STATISTICAL INFERENCE FOR OPTIMIZATION MODELS: SENSITIVITY ANALYSIS AND UNCERTAINTY QUANTIFICATION

A Dissertation
Presented to
The Academic Faculty

By

Stewart Curry

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology

December 2019

STATISTICAL INFERENCE FOR OPTIMIZATION MODELS: SENSITIVITY ANALYSIS AND UNCERTAINTY QUANTIFICATION

Approved by:

Dr. Nicoleta Serban, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Arkadi Nemirovski
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Pinar Keskinocak
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. V. Roshan Joseph
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Ilbin Lee
Alberta School of Business
University of Alberta

Date Approved: August 23, 2019

To my parents, Sam and Sara Curry. I love you both.

ACKNOWLEDGEMENTS

I am indebted to many individuals for their kindness, knowledge, and support during my graduate studies at Georgia Tech. First, I would like to thank Dr. Nicoleta Serban, for this dissertation would not be possible without her expertise and many hours of advisement. Similarly, this thesis would not be possible without the dedicated efforts of Dr. Ilbin Lee, with whom I worked closely when he was a postdoctoral researcher at Georgia Tech. I would also like to that the members of my dissertation committee, Dr. Nemirovski, Dr. Keskinocak, and Dr. Joseph, for their service, as well as the numerous ISyE faculty members who taught, mentored, and guided me over the past five years.

Prior to coming to Georgia Tech, the faculty members of the mathematics departments at Presbyterian College and Wake Forest University fostered my love of mathematics, encouraged my intellectual curiosity, and challenged me to be a better student. Thank you to each for being wonderful teachers.

Finally, the completion of this dissertation would not have been possible without the love and support of my parents. Words simply cannot express my gratitude and appreciation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In recent years, the optimization, statistics and machine learning communities have built momentum in bridging methodologies across domains by developing solutions to challenging optimization problems arising in advanced statistical modeling. While the field of optimization has contributed with general methodology and scalable algorithms to modern statistical modeling, fundamental statistics can also bring established statistical concepts to bear into optimization. In the operations research literature, sensitivity analysis is often used to study the sensitivity of the optimal decision to perturbations in the input parameters. Providing insights about how uncertain a given optimal decision might be is a concept at the core of statistical inference. Such inferences are essential in decision making because in some cases they may suggest that more data need to be acquired to provide stronger evidence for a decision; in others, they may prompt not making a decision at all because of the high uncertainty of the decision environment. Statistical inference can provide additional insights in decision making by quantifying how uncertainty in input data propagates into decision making.

In this dissertation, we propose a methodological and computational framework for statistical inference on the decision solutions derived using optimization models, particularly, high-dimensional linear programming (LP). In Chapter 2, we explore the theoretical geometric properties of critical regions, an important concept from classical sensitivity analysis and parametric linear programming, and suggest a statistical tolerance approach to sensitivity analysis which considers simultaneous variation in the objective function and constraint parameters. Using the geometric properties of critical regions, in Chapter 3, we develop an algorithm that solves LPs in batches for sampled values right-hand-side parameters (i.e. b of $Ax = b$ in the constraints). Moreover, we suggest a data-driven version of our algorithm that uses the distribution of the bs and empirically compare our approach to other methods on various problem instances. Finally, in Chapter 4, we suggest a unified framework for

statistical inference on the decision solutions and propose the remaining work, including the implementation of the framework to making statistical inferences on spatial disparities in access to dental care services.

CHAPTER 1
INTRODUCTION

In recent years, the optimization, statistics and machine learning communities have built momentum in bridging methodologies across domains by developing solutions to challenging optimization problems arising in advanced statistical modeling [1, 2]. While the field of optimization has contributed with general methodology and scalable algorithms to modern statistical modeling, fundamental statistics can also bring established statistical concepts to bear into optimization.

In deterministic optimization, the input parameters specifying the constraints or the objective function are not exact physical parameters; commonly, they are estimates derived from data about systems' behavior. Deterministic optimization ignores the uncertainty in the input parameters, assuming they are fixed; thus the derived decision solution is also fixed. Stochastic programming and robust optimization handles data uncertainty by finding an optimal solution that accounts for the uncertainty in the input parameters [3, 4, 5, 6]. A decision is made prior to the realization of random parameters in such a way to perform well on average, or to guarantee a worst-case performance. However, the derived optimal decision is still assumed to be one "best" decision.

In the operations research literature, sensitivity analysis is often used to study the sensitivity of the optimal decision to perturbations in the input parameters. Classical sensitivity methods quantify uncertainty due to variations in input parameters either marginally for each parameter [7, 8, 9] or for a small subset of parameters [10, 11, 12]. Given either the distribution of uncertain parameters or a sample from the distribution, Wagner [13] proposed apportioning the variance of the optimal objective value to each input parameter as a measure of sensitivity. This dissertation proposal goes beyond existing sensitivity analysis approaches, focusing on statistical inference for high-dimensional LP models.

Providing insights about how uncertain a given optimal decision might be is a concept at the core of statistical inference. Insights on the uncertainty of a decision can come in various forms in statistical inference, including an estimate of the risk of making the decision (e.g., the standard error of an estimator), a measure of the plausibility of the decision (e.g., p-values in hypothesis testing), or a subset of plausible decisions (e.g., confidence set estimation). Such inferences are essential in decision making because in some cases they may suggest that more data need to be acquired to provide stronger evidence for a decision; in others, they may prompt not making a decision at all because of the high uncertainty of the decision environment [14]. Statistical inference can provide additional insights in decision making by quantifying how uncertainty in input data propagates into decision making [15, 13].

In particular, a statistical inference framework is needed in the estimation of access to healthcare services. Optimization modeling has been implemented in the study of access [16, 17, 18, 19], but estimates of access measures can be sensitive to the input parameters [20]. Because such estimates may greatly influence decision makers seeking to target interventions reducing disparities with limited resources, a rigorous framework for studying access is needed to support informed policy making.

In this dissertation proposal, we propose a methodological and computational framework for statistical inference on the decision solutions derived using optimization models, particularly, high-dimensional linear programming (LP). In Chapter 2, we explore the theoretical geometric properties of critical regions, an important concept from classical sensitivity analysis and parametric linear programming. Furthermore, we suggest a statistical tolerance approach to sensitivity analysis, which considers simultaneous variation in the objective function and constraint parameters and illustrate its usefulness using an inventory management problem. Using the geometric properties of critical regions, in Chapter 3, we develop an algorithm that solves LPs in batches for sampled values right-hand-side parameters (i.e. b of $Ax = b$ in the constraints). Moreover, we suggest a data-driven version of

our algorithm that uses the distribution of the bs and empirically compare our approach to other methods on various problem instances. Finally, in Chapter 4, we suggest a unified framework for statistical inference on the decision solutions and propose the remaining work, including the implementation of the framework to making statistical inferences on spatial disparities in access to dental care services.

CHAPTER 2
GLOBAL SENSITIVITY ANALYSIS VIA A STATISTICAL TOLERANCE APPROACH

Sensitivity analysis in optimization modeling studies the stability of optimal solutions in the presence of uncertain input parameters. In this chapter, we consider the problem of global sensitivity analysis for linear programs when the sensitivity is quantified with respect to multiple input parameters simultaneously. A first contribution is the study of geometric properties of critical regions, which are subsets of values of the input parameters sharing the same optimal basis, with emphasis given to the case where the input parameters in both the objective function and the constraints vary jointly. While the theoretical properties have been stated in the parametric programming literature, formal proofs have not been provided to the best of our knowledge. A second contribution is a statistical tolerance approach to sensitivity analysis, which considers simultaneous variations in the objective function and constraint parameters (jointly called the RIM parameters), informed by the distribution of the input parameters. The approach is extended to the generalized tolerance case where the maximum regret is used as a measure of global sensitivity. In both cases, the tolerance regions obtained can be considered confidence sets for the input parameters. We conclude with a brief application of the proposed approach to sensitivity analysis for an inventory management problem.

2.1 Introduction

Variations in the input parameters of (linear) optimization models and their impact on the optimal outputs have been extensively studied in the operations research literature through the paradigms of sensitivity analysis and parametric programming. In general, sensitivity analysis studies small, local perturbation of parameter values, while parametric program-

ming aims to completely characterize the optimal objective value and optimal solutions as functions over the parameter space.

There are broadly two classes of classical sensitivity measures in the literature: deterministic (including classical sensitivity analysis) and probabilistic (e.g., [21]). Summaries of various approaches to deterministic sensitivity analysis can be found in [22] and [23]. Deterministic methods study the stability of an optimal solution with respect to changes in the input parameters of a linear program by determining bounds on the input parameters within which the optimal basis is maintained. The classical deterministic sensitivity analysis, studying sensitivity with respect to one input parameter at a time, provides limited insights for complex systems since it does not capture joint perturbations of uncertain parameters ([24, 25]). Several deterministic methods have been developed to examine perturbations of multiple input parameters simultaneously; notably, the 100% rule ([26]) and the tolerance approach ([27, 12]) allow for sensitivity analysis due to simultaneous variability in input parameters in the objective function and in the right hand sides (RHS) of constraints, which are jointly called the *RIM parameters*. These methods, along with classical sensitivity analysis, have the benefit of being easily interpretable by practitioners, and of requiring relatively small computational effort. However, all deterministic methods suffer from the fact that they make deterministic statements about the variation of the outputs over certain, pre-determined sets of inputs, taking into account neither how likely each combination of input values is nor how variations of a single input parameter may affect variations of other parameters.

In contrast, probabilistic sensitivity analysis assumes probabilistic models for the input parameters, viewed as realizations of random variables, to derive insights into the model behavior, in particular, how the variations in the input parameters propagate into the variations in the optimal solution. Some early works in stochastic programming literature studied the related problem of describing the distribution of the optimal value given distributional assumptions on the input parameters ([28, 29, 30, 31, 32]). A seminal work in probabilistic

sensitivity analysis was by [13], which used sampling to quantify how much of the total variability in the optimal value is attributed to each input parameter. There have been recent works related to the method introduced by [13], such as [33], which derived a closed-form formula of Wagner's sensitivity measure for a special case and [34], which took a Bayesian view similar to Wagner's framework. However, the number of samples required for reliable estimation increases exponentially in the number of random parameters. In addition, sampling-based approaches do not apply to the use of parametric programming in many applications, for example, for the predictive control model, which is another motivating application of this paper.

Recent research in deterministic sensitivity analysis has also explored ways to incorporate distributional information of the random input parameters. [35] considered functional relationships between parameters. In a subsequent paper, [36] used principal component analysis (PCA) to determine linear functions describing relationships between random input parameters for the case where RIM parameters are correlated with each other. However, these works focused on the marginal effect on the optimal value of perturbing one parameter at a time and are limited to perturbations for which an optimal basis does not change.

In this paper, we bring the benefits of the two classes of sensitivity analysis into one approach. We revisit the tolerance approach for the case where input RIM parameters follow a multivariate probability distribution. Specifically, we assume that the parameters follow a multivariate normal distribution; however, the approach can be easily modified for any symmetric distribution. Under the proposed tolerance approach, we can derive the probability of input RIM parameters varying within the tolerance region. Thus, our contribution lies in theory and methods for extending the tolerance approach to derive a tolerance region that accounts for the distribution of input RIM parameters, including possible dependence between the parameters. We present the applicability of this approach within a statistical modeling framework. Specifically, the input parameters may be estimators or forecasts obtained using uncertain data. Given the so-called sampling distribution of the estimated or

forecast input parameters, the tolerance region becomes a confidence set for the uncertain input parameters. The input random parameters can also be assumed to follow a distribution, for example, estimated or specified using prior data. Given the distribution of the input parameters, the confidence level of the tolerance region can be used to infer how much of the uncertainty in the input parameters is covered and we propose a method to characterize optimal solutions for all parameter values in the tolerance region by searching over a subset of critical regions, where a *critical region* of a basis is defined as the set of input parameters for which the basis is optimal; for a formal definition, see Section 2.4. Thus this paper addresses an important limitation of sensitivity analysis, specifically, determining how stable (in a probabilistic sense) the optimal basis and optimal value are when the input parameters follow a statistical distribution.

When estimates of RIM parameters have high uncertainty, a challenge in the proposed approach is the need to find critical regions covering the tolerance region. To address this challenge, it is necessary to understand the geometry of critical regions in the parameter space. In this paper, we contribute to the existing theory of parametric programming by establishing additional results, in particular, extending the theoretical study of critical regions in our previous work ([37]) to the case where all RIM parameters (i.e., both the RHS of constraints and the objective coefficients) may vary. Variations of RIM parameters were studied by [11] and [38], but geometric properties of critical regions in the RIM parameter space have not been formally studied to the best of our knowledge.

We also describe how the proposed framework can be applied to model predictive control (MPC). Instead of solving a control optimization problem in real time when the input parameters are observed, MPC solves the optimization problem for a set of input parameters in advance, thereby reducing the online computational effort ([39, 40, 41]). In MPC problems, the input parameters are often forecasts, thus uncertain, following a statistical distribution. Therefore, the proposed tolerance approach can guide determining the set of input parameters to be solved using the distributional information. For example, [42] pre-

sented an optimal power flow problem which includes demand forecasts on the RHS of constraints. After determining a tolerance region, the proposed algorithm helps solving the control optimization problem for all parameter vectors in the region.

The paper is organized as follows. After some preliminaries in section 2, we introduce the PCA tolerance approach in section 3, first focusing on variations of RIM parameters within the initial critical region. In section 4, we explore properties of critical regions and how they relate to one another geometrically. We conclude section 4 by describing an algorithm for finding all critical regions which cover a tolerance region for a pre-specified tolerance. In section 5, we show the applicability of the PCA tolerance approach and the theory of critical regions to global sensitivity analysis and MPC, where the RIM parameters vary beyond the initial critical region. Section 6 concludes the paper.

2.2 Preliminary Notation and Concepts

The linear program problem in standard form (P) is defined as

$$\begin{aligned} \text{(P)} \ \min_{x} \ & c^T x \\ \text{s.t. } & Ax = b, \\ & x \geq 0, \end{aligned}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, c and $x \in \mathbb{R}^n$, and A has full row rank. The dual problem is

$$\begin{aligned} \text{(D)} \ \max_{y,s} \ & b^T y \\ \text{s.t. } & A^T y + s = c, \\ & s \geq 0. \end{aligned}$$

We refer to c as the objective function parameters and b as the right-hand-side (RHS) parameters, while referring to b and c jointly as the *RIM parameters* as in the parametric

programming literature ([11]). We denote the RIM parameters jointly by $r = [b^T c^T]^T$. Although b and c are vectors, we denote $r = (b, c)$ with a slight abuse of notation for simplicity. Let ($P^{b,c}$) denote the LP (P) with specific values of b and c. We define ($D^{b,c}$) similarly.

A *basis* B is a subset of indexes $\mathcal{I} = \{1, \ldots, n\}$ such that the corresponding columns of A are linearly independent and $|B| = m$. Correspondingly, $N = \mathcal{I} \setminus B$ is the set of nonbasic indexes. For a matrix M, M_B denotes the matrix comprised of the columns of M corresponding to the index set B and for a vector v, v_B denotes the sub-vector corresponding to B. M_B^T denotes the transpose of the sub-matrix M_B and if M_B is invertible, M_B^{-T} denotes $(M_B^{-1})^T = (M_B^T)^{-1}$. A basis B is optimal for a given problem ($P^{b,c}$) if and only if the following conditions are met:

$$A_B^{-1} b \geq 0, \tag{2.1}$$

$$c_N^T - c_B^T A_B^{-1} A_N \geq 0. \tag{2.2}$$

For the primal problem ($P^{b,c}$), the first condition (2.1) is known as the *feasibility condition* and the second condition (2.2) is known as the *optimality condition*.

For a matrix M, $M_{(i,j)}$ is the entry in the ith row and jth column of M. Also, $M_{(\cdot,j)} \in \mathbb{R}^{m \times 1}$ and $M_{(i,\cdot)} \in \mathbb{R}^{1 \times n}$ is the jth column and ith row of M, respectively. We will use I to denote an identity matrix. For a vector v, we denote the ith entry by v_i, without using parentheses.

A k-dimensional random vector R following a *multivariate distribution*, $\mathcal{F}$, with mean vector μ and covariance matrix Σ is denoted as $R \sim \mathcal{F}(\mu, \Sigma)$. In this paper, we only consider multivariate distributions for which the expectation and covariance matrix exist. We let $R = (B, C)$ denote the random input parameters and $\hat{r} = (\hat{b}, \hat{c})$ denote realizations of the random parameters. In general, the individual components of R may not be independent, with the covariance of R_i and R_j given by $\Sigma_{(i,j)}$. Of particular interest in

this paper will be the case where $\mathcal{F}$ is the multivariate normal distribution, for which we will write $R \sim \mathcal{N}(\mu, \Sigma)$. In this case, the R vector may be represented as a transformation of independent normal random variables, i.e., $R = AZ + \mu$, where the elements of Z are independent, standardized normal random variables and A is an appropriately chosen matrix.

2.3 PCA Tolerance Approach

In this section, we introduce a tolerance approach incorporating distributional information of the RIM parameters. Here we take a statistical modeling perspective; specifically, we assume the RIM parameters are specified using uncertain data, thus they can be viewed as random variables, denoted by R, with realizations from this distribution denoted by $\hat{r}$. For example, they can be expected costs; in this case, uncertainty is specified by the so called sampling distribution of the estimates. The uncertainty in the RIM random parameters can also be specified by an estimated distribution based on prior knowledge about the behavior of the random parameters, such as forecasts of demand where the distribution of demand is estimated from previous sales. The estimate or forecast, $\hat{r}$, is used as the input parameter to the optimization problem, thus specifying a particular linear program to be solved. We will call this estimate the "baseline value" and examine how deviations from the baseline affect the stability of the solution. For simplicity of illustration of the overall approach, we assume that the sampling or estimated distribution follows a multivariate symmetric distribution, however, this could be restrictive in some cases.

The tolerance approach ([27, 12]) is a sensitivity analysis tool for examining perturbations in the RIM parameters. The maximum tolerance is defined as the largest percentage within which the RIM parameters may vary simultaneously and independently from the baseline values while the optimal basis is maintained. The tolerance region is the set of input parameters within the maximum tolerance of perturbation, which is a hyperbox symmetric around some baseline input parameters. Thus this is a deterministic approach to

sensitivity analysis. Several extensions of the tolerance approach exist ([43, 44, 45, 46, 47, 48]), mainly focused on relaxing the symmetry condition of the tolerance region in order to enlarge the region. One particular extension is mentioned briefly on page 12 of Ward and Wendell (1990), allowing for deterministic relationships between the parameters. This idea is of particular interest to this paper as we extend this idea to account for stochastic relationships between parameters.

2.3.1 PCA Tolerance Region

The tolerance approach introduced in this paper utilizes distributional information of input parameters in the following way. Consider a standard form LP (P) where the RIM input random parameters follow a multivariate symmetric distribution, i.e., $\hat{r} = (\hat{b}, \hat{c})$ is a realization from $\mathcal{F}(\mu, \Sigma)$. We assume that the covariance matrix Σ is known, or well approximated. For example, if R is a sample mean from a random sample of size k, we can approximate Σ by $k^{-1}S$, where S is the sample covariance matrix of the data. For cases when the uncertainty is specified by an estimated distribution, then the distribution parameters are specified using prior information. For generality of the description of the PCA tolerance approach we assume Σ is known although in the first case, we replace Σ by the sample covariance. Under this setting we seek to evaluate how perturbations from the (unknown) expected RIM parameters impact the optimal value and optimal solution.

Using the eigenvalue decomposition, we can write $\Sigma = D\Lambda D^T$, where D is a matrix with eigenvector columns and Λ is a diagonal matrix of eigenvalues. In order to consider variations from the RIM input parameters $\hat{r}$, we re-parameterize the RIM vector as $\hat{r} + D\Lambda^{1/2}p$, where $p \in \mathbb{R}^{m+n}$. The resulting RIM multiparametric linear program is

$$\begin{aligned} &\min_x \ (c + S_c p)^T x \\ &\text{s.t. } Ax = b + S_b p, \\ &x \geq 0, \end{aligned}$$

where S_b and S_c are the first m rows and final n rows of $D\Lambda^{1/2}$, respectively. Note that this is the original problem ($\mathrm{P}^{b,c}$) for $p = 0$.

Decomposing the covariance matrix as provided above is similar to the approach used in principal component analysis (PCA). PCA is a multivariate statistical method for identifying an orthogonal linear basis that most meaningfully describes a data set. A key benefit of finding such an orthogonal basis is that the eigenvectors define a set of directions along which the RIM parameters vary independently, while eigenvalues represent the variance along each eigenvector. Note that the eigenvectors are scaled to unit length, so that the product $D\Lambda^{1/2}$ is the matrix of eigenvectors scaled according to the standard deviation along each direction. If some of the eigenvalues are zeros, then the parameter vector varies in a lower-dimensional affine set of dimension h, where h is the number of nonzero eigenvalues. We can then re-define D and Λ as $(m+n)\times h$ and $h\times h$ matrices, respectively, using only the nonzero eigenvalues and the corresponding eigenvectors. From this point forward, we will assume all eigenvalues are strictly positive for brevity of presentation. We also note that if the multivariate symmetric distribution is an approximation of the uncertainty, then the independence between directions is also approximate.

Given the baseline RIM parameter $\hat{r}$, we define the *tolerance region* $V(\tau) = \{\hat{r} + D\Lambda^{1/2}p : \|p\|_\infty \leq \tau\}$ for a finite, nonnegative number τ, known as a *tolerance*. The baseline parameters can be estimators or forecasts of the input parameters as well as values reflecting some prior believe on the behavior of the inputs of interest. The tolerance region is then a hyperbox, centered at $\hat{r}$, that scales symmetrically in τ. The tolerance τ represents the maximum number of standard deviations that $\hat{r} + D\Lambda^{1/2}p \in V(\tau)$ falls away from the baseline parameter, along the directions provided by the eigenvectors. Since variations along each of the eigenvectors are independent, the likelihood of the expected RIM parameters μ belonging in the tolerance region can be easily related to the tolerance τ.

We provide two propositions relating the tolerance τ to the confidence level of the tolerance region as a confidence set for μ in the case where R follows a multivariate normal

distribution, i.e. $R \sim \mathcal{N}(\mu, \Sigma)$. The first proposition explicitly relates the tolerance τ to the probability of the tolerance region covering μ, i.e., $\mathbb{P}(\mu \in V(\tau))$. The second proposition provides guidance for finding the largest ellipse contained in a tolerance region, since confidence regions obtained from multivariate normal distributions are often given as an ellipse, rather than a hyperbox.([**Chew**]) We show that similar results can be obtained for other symmetric multivariate distributions in Appendix A.1.

Proposition 2.3.1 *The tolerance region $V(\tau)$ is an approximate $(1-\alpha)\times 100\%$ confidence box for $\mu = E[R]$, where $\alpha = 1 - P(|z| \leq \tau)^h$ for $Z \sim \mathcal{N}(0,1)$.*

Proof: $P(\mu \in V(\tau)) = P(\|\Lambda^{-1/2}D^T(R-\mu)\|_\infty \leq \tau)$ since $R + D\Lambda^{1/2}\Lambda^{-1/2}D^T(\mu - R) = \mu$. Since $R \sim N(\mu, \Sigma)$ and $\Lambda^{-1/2}D^T(R-\mu) \sim \mathcal{N}(0, I)$. Then, $P(\|\Lambda^{-1/2}D^T(\mu - R)\|_\infty \leq \tau) = \prod_{i=1}^{h} P(\left|(\Lambda^{-1/2}D^T(\mu - R))_i\right| \leq \tau) = P(|Z| \leq \tau)^h$, where the first equality relies on the fact that variations along each eigenvector are independent. Therefore, the tolerance region $V(\tau)$ is a random set that covers μ with probability $P(|Z| \leq \tau)^h$. □

Proposition 2.3.2 *Let C be a $(1-\alpha)\%$ confidence ellipse for μ given by $C = \{r' : (r' - \hat{r})^T\Sigma^{-1}(r' - \hat{r}) \leq (\tau)^2\}$, where $\alpha = 1 - P(\chi_h^2 \leq \tau)$ and χ_h^2 follows a chi-square distribution with h degrees of freedom. Then $C \subset V(\tau)$.*

Proof: It is well known that C is a $(1-\alpha)\%$ confidence ellipse for μ, where α is defined as above. Let $r' \in C$ be given. Choose $\gamma = \Lambda^{-1/2}D^T(r' - \hat{r})$ so that $(r' - \hat{r}) = D\Lambda^{1/2}\gamma$. Since $r' \in C$, we have $\gamma^T\Lambda^{1/2}D^T\Sigma^{-1}D\Lambda^{1/2}\gamma \leq (\tau)^2 \Rightarrow \gamma^T\gamma \leq (\tau)^2 \Rightarrow \|\gamma\|_2 \leq \tau \Rightarrow \|\gamma\|_\infty \leq \tau$. By definition of τ, $r' = \hat{r} + D\Lambda^{1/2}\gamma \in V(\tau)$. □

2.3.2 PCA Tolerance Region Maintaining Optimal Basis

Following Wendell's approach, we first consider the PCA tolerance region that maintains an optimal basis. We denote the corresponding tolerance level as τ^* and it can be ob-

tained by solving an optimization problem $\tau^* = \max_\tau\{\tau : \|p\|_\infty \leq \tau \Rightarrow (b, c) = \hat{r} + D\Lambda^{1/2}p, A_B^{-1}b \geq 0, c_N^T - c_B^T A_B^{-1} A_N \geq 0\}$. Lemma A.2 of [12] provides a way for deriving a closed-form formula for τ^*.

$$\tau^* = \min\left\{\min_{k=1,\ldots,m}\left\{\frac{(A_B^{-1})_{(k,\cdot)}b}{\sum_{j=1}^h \left|((A_B^{-1})_{(k,\cdot)}S_b)_j\right|}\right\}, \min_{k=1,\ldots,n-m}\left\{\frac{((I_N - I_B A_B^{-1}A_N)^T)_{(k,\cdot)}c}{\sum_{j=1}^h \left|((I_N - I_B A_B^{-1}A_N)^T)_{(k,\cdot)}S_c)_j\right|}\right\}\right\} \tag{2.3}$$

This approach offers several advantages to the original tolerance approach. First, in the proposed approach, the maximum tolerance can be directly related to the likelihood of such variations using Propositions 2.3.1 and 2.3.2. Second, while the original tolerance approach only considers independent variations of the input parameters, our approach allows for possible dependencies in the data.

An important limitation of this tolerance region is that it is limited to those input parameters that maintain an optimal basis of the baseline parameter. When a desired tolerance value (in other words, a desired coverage of the distribution) does not allow the corresponding tolerance region to maintain the same optimal basis, one needs a computational procedure to find optimal solutions for all input parameters in the tolerance region. The next section establishes theoretical results that are necessary for developing such a method. Based on these theoretical results, we present an algorithm for identifying optimal solutions of all input parameters within the tolerance region.

We conclude this section with an example illustrating the PCA tolerance region introduced in this section (maintaining the optimal basis). This example is modified from the one considered by [49] and [12]. Consider the following LP

$$\begin{aligned}
\min_x \ & -12x_1 - 20x_2 - 18x_3 - 40x_4 \\
\text{s.t. } & 4x_1 + 9x_2 + 7x_3 + 10x_4 + x_5 = B_1 \\
& x_1 + x_2 + 3x_3 + 40x_4 + x_6 = B_2 \\
& x \geq 0,
\end{aligned}$$

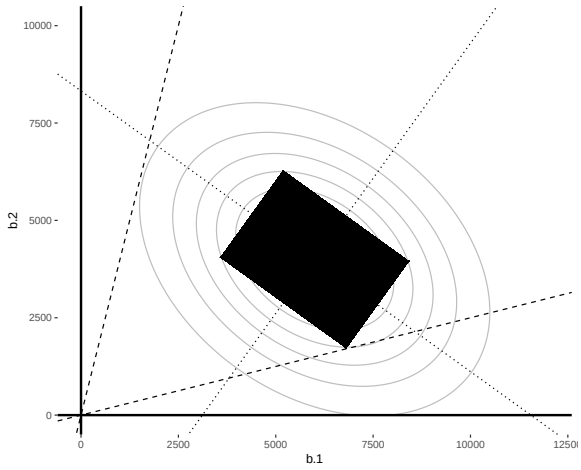


Figure 2.1: PCA Tolerance illustrated: critical region, contour plot and tolerance region

where $B \sim \mathcal{N}(\mu, \Sigma)$ and

$$\Sigma = \begin{bmatrix} 5000000 & -1500000 \\ -1500000 & 4000000 \end{bmatrix}.$$

For estimated values $\hat{b}_1 = 6000$ and $\hat{b}_2 = 4000$, the optimal basis is $B = \{1, 4\}$ with

$$A_B^{-1} = \begin{bmatrix} \frac{4}{15} & \frac{-1}{15} \\ \frac{-1}{150} & \frac{2}{75} \end{bmatrix}.$$

By (2.1), the set of parameters b's for which B is optimal is $\{b : A_B^{-1} b \geq 0\}$, shown in Figure 2.1 by the area between the two dashed lines. A contour plot of the bivariate normal likelihood, centered at the baseline values $\hat{b}_1 = 6000, \hat{b}_2 = 4000$ is overlaid. Notice that the negative covariance between B_1 and B_2 implies that the two values tend to vary in opposite directions. The tolerance region is shown by the rectangle, oriented along the directions of the eigenvectors of the covariance matrix, shown by the two dotted lines. The maximum tolerance is $\tau^* = 0.81$, which means that the parameter vector may vary

up to 0.81 standard deviations simultaneously along the direction of each eigenvector and still maintains the optimal basis. Using Proposition 2.3.1, we determine that the tolerance region covers the (unknown) expected input parameters, μ, with $P(|Z| \leq \tau^*)^2 = 34\%$ confidence.

2.4 Theory and Method: Critical Regions and Neighbors

Given a pre-specified tolerance value, the corresponding tolerance region may span several *critical regions*, where a critical region is defined as a set of (b, c) vectors for which a certain basis is optimal. The task of finding optimal bases and solutions for each $r \in V(\tau)$ can be reduced to finding a set of critical regions that cover $V(\tau)$. To provide a theoretical base for the exploration of the tolerance region introduced in the previous section, this section examines the geometric properties of critical regions and their *neighbors*. After introducing some notation, we review the theoretical properties of critical regions in the space of the RHS (b in (P)), and then establish similar results for critical regions in the space of objective coefficients (c in (P)). Based on these properties, we study geometric properties of critical regions in the space of (b, c), i.e., the RIM parameters. Throughout, we give particular attention to the way in which critical regions relate to each other and delineate the parameter space. Lastly, we provide an algorithm for finding critical regions that cover the tolerance region for a given tolerance.

2.4.1 Critical Regions: Definitions and Review

For a basis B, *critical regions* Rb^B, Rc^B, and R^B are defined as follows

$$Rb^B = \{b \in \mathbb{R}^m \mid A_B^{-1} b \geq 0\}; \tag{2.4}$$

$$Rc^B = \{c \in \mathbb{R}^n \mid c_N^T - c_B^T A_B^{-1} A_N \geq 0\} = \{c \in \mathbb{R}^n \mid (I_N - I_B A_B^{-1} A_N)^T c \geq 0\}; \tag{2.5}$$

$$R^B = Rb^B \times Rc^B = \{r = (b, c) \in \mathbb{R}^{m+n} \mid \mathbb{B}r \geq 0\} \tag{2.6}$$

where

$$\mathbb{B} = \begin{bmatrix} A_B^{-1} & 0 \\ 0 & (I_N - I_B A_B^{-1} A_N)^T \end{bmatrix} \tag{2.7}$$

.

