate case are in Figure 1b.

2.3 Problem formulation

Each citizen deals with his data sequence $d_c(\mathring{t})$, $c \in c^* \equiv \{1, \ldots, \mathring{c}\}$. The respective data records $d_{c;t}$ consist of citizen observable consequences $\Delta_{c;t}$ and decisions $a_{c;t}$, i.e., $d_{c;t} = (\Delta_{c;t}, a_{c;t})$. The c-th citizen is assumed to express his objectives in a way that can be translated into the ideal pdf $\prod_{t \in t^*} F(d_{c;t}|d_c(t-1), c)$. The construction, discussed practically in Section 5, relies on the ability of individuals to generate desirable ranges on respective data and their changes. Grouping of individual opinions with the *common data* $d_c(\mathring{t})$ *and similar desired ranges* leads to a pdf describing a cluster of similar personal wishes. Consequently, we can deal with groups of citizens (with similar preferences) instead of individual citizens. Thus, the clustering allows $\mathring{c}$ to be kept small enough.

A citizen would be advised to apply the strategy optimal in the fully probabilistic sense, Proposition 1, but ultimately the choice is fully his responsibility.

The *government* is assumed to generate decisions $a_g(\mathring{t})$ that influence citizens indirectly, i.e. the citizen observable consequences $\Delta_{c;t}$ can be split into observable consequences without the government decision $a_{g;t}$, denoted $\delta_{c;t}$, and government decisions $a_{g;t}$

$$\Delta_{c;t} = (\delta_{c;t}, a_{g;t}). \tag{6}$$

Note that $\delta_{c;t}$ may contain also a part of the government observable consequences $\Delta_{g;t}$. Thus, citizens may be influenced both by the chosen government decisions and by their consequences at the global level.

Thus, the problem can be formulated as follows.

A government strategy $d_g^*(t-1) \rightarrow a_{g;t}^*$ *is required that leads to a fair compromise among wishes of different citizens.*

For simplicity, common sampling time points $t \in t^*$ as well as a common joint horizon $\mathring{t}$ are considered. It represents no real restriction when the finest of the involved time grids is adopted and the piece-wise constant approximations are used for other decision makers.

2.4 Data sets and extension of citizen ideal pdfs

Information extents available to government and particular citizens always differ. Proper treatment of this always present difference requires formalisation of this fact. Formally, the set d_g^* of the data d_g available to the government can be related to the data set d_c^* of the c-th citizen as follows

$$d_g^* = \underbrace{d_{gc+}^*}_{\text{surplus } g\text{-data}} \cup \underbrace{d_{gc}^*}_{\text{common data}} \cup \underbrace{d_{gc-}^*}_{\text{surplus } c\text{-data}}$$
$$d_{gc+}^* \cap d_{gc}^* = \emptyset,\ d_{gc}^* \cap d_{gc-}^* = \emptyset,\ d_{gc+}^* \cap d_{gc-}^* = \emptyset,$$

where the *surplus g-data* d_{gc+} are data available to the government but not to the citizen, the *common data* d_{gc} are available both to the government and the citizen, the *surplus c-data* d_{gc-} are available to the citizen but not to the government.

The government cannot rationally influence the unknown surplus c-data. Its decision making will be the same irrespectively whether d_{gc-}^* is empty or not. Thus, without loss of generality, we can and will assume that $d_{gc-}^* = \emptyset$, i.e. $d_c^* \subset d_g^*$, $\forall c \in c^*$.

The c-th citizen specifies his ideal pdf on his data d_c only. Thus, he leaves the surplus g-data d_{gc+} to their fate. Consequently, he has to accept the pdf that arises from the applied government strategy as the ideal pdf on the surplus data. Altogether, $d_g^* = d_{gc+}^* \cup d_c^*$, $d_{gc+}^* \cap d_c^* = \emptyset$ and the extension $F(d_{g;t}|d_g(t-1), c)$ of the ideal pdf

$F(d_{c;t}|d_c(t-1), c)$ of c-th citizen on the government data d_g is

$$F(d_{g;t}|d_g(t-1), c) \equiv F(d_{c;t}|d_c(t-1), c) f(d_{gc+;t}|d_g(t-1)), \tag{7}$$

where $f(d_{gc+;t}|d_g(t-1))$ is the marginal pdf of the pdf $f(d_{g;t}|d_g(t-1))$ describing the data d_g inspected and influenced by the government.

2.5 Government ideal pdf and optimal strategy

The identity (7) extends the ideal pdf of the c-th citizen on the government data. Note that government decisions $a_{g;t}$ are supposed to be known to citizens so that the pdf $f(d_{gc+;t}|d_g(t-1))$ can be determined from the government model irrespective of the government strategy used.

The overall "fair" ideal government pdf is taken as a *weighted sum of the extended ideal pdfs of citizens (i.e., a probabilistic mixture)*

$$\begin{aligned} F(d_{g;t}|d_g(t-1)) &= \sum_{c \in c^*} \alpha_c F(d_{g;t}|d_g(t-1), c) = \\ &= \sum_{c \in c^*} \alpha_c F(d_{c;t}|d_c(t-1), c) f(d_{gc+;t}|d_g(t-1)). \end{aligned} \tag{8}$$

The practical construction of the probabilistic weights α_c is discussed in Section 5. Intuitively, the fair probabilities

$$\alpha = [\alpha_1, \ldots, \alpha_{\mathring{c}}]' \in \alpha^* \equiv \left\{ [\alpha_1, \ldots, \alpha_{\mathring{c}}]' \middle| \alpha_c \geq 0, \sum_{c \in c^*} \alpha_c = 1 \right\}$$

should represent portions of different stakeholders (labelled by c) within the overall population. Stakeholders are individuals with the same data sets and very similar preferences. Their practical grouping is also discussed in Section 5. Formally,

$$\alpha_c = \frac{\mathring{\imath}_c}{\sum_{c \in c^*} \mathring{\imath}_c},$$

where $\mathring{\imath}_c$ is the number of citizens belonging to the c-th group of stakeholders.

The government may treat itself as a specific stakeholder with its specific ideal pdf and combine it with ideal pdfs of other stakeholders into the overall ideal pdf (8). It gives the government a chance to respect the aspects (data) that are unavailable to majority of citizens. A wise government has to, however, assign to itself a "fair" α-weight.

