manuscript

Joint Estimation and Robustness Optimization

Taozeng Zhu
School of Management, University of Science and Technology of China, Hefei, China 230026, zhutozng@mail.ustc.edu.cn
Jingui Xie
School of Management, University of Science and Technology of China, Hefei, China 230026, xiej@ustc.edu.cn
Melvyn Sim
NUS Business School, National University of Singapore, Singapore 119245, dscsimm@nus.edu.sg

Many real-world optimization problems have input parameters estimated from data whose inherent imprecision can lead to fragile solutions that may impede desired objectives and/or render constraints infeasible. We propose a *joint estimation and robustness optimization* (JERO) framework to mitigate estimation uncertainty in optimization problems by seamlessly incorporating both the parameter estimation procedure and the optimization problem. Toward that end, we construct an uncertainty set that incorporates all of the data, where the size of the uncertainty set is based on how well the parameters would be estimated from that data when using a particular estimation procedure: regressions, the least absolute shrinkage and selection operator, and maximum likelihood estimation (among others). The JERO model maximizes the uncertainty set's size and so obtains solutions that—unlike those derived from models dedicated strictly to robust optimization—are immune to parameter perturbations that would violate constraints or lead to objective function values exceeding their desired levels. We describe several applications and provide explicit formulations of the JERO framework for a variety of estimation procedures. To solve the JERO models with exponential cones, we develop a second-order conic approximation that limits errors beyond an operating range; with this approach, we can use state-of-the-art SOCP solvers to solve even large-scale convex optimization problems. Finally, we apply the JERO model to a case study, thereby addressing a health insurance reimbursement problem with the aim of improving patient flow in the healthcare system while hedging against estimation errors.

Key words: Robustness optimization; robust optimization; parameter estimation; data-driven optimization

1. Introduction

Data are used to estimate the input parameters for many optimization problems. Because such data are seldom sufficient to estimate those parameters precisely, the solutions to optimization problems may be so fragile that they fail to serve their intended purposes or to satisfy relevant constraints. Ben-Tal et al. (2009) points out that, in linear optimization problems, a small perturbation of the parameters—that is, of no more that 0.1% from their nominal values—can render the concept of an "optimal" solution practically meaningless given the consequent potential for severe violations

Electronic copy available at: https://ssrn.com/abstract=3335889

of the problem's constraints. Hence, there is a clear need to consider a framework that mitigates the inherent inaccuracy of parameter estimation.

Robust optimization (RO), an approach that has witnessed an explosive growth in the past two decades, is well suited for this purpose. It was introduced by Soyster (1973) and popularized by Ben-Tal and Nemirovski (1998), El Ghaoui et al. (1998) and Bertsimas and Sim (2004). In classical robust optimization, the input parameters are not specified exactly but instead are characterized by a so-called uncertainty set. Simple uncertainty sets, such as the "ellipsoidal" version of Ben-Tal and Nemirovski (1998) and the "budgeted" version of Bertsimas and Sim (2004), are ubiquitous in robust optimization models thanks to their computational tractability. More recently, Bertsimas et al. (2018a) use data to design uncertainty sets for robust optimization that are based on statistical confidence intervals.

Another recent development is distributionally robust optimization (DRO), which enhances the modeling of uncertainty by defining an ambiguity set of probability distributions that are constrained by their moments and/or statistical distances. Van Parys et al. (2017) offer an elegant justification of DRO models. Scarf (1957) was the first to apply the moment-based approach to study a single-item newsvendor problem, which has now been extended to more general optimization frameworks (see e.g., Breton and El Hachem 1995, Shapiro and Kleywegt 2002, Delage and Ye 2010, Wiesemann et al. 2014, Bertsimas et al. 2018b). There is also considerable interest in DRO models based on statistical distance, which—much as in our proposed *joint estimation and robustness optimization* (JERO) model—incorporate data in the ambiguity set. For discrete distributions, Ben-Tal et al. (2013) study robust optimization problems with uncertain probabilities defined by ϕ-divergences (see also Wang et al. 2016). For continuous distributions, the use of statistical distance based on the Wasserstein metric has been popular in DRO models (see, e.g., Esfahani and Kuhn 2018, Zhao and Guan 2018).

The JERO framework developed here incorporates both the parameter estimation procedure and the optimization problem. The uncertainty set in this model is based on how well the parameters are estimated from data via particular estimation procedures; examples include, *inter alia*, regressions, the least absolute shrinkage and selection operator (LASSO), and maximum likelihood estimation (MLE). In contrast to models of robust optimization, JERO's robustness optimization maximizes the size of the uncertainty set in order to derive solutions that are far less subject to estimation errors. So unlike classical robust optimization techniques, our procedure does not require stipulating the exact size of the uncertainty set; such exactness can be elusive, even with the help of performance bounds, because the bounds are often weak and/or reliant on assumptions that need not apply to the estimation problem at hand (see, for instance, Bertsimas and Sim 2004).

Electronic copy available at: https://ssrn.com/abstract=3335889

Our JERO framework is derived from both RO and DRO models.Yet it differs from those models by specifically, and crucially, incorporating the parameter estimation procedure.

Our paper's contributions can be summarized as follows.

1. We propose a *joint estimation and robustness optimization* framework, which seamlessly incorporates both the parameter estimation procedure and the optimization problem in the same model. The uncertainty set used in this JERO framework is based on how well the parameters are estimated—from the available data—by various estimation procedures. Given that estimation, the JERO model maximizes the uncertainty set's size so as to obtain solutions that are relatively immune to estimation errors.
2. We detail a number of applications that involve explicit JERO formulations for several different standard estimation procedures.
3. We use a real-world data set in adapting the JERO framework to address a health insurance reimbursement problem. The goal of this case-study exercise is to improve patient flow in a healthcare system while hedging against estimation errors.
4. To address the computational issues that arise in some of the JERO applications, we exploit practical second-order conic (SOC) approximations of exponential cones and thereby reduce the approximation errors associated with extreme values. Hence we can use state-of-the-art second-order conic programming (SOCP) solvers (e.g., CPLEX, Gurobi) and find solutions even to large-scale convex optimization problems. As our SOC approximation of an exponential cone is of independent interest, we present our approach in Appendix B.

The rest of our paper proceeds as follows. Section 2 develops our framework for *joint estimation and robustness optimization*. In Section 3, we describe some applications and related formulations of the JERO framework for a variety of estimation procedures. Section 4 presents a case study, based on real data, of improving patient flow in a healthcare system. We conclude in Section 5 with a brief summary. The proofs of all formal propositions are given in the Appendix.

Notation. We use boldface lowercase letters, such as $\boldsymbol{x} \in \mathbb{R}^N$, to represent vectors; we use x_i to denote the ith element of the vector $\boldsymbol{x}$. In addition, boldface uppercase letters—such as $\boldsymbol{A} \in \mathbb{R}^{M\times N}$—are used to denote matrices while $\boldsymbol{A}_i$ denotes the ith row of the matrix $\boldsymbol{A}$. Special vectors include $\boldsymbol{0}$, $\boldsymbol{e}$, and $\boldsymbol{e}_i$; these represent (respectively) the vector of 0s, the vector of 1s, and the standard unit basis vector. We denote by $[N]$ the set of positive running indices up to N; thus, $[N] = \{1, 2, \cdots, N\}$. We use $\mathbb{S}_+^M$ to denote the set of symmetric positive semidefinite $M \times M$ matrices, and use $\mathbb{S}_{++}^M$ to denote the set of symmetric positive definite $M \times M$ matrices. Given a set $\mathcal{S}$, we denote its relative interior by $\mathrm{ri}(\mathcal{S})$. Finally, we follow the convention that $0\ln(0/w) = 0$ if $w \geq 0$.

Electronic copy available at: https://ssrn.com/abstract=3335889

2. Joint Estimation and Robustness Optimization Model

We consider the following optimization problem:

$$\begin{aligned} Z = \min\ & a_0(\boldsymbol{x}, \hat{\boldsymbol{\beta}}) \\ \text{s.t. } & a_i(\boldsymbol{x}, \hat{\boldsymbol{\beta}}) \le \tau_i \ \forall i \in [I], \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{1}$$

Here $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^N$ is the decision variable, $\hat{\boldsymbol{\beta}} \in \mathcal{W} \subseteq \mathbb{R}^M$ is the input parameter, and the $a_i \colon \bar{\mathcal{X}} \times \bar{\mathcal{W}} \mapsto \mathbb{R}$ are *saddle functions* on the domain $\bar{\mathcal{X}} \times \bar{\mathcal{W}} \subseteq \mathbb{R}^N \times \mathbb{R}^M$ (note $\mathcal{X} \subseteq \bar{\mathcal{X}}, \mathcal{W} \subseteq \bar{\mathcal{W}}$), *i.e.*, $a_i(\boldsymbol{x}, \boldsymbol{\beta})$ being concave in $\boldsymbol{\beta}$ given $\boldsymbol{x} \in \bar{\mathcal{X}}$ and convex in $\boldsymbol{x}$ given $\boldsymbol{\beta} \in \bar{\mathcal{W}}$.

The optimization problem's input parameter, $\hat{\boldsymbol{\beta}}$, is determined from data via an estimation procedure. Conceivably, when the estimated parameter differs from the true value, the objective value of problem (1) may deviate and some of the constraints may become infeasible. To address this issue, classical robust optimization solves the following semi-infinite optimization problem:

$$\begin{aligned} Z_R(r) = \min\ & \tau \\ \text{s.t. } & a_0(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau \ \ \forall \boldsymbol{\beta} \in \mathcal{U}(r), \\ & a_i(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau_i \ \forall \boldsymbol{\beta} \in \mathcal{U}(r), \forall i \in [I], \\ & \boldsymbol{x} \in \mathcal{X}; \end{aligned} \tag{2}$$

here $\mathcal{U}(r)$ is the *uncertainty set*, which is usually a normed ball of radius r centered at the estimate $\hat{\boldsymbol{\beta}}$, defined by

$$\mathcal{U}(r) \triangleq \{\boldsymbol{\beta} \in \bar{\mathcal{W}} \mid \|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\| \le r\}.$$

One challenge faced by the robust optimization model is to determine the size of the uncertainty set, r, which is typically difficult for the decision maker to specify. Although that size has some connection to probability constraints, the bounds are usually loose and strongly dependent on the imposed distributional assumptions and the functions a_i. We therefore propose the following *robustness optimization* model:

$$\begin{aligned} Z_S(\tau_0) = \max\ & r \\ \text{s.t. } & a_0(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau_0 \ \ \forall \boldsymbol{\beta} \in \mathcal{U}(r), \\ & a_i(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau_i \ \ \forall \boldsymbol{\beta} \in \mathcal{U}(r), \forall i \in [I], \\ & \boldsymbol{x} \in \mathcal{X}, \ r \ge 0. \end{aligned} \tag{3}$$

In this robustness optimization, the objective function is the size of the uncertainty set and the goal is to find a solution $\boldsymbol{x}$ that remains feasible for the largest possible uncertainty set. Observe that we must specify the cost target τ_0 in problem (3), a task that is more intuitive than specifying the uncertainty set's size r in problem (2). For instance, we may specify the the cost target τ_0 relative

Electronic copy available at: https://ssrn.com/abstract=3335889

to the optimum cost of problem (1), Z^*. Since the uncertainty set $\mathcal{U}(r)$ is nondecreasing in r, it follows that we can obtain the solution to the robustness optimization problem (3) by solving—via binary search—a sequence of robust optimization problems (2) until $r = r^*$ such that $Z_R(r^*) = \tau_0$.

Given a data set $\mathcal{D}$, we obtain the parameter $\hat{\boldsymbol{\beta}}$ by using an estimation procedure that solves the following optimization problem:

$$\begin{aligned} \hat{\boldsymbol{\beta}} = \arg\min\ & \rho\left(\boldsymbol{\beta}; \mathcal{D}\right) \\ \text{s.t. } & \boldsymbol{\beta} \in \mathcal{W}; \end{aligned} \tag{4}$$

here $\rho(\boldsymbol{\beta}; \mathcal{D})$ is referred to as the *estimation metric*. In ordinary least-squares (OLS) estimation, for example, ρ is the sum of the squares of prediction errors.

Definition 1 (Estimate Uncertainty Set). We define the uncertainty set of an estimate as

$$\mathcal{E}(r; \mathcal{D}) \triangleq \{\boldsymbol{\beta} \in \mathcal{W} \mid \rho(\boldsymbol{\beta}; \mathcal{D}) \leq \hat{\rho} + r\}, \tag{5}$$

where $\hat{\rho} = \rho(\hat{\boldsymbol{\beta}}; \mathcal{D})$ is the minimal value of the estimation metric and $r \geq 0$ is the gap from the optimal value $\hat{\rho}$.

Thus $\mathcal{E}(r; \mathcal{D})$ denotes the set of estimates that are feasible within the optimality gap, where $\mathcal{E}(0; \mathcal{D}) = \{\hat{\boldsymbol{\beta}}\}$. The corresponding *joint estimation and robustness optimization* (JERO) model is given by

$$\begin{aligned} Z_E(\tau_0) = \max\ & r \\ \text{s.t. } & a_0(\boldsymbol{x}, \boldsymbol{\beta}) \leq \tau_0 \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}) \\ & a_i(\boldsymbol{x}, \boldsymbol{\beta}) \leq \tau_i \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}), \forall i \in [I] \\ & \boldsymbol{x} \in \mathcal{X},\ r \geq 0. \end{aligned} \tag{6}$$

This model incorporates the estimation procedure and the optimization problem in a seamless framework.

For a fixed $r \geq 0$, the constraint in model (6) is convex in the decision variable $\boldsymbol{x}$—althougth, the model might not be jointly convex in r and $\boldsymbol{x}$. Nevertheless, we can obtain the optimal solution by solving a sequence of subproblems as follows:

$$\begin{aligned} Z_E^r(\tau_0) = \min\ & t - \tau_0 \\ \text{s.t. } & a_0(\boldsymbol{x}, \boldsymbol{\beta}) \leq t \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}), \\ & a_i(\boldsymbol{x}, \boldsymbol{\beta}) \leq \tau_i \ \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}), \forall i \in [I], \\ & \boldsymbol{x} \in \mathcal{X}; \end{aligned} \tag{7}$$

here $\boldsymbol{\beta} \in \mathcal{W}$ is the input parameter and $\mathcal{E}(r; \mathcal{D})$ is the uncertainty set of estimates. Assuming that there exists a method for finding the optimal solution of model (7), we propose the following binary search to obtain the optimal solution for model (6).

Electronic copy available at: https://ssrn.com/abstract=3335889

Algorithm 1 Binary Search

Input A routine that solves model (7) optimally, $\Delta > 0$, and $\bar{r}$ is a large enough positive number.

Output: $\boldsymbol{x}$

Step 1. Set $r_1 = 0$ and $r_2 = \bar{r}$.