Rb^B is the set of b's for which B satisfies (2.1), Rc^B is the set of c's for which B satisfies (2.2), and R^B is the set of (b, c)'s for which B is optimal for (P^{b,c}). We refer to Rb^B and Rc^B as a *critical region in* b and a *critical region in* c, respectively. From this point forward, we reserve the term *critical region*, without reference to b or c, for R^B. A basis B is an *optimal basis* if there exists (b, c) for which B is optimal for (P^{b,c}), i.e., R^B is nonempty. Definitions (2.4), (2.5), and (3.3) provide that critical regions, in b, in c, or in both are polyhedral cones pointed at the origin. The dimensionality of Rb^B, Rc^B, and R^B are m, n, and $m + n$, respectively. For a cone defined as $\{v \in \mathbb{R}^k : Mv \geq 0\}$, where $M \in \mathbb{R}^{p \times q}$, we refer to the facet defined by the jth inequality $M_{(j,\cdot)}v \geq 0$ as the *jth facet* of the cone. In parametric programming literature, there are definitions of critical region that are different from the above ([50, 51, 47]), for example, as the set of parameter vectors for which the same set of constraints are binding. This paper focuses on the above definitions (2.4), (2.5), and (3.3).

The following well-known result illustrates why the concept of critical region is useful in analyzing variations in the LP RIM parameters. Further discussions about the below result are in [11, 23]. Let $v(b, c)$ denote the optimal value of (P^{b,c}).

Theorem 2.4.1 *[cf. Sections 2.4 and 3.1 of [52]] (a) The optimal value function $v(b, c)$ is piecewise quadratic and continuous;*

(b) the optimal value function $v(b, c)$ is convex in b and concave in c;

(c) in each critical region R^B, $v(b, c)$ is quadratic and there exists an optimal solution

function $x^\star(b,c) \in X^\star(b,c)$ *that is linear in* R^B*, namely, for* $(b,c) \in R^B$*,*

$$v(b,c) = c_B^T A_B^{-1} b \tag{2.8}$$

and

$$x^*(b,c)_B = A_B^{-1} b,\ x^*(b,c)_N = 0. \tag{2.9}$$

The following additional notation about simplex pivots are useful in our theoretical discussion. We assume throughout that while performing simplex pivots, rows of the simplex tableau are not re-ordered. That is, the ordering of the basic and nonbasic variables given by the tableau remain the same during a pivot operation, and only the entering and exiting variables swap positions. To track the ordering of the basic and nonbasic variables, we will use the following notation. For a basis B, let $B(\cdot)$ be the bijective mapping from the set B to the set $\{1,\ldots,m\}$ which returns the row index in the tableau corresponding to each basic variable. Similarly, let $N(\cdot)$ be the bijective mapping from the set N to $\{1,\ldots,n-m\}$ which gives an ordering to the nonbasic variables, returning the column index in the reduced tableau corresponding to each nonbasic variable. If B_2 is a basis obtained from a basis B_1 by exchanging $i \in B_1 \setminus B_2$ for $j \in B_2 \setminus B_1$, then the following properties hold: (1) $B_1(i) = B_2(j)$; (2) $N_1(j) = N_2(i)$; (3) $B_1(k) = B_2(k)$ for $k \in B_1 \cap B_2$; and (4) $N_1(k) = N_2(k)$ for $k \in N_1 \cap N_2$. We will denote the pivoting element of a pivot that exchanges an exiting variable i for an entering variable j as $\bar{a}_{ij}$.

2.4.2 Geometry of Critical Regions in b

Definitions In the case where only b varies but c is fixed, two bases B_1 and B_2 are said to be *dual neighbors* if and only if it is possible to pass from B_1 to B_2 by a dual simplex pivot and vice versa. Two critical regions in b are defined to be *dual neighbors* if their bases are dual neighbors. Two critical regions in b are said to be *geometric neighbors* if and only if they share a facet.

In [37], properties of critical regions in b were studied while assuming A and c fixed.

Theorem 2.4.2 *[cf. Theorem 2 and 4 in [37]] Two critical regions in b are dual neighbors if and only if they are geometric neighbors.*

2.4.3 Geometry of Critical Regions in c

Definitions In the case where only c varies, but b is fixed, two bases B_1 and B_2 are said to be *primal neighbors* if and only if it is possible to pass from B_1 to B_2 by a primal simplex pivot and vice versa. Two critical regions in c are *primal neighbors* if their bases are primal neighbors. The definition of geometric neighbor remains the same: two critical regions in c are *geometric neighbors* if they share a facet.

The fact that c is the RHS of the dual problem ($D^{b,c}$) suggests that all of the results given in [37] may be extended to the case where c varies while A, b are fixed. We provide a full exploration of the properties of critical regions in c in Appendix A.2. As a consequence of Theorems A.2.1 and A.2.2, we have the following corollary, analogous to Theorem 2.4.2.

Corollary 2.4.3 *Two critical regions in c are primal neighbors if and only if they are geometric neighbors in c.*

2.4.4 Geometry of Critical Regions

Recall that we refer to R^B defined in (3.3) simply as a *critical region* without any reference to b or c. Given a basis, its *critical region* is the Cartesian product of its critical regions in b and c. The results in [37] show that two bases are dual neighbors if and only if their critical regions in b are geometric neighbors. Also, from the previous section, we know that two bases are primal neighbors if and only if their critical regions in c are geometric neighbors. In this section, we assume that neither b nor c is fixed and study the geometry of critical regions. Previously, we defined primal and dual neighbors where either a fixed c or b is given, respectively. As neither b nor c is fixed in this section, we re-define primal and dual

neighbors. We will keep using the same terms, but it will be clear from the context which definition is used.

Definition Two bases B_1 and B_2 are *primal neighbors* if and only if it is possible to pass from B_1 to B_2 by a primal simplex pivot for some fixed $b \in Rb^{B_1}$ and vice versa. Similarly, B_1 and B_2 are *dual neighbors* if and only if it is possible to pass from B_1 to B_2 by a dual simplex pivot for some fixed $c \in Rc^{B_2}$ and vice versa. Finally, B_1 and B_2 are *neighbors* if and only if they are either primal or dual neighbors.

We emphasize the difference between the definitions of neighbors in the previous section and in this section. In the previous section, in a primal pivot, any nonbasic variable with a positive entry in the corresponding column of the simplex tableau may be chosen to enter the basis and the leaving variable is determined by the minimum ratio test for the fixed b value. Recall that the purpose of the minimum ratio test in the simplex method is to determine the exiting (entering) variable in such a way as to ensure that the primal (dual) simplex pivot maintains primal feasibility (optimality). However, in the above definition, the choice of a leaving variable in a primal pivot may vary depending on the value of b. In other words, after choosing a nonbasic variable to enter the basis, multiple neighbors may be reached depending on the b value used for the minimum ratio test. Later, we will demonstrate which neighbors are reachable. We start with a basic lemma.

Lemma 2.4.4 *Two critical regions cannot be both primal and dual neighbors.*

The proof follows from the fact that primal pivoting elements are positive, while dual pivoting elements are negative.

We now begin our exploration of geometric properties of critical regions. Let B_1 and B_2 be two bases differing in only one index and suppose that there is a pivot operation (primal or dual) that exchanges variable $i \in B_1 \setminus B_2$ for variable $j \in B_2 \setminus B_1$. It is clear from the proofs of Theorems 2.4.2 and A.2.1 that a facet is shared by the pair of critical regions in b, Rb^{B_1} and Rb^{B_2}, and the pair of critical regions in c, Rc^{B_1} and Rc^{B_2}. A key observation is that the sign of the pivoting element determines whether these critical regions in b and c lie

on the same side or opposing sides of the corresponding facet. If the pivot is primal, then the pivot element is positive, Rc^{B_1} and Rc^{B_2} lie on opposing sides of a shared facet, and Rb^{B_1} and Rb^{B_2} lie on the same side of a shared facet. This case is illustrated in Figure 2.2. For a dual pivot, the pivot element is negative and the critical regions in c and b are oriented conversely.

(a) Rb^{B_1} (red) and Rb^{B_2} (blue) (b) Rc^{B_1} (red) and Rc^{B_2} (blue)

Figure 2.2: Critical Regions before and after Primal Simplex Pivot

If R^{B_1} and R^{B_2} are neighbors, then the pivoting element is nonzero and either the pair Rb^{B_1} and Rb^{B_2}, or the pair Rc^{B_1} and Rc^{B_2} is separated by a hyperplane in their corresponding space. This implies that R^{B_1} and R^{B_2} are separated by a hyperplane in $\mathbb{R}^{n+m}$. We formalize the geometric relationship between neighbors with the concept of partially shared facets.

Definition Two critical regions, R^{B_1} and R^{B_2}, *partially share a facet* if and only if $R^{B_1} \cap R^{B_2}$ is contained in a facet of both R^{B_1} and R^{B_2} and has dimension $n + m - 1$.

The following theorem is an extension of Theorem A.2.1 to the RIM case.

Theorem 2.4.5 *If two critical regions are neighbors, then they partially share a facet.*

Proof: In A.4

A second key observation is that in a primal (dual) pivot the choice of a leaving (entering) variable determined by the minimum ratio test depends on the value of b (c) used. The

next lemma states that at a given basis B, for any positive (negative) entry in the tableau, there exists a value of b (c) at which the entry is chosen to be the pivoting element for a primal (dual) pivot, i.e., the entry wins the minimum ratio test.

Lemma 2.4.6 *Let B be a basis. We then have the following results:*

1. *Let $j \in N$ be a nonbasic variable selected to be the entering variable in a primal pivot. Then, for every positive entry in the jth column of the tableau, $(A_B^{-1}A)_{(\cdot,j)}$, there exists $b \in Rb^B$ such that the corresponding basic variable wins the minimum ratio test.*

2. *Let $i \in B$ be a basic variable selected to be the exiting variable in a dual pivot. Then, for every negative entry in the $B(i)$th row of the tableau, $(A_B^{-1}A)_{(B(i),\cdot)}$, there exists $c \in Rc^B$ such that the corresponding nonbasic variable wins the minimum ratio test.*

Proof: We prove the first statement only as the second statement follows analogously. Let $j \in N$ be the entering variable for a primal pivot. Recall that for a given b, the minimum ratio test selects the exiting variable

$$\arg\min_{i \in B}\{\frac{(A_B^{-1}b)_{B(i)}}{(A_B^{-1}A)_{(B(i),j)}} : (A_B^{-1}A)_{(B(i),j)} > 0\}.$$

Now suppose $(A_B^{-1}A)_{(B(k),j)} > 0$ for some $k \in B$. It is clear that a nonnegative $z \in \mathbb{R}^m$ may be selected such that

$$k = \arg\min_{i \in B}\{\frac{z_{B(i)}}{(A_B^{-1}A)_{(B(i),j)}} : (A_B^{-1}A)_{(B(i),j)} > 0\}$$

since the $B(k)$th entry of z can be made arbitrarily small. Let $b^* = A_B z$. Since $A_B^{-1}b^* = z \geq 0$, we have $b^* \in Rb^B$. By construction, it is clear that at b^*, k wins the minimum ratio test, and this concludes the proof. $\square$

Using this lemma, we show below that a critical region may share a facet not only with one neighbor, but possibly with multiple neighbors. This result formally establishes the geometric relationship between neighboring critical regions in the RIM parameter space. Recall that in the definition of a critical region R^B in (2.7), the first m facets are on its projection on the space of b (i.e., Rb^B) and the next $n-m$ facets define its projection on the space of c (i.e., Rc^B).

Theorem 2.4.7 *Let F be the wth facet of R^B. Then, R^B shares the facet F with $R^{B_1}, \ldots, R^{B_k}$ and we have $F = \bigcup_{i=1}^{k}(R^B \cap R^{B_i})$, where $B_1, \ldots, B_k$ is the set of bases obtained as follows:*

1. *if $w \in \{1, \ldots, m\}$, then $B_1, \ldots, B_k$ are the bases obtained by a dual pivot replacing the wth basic variable by any nonbasic variable with a negative entry on the wth row of the simplex tableau at B,*

2. *if $w \in \{n-m, \ldots, n\}$, then the bases are obtained by a primal pivot selecting the $(w-m)$th nonbasic variable to enter the basis and dropping any basic variable with a positive entry on the $(w-m)$th nonbasic column of the simplex tableau at B.*

Proof: We prove the case of $w \in \{1, \ldots, m\}$ only as the case of $w \in \{m+1, \ldots, n\}$ follows analogously. We have

$$F = R^B \cap \{(b,c) \in \mathbb{R}^{m+n} : (A_B^{-1})_{(w,\cdot)} b = 0\} = (Rb^B \cap \{b \in \mathbb{R}^m : (A_B^{-1})_{(w,\cdot)} b = 0\}) \times Rc^B,$$

so $F = F_b \times Rc^B$, where F_b is the wth facet of Rb^B. For each B_i which can be obtained by a dual pivot dropping the wth basic variable from B, we have $R^B \cap R^{B_i} = (Rb^B \cap Rb^{B_i}) \times (Rc^B \cap Rc^{B_i}) = F_b \times (Rc^B \cap Rc^{B_i})$, since Rb^{B_i} and Rb^B are geometric neighbors. Note that for any $c \in Rc^B$, one of the B_i will maintain dual feasibility (the minimum ratio test guarantees this), that is, there exists $i \in \{1, \ldots, k\}$ such that $c \in Rc^{B_i}$. Thus, $\bigcup_i (Rc^B \cap Rc^{B_i}) = Rc^B$ and therefore, $\bigcup_i (R^B \cap R^{B_i}) = \bigcup_i F_b \times (Rc^B \cap Rc^{B_i}) =$

$F_b \times Rc^B = F$. $\square$

This result can be interpreted as follows. Consider the wth facet of R^B. If $w \in \{1, \ldots, m\}$, then the facet is partitioned by all of the critical regions that can be obtained by performing a dual pivot on a negative entry in the wth row of $A_B^{-1}A_N$. If $w \in \{m+1, \ldots, n\}$, then the facet is partitioned by all of the critical regions that can be obtained by performing a primal pivot on a positive entry in the $(w-m)$th column of $A_B^{-1}A_N$. Our next result shows that any two critical regions that partially share a facet are either primal neighbors or dual neighbors.

Theorem 2.4.8 *Suppose that R^{B_1} and R^{B_2} are critical regions that partially or fully share a facet. Then, R^{B_1} and R^{B_2} are neighbors.*

Proof: In A.5

As a consequence of Theorems 2.4.5 and 2.4.8, we have the following corollary.

Corollary 2.4.9 *Two critical regions are neighbors if and only if they partially share a facet.*

2.4.5 Covering a Tolerance Region by Critical Regions

Based on our understanding of critical regions and their neighbors developed in the preceding sections, in this section we present an algorithm to find all critical regions covering the tolerance region for a pre-specified tolerance. The tolerance τ may be selected using Propositions 2.3.1 or 2.3.2 for a desired level of confidence.

Algorithm 1 begins with a set $\mathcal{B}$ containing the bases that are optimal for the baseline parameter vector, i.e., optimal for ($P^{\hat{b},\hat{c}}$). It needs the input of the baseline input parameters $\hat{r}$ for the RIM parameters and the covariance structure of the dependence between them. The set $\mathcal{B}$ may contain a single optimal basis or, in the degenerate case, multiple optimal bases. As the tolerance region is enlarged, it begins to intersect critical regions whose

bases are not in $\mathcal{B}$. By our theoretical results, we know that the first intersected critical region is a neighbor of a basis in $\mathcal{B}$. In each iteration, the algorithm finds the first neighbor that is intersected when the current tolerance region expands. To find such a neighbor, the algorithm maintains $\mathcal{N}(\mathcal{B})$, the set of neighbors of bases in $\mathcal{B}$ that are not in $\mathcal{B}$, and finds the first intersected critical region $\bar{B}$. In addition, the algorithm computes the tolerance value of the largest tolerance region that fits into the union of critical regions of bases in $\mathcal{B}$, which is denoted as t. While computing t, it computes t^B for each $B \in \mathcal{N}(\mathcal{B})$, which is the minimum tolerance level at which the tolerance region intersects the critical region of B. Note that each t^B value can be obtained by solving a linear program. Then, it adds $\bar{B}$ to $\mathcal{B}$, updates $\mathcal{N}(\mathcal{B})$, and go to the next iteration. It terminates when the tolerance value t reaches the target tolerance τ. Based on the theoretical results in the previous sections, the algorithm can be seen as growing a polytope by adding critical regions in the order the tolerance region intersects as the tolerance level increases. Recall that D and Λ are the eigenvector matrix and the diagonal matrix of the eigenvalue decomposition of Σ, respectively.

Algorithm 1 Algorithm for Finding Critical Regions Covering a Given Tolerance Region

Input: $\hat{r}$, Σ, and a desired tolerance value $\tau > 0$

Initialize:

Set $\mathcal{B}$ as the set of bases optimal for (P_μ)

$\mathcal{N}(\mathcal{B}) \leftarrow \{B : B \text{ is a neighbor of } B' \in \mathcal{B}\} \setminus \mathcal{B}$

$t \leftarrow 0$

while $t < \tau$ **do**

  Find the Next Critical Region:

  for $B \in \mathcal{N}(\mathcal{B})$ **do**

    Compute $t^B \leftarrow \min_p\{\|p\|_\infty : \mathbb{B}(\hat{r} + D\Lambda^{1/2}p) \geq 0\}$

  $t \leftarrow \min_B\{t^B : B \in \mathcal{N}(\mathcal{B})\}$, $\bar{B} \leftarrow \arg\min_B\{t^B : B \in \mathcal{N}(\mathcal{B})\}$

  Update Sets of Bases:

  $\mathcal{B} \leftarrow \mathcal{B} \cup \{\bar{B}\}$

  $\mathcal{N}(\mathcal{B}) \leftarrow \{B : B \text{ is a neighbor of } B' \text{ for some } B' \in \mathcal{B}\} \setminus \mathcal{B}$

return $\mathcal{B}$.

2.5 Applications

In this section, we present a series of applications of the PCA tolerance approach with the implementation of Algorithm 1 presented in the previous section. First, however, we pause for a general discussion about nature of uncertainty in the RIM parameters and the applicability of the PCA tolerance approach.

There are two major categories of situations where modelers may wish to consider stochastic uncertainty in the RIM input parameters. The first is the case where the RIM input parameter is an estimate of a distributional parameter. Using the sample mean to estimate the expected value is such an example. The second is the case where the RIM input parameter is a realization from the distribution itself. For instance, the RIM input

parameter may reflect real time cost/demand information, viewed as a single realization from the distribution of costs/demands. This second scenario includes forecasting, where historical data is used to approximate the distribution from which future input data are generated.

In post-optimality analysis, after the input parameters are specified and a solution is obtained, the stability of the solution is examined by determining a range of RIM parameters for which the current basis is optimal. Crucially, the decision is made under uncertainty, so an examination of how uncertainty affects the decisions that we have made is important. ([25, 7]) In the case where only c varies, we may go beyond the initial critical region, examining the loss in optimality of the current solution against alternatives. However, in the case where b varies, the solution may become infeasible. Therefore, as a tool for sensitivity analysis, the tolerance approach is most appropriate when b is either fixed or when the uncertainty in b is small, and confined to the initial critical region.

The PCA tolerance approach, however, can also be used as a tool of a priori analysis. In this case, we wish to examine the decisions that we might make in the future, once the input parameters are realized. In other words, any decisions made are made under certainty, once the input data is determined. Such analysis can be used to make decisions auxilary to the optimization problem or save in real time computational costs. For instance, in model predictive control, it is often the case that we wish to solve the optimization problem offline over all likely input values so that the solutions can be saved and real-time optimization is not needed once the input data is realized. In these cases, the PCA tolerance approach can be used to determine the set of likely input values and obtain the corresponding solutions.

In the following applications, we illustrate these two contexts, first highlighting a generalized PCA tolerance approach for sensitivity analysis and second example from model predictive control.

2.5.1 Generalized PCA Tolerance Approach

One limitation of the tolerance approach in Section 2.3 is that it considers only variations within the initial critical region. The generalized tolerance approach ([53]) extends the original tolerance approach for perturbations in c beyond the initial critical region corresponding to the optimal basis for the baseline parameters of the c inputs in the objective function. Below, we briefly describe how the PCA tolerance approach and the algorithm for exploring critical regions (Algorithm 1) can be applied to derive a sensitivity measure for the case where the tolerance region for a given tolerance value spans multiple critical regions.

Consider a tolerance value τ and the corresponding tolerance region $V(\tau) = \{\hat{c} + D\Lambda^{1/2}p : \|p\|_\infty \leq \tau\}$. This is a special case of the PCA tolerance region where b is fixed. For $p = 0$, let B_0 be an optimal basis and x^* be an optimal solution. Suppose that the tolerance region intersects multiple critical regions in c, say $Rc^{B_1}, Rc^{B_2}, \ldots, Rc^{B_k}$ in addition to Rc^{B_0}. Let $x^1, x^2, \ldots, x^k$ be optimal solutions for the k critical regions in c, respectively. For any $\hat{c} + D\Lambda^{1/2}p \notin Rc^{B_0}$, x^* is suboptimal and the objective could be improved by using one of the k solutions. The maximum regret $\alpha^*(\tau)$ is the maximum amount that the objective function could be improved from that of x^* by using one of the k solutions where c varies over $V(\tau)$, i.e.,

$$\alpha^*(\tau) = \max_j \{\alpha_j(\tau) : j = 1, \ldots, k\} \tag{2.10}$$

where

$$\alpha_j(\tau) = \max_p \{(\hat{c} + D\Lambda^{1/2}p)^T(x^* - x^j) : \|p\|_\infty \leq \tau, c + D\Lambda^{1/2}p \in Rc^{B_j}\}.$$

Note that each $\alpha_j(\tau)$ can be obtained for a given value of τ by solving an LP.

Once Algorithm 1 finds the set of bases $\mathcal{B}$ such that $V(\tau) \subseteq \bigcup_{B \in \mathcal{B}} Rc^B$, the maximum

regret $\alpha^*(\tau)$ can be calculated because we have an optimal basis for every parameter vector in the tolerance region following the theoretical results in the previous section. Note that it is also possible to construct a maximum regret curve $\alpha^*(t)$ for $t \leq \tau$ by calculating $\alpha^*(t)$ in each iteration of Algorithm 1. By Propositions 2.3.1 and 2.3.2, $V(\tau)$ is a confidence region for μ, where the confidence level is a function of the tolerance τ and the dimensionality of the tolerance region h. This allows us to make confidence statements about the maximum regret. For instance, with at least $P(|Z| \leq \tau)^h \times 100\%$ confidence, the regret is less than or equal to $\alpha^*(\tau)$. If there is a desired confidence level, the corresponding tolerance value τ can be determined using Proposition 2.3.1.

Dantzig Example

For illustration of the generalized PCAT approach, let us return to the example given in Section 2.3.2. This time, we fix $b_1 = 6000$ and $b_2 = 4000$, and consider the case where only the objective function coefficients vary. Specifically, we consider the case where the first four coefficients, C_1, C_2, C_3, and C_4 vary according to a multivariate normal distribution with mean vector $\mu = [12\ 20\ 18\ 40]^T$ and covariance matrix Σ where

$$\Sigma_{(i,j)} = \begin{cases} \sqrt{\mu_i} & i = j, \\ \frac{1}{2} & i \neq j. \end{cases}$$

We take μ to be the baseline. Using (2.3), we see that the maximum tolerance which maintains the initial optimal basis is $\tau^* = 0.68$. For values of the objective function coefficients outside the initial critical region, the optimal solution changes and the maximum regret is therefore nonzero.

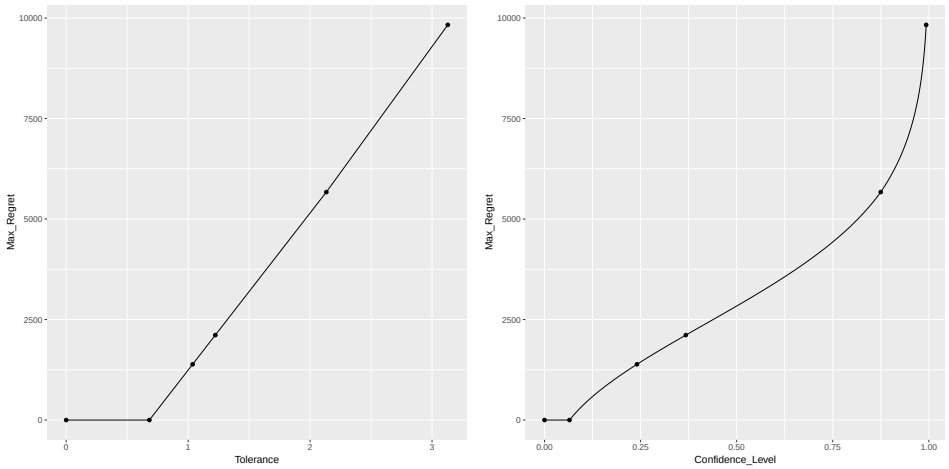


Figure 2.3: Generalized Tolerance

A plot of the maximum regret function against the tolerance value is provided in the left panel of Figure 2.3. The tolerance value can be translated to a confidence level using Proposition 2.3.1. A plot of the maximum regret function against the confidence level is provided in the right panel of Figure 2.3. As can be seen in the right panel, the loss in optimality is almost surely less than 10,000, as the tolerance region containing 99% of the density has a maximum regret of approximately 9384.13. The maximum regret as a function of the tolerance is piecewise-linear and nondecreasing. As a function of the confidence level, the regret is nondecreasing and continuous, but not piecewise-linear.

2.5.2 Application to Model Predictive Control

Multiparametric programming is used in model predictive control (MPC) to solve a control optimization problem for a set of input parameters that span beyond the initial critical region. Since the total number of critical regions can be exponential in problem size ([54]), it is desirable to focus on a set of input parameters that are more likely than others.

MPC LP problems ([55]) can often be written in the following way:

$$\begin{aligned} &\min_{x} c^T x \\ &\text{s.t. } Ax = b + Fy(t), \\ &\quad x \geq 0, \end{aligned}$$

where $y(t)$ is a vector of forecasts at time t. These forecasts may represent future demands, economic conditions, or other types of unknown future system states. Often, statistical models are used to obtain such forecasts from past data. ([42, 56, 57]) In these cases, the distribution of the forecast can be used to determine a region in which the parameters are likely to fall, which saves the computational load to explore all possible values of the input parameters.

Consider an input RIM parameter vector $\hat{r}$ observed from $\mathcal{N}(\mu, \Sigma)$. By Proposition 2.3.1, we can use the predictive distribution to construct a tolerance region that contains the future value of the parameters with a given probability. In particular, the PCA tolerance approach can target a subset of the RIM parameter space to examine based on the sampling distribution of the estimated RIM parameters. Algorithm 1 finds all critical regions covering the tolerance region and then we have an optimal control solution for all input parameters in the tolerance region.

Inventory Control Example

We illustrate the applicability of our approach to MPC using a multi-commodity inventory control problem with parameter forecasts. The problem is to plan production of J products over T time periods, where the cost and demand parameters are forecasts calculated using time series modeling techniques. Similar MPC and forecasting frameworks have been used by [57, 42] and [56] and specifically for inventory control by [58] and [59].

Let q_{tj}, x_{tj}, and b_{tj} denote the production quantity, the on-hand inventory at the end

of time period, and backorders, for product j at time period t. The initial inventory and backorder levels are fixed and denoted by x_{0j} and b_{0j} for product j, respectively. Production cost, holding cost, and the penalty for backlogged demand for product j in period t are denoted by C_{tj}, h_{tj}, and p_{tj}, respectively. Finally, D_{tj} denotes the demand at time t for product j and K is the inventory capacity.

$$
\begin{aligned}
(\mathrm{P}^{C,D}) \min_{q,x,b} \; & \sum_{j=1}^{J} \sum_{t=1}^{T} C_{tj} q_{tj} + h_{tj} x_{tj} + p_{tj} b_{tj} \\
\text{s.t. } & x_{t-1,j} + q_{tj} + b_{tj} = D_{tj} + x_{tj} + b_{t-1,j} \text{ for } t = 2, \ldots, T \text{ and } j = 1, \ldots, J \\
& \sum_{j=1}^{J} x_{tj} \leq K \text{ for } t = 1, \ldots, T \\
& q_{tj}, x_{tj}, b_{tj} \geq 0, \text{ for } t = 1, \ldots, T \text{ and } j = 1, \ldots, J
\end{aligned}
$$

The goal is to make the optimal control decision minimizing the predicted system costs, while meeting demand forecasts. We consider the inventory management problem with $J = 3$ products, with costs and demands predicted over $T = 7$ time periods. We assume fixed holding costs $h_{tj} = 1$, fixed penalties for backlogged demands $p_{tj} = 10$, and fixed inventory capacity $K = 15$.

Consider forecasting demand and production costs for each time period using a vector autoregression (VAR) approach. Let y_t denote the vector of demand and production costs at time t, i.e.,

$$y_t = [C_{t1} \ldots C_{tJ} \; D_{t1} \ldots D_{tJ}]^T.$$

Assuming that y_t follows a VAR(1) process, we have

$$y_t = g + M y_{t-1} + \epsilon_t. \tag{2.11}$$

The matrix M describes the correlation structure between the elements of y_t across time,

while the error terms in ϵ_t are independent and identically distributed normal random variables with mean zero, i.e., $\epsilon_t \sim \mathcal{N}(0, \Sigma)$, where Σ is a diagonal covariance matrix.

Suppose that we have an observation y_0. Given estimates $\hat{g}$ and $\hat{M}$, the estimate of y_t for $t = 1, \ldots, T$ is given recursively by

$$\hat{y}_t = \hat{g} + \hat{M}\hat{y}_{t-1}.$$

The forecast demands and costs are then used to solve the control problem ($\mathrm{P}^{\hat{C},\hat{D}}$). The model (2.11) implies a correlation structure on the demands and costs, allowing for correlations to exist among costs, among demands, and between them, over time. More details and parameter values are provided in the A.6.

In this example, the tolerance region defined in Section 3 is a prediction interval for the demand and cost forecasts. Let B_0 be an optimal basis and x^* be an optimal solution at the predicted cost and demand values, i.e., optimal for ($\mathrm{P}^{\hat{C},\hat{D}}$). The objective value of this LP is \$1775.79. We examine the tolerance of this solution to variations in the cost and demand parameters within the tolerance region given by the prediction model. The PCA maximum tolerance defined in Section 3 is $\tau^* = 1.2805$. This means that the production costs and demands can simultaneously vary up to 1.2805 standard deviations from their forecasts along the directions provided by the eigenvectors of the covariance matrix and maintain the optimal basis B_0. From Proposition 2.3.1, we see that B_0 is optimal with probability $P(|Z| \leq 1.2805)^{42} \approx 0.00834\%$.

When we consider variations of the demand and cost parameters beyond the maximum tolerance, the tolerance region will go out of the initial critical region and the optimal basis will change. For a tolerance value of $\tau = 2.602$, the tolerance region is covered by 501 critical regions which were found by Algorithm 1. This tolerance region is a $P(|Z| \leq 2.602)^{42} \approx 67.622\%$ prediction interval for the demand and cost parameters.

2.6 Conclusion

Sensitivity analysis remains an important tool for practitioners in making optimal decisions. In this paper, we focus on sensitivity analysis for linear programs when the parameters in both the objective function and the constraints, jointly called the RIM parameters, vary with a given statistical distribution.

In this paper, we revisit the tolerance region approach for sensitivity analysis but under a more general setting, assuming statistical distributions on the uncertainty of the random input parameters. The approach yields a tolerance region which is a symmetric confidence set for given baseline input parameters, providing practitioners a sense of both how much the parameters can vary while maintaining stability of a solution and how likely such variations are. We also extend the suggested tolerance approach beyond the initial critical region. In this approach, a confidence set is constructed over which the maximum regret sensitivity measure can be calculated as a measure of global sensitivity. We illustrated the approach using multiple applications, particularly, for model predictive control and an inventory management problem.