The *optimal government strategy*, generating its decisions

$$d_g^*(t-1) \to a_{g;t}^*,\ t \in t^*,$$

is taken as the outcome of the fully probabilistic design, Proposition 1, with the government ideal pdf (8).

2.6 Government model and citizen ideal pdfs

The complete problem formulation requires specification of the government model $\{f(\Delta_{g;t}|a_{g;t}, d_g(t-1))\}_{t\in t^*}$. Complexity of the overall environment calls for a simplified modelling, local both in time and data space. In other words, an adaptively estimated model is highly desirable (Kárný, 1998). The government view has to be rather aggregated and simplified as the length $\mathring{d}_g$ is, as a rule, much higher than $\mathring{d}_c$. Also, the relatively long period between changes in government decisions, determined by a significant inertia in large scale systems, implies that just the dominant dynamics have to be modelled. Thus, a low order model can be used. Often, even uni-modality can be assumed. A mixture model can also be estimated and used.

When constructing his decision strategy, each citizen should specify his personal model giving, ideally, a pdf $f(\Delta_{c;t}|a_{c;t}, d_c(t-1))$. For the addressed problem, however, we need not force him do that and even we need not to know how he selects his decisions. For the design of the government strategy, the knowledge of ideal pdfs $F(d_{c;t}|d_c(t-1)),\ c \in c^*$, is needed only. The ideal pdfs of respective citizens may differ both in their form and mutual compatibility of assigned preferences to various data configurations. It should be stressed that the government decisions $a_{g;t}$ form a part of citizen's observable consequences $\Delta_{c;t}$, i.e., each citizen should predict decisions of his government and specify his wishes for them. It is also obvious that rational government is possible when citizen observable consequences $\Delta_{c;t}$ do not reduce to government decisions $a_{g;t}$ only, i.e. when the information exchange is bi-directional.

3 Jensen approximation of the design

We want to design the fully probabilistic government strategy $\{d^*_g(t-1) \to a^*_{g;t}\}_{t\in t^*}$. For it, we need not distinguish citizen decisions and observable consequences. Thus, we can set, cf. (6),

$$\Delta_{c;t} = d_{c;t} = (\delta_{c;t}, a_{g;t}). \tag{9}$$

Looking at the definition of the KL distance (1), it is obvious that it cannot be evaluated exactly for the government ideal pdf given as a mixture (8). Thus, instead of it, we minimize its upper bound implied by the Jensen inequality $\mathcal{D}(f||F) \equiv$

$$\begin{aligned}
&\equiv \sum_{t\in t^*} \mathcal{E} \int f(d_{g;t}|d_g(t-1)) \ln\left(\frac{f(d_{g;t}|d_g(t-1))}{\sum_{c\in c^*} \alpha_c F(d_{g;t}|d_g(t-1),c)}\right) dd_{g;t} \leq \\
&\leq \sum_{t\in t^*, c\in c^*} \alpha_c \mathcal{E}\left[\int f(d_{g;t}|d_g(t-1)) \ln\left(\frac{f(d_{g;t}|d_g(t-1))}{F(d_{g;t}|d_g(t-1),c)}\right) dd_{g;t}\right] = \\
&\underbrace{=}_{(8),(9)} \sum_{t\in t^*, c\in c^*} \alpha_c \mathcal{E}\left[\int f(\Delta_{g;t}|a_{g;t}, d_g(t-1)) f(a_{g;t}|d_g(t-1)) \times\right. \\
&\left.\times \ln\left(\frac{f(\delta_{c;t}|a_{g;t}, \Delta_{gc+;t}, d_g(t-1)) f(a_{g;t}|d_g(t-1))}{F(\delta_{c;t}|a_{g;t}, d_c(t-1),c) F(a_{g;t}|d_c(t-1),c)}\right) dd_{g;t}\right].
\end{aligned}$$

The approximation corresponds with the choice of the government ideal pdf equal to the *geometric mean* of the citizen ideal pdfs (8).

Proposition 2 (Government fully probabilistic design) *The optimal strategy* *minimizing the Jensen upper bound on the KL distance has the form*

$$
\begin{aligned}
f^o(a_{g;t}|d_g(t-1)) &= \prod_{c\in c^*} [F(a_{g;t}|d_c(t-1),c)]^{\alpha_c} \frac{\exp[-\omega_\gamma(a_{g;t},d_g(t-1))]}{\gamma(d_g(t-1))} \\
\gamma(d_g(t-1)) &\equiv \int \prod_{c\in c^*} [F(a_{g;t}|d_c(t-1),c)]^{\alpha_c} \exp[-\omega_\gamma(a_{g;t},d_g(t-1))]\, da_{g;t} \\
\omega_\gamma(a_{g;t},d(t-1)) &\equiv \sum_{c\in c^*} \alpha_c \int f(\Delta_{g;t}|a_{g;t},d_g(t-1)) \times \\
& \times \ln\left(\frac{f(\delta_{c;t}|a_{g;t},\Delta_{gc+;t},d(t-1))}{\gamma(d_g(t))F(\delta_{c;t}|a_{g;t},d_c(t-1),c)}\right) d\Delta_{g;t} \\
\gamma(d_g(\mathring{t})) &= 1. \qquad (10)
\end{aligned}
$$

The solution is performed against the time course, starting at $t = \mathring{t}$.

For proof see Appendix.

The support of the optimal strategy is in the intersection of supports of the citizen ideal pdfs. Thus, a well defined optimal strategy exists for the non-empty intersection. Otherwise, some citizen restrictions, given by supports of their ideal pdfs, have to be neglected.

4 Application to normal ARX models

The functional recursion for the optimal strategy (Proposition 2) can be solved numerically in small problems. In large scale cases, that are inherent to our interpretation, mixtures of normal auto-regression models with exogenous variables (ARX) represent the dominant class for which a feasible solution is available. Moreover, such models can be continually updated in a Bayesian sense: weighted, recursive least squares are run (Peterka, 1981; Kárný et al., 1998).

For simplicity, we present the solution in which the government uses a single multivariate ARX model and citizen ideal pdfs are multivariate ARX models, too.