Step 2. If $r_2 - r_1 \leq \Delta$, stop. Output: $\boldsymbol{x}$.

Step 3. Let $r := (r_1 + r_2)/2$. Compute $Z_E^r(\tau_0)$ from model (7) and obtain the corresponding optimal solution.

Step 4. If $Z_E^r(\tau_0) \leq 0$, update $r_1 := r$; otherwise, update $r_2 := r$.

Step 5. Go to Step 2.

PROPOSITION 1. *Suppose that model* (6) *is feasible. Then Algorithm 1 finds a solution* $\boldsymbol{x}$ *with objective* $\tilde{r}$ *satisfying* $|\tilde{r} - r^*| < \Delta$ *in at most* $\lceil \log_2(\bar{r}/\Delta) \rceil$ *computations of the subproblem* (7). *Here* r^* *is the optimal objective of model* (6), $\Delta > 0$ *is an arbitrary number, and* $\bar{r} > 0$ *is a sufficiently large number.*

The tractability of JERO depends on a tractable representation of what is known as the *robust counterpart*. We focus here on JERO's nonlinear constraints,

$$a_i(\boldsymbol{x}, \boldsymbol{\beta}) \leq \tau_i, \quad i \in [I] \cup \{0\}, \tag{8}$$

where $\boldsymbol{x} \in \bar{\mathcal{X}}$ is the decision variable and $\boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D})$ is the uncertain estimate. The robust counterpart of (8) is then

$$a_i(\boldsymbol{x}, \boldsymbol{\beta}) \leq \tau_i \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}) \tag{9}$$

or, equivalently,

$$\max_{\boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D})} a_i(\boldsymbol{x}, \boldsymbol{\beta}) \leq \tau_i.$$

Deriving the robust counterpart requires our next proposition.

PROPOSITION 2. *Assume that* $\mathrm{ri}\,(\mathcal{E}(r; \mathcal{D})) \neq \emptyset$ *and that* $a_i(\boldsymbol{x}, \cdot)$ *is closed concave for all* $\boldsymbol{x} \in \bar{\mathcal{X}}$. *Then* $\boldsymbol{x}$ *satisfies* (9) *if and only if* $\boldsymbol{x}$ *and* $\boldsymbol{\nu}$ ($\boldsymbol{x} \in \bar{\mathcal{X}}$, $\boldsymbol{\nu} \in \bar{\mathcal{W}}$) *satisfy*

$$\delta_r^*(\boldsymbol{\nu}) - a_i^*(\boldsymbol{x}, \boldsymbol{\nu}) \leq \tau_i. \tag{10}$$

Here $a_i^*(\boldsymbol{x}, \boldsymbol{\nu})$ *is the concave conjugate of* $a_i(\boldsymbol{x}, \boldsymbol{\nu})$, *defined as*

$$a_i^*(\boldsymbol{x}, \boldsymbol{\nu}) \triangleq \inf_{\boldsymbol{\beta} \in \bar{\mathcal{W}}} \{\boldsymbol{\beta}'\boldsymbol{\nu} - a_i(\boldsymbol{x}, \boldsymbol{\beta})\};$$

and $\delta_r^*(\boldsymbol{\nu})$ *is the support function of the set* $\mathcal{E}(r; \mathcal{D})$, *defined as*

$$\delta_r^*(\boldsymbol{\nu}) \triangleq \sup_{\boldsymbol{\beta} \in \mathcal{W}} \{\boldsymbol{\beta}'\boldsymbol{\nu} \mid \rho(\boldsymbol{\beta}; \mathcal{D}) \leq \hat{\rho} + r\}.$$

Electronic copy available at: https://ssrn.com/abstract=3335889

More interestingly, we show in the next result that the support function of $\mathcal{E}(r;\mathcal{D})$ can be obtained by computing the conjugate function of the corresponding estimation metric.

PROPOSITION 3. *Assume that $r>0$ and that $\rho(\boldsymbol{\beta};\mathcal{D})$ is convex in $\boldsymbol{\beta}$. Then the support function of $\mathcal{E}(r;\mathcal{D})$ can be represented as*

$$\delta_r^*(\boldsymbol{\nu}) = \inf_{\mu>0}\{(\hat{\rho}+r)\mu + \mu\rho^*(\boldsymbol{\nu}/\mu;\mathcal{D})\}; \tag{11}$$

here $\rho^*(\boldsymbol{\nu};\mathcal{D})$ is the convex conjugate of the estimation metric $\rho(\boldsymbol{\beta};\mathcal{D})$ and is defined as

$$\rho^*(\boldsymbol{\nu};\mathcal{D}) \triangleq \sup_{\boldsymbol{\beta}\in\bar{\mathcal{W}}}\{\boldsymbol{\beta}'\boldsymbol{\nu} - \rho(\boldsymbol{\beta};\mathcal{D})\}.$$

In the next section, we focus on the estimation metrics and support functions of estimate sets $\mathcal{E}(r;\mathcal{D})$. We refer to Ben-Tal et al. (2015), and the references therein, for computing conjugate functions.

3. Estimation Procedures, Explicit Formulations and Applications

The tractability of the robust counterpart (9) depends on the chosen estimation metric. We now describe several applications with explicit formulations of the JERO framework for a variety of estimation procedures.

3.1. Regression-based or Least squares estimation metric

OLS. Suppose that data $\mathcal{D}=\{\boldsymbol{z},\boldsymbol{Y}\}$ are given, where $\boldsymbol{z}\in\mathbb{R}^P$ is a vector of P observations of the response variable and where $\boldsymbol{Y}=\{\mathbf{1},\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{M-1}\}\in\mathbb{R}^{P\times M}$ is a full column rank matrix of observations of $M-1$ dependent variables. Then ordinary least squares (OLS) solves the following optimization problem:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^M} \|\boldsymbol{z}-\boldsymbol{Y}\boldsymbol{\beta}\|_2^2,$$

where $\boldsymbol{\beta}\in\mathbb{R}^M$ is a vector of regression coefficients. Hence the estimation metric is

$$\rho(\boldsymbol{\beta};\mathcal{D}) = \|\boldsymbol{z}-\boldsymbol{Y}\boldsymbol{\beta}\|_2^2. \tag{12}$$

It is known that the OLS estimate is $\hat{\boldsymbol{\beta}}=(\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{z}$ and $\hat{\rho}=\|\boldsymbol{z}-\boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{z}\|_2^2$.

PROPOSITION 4. *Let $r>0$. Then the corresponding support function of the estimate uncertainty set $\mathcal{E}(r;\mathcal{D})$, given the estimation metric* (12)*, is*

$$\begin{aligned} \delta_r^*(\boldsymbol{\nu}) = \min\ & \sqrt{\hat{\rho}+r}\|\boldsymbol{w}\|_2 + \boldsymbol{z}'\boldsymbol{w} \\ s.t.\ & \boldsymbol{Y}'\boldsymbol{w} = \boldsymbol{\nu}, \\ & \boldsymbol{w}\in\mathbb{R}^P. \end{aligned} \tag{13}$$

Electronic copy available at: https://ssrn.com/abstract=3335889

LASSO. Given the same data $\mathcal{D}=\{\boldsymbol{z},\boldsymbol{Y}\}$ as before, the least absolute shrinkage and selection operator (LASSO) introduced by Tibshirani (1996) solves the l_1 penalized regression problem

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^M}\left\{\frac{1}{P}\|\boldsymbol{z}-\boldsymbol{Y}\boldsymbol{\beta}\|_2^2+\lambda\|\boldsymbol{\beta}\|_1\right\},$$

where $\boldsymbol{\beta}\in\mathbb{R}^M$ is a vector of regression coefficients and $\lambda\in\mathbb{R}_+$ is a prespecified parameter. The corresponding estimation metric is

$$\rho(\boldsymbol{\beta};\mathcal{D})=\frac{1}{P}\|\boldsymbol{z}-\boldsymbol{Y}\boldsymbol{\beta}\|_2^2+\lambda\|\boldsymbol{\beta}\|_1. \tag{14}$$

Note that LASSO reduces to OLS when $\lambda=0$.

Proposition 5. *Let $r>0$. Then the corresponding support function of the estimate uncertainty set $\mathcal{E}(r;\mathcal{D})$, given the estimation metric* (14)*, is*

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu})=\min\ & (\hat{\rho}+r)s+u-v+\boldsymbol{z}'\boldsymbol{w}\\ s.t.\ & \boldsymbol{Y}'\boldsymbol{w}-\boldsymbol{t}=\boldsymbol{\nu},\\ & \left\|\begin{matrix}2v\\ \boldsymbol{w}\end{matrix}\right\|_2\le 2u,\\ & \|\boldsymbol{t}\|_\infty\le\lambda s,\\ & Pu+Pv\le s,\\ & u,s\in\mathbb{R}_+,\ v\in\mathbb{R},\ \boldsymbol{w}\in\mathbb{R}^P,\ \boldsymbol{t}\in\mathbb{R}^M;\end{aligned} \tag{15}$$

here $\hat{\rho}$ is the optimal value of the corresponding estimation metric.

Example 1 (Store Location). Consider the problem of setting up stores in N different areas. There is a fixed cost c_i of putting a store in area i. Suppose that data $\mathcal{D}=\{\boldsymbol{z}_t,\boldsymbol{Y}_t\}_{t=1}^P$ are given, where $\boldsymbol{z}_t\in\mathbb{R}^N$ is a vector of demand observations for N areas in period t and where $\boldsymbol{Y}_t=\{\boldsymbol{1},\boldsymbol{y}_1^{(t)},\ldots,\boldsymbol{y}_{M-1}^{(t)}\}\in\mathbb{R}^{N\times M}$ is a matrix of observations in period t for factors that may affect demand. We can apply the following JERO framework to solve this store location problem:

$$\begin{aligned}\max\ & r\\ \text{s.t. }& \textstyle\sum_{i=1}^N\boldsymbol{\beta}'\boldsymbol{f}_i x_i\ge\tau\ \ \forall\boldsymbol{\beta}\in\mathcal{E}(r;\mathcal{D}),\\ & x_i+x_j\le 1\qquad\forall(i,j)\in\mathcal{A},\\ & \textstyle\sum_{i=1}^N c_i x_i\le B,\\ & \boldsymbol{x}\in\{0,1\}^N.\end{aligned}$$

In this problem, $\boldsymbol{x}$ is the decision variable, τ is the target for service level, $\boldsymbol{\beta}\in\mathbb{R}^M$ is a vector of regression coefficients, and $\mathcal{A}$ is a set of arcs denoting two locations with at most one store. In addition, B is the total budget available for opening stores and $\boldsymbol{f}_i\in\mathbb{R}^M$ is a vector of values for

Electronic copy available at: https://ssrn.com/abstract=3335889

factors at current that may influence demand in area i. Using the estimation metric (14), we can reformulate this optimization problem as follows::

$$
\begin{aligned}
\max\ & r \\
\text{s.t. } & (\hat{\rho}+r)s + u - v + \textstyle\sum_{i=1}^{N} \boldsymbol{z}_i' \boldsymbol{w}_i + \tau \le 0, \\
& \textstyle\sum_{i=1}^{P} \boldsymbol{Y}_i' \boldsymbol{w}_i + \sum_{i=1}^{N} \boldsymbol{f}_i x_i - \boldsymbol{t} = \boldsymbol{0}, \\
& \left\| \begin{array}{c} 2v \\ \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_P \end{array} \right\|_2 \le 2u, \\
& \|\boldsymbol{t}\|_\infty \le \lambda s, \\
& PNu + PNv \le s, \\
& x_i + x_j \le 1 && \forall (i,j) \in \mathcal{A}, \\
& \textstyle\sum_{i=1}^{N} c_i x_i \le B, \\
& \boldsymbol{x} \in \{0,1\}^N, \\
& u, s \in \mathbb{R}_+,\ v \in \mathbb{R},\ \boldsymbol{t} \in \mathbb{R}^M, \\
& \boldsymbol{w}_i \in \mathbb{R}^N && \forall i \in [P];
\end{aligned}
$$

here $\hat{\rho}$ is the optimal value of the corresponding estimation metric. Observe that the optimization problem is a mixed-integer SOCP, which can be addressed by state-of-the-art solvers such as CPLEX and Gurobi.

3.2. Maximum Likelihood Estimation Metric

Given the data $\mathcal{D}$, which consists of P independent and identically distributed (i.i.d.) observations $\boldsymbol{z}_i$ and a family of density functions $f(\cdot \,|\, \boldsymbol{\beta})$, the average log-likelihood function is defined as

$$\ell(\boldsymbol{\beta}; \mathcal{D}) \triangleq \frac{1}{P} \ln \left(\prod_{i=1}^{P} f(\boldsymbol{z}_i \,|\, \boldsymbol{\beta}) \right).$$

We can now derive different estimate uncertainty sets (and their support functions) under various distributions, as described next.

Multivariate Normal. Suppose we are given data $\mathcal{D} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_P\}$ with $\boldsymbol{z}_i \in \mathbb{R}^M$. Let these data be characterized by a multivariate normal distribution with mean $\boldsymbol{\beta}$ and covariance $\boldsymbol{\Sigma}$, and let the sample covariance matrix be positive definite. Then the average log-likelihood function can be written as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}; \mathcal{D}) = -\frac{1}{2} \left(M \ln(2\pi) + \ln \det \boldsymbol{\Sigma} + \frac{1}{P} \sum_{i=1}^{P} (\boldsymbol{z}_i - \boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\beta}) \right),$$

Electronic copy available at: https://ssrn.com/abstract=3335889

which is not a jointly concave function with respect to the parameters $\boldsymbol{\beta} \in \mathbb{R}^M$ and $\boldsymbol{\Sigma} \in \mathbb{S}_+^M$. Suppose, we are keen in estimating the mean, $\boldsymbol{\beta}$, then the corresponding estimation metric would be

$$\rho(\boldsymbol{\beta}; \mathcal{D}) = \min_{\boldsymbol{\Sigma} \in \mathbb{S}_+^M} \frac{1}{2} \left(\ln \det \boldsymbol{\Sigma} + \frac{1}{P} \sum_{i=1}^{P} (\boldsymbol{z}_i - \boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\beta}) \right).$$

It is known that the maximum likelihood estimates are

$$\hat{\boldsymbol{\beta}} = \frac{1}{P} \sum_{i=1}^{P} \boldsymbol{z}_i, \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{P} \sum_{i=1}^{P} \left(\boldsymbol{z}_i - \hat{\boldsymbol{\beta}} \right) \left(\boldsymbol{z}_i - \hat{\boldsymbol{\beta}} \right)',$$

and $\hat{\rho} = \frac{1}{2}(\det \hat{\boldsymbol{\Sigma}} + M)$.