The implementation of the proposed tolerance approach relies on a series of theoretical results presented in this paper. We first contribute to the theory of critical regions, which is an important concept in the parametric programming literature used in global sensitivity analysis methods. We provide theoretical insights into the geometric properties of critical regions in the RIM case for which a thorough theoretical treatment was previously lacking in the literature.

The results presented here can be extended in several ways. First, the approach of using PCA to translate a given distribution of the uncertain parameters into the construction of the tolerance region can be extended to the cases where the tolerance region is asymmetric. While the resulting confidence sets would be asymmetric, calculating the corresponding confidence levels would still be rather straightforward due to the fact that variations along

each of the directions provided by PCA are independent. Extensions to asymmetric distributions can be achieved by defining the tolerance set based on the distribution and then solving the corresponding multiparametric program over the region. ([38]) This approach may be especially important for the case when uncertainty in the RIM parameters is specified by an estimated distribution rather than the sampling distribution of statistical estimators, since the symmetry distribution in these cases may be especially restrictive. Last, an algorithm searching for critical regions such as Algorithm 1 can also identify which parameter value makes a given problem infeasible ([37]). Analyzing how the theory and method of this paper relate to the concept of condition number of linear programs is a future research direction.

Acknowledgments

Research reported in this publication was supported by the National Institute of Dental and Carniofacial Research of the National Institutes of Health under Award Number R01DE028283. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CHAPTER 3

SOLVING LARGE BATCHES OF LINEAR PROGRAMS

Solving a large batch of linear programs (LPs) with varying parameters is needed in stochastic programming and sensitivity analysis among other modeling frameworks. The common approach is solving the LPs for all combinations of given parameter values, called the brute-force approach, which can be computationally infeasible when the parameter space is high-dimensional and/or the underlying LP is computationally challenging. This chapter introduces a computationally efficient approach for solving a large number of LPs that differ only in the right hand side of the constraints ($\mathbf{b}$ of $\mathbb{A}x = \mathbf{b}$). The computational approach builds on theoretical properties of the geometry of the space of critical regions, where a critical region is defined as the set of $\mathbf{b}$'s for which a basis is optimal. To formally support our computational approach we provide proofs of geometric properties of neighboring critical regions. While these theoretical properties have been stated and applied in the existing literature of parametric programming, their formal proofs have not been provided to the best of our knowledge. Based on the geometric properties of critical regions, we develop an algorithm that solves the LPs in batches by finding critical regions that contain multiple $\mathbf{b}$'s. Moreover, we suggest a data-driven version of our algorithm that uses the distribution (e.g., shape) of a sample of $\mathbf{b}$'s for which the LPs need to be solved. The experimental results on a benchmark problem show that our approach can be more efficient and scale better in the number of $\mathbf{b}$'s than the brute-force, but also indicate some limitations of the algorithm.

3.1 Introduction

The need to solve a large number of linear programs (LPs) with varying model parameters arises in various modeling frameworks. There are several popular approaches to multi-

stage stochastic programs (SPs) that require solving a large number of LPs with varying parameter values. One such example is the solution algorithm for two-stage SPs by [60], a systematic approach for generating cutting planes that exclude first-stage decisions that either are non-optimal or make the second-stage problem infeasible, which requires solving the second-stage LP for each realization of random parameters at each iteration. A second example is the sample average approximation approach for two-stage SPs with recourse [61], where the second-stage LP is solved for sampled parameters at each iteration. A third example is the simulation-based scheme called bagging to obtain a solution whose optimality gap exhibits a specific distributional behavior [62]; to obtain the optimality gap, multiple LPs with sampled parameter values need to be solved.

Global sensitivity analysis is another modeling framework in which a large number of LPs with different parameters needs to be solved [13]. Classical sensitivity analysis evaluates variations in the optimal objective function value (called herein the optimal value) with respect to perturbations of a parameter. [13] proposed estimating the variability of the optimal value attributed to varying parameter values and using the variance as a measure of sensitivity. Given a sample of parameter values, the LP for each sampled parameter is solved separately to obtain the optimal value, from which the variance of the optimal value is approximated. Global sensitivity or parametric programming hints broadly to quantification of the uncertainty in the optimal value derived from optimization models, particularly when the parameters follow a multivariate distribution. This approach can be applied to quantify the uncertainty in the optimal solutions and outcome functions of the optimal solutions.

Generally, we consider a standard form LP:

$$\begin{aligned} \text{(P)} \ \min_{\mathbf{x}} \ & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \ & \mathbb{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq 0, \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n, \mathbb{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n$. In this paper, we focus on the case where a set of right hand side parameter values, $\Theta = \{\mathbf{b}^1, \ldots, \mathbf{b}^K\}$, is given while $\mathbb{A}, \mathbf{c}$ are fixed. By duality, all results in this paper extend to solving LPs for $\mathbb{A}, \mathbf{b}$ fixed but varying parameter values for $\mathbf{c}$. These are common settings in the aforementioned examples.

Let ($\text{P}_\mathbf{b}$) denote the LP (P) with right hand side $\mathbf{b}$. Given the sample of parameter values in Θ, the objective is to develop a computationally efficient method for solving the collection of LPs ($\text{P}_{\mathbf{b}^1}$), ($\text{P}_{\mathbf{b}^2}$), . . ., ($\text{P}_{\mathbf{b}^K}$). Solving each individual LP separately, which we call the *brute-force* approach, can easily become computationally prohibitive for a large K and as the dimensionality of the LPs grows. The proposed method obtains optimal solutions for the collection of LPs by solving possibly a much smaller number of LPs than K (as small as one). The underlying idea is to take advantage of the "similarity" between the parameter values in the sample, thus solving the LP for only one or a few parameter values and deriving optimal solutions for similar parameter values from the solutions of the LPs solved.

A straightforward approach is to solve the LP for a parameter vector, choose another parameter vector to solve next, solve the LP starting from the solution for the previous parameter vector (called 'the warm start'), and repeat the process (e.g., [63, 64, 60]). The challenge of this approach is deciding on the order of parameter vectors to solve the LPs, particularly when a large number of parameter vectors is considered and when the parameter vectors lie in a high-dimensional space. In this paper, we develop a novel approach

utilizing the idea of solving LPs from solutions of other LPs in order to solve an LP for an extremely large number of parameter values (e.g., millions) while dealing with high-dimensionality of the space of the parameter values.

The proposed computational approach for solving the collection of LPs relies on theoretical results from parametric programming literature, particularly, properties of optimal bases and critical regions (e.g., see [52, 11] and references therein). Parametric programming is an extension of sensitivity analysis that quantifies variations of the optimal value and optimal solution under perturbations in the model parameters. A basis is optimal for a given $\mathbf{b}$ if and only if the corresponding basic solution is feasible (the *feasibility condition*) and the reduced costs are nonnegative (the *optimality condition*). A basis that is optimal for some $\mathbf{b}$ is called an *optimal basis*. The *critical region* of an optimal basis is defined as the set of $\mathbf{b}$'s for which the basic solution is feasible, thus, optimal. One important property is that the LPs whose $\mathbf{b}$'s belong to the same critical region can be solved together by finding the optimal basis of the critical region; this property was used in [60] for two-stage SPs to solve the second-stage LP for all realizations of random parameters. A second property is that dual simplex pivots can be used to generate optimal bases; this property was used by [65] to solve the second-stage LPs in batches. However, they did not study geometric properties of critical regions found by dual simplex pivots. By understanding properties of critical regions and the dual simplex method, we will thus develop a computationally efficient approach to find critical regions covering the sample of parameter values.

Geometric properties of critical regions have been previously introduced in the parametric programming literature. In [52, 11] and [66], a *neighboring critical region* of a critical region R is defined as a critical region whose basis can be reached by a dual simplex pivot from B, where B is the optimal basis of R. While this is an established definition, geometric properties of neighboring critical regions have not been studied thoroughly. The most detailed discussion on the geometry of neighboring critical regions was given by [67], which stated properties of neighbors including that neighbors share a facet, i.e., they are

also "geometrically neighbors". However, they did not provide proofs for their statements or prior references supporting them. In [66], it was shown that neighboring critical regions had nonempty intersection and lay on opposite sides of a hyperplane, but further geometric properties were not studied. In this paper, we provide proofs of the geometric properties of neighboring critical regions and simplex tableaus stated in [67] and establish geometric properties in addition to those in the previous works. Our proofs also provide a new geometric insight for the simplex tableau that is used to find neighboring critical regions.

Furthermore, we develop a data-driven computational framework to solve batches of LPs using the theoretical properties of neighboring critical regions and the distributional properties of the sample of parameter values for which the LPs are solved. We provide insights into the computational complexity of the proposed algorithms under different multivariate distributions of the sample parameter values along with current limitations of the algorithms. We provide empirical comparisons of our approach and three other methods for various instances generated based on benchmark problems from the existing literature. Lastly, we evaluate the proposed computational approach for an LP from a specific application, the measurement of spatial access to healthcare services [17, 20] and the results suggest a limitation of our approach and future research directions.

3.2 Theoretical Properties

Notation. Let $\mathcal{F} = \{\mathbf{b} \in \mathbb{R}^m \mid (\mathrm{P}_\mathbf{b}) \text{ is feasible}\}$, the set of all parameter values $\mathbf{b}$ for which the LP is feasible, which is a convex polyhedron (e.g., see [52]). Let $V(\mathbf{b})$ and $X^\star(\mathbf{b})$ denote the optimal value and the set of optimal solutions of $(\mathrm{P}_\mathbf{b})$, respectively. For a matrix M, M_i denotes its ith column and $M_{(i)}$ denotes its ith row. We use $B = \{B(1), B(2), \ldots, B(m)\}$ to denote a basis, which is a set of indices of basic variables or equivalently, columns of $\mathbb{A}$. Let N denote the set of nonbasic indices. Subscript B on a vector (a matrix) denotes its subvector (its submatrix) of those elements (columns) corresponding to the indices in B; subscript N is defined in the same way.

A basis B is optimal for ($P_{\mathbf{b}}$) if and only if

$$\mathbb{A}_B^{-1}\mathbf{b} \geq 0, \tag{3.1}$$

$$\mathbf{c} - \mathbf{c}_B'\mathbb{A}_B^{-1}\mathbb{A} \geq 0. \tag{3.2}$$

The inequality (3.1) is called the *feasibility condition* and (3.2) is called the *optimality condition*. We make the following basic assumption.

Basic Assumption. There exists $\mathbf{b} \in \mathcal{F}$ for which ($P_{\mathbf{b}}$) has a finite optimum.

If this assumption does not hold, then ($P_{\mathbf{b}}$) is unbounded for any feasible b, thus the goal of this paper becomes uninteresting. By duality, the assumption implies that for every $\mathbf{b} \in \Theta$, ($P_{\mathbf{b}}$) is either infeasible or has a finite optimal solution.

3.2.1 Critical Region and Neighbor

In this section, we first introduce the definition of a critical region and then study its properties as provided in the parametric programming literature. A basis B is defined to be an *optimal basis* if there exists $\mathbf{b} \in \mathcal{F}$ for which B is optimal for ($P_{\mathbf{b}}$). For an optimal basis B, its *critical region* R^B is defined as

$$R^B = \{\mathbf{b} \in \mathbb{R}^m \mid \mathbb{A}_B^{-1}\mathbf{b} \geq 0\}. \tag{3.3}$$

That is, R^B is the set of $\mathbf{b}$'s for which B is optimal for ($P_{\mathbf{b}}$). From (3.3), we know that a critical region is a polyhedral cone pointed at the origin. In addition, a critical region has the following properties.

Proposition 3.2.1 *(a) A critical region R^B is the set of conic combinations of the columns of $\mathbb{A}_B$. Moreover, each column of $\mathbb{A}_B$ is an extreme ray of R^B.*

(b) There is no redundant inequality in the description (3.3) of a critical region.

(c) Any critical region is full-dimensional, and thus, has an interior point.

Proof: A critical region R^B can be also written as

$$R^B = \{\mathbf{b} \in \mathbb{R}^m \mid \mathbb{A}_B^{-1}\mathbf{b} \geq 0\} = \{\mathbb{A}_B\mathbf{z} \mid \mathbf{z} \in \mathbb{R}^m, \mathbf{z} \geq 0\},$$

and the columns of $\mathbb{A}_B$ are linearly independent, thus, (a) holds. Suppose that an inequality in (3.3) is redundant, i.e., the inequality is a linear combination of the other inequalities. However, the inequalities defining R^B are linearly independent since $\mathbb{A}_B^{-1}$ is nonsingular, which proves (b). Suppose that a critical region is not full-dimensional, i.e., it is contained in a hyperplane. It implies that the inequalities defining a critical region should imply an equality, which again contradicts the fact that the inequalities are linearly independent, and this proves (c). □

Remark: In this paper, we consider the right hand side b be a parameter vector. On the other hand, most of the prior research in the parametric programming literature such as [52] defined b as a linear function of parameters (often denoted as λ). This difference makes our overall analysis simpler and enables us to establish additional results. For example, in the previous works, a critical region defined therein may not be full-dimensional.

We review some definitions from polyhedral geometry. A *face of a polyhedron* is defined as the intersection of the polyhedron and a supporting hyperplane. A *facet of a polyhedron* is defined as a face whose dimension is one less than that of the polyhedron. From the above proposition, we know that each inequality in the definition (3.3) of a critical region defines a facet. We refer to the facet defined by the ith inequality $(\mathbb{A}_B^{-1})_{(i)}\mathbf{b} \geq 0$ as the ith facet.

The following result from parametric LP literature [68, 52, 66] shows that $V(\mathbf{b})$ is linear in each critical region and there exists $x^\star(\mathbf{b}) \in X^\star(\mathbf{b})$ for $\mathbf{b} \in \mathcal{F}$ such that $x^\star(\mathbf{b})$ is a linear function in each critical region.

Theorem 3.2.2 *[cf. Section 2.2 of [52]] (a) The optimal value function $V(\mathbf{b})$ is piecewise*

linear, convex, and continuous; and

(b) in each critical region R^B, $V(\mathbf{b})$ is linear and there exists an optimal solution function $x^\star(\mathbf{b}) \in X^\star(\mathbf{b})$ that is linear in R^B, namely, for $\mathbf{b} \in R^B$,

$$V(\mathbf{b}) = \mathbf{c}'_B \mathbb{A}_B^{-1} \mathbf{b} \tag{3.4}$$

and

$$x^\star(\mathbf{b})_B = \mathbb{A}_B^{-1}\mathbf{b} \text{ and } x^\star(\mathbf{b})_N = 0. \tag{3.5}$$

Theorem 3.2.2 implies that for parameter values of $\mathbf{b}$ in one critical region, the optimal values and optimal solutions can be computed all together by finding the optimal basis of the critical region and computing the inverse of the basis matrix. For example, after solving an LP ($P_{\mathbf{b}^1}$) for a parameter value $\mathbf{b}^1$ (e.g., by simplex method), the optimal basis B^1 is obtained and the critical region R^{B^1} is computed as in (3.3). Then, for all $\mathbf{b} \in R^{B^1}$, $V(\mathbf{b})$ and $x^\star(\mathbf{b})$ are given as in (3.4) and (3.5). Therefore, Theorem 3.2.2 provides a way to solve a batch of LPs whose right hand sides belong to the same critical region while solving only one LP. This is the first theoretical result supporting our computational approach for solving a large collection of LPs in batches.

A second result supporting the proposed computational approach that we will prove is that given a critical region R^{B^1} and the inverse of the basis matrix $A_{B^1}^{-1}$, we can obtain its surrounding critical regions and compute optimal solutions for the parameter values in the surrounding critical regions without solving another LP or inverting another matrix. This property relies on the concept of *neighboring critical regions*. Any two bases B^1 and B^2 are said to be *neighboring bases* if and only if it is possible to pass from B^1 to B^2 by one pivot operation of dual simplex method and vice versa. Any two critical regions R^1, R^2 are said to be *neighbors* if and only if the corresponding optimal bases B^1, B^2 are neighbors.

Remark: There is a key difference between a pivot operation of typical dual simplex methods and that of the above definition. In an iteration of dual simplex method (e.g., see

Chapter 5 of [69]), a *negative* basic variable is found and if the corresponding row of the tableau has a negative entry, then the basic variable leaves the basis. On the other hand, in the above definition, *any* basic variable whose corresponding row has a negative entry can leave the basis to find a neighboring critical region. This difference is due to the fact that the goal of typical dual simplex methods is to find a basis optimal for a given $\mathbf{b}$, whereas, a dual simplex pivot of the above definition finds a basis that is optimal for $\mathbf{b}$s not in the current critical region.

3.2.2 Geometry of Neighbors

In this section, we prove that neighboring critical regions are also neighbors in a geometric sense and vice versa. Also, we introduce a geometric relationship between critical region and simplex tableau. We begin by defining geometric neighbors. Two critical regions are said to be *geometric neighbors* if their intersection is a facet of each of them, i.e., they share a facet. We refer to the concept of neighboring critical regions defined in the previous section as *algebraic neighbors* when it is necessary to distinguish the two definitions.

It is a well-established result in the parametric LP literature that two algebraically neighboring critical regions have nonempty intersection and lie in opposite halfspaces as it follows from [66]. The next theorem shows that the intersection of neighbors is a facet of each of them, thus they are geometric neighbors. The most detailed discussion about geometric relationships of algebraic neighbors was by [67]. They stated the two theorems provided below however without formal proofs. We provide proofs for both theorems, thus formally establishing these results.

Key fundamental results used in the proofs are the following. $\mathcal{F}$ is the set of conic combinations of the columns of $\mathbb{A}$ and a critical region R^B of an optimal basis B is the set of conic combinations of the columns of $\mathbb{A}_B$. In the simplex tableau at a basis B, the entries in the jth column are the coefficients of the unique linear combination representation of $\mathbb{A}_j$ by the columns of $\mathbb{A}_B$. Thus, a negative entry in (i, j) position in the tableau indicates that

$\mathbb{A}_j$ is not a conic combination of the columns of $\mathbb{A}_B$, and that replacing the ith column of $\mathbb{A}_B$ by $\mathbb{A}_j$ leads to a new set of conic combinations.

Theorem 3.2.3 *If two critical regions are algebraic neighbors, then they are geometric neighbors.*

Proof: The proof is similar to the one of Theorem 2.4 in [66].

Consider two critical regions R^1 and R^2 that are algebraic neighbors. Let B^1, B^2 be the optimal bases corresponding to R^1, R^2, respectively. Suppose that the pivot operation obtaining B^2 from B^1 replaces the ith basic variable $x_{B^1(i)}$ by x_j where $j \in B^2 \setminus B^1$. Let H be the hyperplane defined by $(\mathbb{A}_{B^1}^{-1})_{(i)}\mathbf{b} = 0$. We denote the ith facet of R^1 in H as F^1. We will first show that R^2 also has a facet F^2 contained in H and then, show that $F^1 = F^2$. For simplicity of notation, let $\bar{\mathbb{A}} = \mathbb{A}_{B^1}^{-1}\mathbb{A}$. Then, we know that $\bar{\mathbb{A}}_{ij} < 0$ because x_j enters the basis.

Assume that we do not reorder rows while performing the pivot operation. Then, while computing the simplex tableau at basis B^2 from the tableau at B^1, the elementary row operation performed on the ith row is dividing it by $\bar{\mathbb{A}}_{ij}$. Also, recall that $\mathbb{A}_{B^2}^{-1}$ is obtained from $\mathbb{A}_{B^1}^{-1}$ by performing the same elementary row operations. Thus, $(\mathbb{A}_{B^1}^{-1})_{(i)} = \bar{\mathbb{A}}_{ij}(\mathbb{A}_{B^2}^{-1})_{(i)}$. By Proposition 3.2.1(b), R^2 has a facet F^2 defined by $0 \leq (\mathbb{A}_{B^2}^{-1})_{(i)}\mathbf{b} = \frac{1}{\bar{\mathbb{A}}_{ij}}(\mathbb{A}_{B^1}^{-1})_{(i)}\mathbf{b}$. Therefore, the facets F^1 and F^2 lie in the hyperplane H. Since $\bar{\mathbb{A}}_{ij} < 0$, R^1 and R^2 lie on the opposite sides of H. Moreover, for $k = 1, 2$,

$$F^k = R^k \cap H = R^k \cap \{\mathbf{b} \mid (\mathbb{A}_{B^k}^{-1})_{(i)}\mathbf{b} = 0\} = \{\mathbb{A}_{B^k}\mathbf{z} \mid \mathbf{z} \in \mathbb{R}^m, \mathbf{z} \geq 0, \mathbf{z}_i = 0\}.$$

Since B^1 and B^2 differ only at their ith index, we have $F^1 = F^2$. Therefore, $R^1 \cap R^2 = F^1 = F^2$. □

The above proof states that if a neighbor basis B^2 is obtained from B^1 by a dual simplex pivot with the ith row as a pivot row, then the ith constraint of R^1, $(\mathbb{A}_{B^1}^{-1})_{(i)}\mathbf{b} \geq 0$ defines

the hyperplane between R^1 and R^2. In other words, if one can perform a dual simplex pivot with the ith row as a pivot row, then there is a (geometrically) neighboring critical region on the other side of the hyperplane. The next theorem states that if one cannot perform a pivot operation with the ith row as a pivot row, then there is no feasible point $\mathbf{b} \in \mathcal{F}$ on the other side of the hyperplane.

Theorem 3.2.4 *Consider the simplex tableau at a basis B and let R denote the critical region. If a dual simplex pivot cannot be performed with the ith row as a pivot row, then there is no feasible point $\mathbf{b} \in \mathcal{F}$ on the other side of the hyperplane defining the ith facet of R.*

Proof: Assume that there exists $b \in \mathcal{F}$ such that $(\mathbb{A}_B^{-1})_{(i)}\mathbf{b} < 0$. Since $\mathcal{F}$ is the set of conic combinations of columns of $\mathbb{A}$, there exists $\mathbb{A}_j$ such that $(\mathbb{A}_B^{-1})_{(i)}\mathbb{A}_j < 0$. Thus, the ith row of the tableau has a negative entry at the jth column. Therefore, one can perform a dual simplex pivot with the ith row as a pivot row and the resulting (algebraic) neighbor is on the other side of the ith facet. □

This theorem implies that, if there is a feasible parameter value $\mathbf{b} \in \mathcal{F}$ on the other side of a hyperplane defining a facet, then it is possible to perform a dual simplex pivot to find a (algebraic) neighbor across the facet. This theorem is central to the proposed computational approaches in Section 3.3 and 3.4. A given LP may not be feasible for all given parameter values in many realistic modeling frameworks and the theorem suggests an approach to find boundaries of $\mathcal{F}$ and to exclude those parameter values that make ($\mathrm{P}_{\mathbf{b}}$) infeasible while exploring critical regions. If the simplex tableau at an optimal basis has a row with no negative entry, then the corresponding hyperplane also defines a facet of $\mathcal{F}$ and we can exclude those parameter values on the other side of the hyperplane as they make ($\mathrm{P}_{\mathbf{b}}$) infeasible.

In addition to the above two theorems stated in [67], we add the following theorem,

which completes proving the equivalence between algebraic neighbors and geometric neighbors.

Theorem 3.2.5 *If two critical regions are geometric neighbors, then they are also algebraic neighbors.*

Proof: Consider the critical regions R^1 and R^2 assumed to be geometric neighbors and let B^1 and B^2 be their corresponding optimal bases, respectively. Let $F = R^1 \cap R^2$ be the shared facet. Let F be the ith facet of R^1 and the i'th facet of R^2, i.e.,

$$F = \{\mathbb{A}_{B^1}\mathbf{z} \mid \mathbf{z} \in \mathbb{R}^m, \mathbf{z} \geq 0, \mathbf{z}_i = 0\} = \{\mathbb{A}_{B^2}\mathbf{z} \mid \mathbf{z} \in \mathbb{R}^m, \mathbf{z} \geq 0, \mathbf{z}_{i'} = 0\}.$$

Using this equality and the fact that all columns of $\mathbb{A}_{B^1}$ and $\mathbb{A}_{B^2}$ are extreme rays of F, it is easy to show that $B^1 \setminus \{B^1(i)\} = B^2 \setminus \{B^2(i')\}$, that is, B^1 and B^2 differ only by one basic variable. Let $B^2(i') = j$.

Let H denote the hyperplane containing the facet F. Since R^1 and R^2 lie on the opposite sides of H, for all $\mathbf{b} \in R^2$, $(\mathbb{A}_{B^1}^{-1})_{(i)}\mathbf{b} \leq 0$. Since R^2 has an interior point by Proposition 3.2.1, there exists $\hat{\mathbf{b}} \in R^2$ such that $(\mathbb{A}_{B^1}^{-1})_{(i)}\hat{\mathbf{b}} < 0$ and since $B^2 \setminus B^1 = \{j\}$, we have $(\mathbb{A}_{B^1}^{-1})_{(i)}\mathbb{A}_j < 0$. This implies that the (i, j) element of the simplex tableau at B^1 (which is the matrix $\mathbb{A}_{B^1}^{-1}\mathbb{A}$) is negative. Thus, we can perform a dual simplex pivot operation at B^1 removing its ith basic variable. Moreover, we can show that in the pivot operation, x_j wins the min ratio test (possibly allowing a tie). In short, if we assume that x_j does not win the min ratio test, then we can easily show that there is a nonbasic variable with negative reduced cost at B^2, that is, B^2 is not an optimal basis, thus a contradiction. Therefore, x_j can enter the basis and B^2 is obtained from B^1 by a dual simplex pivot. □

By Theorems 3.2.3 and 3.2.5, we have the equivalence of neighbors and geometric neighbors and thus the following corollary.

Corollary 3.2.6 *Two critical regions are algebraic neighbors if and only if they are geometric neighbors.*

The partition of $\mathcal{F}$ into critical regions depends on degeneracy of the dual of (P). If the dual of (P) is nondegenerate, i.e., ($P_{\mathbf{b}}$) has a unique optimal solution for any $\mathbf{b} \in \mathcal{F}$, then $\mathcal{F}$ is uniquely partitioned by the collection of all critical regions. They can be found by starting at a critical region, and finding its neighbors, neighbors of neighbors, and so on [66].

Under degeneracy of the dual of (P), there can be multiple partitions of $\mathcal{F}$ by critical regions as demonstrated by [11]. Under dual degeneracy, there is a tie in the min ratio test of a dual simplex pivot in the search for neighbors. Specifically, say at a basis B^1, there is a tie between two nonbasic variables in the min ratio test and let B^2 and B^3 be the two resulting bases. Then, for each of B^2 and B^3, there is a nonbasic variable with zero reduced cost. Note that B^2 can be obtained from B^3 by a primal simplex pivot making the nonbasic variable with zero reduced cost basic and vice versa, but they are <u>not</u> neighbors according to the definition of neighboring bases. In fact, by using arguments similar to the proof of Theorem 3.2.3, we can show that their corresponding critical regions share a facet, but they lie on the same side of the facet and thus overlap. This implies that, in case of dual degeneracy, there can be multiple partitions of $\mathcal{F}$ by critical regions. However, any tie-breaking rule for the min ratio test gives a unique collection of critical regions partitioning $\mathcal{F}$ [11].

3.3 Solving Large Batches of LPs: A General Algorithm

Based on Theorems 3.2.2 and 3.2.4, and Corollary 3.2.6, we develop an algorithm that solves a large number of LPs with different right hand sides. Our algorithm does not solve each individual LP separately. Instead, it solves LPs in batches, each of which contains LPs with right hand sides belonging to the same critical region.

Suppose that for a data point $\mathbf{b}^1$, ($P_{\mathbf{b}^1}$) is solved by the simplex method (from now on,

we also call the parameter values data points as they are points in the parameter space). Let B^1 be the optimal basis. The inverse of the basis matrix A_{B^1} is obtained as a byproduct of the simplex method. Then, the critical region R^{B^1} is given as in (3.3) and for data points in R^{B^1}, $V(\mathbf{b})$ and an optimal solution $x^\star(\mathbf{b})$ are obtained as in (3.4) and (3.5). For any (geometric) neighbor of R^{B^1}, the optimal basis is obtained from B^1 by a dual simplex pivot. Let B^2 be a neighbor basis of B^1. The inverse of the basis matrix $\mathbb{A}_{B^2}$ is obtained from the inverse of $\mathbb{A}_{B^1}$ by performing elementary row operations to obtain the tableau of B^2 from the one of B^1. Specifically, a matrix $\mathbb{E}$ such that $\mathbb{E}\mathbb{A}_{B^1}^{-1} = \mathbb{A}_{B^2}^{-1}$, representing the elementary row operations, can be constructed. Then, the critical region R^{B^2} is obtained and for parameter values in R^{B^2}, $V(\mathbf{b})$ and $x^\star(\mathbf{b})$ are also obtained. Therefore, after solving one LP at the beginning, we can find critical regions covering sets of parameter values and compute $V(\mathbf{b})$ and $x^\star(\mathbf{b})$ in the critical regions without solving an additional LP or inverting another matrix. Thus, obtaining an additional (neighboring) critical region is computationally inexpensive in comparison to solving an LP, which is a key to the efficiency of the proposed algorithm.

This property of critical regions is well established in parametric programming literature, but it has only been used to find all critical regions that partition $\mathcal{F}$ so that the functions $V(\mathbf{b})$ and $x^\star(\mathbf{b})$ are completely constructed in $\mathcal{F}$ [52, 66]. For example, [66] suggested an algorithm that finds a collection of critical regions that partitions $\mathcal{F}$ and their corresponding bases. However, the number of critical regions in $\mathcal{F}$ can be exponential in the problem size [54]. Our goal in this paper is not to find all critical regions partitioning $\mathcal{F}$, but to instead find only those critical regions that cover the given set of parameter values in $\Theta = \{\mathbf{b}^1, \ldots, \mathbf{b}^K\}$.

Motivated by our understanding of critical regions and neighbors, we first present our algorithm for solving batches of LPs with differing RHSs in its most general form (Algorithm 1). The algorithm starts with an initial critical region (line 4), solves LPs whose RHSs belong to the critical region (line 7-9), checks if any facet of it is a facet of $\mathcal{F}$, and

if so, then excludes those parameter values on the other side of the hyperplane based on Theorem 3.2.4 (line 12-14), finds a neighbor of a previously explored critical region (line 16), and then repeats the process, until it finds all of the critical regions covering the parameter values in Θ. We emphasize again that $\mathbb{A}_B^{-1}\mathbb{A}$ in line 13 is available from the simplex tableau of the current basis B and that $\mathbb{A}_B^{-1}$ in line 8 is also obtained from the inverse of the previous basis matrix by simply performing the elementary row operations as explained above.

The starting point of the algorithm needs to be selected based on the set of parameter values that need to be covered. For example, if the set of parameter values is a sample from a unimodal probability distribution, we can start the covering search at the (sample) mode of the distribution and find critical regions near the mode to cover the sample parameter values. Since we expect the majority of the sample values to be located around the mode, we also expect the majority of the values to belong to critical regions near the mode. If the distribution is multimodal, one can consider partitioning the given points using a clustering algorithm in a way that each cluster forms a unimodal distribution and applying the proposed method to each cluster, possibly in a parallel fashion. For ease of presentation, we will focus on unimodal distributions and hence one starting point.