It means

$$f(\Delta_{g;t}|a_{g;t}, d_g(t-1)) = \prod_{i=1}^{\mathring{\Delta}_g} \mathcal{N}_{\Delta_{ig;t}}(\theta'_{ig}\psi_{ig;t}, r_{ig}), \tag{11}$$

where $\mathcal{N}_{\Delta}(\mu, r) \equiv (2\pi r)^{-0.5} \exp[-0.5(\Delta-\mu)^2 r^{-1}]$, θ_{ig} are vectors of regression coefficients, regression vectors are $\psi_{ig;t} \equiv [d_{(i+1)g;t}, \psi'_{(i+1)g;t}]'$ with $\psi_{\mathring{\Delta}_g g;t} = [a'_{g;t}, \phi'_{g;t-1}]'$ and $\phi'_{g;t-1} = [d'_{g;t-1}, \ldots, d'_{g;t-\partial_g}, 1]$; $\partial_g \in \{0, 1, \ldots\}$ is given. We use the definition of $\psi_{ig;t}$ for $i = 0, 1, \ldots, \mathring{d}_g$. Among others, it gives $\Psi_{g;t} = \psi_{0g;t}$ and $\phi_{g;t-1} = \psi_{\mathring{d}_g g;t}$.

The citizen ideal pdf models the desired behavior of the data records available to him

$$F(d_{c;t}|d_c(t-1), c) = \prod_{i=1}^{\mathring{d}_c} \mathcal{N}_{d_{ic;t}}(\theta'_{ic}\psi_{ic;t}, r_{ic}) \tag{12}$$

with its elements mimic to that of government model (11).

In order to get a compact algorithm, we assume that

- $\phi_{g;t-1} = \phi_{c;t-1} \equiv \psi_{\mathring{d}_c c;t}$: it is reached by inserting zeros into θ_{ig} and θ_{ic}; consequently, the subscript g can be dropped at Ψ, ψ, and ϕ,
- $d_{g;t} = (\Delta_{g;t}, a_{g;t}) = (\delta_{c;t}, \Delta_{gc+;t}, a_{g;t})$, $d_{c;t} = (\delta_{c;t}, a_{g;t})$: the identities depend on the ordering only.

Proposition 3 (Government design for ARX models) *The optimal strategy minimizing the Jensen upper bound on the KL distance (10) with models (11), (12) is described by the normal pdfs*

$$f^o(a_{g;t}|d_g(t-1)) = \prod_{i=1}^{\mathring{a}_g} \mathcal{N}_{a_{ig;t}}\left(\theta^o_{i;t}\psi'_{\mathring{\Delta}_g+i;t}, r^o_{i;t}\right) \text{ with} \tag{13}$$

$\psi_{\mathring{\Delta}_g+i;t} \equiv [a_{(i+1)g;t} \ldots, a_{\mathring{a}_g g;t}, \phi'_{t-1}]'$ *and recursively generated,*

time-variant, regression coefficients $\theta^o_{i;t}$ *and variances* $r^o_{i;t}$.

The solution is performed against the time course, starting at $t = \mathring{t}$ *with* $L_{\gamma;\mathring{t}} = I_{\mathring{\phi}}$, $D_{\gamma;\mathring{t}} = 0$, $k_{\gamma;\mathring{t}} = 0$.

$$\textit{For} \quad t = \mathring{t}, \mathring{t} - 1, \ldots, 1$$

$$L_{\mathring{\Delta}_g} = I_{\mathring{\Psi}_{\mathring{\Delta}_g}}, \;\; D_{\mathring{\Delta}_g} = 0$$

$$\textit{For} \quad c = 1, \ldots, \mathring{c}$$

$$k_{0c} = -\mathring{\delta}_c + k_{\gamma;\mathring{t}}, \;\; L_{0c} D_{0c} L'_{0c} = \mathcal{K} L_{\gamma;t} D_{\gamma;t} L'_{\gamma;t} \mathcal{K}'$$

$$\textit{For} \;\; i = 1, \ldots, \mathring{\Delta}_g$$

$$L_{ic} D_{ic} L'_{ic} = {}^{\lfloor\psi} L_{(i-1)c} {}^{\lfloor\psi} D_{(i-1)c} {}^{\lfloor\psi} L'_{(i-1)c} +$$
$$+ \left(\theta_{ig} + {}^{\lfloor\Delta\psi} L_{(i-1)c}\right) {}^{\lfloor\Delta} D_{(i-1)c} \left(\theta_{ig} + {}^{\lfloor\Delta\psi} L_{(i-1)c}\right)' +$$
$$+ \chi\left(i \le \mathring{\delta}_c\right) \frac{(\theta_{ig} - \theta_{ic})(\theta_{ig} - \theta_{ic})'}{r_{ic}}$$
$$k_{ic} = k_{(i-1)c} + {}^{\lfloor\Delta} D_{(i-1)c} r_{ig} + \chi\left(i \le \mathring{\delta}_c\right) \left[\ln\left(\frac{r_{ic}}{r_{ig}}\right) + \frac{r_{ig}}{r_{ic}}\right]$$

end of the cycle over i

$$L_{\mathring{\Delta}_g} D_{\mathring{\Delta}_g} L'_{\mathring{\Delta}_g} = L_{\mathring{\Delta}_g} D_{\mathring{\Delta}_g} L'_{\mathring{\Delta}_g} + \alpha_c L_{\mathring{\Delta}_g c} D_{\mathring{\Delta}_g c} L'_{\mathring{\Delta}_g c}$$

$$k_{\mathring{\Delta}_g} = k_{\mathring{\Delta}_g} + \alpha_c k_{\mathring{\Delta}_g c}$$

end of the cycle over c

$$\textit{For} \quad i = \mathring{\Delta}_g + 1, \ldots, \mathring{d}_g$$

$$\tilde{L}_i \tilde{D}_i \tilde{L}'_i = L_{i-1} D_{i-1} L'_{i-1}, \;\; \ln(r_i) = 0$$

$$\textit{For} \;\; c = 1, \ldots, \mathring{c}$$

$$\tilde{L}_i \tilde{D}_i \tilde{L}'_i = \tilde{L}_i \tilde{D}_i \tilde{L}'_{i-1} + \alpha_c [-1, \theta'_{ic}]' r_{ic}^{-1} [-1, \theta'_{ic}]$$