PROPOSITION 6. *The corresponding estimation metric $\rho(\boldsymbol{\beta}; \mathcal{D})$ is equivalent to*

$$\rho(\boldsymbol{\beta}; \mathcal{D}) = \frac{1}{2} \big(M + \ln \det \hat{\boldsymbol{\Sigma}} + \ln(1 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) \big), \tag{16}$$

and the corresponding support function of the estimate uncertainty set $\mathcal{E}(r; \mathcal{D})$, given the estimation metric (16)*, is*

$$\delta_r^*(\boldsymbol{\nu}) = \hat{\boldsymbol{\beta}}' \boldsymbol{\nu} + \sqrt{e^{2r} - 1} \| \hat{\boldsymbol{\Sigma}}^{\frac{1}{2}} \boldsymbol{\nu} \|_2. \tag{17}$$

EXAMPLE 2 (ROBUST PORTFOLIO OPTIMIZATION). Consider an instance of portfolio optimization with N assets or stocks held over a period of time. Suppose we have the data $\mathcal{D} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_P\}$, where $\boldsymbol{z}_i \in \mathbb{R}^N$ is an observation of the returns on N assets. To derive a robust solution that will meet a prespecified target even in the worst-case scenario, we can solve the JERO problem

$$\begin{aligned} \max \; & r \\ \text{s.t. } & \boldsymbol{\beta}' \boldsymbol{x} \geq \tau \qquad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}), \\ & \boldsymbol{x}' \boldsymbol{e} = 1, \\ & \boldsymbol{x} \in \mathbb{R}_+^N, \, r \in \mathbb{R}_+, \end{aligned}$$

where $\boldsymbol{x} \in \mathbb{R}_+^N$ is the investment portfolio, τ is the target for portfolio, and $\boldsymbol{\beta} \in \mathbb{R}^N$ is the estimate (i.e., the mean value) of the portfolio. If we use the estimation metric (16), then the optimization problem would be equivalent to

$$\begin{aligned} \max \; & r \\ \text{s.t. } & -\hat{\boldsymbol{\beta}}' \boldsymbol{x} + \sqrt{e^{2r} - 1} \| \hat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{x} \|_2 + \tau \leq 0, \\ & \boldsymbol{x}' \boldsymbol{e} = 1, \\ & \boldsymbol{x} \in \mathbb{R}_+^N, r \in \mathbb{R}_+, \end{aligned}$$

where $\hat{\boldsymbol{\beta}} = \frac{1}{P} \sum_{i=1}^{P} \boldsymbol{z}_i$ and $\hat{\boldsymbol{\Sigma}} = \frac{1}{P} \sum_{i=1}^{P} \left(\boldsymbol{z}_i - \hat{\boldsymbol{\beta}} \right) \left(\boldsymbol{z}_i - \hat{\boldsymbol{\beta}} \right)'$. Note that

$$-\hat{\boldsymbol{\beta}}' \boldsymbol{x} + \sqrt{e^{2r} - 1} \| \hat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{x} \|_2 + \tau \leq 0$$

Electronic copy available at: https://ssrn.com/abstract=3335889

can be rewritten as

$$S(\boldsymbol{x}) \triangleq \frac{\hat{\boldsymbol{\beta}}'\boldsymbol{x} - \tau}{\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{x}\|_2} \geq \sqrt{e^{2r} - 1},$$

where $S(\boldsymbol{x})$ is the Sharpe ratio (introduced by Sharpe 1966). The JERO here is equivalent to maximizing the Sharpe ratio of the portfolio selection with respect to $N(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ (Sharpe 1994).

Finite Support. Suppose we are given data $\mathcal{D} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_P\}$ with $\boldsymbol{z}_i \in \mathbb{R}^N$. Let the underlying distribution of these data be the finite support distribution on support Ξ with probability distribution $\boldsymbol{\beta} \in \mathbb{R}_+^M$ in a probability simplex, where $\Xi \triangleq \{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_M\}$ and $\boldsymbol{\xi}_i \in \mathbb{R}^N$. Then the average log-likelihood function is given by

$$\ell(\boldsymbol{\beta}; \mathcal{D}) = \frac{1}{P} \sum_{i=1}^{M} P_i \ln(\beta_i),$$

where $P_i = \sum_{j=1}^{P} \mathbb{I}(\boldsymbol{z}_j = \boldsymbol{\xi}_i)$ for $i \in [M]$. The corresponding estimation metric would be

$$\rho(\boldsymbol{\beta}; \mathcal{D}) = -\frac{1}{P} \sum_{i=1}^{M} P_i \ln(\beta_i). \tag{18}$$

It is known that the maximum likelihood estimates are $\hat{\beta}_i = P_i/P$, $i \in [M]$, from which it follows that the optimal value of the estimation metric is $\hat{\rho} = \frac{1}{P} \sum_{i=1}^{M} P_i \ln(P/P_i)$. Note that the estimation metric here is similar to the Burg entropy divergence (see Ben-Tal et al. 2013).

PROPOSITION 7. *Let $r > 0$. Then the corresponding support function of the estimate uncertainty set $\mathcal{E}(r; \mathcal{D})$, given the estimation metric* (18)*, is*

$$\begin{aligned} \delta_r^*(\boldsymbol{\nu}) = \min\ & (\hat{\rho} + r)Pu + v + \boldsymbol{e}'\boldsymbol{s} \\ s.t.\ & v - w_i \geq \nu_i && \forall i \in [M], \\ & P_i u \ln(P_i u / w_i) - P_i u - s_i \leq 0 && \forall i \in [M], \\ & u \in \mathbb{R}_+,\ v \in \mathbb{R},\ \boldsymbol{w} \in \mathbb{R}_+^M,\ \boldsymbol{s} \in \mathbb{R}^M. \end{aligned} \tag{19}$$

EXAMPLE 3 (TWO-STAGE STOCHASTIC PROGRAMMING). Consider an instance of two-stage stochastic programming (cf. Shapiro et al. 2009). The first-stage, "here and now" decision is $\boldsymbol{x} \in \mathbb{R}^{N_1}$, which is chosen over the feasible set $\mathcal{X} \subseteq \mathbb{R}^{N_1}$. The cost incurred during the first stage is deterministic and given by $\boldsymbol{c}'\boldsymbol{x}$ with $\boldsymbol{c} \in \mathbb{R}^{N_1}$. In the second stage, the random variable with support $\Xi = \{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_M\}$ is realized, where $\boldsymbol{\xi}_i = (\boldsymbol{d}^{(i)}, \boldsymbol{A}^{(i)}, \boldsymbol{B}^{(i)}, \boldsymbol{b}^{(i)})$, $\boldsymbol{d}^{(i)} \in \mathbb{R}^{N_2}$, $\boldsymbol{A}^{(i)} \in \mathbb{R}^{I \times N_1}$, $\boldsymbol{B}^{(i)} \in \mathbb{R}^{I \times N_2}$, and $\boldsymbol{b}^{(i)} \in \mathbb{R}^I$. Given the realization of that variable, we can determine the cost incurred in the second stage. For a decision $\boldsymbol{x}$ and a realization $\boldsymbol{\xi} = (\boldsymbol{d}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{b})$ of the random variable, we assess the second-stage costs via the optimization problem

$$\begin{aligned} \min\ & \boldsymbol{d}'\boldsymbol{y} \\ \text{s.t.}\ & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}, \\ & \boldsymbol{y} \in \mathbb{R}^{N_2}. \end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=3335889

Suppose we are given data $\mathcal{D} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_P\}$ that amounts to second-stage realizations of the random variable. Then we can solve the following JERO problem:

$$\begin{array}{ll} \max & r \\ \text{s.t.} & \boldsymbol{c}'\boldsymbol{x} + \sum_{i=1}^{M} \boldsymbol{y}_i' \boldsymbol{d}^{(i)} \beta_i \leq \tau \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}), \\ & \boldsymbol{A}^{(i)}\boldsymbol{x} + \boldsymbol{B}^{(i)}\boldsymbol{y}_i \geq \boldsymbol{b}^{(i)} \qquad \forall i \in [M], \\ & \boldsymbol{y}_i \in \mathbb{R}^{N_2} \qquad\qquad\qquad\quad \forall i \in [M], \\ & \boldsymbol{x} \in \mathcal{X},\, r \in \mathbb{R}_+; \end{array}$$

here τ is the cost target and $\boldsymbol{\beta} \in \mathbb{R}_+^M$ is the probability distribution over Ξ. If we use the estimation metric (18), then this optimization problem would be equivalent to

$$\begin{array}{lll} \max & r & \\ \text{s.t.} & \boldsymbol{c}'\boldsymbol{x} + (\hat{\rho} + r) P u + v + \boldsymbol{e}'\boldsymbol{s} \leq \tau, & \\ & v - w_i \geq \boldsymbol{y}_i' \boldsymbol{d}^{(i)} & \forall i \in [M], \\ & P_i u \ln(P_i u / w_i) - P_i u - s_i \leq 0 & \forall i \in [M], \\ & \boldsymbol{A}^{(i)}\boldsymbol{x} + \boldsymbol{B}^{(i)}\boldsymbol{y}_i \geq \boldsymbol{b}^{(i)} & \forall i \in [M], \\ & \boldsymbol{y}_i \in \mathbb{R}^{N_2} & \forall i \in [M], \\ & \boldsymbol{x} \in \mathcal{X},\, u, r \in \mathbb{R}_+,\, v \in \mathbb{R},\, \boldsymbol{w} \in \mathbb{R}_+^M,\, \boldsymbol{s} \in \mathbb{R}^M, & \end{array}$$

where $\hat{\rho}$ is the optimal value of the corresponding estimation metric.

Linear Regression. Suppose we have the data $\mathcal{D} = \{\boldsymbol{z}, \boldsymbol{Y}\}$, where $\boldsymbol{z} \in \mathbb{R}^P$ is a vector of observations of the response variable and $\boldsymbol{Y} = \{\boldsymbol{1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{M-1}\} \in \mathbb{R}^{P \times M}$ is a full column rank matrix of observations of dependent variables. We consider the linear regression model

$$\boldsymbol{z} = \boldsymbol{Y}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{20}$$

where $\boldsymbol{\beta} \in \mathbb{R}^M$ is a vector of regression coefficients and the ϵ_i are i.i.d. normal noises. The corresponding average log-likelihood function is given by

$$\ell(\boldsymbol{\beta}, \sigma^2; \mathcal{D}) = -\frac{1}{2}\left(\ln(2\pi\sigma^2) + \frac{1}{P\sigma^2} \|\boldsymbol{z} - Y\boldsymbol{\beta}\|_2^2 \right).$$

The corresponding estimation metric would be

$$\begin{aligned} \rho(\boldsymbol{\beta}; \mathcal{D}) &= \min_{\sigma^2 \in \mathbb{R}_+} \frac{1}{2}\left(\ln(\sigma^2) + \frac{1}{P\sigma^2} \|\boldsymbol{z} - Y\boldsymbol{\beta}\|_2^2 \right) \\ &= \frac{1}{2}\left(\ln \frac{t\|\boldsymbol{z} - Y\boldsymbol{\beta}\|_2^2}{P} + 1 \right). \end{aligned} \tag{21}$$

It is known that the maximum likelihood estimate is $\hat{\boldsymbol{\beta}} = (\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{z}$. Hence the optimal value of the estimation metric is $\hat{\rho} = \frac{1}{2}\left(\ln\left(\|\boldsymbol{z} - \boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{z}\|_2^2 / P\right) + 1\right)$. Note that the ML estimate

Electronic copy available at: https://ssrn.com/abstract=3335889

$\hat{\boldsymbol{\beta}}$ here is exactly the same as the OLS estimate derived when using the estimation metric (12). Moreover, the uncertainty set of estimates here has the same geometric characterization as the set defined by (12).

PROPOSITION 8. *Let $r > 0$. Then the corresponding support function of the estimate uncertainty set $\mathcal{E}(r; \mathcal{D})$, given the estimation metric* (21)*, is*

$$\begin{aligned} \delta_r^*(\boldsymbol{\nu}) = \min\ & \gamma(r)\|\boldsymbol{w}\|_2 + \boldsymbol{z}'\boldsymbol{w} \\ \text{s.t. } & \boldsymbol{Y}'\boldsymbol{w} = \boldsymbol{\nu}, \\ & \boldsymbol{w} \in \mathbb{R}^P, \end{aligned} \tag{22}$$

where $\gamma(r) = \sqrt{P\exp(2\hat{\rho} + 2r - 1)}$.

EXAMPLE 4 (ROBUST NEWSVENDOR). Here we consider a single-product newsvendor problem in which the newsvendor, who faces uncertain demand, must simultaneously determine the stocking quantity and the selling price (see, for instance, Kunreuther and Schrage 1973, Federgruen and Heching 1999, Petruzzi and Dada 1999, Ramachandran et al. 2018). Suppose the given data are $\mathcal{D} = \{\boldsymbol{z}, \boldsymbol{Y}\}$, where $\boldsymbol{z} \in \mathbb{R}^P$ is a vector of demand observations for the past P periods, and $\boldsymbol{Y} = \{\mathbf{1}, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_{M-1}\}$ is a matrix of observations of factors—such as selling price, advertising, promotion, and so forth—that affect demand. To derive a robust solution that can control the overall targeted reward even in worst-case scenarios, the newsvendor could solve the following JERO problem:

$$\begin{aligned} \max\ & r \\ \text{s.t. } & p\min\{x, \boldsymbol{\beta}'\boldsymbol{f}\} - cx \geq \tau \qquad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}), \\ & f_1 = p, \\ & (x, \boldsymbol{f}) \in \mathcal{X},\ p \in \mathcal{L},\ r \in \mathbb{R}_+. \end{aligned}$$

Here x is the stocking quantity; $\boldsymbol{f}$ is a vector of factors, which include the selling price p, τ is the reward target; and $\boldsymbol{\beta} \in \mathbb{R}^M$ is a vector of regression coefficients. In addition, $c > 0$ is the ordering cost per unit and $\mathcal{L} = \{p_1, \ldots, p_K\}$, $k \in \mathbb{N}_+$. If we use the estimation metric (21), then the JERO problem has the tractable reformulation

$$\begin{aligned} \max\ & r \\ \text{s.t. } & px - cx \geq \tau, \\ & \gamma(r)\|\boldsymbol{w}\|_2 + \boldsymbol{z}'\boldsymbol{w} + cx + \tau \leq 0, \\ & \boldsymbol{Y}'\boldsymbol{w} + p\boldsymbol{f} = \mathbf{0}, \\ & f_1 = p, \\ & (x, \boldsymbol{f}) \in \mathcal{X},\ p \in \mathcal{L},\ r \in \mathbb{R}_+,\ \boldsymbol{w} \in \mathbb{R}^P, \end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=3335889

where $\gamma(r) = \sqrt{P\exp(2\hat{\rho}+2r-1)}$ and $\hat{\rho}$ is the optimal value of the corresponding estimation metric. The optimal solution can be obtained by fixing the price and solving the corresponding sequence of SOCPs for each $p \in \mathcal{L}$.