Before analyzing the computational complexity of this algorithm, we provide an alternative approach to the computation in line 8. Each time a new critical region is found, the algorithm needs to check if data points that are yet to be covered belong to the newly found critical region, i.e., the inequality $\mathbb{A}_B^{-1}\mathbf{b} \geq 0$ in line 8. The computational burden of checking the inequality is $\mathcal{O}(n^2)$ for each $\mathbf{b}$ given the inverse $\mathbb{A}_B^{-1}$. However, for two neighboring bases B and B', we have $\mathbb{A}_B^{-1}\mathbf{b} = \mathbb{E}\mathbb{A}_{B'}^{-1}\mathbf{b}$, where $\mathbb{E}$ is the matrix of the elementary row operation obtaining the tableau of B from that of B', and thus, if we have $\mathbb{A}_{B'}^{-1}\mathbf{b}$ already computed and stored for $\mathbf{b} \in \Theta$ that needs to be checked at B, then the $\mathbb{A}_B^{-1}\mathbf{b}$ values can be obtained by performing the elementary row operations on $\mathbb{A}_{B'}^{-1}\mathbf{b}$, which requires only $\mathcal{O}(n)$ operations for each $\mathbf{b}$. Thus, by storing the values of $\mathbb{A}_B^{-1}\mathbf{b}$ for $\mathbf{b}$'s that are not yet covered,

Algorithm 2 A Batch Solution Algorithm for LPs with Different RHSs

Input: A set of data points $\Theta = \{\mathbf{b}^1, \ldots, \mathbf{b}^K\}$ and an initial point $\mathbf{b}^0$
Initialize:
Set of explored bases $\mathcal{P} \leftarrow \{\}$
Solve $(\mathrm{P}_{\mathbf{b}^0})$ to find the optimal basis B
while $\Theta \neq \emptyset$ **do**
  Solve a Batch:
  for $\mathbf{b} \in \Theta$ **do**
    if $\mathbb{A}_B^{-1}\mathbf{b} \geq 0$ **then**
      Store $\mathbb{A}_B^{-1}\mathbf{b} = x^\star(\mathbf{b})_B$ as an optimal solution of $(\mathrm{P}_{\mathbf{b}})$ and exclude $\mathbf{b}$ from Θ
  $\mathcal{P} \leftarrow \mathcal{P} \cup \{B\}$
  Feasibility Check:
  for $i = 1, \ldots, m$ **do**
    if $(\mathbb{A}_B^{-1}\mathbb{A})_{(i)} \geq 0$ **then**
      Exclude from Θ those $\mathbf{b}$s satisfying $(\mathbb{A}_B^{-1})_{(i)}\mathbf{b} < 0$ and mark them 'infeasible'
  Choose a Neighbor:
  Choose a previously explored basis $B' \in \mathcal{P}$ and find a previously unexplored neighboring basis B'' of B' that is not in $\mathcal{P}$
  $B \leftarrow B''$

we can save computations in testing to which critical region the parameter values belong for further search. The computational improvement comes at the expense of memory usage since this technique requires the values $\mathbb{A}_B^{-1}\mathbf{b}$ to be stored for each critical region explored and for each $\mathbf{b} \in \Theta$ uncovered yet.

The worst case computational complexity of the above algorithm is $\mathcal{O}(Ln^{2.5} + Qn^2 + Kn^2 + KQn)$, where Q is the total number of critical regions the algorithm explores and L is the maximum size of the LP $(\mathrm{P}_{\mathbf{b}})$ in binary coding (for simplicity of discussion, we assume in the rest of this section that the LP is feasible for all parameter values given, so the feasibility check part is omitted). This bound is conservative in the sense that it assumes that all points are covered by the last critical region found. The first term is the complexity of solving one LP by the interior point method to find an initial critical region, which cannot be further reduced [70]. The second is the complexity of identifying neighbors and obtaining the corresponding optimal bases. The third and the final term are the complexity of checking to which critical region each $\mathbf{b} \in \Theta$ belongs, discussed in the previous paragraph, for the initial critical region and the later ones, respectively. For comparison, the

computational complexity of the brute-force approach, solving each individual LP using the interior point method, is $\mathcal{O}(KLn^{2.5})$. Note that if we sample a certain number of values for each entry of $\mathbf{b}$, then K equals the total number of combinations of the sampled entries, thus is exponential in n. If we consider the maximum number of bits to store each entry of $\mathbb{A}, \mathbf{b}, \mathbf{c}$ to be fixed, then $L = \mathcal{O}(n^2)$. Then, the computational complexity of Algorithm 1 is $\mathcal{O}(n^{4.5} + Qn^2 + Kn^2 + KQn)$ as opposed to $\mathcal{O}(Kn^{4.5})$ of the brute-force approach. Apparently, the computational efficiency of our approach will depend on Q, the number of critical regions explored to cover the parameter vectors. If the order of Q is bigger than $\mathcal{O}(Kn^{2.5})$ or $\mathcal{O}(n^{3.5})$, then the theoretical computational complexity of the proposed approach will be worse than that of the brute-force. The total number of critical regions partitioning $\mathcal{F}$ can be exponential in problem size [54]. Thus, as the number of critical regions that have to be found to cover given parameter vectors increases, the proposed method may be asymptotically slower than the brute-force approach.

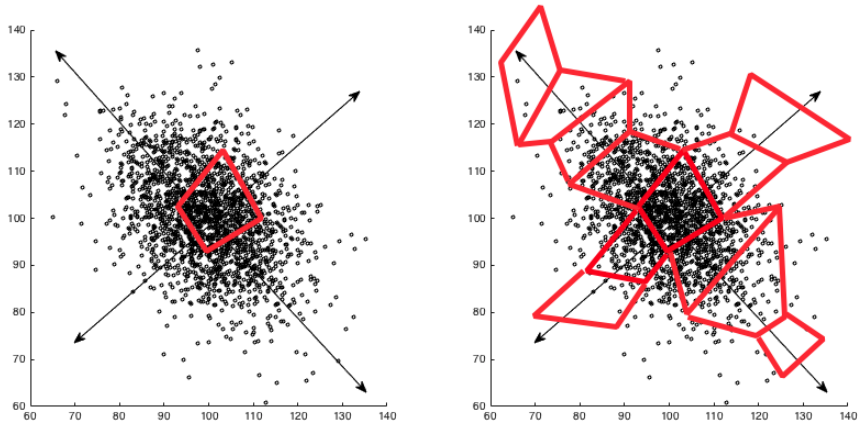
However, the number of critical regions that contain the parameter values can be relatively small in practice. The efficiency of the proposed algorithm is determined by how many critical regions the algorithm finds to obtain those critical regions covering all or most of the parameter vectors, which depends on the specifications of an implementation: the choice of the starting point $\mathbf{b}^0$ (line 1), the method of choosing a previously explored basis $B' \in P$, and the method of choosing a previously unexplored neighboring basis B'' of B' (both in line 16). Depending on the underlying distribution of the sample parameter values (e.g., symmetric but with different spreadness across dimensions; or symmetric in some dimensions but skewed in other dimensions), we can consider different strategies in exploring critical regions. One advantage of the proposed algorithm is that we can improve on its efficiency (i.e., reduce Q) by accounting for the distribution of the sample parameter values. We present such a data-driven approach in Section 3.4 and demonstrate its effectiveness by experimental results in Section 3.5.

Lastly, it is illustrative to view our algorithm using graph theory (see Figure 3.2 in

the next section or Figure 4-13 on page 155 of [11]). Consider a graph G whose nodes correspond to optimal bases and in which there is an edge between two nodes if and only if the two optimal bases are neighbors. Then, the above search algorithm reduces to a node traversal algorithm on the unknown graph G where only a single starting node is initially known. Finding a neighboring critical region corresponds to moving along an edge to discover a new node in G. We also remark that a node in the context of our search algorithm corresponds to an entire critical region, and hence represents a subset of points in the search space, while nodes of a graph often refer to points in the search space in the classical graph theory. Let us refer to an optimal basis as being at level k if the shortest path to the initial optimal basis we find is $k-1$ edges (thus, the initial basis is at level 1 and its neighbors are level 2, etc.). Then, the above complexity comparison indicates that roughly speaking, if our algorithm has to find level 5 neighbors, i.e., Q is in the order of n^4 or higher, then its complexity may become worse than that of the brute-force approach; this can guide a stopping criterion of searching for critical regions containing parameter values in Θ. The goal of the next section is developing a search strategy that limits the order of Q, while covering as many parameter values as possible.

3.4 Solving a Collection LPs: A Data-Driven Algorithm

There is a fundamental challenge in the general search algorithm introduced in the previous section: the geometry of the critical regions is learned with each step in the searching procedure. Thus, the search is over an unknown graph. As the algorithm progresses, it finds a combination of columns of $\mathbb{A}$ that forms an optimal basis, obtains the inverse of the basis matrix, and then determines what portion of Θ is covered by the critical region as in (3.3). Prior to performing the search, however, it is unknown which bases are optimal bases, so one cannot know the shape and location of critical regions to be found. Consequently, one can anticipate neither how many parameter values each critical region would cover nor how many critical regions must be explored to cover all or a specified proportion of sample



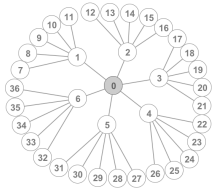
(a) Initial Critical Region (b) Critical Regions on Skeleton

Figure 3.1: Illustration of the Skeleton Building

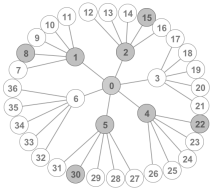
parameter values.

The distribution of the sample parameter values can, however, inform an implementation of our algorithm to guide our search toward the critical regions of interest. In this section, we present an implementation of the general algorithm that exploits the distribution of the sample parameter values using the sample covariance matrix and its spectral decomposition. The sample covariance matrix of the data points, denoted as Σ, along with its eigenvectors, describes how the data points are distributed around the sample mean. In particular, the eigenvector with the largest eigenvalue gives the direction along which the data points are distributed farthest away from the sample mean. Our new algorithm will first search along the directions given by the eigenvectors with the largest eigenvalues, using the tool of one-dimensional parametric RHS LP [11]. Thus, the first phase of our algorithm can be viewed as building a "skeleton" of critical regions, following the directions along which the data points are distributed far away from the sample mean. The next phase will find neighbors of the critical regions on the "skeleton" to cover the rest of the data points. A graphical representation of the skeleton building process is given in Figures **??** and 3.2.

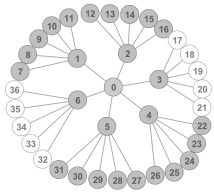
Let $\mathbf{v}_1, ..., \mathbf{v}_m$ be the eigenvectors of Σ, sorted in decreasing order of their corresponding eigenvalues. For each eigenvector of the I largest eigenvalues, where I is an input parameter, we solve a scalar parametric RHS LP along the direction of the eigenvector. For



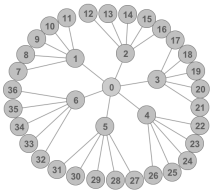
(a) Initial Critical Region



(b) Skeleton



(c) Neighbors of Skeleton Nodes



(d) Neighbors of Neighbors

Figure 3.2: Graph Representation of the Data-Driven Algorithm

a unit eigenvector $\mathbf{v}_i$, consider the following scalar parametric RHS LP: solve

$$\begin{aligned} (\mathrm{P}_{i,\theta}) \ \min_{\mathbf{x}} \ & \mathbf{c}'\mathbf{x} \\ \text{s.t. } & \mathbb{A}\mathbf{x} = \bar{\mathbf{b}} + \theta \mathbf{v}_i \\ & \mathbf{x} \geq 0, \end{aligned}$$

for $\theta \in [\underline{D}_i, \bar{D}_i]$, where $\bar{\mathbf{b}}$ is the sample mean, $\underline{D}_i = \min_{\mathbf{b}\in\Theta}\{\mathbf{v}_i^T(\mathbf{b} - \bar{\mathbf{b}})\}$, and $\bar{D}_i = \max_{\mathbf{b}\in\Theta}\{\mathbf{v}_i^T(\mathbf{b} - \bar{\mathbf{b}})\}$. The lower and upper bounds of θ ($\underline{D}_i$ and $\bar{D}_i$) are, respectively, the minimal and maximal scalar projections of $\mathbf{b} - \bar{\mathbf{b}}$ onto $\mathbf{v}_i$ for $\mathbf{b} \in \Theta$, yielding the farthest projected distances between data points and the sample mean onto the two opposite directions given by $\mathbf{v}_i$.

Solution methods for scalar parametric RHS LPs have been studied extensively in parametric programming literature (see [11] and references therein). The solution algorithm amounts to finding neighbors following the procedure described in the previous section, while the RHS moves along the line segment. Note that, necessarily, $\mathbf{0} \in [\underline{D}_i, \bar{D}_i]$ and problems $(\mathrm{P}_{i,0})$ and $(\mathrm{P}_{\bar{\mathbf{b}}})$ are identical for all $i = 1, \ldots, m$. This means that after solving

an initial LP $(\mathrm{P}_{\bar{\mathbf{b}}})$, we have a starting critical region for all I scalar parametric problems. Solving $(\mathrm{P}_{i,\theta})$ for $\theta \in [\underline{D}_i, \bar{D}_i]$ finds critical regions that intersect the line segment. Since eigenvectors with the largest eigenvalues give the directions with highest variances, finding those critical regions intersecting the line segments along the eigenvectors advances our search away from the initial critical region in the direction where data points are located.

Motivated by this geometric idea for the search, we present a data-driven algorithm for solving batches of LPs with differing $\mathbf{b}$'s sampled from a unimodal probability distribution (Algorithm 2 and 3). Similarly to the general algorithm, the data-driven algorithm starts with an initial critical region, then finds neighboring critical regions in three phases. The first phase performs the one-dimensional parametric search to 'Build the Skeleton'. The second phase ('Search near the Skeleton') finds critical regions around the "skeleton", guided by location of data points yet uncovered. It chooses an explored critical region that covers the most data points, finds the facet of the critical region that has the most uncovered data points on the other side and leads to an unexplored neighbor, explores the neighbor across the facet, then repeats the process. In the last phase ('Cover Outliers'), the LPs for the remaining $\mathbf{b}$'s are solved.

The 'Build the Skeleton' phase of Algorithm 2 can be viewed as giving 'depth' to our search, advancing our exploration in the directions of highest variance and providing a "skeleton" for the remainder of the search. Using the "skeleton" provided, the next phase gives 'breadth' to our exploration, prioritizing the neighbors near the previously found critical regions, and the last phase covers outliers. For the stopping criterion at line 10 in Algorithm 2, we used the following rule in our implementation: as the search goes on, we compute the number of parameter values covered by the most recent q critical regions discovered, denoted as c_q, and stop if $\frac{c_q}{q} < rK$ where r is a pre-specified threshold.

Algorithm 3 A Data-Driven Algorithm for Solving LPs with Different RHSs

Input: A set of data points $\Theta = \{\mathbf{b}^1, \ldots, \mathbf{b}^K\}$ and a positive integer $I \leq m$
Initialize:
Solve $(\mathrm{P}_{\bar{\mathbf{b}}})$ to find the optimal basis B^0
$\Theta \leftarrow$ BatchSolve(Θ, B^0) (see Algorithm 3)
Set of explored bases $\mathcal{P} \leftarrow \{B^0\}$.
Build the Skeleton:
for $i = 1, \ldots, I$ **do**
    Solve $(\mathrm{P}_{i,\theta})$ for $\theta \in [\underline{D}_i, \bar{D}_i]$ and for each optimal basis B obtained, run $\Theta \leftarrow$ BatchSolve(Θ, B) and update $\mathcal{P} \leftarrow \mathcal{P} \cup \{B\}$
Search near the Skeleton:
while a stopping criterion is not met **do**
    Find the basis $B \in \mathcal{P}$ whose critical region covers the most data points
    Find the i^*th neighbor B' of B such that B' is unexplored and the other side of the i^*th facet of B has the most uncovered data points
    $\Theta \leftarrow$ BatchSolve(Θ, B')
    $\mathcal{P} \leftarrow \mathcal{P} \cup \{B'\}$
Cover Outliers:
while Θ is not empty **do**
    Choose a $\mathbf{b} \in \Theta$ and solve $(\mathrm{P}_{\mathbf{b}})$ to obtain its optimal basis B
    $\Theta \leftarrow$ BatchSolve(Θ, B)
    $\mathcal{P} \leftarrow \mathcal{P} \cup \{B\}$

Algorithm 4 BatchSolve(Θ, B)

for $\mathbf{b} \in \Theta$ **do**
    if $\mathbb{A}_B^{-1}\mathbf{b} \geq 0$ **then**
        Store $\mathbb{A}_B^{-1}\mathbf{b} = x^\star(\mathbf{b})_B$ as an optimal solution of $(\mathrm{P}_{\mathbf{b}})$ and $\Theta \leftarrow \Theta \setminus \{\mathbf{b}\}$
for $i = 1, \ldots, m$ **do**
    if $(\mathbb{A}_B^{-1}\mathbb{A})_{(i)} \geq 0$ **then**
        Exclude from Θ those $\mathbf{b}$s satisfying $(\mathbb{A}_B^{-1})_{(i)}\mathbf{b} < 0$ and mark them 'infeasible'
return Θ

3.5 Experimental Results

In this section, we compare the empirical performance of the data-driven approach (DD in short), the brute-force, and two additional approaches using the idea of warm start. The 'warm start' (WS in short) first solves the LP for a randomly chosen parameter vector and then repeats the following two steps: randomly chooses an unsolved parameter vector and solves the LP starting from the optimal solution of the previous parameter vector. The 'warm start with batch solve' (WSBS in short) is the same as the warm start except that every time it solves an LP, it obtains the critical region of the optimal basis and checks if other parameter vectors belong to the critical region [60]. Both of WSBS and DD cover points by critical regions, but WSBS finds critical regions in a random manner while DD uses the distributional information of the sampled data points. The objective of this experimental study is to compare the four approaches and to highlight how properties of given data affect their efficiency.

3.5.1 Problem Instances

In this experimental study, one problem instance requires input of A, c, and Θ, the set of bs to solve the LP. We generated problem instances based on five existing LPs (Table 3.1), including four benchmark problems from the MIPLIB 2010 repository [71] and a linear optimization problem from a real application whose details are explained in Section 3.5.7. Since the problems from MIPLIB are integer programs, we relaxed the integer constraints and solved their LP relaxations. Table 3.1 shows the number of variables and the number of constraints for the five base problems. Based on these five problems, we produced problem instances by generating Θs with different sizes and varying distributional properties. By generating instances in such a way, we evaluated how the performance of the four approaches vary as the size of Θ increases, as the distribution of Θ changes, and as the rest of the problem structure given by A and c varies. In total, we tested for 77 (A, c, Θ) triplets.

Table 3.1: Base problems for generating instances

Problem	# constraints	# variables	Source
enlight13	169	338	MIPLIB
enlight15	225	450	MIPLIB
mik-250-1-100-1	401	652	MIPLIB
roll3000	3,460	4,626	MIPLIB
Access problem	285	15,120	real application

3.5.2 Implementation Setup

We set the following parameters for our implementation of DD (Algorithm 2). First, we choose I, the number of eigenvectors for the algorithm building the "skeleton", to be the smallest positive integer I such that $(\sum_{i=1}^{I} \lambda_i)/(\sum_{i=1}^{m} \lambda_i) \geq 1-\alpha$ where we chose $\alpha = 0.4$ and λ_i is the ith largest eigenvalue of the sample covariance matrix. This approach is common in principal component analysis, where I is interpreted as the smallest number of eigenvectors explaining at least $(1-\alpha) \times 100\%$ of the total variance. For the stopping criterion of the data-driven algorithm, we used $q = 50$ and a threshold $r = 0.001$.

We implemented Algorithm 2 using MATLAB 2012b. For solving LPs in Algorithm 2 (line 3 and 17), we used CPLEX 12.6.3.0. For the other three approaches, we also used CPLEX for solving LPs. We let CPLEX choose a solution method (primal simplex method, dual simplex method, or interior point method), and it chose one of the simplex methods for all instances we tried. We used a computer with Intel Xeon 1.8 GHz CPU and 64 GB RAM.

3.5.3 Evaluating the Data-Driven Approach for Varying Distributions

We evaluated DD and its three phases using the 'enlight15' problem and 10 samples of parameter vectors with size $K = 100,000$. The samples were generated from a multivariate normal distribution of dimension $m = 225$ with mean $\mu = [200, \ldots, 200]^T$ and covariance

matrix Σ, where

$$\Sigma_{i,j} = \begin{cases} 60 & \text{if } i = j = 1, \\ 1.01^i & \text{if } i = j \in \{2, \ldots, 225\}, \\ 0 & \text{if } i \neq j. \end{cases} \tag{3.6}$$

For each of the 10 samples, we ran DD to solve the 100,000 LPs. In Figure 3.3, we show the percentage of data points covered as a function of CPU time for each of the 10 samples. The two vertical lines indicate the average time of the transitions between the 3 phases of our algorithm: the first line at the transition between the skeleton building and the search near the skeleton, and the second line at the transition between the search near the skeleton and covering outliers.

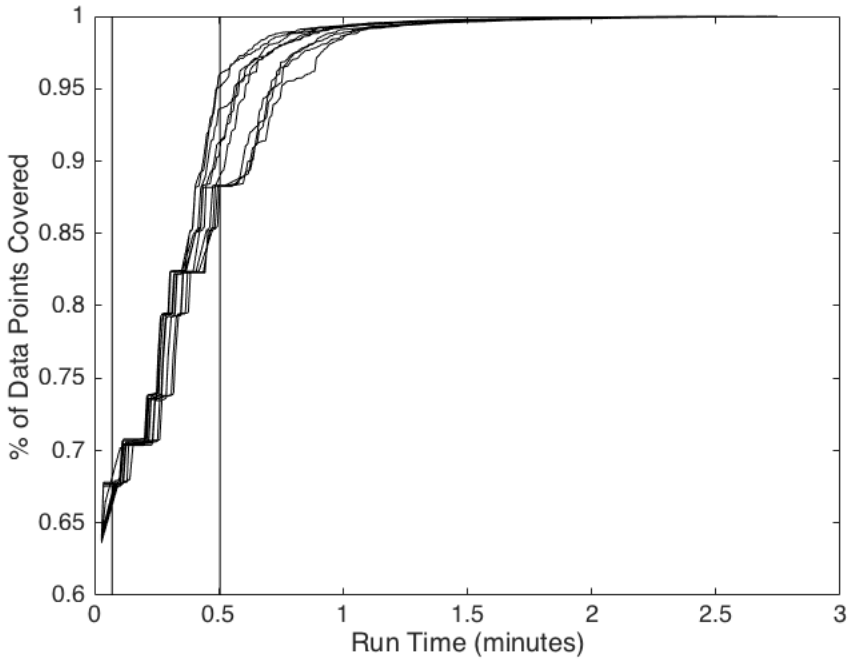


Figure 3.3: Coverage % vs. CPU time of the data-driven algorithm for normally distributed data

In Figure 3.3, over 60% of the sample parameter values were covered by the skeleton building phase after 4.28 seconds on average (the first vertical line in the figure). When

the next phase, in which the algorithm searches near the Skeleton, ends (the second vertical line in the figure), over 85% sample points are covered in 30.32 seconds on average. On average, the data-driven algorithm took 2.57 minutes to solve the 100,000 LPs and explored 458.9 critical regions on average, among which an average of 379.6 critical regions contained data points.

We also evaluated DD for parameter vectors generated from an exponential distribution, which is not symmetric as opposed to the normal distribution in the previous experiment. We again obtained 10 samples of size $K = 100,000$ from the following distribution:

$$\mathbf{b}_i \sim \begin{cases} 200 - \sqrt{60} + \text{Expo}(\sqrt{60}) & \text{if } i = 1 \\ 200 - \sqrt{1.01^i} + \text{Expo}(\sqrt{1.01^i}) & \text{if } i \in \{2, \ldots, 225\}, \end{cases}$$

where $\mathbf{b}_i$ denotes the ith component of $\mathbf{b}$ and Expo(μ) is the exponential distribution with mean μ. Note that the the first two moments of this distribution are the same as those of the normal distribution in the previous experiment.

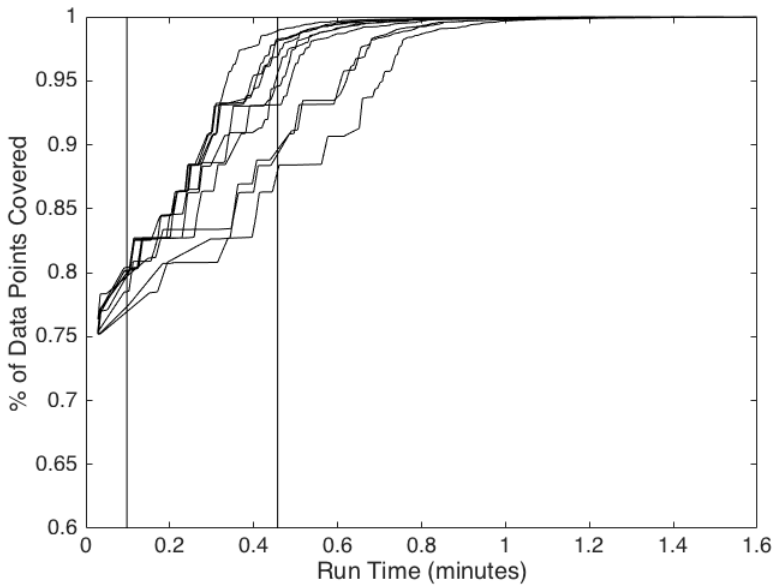


Figure 3.4: Coverage % vs. CPU time for exponentially distributed data

The results were similar to those of the normal distribution. Our algorithm took an average 1.38 minutes to solve 100,000 LPs, exploring 295.0 critical regions on average to find 221.9 critical regions containing data points. In Figure 3.4, we present the coverage percentages of sample parameter vectors by CPU time. Comparing Figures 3.3 and 3.4, we find that the percentage of data points covered at any given time varies more in the exponential case than in the normal case. The skewness of the exponential distribution causes some data points to be located further from those critical regions in the skeleton than they are in the normal case. The second phase of DD (Search near the Skeleton) is guided only by the notion that we should explore neighbors near those critical regions where data points were previously covered. Therefore, a wider spread of data points could lead to less consistent performance during the second phase.

3.5.4 Comparing the Four Approaches for Varying Sample Sizes

In this section, we compare the four approaches for varying Ks. We ran these experiments using 'enlight15' with two kinds of distributions of $\mathbf{b}$, the first distribution is the same as the normal distribution in the previous section and the second is a normal distribution with the same mean vector but all entries of $\mathbf{b}$ have an equal variance. We call the first distribution 'ellipse-shaped' as the entries of $\mathbf{b}$ have different variances and the second distribution as 'sphere-shaped'.

Ellipse-shaped data.

Table 3.2: CPU time comparison for varying Ks and ellipse-shaped data

K	CRs	CPU time (min)			
		DD	WSBS	WS	Brute-Force
100	18	0.01	0.01	0.02	0.05
500	49	0.04	0.03	0.04	0.27
1,000	58	0.05	0.05	0.14	0.56
5,000	107	0.16	0.14	1.15	3.62
10,000	149	0.29	0.25	2.83	8.66
50,000	290	1.24	1.12		
100,000	383	2.59	2.20		
500,000	680	13.18	11.71		
1,000,000	834	28.17	25.31		

Table 3.3: Comparison of DD and WSBS in pivots and membership checks for varying Ks and ellipse-shaped data

K	# Pivots		# Membership checks	
	DD	WSBS	DD	WSBS
100	379	387	204	382
500	584	584	1,825	3,143
1,000	598	614	3,288	5,769
5,000	882	878	21,303	37,191
10,000	1,127	1,141	39,214	60,760
50,000	1,985	1,949	246,232	406,004
100,000	2,655	2,642	605,422	829,656
500,000	4,891	4,844	3,012,348	3,947,233
1,000,000	5,997	5,997	6,369,859	8,627,948

We generated data sets of varying sizes from the ellipse-shaped normal distribution and compared the four approaches. Table 3.2 shows the CPU time of the four approaches. It also shows the number of critical regions containing points for each instance, in the column titled 'CRs'. DD and WSBS are much faster than the other two approaches and WS is faster than the brute force. Since the CPU times of the two slower approaches get large quickly as K increases, we measured them only up to $K = 10,000$. The CPU time of DD is slightly higher than that of WSBS. For K up to 50,000, the CPU times of the two faster approaches grow at a slightly slower rate than the rate at which K grows. For larger Ks, their CPU times grows at higher rates, possibly due to the burden of storing solutions for the large number of points.

For the two faster approaches, DD and WSBS, the majority of their computation consists of performing pivot operations and checking which $\mathbf{b}$s belong to each newly found critical region (called herein 'the membership check'). Table 3.3 shows the number of

pivot operations and the number of membership checks of the two approaches for varying Ks. For WSBS, the number of pivots is the total number of pivots the CPLEX simplex method performed to solve all given LPs. In Table 3.3, DD requires much fewer membership checks than WSBS for all Ks. On the other hand, they performed similar numbers of pivot operations. Since DD uses a search strategy based on the distributional information of the sampled points, it finds critical regions that contain many points earlier than WSBS does. However, because both the pivot operation and the membership check parts of DD were implemented by the authors for research purpose while the pivot operation part of WSBS is done by CPLEX, hence using computing techniques more effectively, the CPU time of DD is only comparable to that of WSBS.

Sphere-shaped data.

Table 3.4: CPU time comparison for varying Ks and sphere-shaped data

K	CRs	CPU time (min)			
		DD	WSBS	WS	Brute-Force
100	12	0.01	0.01	0.02	0.06
500	24	0.02	0.01	0.07	0.26
1,000	27	0.02	0.02	0.15	0.59
5,000	57	0.06	0.06	1.01	3.78
10,000	76	0.11	0.10	2.69	9.14
50,000	142	0.42	0.41		
100,000	173	0.92	0.7		
500,000	281	4.87	4.38		
1,000,000	362	11.19	9.80		

Table 3.5: Comparison of DD and WSBS in pivots and membership checks for varying Ks and sphere-shaped data

K	# Pivots		# Membership checks	
	DD	WSBS	DD	WSBS
100	381	386	135	212
500	443	444	892	1,586
1,000	466	474	1,666	2,866
5,000	618	609	10,186	16,523
10,000	715	716	22,018	33,746
50,000	1,107	1,094	115,390	177,417
100,000	1,266	1,253	244,485	405,349
500,000	1,968	1,967	1,176,249	1,820,552
1,000,000	2,447	2,465	2,337,547	3,834,952

We repeated the experiment of the previous section, but with Θ generated using a different covariance matrix $\Sigma = 3.1347I$, where I is the identity matrix. The scalar 3.1347 was chosen so that the determinant of the covariance matrix remains the same as that of the previous experiment. Table 3.4 shows the CPU time of the four approaches and Table 3.5 shows the number of pivot operations and the number of membership checks for DD and WSBS. Their overall comparison is similar to that of the case of ellipse-shaped data. Again, the number of membership checks is consistently lower for DD, implying that DD finds critical regions covering many points faster than WSBS.

3.5.5 Comparing DD and WSBS for Varying Benchmark Problems

Since it is clear that DD and WSBS are more efficient than WS and the brute-force, we compared DD and WSBS in the remaining experiments. In this section, we compare them for the four benchmark problems from MIPLIB. For each benchmark problem, we gener-

ated Θ of size $K = 50,000$ from an ellipse-shaped multivariate normal distribution. For 'enlight15', we used the same normal distribution as in Section 3.5.3. For the other three problems, we also used similar distributions, but details are slightly different because the scale of the constraints varies across problem. The details of the sampling normal distributions for the other three problems are given in Appendix A.

Table 3.6: Comparison of DD and WSBS for varying benchmark problems, all ellipse-shaped data, $K = 50,000$

Problem	CRs	# Pivots		# Membership checks		CPU time	
		DD	WSBS	DD	WSBS	DD	WSBS
enlight13	125	752	779	98,736	154,801	0.26	0.19
enlight15	290	1,985	1,949	246,232	406,004	1.24	1.12
mik-250-1-100-1	1,288	6,894	6,964	4,000,122	4,525,461	6.71	6.40
roll3000	3	803	609	898,900	114,795	19.63	1.03

Table 3.6 compares DD and WSBS for the four instances in the number of pivots, the number of membership checks, and the CPU time. For the first three instances, the two approaches perform similar numbers of pivots and DD requires a less number of membership checks but is slightly slower than WSBS in CPU time. However, the pattern does not hold for 'roll3000', for which DD performs by far more pivots and is much slower than WSBS. Note that the 50,000 points belong to only three critical regions. Since WSBS chooses an uncovered point and obtains the critical region covering the point, WSBS quickly finds those three critical regions. On the other hand, DD obtains all critical regions in the skeleton and searches some neighbors of the skeleton even though only three critical regions contain points. While the results for the other instances show the effectiveness of the data-driven search strategy, there are problems such as the instance of 'roll3000' in which DD does not perform comparatively or better than WSBS because the distribution of sampled points matches the geometry of critical regions in a particular way. For more results of

'enlight13', 'mik-250-1-100-1', and 'roll3000', see Appendix A.