$$\ln(r_i) = \ln(r_i) + \alpha_c \ln(r_{ic})$$

end of the cycle over c

$$\tilde{L}_i = \begin{bmatrix} 1 & \\ -\theta^o_{i;t} & L_i \end{bmatrix}, \ \tilde{D}_i = \mathrm{diag}\left[r^{o-1}_{i;t}, D_i\right]$$

$$k_i = k_{i-1} + \ln\left(r^o_{i;t}\right) - \ln(r_i)$$

end of the cycle over i

$$L_{\gamma;t-1} = L_{\mathring{d}_g}, \ D_{\gamma;t-1} = D_{\mathring{d}_g}, \ k_{\gamma;t-1} = k_{\mathring{d}_g}$$

end of the cycle over t.

Proof is given in Appendix.

5 Elicitation of the involved elements

Ideal pdfs of stakeholders, their weights in the overall population and government model relating the considered data are key ingredients for deciding on the government strategy. Their possible elicitation is outlined here. It is practicable only when the questions discussed below are predominantly communicated in an electronic way, mostly via the Internet.

5.1 Data and model available to government

Data available to the government are collected by a version of the State Statistical Office that groups them according to more and more standardized classification. For instance, Classification of Individual Consumption by Purpose may be relevant to the addressed task. Those commodity groups that are felt to be relevant to the addressed problem and that are reasonably populated by historical data should be taken as the government data. Among them the variables that can be and should be chosen by the government have to be present.

Bayesian processing of historical data together with the available economical and sociological information provide the government model. Efficient procedure for ARX models can be found in (Peterka, 1981), for normal mixtures in (Kárný et al., 2003). Note that the Bayesian paradigm is (almost) a necessity as the lengths of the time series available are (almost) always too small for asymptotically based analysis. The ever present lack of data is caused by both their high inherent dimensionality and by relatively sparse sampling.

5.2 Citizen data and ideal pdfs

The sub-selection of commodity categories should be presented as a questionnaire to individual citizens or to their sufficiently rich and representative sample. The questionnaire should ask on desired ranges within the individual categories:

> What is your smallest acceptable value ... of the commodity X? What is the sufficiently satisfactory value ... of the commodity X?

The answers provide the pairs of data items for building the corresponding static marginal pdfs in the citizen and consequently government ideal pdfs. By answering the question related to a commodity X that has meaning of a change rate, a dynamic, first order marginal ideal pdf can be specified.

It is important that the citizen is not forced to answer all questions: a lot of data values he may feel are irrelevant to him. Consequently, the questionnaire can be relatively extensive. Then, the questions should be presented according the art developed in connection with design of human computer interface (grouping of related questions with possibility to jump between boxes; answering by a simple ticking of a box; explanatory help, etc.).

The ranges specified by a citizen are complemented to the full data record of government data by automatically filling the technically lowest and highest values for the unspecified ranges. It is justified by the fact that the extended citizen ideal acts as the ideal with a very wide range on data out of the citizen interest.

After this completion, a single normal dynamic mixture is fitted to these data (Kárný et al., 2003), using static or first order marginal pdfs according to the character of individual commodities. The referred fitting includes a complete structure and parameter estimation of the mixture. Thus, it reveals respective stakeholders (described by a single component within the mixture) as well as their probabilistic weights α. The resulting mixture estimate is then the complete government ideal needed.

6 Concluding remarks

The presented formalization implies immediately the following qualitative conclusions that can be converted into quantitative studies.

- The extent of surplus data spaces should be minimized. The data about which citizens are not aware are governed solely by the government strategy. Unsaid wishes can hardly be respected. Consequently, citizens may be quite unhappy about the government strategy adopted. In other words, final quality of the resulting strategy is strongly influenced by the discussed information flow.
- The ideals of citizens have to be known to the government. Formally, it follows from the fact that the government ideal pdf is made of the citizens ideal pdfs.

Practically, it means that a good government can rule well only when it knows the wishes of the citizens.

- The optimal compromise can be found only when the intersection of supports of the citizen ideal pdfs is non-empty, when antagonistic contradictions among citizens are smoothed up. This is formally seen from the form of the optimal government strategy (10).

- The change in the extent of citizen groups should lead to a revision of the government policy. Formally, it is implied by influence of component weights α on government strategy. Practically, it means changes of citizen ideals and shifts among various groups of citizens should be respected. Thus, citizens' ideal pdfs and consequently government strategy should be revised whenever significant changes are expected in population wishes. Note that it does not advocate populism that follows even local changes of opinions. We conjecture that the ideal *pdfs* have rather stable nature and their changes means real changes in population.

- The optimal strategy is non-stationary and depends on closeness to the horizon. It is formally seen from the strategy description. Practically, it means that any planning with a short horizon is dangerous with respect to the achieved performance and may lead to instability of the overall system. For an explicit support of this statement, see (Kárný et al., 1985).

The majority of above points correspond with a common sense. It indicates that the formal treatment, from which they follow, reflects well the basic features of the fair government.

7 Appendix: Proofs and Auxiliary Propositions

Here, outlines of proofs of Propositions presented in body of the text are given together with auxiliary propositions and their proofs.