Poisson Regression. Suppose we are given data $\mathcal{D} = \{\boldsymbol{z}, \boldsymbol{Y}\}$, where $\boldsymbol{z} \in \mathbb{N}^P$ is a vector of observations of the response variable and $\boldsymbol{Y} = \{\mathbf{1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{M-1}\} \in \mathbb{R}^{P\times M}$ is a matrix of observations of dependent variables. Then we consider the linear model (20) with respect to a Poisson distribution. The corresponding average log-likelihood function is given by

$$\ell(\boldsymbol{\beta}; \mathcal{D}) = \frac{1}{P}\sum_{i=1}^{P}\big(z_i \ln(\boldsymbol{\beta}'\boldsymbol{Y}_i) - \boldsymbol{\beta}'\boldsymbol{Y}_i - \ln(z_i!)\big),$$

where $\boldsymbol{\beta} \in \mathbb{R}^M$ is a vector of regression coefficients and $\boldsymbol{Y}_i$ is the ith row of the matrix $\boldsymbol{Y}$. The corresponding estimation metric would be

$$\rho(\boldsymbol{\beta}; \mathcal{D}) = \frac{1}{P}\sum_{i=1}^{P}\big(\boldsymbol{\beta}'\boldsymbol{Y}_i - z_i \ln(\boldsymbol{\beta}'\boldsymbol{Y}_i)\big). \tag{23}$$

PROPOSITION 9. *Let $r > 0$. Then the corresponding support function of the estimate uncertainty set $\mathcal{E}(r; \mathcal{D})$, given the estimation metric* (23)*, is*

$$\begin{aligned} \delta_r^*(\boldsymbol{\nu}) = \min\ & (\hat{\rho}+r)Pu + \boldsymbol{e}'\boldsymbol{w} \\ s.t.\ & \boldsymbol{\nu} + \textstyle\sum_{i=1}^{P} v_i\boldsymbol{Y}_i - u\sum_{i=1}^{P}\boldsymbol{Y}_i = \mathbf{0}, \\ & z_i u\ln(z_i u/v_i) - z_i u - w_i \le 0 \qquad \forall i \in [P], \\ & u \in \mathbb{R}_+,\ \boldsymbol{v} \in \mathbb{R}_+^P,\ \boldsymbol{w} \in \mathbb{R}^P, \end{aligned} \tag{24}$$

where $\hat{\rho}$ is the optimal value of the corresponding estimation metric.

REMARK 1. The nonlinear (or, equivalently, exponential) constraints

$$z_i u \ln\left(\frac{z_i u}{v_i}\right) - z_i u - w_i \le 0 \quad \forall i \in [P]$$

in (24)—as well as those in the following propositions and examples—are generally challenging. We address this issue by constructing an SOC approximation for the exponential cone in Appendix B.

Logistic Regression Suppose we have the data $\mathcal{D} = \{\boldsymbol{z}, \boldsymbol{Y}\}$, where $\boldsymbol{z} \in \{0,1\}^P$ is a vector of observations of the binary response variable and $\boldsymbol{Y} = \{\mathbf{1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{M-1}\} \in \mathbb{R}^{P\times M}$ is a matrix of observations of dependent variables. Assume that values of the dependent variable exhibit a Bernoulli distribution whose probability is of the form

$$p(\boldsymbol{x}) = \frac{\exp(\boldsymbol{\beta}'\boldsymbol{x})}{1+\exp(\boldsymbol{\beta}'\boldsymbol{x})}, \tag{25}$$

where $\boldsymbol{x} \in \mathbb{R}^M$ is a vector of factors and $\boldsymbol{\beta} \in \mathbb{R}^M$ is a vector of regression coefficients.

Electronic copy available at: https://ssrn.com/abstract=3335889

For simplicity, we have re-ordered the data such that $z_1 = z_2 = \ldots = z_Q = 1$ for $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_Q$ and such that $z_{Q+1} = z_{Q+2} = \ldots = z_P = 0$ for $\boldsymbol{Y}_{Q+1}, \boldsymbol{Y}_{Q+2}, \ldots, \boldsymbol{Y}_P$; here $\boldsymbol{Y}_i$ is the ith row of the matrix $\boldsymbol{Y}$. The corresponding average log-likelihood function is then given by

$$\ell(\boldsymbol{\beta}; \mathcal{D}) = \frac{1}{P}\sum_{i=1}^{Q} \boldsymbol{\beta}'\boldsymbol{Y}_i - \frac{1}{P}\sum_{i=1}^{P} \ln(1 + \exp(\boldsymbol{\beta}'\boldsymbol{Y}_i)),$$

and the corresponding estimation metric would be

$$\rho(\boldsymbol{\beta}; \mathcal{D}) = \frac{1}{P}\sum_{i=1}^{P} \ln(1 + \exp(\boldsymbol{\beta}'\boldsymbol{Y}_i)) - \frac{1}{P}\sum_{i=1}^{Q} \boldsymbol{\beta}'\boldsymbol{Y}_i. \tag{26}$$

PROPOSITION 10. *The corresponding support function of the estimate uncertainty set $\mathcal{E}(r; \mathcal{D})$, given the estimation metric* (26)*, is*

$$\begin{aligned} \delta_r^*(\boldsymbol{\nu}) = \min\ & (\hat{\rho} + r)Pu + (\boldsymbol{v} + \boldsymbol{f} + \boldsymbol{t})'\boldsymbol{e} \\ s.t.\ & u\sum_{i=1}^{Q} \boldsymbol{Y}_i - \sum_{i=1}^{P} s_i \boldsymbol{Y}_i + \boldsymbol{\nu} = \boldsymbol{0}, \\ & u - w_i - s_i = 0 \qquad \forall i \in [P], \\ & w_i \ln(w_i/v_i) - w_i - f_i \leq 0 \qquad \forall i \in [P], \\ & s_i \ln(s_i/v_i) - s_i - t_i \leq 0 \qquad \forall i \in [P], \\ & u \in \mathbb{R}_+,\ \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{s} \in \mathbb{R}_+^P,\ \boldsymbol{f}, \boldsymbol{t} \in \mathbb{R}^P, \end{aligned} \tag{27}$$

where $\hat{\rho}$ is the optimal value of the corresponding estimation metric.

EXAMPLE 5 (EPIDEMIC MANAGEMENT). Consider an epidemic management problem, where a decision maker aims to control the spread of disease by interventions that depend on various risk factors (Prentice and Pyke 1979). Suppose that the relevant data are $\mathcal{D} = \{\boldsymbol{z}, \boldsymbol{Y}\}$, where $\boldsymbol{z} \in \{0,1\}^P$ is a vector of binary observations of disease outcomes and $\boldsymbol{Y} = \{\boldsymbol{1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{M-1}\} \in \mathbb{R}^{P \times M}$ is a matrix of observations of risk factors. To undertake a robust intervention, the decision maker could solve the following JERO problem:

$$\begin{aligned} \max\ & r \\ \text{s.t. } & \boldsymbol{\beta}'\boldsymbol{x} \leq \tau \qquad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}), \\ & \boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^M,\ r \in \mathbb{R}_+; \end{aligned}$$

In this problem, $\boldsymbol{x}$ is the decision variable, $e^\tau/(1+e^\tau)$ is the target incident rate, and $\boldsymbol{\beta} \in \mathbb{R}^M$ is a vector of regression coefficients. If we use the estimation metric (26), then the optimization

Electronic copy available at: https://ssrn.com/abstract=3335889

problem would be equivalent to

$$\begin{aligned}
&\max\ r \\
&\text{s.t. } (\hat{\rho}+r)Pu+(\boldsymbol{v}+\boldsymbol{f}+\boldsymbol{t})'\boldsymbol{e}\le\tau, \\
&\quad u\sum_{i=1}^{Q}\boldsymbol{Y}_i-\sum_{i=1}^{P}s_i\boldsymbol{Y}_i+\boldsymbol{\nu}=\boldsymbol{0}, \\
&\quad u-w_i-s_i=0 && \forall i\in[P], \\
&\quad w_i\ln\left(\tfrac{w_i}{v_i}\right)-w_i-f_i\le 0 && \forall i\in[P], \\
&\quad s_i\ln\left(\tfrac{s_i}{v_i}\right)-s_i-t_i\le 0 && \forall i\in[P], \\
&\quad \boldsymbol{x}\in\mathcal{X}\subseteq\mathbb{R}^N, \\
&\quad r,u\in\mathbb{R}_+,\ \boldsymbol{v},\boldsymbol{w},\boldsymbol{s}\in\mathbb{R}_+^P,\ \boldsymbol{f},\boldsymbol{t}\in\mathbb{R}^P,
\end{aligned}$$

where $\hat{\rho}$ is the optimal value of the corresponding estimation metric.

Independent Marginal. Suppose that our data are given by $\mathcal{D}$ and that the underlying distribution consists of $|J|$ independent marginal distributions. Then the corresponding average log-likelihood function and estimation metric for the jth marginal distribution are, respectively, $\ell_j(\boldsymbol{\beta};\mathcal{D})$ and $\rho_j(\boldsymbol{\beta};\mathcal{D})$. Instead of using the average log-likelihood, we can use the weighted average log-likelihood, $\ell(\boldsymbol{\beta};\mathcal{D})=\sum_{j=1}^{J}\ell_j(\boldsymbol{\beta};\mathcal{D})$. There are many alternatives for the estimation metric. For example, one could define it as $\rho(\boldsymbol{\beta};\mathcal{D})=\sum_{j=1}^{J}\rho_j(\boldsymbol{\beta};\mathcal{D})$. Here we focus on the estimation metric defined as $\rho(\boldsymbol{\beta};\mathcal{D})=\max_{j\in[J]}\rho_j(\boldsymbol{\beta};\mathcal{D})$.

Proposition 11. *Assume that* $\operatorname{ri}(\mathcal{E}(r;\mathcal{D}))\neq\emptyset$ *and that* $\rho_j(\boldsymbol{\beta};\mathcal{D})$ *is convex in* $\boldsymbol{\beta}$ *for all* $j\in[J]$*; then the support function of* $\mathcal{E}(r;\mathcal{D})$ *is*

$$\delta_r^*(\boldsymbol{\nu})=\inf\left\{\sum_{j=1}^{J}(\delta_{r_j}^j)^*(\boldsymbol{\nu_j})\,|\,\sum_{j=1}^{J}\boldsymbol{\nu}_j=\boldsymbol{\nu}\right\}. \tag{28}$$

Here $(\delta_r^j)^*(\boldsymbol{\nu_j})$ *is the corresponding support function of the estimate set defined by* $\rho_j(\boldsymbol{\beta};\mathcal{D})$*;* $\hat{\rho}$ *and* $\hat{\rho}_j$ *are the respective optimal values of the estimation metrics* $\rho(\boldsymbol{\beta};\mathcal{D})$ *and* $\rho_j(\boldsymbol{\beta};\mathcal{D})$*; and* $r_j=\hat{\rho}-\hat{\rho}_j+r$*.*

We now provide an example of independent marginal distributions.

Example 6 (Mixture Independent Case). Consider the data $\mathcal{D}=\{\mathcal{D}_1,\mathcal{D}_2\}$ and a distribution consisting of two independent marginal distributions: a finite support distribution on $\Xi=\{\xi_1,\xi_2,\cdots,\xi_M\}$ with data $\mathcal{D}_1=\{z_1^{(1)},\cdots,z_{P^{(1)}}^{(1)}\}$ and a Poisson distribution with data $\mathcal{D}_2=\{z_1^{(2)},\cdots,z_{P^{(2)}}^{(2)}\}$. Let

$$\rho_1(\boldsymbol{\beta}^{(1)};\mathcal{D}_1)=-\frac{1}{P^{(1)}}\sum_{i=1}^{M}P_i^{(1)}\ln(\beta_i^{(1)}),$$

and

$$\rho_2(\beta^{(2)};\mathcal{D}_2)=\beta^{(2)}-\bar{z}^{(2)}\ln\beta^{(2)},$$

Electronic copy available at: https://ssrn.com/abstract=3335889

where $\boldsymbol{\beta}^{(1)} \in \mathbb{R}_+^M$ is the probability distribution of the finite support distribution, $\beta^{(2)} \in \mathbb{R}_+$ is the mean of the Poisson distribution, $P_i^{(1)} = \sum_{j=1}^{P^{(1)}} \mathbb{I}\left(z_j^{(1)} = \xi_i\right)$ for $i \in [M]$, and $\bar{z}^{(2)} = \frac{1}{P^{(2)}} \sum_{i=1}^{P^{(2)}} z_i^{(2)}$. Then the estimation metric would be

$$\rho(\boldsymbol{\beta}^{(1)}, \beta^{(2)}; \mathcal{D}) = \max\left\{-\frac{1}{P^{(1)}} \sum_{i=1}^{M} P_i^{(1)} \ln(\beta_i^{(1)}),\, \beta^{(2)} - \bar{z}^{(2)} \ln \beta^{(2)}\right\}.$$

Let $r > 0$. Then the corresponding support function of $\mathcal{E}(r; \mathcal{D})$ becomes

$$\begin{aligned}
&\min\ (\hat{\rho} + r) P^{(1)} u + v + \boldsymbol{e}'\boldsymbol{s} + (\hat{\rho} + r) h + g \\
&\text{s.t. } v - w_i \geq \nu_i^1 && \forall i \in [M], \\
&\quad P_i^{(1)} u \ln\left(\frac{P_i^{(1)} u}{w_i}\right) - P_i^{(1)} u - s_i \leq 0 && \forall i \in [M], \\
&\quad h - d \geq \nu^2, \\
&\quad \bar{z}^{(2)} h \ln(\bar{z}^{(2)}/d) - \bar{z}^{(2)} h - g \leq 0, \\
&\quad d, h, u \in \mathbb{R}_+,\, g, v \in \mathbb{R},\, \boldsymbol{w} \in \mathbb{R}_+^M,\, \boldsymbol{s} \in \mathbb{R}^M,
\end{aligned}$$

where $\hat{\rho}$ is the optimal value of the corresponding estimation metric.