3.5.6 Effect of Variability in $\mathbf{b}$

We evaluated how the variance of $\mathbf{b}$ affects the efficiency of DD and WSBS. We sampled $\mathbf{b}$ from the same multivariate normal distribution as in Section 3.5.3, but for the covariance matrix we used $t\Sigma$ for $t = 0.5, 0.6, \ldots, 1.5$, where Σ is the original covariance matrix defined as (3.6). For each t value, we generated a data set of size $K = 100,000$ for which we solved the 'enlight15' problem using DD and WSBS.

Table 3.7: Comparison of DD and WSBS for varying constants multiplied to the covariance matrix, $K = 100,000$

t	CRs	# Pivots		# Membership checks		CPU time	
		DD	WSBS	DD	WSBS	DD	WSBS
0.5	74	659	669	128,973	175,059	0.33	0.29
0.6	121	933	941	158,319	248,565	0.56	0.5
0.7	176	1,329	1,304	241,501	354,040	0.95	0.78
0.8	222	1,521	1,520	291,515	470,933	1.22	1.06
0.9	293	2,034	2,014	369,732	570,993	1.71	1.52
1	383	2,647	2,623	518,651	793,193	2.46	2.25
1.1	488	3,511	3,494	664,643	1,006,703	3.45	3.17
1.2	614	4,241	4,311	1,021,733	1,308,266	4.99	4.51
1.3	716	5,242	5,245	1,534,780	1,706,076	6.18	5.84
1.4	868	6,352	6,358	2,126,265	2,253,058	8.37	8.06
1.5	990	7,162	7,138	2,997,976	2,605,891	11.09	10.01

Table 3.7 compares DD and WSBS for varying scalars multiplied to the covariance matrix. The number of pivots is similar for the two approaches across all t values. However, the number of membership checks of DD is consistently lower except for $t = 1.5$. The CPU

time of the two approaches is similar.

3.5.7 Access Measure Application

We also evaluated our approach for an LP from a specific application, measuring spatial access to healthcare services over a geographic area [17, 20]. Because spatial access is a manifestation of the dynamics between a service provider network and the need or demand for the service, its estimation requires modeling the 'matching' of service providers with need/demand. Optimization models have been previously considered to derive such matching that optimizes an overall system performance such as the total travel distance under a series of realistic constraints related to need and supply. The resulting optimization problem is an LP and the output solution represents the matching between supply and demand. Measures of spatial access are formalized as linear functions of the output solution, thus subject to variations of the parameters in the optimization model.

We considered a simplified version of the model in [17] for estimating access to pediatric specialized asthma care within the geographic area of Savannah, Georgia. The LP has a variable x_{ij} for each pair of a census tract $i \in I$ and a healthcare provider $j \in J$, representing the assignment of the provider's capacity for care to patients living at the census tract. Then, the following LP is solved to find a matching between demands in census tracts and providers:

$$\begin{aligned} \min \ & \sum_{i \in I} \sum_{j \in J} d_{ij} x_{ij} \\ \text{s.t.} \ & \sum_{i \in I} x_{ij} \leq u_j \text{ for } j \in J \\ & \sum_{j \in J} x_{ij} = v_i \text{ for } i \in I \\ & x \geq 0, \end{aligned}$$

where d_{ij} is the travel distance between census tract i and provider j, u_j is the capacity of

provider j, and v_i is the demand of census tract i. The LP for Savannah has 15,120 variables and 285 equality constraints in standard form. The capacity parameters (u_j's) were sampled from a discrete distribution, while the demand parameters (v_i's) were sampled from a multivariate normal distribution. Details of the sampling procedures for both sets of parameters and their data sources can be found in the online supplementary material. We generated samples of varying sizes, and compared DD and WSBS. The results are shown in Table 3.8.

Table 3.8: Comparison of DD and WSBS for the access measure LP

K	CRs	# Pivots		# Membership checks		CPU time	
		DD	WSBS	DD	WSBS	DD	WSBS
100	12	156	105	321	307	0.19	0.03
500	20	135	129	1,188	1,772	0.18	0.03
1,000	20	167	108	4,665	2,761	0.25	0.05
5,000	23	183	138	19,893	15,189	0.3	0.05
10,000	23	174	150	39,683	34,040	0.31	0.08
50,000	31	160	166	97,252	146,557	0.39	0.16

In Table 3.8, DD is slower than WSBS in CPU time across all Ks. Although not shown in the table, we also tested WS, but it took about 2 hours to solve 10,000 LPs as opposed to DD and WSBS which took less than a minute. DD also performed more pivots and membership checks than WSBS for Ks up to 10,000. However, for $K = 50,000$, DD performed less pivots and much less membership checks than WSBS.

The performance of DD can be explained by the mismatch between the distribution of sampled data points and the type of the distributions DD is designed for. The RHSs of the access problem are sampled from both discrete and normal distributions, thus there may not be many sampled $\mathbf{b}$s near the sample mean. Moreover, the data points may not be located along the eigenvectors of the sample covariance matrix around the sample mean.

In contrast, Algorithm 2 was designed mainly for unimodal continuous distributions and demonstrated good performance for such distributions. However, as K gets large (such as $K = 50,000$ in Table 3.8), the performance of DD gets better. It is possible that for a small K the sampled points are well separated into several groups due to the discrete distribution of some part of $\mathbf{b}$, but as K increases the sampled points are less separated because of the normal distribution part of $\mathbf{b}$, and thus the performance of DD improves. In sum, the above results illustrate the importance of matching the premises of a specific implementation of our general approach and the underlying distribution of the given data.

3.6 Conclusion

Solving a large collection of LPs with different parameters arises in various contexts, such as sensitivity analysis and stochastic programming. We developed a new approach to solving a large number of LPs with the same dimension and structure, but different parameter values, by borrowing information across "similar" LPs where similarity is in the sense of geometric proximity of the critical regions. We formally established the geometric properties of critical regions, and using the results, introduced a general approach for solving a collection of LPs in batches. Further, we developed a data-driven algorithm based on our general approach. The experimental results on a set of benchmark problems with varying data sets demonstrate the computational efficiency of the proposed approach over other approaches. The results for various problem instances show that the computational load (the numbers of pivots and membership checks) of our approach grows more slowly than other approaches as K, the sample size, increases. However, when the variance is large, the batch solution methods, DD and WSBS, become less efficient while the performance of the brute-force, solving each individual LP from scratch, can provide similar results. In addition, the advantage of our approach diminishes when the underlying distribution of the sampled parameter values do not fall into the class of distributions for which our specific implementation is intended.

There are a number of future research directions. First, the empirical comparison of the four approaches raises a fundamental question for practice: for what kind of linear programs and sample parameter values the proposed approach is better (or can be tailored to be better) than the other approaches? Our experimental results show that the specifics of the proposed approach should be tailored to take into account the distribution of the sample parameter values, resulting in a more efficient algorithm than the other approaches. In cases where sampled parameters values are generated from a multimodal distribution, one can consider clustering given parameter values using a clustering algorithm and then applying our data-driven approach to each cluster, possibly using distributed computing. Another future direction is further theoretical analysis comparing the worst case guarantee of the computational complexity for the various algorithms. If one can derive a condition under which Q (the number of critical regions found by the proposed method) is bounded by some expression of problem size and other parameters ($\mathbb{A}$ and $\mathbf{c}$), then we will be able to better choose which method to use under the conditions. Lastly, we considered only the cases where uncertain parameters reside in the right hand side of the constraints, but in practice, the coefficients in the objective function may be uncertain as well. Thus, we need to solve an LP for a large number of given (c, b) pairs, which is called the RIM case in parametric programming [72]. Extending the theory of critical region and the computational method to the RIM case is another potential future research direction.

CHAPTER 4
STATISTICAL INFERENCE FOR OPTIMIZATION MODELING: IMPLICATIONS IN DECISION MAKING FOR IMPROVING ACCESS TO DENTAL CARE

4.1 Introduction

Optimization is a mathematical model that is widely researched in operations research (OR) and applied to many areas including healthcare, transportation, and manufacturing. [73, 74] In healthcare, optimization models have been used to determine the best location for new clinics [75, 76], to ensure appropriate coverage by ambulances across a network [77], to route nurses for home health services [78], and to evaluate policies for pandemic influenza [79], for evaluating service access [80], among other examples.

In deterministic optimization, one of the most common optimization approaches in healthcare OR, the input parameters specifying the model are assumed exact physical parameters; however, they commonly are derived from uncertain data about systems' behavior. Thus using deterministic optimization, the derived optimal decision is assumed to be one "best" decision with no uncertainty. While making one best decision given limited information is key to the operation of any system, it is also important to provide additional insights on how uncertain such decision might be, a concept at the core of statistical inference. This is essential in decision making because in some cases statistical inference may suggest that more data need to be acquired to provide stronger evidence for a decision; in others, they may prompt not making a decision at all because of the high uncertainty of the decision environment. [81, 13].

In this paper, we introduce a statistical inference approach for informed decision making on healthcare access using optimization modeling. The focus is on two dimensions of

healthcare access, specifically, accessibility and availability of services, together referring to *spatial access*. The access model is an assignment optimization, matching the underlying networks of need (patients) and supply (providers) of healthcare to quantify access for communities within a geographic region, jointly for subpopulation with different financial access (Medicaid vs. non-Medicaid vs. without affordability of care) [82, 83, 20].

Uncertainty propagates into the derived access measures through imprecise knowledge of the supply and demand/need inputs of the optimization model. We develop a sampling approach to obtain a sample of instances for the supply and need inputs to quantify how uncertainty in the input data impacts informed decision making for improving access. The sampling distributions are established separately for the need of and supply inputs; the distributions are specified using a hierarchical model with hyperparameters informed by prior information derived from multiple data sources. For each sampled need and supply input, we derived the access measure using the access model with the sampled input. The final output consists of an empirical sample of of plausible values for the access measure for each community in a geographic region of interest rather than one access estimate as provided in the existing access modeling approaches. Using this empirical sample, we can derive summary statistics as well as confidence intervals for the access measure for each community. Furthermore, we apply simultaneous inference techniques to make inference across geographically related communities.

This study is motivated by the need of addressing rigor in estimation of access, and in evaluating and designing interventions to improve access to dental care for children.[84] The Centers for Medicare and Medicaid Services (CMS) promoted the Oral Health Initiative to increase utilization of dental services, citing lack of access to dental care as the most frequently reported reason for children not receiving dental care services.[85] Lack of access is associated with poor oral health outcomes [86], which have been associated with diminished overall health.[87] We illustrate the approach using the state of Colorado, one of the most progressive states in terms of oral health policy.

4.2 Methods

4.2.1 Data Sources

We used data from the 2017 American Community Survey (ACS) of the Census Bureau to obtain the study population consisting of Colorado children, aged 0 to 18 years at the census tract level. We used the ACS to divide the study population by financial access, differentiated into four categories: uninsured and without affordability to pay out-of-pocket care expenses (herein called *without financial access*); Medicaid-insured; Children's Health Insurance Program (CHIP)-insured; or with other affordability, including commercial insurance or out-pocket expenses (herein called *private financial access*). See Web-Appendix C.1.4 for details. We note here that CHIP is referred to in Colorado as Children's Health Plan Plus (CHP+). We us the abbreviation CHP+ for the remainder of the paper.

To derive the provider-level information on the dental care providers, including dentists and dental hygienists, we used the data from the Board of Dentistry and National Provider Plan and Enumeration (NPPES). We linked these data to the InsureKidsNow.gov data to derive the providers' participation in public insurance programs, including Medicaid and CHP+. See Web-Appendix C.1.5 for details.

The sampling model for the provider-level caseload was informed by multiple data sources including the 2012 Medicaid Analytical eXtract (MAX) claims, the ACS, the Health Policy Institute (HPI) of the American Dental Association reports on supply, and the Rural-Urban Continuum Codes (RUCCs) developed by the United States Department of Agriculture.

4.2.2 Access Model

We focused on two access dimensions: availability and accessibility. *Availability* pertains to scarcity or congestion of service providers. *Accessibility* pertains to travel impedance between demand/need points and service sites.

The access model replicated how supply of dental care is accessed by children in need of dental care by matching need and supply of dental care services to minimize the total distance traveled under access constraints using optimization modeling. Similar models have been implemented for measuring healthcare access to pediatricians [20], pediatric primary care [18], adult primary care [19], pediatric asthma care [17] and dental care in Georgia [82] Because some providers do not accept or limit the number of Medicaid and/or CHP+ patients [88, 89], the model accounted for the differentiation of the study population by financial access. The access model also accounted for the spatial dependence due to geographic similarities in access for nearby communities [90]. Details of the optimization model are provided in Web Appendix C.1.1.

The access measure of interest was the average travel distance to services, a linear function of the matching derived from the optimization model. [91, 83] The access measure was derived at the census tract level. Details are provided in the Web-Appendix C.1.2.

4.2.3 Statistical Inference for Access Measures

The access model provides (point) estimates of the decisions patients make to access dental care providers, further used to estimate the access measures. Because the estimated decisions and access measures are derived from a model with uncertain inputs, they are also uncertain; in statistical terms, the estimates are realizations of random variables following a sampling distribution. Determining the sampling distribution of these point estimators requires understanding how uncertainty propagates from the input parameters of the optimization model into the output solutions.

We used a Bayesian approach because of the limited data available to derive the sampling distribution. Specifically, we modeled the demand/need and the provider caseload parameters using a hierarchical Bayesian approach, leveraging data from various sources to inform the empirical prior distributions of the hyperparameters as described in the next sub-section. The sampling distribution was then the posterior predictive distribution.

Not all model parameters were assumed to be uncertain; some of the parameters of the optimization model were assumed fixed, particularly those that could be specified by policies and guidelines. A complete list of the optimization parameters is in Web-Appendices C.1.1 and C.1.1.

Sampling Distribution of the Population Parameters

Demand data entered the access model through the total number of visits per-year needed by the study population within each census tract, computed as the product of the number of children in that tract and the average visits demanded per-child (assumed fixed and equal to 1.5).

The total population of the ith census tract and of the entire state of Colorado were denoted P_i and POP, respectively, for $i \in \{1, \ldots, \mathbb{C}\}$. The total population POP was sampled from a negative binomial distribution, with hyperparameters chosen to mimic the uncertainty level of the total population, given by the confidence interval provided by the ACS data.

Given the sampled POP, the census tract population was sampled from a multinomial distribution: $(P_1, \ldots, P_{\mathbb{C}}) \sim Mulitnomial(POP, \pi)$, where π is the vector of population proportions of all census tracts. The prior distribution placed on π is Dirichlet, with hyperparameters estimated from ACS data using an empirical Bayes approach.

We denote the number of children without financial access, with Medicaid, CHP+ or private financial access of the ith census tract by $N_i^U, N_i^M, N_i^H, N_i^O$, respectively. Given the sampled population of the ith census tract, we modeled the population parameters of the ith census tract by $(N_i^O, N_i^M, N_i^H, N_i^U) \sim Multinomial(P_i, \tau_i)$, where τ_i is the vector of population proportion hyperparameters.

Details of the sampling approach are found in Web-Appendix C.1.4.

Sampling Distribution of the Provider Caseload Parameters

The provider-level pediatric caseload was measured in number of visits dedicated to child patients. The provider-level caseload entered the optimization model through the inputs C_j, C_j^M, C_j^H, where C_j is the total number of visits offered and C_j^M, C_j^H are the caseload dedicated to the publicly insured, differentiated by financial access (Medicaid and CHP+, respectively) from provider j for $j = 1, \ldots, \mathbb{J}$.

We assumed that the overall pediatric caseload for the jth provider follows a Poisson distribution with parameter $\lambda_{j|t}$ given provider's t taxonomy. The prior distribution of λ_j was assumed a gamma distribution, $\lambda_{j|t} \sim Gamma(a_t, b_t)$, with hyperparameters a_t and b_t depending on the provider taxonomy t of the jth provider. We used an empirical Bayes method to inform the selection of hyperparmeters of the priors using the Medicaid caseload data. Specifically, the caseload dedicated to publicly-insured children versus other children was modeled as a binomial distribution, where the percentage of the caseload dedicated to Medicaid patients is sampled from a Beta distribution. We used the MAX claims data for deriving the sampling distribution for the caseloads for Medicaid-insured and CHP+-insured sub-populations. We derived the realized caseload for the Medicaid participating providers, by their taxonomy and urbanicity using prior research [92].

Details are given in Web-Appendix C.1.5.

Statistical Inference on Access Measures

We sampled the posterior predictive distributions described in Sections 4.2.3 and 4.2.3 to obtain $B = 1000$ instances of the input parameters of the optimization model. For each instance of the input parameters, we solved the optimization problem and derived the access measures. We used these samples of the access measures: (1) to evaluate the uncertainty levels in the access measures by rurality level and by financial access; and (2) to assess policy implications on geographic access and disparities by financial access.

For the first objective, we computed the interquantile range (IQR) of the access mea-

sure samples and of the disparity between publicly-insured children and those with private financial access for each census tract. Each census tract was classified into urban, suburban and rural using the RUCCs of its corresponding county; we classified tracts with RUCCs 1-3 as urban, tract with RUCCs 4-6 as suburban, and tracts with RUCCs 7-9 as rural.

To evaluate georgraphic access, the access measures are differentiated by insurance type, U_i, Z_i, Y_i, and W_i describing the average distance traveled by publicly (Medicaid and CHP+) insured, Medicaid insured, CHP+ insured, and privately insured in census tract i, respectively. For each of sample of input parameters, the optimization model is solved and the access measure are computed, resulting in a sample of B estimates of the access measures for each census tract Let $(\hat{u}_{ib}, \hat{z}_{ib}, \hat{y}_{ib}, \hat{w}_{ib})$ represent the bth sample of the access measures for the ith census tract derived from solving the optimization model using the bth sample of the input parameters from the posterior distribution. Point estimates of access for each financial access and census tract are then taken to be the medians of the B samples.

To evaluate disparities by financial access, the access measures are differentiated by insurance type, U_i and W_i describing the average distance traveled by publicly (Medicaid and CHP+) insured and privately insured in census tract i, respectively. The disparity between the publicly and privately insured is defined by the difference in access between the groups, $\Delta_i = U_i - W_i$. Disparities in spatial access can be estimated in the bth sample by the differences $\delta_{ib} = \hat{u}_{ib} - \hat{w}_{ib}$. Similar disparity measures can be constructed for measuring disparities between the Medicaid and privately insured populations, as well as between the CHP+ and privately insured populations. Since those with private financial access are taken to be the baseline for each of the three comparisons, we refer to disparities by the specific publicly insured population (Medicaid, CHP+, or all public insurance).

We perform simultaneous statistical inference on the access measures and disparities between the two populations across multiple locations while accounting for spatial dependence using simultaneous confidence bands. The simultaneous bands account for both the spatial dependencies and varying levels of spatial uncertainty, while creating bands that

are sufficiently minimal in width to detect significant differences in access. [93, 94] Using simultaneous confidence bands, we can test whether the access measures is significantly more than some access standard T (e.g. 15 miles in distance traveled). Furthermore, we can test whether the disparity in access between those with public insurance and those with private insurance is lower than some intervention threshold value T (e.g., 5 miles in distance traveled). This hypothesis indicates the presence of systematic disparities at the threshold level T. This hypothesis indicates the presence of systematic disparities at the threshold level T. Complete detail on inference procedures are provided in Web-Appendix C.1.6.

4.3 Results

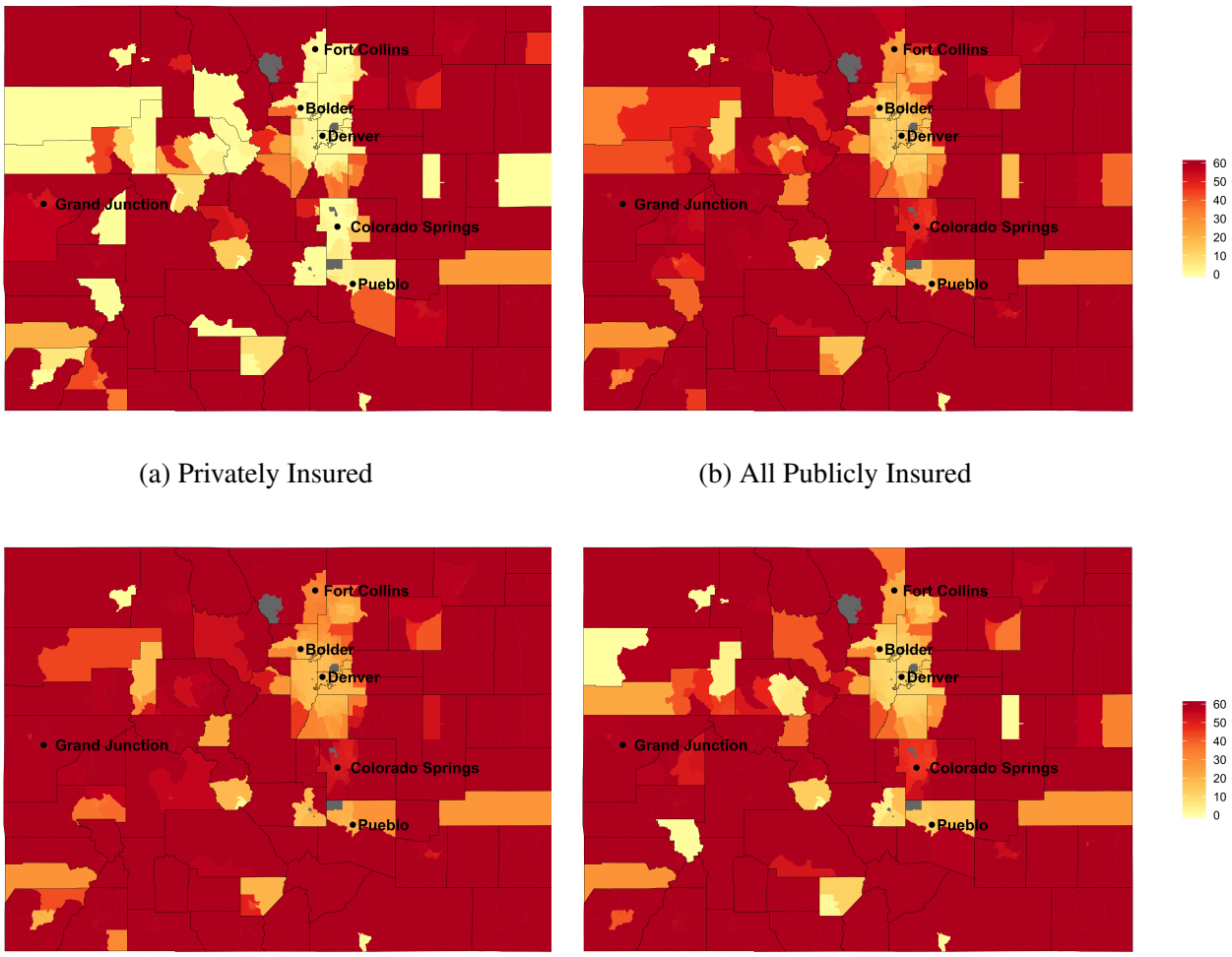
4.3.1 Overall Dental Supply

We identified the full time equivalent of approximately 3,598 general dentists, 182 pediatric dentists, and 263 independent dental hygienists operating across 3,051 provider locations. Among these providers, 67.5% of the general dentists, 86.3% of the pediatric dentists, and 100% of the independent dental hygienists accepted public insurance. These results are roughly equivalent to those found by the Health Policy Institute [95] and the Kaiser Family Foundation [96]. Differentiating by Medicaid and CHP+, we found that 34.4% of general dentists, 64.1% of pediatric dentists, and 88.7% of independent hygienists accepted Medicaid, while 54.9% of general dentists, 61.3% of pediatric dentists, and 27% of independent hygienists accepted CHP+.

4.3.2 Access To Dental Care

Access Measures by Census Tract

Figure 4.1 displays access maps for the median access measure across the $B = 1000$ posterior samples. Median access to dental care was better in urban areas than suburban



(a) Privately Insured

(b) All Publicly Insured

(c) Medicaid Insured

(d) CHP+ Insured

Figure 4.1: Median access to dental care, by financial access

and rural areas, across insurance types. The privately insured population also had better access to care than the publicly insured, with the Medicaid insured population experiencing worse access than those insured by CHP+.

For urban tracts, the median travel distance was 2.46 (25th percentile = 1.38, 75th percentile = 2.71) miles among the privately insured. Among the publicly insured, the median travel distance was 15.56 (13.71, 24.78) miles , with the Medicaid population traveling a median distance of 18.16 (17.75, 31.61) miles and the CHP+ population traveling a median distance of 11.13 (10.83, 20.77) miles.

For suburban tracts, the median travel distance was 43.49 (0.97, 58.02) miles among the privately insured. Among the publicly insured, the median travel distance was 46.94 (24.21, 57.96) miles , with the Medicaid population traveling a median distance of 56.38 (31.18, 59.99) miles and the CHP+ population traveling a median distance of 54.08 (7.65, 59.63) miles.

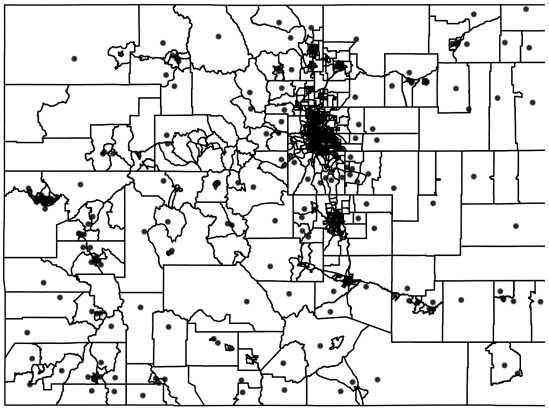
For rural tracts, the median travel distance was 50.52 (0.99, 59.99) miles among the privately insured. Among the publicly insured, the median travel distance was 55.53 (32.80, 60) miles , with the Medicaid population traveling a median distance of 59.12 (48.41, 60) miles and the CHP+ population traveling a median distance of 54.43 (35.64, 60) miles.

Low Access Areas

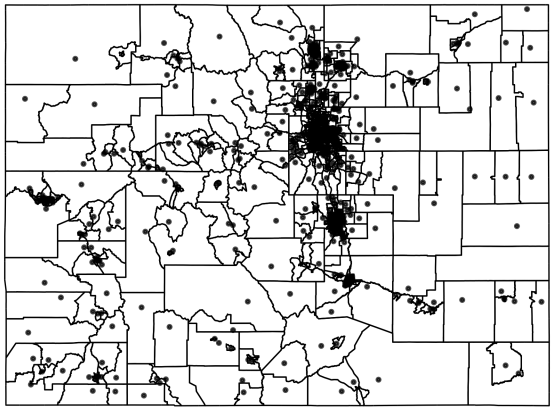
In this section, we identify areas that have access to dental care services significantly below access standards. We use the standards of $T = 15$ miles for urban tracts and $T = 30$ miles for suburban and rural tracts.[97] For the remainder of this section, we will simply refer to the access standard as T, without differentiating by urbanicity.

Points maps highlighting census tracts with median access measures greater than the access standard are displayed in Figure 4.2 by financial access. Then median access measure was greater than the access standard in 15.8%, 59.9%, 94.3%, 35.1% of census tracts for those with private financial access, all publicly insured, those insured through Medicaid, and those insured through CHP+, respectively.

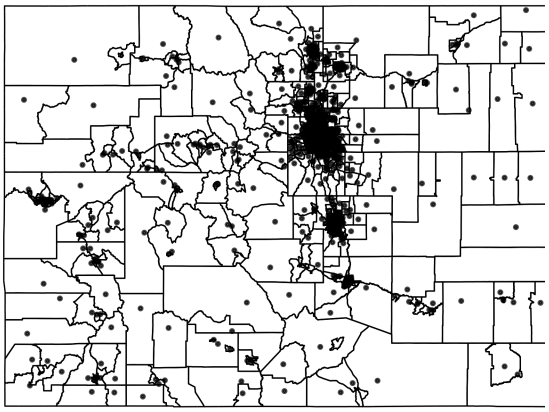
Although the median access measure was greater than the standard in many census tracts, there was, however, considerable variability in the posterior samples of the access measures. Figures C.7 and C.8 in Web Appendix C.1.6 show the census tract level standard deviation and the IQR of the access measures across runs of the optimization model. Because high uncertainty makes identifying areas not meeting access standards more difficult, we account for the varying levels of spatial uncertainty using the simultaneous inference procedure described in Section 4.2.3. A point map showing census tracts detected to have



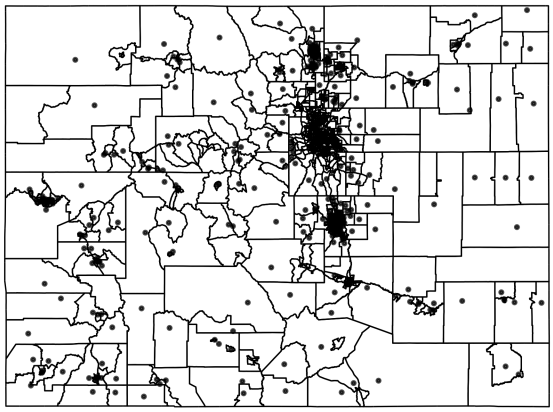
(a) Private Insurance (15.8% of tracts marked)



(b) Public Insurance (59.9% of tracts marked)

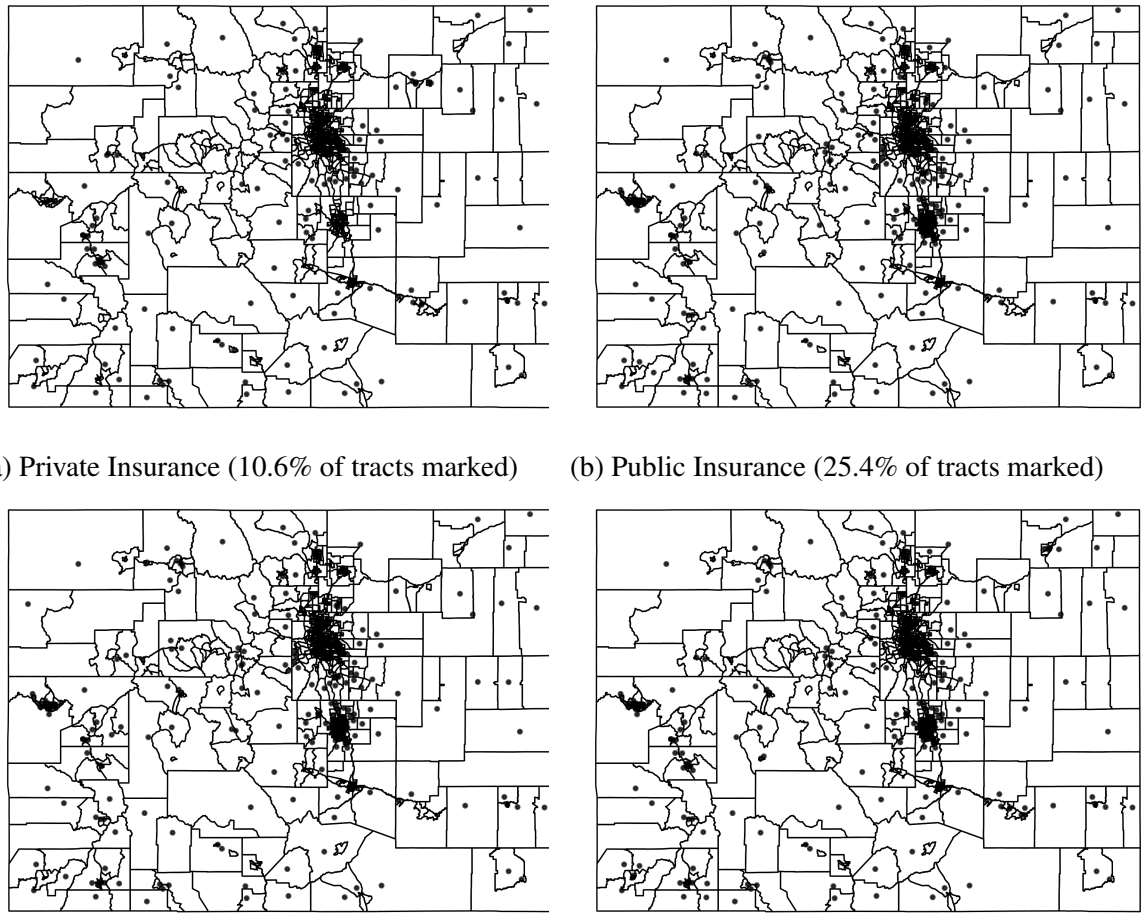


(c) Medicaid (94.3% of tracts marked)



(d) CHP+ (35.1% of tracts marked)

Figure 4.2: Census tracts with median access measures greater than the access standard, by financial access type



(a) Private Insurance (10.6% of tracts marked)

(b) Public Insurance (25.4% of tracts marked)

(c) Medicaid (25.2% of tracts marked)

(d) CHP+ (25.2% of tracts marked)

Figure 4.3: Census tracts with median access measures significantly greater access standard, by financial access type

access measures significantly more than the access standard are displayed in Figure 4.3.