7.1 Proof of Proposition 1

Proof: The chain rule allows to write the KL distance in the form

$$\mathcal{D}\left(f||F\right) = \mathcal{E}\left\{\sum_{t\in t^*}\int f(a_t|d(t-1))\times\right.$$
$$\left.\times\left[\ln\left(\frac{f(a_t|d(t-1))}{F(a_t|d(t-1))}\right)+\omega(a_t,d(t-1))\right]\,da_t\right\}\text{ with}$$
$$\omega(a_t,d(t-1))\equiv\int f(\Delta_t|a_t,d(t-1))\ln\left(\frac{f(\Delta_t|a_t,d(t-1))}{F(\Delta_t|a_t,d(t-1))}\right)\,d\Delta_t.$$

Let us denote $-\ln(\gamma(d(t)))\equiv$

$$\equiv\min_{\{f(a_{\tau+1}|d(\tau))\}_{\tau=t}^{\mathring{t}}}\mathcal{E}\left\{\sum_{\tau=t+1}^{\mathring{t}}\int f(a_\tau|d(\tau-1))\times\right.$$
$$\left.\times\left[\ln\left(\frac{f(a_\tau|d(\tau-1))}{F(a_\tau|d(\tau-1))}\right)+\omega(a_\tau,d(\tau-1))\right]\,da_\tau|d(t)\right\}.$$

Then, this definition implies that $\gamma(d(\mathring{t}))=1$ and $-\ln(\gamma(d(t)))\equiv$

$$\equiv\min_{f(a_{t+1}|d(t))}\int f(a_{t+1}|d(t))\times$$
$$\times\left[\ln\left(\frac{f(a_{t+1}|d(t))}{F(a_{t+1}|d(t))}\right)+\omega_\gamma(a_{t+1},d(t))\right]\,da_{t+1}\text{ with}$$
$$\omega_\gamma(a_{t+1},d(t))\equiv\int f(\Delta_{t+1}|a_{t+1},d(t))\times$$
$$\times\ln\left(\frac{f(\Delta_{t+1}|a_{t+1},d(t))}{\gamma(d(t+1))F(\Delta_{t+1}|a_{t+1},d(t))}\right)\,d\Delta_{t+1}.\text{ It gives}$$
$$-\ln(\gamma(d(t)))\equiv\min_{f(a_{t+1}|d(t))}\int f(a_{t+1}|d(t))\times$$
$$\times\left[\ln\left(\frac{f(a_{t+1}|d(t))}{\frac{F(a_{t+1}|d(t))\exp[-\omega_\gamma(a_{t+1},d(t))]}{\int F(\tilde{a}_{t+1}|d(t))\exp[-\omega_\gamma(\tilde{a}_{t+1},d(t))]\,d\tilde{a}_{t+1}}}\right)\,da_{t+1}-\right.$$
$$\left.-\ln\left(\int F(a_{t+1}|d(t))\exp\left[-\omega_\gamma(a_{t+1},d(t))\right]\,da_{t+1}\right)\right].$$

The first term in the above identity is the KL distance, that reaches its smallest zero value (2) for the claimed pdf $f^o(a_{g;t}|d_g(t-1))$. At the same time, it defines the form of the reached minimum. $\square$

7.2 Proof of Proposition 2

Proof: Let us denote $-\ln(\gamma(d_g(t))) \equiv$

$$\equiv \min_{\{f(a_{g;\tau}|d_g(\tau-1))\}_{\tau=t+1}^{\mathring{t}}} \sum_{\tau=t+1}^{\mathring{t}} \sum_{c\in c^*} \alpha_c \mathcal{E}\left[\int f(\Delta_{g;\tau}|a_{g;\tau}, d_g(\tau-1)) \times \right.$$
$$\times f(a_{g;\tau}|d_g(\tau-1)) \times$$
$$\left.\times \ln\left(\frac{f(\delta_{c;\tau}|a_{g;\tau}, \Delta_{gc+;\tau}, d_g(\tau-1)) f(a_{g;\tau}|d_g(\tau-1))}{F(\delta_{c;\tau}|a_{g;\tau}, d_c(\tau-1), c) F(a_{g;\tau}|d_c(\tau-1), c)}\right) dd_{g;\tau} \middle| d_g(t)\right].$$

Then, using the chain rule, we can write the Jensen upper bound of the KL distance in the form $-\ln(\gamma(d_g(t-1))) =$

$$= \min_{f(a_{g;t}|d_g(t-1))} \mathcal{E}\left\{\int f(a_{g;t}|d_g(t-1)) \times \right.$$
$$\left. \times \left[\ln\left(\frac{f(a_{g;t}|d_g(t-1))}{\prod_{c\in c^*}[F(a_{g;t}|d_c(t-1), c)]^{\alpha_c}}\right) + \omega_\gamma(a_{g;t}, d_g(t-1))\right] da_{g;t} | d_g(t-1)\right\}$$
$$\omega_\gamma(a_{g;t}, d_g(t-1)) \equiv \sum_{c\in c^*} \alpha_c \int f(\Delta_{g;t}|a_{g;t}, d_g(t-1)) \times$$
$$\times \ln\left(\frac{f(\delta_{c;t}|a_{g;t}, \Delta_{gc+;t}, d_g(t-1))}{\gamma(d_g(t)) F(\delta_{c;t}|a_{g;t}, d_c(t-1), c)}\right) d\Delta_{g;t}.$$

The rest of the proof is identical with that of Proposition 1. $\square$

7.3 Propositions on Expectation of Quadratic Forms

For an algorithmic solution of fully probabilistic design with normal ARX models, Proposition 3, and its proof in subsection 7.4, we need the following propositions.

Proposition 4 (Expected quadratic form) *Let us consider a normal ARX model (11) with the subscript g suppressed. Let us assume a quadratic form in $\psi_{0;t} \equiv \Psi_t$*

with the decomposed kernel $L_0 D_0 L_0'$. Here, L_0 is a given lower triangular matrix with unit diagonal and D_0 is a given diagonal matrix with non-negative diagonal entries.

Then, the expected quadratic form lifted by a constant k_0 is

$$\mathcal{E}[k_0 + \Psi_t' L_0 D_0 L_0' \Psi_t | a_t, \phi_{t-1}] \equiv \mathcal{E}[k_0 + \psi_{0;t}' L_0 D_0 L_0' \psi_{0;t} | \psi_{\mathring{\Delta};t}] =$$

$$= k_{\mathring{\Delta}} + \psi_{\mathring{\Delta};t}' L_{\mathring{\Delta}} D_{\mathring{\Delta}} L_{\mathring{\Delta}}' \psi_{\mathring{\Delta};t}, \quad \textit{where, for } i = 0, \ldots, \mathring{\Delta} - 1,$$

$$L_{i+1} D_{i+1} L_{i+1}' = {}^{\lfloor \psi}L_i \, {}^{\lfloor \psi}D_i \, {}^{\lfloor \psi}L_i' + \left(\theta_{i+1} + {}^{\lfloor \Delta\psi}L_i\right) {}^{\lfloor \Delta}D_i \left(\theta_{i+1} + {}^{\lfloor \Delta\psi}L_i\right)'$$

$$L_i \equiv \begin{bmatrix} 1 & 0 \\ {}^{\lfloor \Delta\psi}L_i & {}^{\lfloor \psi}L_i \end{bmatrix}, \; D_i \equiv \operatorname{diag}\left[{}^{\lfloor \Delta}D_i, \; {}^{\lfloor \psi}D_i\right], \; k_{i+1} = k_i + {}^{\lfloor \Delta}D_i r_{i+1}$$

where ${}^{\lfloor \Delta}D_i$ is scalar and $\mathring{D}_{i+1} = \mathring{D}_i - 1$.