Example 7 (Queue Management). Consider a queueing system of N parallel $M/M/1$ queues. In queue $i \in [N]$, the exogenous arrival rate is λ_i and the service rate is modeled as $\mu_i = \boldsymbol{\beta}_i' \boldsymbol{x}_i$, where $\boldsymbol{\beta}_i \in \mathbb{R}^M$ is a vector of coefficients affecting the service rate and $\boldsymbol{x}_i \in \mathbb{R}^M$ is a vector of resources allocated to queue i. On a medical team, for example, the resources may include the number of physicians and nurses with different skill sets, medical equipment and materials, and so forth. Suppose the data are $\mathcal{D} = \{\boldsymbol{z}^i, \boldsymbol{Y}^i\}_{i=1}^N$, where $\boldsymbol{z}^i \in \mathbb{R}^{P^{(i)}}$ is a vector of service rate observations for queue i and $\boldsymbol{Y}^i \in \mathbb{R}^{P^{(i)} \times M}$ is a matrix of historical values of resources allocated to that queue. Then the service system manager could solve the following JERO problem to obtain a robust solution—for resource allocation—such that each queue meets the prespecified service level:

$$\begin{aligned}
&\max\ r \\
&\text{s.t. } \boldsymbol{\beta}_i' \boldsymbol{x}_i - \lambda_i \geq 1/\tau_i && \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}),\, \forall i \in [N], \\
&\quad \textstyle\sum_{i=1}^N \boldsymbol{x}_i \leq \boldsymbol{c}, \\
&\quad (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \in \mathcal{X},\, r \in \mathbb{R}_+.
\end{aligned}$$

Here $\boldsymbol{x}_i \in \mathbb{R}^M$ is the decision variable denoting resources allocated to queue i and $\boldsymbol{c}$ is a vector of constraints on different resources; and τ_i is the targeted average response time or average sojourn time, where the latter is the total time a customer spends in the system. The estimation metric in this case is given by

$$\rho(\boldsymbol{\beta}, \mathcal{D}) = \max_{i \in [N]} \left\{\frac{1}{P^{(i)}} \sum_{j=1}^{P^{(i)}} (z_j^i \boldsymbol{\beta}_i' \boldsymbol{Y}_j^i - \ln(\boldsymbol{\beta}_i' \boldsymbol{Y}_j^i))\right\},$$

Electronic copy available at: https://ssrn.com/abstract=3335889

where $\boldsymbol{Y}_j^i$ is the jth row of the matrix $\boldsymbol{Y}^i$. If we use the estimation metric just given, then the optimization problem would be equivalent to

$$\begin{aligned}
&\max\ r \\
&\text{s.t. } P^{(i)}(\hat{\rho}+r)u_i + \boldsymbol{e}'\boldsymbol{w}_i + \tau_i\lambda_i + 1 \le 0 \quad \forall i \in [N], \\
&\quad \textstyle\sum_{j=1}^{P^{(i)}}(z_j^i u_i - v_{ij})\boldsymbol{Y}_j^i + \tau_i \boldsymbol{x}_i = \boldsymbol{0} \quad \forall i \in [N], \\
&\quad u_i \ln(u_i/v_{ij}) - u_i - w_{ij} \le 0 \quad \forall i \in [N], \forall j \in [P^{(i)}], \\
&\quad \textstyle\sum_{i=1}^{N} \boldsymbol{x}_i \le \boldsymbol{c}, \\
&\quad \boldsymbol{w}_i \in \mathbb{R}^{P^{(i)}},\ \boldsymbol{v}_i \in \mathbb{R}_+^{P^{(i)}} \quad \forall i \in [N], \\
&\quad r \in \mathbb{R}_+,\ \boldsymbol{u} \in \mathbb{R}^N,\ (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N) \in \mathcal{X},
\end{aligned}$$

where $\hat{\rho}$ is the optimal value of the corresponding estimation metric. It is worth noting that some components of the decision variable here could be integers.

4. An Application to Improve Patient Flow

In the healthcare system that is the subject of our case study, insurance reimbursement is an important problem. Since every beneficiary pays a premium and shares the cost of treatment, health insurance can reduce the financial risk faced by covered families. Here we illustrate how the proposed framework can be used to solve a design problem of health insurance reimbursement and to improve patient flow in the healthcare system. We then test the viability of our proposed JERO framework with real-world data. To compare the JERO model's performance with the traditional model while using deterministic input parameters, we separate the original data into a training set and a test set according as whether the data were collected on an odd- or even-numbered day. Our criterion for the optimization model's performance is the number of constraints violated under the test set.

4.1. Data Set

The data used in our study comes from the New Rural Cooperative Medical Insurance (NRCMI) in the Anhui province of China. In this case study, we include 26 hospitals that admitted a total number of 25,500 cerebral infarction (CI) patients during 2015. Each data record includes patient information (e.g., patient ID, age, gender, cost, reimbursement, admission data, discharge date, length of stay, location) and hospital information (hospital ID, hospital level, number of beds, location, etc.). In these 26 hospitals, the total medical expenditure in 2015 for these admitted CI patients was about ¥156 million while the total amount of reimbursement from health insurance companies amounted to some ¥102 million; thus the average reimbursement ratio was 65.41%, and it varied from 43.81% to 82.37% depending on the hospital. The number of CI patients admitted to a hospital in 2015 ranged between 519 and 3,813 and the average length of stay varied from 5.47 to 20.25 days. The average cost per inpatient ranged between ¥1,352 and ¥21,405. Basic statistics are summarized in Table 1.

Electronic copy available at: https://ssrn.com/abstract=3335889

Table 1 Basic statistics for data

Hospital ID	Number of patients	Average LOS (days)	Number of beds	Average cost (¥)	Average reimbursement (¥)	Reimbursement ratio (%)
485926898	1,601	9.00	424	6,550	4,355	66.49
486177120	659	7.76	380	4,518	3,240	71.72
05445776-X	930	9.74	210	4,945	3,832	77.49
08520520-7	643	10.36	60	3,150	2,320	73.64
48500116-6	709	13.27	1,214	21,405	9,378	43.81
48500425-2	511	20.25	300	16,819	9,150	54.40
48520931-1	707	7.49	300	4,151	3,154	75.97
48539625-X	810	8.09	80	3,975	3,146	79.14
48539640-1	704	6.93	260	3,729	2,855	76.56
48570467-0	876	9.04	300	4,339	3,061	70.54
48571563-6	755	7.89	341	6,659	4,854	72.90
48585855-7	1,059	12.76	1,070	8,813	4,433	50.30
48593620-1	1,896	12.00	560	8,339	5,511	66.09
48593738-6	519	13.40	80	1,432	1,143	79.82
48593799-2	725	5.47	50	1,352	1,086	80.28
48596897-5	3,813	9.00	700	6,652	4,716	70.90
48604020-0	1,279	11.23	570	4,766	3,578	75.08
48616610-6	786	5.72	150	3,464	2,782	80.31
48617707-5	537	13.57	200	7,372	5,699	77.31
48617713-9	736	12.52	50	1,886	1,491	79.07
48641950-5	1,049	7.44	800	4,493	2,948	65.62
48641952-1	689	9.82	300	3,996	2,816	70.47
73003174-9	1,213	10.20	165	5,231	3,857	73.74
73003431-7	616	8.85	80	1,638	1,349	82.37
78493096-9	949	9.84	400	7,650	4,822	63.04
N031246	729	8.20	600	8,633	4,153	48.11
Total [Avg.]	25,500			156,239,029	102,196,990	[65.41]

4.2. Regression Model

Reimbursement is widely acknowledged to be a key factor affecting hospital demand, and regression methods can be used to estimate the extent to which that demand is influenced by insurance

Electronic copy available at: https://ssrn.com/abstract=3335889

reimbursement levels. For example, one can calculate that if the reimbursement were increased by ¥1,000, then the number of bed-days would increase by 151. In this case study, we use a linear model to characterize the relationship between hospital demand (measured in bed-days) and four factors: length of stay (LOS), medical cost, hospital capacity and reimbursement,

$$Bed-days = \beta_0 + \beta_1 LOS + \beta_2 Cost + \beta_3 Capacity + \beta_4 Reimbursement;$$

in this expression, $\boldsymbol{\beta} \in \mathbb{R}^5$ is the coefficient to be estimated.

Table 2 reports the coefficients estimated from data—which include the training set (odd days) and the test set (even days)—using the proposed estimation metrics (21) and (23). When we use linear regression, the coefficient β_5 for reimbursement is 151.11 with the training data set and 139.94 with the test data set. When we use Poisson regression, the coefficient β_5 for reimbursement is 104.56 with the training set and 100.73 with the test set; these coefficients are smaller but more robust than those estimated via linear regression.

Table 2 Deterministic estimates from data

Data set	Estimation metric	Intercept β_1	LOS β_2	Cost β_3	Capacity β_4	Reimbursement β_5
Training set	Linear regression	-18.71	2.26	-92.63	0.05	151.11
	Poisson regression	–8.61	1.90	-70.79	0.05	104.56
Test set	Linear regression	-15.47	2.19	-85.93	0.04	139.94
	Poisson regression	–7.43	1.97	-68.31	0.04	100.73

4.3. Traditional Deterministic Optimization Model

Using coefficients estimated from data as input parameters, we can formulate an optimization problem. The decision variable is the average reimbursement value per bed-day (or, equivalently, the reimbursement ratio) for each hospital, and the objective is to minimize the total amount (budget) of reimbursement while serving no fewer than the data set's 25,500 patients. We set the reimbursement ratio's lower bound at 40%, which is a little under the lowest value (43.81%) in the data. Note that each hospital must also satisfy some patient flow constraints. The deterministic optimization model is given by

$$\begin{aligned}
\min\ & \textstyle\sum_{i=1}^{4} \boldsymbol{x}'\boldsymbol{y}_i\hat{\beta}_i + \hat{\beta}_5\boldsymbol{x}'\boldsymbol{x} \\
\text{s.t.}\ & \textstyle\sum_{i=1}^{4} \boldsymbol{d}'\boldsymbol{y}_i\hat{\beta}_i + \hat{\beta}_5\boldsymbol{d}'\boldsymbol{x} \geq 25500, \\
& \underline{\boldsymbol{\tau}} \leq \textstyle\sum_{i=1}^{4} \mathrm{diag}(\boldsymbol{d})\boldsymbol{y}_i\hat{\beta}_i + \hat{\beta}_5\mathrm{diag}(\boldsymbol{d})\boldsymbol{x} \leq \bar{\boldsymbol{\tau}}, \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u},\ \boldsymbol{x} \in \mathbb{R}^{26},
\end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=3335889

where $\boldsymbol{x} \in \mathbb{R}^{26}$ is a decision variable denoting the average reimbursement value per bed-day, $\boldsymbol{d} \in \mathbb{R}^{26}$ represents the reciprocals of the length of stay in each hospital, and $\boldsymbol{l}, \boldsymbol{u} \in \mathbb{R}^{26}$ are (respectively) the lower bound and upper bound on the reimbursement. Also, $\boldsymbol{y}_1 \in \mathbb{R}^{26}$ is the vector of 1s, $\boldsymbol{y}_2 \in \mathbb{R}^{26}$ is the LOS vector, $\boldsymbol{y}_3 \in \mathbb{R}^{26}$ is the vector of the cost, $\boldsymbol{y}_4 \in \mathbb{R}^{26}$ is the vector representing the number of beds in each hospital, and $\underline{\boldsymbol{\tau}}, \bar{\boldsymbol{\tau}} \in \mathbb{R}^{26}$ are (respectively) the lower bound and upper bound on the patient flow.

Solving this optimization model yields, for each hospital, the reimbursement value that minimizes the budget: about ¥91 million, which is 10.78% less than the average budget in the data (The reimbursement rates are reported in Table 3; see Section 4.5.). Absent any consideration of uncertainty in the coefficient estimation, the solution obtained from the deterministic model is quite attractive. That said, the actual parameter almostly certaintly differs from the ML estimate. Hence the actual objective value may escalate and render some of the constraints infeasible.

4.4. JERO Model

As an alternative to the deterministic model, for which the exclusion of uncertainty casts some doubt on derived solutions, we solve the corresponding JERO problem:

$$
\begin{aligned}
&\max\ r \\
&\text{s.t. } \textstyle\sum_{i=1}^{4} \boldsymbol{x}'\boldsymbol{y}_i\beta_i + \beta_5 \boldsymbol{x}'\boldsymbol{x} \le B && \forall \boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D}), \\
&\quad \textstyle\sum_{i=1}^{4} \boldsymbol{d}'\boldsymbol{y}_i\beta_i + \beta_5 \boldsymbol{d}'\boldsymbol{x} \ge 25500 && \forall \boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D}), \\
&\quad \underline{\boldsymbol{\tau}} \le \textstyle\sum_{i=1}^{4} \operatorname{diag}(\boldsymbol{d})\boldsymbol{y}_i\beta_i + \beta_5 \operatorname{diag}(\boldsymbol{d})\boldsymbol{x} \le \bar{\boldsymbol{\tau}} && \forall \boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D}), \\
&\quad \boldsymbol{l} \le \boldsymbol{x} \le \boldsymbol{u},\ \boldsymbol{x} \in \mathbb{R}^{26};
\end{aligned}
$$

here B is an acceptable budget, such as ¥102 millions (the current budget).

Normal Demand The tractability and solution of this robustness optimization problem depend on how we characterize the estimate uncertainty set. One possibility is to derive the estimation metric via linear regression while assuming that the demand for hospital services is normally distributed. Under this assumption, we will show that the JERO problem can be reformulated as a tractable optimization problem. Here we consider the JERO problem based on linear regression with the estimate set $\mathcal{E}(r;\mathcal{D})$ given by

$$
\mathcal{E}(r;\mathcal{D}) = \left\{ \boldsymbol{\beta} \in \mathbb{R}^5 \;\middle|\; \begin{array}{l} \frac{1}{2}\ln\left(\|\boldsymbol{z} - \boldsymbol{Y}\boldsymbol{\beta}\|_2^2/P\right) + \frac{1}{2} \le \hat{\rho} + r \\ \beta_5 \ge 0 \end{array} \right\},
$$

Electronic copy available at: https://ssrn.com/abstract=3335889

where $\boldsymbol{z} \in \mathbb{R}^{26}$ is the vector representing the average bed days in each hospital, $\boldsymbol{y}_5 \in \mathbb{R}^{26}$ is the vector of average reimbursement in the data, $\hat{\rho} = \frac{1}{2}\left(\ln\left(\|\boldsymbol{z} - \boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{z}\|_2^2/P\right) + 1\right)$, and $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4, \boldsymbol{y}_5)$. The optimization problem under this estimate set can be reformulated as

$$
\begin{array}{lll}
\max\ r & & \\
\text{s.t. } \gamma\|\boldsymbol{u}\|_2 + \boldsymbol{z}'\boldsymbol{u} \le B, & & \\
\quad \boldsymbol{y}_i'\boldsymbol{u} = \boldsymbol{y}_i'\boldsymbol{x} & \forall i \in \{1,2,3,4\}, & \\
\quad \|\boldsymbol{x}\|_2^2 \le \boldsymbol{y}_5'\boldsymbol{u}, & & \\
\quad \gamma\|\boldsymbol{v}\|_2 + \boldsymbol{z}'\boldsymbol{v} + 25500 \le 0, & & \\
\quad \boldsymbol{y}_i'\boldsymbol{u} + \boldsymbol{y}_i'\boldsymbol{d} = 0 & \forall i \in \{1,2,3,4\}, & \\
\quad \boldsymbol{y}_5'\boldsymbol{v} + \boldsymbol{d}'\boldsymbol{x} \ge 0, & & \\
\quad \gamma\|\boldsymbol{w}_j\|_2 + \boldsymbol{z}'\boldsymbol{w}_j \le \bar{\tau}_j & \forall j \in \{1,\ldots,26\}, & \\
\quad \boldsymbol{y}_i'\boldsymbol{w}_j = d_j y_{ij} & \forall i \in \{1,2,3,4\},\ \forall j \in \{1,\ldots,26\}, & \\
\quad \boldsymbol{y}_5'\boldsymbol{w}_j \ge d_j x_j & \forall j \in \{1,\ldots,26\}, & \\
\quad \gamma\|\boldsymbol{s}_j\|_2 + \boldsymbol{z}'\boldsymbol{s}_j + \underline{\tau}_j \le 0 & \forall j \in \{1,\ldots,26\}, & \\
\quad \boldsymbol{y}_i'\boldsymbol{s}_j + d_j y_{ij} = 0 & \forall i \in \{1,2,3,4\},\ \forall j \in \{1,\ldots,26\}, & \\
\quad \boldsymbol{y}_5'\boldsymbol{s}_j + d_j x_j \ge 0 & \forall j \in \{1,\ldots,26\}, & \\
\quad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{26},\ \boldsymbol{w}_j, \boldsymbol{s}_j \in \mathbb{R}^{26} & \forall j \in \{1,\ldots,26\}. &
\end{array}
\tag{29}
$$

Here $\gamma = \sqrt{26\exp(2\hat{\rho} + 2r - 1)}$ and y_{ij} is the jth element of the vector $\boldsymbol{y}_i$. We remark that the feasibility of this model for a given $r \ge 0$ can be checked by solving the corresponding optimization problem, which can be accomplished using SOCP.