Because of the high uncertainty the access measure for some census tracts, particularly in the western portion of Colorado, we did not detect access measures significant worse than the standard in many of the census tracts identified as having large differences in Figure 4.2 when performing simultaneous inference. We do, however, detect census tracts with access to care significantly worse than the standard. We observe that 25.4% of census tracts had median access measures significantly above the access standard for the publicly insured, while only 10.6% of census tracts had median access measures significantly over the access standard for those with private financial access. Among the publicly insured,

Table 4.1: Percentage of census tracts with median access measures greater than and *significantly* greater than the access standard, by urbanicity and type of financial access

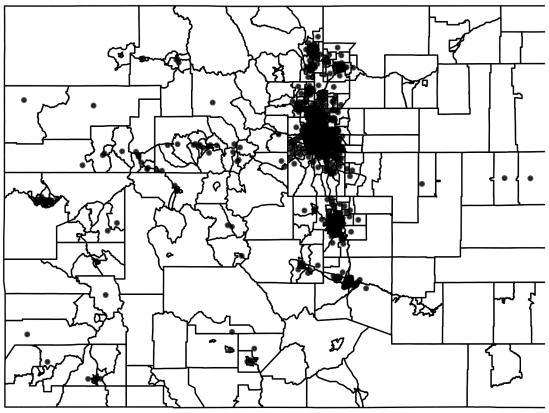
	Urban		Suburban		Rural	
Financial Access	% Above Std	% Sig	% Above Std	% Sig	% Above Std	% Sig
Private	6.3%	3.2%	58.2%	35.8%	54.3%	40%
Public	54.9%	19.8%	74.6%	35.8%	84.3%	47.1%
Medicaid	95.8%	19.4%	77.6%	34.4%	85.7%	47.1%
CHP+	26.5%	18.3%	68.7%	52.2%	77.1%	57.1%

Medicaid and CHP+ populations were observed to have access measures significantly over access standard in 25.2% of census tracts, each. Furthermore, we detected many more suburban and rural census tracts failing to meet the access standards than urban census tracts. Table 4.1 provides a full breakdown of the percentage of census tracts, by urbanicity, with median access measures above the access standard, as well as the percentage of census tracts found to have access measures significantly above the standard.

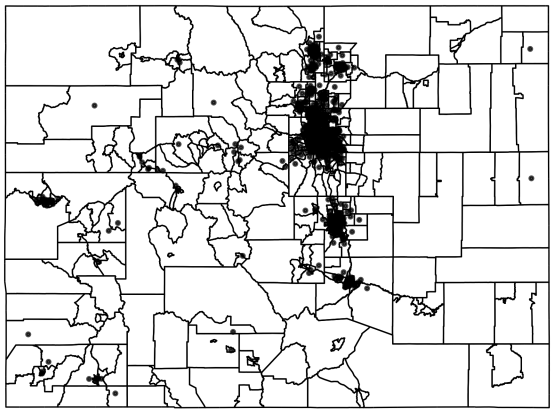
4.3.3 Disparities in Access to Dental Care

We define disparities in access for each census as the difference in distance average traveled between a publicly insured population (Medicaid, CHP+, or all publicly insured) and those with private financial access. Thus, for the bth sample of input parameters, disparities are estimated by the differences, δ_{ib} (see Section 4.2.3). The median disparity was typically positive for the publicly insured, with the Medicaid insured population experiencing even greater disparities than those insured through CHP+. Web Appendix C.1.6 displays disparity maps for the median disparity across the $B = 1000$ posterior samples (see Figure C.9) and provides a full description of the median disparity by census tract.

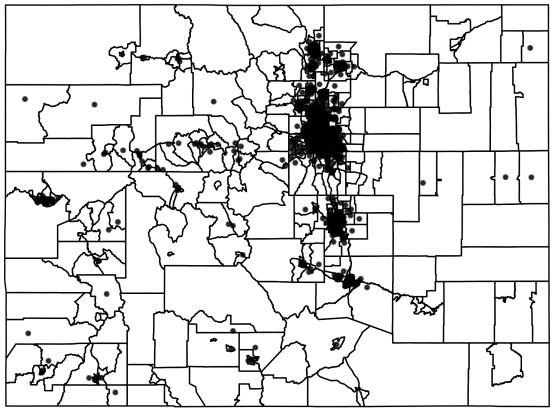
We now seek to identify disparities which are especially large by examining which disparities are greater than some threshold T. Points maps highlighting census tracts with median disparity in access greater than $T = 5$ miles are displayed in Figure 4.4. The median disparity was greater than $T = 5$ miles in 87.2%, 87.3%, and 84.3% of census tracts for all publicly insured, Medicaid insured, and CHP+ insured, respectively. Web



(a) Medicaid (87.3% of tracts marked)



(b) CHP+ (84% of tracts marked)



(c) Public Insurance (87.2% of tracts marked)

Figure 4.4: Census tracts with median disparities greater than $T = 5$ miles, by financial access

Appendix C.1.6 reports the corresponding percentages for thresholds of $T = 2$ and $T = 10$ miles.

There was, again, considerable variability in the posterior samples of the disparity measures for many census tracts. Figures C.10 and C.11 in Web Appendix C.1.6 show the census tract level standard deviation and the IQR disparity measure. Importantly, the variability in the disparity measures was considerably higher in rural and suburban areas than in urban areas. Figure 4.5 shows the census tract level IQR of the disparity measures for public insurance, grouped by urbanicity. The variability was similar examining disparities for by Medicaid and CHP+ (see Figure C.12 in Web Appendix C.1.6).

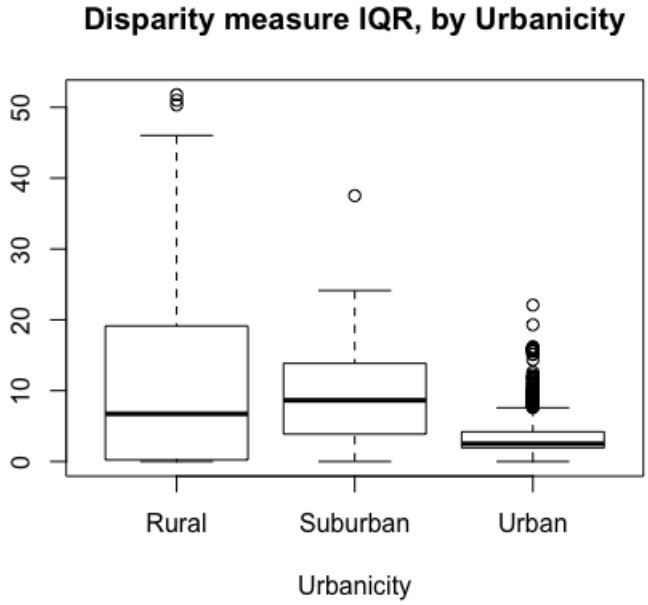
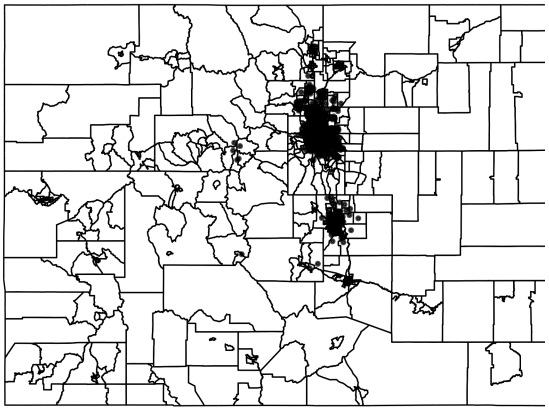


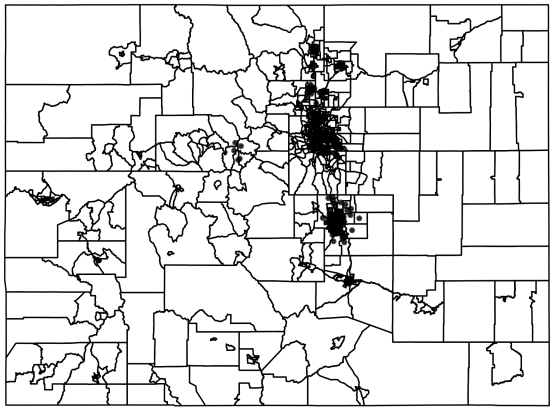
Figure 4.5: Boxplot of census tract IQR of disparity measures for publicly insured, grouped by urbanicity

We again applied the simultaneous inference procedure described in Section 4.2.3 to detect census tracts with statistically significant disparities in dental care access. Census tracts with significant disparities of at least 5 miles are displayed in Figure 4.6. Similar figures for disparities significantly over 2 and 10 miles are provided in Figures C.13, C.14, and C.15 of Web Appendix C.1.6.

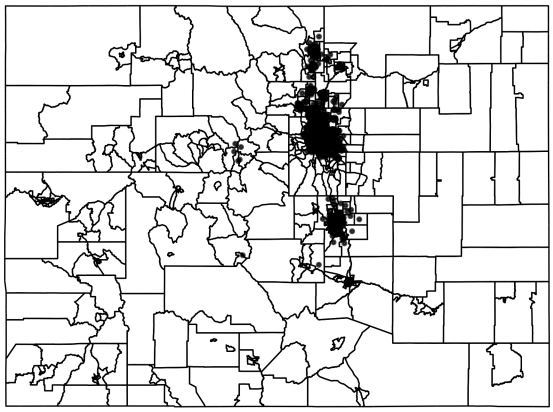
Because of the high uncertainty the access measure for some census tracts, particularly in suburban and rural census tracts, we do not detect significant disparities so many of the communities identified as having large differences in Figure 4.4 when performing simultaneous inference. We do, however, detect significant disparities in access to dental care between the publicly and privately insured clustered around the Denver and Colorado Springs metropolitan areas. More generally, we find statistically significant disparities of at least $T = 5$ miles for the publicly insured in 73.2% of urban census tracts, while only finding significant disparities in 4.5% and 5.7% of suburban and rural census tracts, respectively. Table 4.2 provides a full breakdown of the percentage of census tracts, by urbanicity,



(a) Medicaid (63.1% of tracts marked)



(b) CHP+ (12.5% of tracts marked)



(c) Public Insurance (61.7% of tracts marked)

Figure 4.6: Census tracts with disparities significantly greater than $T = 5$ miles, by financial access

Table 4.2: Percentage of census tracts with disparity measures greater than and *significantly* greater than $T = 5$ miles, by urbanicity and type of financial access

	Urban		Suburban		Rural	
Financial Access	% Above Std	% Sig	% Above Std	% Sig	% Above Std	% Sig
Public	80.5%	60%	2.8%	0.2%	2.5%	0.2%
Medicaid	81.1%	62.3%	2.9%	0.3%	2.2%	0.3%
CHP+	80%	12%	1.5%	0.2%	1.5%	0.2%

with median disparity measures above the threshold $T = 5$ miles, as well as the percentage of census tracts found to have disparities significantly above the 5 miles. Tables C.8 and C.9 of Web Appendix C.1.6 repeat the analysis using thresholds of $T = 2$ and $T = 10$ miles, respectively.

Those insured through CHP+ tend to experience fewer disparities than those insured through Medicaid. Overall, we detect disparities significantly greater than $T = 5$ miles in 61.7% of census tracts and greater than 10 miles in 12.2% of census tracts for all forms of public insurance. However, when we examine Medicaid and CHP+ separately, we find that 63.1% of census tracts have Medicaid populations experiencing significant disparities greater than $T = 5$ miles, while only 12.5% of census tracts have CHP+ populations experiencing significant disparities greater than $T = 5$ miles.

4.4 Discussion

In this paper, we introduced a statistical inference scheme for informed decision making on healthcare access. This approach builds on the optimization methodology for measuring access recently developed for studying access [20, 17, 19, 82, 18] by offering a rigorous framework for understanding how uncertainty propagates from uncertainty input parameters into the optimal solutions and, ultimately, the measurement of spatial access. Early attempts to derive the distribution of the optimal objective value for LP models have been considered for stochastic optimization [28, 31, 32, 30, 29]. These formulations are impractical for high dimensional LPs, such as the access model which employs millions of

variables, suggesting an empirical approach [98], such as the one developed in this paper.

Overall, we find that Colorado meets access standards for the privately insured population in urban areas; however, many suburban and rural communities fail to meet access standards, particularly for publicly insured populations. Although suburban and rural communities experience more issues access dental care, disparities in access between the publicly and privately insured populations are most acute in the urban centers. In particular, the Medicaid population experiences more disparities than those insured through CHP+, as the supply of dental care visits provided by providers accepting CHP+ is greater in Colorado than the supply provided by those accepting Medicaid.

Limitations of this study center around the assumptions made to estimate access and the limited availability of detailed data. In particular, we made assumptions about the total capacity of providers in order to estimate the sampling distribution hyperparamters from MAX data, although these assumptions were informed by HPI data.[99] Details are given in Web Appendix C.1.5. Furthermore, we made several assumptions using the notion that CHP+ providers provided caseloads similar to Medicaid providers.

The methodology in this study can be used to help decision makers evaluate the potential of policies to improve health care access, providing a rigorous framework for the study of access in a given state or other geographic entity while accounting for factors such as geography, the need for care, and the supply of care. Furthermore, this framework can be extended to evaluate how changes in the supply of care, such as raising the acceptance rate of public insurance or open new provider locations, effect access. For dental care in particular, increasing access to preventative dental care services could be cost saving. [100, 101]

APPENDIX A

APPENDIX TO CHAPTER 2

A.1 Alternative Multivariate Distributions for Proposition 2.3.1

Applying the definition of the PCA tolerance region, the proof proposition 2.3.1 shows that$P(\mu \in V(\tau)) = P(\|\Lambda^{-1/2}D^T(R-\mu)\|_\infty \leq \tau)$ where μ is the expectation of R and $\Sigma = D\Lambda D^T$ is the covariance matrix. Thus for any distribution on R, determining the confidence level of the tolerance region requires determining the distribution of $\Lambda^{-1/2}D^T(R-\mu)$. This transformation of R "standardizes" the random variable, clearly having expectation 0 and an identity covariance matrix.

For example, suppose R follows a multivariate t-distribution with v degrees of freedom, i.e. $R \sim t_v(\mu, \Sigma)$ with density function given by

$$f_R(r; v, \mu, \Sigma) = \frac{\Gamma((v+m+n)/2)}{\Gamma(v/2)(v\pi)^{(m+n)/2}\,|\Sigma|^{1/2}}[1 + v^{-1}(r-\mu)^T\Sigma^{-1}(r-\mu)]$$

It is easy to show that then that $\Lambda^{-1/2}D^T(R-\mu) \sim t_v(0, I)$, where I is the identify matrix. Thus, we have

$$P(\mu \in V(\tau)) = P(\|\Lambda^{-1/2}D^T(R-\mu)\|_\infty \leq \tau) = P(|t_v| \leq \tau)^h$$

where t_v follows a univariate t-distribution with v degrees of freedom.

Similar results are possible for other symmetric multivariate distributions. The symmetric property allows for ease in the calculation of the confidence since the tolerance region is symmetric as well. It is possible to derive results for non-symmetric distributions, although some of the simplifications are lost.

A.2 Theoretical Results: Critical Regions in c

A.2.1 Geometry of Critical Regions in c

In this section, we assume that A and b are fixed.

Definitions Two bases B_1 and B_2 are said to be *primal neighbors* if and only if it is possible to pass from B_1 to B_2 by a primal simplex pivot and vice versa. Two critical regions in c are *primal neighbors* if their bases are primal neighbors. The definition of geometric neighbor remains the same: two critical regions in c are *geometric neighbors* if they share a facet.

The fact that c is the RHS of the dual problem (D^{b,c}) suggests that all of the results given in [37] may be extended to the case where c varies while A, b are fixed. However, directly applying the proofs of the results requires consideration of optimal bases of the dual problem written in standard form. It is well known that if B is an optimal basis to (P^{b,c}), then the following solution is optimal to (D^{b,c})

$$y^*(b,c) = A_B^{-1} c_B, \tag{A.1}$$

$$s_B^*(b,c) = 0,\ s_N^*(b,c) = c_N^T - c_B^T A_B^{-1} A_N. \tag{A.2}$$

In (D^{b,c}), a basis for the dual problem is of size n. For an optimal basis B for (P^{b,c}), the m elements of y and the $n-m$ elements of s_N form an optimal basis for the dual. In this way, optimal bases of the dual problem are fully characterized by optimal bases of the primal problem. The basis for the primal problem is determined by the basic indices B, while the corresponding basis for the dual problem is determined by the nonbasic indices N.

The dual basis related to a primal basis B yields a feasible solution to the dual problem if $([I_N\ A^T]^{-1})_{(k,\cdot)} c \geq 0$ for $k = 1, \ldots, n-m$. One can easily show by matrix algebra that the first $n-m$ rows of $[I_N\ A^T]^{-1}$ are precisely the matrix $(I_N - I_B A_B^{-1} A_N)^T$; thus, for a primal basis satisfying the optimality condition (2.2), the corresponding dual basis is dual

feasible. Also, a critical region in c can be written as below:

$$Rc^B = \{c \in \mathbb{R}^n \mid (I_N - I_B A_B^{-1} A_N)^T c \geq 0\}; \tag{A.3}$$

$$= \{c \in \mathbb{R}^n : ([I_N \; A^T]^{-1})_{(k,\cdot)} c \geq 0 \text{ for } k = 1, \ldots, n-m\}. \tag{A.4}$$

That is, Rc^B is the set of linear combinations of the columns of I_N and A^T with nonnegative coefficients for the columns of I_N. This interpretation is useful in proving geometric properties of neighboring critical regions in c, to which we move on now.

Remark: We note that the simplex pivot operations described in the definitions of primal and dual neighbors are different from the pivot operations of the primal and dual simplex algorithm (e.g., see [8]). In an iteration of the primal simplex algorithm, a nonbasic variable with negative reduced cost is selected to enter the basis. In the definition of primal neighbors, however, any nonbasic variable with a positive entry in its corresponding column may be chosen to enter the basis. This difference is due to the fact that the primal simplex method aims to solve a single LP for a specific c value. On the other hand, our goal for performing a primal pivot is to find a basis that is optimal for a new set of c vectors.

Theorem A.2.1 *If two critical regions in c are primal neighbors, then they are geometric neighbors in c.*

Proof: Let B_1 and B_2 be primal neighbors and Rc^{B_1} and Rc^{B_2} are their corresponding critical regions in c. Let $i \in B_1 \setminus B_2$ and $j \in B_2 \setminus B_1$ be the leaving and entering variables in the primal pivot operation from B_1 to B_2, respectively. We denote the pivoting element by $\bar{a}_{ij} = (A_{B_1}^{-1} A)_{(B_1(i),j)}$.

The ability to perform a primal pivot operation implies the pivoting element $\bar{a}_{ij} > 0$. Let H be the hyperplane defined by $(I_{N_1}^T - A_{N_1}^T A_{B_1}^{-T} I_{B_1}^T)_{(N_1(j),\cdot)} c = 0$. Furthermore, let F_1 and F_2 be the $N_1(j)$th (equivalently the $N_2(i)$th) facets of Rc^{B_1} and Rc^{B_2}, respectively. Then, $F_1 \subseteq H$ by definition. We will prove the theorem by first showing $F_2 \subset H$ and then that F_1 and F_2 coincide.

By matrix algebra, we can derive the following equality (detailed steps are in A.3)

$$-\bar{a}_{ij}(I^T_{N_2} - A^T_{N_2}A^{-T}_{B_2}I^T_{B_2})_{(N_2(i),k)} = (I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),k)}, \tag{A.5}$$

for any k. Therefore, $F_2 \subseteq H$ and since $\bar{a}_{ij} > 0$, Rc^{B_1} and Rc^{B_2} lie on opposite sides of H. In addition, we have for $h \in \{1, 2\}$,

$$F_h = Rc^{B_h} \cap H = Rc^{B_h} \cap \{c \in \mathbb{R}^n : ([I_{N_h}\ A^T]^{-1}c)_{(N_1(j),\cdot)} = 0\}$$

$$= \{[I_{N_h}\ A^T]z : z \in \mathbb{R}^n, z_k \geq 0 \text{ for } k \in \{1, \ldots, n-m\} \setminus \{N_1(j)\}, z_{N_1(j)} = 0\}$$

Since N_1 and N_2 only differ in their $N_1(j)$th index, we have $F_1 = F_2$. □

We note here that if $\bar{a}_{ij} < 0$, as is the case in a dual pivot, then Rc^{B_1} and Rc^{B_2} share a facet but lie on the same side of the hyperplane H. This will be an important observation in Section 2.4.4. We now present the converse of Theorem A.2.1.

Theorem A.2.2 *If two critical regions in c are geometric neighbors, then they are primal neighbors.*

Proof: Let B_1 and B_2 be bases, and Rc^{B_1} and Rc^{B_2} denote their corresponding critical regions in c. Assume that they are geometric neighbors. Let $F = Rc^{B_1} \cap Rc^{B_2}$ be the shared facet and say that F is the $N_1(j)$th facet of Rc^{B_1} and the $N_2(i)$th facet of Rc^{B_2}. That is,

$$F = \{[I_{N_1}\ A^T]z : z \in \mathbb{R}^n, z_k \geq 0 \text{ for } k \in \{1, \ldots, n-m\} \setminus \{N_1(j)\}, z_{N_1(j)} = 0\}$$

$$= \{[I_{N_2}\ A^T]z : z \in \mathbb{R}^n, z_k \geq 0 \text{ for } k \in \{1, \ldots, n-m\} \setminus \{N_2(i)\}, z_{N_2(i)} = 0\}$$

Using this equality and the fact that all columns of I_{N_1} and I_{N_2} are extreme rays of F,

we have that $N_1 \setminus \{j\} = N_2 \setminus \{i\}$. That is, N_1 and N_2 differ only by one variable.

Now, let H be the hyperplane containing F. Since Rc^{B_1} and Rc^{B_2} lie on opposing sides of H, we have $(I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),\cdot)}c \leq 0$ for all $c \in Rc^{B_2}$. Since interiors of the two critical regions in c do not overlap, there exists $\hat{c} \in Rc^{B_2}$ such that $(I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),\cdot)}\hat{c} < 0$.

Since $\hat{c} \in Rc^{B_2}$, (A.4) implies that there exists $\hat{z} \in \mathbb{R}^n$ such that $\hat{c} = [I_{N_2}\ A^T]\hat{z}$ where $\hat{z}_{(k,\cdot)} \geq 0$ for $k \in \{1, \ldots, n-m\}$. By lemma A.3.1, for $k \in N_1 \cap N_2$, we have $(I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),\cdot)}I_{(\cdot,k)} = (I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),k)} = 0$. Also note that $(I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})A^T = 0_{(n-m)\times m}$. Therefore, it must be that $(I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),\cdot)}\hat{c} = (I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),\cdot)}[I_{N_2}\ A^T]\hat{z} = (I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),\cdot)}I_{(\cdot,i)}\hat{z}_{(N_1(j),\cdot)} < 0$. Since $\hat{z}_{(N_1(j),\cdot)} > 0$, we have $(I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),\cdot)}I_{(\cdot,i)} = (I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),i)} < 0$.

Again, by lemma A.3.1, $(I^T_{N_1} - A^T_{N_1}A^{-T}_{B_1}I^T_{B_1})_{(N_1(j),i)} = -(A^{-1}_{B_1}A)_{(B_1(i),j)} = -\bar{a}_{ij}$. Therefore, the pivoting element $\bar{a}_{ij}$ is positive and a primal simplex pivot can be performed by removing the ith variable.

It now remains to show that i enters the basis by the min ratio test when j is selected as the exiting variable. We show this by contradiction. That is, assume $k \in B_1 \setminus \{i\}$ enters the basis by the min ratio test. We will show that dual feasibility is not maintained when allowing i to be the entering variable instead of k.

Since k enters the basis by the min ratio test, it follows that

$$\frac{(A^{-1}_{B_1}b)_{(B_1(k),\cdot)}}{(A^{-1}_{B_1}A)_{(B_1(k),j)}} < \frac{(A^{-1}_{B_1}b)_{(B_1(i),\cdot)}}{(A^{-1}_{B_1}A)_{(B_1(i),j)}} \tag{A.6}$$

.

During a primal pivot, dual feasibility must be maintained. Row operations are used to update the tableau, i.e., there exists a matrix E such that $A^{-1}_{B_1} = EA^{-1}_{B_2}$. Specifically, in a

pivot from B_1 to B_2, we have that

$$(A_{B_2}^{-1}b)_{(B_1(k),\cdot)} = (A_{B_1}^{-1}b)_{(B_1(k),\cdot)} - \frac{(A_{B_1}^{-1}A)_{(B_1(k),j)}}{(A_{B_1}^{-1}A)_{(B_1(i),j)}}(A_{B_1}^{-1}b)_{(B_1(i),\cdot)} \tag{A.7}$$

.

Combining (A.6) and (A.7), we see that dual feasibility is not maintained, thus a contradiction has been reached. □

As a consequence of Theorems A.2.1 and A.2.2, we have the following corollary.

Corollary A.2.3 *Two critical regions in c are primal neighbors if and only if they are geometric neighbors in c.*

Finally, we show that if no pivot operation is possible with the jth column as a pivot column, then all c's that lie on the opposite side of the corresponding hyperplane yield an LP which is dual infeasible.

Theorem A.2.4 *If there is no positive entry on the jth column of the simplex tableau at a basis, then there is no c vector that makes the LP dual feasible on the other side of the hyperplane defining the $N(j)$th facet of the critical region in c.*

Proof: Let $\mathcal{K}_c$ be the set of cost vectors that makes the dual feasible. Note that we have

$$\mathcal{K}_c = \left\{ c \in \mathbb{R}^n \mid c = \begin{bmatrix} I & A^T \end{bmatrix} \begin{bmatrix} s \\ y \end{bmatrix}, s \geq 0 \right\},$$

thus, $\mathcal{K}_c$ is the set of linear combinations of the columns of I and A^T with nonnegative weights for the columns of I.

Let B be an optimal basis and N be the corresponding set of non-basic indices. Suppose that for a non-basic index $j \in N$, there is no positive entry on the jth column of the simplex tableau at B, i.e., $\bar{a}_{ij} \leq 0 \; \forall \; i$.

The critical region in c of B is

$$\begin{aligned} Rc^B &= \{c \in R^n \mid (I_N - I_B A_B^{-1} A_N)^T c \geq 0\} \\ &= \{[I_N \; A^T] z \mid z \in \mathbb{R}^n, z_k \geq 0 \text{ for } k = 1, \ldots, n-m\}, \end{aligned}$$

and thus Rc^B is the set of linear combinations of the columns of I_N and A^T with nonnegative weights for the columns of I_N.

The $N(j)$th facet of Rc^B is defined by the inequality $(I_N^T - A_N^T A_B^{-T} I_B^T)_{(N(j),\cdot)} c \geq 0$. By the above definition of Rc^B, all columns of I_N and A^T satisfy this inequality. Using Lemma A.3.1, we can also check whether the columns of I_B satisfy the inequality: for $i \in B$, let $e_i \in R^n$ be the unit vector with one at the ith entry and zeros at the others, then

$$(I_N^T - A_N^T A_B^{-T} I_B^T)_{(N(j),\cdot)} e_i = (I_N^T - A_N^T A_B^{-T} I_B^T)_{(N(j),i)} = -(A_B^{-1} A)_{(B(i),j)} = -\bar{a}_{ij} \geq 0.$$

Therefore, all of the columns of I_B are on the same side of the hyperplane defining the $N(j)$th facet as Rc^B. Thus, the set of all feasible c's, $\mathcal{K}_c$ is also on the same side of the hyperplane, that is, there is no c on the other side of the hyperplane that makes the LP dual feasible. □

A.3 Details of the Proof of Theorem A.2.1

To establish the equality (A.5), we need two lemmas.

Lemma A.3.1 *Let B be a basis with $i \in N$. Then the $N(i)$th row vector of $I_N^T - A_N^T A_B^{-T} I_B^T$ is*

$$(I_N^T - A_N^T A_B^{-T} I_B^T)_{(N(i),k)} = \begin{cases} -(A_B^{-1} A)_{(B(k),i)}, & k \in B \\ 1, & k = i \\ 0, & k \in N \setminus i \end{cases}$$

Proof: Due to basic matrix algebra. □

Lemma A.3.2 *Suppose B_1 and B_2 are primal neighbors. Further suppose the primal pivot step exchanges the variable $i \in B_1 \setminus B_2$ for $j \in B_2 \setminus B_1$ and denote the pivoting element $\bar{a}_{ij} = (A_{B_1}^{-1}A)_{(B_1(i),j)}$. Then*

$$(A_{B_2}^{-1}A)_{(B_2(k),i)} = \begin{cases} -(\bar{a}_{ij})^{-1}(A_{B_1}^{-1}A)_{(B_1(k),j)}, & k \in B_1 \cap B_2 \\ (\bar{a}_{ij})^{-1}, & k = j \end{cases}$$

Proof: The tableau is updated by row operations. Specifically, the pivoting row, $B_2(j)$, is updated by dividing by the pivoting element $\bar{a}_{ij}$. Since $(A_{B_2}^{-1}A)_{(B_2(j),i)} = 1$, the result is proven for $k = j$. Each non-pivoting row is updated by adding a multiple of the pivoting row to it. Since $(A_{B_2}^{-1}A)_{(B_2(k),i)} = 0$ for $k \in B_1 \cap B_2$ and the multiple is $-(\bar{a}_{ij})^{-1}(A_{B_1}^{-1}A)_{(B_1(k),j)}$, the proof is completed. □

Lemma A.3.1 describes the rows of $(I_N - I_B A_B^{-1} A_N)^T$ while Lemma A.3.2 describes how the pivot column is updated in a primal pivot.