The evaluation of the LDL' decomposition can be performed by an efficient rank-one updating (Bierman, 1977).

Proof: The expectation is taken over entries of Δ_t as the remaining part of the data vector Ψ_t is fixed by the condition $\psi_{\mathring{\Delta};t} \equiv [a_t', \phi_{t-1}']'$. The chain rule for expectations implies that we can evaluate conditional expectations of individual entries in the vector Δ_t one-by-one, starting from the first one. Taking a generic step and using the identity

$$\mathcal{E}\left[\Delta_{i+1}^2 | \psi_{i+1}\right] = r_{i+1} + \{\mathcal{E}[\Delta_{i+1} | \psi_{i+1}]\}^2, \quad \text{we have}$$

$$\mathcal{E}\left[k_i + \psi_{i;t}' L_i D_i L_i' \psi_{i;t} | \psi_{i+1;t}\right] = \underbrace{k_i + {}^{\lfloor \Delta}D_i r_{i+1}}_{k_{i+1}} +$$

$$+\psi_{i+1;t}' \begin{bmatrix} \theta_{i+1}' \\ I_{\mathring{\psi}_{i+1}} \end{bmatrix}' L_i D_i L_i' \begin{bmatrix} \theta_{i+1}' \\ I_{\mathring{\psi}_{i+1}} \end{bmatrix} \psi_{i+1;t} = k_{i+1} +$$

$$+\psi_{i+1;t}' \underbrace{\left[{}^{\lfloor \psi}L_i \, {}^{\lfloor \psi}D_i \, {}^{\lfloor \psi}L_i' + \left(\theta_{i+1} + {}^{\lfloor \Delta\psi}L_i\right) {}^{\lfloor \Delta}D_i \left(\theta_{i+1} + {}^{\lfloor \Delta\psi}L_i\right)'\right]}_{L_{i+1} D_{i+1} L_{i+1}'} \psi_{i+1;t}.$$

$\square$

Proposition 5 (The weighted conditional KL distance) *Let us consider models (11), (12) with regression coefficient complemented by zeros so that regression vectors $\psi_{i;t}$ of i-th factors coincide and the subscript at them can be dropped.*

Let the function $\gamma(d_g(t))$ (10) equal to $\gamma(\phi_t)$ with

$$\gamma(\phi_t) \equiv \exp\left[-0.5(k_\gamma + \phi_t' L_\gamma D_\gamma L_\gamma' \phi_t)\right], \quad \textit{where}$$

$$\phi_t \equiv [d_t', \ldots, d_{(t-\partial+1)}', 1]' \Rightarrow \psi_{0;t} \equiv \Psi_t \equiv [\Delta_{g;t}', a_{g;t}', \phi_{t-1}']',$$

$$L_\gamma \equiv \text{ a lower triangular matrix with unit diagonal}$$

$$D_\gamma \equiv \text{ a diagonal matrix with non-negative diagonal.}$$

$$\textit{Then,} \quad \omega_\gamma(c, a_{g;t}, \phi_{t-1}) \equiv \tag{14}$$

$$\equiv 2\int f(\Delta_t|a_{g;t}, \phi_{t-1}) \ln\left(\frac{f(\delta_{c;t}|a_{g;t}, \Delta_{gc+;t}, \phi_{t-1})}{\gamma(\phi_t) F(\delta_{c;t}|a_{g;t}, \phi_{t-1}, c)}\right) d\Delta_t =$$

$$= k_{\mathring{\Delta}c} + \psi_{\mathring{\Delta};t}' L_{\mathring{\Delta}c} D_{\mathring{\Delta}c} L_{\mathring{\Delta}c}' \psi_{\mathring{\Delta};t}, \text{ is determined recursively}$$

$$L_{ic} D_{ic} L_{ic}' = {}^{\lfloor\psi}L_{(i-1)c} {}^{\lfloor\psi}D_{(i-1)c} {}^{\lfloor\psi}L_{(i-1)c}' +$$

$$+ \left(\theta_{ig} + {}^{\lfloor\Delta\psi}L_{(i-1)c}\right) {}^{\lfloor\Delta}D_{(i-1)c} \left(\theta_{ig} + {}^{\lfloor\Delta\psi}L_{(i-1)c}\right)' +$$

$$+\chi\left(i \leq \mathring{\delta}_c\right) \frac{(\theta_{ig} - \theta_{ic})(\theta_{ig} - \theta_{ic})'}{r_{ic}}, \; i = 1, \ldots, \mathring{\Delta}_g$$

$$k_{ic} = k_{(i-1)c} + {}^{\lfloor\Delta}D_{(i-1)c} r_{ig} + \chi\left(i \leq \mathring{\delta}_c\right)\left[\ln\left(\frac{r_{ic}}{r_{ig}}\right) + \frac{r_{ig}}{r_{ic}}\right]$$

$$k_{0c} = -\mathring{\delta}_c + k_\gamma, \; L_{0c} D_{0c} L_{0c}' = \mathcal{K} L_\gamma D_\gamma L_\gamma' \mathcal{K}', \; \forall c \in c^*$$

$$\mathcal{K}' \equiv \begin{bmatrix} I_{\mathring{d}(\partial-1)} & 0 & 0 \\ 0 & 0_{1,\mathring{d}} & 1 \end{bmatrix}, \quad L_{ic} \equiv \begin{bmatrix} 1 & 0 \\ {}^{\lfloor\Delta\psi}L_{ic} & {}^{\lfloor\psi}L_{ic} \end{bmatrix}$$

$$D_{ic} \equiv \operatorname{diag}\left[{}^{\lfloor\Delta}D_{ic}, {}^{\lfloor\psi}D_{ic}\right], \; {}^{\lfloor\Delta}D_{ic} \text{ is scalar, } \mathring{D}_{(i+1)c} = \mathring{D}_{ic} - 1.$$

The correction of the LDL' decomposition needs double, if $\chi(\cdot) = 1$, or single, if $\chi(\cdot) = 0$, rank-one updating (Bierman, 1977).