Poisson Demand After careful reflection, we may decide that it would be preferable to use a Poisson rather than a normal distribution—that is, because hospital demand is measured in bed-days, which are integers. Recall from Section 4.2 that Poisson regression is more robust than linear regression. Hence we now consider the JERO problem based on Poisson regression, using the estimate set $\mathcal{E}(r;\mathcal{D})$ given by

$$
\mathcal{E}(r;\mathcal{D}) = \left\{ \boldsymbol{\beta} \in \mathbb{R}^5 \;\middle|\; \begin{array}{l} \frac{1}{26}\sum_{i=1}^{26}(\boldsymbol{Y}_i\boldsymbol{\beta} - z_i \ln(\boldsymbol{Y}_i\boldsymbol{\beta})) \le \hat{\rho} + r \\ \boldsymbol{Y}_i\boldsymbol{\beta} \ge 0\ \forall i \in \{1,\ldots,26\} \\ \beta_5 \ge 0 \end{array} \right\};
$$

here

$$
\hat{\rho} = \min\left\{ \frac{1}{26}\sum_{i=1}^{26}(\boldsymbol{Y}_i\boldsymbol{\beta} - z_i \ln(\boldsymbol{Y}_i\boldsymbol{\beta})) \mid \beta_5 \ge 0,\ \boldsymbol{Y}_i\boldsymbol{\beta} \ge 0\ \forall i \in \{1,\ldots,26\} \right\}
$$

Electronic copy available at: https://ssrn.com/abstract=3335889

and $\boldsymbol{Y}_i$ is the ith row of the matrix $\boldsymbol{Y}$. The optimization problem under this estimate set can be reformulated as

$$
\begin{array}{lll}
\max\ r & & \\
\text{s.t. } 26(\hat{\rho}+r)u+\boldsymbol{e}'\boldsymbol{w}\le B, & & \\
\quad \boldsymbol{y}_i'\boldsymbol{x}+\sum_{j=1}^{26}(v_j-u)Y_{ji}=0 & \forall i\in\{1,2,3,4\}, & \\
\quad \|\boldsymbol{x}\|_2^2+\sum_{j=1}^{26}(v_j-u)Y_{j5}\le 0, & & \\
\quad z_j u\ln(z_j u/v_j)-z_j u-w_j\le 0 & \forall j\in\{1,\ldots,26\}, & \\
\quad 26(\hat{\rho}+r)f+\boldsymbol{e}'\boldsymbol{h}+25500\le 0, & & \\
\quad \boldsymbol{y}_i'\boldsymbol{d}+\sum_{j=1}^{26}(f-g_j)Y_{ji}=0 & \forall i\in\{1,2,3,4\}, & \\
\quad \boldsymbol{d}'\boldsymbol{x}+\sum_{j=1}^{26}(f-g_j)Y_{j5}\le 0, & & \\
\quad z_j f\ln(z_j f/g_j)-z_j f-h_j\le 0 & \forall j\in\{1,\ldots,26\}, & \\
\quad 26(\hat{\rho}+r)k_j+\boldsymbol{e}'\boldsymbol{q}_j\le\bar{\tau}_j & \forall j\in\{1,\ldots,26\}, & \\
\quad d_j y_{ij}+\sum_{j'=1}^{26}(p_{jj'}-k_j)Y_{j'i}=0 & \forall i\in\{1,2,3,4\},\ \forall j\in\{1,\ldots,26\}, & \\
\quad d_j x_j+\sum_{j'=1}^{26}(p_{jj'}-k_j)Y_{j'5}\le 0 & \forall j\in\{1,\ldots,26\}, & \\
\quad z_j k_i\ln(z_j k_i/p_{ij})-z_j k_i-q_{ij}\le 0 & \forall i\in\{1,\ldots,26\},\ \forall j\in\{1,\ldots,26\}, & \\
\quad 26(\hat{\rho}+r)s_j+\boldsymbol{e}'\boldsymbol{\alpha}_j+\underline{\tau}_j\le 0 & \forall j\in\{1,\ldots,26\}, & \\
\quad d_j y_{ij}+\sum_{j'=1}^{26}(s_j-t_{jj'})Y_{j'i}=0 & \forall i\in\{1,2,3,4\},\ \forall j\in\{1,\ldots,26\}, & \\
\quad d_j x_j+\sum_{j'=1}^{26}(s_j-t_{jj'})Y_{j'5}\le 0 & \forall j\in\{1,\ldots,26\}, & \\
\quad z_j s_i\ln(z_j s_i/t_{ij})-z_j s_i-\alpha_{ij}\le 0 & \forall i\in\{1,\ldots,26\},\ \forall j\in\{1,\ldots,26\}, & \\
\quad f,u\in\mathbb{R}_+,\ \boldsymbol{g},\boldsymbol{k},\boldsymbol{s},\boldsymbol{v}\in\mathbb{R}_+^{26},\ \boldsymbol{h},\boldsymbol{w}\in\mathbb{R}^{26}, & & \\
\quad \boldsymbol{p}_j,\boldsymbol{t}_j\in\mathbb{R}_+^{26},\ \boldsymbol{q}_j,\boldsymbol{\alpha}_j\in\mathbb{R}^{26} & \forall j\in\{1,\ldots,26\}. &
\end{array}
\tag{30}
$$

In this formulation, there are 1,404 exponential constraints. Exponential cone solvers such as MOSEK, SCS, ECOS are unable to handle such large-scale problems. However, we can exploit Proposition 12 (see Appendix B) to replace all the exponential constraints in problem (30) with $1404(L+3)$ SOC constraints. Then, for L of moderate size, this problem can be solved by using state-of-the-art solvers such as Gurobi and CPLEX. Although one might expect that assuming a Poisson distribution of demand would complicate the computation, numerical results reported in the next section reveal that doing so actually yields better solutions.

4.5. Numerical Results

We now evaluate, via binary search, the output of problem (30) and compare it with the output of problem (29). We start by optimizing the budget of reimbursement in the deterministic model using estimates from the training set. The optimal budget is about ¥91.2 million under a Poisson distribution of demand and about ¥91.7 million under a normal distribution. Next we use our JERO model to solve the problem directly for a budget about ¥102 million, which is 0.045% less

Electronic copy available at: https://ssrn.com/abstract=3335889

Table 3 Optimal reimbursement rates under different optimization models

Hospital	Optimal reimbursement rate (%)				
ID	Current	Normal	Poisson	JERO-Normal	JERO-Poisson
485926898	66.49%	62.99%	62.52%	70.14%	66.84%
486177120	71.72%	74.61%	77.14%	72.97%	75.36%
05445776-X	77.49%	78.14%	77.71%	87.74%	83.49%
08520520-7	73.64%	100.00%	100.00%	100.00%	100.00%
48500116-6	43.81%	40.98%	40.00%	40.00%	40.00%
48500425-2	54.40%	54.00%	55.24%	40.00%	58.50%
48520931-1	75.97%	79.45%	83.02%	77.57%	80.85%
48539625-X	79.14%	97.56%	99.99%	94.24%	99.93%
48539640-1	76.56%	81.38%	84.86%	79.30%	82.52%
48570467-0	70.54%	83.54%	82.56%	83.24%	88.38%
48571563-6	72.90%	66.28%	67.60%	70.27%	71.61%
48585855-7	50.30%	40.00%	40.00%	40.00%	40.00%
48593620-1	66.09%	46.95%	43.38%	52.68%	47.12%
48593738-6	79.82%	100.00%	100.00%	100.00%	99.86%
48593799-2	80.28%	100.00%	100.00%	100.00%	100.00%
48596897-5	70.90%	56.40%	53.89%	57.76%	57.16%
48604020-0	75.08%	62.10%	54.75%	69.68%	61.70%
48616610-6	80.31%	83.07%	87.33%	80.51%	84.18%
48617707-5	77.31%	61.40%	60.01%	61.59%	62.09%
48617713-9	79.07%	100.00%	100.00%	100.00%	99.87%
48641950-5	65.62%	50.98%	45.45%	47.32%	40.57%
48641952-1	70.47%	85.56%	82.91%	88.83%	90.43%
73003174-9	73.74%	75.89%	75.58%	84.93%	81.00%
73003431-7	82.37%	100.00%	100.00%	100.00%	99.99%
78493096-9	63.04%	57.21%	56.62%	63.32%	60.29%
N031246	48.11%	53.72%	53.85%	59.08%	57.00%
Budget	¥102,196,990	¥91,724,427	¥91,206,700	¥102,151,504	¥102,151,504

than the current budget. Table 3 summarizes the results for problems (29) and (30). Each row gives the reimbursement rates of a hospital in the current data set (second column) and under four different optimization models (third–sixth columns).

Electronic copy available at: https://ssrn.com/abstract=3335889

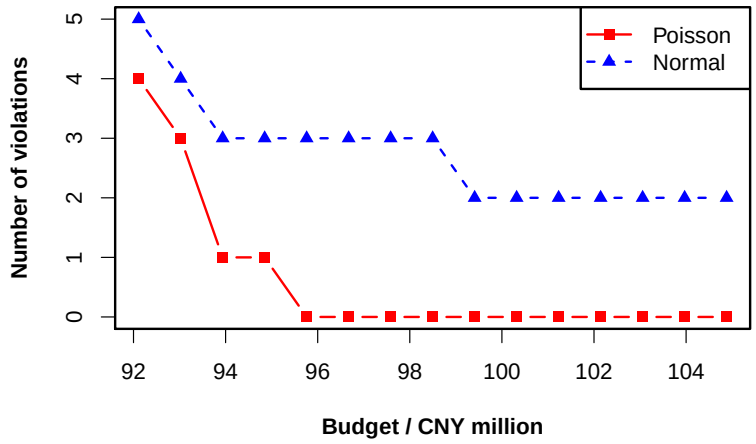

Figure 1 Robustness of solutions

In Figure 1, we use the test data set to show how the number of constraints violated declines when the corresponding budget increases. We use 15 different budgets (ranging from ¥92.1 million to ¥104.9 million) in testing the robustness of our solutions to problems (29) and (30). The number of constraints that are violated falls from 4 to 0 in problem (30) and falls from 5 to 2 in problem (29). Figure 1 reveals that the solution obtained from (30) is more robust than the solution obtained from (29). The reason may well be that problem (30) benefits from the assumption of an underlying Poisson distribution.

5. Conclusion

We propose a new framework that accounts for the uncertainty in parameter estimation and incorporates a problem's entire data set during optimization. We also propose a new method for approximating the exponential cone—via a sequence of second-order cones—that can be instrumental in rendering larger-scale optimization problems more tractable. Finally, we present a case study that demonstrates how the the proposed framework can be applied to solve a design problem of health insurance reimbursement and to balance patient flow in a healthcare system.

Electronic copy available at: https://ssrn.com/abstract=3335889

A. Proofs of Propositions

Proof of Proposition 1. Observe that each loop in Algorithm 1 reduces the gap between r_2 and r_1 by half (see Step 3). We now establish the correctness of this binary search. Suppose that $Z_E^r(\tau_0) \leq 0$. Then r is feasible in model 6 and so $r^* \geq r$, since otherwise r would be infeasible in 6. Because the function $Z_E^r(\tau_0)$ is nondecreasing in r, we have $r^* < r$. Hence the number of computations K must satisfy $\bar{r}/2^K < \Delta$; that is, $K \geq \lceil \log_2(\bar{r}/\Delta) \rceil$, which completes the proof. □

Proof of Proposition 2. Let $\delta_r(\boldsymbol{\beta})$ the indicator function on $\mathcal{E}(r;\mathcal{D})$ defined by

$$\delta_r(\boldsymbol{\beta}) \triangleq \begin{cases} 0, & \text{if } \boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D}) \\ \infty, & \text{otherwise.} \end{cases}$$

Then, we have

$$\begin{aligned} &\max_{\boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D})} a_i(\boldsymbol{x}, \boldsymbol{\beta}) \\ &= \max_{\boldsymbol{\beta} \in \mathbb{R}^M} \{a_i(\boldsymbol{x}, \boldsymbol{\beta}) - \delta_r(\boldsymbol{\beta})\} \\ &= \min_{\boldsymbol{\nu} \in \mathbb{R}^M} \{\delta_r^*(\boldsymbol{\nu}) - a_i^*(\boldsymbol{x}, \boldsymbol{\nu})\}, \end{aligned} \tag{31}$$

where the first equality is due to the definition of indicator function, and the second equality is due to Fenchel's duality theorem in Rockafellar (2015). The assertion now follows if we apply (31) to the robust counterpart (9). □

Proof of Proposition 3. Observe that

$$\begin{aligned} \delta_r^*(\boldsymbol{\nu}) &= \sup_{\boldsymbol{\beta} \in \bar{\mathcal{W}}} \{\boldsymbol{\beta}'\boldsymbol{\nu} \mid \rho(\boldsymbol{\beta};\mathcal{D}) \leq \hat{\rho} + r\} \\ &= \inf_{\mu > 0} \left\{ \sup_{\boldsymbol{\beta} \in \bar{\mathcal{W}}} \{\boldsymbol{\beta}\boldsymbol{\nu} + u(\hat{\rho} + r - \rho(\boldsymbol{\beta};\mathcal{D}))\} \right\} \\ &= \inf_{\mu > 0} \left\{ u(\hat{\rho} + r) + \sup_{\boldsymbol{\beta} \in \bar{\mathcal{W}}} \{\boldsymbol{\beta}\boldsymbol{\nu} - \mu\rho(\boldsymbol{\beta};\mathcal{D})\} \right\} \\ &= \inf_{\mu > 0} \{u(\hat{\rho} + r) + \mu\rho^*(\boldsymbol{\nu}/\mu;\mathcal{D})\}. \end{aligned}$$