Applying lemmas A.3.1 and A.3.2, we see that the kth element of the row vector $-\bar{a}_{ij}(I_{N_2}^T - A_{N_2}^T A_{B_2}^{-T} I_{B_2}^T)_{(N_2(i),\cdot)}$ is given by

$$-\bar{a}_{ij}(I_{N_2}^T - A_{N_2}^T A_{B_2}^{-T} I_{B_2}^T)_{(N_2(i),k)} = \begin{cases} \bar{a}_{ij}(A_{B_2}^{-1}A)_{(B_2(k),i)}, & k \in B_2 \\ -\bar{a}_{ij}, & k = i \\ 0, & k \in N_1 \cap N_2 \end{cases}$$

$$= \begin{cases} \bar{a}_{ij}(A_{B_2}^{-1}A)_{(B_2(k),i)}, & k \in B_1 \cap B_2 \\ \bar{a}_{ij}(A_{B_2}^{-1}A)_{(B_2(k),i)}, & k = j \\ -\bar{a}_{ij}, & k = i \\ 0, & k \in N_1 \cap N_2 \end{cases} = \begin{cases} -(A_{B_1}^{-1}A)_{(B_1(k),j)}, & k \in B_1 \cap B_2 \\ 1, & k = j \\ -\bar{a}_{ij}, & k = i \\ 0, & k \in N_1 \cap N_2 \end{cases}$$

$$= \begin{cases} -(A_{B_1}^{-1}A)_{(B_1(k),j)}, & k \in B_1 \\ 1, & k = j \\ 0, & k \in N_1 \cap N_2 \end{cases} = (I_{N_1}^T - A_{N_1}^T A_{B_1}^{-T} I_{B_1}^T)_{(N_1(j),k)},$$

and thus we have the equality. □

A.4 Proof of Theorem 2.4.5

Let B_1 and B_2 be neighbors with $\mathbb{B}^1$ and $\mathbb{B}^2$ defined as in (2.7). Suppose the pivot operation exchanges variable $i \in B_1 \setminus B_2$ for variable $j \in B_2 \setminus B_1$. By lemma 2.4.4, there are exactly two cases.

Case 1: R^{B_1} and R^{B_2} are primal neighbors.

Note $R^{B_1} \cap R^{B_2} = (Rb^{B_1} \cap Rb^{B_2}) \times (Rc^{B_1} \cap Rc^{B_2})$. Since Rb^{B_1}, Rb^{B_2} lie on the same side of a shared facet, their intersection has dimension m. Since Rc^{B_1}, Rc^{B_2} lie on opposite sides of a shared facet, their intersection has dimension $n-1$. Thus, $R^{B_1} \cap R^{B_2}$ has dimension $n+m-1$.

The proof of Theorem A.2.1 shows that for some row $k \in \{m+1, \ldots, n\}$, $\mathbb{B}^1_{(k,\cdot)} = -\bar{a}_{ij}\mathbb{B}^2_{(k,\cdot)}$. Let H be the hyperplane defined by $\{r \in \mathbb{R}^{n+m} : \mathbb{B}^1_{(k,\cdot)} r = 0\} = \mathbb{R}^m \times H_c$ where $H_c = \{c \in \mathbb{R}^n : (I_{N_1}^T - A_{N_1}^T A_{B_1}^{-T} I_{B_1}^T)_{(k-m,\cdot)} c = 0\}$.

The kth facets of R^{B_1} and R^{B_2} can thus be written, respectively, $F_1 = R^{B_1} \cap H =$

$Rb^{B_1} \times (Rc^{B_1} \cap H_c)$ and $F_2 = R^{B_2} \cap H = Rb^{B_2} \times (Rc^{B_2} \cap H_c)$.

Furthermore, Theorem A.2.1 shows that $Rc^{B_1} \cap Rc^{B_2} = Rc^{B_1} \cap H_c = Rc^{B_2} \cap H_c$, so that $R^{B_1} \cap R^{B_2} \in F_1$ and $R^{B_1} \cap R^{B_2} \in F_2$.

Case 2: R^{B_1} and R^{B_2} are dual neighbors.

Case 2 is analogous to Case 1. We have $R^{B_1} \cap R^{B_2} = (Rb^{B_1} \cap Rb^{B_2}) \times (Rc^{B_1} \cap Rc^{B_2})$. Since Rc^{B_1}, Rc^{B_2} lie on the same side of a shared facet, their intersection has dimension n. Since Rb^{B_1}, Rb^{B_2} lie on opposite sides of a shared facet, their intersection has dimension $m-1$. Thus, $R^{B_1} \cap R^{B_2}$ has dimension $n+m-1$.

The proof of Theorem 2 of [37] shows that for some row $k \in \{1, \ldots, m\}$, $\mathbb{B}^1_{(k,\cdot)} = \bar{a}_{ij}\mathbb{B}^2_{(k,\cdot)}$. Let H be the hyperplane defined by $\{r \in \mathbb{R}^{n+m} : \mathbb{B}^1_{(k,\cdot)} r = 0\} = H_b \times \mathbb{R}^n$ where $H_b = \{b \in \mathbb{R}^m : (A^{-1}_{B_1})_{(k,\cdot)} b = 0\}$.

The kth facets of R^{B_1} and R^{B_2} can thus be written, respectively, $F_1 = R^{B_1} \cap H = (Rb^{B_1} \cap H_b) \times Rc^{B_1}$ and $F_2 = R^{B_2} \cap H = (Rb^{B_2} \cap H_b) \times Rc^{B_2}$.

Furthermore, Theorem 2 of [37] shows that $Rb^{B_1} \cap Rb^{B_2} = Rb^{B_1} \cap H_b = Rb^{B_2} \cap H_b$, so that $R^{B_1} \cap R^{B_2} \in F_1$ and $R^{B_1} \cap R^{B_2} \in F_2$. □

A.5 Proof of Theorem 2.4.8

Suppose that R^{B_1} and R^{B_2} partially share a facet, i.e., $R^{B_1} \cap R^{B_2}$ is of dimension $n+m-1$ and contained in both the ith facet of R^{B_1} and the jth facet of R^{B_2}. Since $R^{B_1} \cap R^{B_2}$ has the same dimension as the facets of R^{B_1} and R^{B_2}, there is a hyperplane H which contains the ith facet of R^{B_1}, the jth facet of R^{B_2}, and $R^{B_1} \cap R^{B_2}$.

We will first show that either $\{i, j\} \subseteq \{1, \ldots, m\}$ or $\{i, j\} \subseteq \{m+1, \ldots, n\}$ should hold. For contradiction and without loss of generality, suppose $i \in \{1, \ldots, m\}$ and $j \in \{m+1, \ldots, n\}$. Then we write H in two ways:

1. $H = \{r \in \mathbb{R}^{m+n} : \mathbb{B}^1_{(i,\cdot)} r = 0\} = \{(b, c) \in \mathbb{R}^{m+n} : (A^{-1}_{B_1})_{(i,\cdot)} b = 0, c \in \mathbb{R}^n\}$

2. $H = \{r \in \mathbb{R}^{m+n} : \mathbb{B}^2_{(j,\cdot)} r = 0\} = \{(b,c) \in \mathbb{R}^{m+n} : (I^T_{N_2} - A^T_{N_2} A^{-T}_{B_2} I^T_{B_2})_{(k,\cdot)} c = 0, b \in \mathbb{R}^m\}$

where $k = j - m$. Note that the second definition places no restrictions on b, while first definition includes the restriction that $(A^{-1}_{B_1})_{(i,\cdot)} b = 0$. This implies that $(A^{-1}_{B_1})_{(i,\cdot)} b = 0$ for all $b \in \mathbb{R}^m$. Thus, $(A^{-1}_{B_1})_{(i,\cdot)}$ is a zero vector and $A^{-1}_{B_1}$ is singular and a contradiction is reached.

Now there are two cases; the first where $\{i,j\} \subseteq \{1,\ldots,m\}$ and the second where $\{i,j\} \subseteq \{m+1,\ldots,n\}$.

Case 1: $\{i,j\} \subseteq \{1,\ldots,m\}$

Since $R^{B_1} \cap R^{B_2} = (Rb^{B_1} \cap Rb^{B_2}) \times (Rc^{B_1} \cap Rc^{B_2})$, and because it is of dimension $n + m - 1$ and contained in facets of R^{B_1} and R^{B_2} which are in the space of b only, we have that $(Rc^{B_1} \cap Rc^{B_2})$ is full dimensional and $(Rb^{B_1} \cap Rb^{B_2})$ is $m - 1$ dimensional.

Furthermore, it is clear from $\{i,j\} \subseteq \{1,\ldots,m\}$ that $(Rb^{B_1} \cap Rb^{B_2})$ is contained in the ith facet of Rb^{B_1} and the jth facet of Rb^{B_2} and that Rb^{B_1} and Rb^{B_2} lie on opposite sides of the hyperplane $H_b = \{b : (A^{-1}_{B_1})_{(i,\cdot)} b = 0\} = \{b : (A^{-1}_{B_2})_{(j,\cdot)} b = 0\}$.

Since $(Rc^{B_1} \cap Rc^{B_2})$ is full-dimensional, it is possible to fix $c \in (Rc^{B_1} \cap Rc^{B_2})$ such that $P_{b,c}$ is not dual degenerate, i.e., $P_{b,c}$ has a unique optimal solution for every $b \in K_b$ where K_b is defined as $K_b = \{b \in \mathbb{R}^m : P_{b,c} \text{ is feasible}\}$. (This is because dual degeneracies occur when c lies simultaneously in the facets of multiple critical regions in c. Since $(Rc^{B_1} \cap Rc^{B_2})$ is full-dimensional, any finite set of facets can be avoided.)

With such a c fixed, $P_{b,c}$ has a unique optimal solution for every $b \in K_b$. Note that for $b \in Rb^{B_1} \cap Rb^{B_2}$, both B_1 and B_2 are optimal bases. Thus, the solutions yielded by B_1 and B_2 are equal, i.e., for all $b \in Rb^{B_1} \cap Rb^{B_2}$,

$$I_{B_1} A^{-1}_{B_1} b = I_{B_2} A^{-1}_{B_2} b \Rightarrow I^T_{B_2} I_{B_1} A^{-1}_{B_1} b = A^{-1}_{B_2} b \tag{A.8}$$

Recall that $Rb^{B_1} \cap Rb^{B_2}$ is $m - 1$ dimensional and contained in the hyperplane defined by

$(A_{B_2}^{-1})_{(j,\cdot)}b = 0$. Thus, we can select a particular $b \in Rb^{B_1} \cap Rb^{B_2}$ such that $(A_{B_2}^{-1})_{(k,\cdot)}b > 0$ for all $k \in \{1, \ldots, m\} \setminus \{j\}$.

Thus, it must be the case that B_1 and B_2 only differ by one basic variable. (Otherwise, $I_{B_2}^T I_{B_1}$ would have multiple zero rows, precluding equality (A.8) for the appropriately selected b.) Since Rb^{B_1} and Rb^{B_2} lie on opposite sides of H_b, it must be that $B_1 \setminus \{B_1^{-1}(i)\} = B_2 \setminus \{B_2^{-1}(j)\}$.

Finally, since the ith facet of Rb^{B_1} and the jth facet of Rb^{B_2} can be written as $\{A_{B_1}z : z \geq 0, z_i = 0\}$ and $\{A_{B_2}z : z \geq 0, z_j = 0\}$, respectively, and these two sets are equal, it is clear that Rb^{B_1} and Rb^{B_2} share a facet and are dual neighbors by Theorem 2.4.2.

Case 2: $\{i, j\} \subseteq \{m+1, \ldots, n\}$

Since $R^{B_1} \cap R^{B_2} = (Rb^{B_1} \cap Rb^{B_2}) \times (Rc^{B_1} \cap Rc^{B_2})$ is of dimension $n + m - 1$ and is contained in facets of R^{B_1} and R^{B_2} pertaining to c only, it must be the case that $(Rb^{B_1} \cap Rb^{B_2})$ is full dimensional, while $(Rc^{B_1} \cap Rc^{B_2})$ is $n - 1$ dimensional.

Furthermore, it is clear from $\{i, j\} \subseteq \{m+1, \ldots, n\}$ that $(Rc^{B_1} \cap Rc^{B_2})$ is contained in the $(i-m)$th facet of Rc^{B_1} and the $(j-m)$th facet of Rc^{B_2} and that Rc^{B_1} and Rc^{B_2} lie on opposite sides of the hyperplane $H_c = \{c : ((I_{N_1} - I_{B_1}A_{B_1}^{-1}A_{N_1})^T)_{(i-m,\cdot)}c = 0\} = \{c : ((I_{N_2} - I_{B_2}A_{B_2}^{-1}A_{N_2})^T)_{(j-m,\cdot)}c = 0\}$.

Since $(Rb^{B_1} \cap Rb^{B_2})$ is full dimensonal, it is possible to fix $b \in (Rb^{B_1} \cap Rb^{B_2})$ such that $P_{b,c}$ is not primal degenerate, i.e., $D_{b,c}$ has a unique optimal solution for every $c \in K_c = \{c \in \mathbb{R}^m : D_{b,c} \text{ is feasible}\}$. (This is because primal degeneracies occur when b lies simultaneously in the facets of multiple critical regions in b. Since $(Rb^{B_1} \cap Rb^{B_2})$ is full dimensional, any finite set of facets can be avoided.)

With such a b fixed, $D_{b,c}$ has a unique optimal solution for every $c \in K_c$. Note that for $c \in Rc^{B_1} \cap Rc^{B_2}$, both B_1 and B_2 are optimal bases. Thus, the dual solutions yielded by B_1 and B_2 are equal, i.e., for all $c \in Rc^{B_1} \cap Rc^{B_2}$. In particular, by equation (A.2) the

optimal solution for the slack variables is given by

$$I_{N_1}(I_{N_1} - I_{B_1}A_{B_1}^{-1}A_{N_1})^T c = I_{N_2}(I_{N_2} - I_{B_2}A_{B_2}^{-1}A_{N_2})^T c$$

$$\Rightarrow I_{N_2}^T I_{N_1}(I_{N_1} - I_{B_1}A_{B_1}^{-1}A_{N_1})^T c = (I_{N_2} - I_{B_2}A_{B_2}^{-1}A_{N_2})^T c \tag{A.9}$$

Recall that $Rc^{B_1} \cap Rc^{B_2}$ is $n-1$ dimensional, with the lone equality constraint being that $((I_{N_2} - I_{B_2}A_{B_2}^{-1}A_{N_2})^T)_{(j-m,\cdot)}c = 0$. Thus, we may select a particular $c \in Rc^{B_1} \cap Rc^{B_2}$ such that $((I_{N_2} - I_{B_2}A_{B_2}^{-1}A_{N_2})^T)_{(j-m,\cdot)}c > 0$ for all $k \in \{m+1, \ldots, m+n\} \setminus \{j-m\}$.

Thus, it must be the case that N_1 and N_2 only differ by one basic variable. (Otherwise, $I_{N_2}^T I_{N_1}$ would have multiple zero rows, precluding equality (A.9) for appropriately selected c.) Since Rc^{B_1} and Rc^{B_2} lie on opposite sides of H_c, it must be that $N_1 \setminus \{N_1^{-1}(i-m)\} = N_2 \setminus \{N_2^{-1}(j-m)\}$.

Finally, since the ith facet of Rc^{B_1} and the jth facet of Rc^{B_2} can be written $\{[I_{N_1}\ A^T]z : z_k \geq 0 \text{ for } k \in \{1, \ldots, n-m\}, z_{N_1(i)} = 0\}$ and $\{[I_{N_2}\ A^T]z : z_k \geq 0 \text{ for } k \in \{1, \ldots, n-m\}, z_{N_2(j)} = 0\}$, respectively, it is clear that Rc^{B_1} and Rc^{B_2} share a facet and are primal neighbors by Theorem A.2.2. □

A.6 Details of VAR Modeling

In Section 2.5.2, we consider an inventory control problem where demand and cost parameters are predicted using a first order vector autoregression (VAR(1)) model. Here we provide the details of the VAR model and the selection of the VAR model coefficients.

Letting $y_t = [D_{t1} \ldots D_{tJ}\ C_{t1} \ldots C_{tJ}]^T$ denote the response vector at time t, the VAR(1)

model can be written

$$y_t = g + My_{t-1} + \epsilon_t \tag{A.10}$$

with $\epsilon_1, \epsilon_2, \ldots, \overset{iid}{\sim} N(0, \Sigma)$ where Σ a diagonal covariance matrix.

The forecast for time $t + \tau$ at time t is given by the expectation of $y_{t+\tau}$ given y_t, i.e.,

$$\hat{y}_{t+\tau|t} = g + M\hat{y}_{t+\tau-1|t} \tag{A.11}$$

where $\hat{y}_{t|t} \equiv y_t$. The variance of the forecast is

$$Var[\hat{y}_{t+\tau|t}] = \sum_{i=1}^{T} A^{\tau-i}\Sigma \tag{A.12}$$

where $A^0 \equiv I$. Furthermore, for any two forecasts where $\tau > \ell$, we have

$$Cov[\hat{y}_{t+\tau|t}, \hat{y}_{t+\ell|t}] = A^{T-\tau}\Sigma \tag{A.13}$$

The joint distribution of all of the forecasts follows a multivariate normal distribution with expectation provided by (A.11) and covariances provided by (A.12) and (A.13).

We select the parameters of the VAR(1) model as follows. We select $g = [30\ 30\ 30\ 3\ 4\ 5]^T$ and

$$\Sigma_{(i,i)} = \begin{cases} 20, & i \in \{1, 2, 3\} \\ 0.005, & i \in \{4, 5, 6\} \end{cases}$$

Finally, the coefficient matrix A is generated randomly by block. Let

$$M = \begin{bmatrix} M_1 & M_2 \\ M_2^T & M_3 \end{bmatrix}$$

Then the elements of M_1 are drawn from a uniform distribution on the interval $(0, 0.1)$, the elements of M_2 are drawn from a uniform distribution on the interval $(0, -0.005)$, and the elements of M_3 are drawn from a uniform distribution on the interval $(0, 0.001)$. These selections are meant to reflect positively correlations among demands and costs, respectively, with negative correlations between demands and costs.

APPENDIX B

APPENDIX TO CHAPTER 3

B.1 Additional Details and Results of the Experiments

B.1.1 Normal Distributions Used for Varying Benchmark Problems

In Section 5, we used the following distributions for the benchmark problems. For 'enlight13', we used the multivariate normal distribution of dimension $m = 169$ with mean $\mu = [200, \ldots, 200]^T$ and covariance matrix Σ, where

$$\Sigma_{i,j} = \begin{cases} 60 & \text{if } i = j = 1, \\ 1.01^i & \text{if } i = j \in \{2, \ldots, 169\}, \\ 0 & \text{if } i \neq j. \end{cases}$$

For 'mik-250-1-100-1', we used the multivariate normal distribution of dimension $m = 401$ with mean as the original b given in the problem and covariance matrix Σ, where

$$\Sigma_{i,j} = \begin{cases} \frac{\mu_i^2}{100} \times 0.98^i & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

For 'roll3000', we used the multivariate normal distribution of dimension $m = 3,460$ with mean as given in the problem and covariance matrix Σ, where

$$\Sigma_{i,j} = \begin{cases} \frac{\mu_i^2}{100{,}000} \times 0.98^i & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

B.1.2 Additional Results

In addition to the experimental results in the body of the paper, we also obtained the following results. Tables B.1, B.2, and B.3 compare DD and WSBS for varying Ks and for 'enlight13', 'mik-250-1-100-1', and 'roll3000', respectively.

Table B.1: Comparison of DD and WSBS for 'enlight13' and varying Ks

K	CRs	# Pivots		# Membership checks		CPU time	
		DD	WSBS	DD	WSBS	DD	WSBS
100	10	267	275	121	179	0.01	0.01
500	24	299	314	692	1,322	0.01	0.01
1,000	32	326	340	1,651	2,825	0.01	0.01
5,000	66	463	491	9.393	17,364	0.04	0.03
10,000	84	531	550	18,775	29,707	0.07	0.04
50,000	125	752	779	98,736	154,801	0.26	0.19
100,000	148	908	941	203,221	342,671	0.48	0.37
500,000	201	1,102	1,169	1,050,262	1,745,294	2.4	2.1
1,000,000	230	1,366	1,396	2,079,501	3,538,638	5.25	4.58

Table B.2: Comparison of DD and WSBS for 'mik-250-1-100-1' and varying Ks

K	CRs	# Pivots		# Membership checks		CPU time	
		DD	WSBS	DD	WSBS	DD	WSBS
100	47	253	255	1,078	1,701	0.02	0.01
500	140	633	664	14,369	18,388	0.05	0.04
1,000	220	1,018	1,054	38,581	48,826	0.1	0.08
5,000	450	2,213	2,230	259,868	310,914	0.55	0.45
10,000	661	3,307	3,343	665,802	780,876	1.3	1.19
50,000	1,288	6,894	6,964	4,000,122	4,525,461	6.71	6.4
100,000	1,682	9.284	9,346	8,122,643	9,312,620	14.56	13.86

Table B.3: Comparison of DD and WSBS for 'roll3000' and varying Ks

K	CRs	# Pivots		# Membership checks		CPU time	
		DD	WSBS	DD	WSBS	DD	WSBS
100	2	761	613	139	139	1.98	0.05
500	2	719	607	1,690	738	1.78	0.05
1,000	2	804	602	4,387	1,483	2.68	0.06
5,000	3	785	618	87,149	11,515	3.98	0.17
10,000	3	779	607	167,073	23,058	5.89	0.26
50,000	3	803	609	898,900	114,795	19.63	1.03

B.2 Sampling Parameters of Access Problem

B.2.1 Sampling Capacity Parameters

Each of the providers may have multiple physicians and the number of physicians at each provider is given in [102]. Given a provider, we first sample the number of visits dedicated to children for each physician located at the provider, and then the provider's capacity is estimated to be the sum of the capacities of physicians.

The sampling procedure for the capacity of an individual physician is as follows:

1. Randomly assign the physician a gender according to a Bernoulli($p = 0.3409$) distribution, where p is the probability of being female, obtained from [103].

2. Using the gender sampled in step 1, randomly assign the physician an age. For female physicians, assign age according to normal distribution $\mathcal{N}(\mu = 48.30, \sigma = 10.11)$. For male physicians, assign age according to $\mathcal{N}(\mu = 50.83, \sigma = 12.70)$. The distributions were obtained from [103].

3. Using the (gender, age) pair of the physician, assign the physician's number of yearly visits for children to be 10% of the corresponding value in Table B.4, which was estimated from [104].

Table B.4: Total number of yearly visits per physician

Age	Males	Females
< 40	8736	6888
40-44	8988	7224
45-49	9240	7644
50-54	9912	8064
55-59	9576	7896
60-64	8904	6216
> 65	7056	7560

B.2.2 Sampling Demand Parameters

The total number of yearly visits required by a census tract was assumed to follow a normal distribution with mean and standard deviation parameters given in Tables B.5 and B.6.

Table B.5: Normal distribution parameters for demand constraints

Census Tract	Mean	Std. Deviation
1.00	37.993	6.6530
3.00	7.487	1.4160
6.01	31.038	5.8110
9.00	0.981	0.1935
11.00	53.626	9.9630
12.00	56.903	10.2665
15.00	5.027	0.8690
20.00	24.108	4.4470
21.00	37.456	6.9650
22.00	105.437	19.4060
23.00	14.732	2.7625
26.00	16.957	3.0020
27.00	42.006	7.8340
28.00	57.792	10.6545
29.00	42.365	7.6550
30.00	9.097	1.6390
33.01	34.136	6.1710
33.02	16.684	3.0275
34.00	41.118	7.5110
35.01	32.113	5.8980
35.02	85.248	16.0035
36.01	47.322	8.5355
36.02	78.764	14.6810
37.00	38.253	7.0470
38.00	47.854	9.0385
39.00	82.340	14.7030
40.01	78.325	14.2415
40.02	26.033	4.8240
41.00	25.739	4.6455
42.07	95.632	17.5040
42.08	80.589	14.3165
42.09	60.987	11.0010
42.10	57.321	10.3630
42.11	71.921	13.4515
42.12	70.250	13.1485
43.00	38.209	6.7765
44.00	61.326	11.3255
45.00	90.873	17.1070
101.01	42.127	7.4860
101.02	22.102	4.0240
102.00	86.945	15.8485
105.01	66.103	12.1135
105.02	113.213	20.6365
106.01	113.243	20.8780
106.03	11.853	2.0670
107.00	200.413	35.8240
108.01	47.592	8.7055
108.02	105.433	18.9695
108.03	208.862	37.8920
108.06	76.487	14.0815
108.07	111.955	20.1585
108.08	134.630	24.4145
108.09	134.496	24.6205
109.01	66.646	12.0295

Table B.6: Normal distribution parameters for demand constraints

Census Tract	Mean	Std. Deviation
110.03	74.908	13.3040
110.04	62.120	11.3940
110.05	46.706	8.2550
110.06	32.368	5.9865
111.03	19.219	3.5655
111.04	81.594	15.1695
111.06	96.672	17.4375
111.07	43.329	8.0905
111.08	68.767	12.7260
111.09	113.250	20.8245
112.00	14.887	2.6180
113.00	36.707	6.6200
114.00	11.181	2.0925
115.00	44.733	8.1535
116.00	51.668	9.6175

APPENDIX C

APPENDIX TO CHAPTER 4

C.1 Supplementary Material

C.1.1 Optimization Modeling Input Data and Notation

Notations

Table C.1: Notations

Notation	Description	Value(s)	Data
$\mathbb{C}$	The number of census tracts in Colorado	1249	2010 SF2 100% census data (for centroid location)
$\mathbb{J}$	The number of provider locations in Colorado	3051	2013 NPI data (for provider location)

Fixed Input Data

Table C.2: Fixed input data

Input Data	Where	Description	Value	Data
ρ	(C.1)	Regularization Parameter	0.55	
d_{ij}	(C.1), (C.9)-(C.11), (C.13)	Distance from centroid of census tract i to provider location j	–	Assumed constant for each year, ArcGIS Network Analyst, 2013 NPI data (for provider location), 2010 SF2 100% Census data (for centroid location)
κ	(C.2)	Minimum percentage of the population to be covered	0.7	
MAX	(C.9)-(C.11), (C.13)	Maximum distance allowed between patient and provider	60	Guidelines established by the U.S. Dept. of Health and Human Services
MAX_{pen}	(C.14)	Maximum distance (miles) between census tracts to penalize for differences in access measure	10	
f	(C.1), (C.6)-(C.8)	Avg. # visit for census tract i	1.5	
ϵ_{ik}	(C.14)	Distance from centroid of census tract i to the centroid of census tract k	–	Assumed constant for each year, ArcGIS Network Analyst, 2010 SF2 100% Census data (for centroid location)

Sampled Input Data

Table C.3: Sampled input data

Input Data	Where	Description	Data
Pop	(C.2)	Total Population of Colorado	Table B17024 of the ACS
p_i	-	Population of census tract i	Table B17024 of the ACS
n_i^M	(C.3)	Medicaid insured population of census tract i	Table B17024 of the ACS
n_i^H	(C.4)	CHP+ insured population of census tract i	Table B17024 of the ACS
n_i^O	(C.5)	Privately insured population of census tract i	Table B17024 of the ACS
n_i^U	(C.2)	Population of census tract i lacking financial access	Table B17024 of the ACS
c_j^M	(C.6)	Yearly preventive dental care capacity at location j devoted to Medicaid insured children	MAX data
c_j^H	(C.7)	Yearly preventive dental care capacity at location j devoted to CHP+ insured children	
c_j	(C.8)	Yearly preventive dental care capacity at location j devoted to children	

Optimization Modeling of Access

The matching parameters to be estimated are θ_{ij}^M, θ_{ij}^H and θ_{ij}^O, the number of children from census tract i, matched to provider j, distinguished by financial access to the dental care

system through insurance types Medicaid (M), CHP+ (H) or other financial access (O). Because children with different types of insurance have access to different networks of providers, our model matches children in the three insurance type groups separately. We note that we do not include children with no financial access in this optimization modeling since financial access is a precursor of spatial access, thus we assume that children with no financial access will not have spatial access.

The location of each census tract $i = 1, \ldots, \mathbb{C}$ is taken to be the centroid of each tract, while the location of each provider is taken to be the address of the individual provider, geocoded using the Texas A&M Geocoding Services.[105] Using the Environmental Systems Research Institute's (Esri) ArcGIS, we measure the distance d_{ij} between each census tract centroid and provider $j = 1, \ldots, \mathbb{J}$. The linear component of the objective function minimizes the total distance traveled under the matching. In formulating the model, we follow the regularized approach of [90] in order to control for the sensitivity of the matchings to perturbations in the input data and to insure smoothness in the spatial domain. The

model is formulated in equations (C.1)-(C.14).

$$\min_{\theta,z,y,w} (1-\rho)\sum_i \sum_j d_{ij} f(\theta_{ij}^M + \theta_{ij}^H + \theta_{ij}^O) + \rho S(z,y,w) \tag{C.1}$$

$$\sum_i \sum_j (\theta_{ij}^M + \theta_{ij}^H + \theta_{ij}^O) \geq \kappa(Pop - \sum_i n_i^U) \tag{C.2}$$

$$\sum_j \theta_{ij}^M \leq n_i^M \; \forall i \in \{1,\ldots,\mathbb{C}\}, \tag{C.3}$$

$$\sum_j \theta_{ij}^H \leq n_i^H \; \forall i \in \{1,\ldots,\mathbb{C}\}, \tag{C.4}$$

$$\sum_j \theta_{ij}^O \leq n_i^O \; \forall i \in \{1,\ldots,\mathbb{C}\}, \tag{C.5}$$

$$\sum_i f\theta_{ij}^M \leq c_j^M \; \forall j \in \{1,\ldots,\mathbb{J}\}, \tag{C.6}$$

$$\sum_i f\theta_{ij}^H \leq c_j^H \; \forall j \in \{1,\ldots,\mathbb{J}\}, \tag{C.7}$$

$$\sum_i f(\theta_{ij}^M + \theta_{ij}^H + \theta_{ij}^O) \leq c_j \; \forall j \in \{1,\ldots,\mathbb{J}\}, \tag{C.8}$$

$$z_i = MAX + \frac{1}{n_i^M}\sum_j (d_{ij} - MAX)\theta_{ij}^M \; \forall i \in \{1,\ldots,\mathbb{C}\}, \tag{C.9}$$

$$y_i = MAX + \frac{1}{n_i^H}\sum_j (d_{ij} - MAX)\theta_{ij}^H \; \forall i \in \{1,\ldots,\mathbb{C}\}, \tag{C.10}$$

$$w_i = MAX + \frac{1}{n_i^O}\sum_j (d_{ij} - MAX)\theta_{ij}^O \; \forall i \in \{1,\ldots,\mathbb{C}\}, \tag{C.11}$$

$$\theta_{ij}^M, \theta_{ij}^H, \theta_{ij}^O \geq 0 \; \forall i \in \{1,\ldots,\mathbb{C}\}, j \in \{1,\ldots,\mathbb{J}\} \tag{C.12}$$

$$\theta_{ij}^M = \theta_{ij}^H = \theta_{ij}^O = 0 \; \forall i,j \text{ s.t. } d_{ij} > MAX \tag{C.13}$$

where

$$S(z,y,w) = \sum_i \sum_{k:\epsilon_{ik} \leq MAX_{pen}} \frac{1}{\epsilon_{ik}}(z_i - z_k)^2 + \sum_i \sum_{k:\epsilon_{ik} \leq MAX_{pen}} \frac{1}{\epsilon_{ik}}(y_i - y_k)^2 + \sum_i \sum_{k:\epsilon_{ik} \leq MAX_{pen}} \frac{1}{\epsilon_{ik}}(w_i - w_k)^2 \tag{C.14}$$

The objective function (C.1) has two components, over which we minimize a weighted sum. The first component requires the minimization of the sum of the total traveled distance in the system. The second component, (C.14), plays the role of smoothing the access measures. The trade off between optimality and smoothness is controlled by the regularization parameter ρ. This trade off is akin to the bias-variance trade off often referred to in the statistics literature.

Without the regularization component given by (C.14), the optimization problem is a linear program. Because of the large number of variables, the optimal solution and access measures obtained using the unregularized model can be overly sensitive to small perturbations of the input parameters.[90] Motivated by the intuition that communities with geographic proximity should experience similar levels of access, we take the regularized objective function of (C.1), which is a quadratic program. By deploying the regularized formulation, we reduce sensitivity and provide a measure of access that is more reliable for decision makers.