Proof: The matrix $\mathcal{K}$ extends ϕ_t to Ψ_t, thus we can define $\phi_t' L_0 D_0 L_0' \phi_t = \Psi_t' \mathcal{K}' L_\gamma D_\gamma L_\gamma' \mathcal{K} \Psi_t$. The transformed kernel serves as an initial condition in the following recursions. With the observable consequences split $\Delta_{g;t} = (\delta_{c;t}, \Delta_{gc+;t})$, $\omega_\gamma(c, a_{g;t}, \phi_{t-1}) \equiv$

$$
\begin{aligned}
&\equiv 2 \int f(\Delta_{g;t} | a_{g;t}, \phi_{t-1}) \times \\
&\times \ln\left(\frac{f(\Delta_{g;t}|a_{g;t}, d_g(t-1))}{\gamma(\phi_t) f(\Delta_{gc+;t}|a_{g;t}, d_g(t-1)) F(\delta_{c;t}|a_{g;t}, d_c(t-1), c)} \right) d\Delta_{g;t} = \\
&= \sum_{i=1}^{\mathring{\delta}_c} \ln\left(\frac{r_{ic}}{r_{ig}}\right) + \sum_{i=1}^{\mathring{\delta}_c} \int f(\Delta_{g;t}|a_{g;t}, \phi_{t-1}) \times \\
&\times \left[-\frac{\left(\Delta_{ig;t} - \theta_{ig}' \psi_{i;t}\right)^2}{r_{ig}} + \frac{(\Delta_{ig;t} - \theta_{ic}' \psi_{i;t})^2}{r_{ic}} - 2\ln(\gamma(\phi_t)) \right] d\Delta_{g;t} = \\
&\underbrace{=}_{\mathcal{E}[\bullet^2] = \mathcal{E}^2[\bullet] + \mathrm{cov}[\bullet]} \sum_{i=1}^{\mathring{\delta}_c} \ln\left(\frac{r_{ic}}{r_{ig}}\right) - \mathring{\delta}_c + \\
&+ \sum_{i=1}^{\mathring{\delta}_c} \int f(\Delta_{g;t}|a_{g;t}, \phi_{t-1}) \left(\frac{(\Delta_{ig;t} - \theta_{ic}' \psi_{i;t})^2}{r_{ic}} - 2\ln(\gamma(\phi_t)) \right) d\Delta_{g;t} = \\
&= -\mathring{\delta}_c + \sum_{i=1}^{\mathring{\delta}_c} \left[\ln\left(\frac{r_{ic}}{r_{ig}}\right) + \frac{r_{ig}}{r_{ic}} \right] + \\
&+ \sum_{i=1}^{\mathring{\delta}_c} \mathcal{E}\left[\psi_{i;t}' \left(\theta_{ig} - \theta_{ic}\right) r_{ic}^{-1} \left(\theta_{ig} - \theta_{ic}\right)' \psi_{i;t} | \psi_{\mathring{\Delta}_g;t} \right] - 2\mathcal{E}[\ln(\gamma(\phi_t)) | \psi_{\mathring{\Delta}_g;t}].
\end{aligned}
$$

Let us define kernel $L_{(i-1)c} D_{(i-1)c} L_{(i-1)c}'$ of the lifted quadratic form $\psi_{i-1;t}' L_{(i-1)c} D_{(i-1)c} L_{(i-1)c}' \psi_{i-1;t} + k_{(i-1)c}$ for which we evaluate $\mathcal{E}[k_{(i-1)c} + \psi_{i-1;t}' L_{(i-1)c} D_{(i-1)c} L_{(i-1)c}' \psi_{i-1;t} | \psi_{\mathring{\Delta};t}]$ needed in evaluation of the expectation of the quadratic form according to Proposition 4. Then, an intermediate lifted quadratic form arises $\tilde{k}_{ic} + \psi_{i;t}' \tilde{L}_{ic} \tilde{D}_{ic} \tilde{L}_{ic}' \psi_{i;t}$ with

$$
\begin{aligned}
\tilde{L}_{ic} \tilde{D}_{ic} \tilde{L}_{ic}' &= {}^{\lfloor\psi}L_{(i-1)c} \, {}^{\lfloor\psi}D_{(i-1)c} \, {}^{\lfloor\psi}L_{(i-1)c}' + \\
&+ \left(\theta_{ig} + {}^{\lfloor\Delta\psi}L_{(i-1)c}\right) {}^{\lfloor\Delta}D_{(i-1)c} \left(\theta_{ig} + {}^{\lfloor\Delta\psi}L_{(i-1)c}\right)' \\
\tilde{k}_{ic} &= k_{(i-1)c} + {}^{\lfloor\Delta}D_{(i-1)c} r_{ic}.
\end{aligned}
$$

While $i \leq \mathring{\delta}_c$, the values $\ln\left(\frac{r_{ic}}{r_{ig}}\right) + \frac{r_{ig}}{r_{ic}}$, and $(\theta_{ig} - \theta_{ic})\, r_{ic}^{-1}\, (\theta_{ig} - \theta_{ic})'$ have to be added to the lift and kernel, respectively,

$$
\begin{aligned}
& k_{ic} = k_{(i-1)c} + {}^{\lfloor\Delta}D_{(i-1)c} r_{ic} + \chi\left(i \leq \mathring{\delta}_c\right)\left[\ln\left(\frac{r_{ic}}{r_{ig}}\right) + \frac{r_{ig}}{r_{ic}}\right] \\
& L_{ic} D_{ic} L'_{ic} = {}^{\lfloor\psi}L_{(i-1)c}\, {}^{\lfloor\psi}D_{(i-1)c}\, {}^{\lfloor\psi}L'_{(i-1)c} + \\
& + \left(\theta_{ig} + {}^{\lfloor\Delta\psi}L_{(i-1)c}\right) {}^{\lfloor\Delta}D_{(i-1)c} \left(\theta_{ig} + {}^{\lfloor\Delta\psi}L_{(i-1)c}\right)' + \\
& + \chi\left(i \leq \mathring{\delta}_c\right) \frac{(\theta_{ig} - \theta_{ic})\,(\theta_{ig} - \theta_{ic})'}{r_{ic}}.
\end{aligned}
$$