Here the second equality is due to strong Lagrangian duality (cf. Bertsekas 1999) and the last equality follows from the definition of $\rho^*(\boldsymbol{\nu};\mathcal{D})$. □

Proof of Proposition 4. By definition, we have

$$\delta_r^*(\boldsymbol{\nu}) = \max_{\boldsymbol{\beta}} \left\{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid \|\boldsymbol{z} - \boldsymbol{Y}\boldsymbol{\beta}\|_2^2 \leq \hat{\rho} + r \right\}.$$

Electronic copy available at: https://ssrn.com/abstract=3335889

Hence strong duality implies that

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \min\ & \sqrt{\hat{\rho}+r}\|\boldsymbol{w}\|_2 + \boldsymbol{z}'\boldsymbol{w} \\ \text{s.t. } & \boldsymbol{Y}'\boldsymbol{w} = \boldsymbol{\nu}, \\ & \boldsymbol{w} \in \mathbb{R}^P,\end{aligned}$$

which completes the proof. □

Proof of Proposition 5. By definition, we have

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) &= \max_{\boldsymbol{\beta}} \left\{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid \frac{1}{P}\|\boldsymbol{z} - \boldsymbol{Y}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \le \hat{\rho} + r \right\} \\ &= \max_{\boldsymbol{\beta}, g\ge 0, h\ge 0} \left\{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid \|\boldsymbol{z} - \boldsymbol{Y}\boldsymbol{\beta}\|_2^2 \le g,\ \|\boldsymbol{\beta}\|_1 \le h,\ \frac{1}{P}g + \lambda h \le \hat{\rho} + r \right\}.\end{aligned}$$

It follows from strong duality that

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \min\ & (\hat{\rho}+r)s + u - v + \boldsymbol{z}'\boldsymbol{w}, \\ \text{s.t. } & \boldsymbol{Y}'\boldsymbol{w} - \boldsymbol{t} = \boldsymbol{\nu}, \\ & \left\| \begin{matrix} 2v \\ \boldsymbol{w} \end{matrix} \right\|_2 \le 2u \\ & \|\boldsymbol{t}\|_\infty \le \lambda s, \\ & Pu + Pv \le s, \\ & u, s \in \mathbb{R}_+,\ v \in \mathbb{R},\ \boldsymbol{w} \in \mathbb{R}^P,\ \boldsymbol{t} \in \mathbb{R}^M.\end{aligned}$$

This completes the proof. □

Proof of Proposition 6. We first consider the estimation metric

$$\begin{aligned}\rho(\boldsymbol{\beta}; \mathcal{D}) &= \min_{\boldsymbol{\Sigma}\in\mathbb{S}_+^M} \frac{1}{2}\left(\ln\det\boldsymbol{\Sigma} + \frac{1}{P}\sum_{i=1}^P (\boldsymbol{z}_i - \boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{\beta}) \right) \\ &= \min_{\boldsymbol{\Sigma}\in\mathbb{S}_+^M} \frac{1}{2}\left(\ln\det\boldsymbol{\Sigma} + \frac{1}{P}\sum_{i=1}^P (\boldsymbol{z}_i - \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right) \\ &= \min_{\boldsymbol{\Sigma}\in\mathbb{S}_+^M} \frac{1}{2}\left(\ln\det\boldsymbol{\Sigma} + \frac{1}{P}\sum_{i=1}^P (\boldsymbol{z}_i - \hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \hat{\boldsymbol{\beta}}) + \frac{1}{P}\sum_{i=1}^P (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right) \\ &= \min_{\boldsymbol{\Sigma}\in\mathbb{S}_+^M} \frac{1}{2}\left(\ln\det\boldsymbol{\Sigma} + \mathbf{tr}(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right) \\ &= \min_{\boldsymbol{S}\in\mathbb{S}_+^M} \frac{1}{2}\left(\mathbf{tr}(\boldsymbol{S}\hat{\boldsymbol{\Sigma}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\boldsymbol{S}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - \ln\det\boldsymbol{S} \right), \qquad (32)\end{aligned}$$

where the third equality is due to regular computations and the last equality follows from the change-of-variable trick.

Electronic copy available at: https://ssrn.com/abstract=3335889

Define a new function

$$f(\boldsymbol{\beta}, \boldsymbol{S}) = \frac{1}{2}\big(\mathbf{tr}(\boldsymbol{S}\hat{\boldsymbol{\Sigma}}) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\boldsymbol{S}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) - \ln\det \boldsymbol{S}\big)$$

for $\boldsymbol{\beta} \in \mathbb{R}^M$ and $\boldsymbol{S} \in \mathbb{S}_{++}^M$. Note that $f(\boldsymbol{\beta}, \boldsymbol{S})$ is convex in $\boldsymbol{S}$. If we set to zero the gradient of $f(\boldsymbol{\beta}, \boldsymbol{S})$ with respect to $\boldsymbol{S}$, then

$$\frac{\partial f(\boldsymbol{\beta}, \boldsymbol{S})}{\partial \boldsymbol{S}} = \frac{1}{2}\left(\hat{\boldsymbol{\Sigma}} + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})' - \boldsymbol{S}^{-1}\right) = 0,$$

which has only one solution:

$$\hat{\boldsymbol{S}} = \left(\hat{\boldsymbol{\Sigma}} + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\right)^{-1}. \tag{33}$$

Applying (33) to (32) and using the optimal condition of convex optimization, we obtain

$$\begin{aligned}
\rho(\boldsymbol{\beta}; \mathcal{D}) &= \frac{1}{2}\big(\mathbf{tr}(\hat{\boldsymbol{S}}\hat{\boldsymbol{\Sigma}}) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{S}}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) - \ln\det \hat{\boldsymbol{S}}\big) \\
&= \frac{1}{2}\big(\mathbf{tr}(\hat{\boldsymbol{S}}\hat{\boldsymbol{\Sigma}}) + \mathbf{tr}(\hat{\boldsymbol{S}}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})') - \ln\det \hat{\boldsymbol{S}}\big) \\
&= \frac{1}{2}\big(\mathbf{tr}(\hat{\boldsymbol{S}}(\hat{\boldsymbol{\Sigma}} + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})')) - \ln\det \hat{\boldsymbol{S}}\big) \\
&= \frac{1}{2}\big(M + \ln\det(\hat{\boldsymbol{\Sigma}} + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})')\big) \\
&= \frac{1}{2}\big(M + \ln\det(\hat{\boldsymbol{\Sigma}}^{1/2}(\boldsymbol{I} + \hat{\boldsymbol{\Sigma}}^{-1/2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Sigma}}^{-1/2})\hat{\boldsymbol{\Sigma}}^{1/2})\big) \\
&= \frac{1}{2}\big(M + \ln\det\hat{\boldsymbol{\Sigma}} + \ln\det(\boldsymbol{I} + \hat{\boldsymbol{\Sigma}}^{-1/2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Sigma}}^{-1/2})\big) \\
&= \frac{1}{2}\big(M + \ln\det\hat{\boldsymbol{\Sigma}} + \ln(1 + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}))\big).
\end{aligned}$$

Next we derive the support function of $\mathcal{E}(r; \mathcal{D})$. By definition, we have

$$\begin{aligned}
\delta_r^*(\boldsymbol{\nu}) &= \max_{\boldsymbol{\beta}} \left\{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid \frac{1}{2}\ln(1 + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})) + \frac{M}{2} + \frac{1}{2}\ln\det\hat{\boldsymbol{\Sigma}} \le \hat{\rho} + r \right\} \\
&= \max_{\boldsymbol{\beta}} \left\{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid \frac{1}{2}\ln(1 + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})) \le r \right\} \\
&= \max_{\boldsymbol{\beta}} \{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) \le e^{2r} - 1 \} \\
&= \hat{\boldsymbol{\beta}}'\boldsymbol{\nu} + \sqrt{e^{2r}-1}\,\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{\nu}\|_2.
\end{aligned}$$

This completes the proof. □

Proof of Proposition 7. By definition,

$$\begin{aligned}
\delta_r^*(\boldsymbol{\nu}) = \max\ & \boldsymbol{\beta}'\boldsymbol{\nu} \\
\text{s.t. } & \tfrac{1}{P}\textstyle\sum_{i=1}^M P_i \ln(\beta_i) + \hat{\rho} + r \ge 0, \\
& \textstyle\sum_{i=1}^M \beta_i = 1, \\
& \boldsymbol{\beta} \in \mathbb{R}_+^M;
\end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=3335889

this problem is equivalent to

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \max\ & \boldsymbol{\beta}'\boldsymbol{\nu} \\ \text{s.t. } & \tfrac{1}{P}\textstyle\sum_{i=1}^M P_i t_i + \hat{\rho} + r \geq 0, \\ & \ln(\beta_i) \geq t_i \qquad\qquad \forall i \in [M], \\ & \textstyle\sum_{i=1}^M \beta_i = 1, \\ & \boldsymbol{\beta} \in \mathbb{R}_+^M,\, \boldsymbol{t} \in \mathbb{R}^M.\end{aligned}$$

It follows from strong duality that, equivalently,

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \min\ & (\hat{\rho}+r)Pu + v + \boldsymbol{e}'\boldsymbol{s} \\ \text{s.t. } & v - w_i \geq \nu_i \qquad\qquad \forall i \in [M], \\ & P_i u \ln\left(\tfrac{P_i u}{w_i}\right) - P_i u - s_i \leq 0 \qquad \forall i \in [M,] \\ & u \in \mathbb{R}_+,\, v \in \mathbb{R},\, \boldsymbol{w} \in \mathbb{R}_+^M,\, \boldsymbol{s} \in \mathbb{R}^M.\end{aligned}$$

This completes the proof. □

Proof of Proposition 8. By definition, we have

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) &= \max_{\boldsymbol{\beta}} \left\{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid \frac{1}{2}\left(\ln \frac{\|\boldsymbol{z} - Y\boldsymbol{\beta}\|_2^2}{P} + 1 \right) \leq \hat{\rho} + r \right\} \\ &= \max_{\boldsymbol{\beta}} \{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid \|\boldsymbol{z} - Y\boldsymbol{\beta}\|_2 \leq \gamma \},\end{aligned}$$

where $\gamma = \sqrt{P\exp(2\hat{\rho} + 2r - 1)}$.

By strong duality, we have

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \min\ & \gamma\|\boldsymbol{w}\|_2 + \boldsymbol{z}'\boldsymbol{w} \\ \text{s.t. } & Y'\boldsymbol{w} = \boldsymbol{\nu}, \\ & \boldsymbol{w} \in \mathbb{R}^P,\end{aligned}$$

which completes the proof. □

Proof of Proposition 9. By definition, we have

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \max\ & \boldsymbol{\beta}'\boldsymbol{\nu} \\ \text{s.t. } & \tfrac{1}{P}\textstyle\sum_{i=1}^P (\boldsymbol{\beta}'\boldsymbol{Y}_i - z_i \ln(\boldsymbol{\beta}'\boldsymbol{Y}_i)) \leq \hat{\rho} + r, \\ & \boldsymbol{\beta} \in \mathbb{R}^M.\end{aligned}$$

This problem is equivalent to

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \max\ & \boldsymbol{\beta}'\boldsymbol{\nu} \\ \text{s.t. } & \tfrac{1}{P}\textstyle\sum_{i=1}^P (\boldsymbol{\beta}'\boldsymbol{Y}_i - z_i f_i) \leq \hat{\rho} + r, \\ & \ln(\boldsymbol{\beta}'\boldsymbol{Y}_i) \geq f_i \qquad\qquad \forall i \in [P], \\ & \boldsymbol{\beta} \in \mathbb{R}^M,\, \boldsymbol{f} \in \mathbb{R}^P.\end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=3335889

It now follows from strong duality that, equivalently,

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \min\ & P(\hat{\rho}+r)u + \boldsymbol{e}'\boldsymbol{w} \\ \text{s.t. } & \boldsymbol{\nu} + \textstyle\sum_{i=1}^P v_i \boldsymbol{Y}_i - u\sum_{i=1}^P \boldsymbol{Y}_i = \boldsymbol{0}, \\ & z_i u \ln\left(\tfrac{z_i u}{v_i}\right) - z_i u - w_i \le 0 \qquad \forall i \in [P], \\ & u \in \mathbb{R}_+,\, \boldsymbol{v} \in \mathbb{R}_+^P,\, \boldsymbol{w} \in \mathbb{R}^P.\end{aligned}$$

This completes the proof. □

Proof of Proposition 10. By definition, we have

$$\delta_r^*(\boldsymbol{\nu}) = \max_{\boldsymbol{\beta}} \left\{ \boldsymbol{\beta}'\boldsymbol{\nu} \mid \frac{1}{P}\sum_{i=1}^P \ln(1+\exp(\beta'\boldsymbol{Y}_i)) - \frac{1}{P}\sum_{i=1}^Q \beta'\boldsymbol{Y}_i \le \hat{\rho}+r \right\},$$

which is equivalent to

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \max\ & \boldsymbol{\beta}'\boldsymbol{\nu} \\ \text{s.t. } & \textstyle\sum_{i=1}^P f_i - \sum_{i=1}^Q \beta'\boldsymbol{Y}_i \le P(\hat{\rho}+r), \\ & g_i + h_i \le 1 && \forall i \in [P], \\ & \exp(-f_i) \le g_i && \forall i \in [P], \\ & \exp(\boldsymbol{\beta}'\boldsymbol{Y}_i - f_i) \le h_i && \forall i \in [P].\end{aligned}$$

By strong duality, we have

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) = \min\ & P(\hat{\rho}+r)u + (\boldsymbol{v}+\boldsymbol{r}+\boldsymbol{t})'\boldsymbol{e} \\ \text{s.t. } & u\textstyle\sum_{i=1}^Q \boldsymbol{Y}_i - \sum_{i=1}^P s_i \boldsymbol{Y}_i + \boldsymbol{\nu} = \boldsymbol{0}, \\ & u - w_i - s_i = 0 && \forall i \in [P], \\ & w_i \ln(w_i/v_i) - w_i - r_i \le 0 && \forall i \in [P], \\ & s_i \ln(s_i/v_i) - s_i - t_i \le 0 && \forall i \in [P], \\ & u \in \mathbb{R}_+,\, \boldsymbol{v}, \boldsymbol{w},\, \boldsymbol{s} \in \mathbb{R}_+^P,\, \boldsymbol{r}, \boldsymbol{t} \in \mathbb{R}^P.\end{aligned}$$

This completes the proof. □

Proof of Proposition 11. By definition, we have

$$\begin{aligned}\delta_r^*(\boldsymbol{\nu}) &= \sup_{\boldsymbol{\beta}\in\mathcal{W}} \{\boldsymbol{\beta}'\boldsymbol{\nu} \mid \rho(\boldsymbol{\beta};\mathcal{D}) \le \hat{\rho}+r\ \forall j \in [J]\} \\ &= \sup_{\boldsymbol{\beta}\in\mathcal{W}} \{\boldsymbol{\beta}'\boldsymbol{\nu} \mid \rho_j(\boldsymbol{\beta};\mathcal{D}) \le \hat{\rho}+r\ \forall j \in [J]\} \\ &= \sup_{\boldsymbol{\beta}_j\in\mathcal{W},\, \boldsymbol{\beta}=\boldsymbol{\beta}_j} \{\boldsymbol{\beta}'\boldsymbol{\nu} \mid \rho_j(\boldsymbol{\beta}_j;\mathcal{D}) \le \hat{\rho}+r\ \forall j \in [J]\} \\ &= \inf_{\sum_{j=1}^J \boldsymbol{\nu}_j = \boldsymbol{\nu}} \left\{ \sup_{\boldsymbol{\beta}_j\in\mathcal{W}} \left\{ \boldsymbol{\beta}'\boldsymbol{\nu} + \sum_{j=1}^J (\boldsymbol{\beta}_j - \boldsymbol{\beta})'\boldsymbol{\nu}_j \mid \rho_j(\boldsymbol{\beta}_j;\mathcal{D}) \le \hat{\rho}+r\ \forall j \in [J] \right\} \right\}\end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=3335889

$$= \inf_{\sum_{j=1}^{J} \boldsymbol{\nu}_j = \boldsymbol{\nu}} \left\{ \sup_{\boldsymbol{\beta}_j \in \mathcal{W}} \left\{ \sum_{j=1}^{J} \boldsymbol{\beta}_j' \boldsymbol{\nu}_j \mid \rho_j(\boldsymbol{\beta}_j; \mathcal{D}) \leq \hat{\rho} + r \ \forall j \in [J] \right\} \right\}$$

$$= \inf_{\sum_{j=1}^{J} \boldsymbol{\nu}_j = \boldsymbol{\nu}} \left\{ \sum_{j=1}^{J} (\delta_{r_j}^{j})^*(\boldsymbol{\nu_j}) \right\}.$$

This completes the proof.