The rank-one updating is to be used in evaluations. □

7.4 Proof of Proposition 3

Proof: By induction, we prove that $-2\ln(\gamma(d_g(t))) = k_t + \phi'_t L'_t D_t L_t \phi_t$ with data independent lift k_t and kernel $L'_t D_t L_t$. Proposition 2, implies that it holds for $t = \mathring{t}$ with $k_t = 0$, $L_t = I$, $D_t = 0$. Performing a generic step t, we both prove the claim and find the recursions for k_t and L_t, D_t. The identities (10), (14) imply that $2\omega_\gamma(a_{g;t}, d_g(t-1)) =$

$$
= \sum_{c \in c^*} \alpha_c \omega_\gamma(c, a_{g;t}, \phi_{t-1}) = \sum_{c \in c^*} \alpha_c [k_{\mathring{\Delta}_g c} + \psi'_{\mathring{\Delta}_g;t} L_{\mathring{\Delta}_g c} D_{\mathring{\Delta}_g c} L'_{\mathring{\Delta}_g c} \psi_{\mathring{\Delta}_g;t}]
$$

with recursively constructed lifts $k_{\mathring{\Delta}_g c}$ and kernels $L_{\mathring{\Delta}_g c} D_{\mathring{\Delta}_g c} L'_{\mathring{\Delta}_g c}$. The formula for the optimal strategy (10) implies that this term has to be increased by weighted sum of exponents of the citizen ideal pdfs on government decisions. It gives logarithms of the constant part of the normalizing factor $-\ln(2\pi) - \sum_{c \in c^*} \alpha_c \ln(r_{ic}) \equiv -\ln(2\pi) - \ln(r_i)$ and exponent $-\frac{1}{2}X$ with

$$
X \equiv \underbrace{\sum_{c \in c^*} \alpha_c k_{\mathring{\Delta}_g c}}_{k_{\mathring{\Delta}_g}} + \psi'_{\mathring{\Delta}_g;t} \underbrace{\sum_{c \in c^*} \alpha_c L_{\mathring{\Delta}_g c} D_{\mathring{\Delta}_g c} L'_{\mathring{\Delta}_g c}}_{L_{\mathring{\Delta}_g} D_{\mathring{\Delta}_g} L'_{\mathring{\Delta}_g}} \psi_{\mathring{\Delta}_g;t} +
$$

$$+\sum_{i=\mathring{\Delta}_g+1}^{\mathring{d}_g} \psi'_{i-1;t} \underbrace{\sum_{c\in c^*} \alpha_c[-1,\theta'_{ic}]' r_{ic}^{-1}[-1,\theta'_{ic}]}_{\bar{L}_i\bar{D}_i\bar{L}'_i} \psi_{i-1;t}.$$

It is quadratic form in entries $a_{ig:t}$. Thus, the resulting pdf is normal and the kernel part related to ϕ_{t-1} defines the $-\ln(\gamma(d(t-1)))$. Specifically, for $i = \mathring{\Delta}_g + 1, \ldots, \mathring{d}_g$ and $\tilde{L}_i\tilde{D}_i\tilde{L}'_i \equiv L_{i-1}D_{i-1}L'_{i-1} + \bar{L}_i\bar{D}_i\bar{L}'_i$, we get

$$\tilde{L}_i = \begin{bmatrix} 1 & \\ -\theta^o_{i;t} & L_i \end{bmatrix}, \; \tilde{D}_i = \mathrm{diag}\left[r^o_{i;t}, D_i\right], \; k_i = k_{i-1} + \ln\left(r^o_{i;t}/r_i\right).$$

The final values define the reproduced form of the function $-\ln(\gamma(\phi_{t-1}))$ with the lift $k_{t-1} = k_{\mathring{d}_g}$ and kernel $L_{t-1} = L_{\mathring{d}_g}$, $D_{t-1} = D_{\mathring{d}_g}$. □

References

Arrow K. 1995. *Social Choice and Individual Values.* New Haven: Yale University Press. 2nd. ed.

Bierman G. 1977. *Factorization Methods for Discrete Sequential Estimation.* Academic Press, New York.

Kárný M. 1996. Towards fully probabilistic control design. *Automatica*, 32(12):1719–1722.

Kárný M. 1998. Adaptive systems: Local approximators? In *Preprints of the IFAC Workshop on Adaptive Systems in Control and Signal Processing*, pages 129–134, Glasgow. IFAC.

Kárný M, Böhm J, Guy TV, Nedoma P. 2003. Mixture-based adaptive probabilistic control. *International Journal of Adaptive Control and Signal Processing*, 17(2):119–132.

Kárný M, Halousková A, Böhm J, Kulhavý R, Nedoma P. 1985. Design of linear quadratic adaptive control: Theory and algorithms for practice. *Kybernetika*, 21. Supplement to No. 3, 4 ,5, 6.

Kárný M, Kadlec J, Sutanto EL. 1998. Quasi-Bayes estimation applied to normal mixture. In Rojíček J, Valečková M, Kárný M, Warwick K. , editors, *Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing*, pages 77–82, Praha. ÚTIA AV ČR.

Kullback S, Leibler R. 1951. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–87.

Nurmi H. 2001. Resolving group choice paradoxes using probabilistic and fuzzy concepts. *Group Decision and Negotiation*, 10:177–198.

Peterka V. 1981. Bayesian system identification. In Eykhoff P. , editor, *Trends and Progress in System Identification*, pages 239–304. Pergamon Press, Oxford.

Ljung L. 1987. *System Identification: Theory for the User*. Prentice-Hall, London.

Acknowledgements: This work was supported by ESF, project TED, AV ČR S1075351 and GAČR 102/03/0049.