□

B. SOC Approximation for Exponential Cone

Because of the logarithm involved in the MLE metric, the JERO model's tractability will depend mainly on the exponential constraint

$$\alpha \exp(x/\alpha) \leq t \quad \text{or} \quad \alpha \ln(t/\alpha) \geq x$$

for some $\alpha > 0$. Current exponential cone solvers—such as Mosek, SCS and ECOS—cannot solve optimization problems on the scale that we consider when there are exponential constraints; and neither, a fortiori, can such problems be solved via mixed-integer programming. To address the computational issue that arises from the JERO model's logarithm-related constraints, we provide a practical second-order conic approximation for exponential cones that curtails the approximation errors at extreme values. This approximation allows us to use state-of-the-art SOCP solvers capable of solving large-scale convex optimization problems.

Our aim in this section is to show that the exponential constraint can be approximately represented in the form of SOCs. Chen and Sim (2009) discuss one method (originally described by Ben-Tal and Nemirovski 2001) of approximating the exponential cone constraint. Thus we can use Taylor's series expansion to write

$$\exp(x) = \exp\left(\frac{x}{2^L}\right)^{2^L} = \left(1 + \frac{x}{2^L} + \frac{1}{2}\left(\frac{x}{2^L}\right)^2 + \cdots + \frac{1}{K!}\left(\frac{x}{2^L}\right)^K + o(x^{K+1})\right)^{2^L},$$

where K and L are positive integers. The precision level of this expansion increases as K and L become larger (see Table 4).

We can, in theory, approximate e^x to an arbitrary level of numerical precision. However, the complexity of using Taylor expansion to approximate the exponential constraint increases with the precision; that is, we then need more auxiliary variables and SOC constraints (see Table 5). Thus, there is, as usual, a tradeoff between the level of numerical precision and the complexity of computation.

Using only a few second-order cones, we can accurately approximate such constraints to a reasonably good level of numerical precision (Chen and Sim 2009). In contrast, using the Taylor

Electronic copy available at: https://ssrn.com/abstract=3335889

Table 4 Relative error of Taylor expansion in the interval [-20, 60]

$K\backslash L$	4	5	6	7	8	9	10
4	>1	7×10^{-1}	2×10^{-1}	2×10^{-2}	1×10^{-3}	9×10^{-5}	6×10^{-6}
6	8×10^{-1}	1×10^{-1}	4×10^{-3}	8×10^{-5}	2×10^{-6}	3×10^{-8}	5×10^{-10}
8	2×10^{-1}	5×10^{-3}	4×10^{-5}	3×10^{-7}	1×10^{-9}	5×10^{-12}	4×10^{-13}
10	3×10^{-2}	1×10^{-4}	3×10^{-7}	5×10^{-10}	6×10^{-13}	2×10^{-13}	4×10^{-13}

Table 5 Complexity of SOC approximation

$K\backslash L$		$L=4$	$L=5$	$L=6$	$L=7$	$L=8$	$L=9$	$L=10$
Number of variables	$K=4$	9	10	11	12	13	14	15
Number of constraints		10	11	12	13	14	15	16
Number of variables	$K=6$	17	18	19	20	21	22	23
Number of constraints		18	19	20	21	22	23	24
Number of variables	$K=8$	25	26	27	28	29	30	31
Number of constraints		26	27	28	29	30	31	32
Number of variables	$K=10$	41	42	43	44	45	46	47
Number of constraints		42	43	44	45	46	47	48

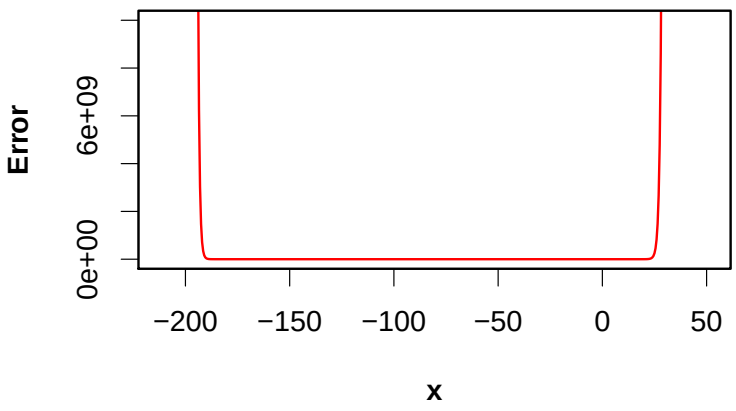

Figure 2 Error of approximation when using Taylor expansion with K=4 and L=6

expansion approximation technique of Chen and Sim (2009) produces significant errors in the tails, as Figure 2 illustrates.

Electronic copy available at: https://ssrn.com/abstract=3335889

We circumvent this problem by first trimming the tails of the exponential function and then using the Taylor expansion only within the truncated interval. Observe that e^x is smaller than 10^{-9} when $x \le -20$ and is larger than 10^{26} when $x \ge 60$. Hence we ease the exposition by trimming the left tail via $f_1(x) \triangleq 0$ with $x \le -20$ and trimming the right tail via $f_3(x) \triangleq \infty$ with $x \ge 60$. Furthermore, we use the following Taylor expansion to approximate e^x for $x \in [-20, 60]$:

$$f_2(x) \triangleq \left(1 + \frac{x}{2^L} + \frac{1}{2}\left(\frac{x}{2^L}\right)^2 + \frac{1}{6}\left(\frac{x}{2^L}\right)^3 + \frac{1}{24}\left(\frac{x}{2^L}\right)^4\right)^{2^L}$$

for L a positive integer. Because $f_3(x) = \infty$ when $x \ge 60$, we use the following infimal convolution of $f_1(x)$ and $f_2(x)$ to approximate $\alpha \exp(x/\alpha)$, $\alpha > 0$:

$$f(x, \alpha) \triangleq \inf\left\{\alpha_1 f_1\left(\frac{x_1}{\alpha_1}\right) + \alpha_2 f_2\left(\frac{x_2}{\alpha_2}\right) \mid x_1 + x_2 = x,\, \alpha_1 + \alpha_2 = \alpha,\, \alpha_1, \alpha_2 > 0\right\}.$$

Then $\alpha \exp(x/\alpha) \le t$ ($\alpha > 0$) can be approximated by $f(x, \alpha) \le t$, which in turn can be represented by a series of SOC constraints.

PROPOSITION 12. *The exponential constraint $\alpha \exp(x/\alpha) \le t$ or $\alpha \ln(t/\alpha) \ge x$, for some $\alpha > 0$, can be approximated by a series of SOC constraints as follows:*

$$\begin{cases} x_1 + x_2 = x,\, \alpha_1 + \alpha_2 = \alpha, \\ y = x_2/2^L,\, z = \alpha_2 + x_2/2^L, \\ \frac{1}{24}(23\alpha_2 + 20y + 6f + h) \le v_1, \\ \left\| \begin{matrix} y \\ (f - \alpha_2)/2 \end{matrix} \right\|_2 \le \frac{f + \alpha_2}{2}, \\ \left\| \begin{matrix} z \\ (g - \alpha_2)/2 \end{matrix} \right\|_2 \le \frac{g + \alpha_2}{2}, \\ \left\| \begin{matrix} g \\ (h - \alpha_2)/2 \end{matrix} \right\|_2 \le \frac{h + \alpha_2}{2}, \\ \left\| \begin{matrix} v_i \\ (v_{i+1} - \alpha_2)/2 \end{matrix} \right\|_2 \le \frac{v_{i+1} + \alpha_2}{2} \qquad \forall i \in [L-1], \\ \left\| \begin{matrix} v_L \\ (t - \alpha_2)/2 \end{matrix} \right\|_2 \le \frac{t + \alpha_2}{2}, \\ x_1/\alpha_1 \le -20,\, -20 \le x_2/\alpha_2 \le 60, \\ \alpha_1, \alpha_2, f, g, h \in \mathbb{R}_+,\, \boldsymbol{v} \in \mathbb{R}_+^L. \end{cases} \tag{34}$$

Electronic copy available at: https://ssrn.com/abstract=3335889

Proof of Proposition 12. First, we approximate $\alpha \exp(x/\alpha)$ by the following infimal convolution:

$$\begin{aligned} \min\ & \alpha_2 \exp(x_2/\alpha_2) \\ \text{s.t. } & x_1 + x_2 = x, \\ & \alpha_1 + \alpha_2 = \alpha, \\ & \tfrac{x_1}{\alpha_1} \le -20,\ -20 \le \tfrac{x_2}{\alpha_2} \le 60, \\ & \alpha_1,\ \alpha_2 \ge 0. \end{aligned}$$

It is then sufficient to approximate $\alpha_2 \exp(x_2/\alpha_2) \le t$ in the form of second-order cones as

$$\begin{cases} y = x_2/2^L, \\ z = \alpha_2 + x_2/2^L, \\ y^2 \le \alpha_2 f,\ z^2 \le \alpha_2 g,\ g^2 \le \alpha_2 h, \\ \frac{1}{24}(23\alpha_2 + 20y + 6f + h) \le v_1, \\ v_i^2 \le \alpha_2 v_{i+1}\ \forall i \in [L-1], \\ v_L^2 \le \alpha_2 t, \\ f, g, h \ge 0,\ \boldsymbol{v} \ge \mathbf{0}. \end{cases}$$

Finally, we use the well-known result that

$$w^2 \le st, \quad s, t \ge 0,$$

is SOC representable as

$$\left\| \begin{matrix} w \\ (s-t)/2 \end{matrix} \right\|_2 \le \frac{s+t}{2}.$$

This completes the proof. □

Electronic copy available at: https://ssrn.com/abstract=3335889

References

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

Ben-Tal, A., D. Den Hertog, and J.-P. Vial (2015). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical programming 149*(1-2), 265–299.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). *Robust optimization*, Volume 28. Princeton University Press.

Ben-Tal, A. and A. Nemirovski (1998). Robust convex optimization. *Mathematics of operations research 23*(4), 769–805.

Ben-Tal, A. and A. Nemirovski (2001). *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, Volume 2. Siam.

Bertsekas, D. P. (1999). *Nonlinear programming*. Athena scientific Belmont.

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

Bertsimas, D. and M. Sim (2004). The price of robustness. *Operations research 52*(1), 35–53.

Bertsimas, D., M. Sim, and M. Zhang (2018b). Adaptive distributionally robust optimization. *Management Science*.

Breton, M. and S. El Hachem (1995). Algorithms for the solution of stochastic dynamic minimax problems. *Computational Optimization and Applications 4*(4), 317–345.

Chen, W. and M. Sim (2009). Goal-driven optimization. *Operations Research 57*(2), 342–357.

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

El Ghaoui, L., F. Oustry, and H. Lebret (1998). Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization 9*(1), 33–52.

Esfahani, P. M. and D. Kuhn (2018). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming 171*(1-2), 115–166.

Federgruen, A. and A. Heching (1999). Combined pricing and inventory control under uncertainty. *Operations research 47*(3), 454–475.

Kunreuther, H. and L. Schrage (1973). Joint pricing and inventory decisions for constant priced items. *Management Science 19*(7), 732–738.

Petruzzi, N. C. and M. Dada (1999). Pricing and the newsvendor problem: A review with extensions. *Operations research 47*(2), 183–194.

Electronic copy available at: https://ssrn.com/abstract=3335889

Prentice, R. L. and R. Pyke (1979). Logistic disease incidence models and case-control studies. *Biometrika 66*(3), 403–411.

Ramachandran, K., N. Tereyağoğlu, and Y. Xia (2018). Multidimensional decision making in operations: An experimental investigation of joint pricing and quantity decisions. *Management Science*.

Rockafellar, R. T. (2015). *Convex analysis*. Princeton university press.

Scarf, H. E. (1957). A min-max solution of an inventory problem. Technical report, RAND CORP SANTA MONICA CALIF.

Shapiro, A., D. Dentcheva, and A. Ruszczyński (2009). *Lectures on stochastic programming: modeling and theory*. SIAM.

Shapiro, A. and A. Kleywegt (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software 17*(3), 523–542.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business 39*(1), 119–138.

Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management 21*(1), 49–58.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research 21*(5), 1154–1157.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Van Parys, B. P., P. M. Esfahani, and D. Kuhn (2017). From data to decisions: Distributionally robust optimization is optimal. *arXiv preprint arXiv:1704.04118*.

Wang, Z., P. W. Glynn, and Y. Ye (2016). Likelihood robust optimization for data-driven problems. *Computational Management Science 13*(2), 241–261.

Wiesemann, W., D. Kuhn, and M. Sim (2014). Distributionally robust convex optimization. *Operations Research 62*(6), 1358–1376.

Zhao, C. and Y. Guan (2018). Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters 46*(2), 262–267.

Electronic copy available at: https://ssrn.com/abstract=3335889