Constraints (C.3) - (C.5) ensure that no more demand is assigned to providers than exists from each census tract, age group pair. Constraint (C.2) provides a lower limit indicating the total number of patients to which must be assigned in the system. We choose κ, the percentage of children which should be assigned overall in the system, to be the largest value for which the problem is consistently feasible as the sampled parameters are repeatedly generated. Constraints (C.6) - (C.8) take providers' capacities into account, for the Medicaid, CHP+, and general populations, respectively. Constraints (C.9), (C.10), and (C.11) define our access measures, z_i, y_i and w_i, for the Medicaid, CHP+, and private fi-

nancial access populations, respectively, for each census tract. Finally, constraint (C.12) is the standard non-negativity constraint, while (C.13) ensures that patients are not assigned to a provider requiring them to travel more than MAX miles.

C.1.2 Measuring Accessibility

After solving the optimization problem assigning patients to providers, we obtain matchings $\hat{\theta}_{ij}^{M}$, $\hat{\theta}_{ij}^{H}$, and $\hat{\theta}_{ij}^{O}$. Based on this matching, the dental care access of each census tract is defined as the average distance that children travel to reach a dental care providers. We define the access measures for children insured through Medicaid, insured through CHP+, and with other financial access for each census tract i in equations (C.15), (C.16), and (C.17), respectively..

$$\hat{z}_i = MAX(1 - \frac{\sum_j \hat{\theta}_{ij}^{M}}{n_i^{M}}) + \frac{\sum_j d_{ij}\hat{\theta}_{ij}^{M}}{n_i^{M}} \tag{C.15}$$

$$\hat{y}_i = MAX(1 - \frac{\sum_j \hat{\theta}_{ij}^{H}}{n_i^{H}}) + \frac{\sum_j d_{ij}\hat{\theta}_{ij}^{H}}{n_i^{H}} \tag{C.16}$$

$$\hat{w}_i = MAX(1 - \frac{\sum_j \hat{\theta}_{ij}^{O}}{n_i^{O}}) + \frac{\sum_j d_{ij}\hat{\theta}_{ij}^{O}}{n_i^{O}} \tag{C.17}$$

Note that equations (C.15), (C.16), and (C.17) are equivalent to constraints (C.9), (C.14), and (C.11), reformulated for clarity. The measures are the weighted average of the travel distance of children in each census tract $i = 1, \ldots, \mathbb{C}$. Children who are not assigned to a provider are assigned a travel distance equal to MAX. Therefore, the access measures range between 0 and MAX. We use a maximum distance of 60 miles to avoid subjecting the access measures to large distance values of those without access.

In order to study access to dental care services of the publicly insured population, we compute the average distance traveled for those insured through Medicaid and CHP+. Equation (C.18), shown below, defines the public insurance access measures.

$$\hat{u}_i = \frac{n_i^M \hat{z}_i + n_i^H \hat{y}_i}{n_i^M + n_i^H} \tag{C.18}$$

C.1.3 Distributions for Sampled Input Data

In order to account for uncertainty in many of the input data of the optimization model, a sampling approach is used. Table C.4 provides a brief overview of the distributions placed on the input data and the methods/sources of data used to construct the distributions.

Table C.4: Sampled Input Parameters Descriptions

Input Data	Distribution	Description
POP	Negative Binomial Distribution	Total number of children in Colorado
N_i^M	Multinomial Distribution	Total Medicaid-insured population, at census tract i
N_i^H	Multinomial Distribution	Total CHP+-insured population, at census tract i
N_i^O	Multinomial Distribution	Total population with other financial access, at census tract i
N_i^U	Multinomial Distribution	Population without financial access, at census tract i
C_j	Poisson Distribution	Yearly preventive dental care capacity at location j devoted to children
C_j^M	Binomial Distribution	Yearly preventive dental care capacity at location j devoted to publicly insured children

C.1.4 Population Sampling

In this section, we describe the modeling and sampling procedures used to sample the population parameters which, along with the average number of visits demanded per child, describe the need for dental care.

Description of the Population Variables

The population parameters are are segmented by census tract and type of financial access. The total number of children in Colorado, Pop, is segmented by census tract such that $Pop = \sum_{i=1}^{\mathbb{C}} p_i$, where p_i is the population of census tract i. Within each census tract, the population is split into Medicaid, privately insured, and uninsured populations, n_i^M, n_i^H, n_i^O, and n_i^U, respectively, such that $p_i = n_i^M + n_i^O + n_i^H + n_i^U$.

Population Data

We use the 2017 ACS 5 year estimates (table B17024) to obtain data on the total number of children in Colorado, the number of children in each census tract, and the number of children in each census tract with each type of financial access.[106] The ACS expressly provides estimates of the the total number of children and the number of children in each census tract, which we use to estimate the hyperparameters of the population sampling distributions.

To estimate the hyperparameters for dividing the population by financial access, we make some assumptions about financial access based on family income. The census data estimates the number of children living in each census tract by family income level. We assume that children from families with incomes less that 142% of the federal poverty line (FPL) are covered by Health First Colorado (Medicaid); similarly, we assume children from families with incomes between 142% and 260% of the FPL are covered by Children Health Plan Plus (CHP+). [107] Chlidren with no financial access are those from households with incomes too high for public insurance programs but with too low to cover out-of-pocket

expenses. We assume that children with no financial access are those from families with incomes between 260% and 400% of the FPL.[82] Finally, we assume that children from families with incomes greater than 400% of the FPL have private financial access, such as private or employer-based insurance.

The end result are census estimates of 1) the number of children in Colorado, Pop, 2) the number of children in each census tract, p_i, and 3) the number of children with each of the four types of financial access in each census tract, $n_i^M, n_i^H, n_i^U, n_i^O$. We use this data to inform the sampling distributions of the population parameters.

Details of Hyperparameter Estimation

The modeling framework for the population parameters takes a hierarchical Bayes approach. Figure C.3 outlines the modeling framework.

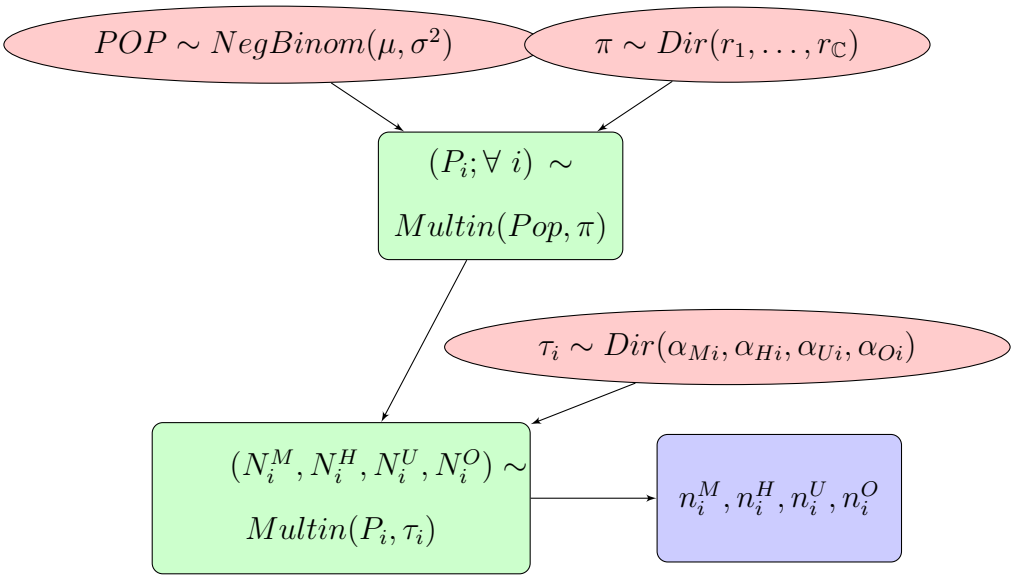


Figure C.1: Hierarchical framework for modeling population parameters

The red nodes represent initial sampling nodes, where the hyperparameters are fixed values, estimated from data. Green nodes represent sampling from distributions whose parameters are themselves samples from a higher level node. The purple node represents the final input data used in the optimization model.

The parameters of the negative binomial distribution, μ and σ^2, used to sample the total population were selected in order to match the estimate and sampling variance provided in the Census Bureau data. Specifically, the ACS estimates the total number of children in Colorado is 1,234,118, with a margin of error of 569, where 1234118 ± 569 represents a 90% confidence interval.[108] Using a normal approximation, this applies that the sampling variance of the population estimate is $(569/1.645)^2 = 119644.5$, where 1.645 is the 95th percentile of the standard normal distribution. The negative binomial distribution, parameterized by the mean and variance, denoted $NegBinom(\mu, \sigma^2)$, is given by

$$p(Pop \mid \mu, \sigma^2) = \binom{Pop + \frac{\mu}{\sigma^2 - \mu} - 1}{Pop} (\frac{\sigma^2 - \mu}{\sigma^2})^{Pop} (\frac{\mu}{\sigma^2})^{\mu/(\sigma^2 - \mu)}$$

The negative binomial parameters are then taken to be $\hat{\mu} = 1234118$ and $\hat{\sigma}^2 = 119644.5$

We then assume that the population in each census tract follows a multinomial distribution with total samples Pop and vector of proportions π. We place a Dirichlet prior on π. That is, we have

$$p(p_1, \ldots, p_{\mathbb{C}} \mid \pi, Pop) = \frac{Pop!}{p_1! \cdots p_{\mathbb{C}!}} \pi_1^{p_1} \ldots \pi_{\mathbb{C}}^{p_{\mathbb{C}}}$$

$$p(\pi \mid r) = \frac{1}{B(r)} \prod_{i=1}^{\mathbb{C}} \pi_i^{r_i}$$

where $B(v) = \frac{\prod_{i=1}^{\mathbb{C}} \Gamma(v_i)}{\Gamma(\sum_{i=1}^{\mathbb{C}} v_i)}$ is the multivariate beta function and r is the vector of hyperparameters $r_1, \ldots, r_{\mathbb{C}}$.

The parameters of the Dirichlet distribution for the proportions of patients in each census tract were estimated using an empirical Bayes approach. That is, we first estimate the hyperparameters of the prior from data, then update the prior distribution using the data again to obtain the posterior distribution. Recall that we obtained census estimates of the total population, Pop, and the number of children in each census tract, $p_1, \ldots, p_{\mathbb{C}}$, denoted collectively by the vector p. We assume that the number of children in each census tract

follows a multinomial distribution and place a Dirichlet prior on the proportions of the total population in each census tract. That is, we place a $p(\pi \mid r_1, \ldots, r_{\mathbb{C}}) = Dir(r_1, \ldots, r_{\mathbb{C}})$ prior on the vector of proportions π. Then, the distribution of the data given the prior hyperparameters is

$$p(p_1, \ldots, p_{\mathbb{C}} \mid r) = \int_{\pi} p(p_1, \ldots, p_{\mathbb{C}} \mid \pi) p(\pi \mid r) d\pi \propto \frac{B(p+r)}{B(r)}$$

Using maximum likelihood estimation, we select the prior hyperparameters which maximize the likelihood conditioned over the hyperparameters r, i.e.

$$r^* = \arg\max_r \frac{B(p+r)}{B(r)}$$

This optimization can be achieved using gradient descent methods. Once the hyperparameters are estimated, we calculate the posterior distribution using Bayes rule, i.e.

$$p(\pi \mid p, r^*) = \frac{p(p \mid \pi) p(\pi \mid r^*)}{\int p(p \mid \pi) p(\pi \mid r^*) d\pi} = Dir(\hat{r}_1, \ldots, \hat{r}_{\mathbb{C}})$$

where $\hat{r} = p + r^*$ is the vector of posterior hyperparameters.

Finally, we assume that the population in each census tract is split into four types of financial access according to another multinomial distribution. In order to estimate the posterior Dirichlet distribution used to sample the proportions of each census tract with the four types of financial access, an empirical Bayes approach is again used. This time, however, we use a regression approach to set the hyperparameters of the priors. This time we assume that the prior distribution takes the form $Dir(\alpha_{iM}, \alpha_{iH}, \alpha_{iU}, \alpha_{iO})$ where $\alpha_{iI} = log(x_i^T \beta_I)$. Here, x_i is a vector of covariates for census tract i and β_I is the financial access type specific vector of regression coefficients, $I \in \{O, M, H, U\}$. The goal is to estimate the regression coefficients which will, in turn, set the prior distributions.

The following variables were considered for covariates in the regression model:

- Poverty: the percentage of the population living below the federal poverty line in

each census tract

- R2: binary variable indicating suburban urbancity, defined as having RUCCs 4, 5, or 6
- R3: binary variable indicating rural urbancity, defined as having RUCCs 7, 8, or 9

Note that urban urbanicity, defined as having RUCCs 1, 2, or 3, is taken to be the baseline state, so variables R2 and R3 control for the differences between suburban and rural, respectively, and urban census tracts. The poverty rate for each census tract was obtained from 2017 ACS 5 year estimates (table S1702).

For each census tract i, we have the conditional likelihood

$$p(n_i^M, n_i^H, n_i^U, n_i^O \mid \beta_M, \beta_H, \beta_U, \beta_O) = \int_\beta p(n_i^M, n_i^H, n_i^U, n_i^O \mid \tau_i) p(\tau_i \mid \beta_M, \beta_H, \beta_U, \beta_O) d\beta$$

$$\propto \frac{\prod_I \Gamma(n_i^I + \exp x_i^T \beta_I)}{\Gamma(P_i + \sum_I \exp x_i^T \beta_I)} \times \frac{\Gamma(\sum_I \exp x_i^T \beta_I)}{\prod_I \Gamma(\sum_I \exp x_i^T \beta_I)}$$

The conditional likelihood across census tract can be obtained taking the product of the individual census tract likelihoods, i.e.

$$p(n_i^M, n_i^H, n_i^U, n_i^O \; \forall i \mid \beta_M, \beta_H, \beta_U, \beta_O) = \prod_{i=1}^{\mathbb{C}} p(n_i^M, n_i^H, n_i^U, n_i^O \mid \beta_M, \beta_H, \beta_U, \beta_O)$$

The 4 regression coefficient vectors are then estimated using maximum likelihood, i.e.

$$\hat{\beta}_M, \hat{\beta}_H, \hat{\beta}_U, \hat{\beta}_O = \underset{\beta_M, \beta_H, \beta_U, \beta_O}{\arg\max} \prod_{i=1}^{\mathbb{C}} \frac{\prod_I \Gamma(n_i^I + \exp x_i^T \beta_I)}{\Gamma(P_i + \sum_I \exp x_i^T \beta_I)} \times \frac{\Gamma(\sum_I \exp x_i^T \beta_I)}{\prod_I \Gamma(\sum_I \exp x_i^T \beta_I)}$$

which can be found using gradient descent methods. The regression parameter estimates are reported in Table C.5.

Table C.5: Regression coefficients for financial access type proportions

	Intercept	Poverty	R2	R3
$\hat{\beta}_O$	1.5992	-0.0450	-0.0343	-0.1155
$\hat{\beta}_M$	-0.1084	0.0624	0.3423	0.3527
$\hat{\beta}_H$	0.4918	0.0229	0.4219	0.3909
$\hat{\beta}_U$	0.6773	0.0054	0.3095	0.3001

Once the regression coefficients are estimated, the prior distributions are set. Posterior calculation proceeds by recognizing that

$$p(\tau_i \mid n_i^M, n_i^H, n_i^U, n_i^O, \hat{\beta}_O, \hat{\beta}_M, \hat{\beta}_H, \hat{\beta}_U, x_i) = Dir(\hat{\alpha}_{Mi}, \hat{\alpha}_{Hi}, \hat{\alpha}_{Oi}, \hat{\alpha}_{Ui})$$

where $\hat{\alpha}_{Ii} = n_i^I + \exp x_i^T \hat{\beta}_I$ for each type of financial access I.

Population Sampling Procedures

Figure C.3 displays the process for sampling the optimization model input parameters. For the bth sample, we first generate the total number of children in Colorado, Pop_b, from a negative binomial distribution with parameters $\hat{\mu}$ and $\hat{\sigma}^2$ and the proportion of the total population belonging to each census tract π_b from a Dirichlet distribution with parameters $\hat{r}_b$. Then, using the values drawn for Pop_b and π_b, we generate the number of children in each census tracts, p_{ib} for $i \in \{1, \ldots, \mathbb{C}\}$, from a multinomial distribution. Finally, for each census tract i, we generate the number of children with each of the four types of financial access, $(n_i^M, n_i^H, n_i^U, n_i^O)_b$, from another multinomial distribution, where the proportions,

τ_{ib}, are drawn from a Dirichlet distribution with parameters $\hat{\alpha}_{Mi}, \hat{\alpha}_{Hi}, \hat{\alpha}_{Ui}, \hat{\alpha}_{Oi}$. The bth sample of population parameters is then $(n_i^M, n_i^H, n_i^U, n_i^O)_b$.

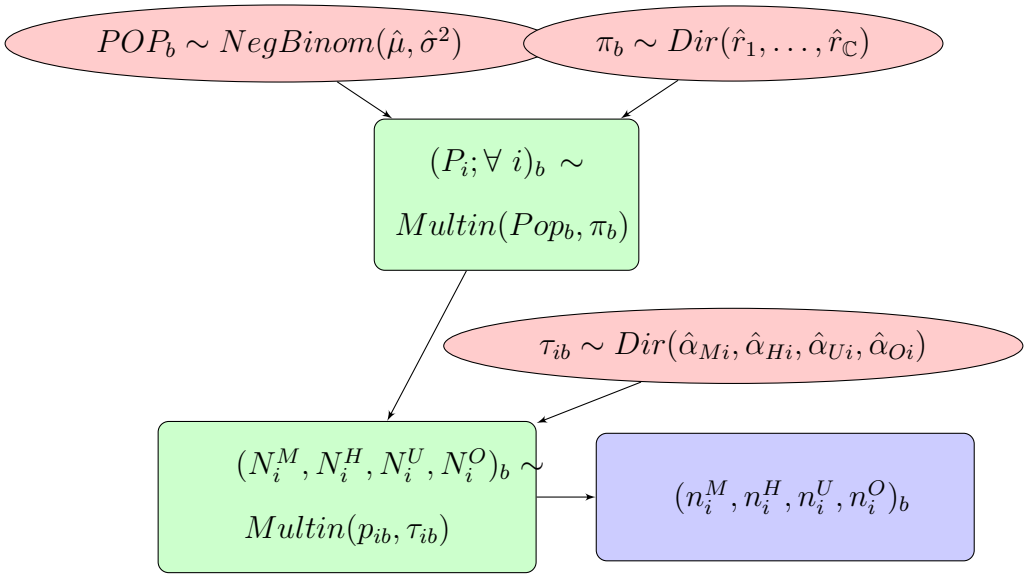


Figure C.2: Hierarchical Bayes sampling of population parameters

C.1.5 Provider Capacities

In this section, we describe the modeling and sampling procedures used to sample the provider capacity parameters which describe the supply of dental care.

Description of Provider Capacity Variables

The capacity parameters are given by the total capacity of visits available from provider j, c_j, the number of visits available to the Medicaid insured population c_j^M, and the number of visits available to the CHP+ insured population c_j^H. By definition, we have that $c_j^M \leq c_j$ and $c_j^H \leq c_j$.

Provider Capacity Data

To estimate the hyperparameters of the sampling distribution, data from the 2012 Medicaid Analytical eXtract (MAX) claims were used. Specifically, we used data detailing the

Medicaid caseloads of providers in Colorado. This data was used previously in examining Medicaid-specific caseload.[84, 92] Although this data is matched to specific provider locations in Colorado, we used this data to inform our sampling distributions, developing caseload profiles for providers by taxonomy.

We used Board of Dentistry (BOD) and National Provider Plan and Enumeration (NPPES) data to consruct a list of the full time equivalent (FTE) number of providers at provider locations across the state of Colorado. For provider locations participating in Medicaid and CHP+, we select the total FTE available to patients of all insurance types to be the maximum of the FTE available to Medicaid and the the FTE available to CHP+.

Details of Hyperparameter Estimation

We estimate the hyperparameters independently for each taxomony t. We classify the providers in the MAX data as either a general dentist or a pediatric dentist. The modeling framework for the capacity parameters for taxonomy t takes a hierarchical Bayes approach. Figure C.3 outlines the modeling framework.

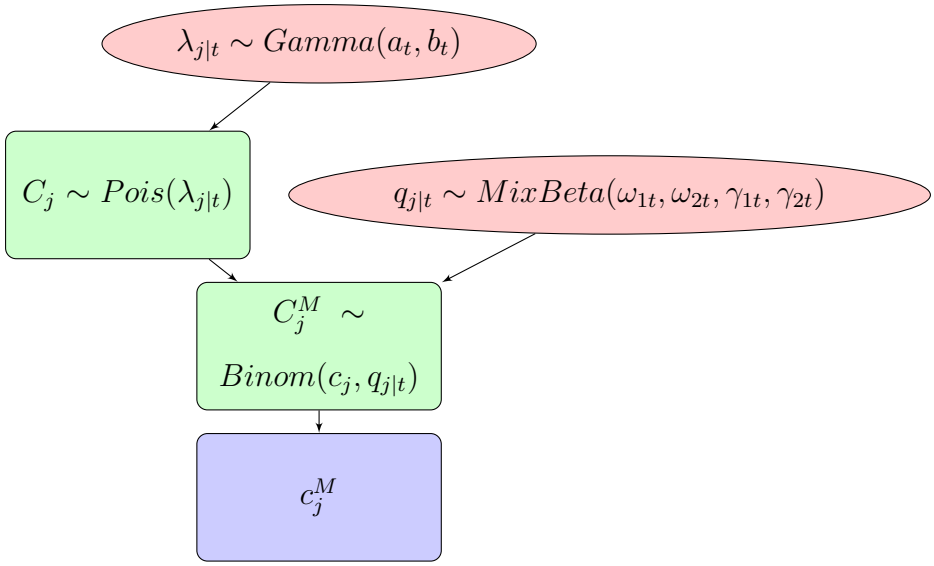


Figure C.3: Hierarchical framework for modeling provider capacity parameters

We assume that the Medicaid caseload data comes from a binomial distribution, where the total caseload, in number of visits, for the jth provider, c_j, is split into Medicaid caseload and non-Medicaid caseload. The expected proportion of the total caseload dedicated to Medicaid is $q_{j|t}$. We assume that the total capacity of provider j is distributed as a Poisson distribution, i.e. $C_j \sim Pois(\lambda_{j|t})$. For estimation of the prior distributions on $\lambda_{j|t}$, we assume that $\lambda_{j|t} \sim Gamma(a_t, b_t)$, where a_t and b_t depend on the provider taxonomy t. We use an empirical Bayes method to inform the selection of hyperparmeters of the priors from Medicaid caseload data. We assume the Medicaid caseload follows a binomial distribution, where the total caseload comes from a Poisson distribution.

$$p(C_j^M \mid c_j, q_j) = \binom{c_j}{C_j^M} q_j^{C_j^M} (1 - q_j)^{c_j - C_j^M}$$

$$p(C_j \mid \lambda_{j|t}) = \frac{\lambda_{j|t}^{C_j}}{C_j!} e^{-\lambda_{j|t}}$$

$$p(\lambda_{j|t} \mid a_t, b_t) = \frac{b_t^{a_t}}{\Gamma(a_t)} \lambda_{j|t}^{a_t - 1} e^{-b_t \lambda_{j|t}}$$

We place a Beta mixture prior on the Medicaid acceptance ratio of provider j, $q_{j|t}$. The mixture distribution accounts for the fact that some of the Medicaid providers will also accept CHP+ patients, while others will not. Since a provider will likely limit the proportion of their caseload dedicated to public insurance, participation in CHP+ will likely affect the proportion of caseload dedicated to Medicaid. The prior is also specific to the provider type t.

Let A_j be a Bernoulli random variable denoting whether provider j accepts CHP+ patients, with 1 denoting acceptance. Note that we cannot determine from the Medicaid caseload data whether each individual provider accepts CHP+. We can, however, see from the FTE data that the percentage of Medicaid providers who also accept CHP+ is around 80%. Therefore, we assume that $P(A_j = 1) = 0.8$ and $P(A_j = 0) = 0.2$. Then, we

assume that the distribution of $q_{j|t}$, conditional on A_j, is as follows:

$$q_{j|t} \mid A_j = 1 \sim Beta(\omega_{1t}, \omega_{2t})$$

$$q_{j|t} \mid A_j = 0 \sim Beta(\gamma_{1t}, \gamma_{2t})$$

The second Beta distribution, with parameters γ_{1t} and γ_{2t}, describes how the total caseload is split between Medicaid caseload and excess caseload available only to the privately insured. The first Beta distribution, with parameters ω_{1t} and ω_{2t}, describes how the total caseload is split between Medicaid caseload and caseload available to patients insured through both private insurance and CHP+. In fact, we may think of the distribution of $q_{j|t} \mid A_j = 1$ as the marginal distribution of the proportion of the total caseload dedicated to Medicaid, where the proportions of caseload dedicated to Medicaid, CHP+, and private insurance, respectively, come from a Dirichlet distribution. We return to this fact in Appendix C.1.5.

The prior on $q_{j|t}$ is then given by a mixture distribution

$$p(q_{j|t} \mid \omega_{1t}, \omega_{2t}, \gamma_{1t}, \gamma_{2t}) = \sum_{a_j=0}^{1} p(q_{j|t} \mid \alpha_{1t}, \alpha_{2t}, \gamma_{1t}, \gamma_{2t}, a_j) p(A_j = a_j)$$

$$= 0.8 \times Beta(\omega_{1t}, \omega_{2t}) + 0.2 \times Beta(\gamma_{1t}, \gamma_{2t})$$

The hyperparameters $\omega_{1t}, \omega_{2t}, \gamma_{1t}$ and γ_{2t} are estimated in the following way. For each of the providers in the Medicaid caseload data with taxonomy t, we divide their caseload by 10,500, trimming caseloads that exceed 10,500. Doing so gives us an ad hoc sample of proportions of caseload dedicated to Medicaid patients under the assumption that a provider can see 10,500 patients per year. The assumption of 10,500 visits per provider is taken from HPI data.[99]. The hyperparameters are then estimated from the data using maximum

likelihood estimation. The final estimates by taxonomy are given in Table C.6.

Table C.6: Medicaid caseload proportion Beta mixture hyperparameters by taxonomy

$\hat{\omega}_{1t}$	$\hat{\omega}_{2t}$	$\hat{\gamma}_{1t}$	$\hat{\gamma}_{2t}$
3.8372768	99.5271701	0.6481611	2.2043669

(a) General Dentist

$\hat{\omega}_{1t}$	$\hat{\omega}_{2t}$	$\hat{\gamma}_{1t}$	$\hat{\gamma}_{2t}$
0.7023452	0.9240403	1.9830447	11.578234

(b) Pediatric Dentist

Once the hyperparameters $\hat{\omega}_{1t}$ $\hat{\omega}_{2t}$ $\hat{\gamma}_{1t}$ $\hat{\gamma}_{2t}$ are obtained, the hyperparameters a_t and b_t are estimated from Medicaid claims data using maximum likelihood estimation. First, we calculate the distribution of the capacities given the hyperparameters. The likelihood of the Medicaid caseloads of providers with taxonomy t can be written

$$
\begin{aligned}
&\mathrm{p}(\mathrm{c}_j^M \mid a_t, b_t, \omega_{1t}, \omega_{2t}, \gamma_{1t}, \gamma_{2t}) \\
&= \textstyle\sum_{c_j=c_j^M}^{\infty} \int_0^1 \int_0^\infty p(c_j^M \mid c_j, q_{j|t}) p(c_j \mid \lambda_{j|t}) p(\lambda_{j|t} \mid a_t, b_t) p(q_{j|t} \mid \omega_{1t}, \omega_{2t}, \gamma_{1t}, \gamma_{2t}) d\lambda_{j|t} dq_{j|t} \\
&= 0.8 \times \\
&\tfrac{\Gamma(c_j^M+\omega_{1t})\Gamma(c_j^M+a_t)\Gamma(\omega_{1t}+\omega_{2t})}{c_j^M!\Gamma(a_t)\Gamma(\omega_{1t})\Gamma(c_j^M+\omega_{1t}+\omega_{2t})} (\tfrac{b_t}{1+b_t})^{a_t} (\tfrac{1}{1+b_t})^{c_j^M} {}_1F_2(\omega_{2t}, c_j^M + a_t; \omega_{1t} + \omega_{2t} + c_j^M; \tfrac{1}{1+b_t}) \\
&+ 0.2 \times \\
&\tfrac{\Gamma(c_j^M+\gamma_{1t})\Gamma(c_j^M+a_t)\Gamma(\gamma_{1t}+\gamma_{2t})}{c_j^M!\Gamma(a_t)\Gamma(\gamma_{1t})\Gamma(c_j^M+\gamma_{1t}+\gamma_{2t})} (\tfrac{b_t}{1+b_t})^{a_t} (\tfrac{1}{1+b_t})^{c_j^M} {}_1F_2(\gamma_{2t}, c_j^M + a_t; \gamma_{1t} + \gamma_{2t} + c_j^M; \tfrac{1}{1+b_t})
\end{aligned}
$$

where ${}_1F_2(\cdot,\cdot;\cdot;\cdot)$ is the hypergeometric function. Then, the likelihood is the product of the densities,

$$L(a_t, b_t) = \prod_{j=1}^{J_t} p(c_j^M \mid a_t, b_t, \omega_{1t}, \omega_{2t}, \gamma_{1t}, \gamma_{2t})$$

,

where J_t is the number of providers of taxonomy t. The hyperparameters are then estimated with maximum likelihood estimation, i.e. a_t and b_t are estimated using gradient

descent solving the optimization problem

$$\hat{a}_t, \hat{b}_t = \arg\max_{a_t, b_t} \log(L(a_t, b_t))$$

The final estimates of the Gamma hyperparameters by taxonomy are given in Table C.7.

Table C.7: Total Caseload Gamma hyperparameters by taxonomy

a_t	b_t
0.3698297	.00005449144

(a) General Dentist

$\hat{a}_t$	$\hat{b}_t$
5.3312234928	0.0006929849

(b) Pediatric Dentist

The hyperparameter estimates $\hat{a}_t, \hat{b}_t, \hat{\omega}_{1t}, \hat{\omega}_{2t}, \hat{\gamma}_{1t}, \hat{\gamma}_{2t}$ are then used to generate capacity estimates for each of the B input data samples.

Capacity Sampling Procedures

To obtain sampled instances of the provider capacities using Monte Carlo simulations, we sampled from the estimated prior distribution. Because we lack data on the capacities of individual providers, the prior distribution cannot be updated. However, the empirical Bayes estimation procedure ensures that the sampled capacities reflect typical provider behavior. For each of the B samples generated, the end result is a set of capacity input parameters $(c_j, c_j^M, c_j^H)_b$ for each provide $j \in \{1, \ldots, \mathbb{J}\}$.

For each sample b and each provider location j, our sampling procedure accounts for both the provider taxonomy t and the full time equivalent number of providers working at that location FTE_j. In order to sample the caseload for the provider locations, we first account for the number of full time equivalent providers by sampling the expected caseload of provider j, $\lambda_{j|t,b}$, from a Gamma distribution with parameters $FTE_j \times \hat{a}_t$ and $\hat{b}_t$, where FTE_j is the full time equivalent number of individual providers at location j. The Gamma($FTE_j \times \hat{a}_t$, $\hat{b}_t$) distribution is the distribution of the sum of FTE_j draws from a Gamma($\hat{a}_t$, $\hat{b}_t$) distribution, reflecting the total caseload available at each provider

location as the sum of the caseload contributed by FTE_j individual providers. Once the sampled expected number of visits, $\lambda_{j|t,b}$, is obtained for each provider location, we sample the total caseloads, C_j, from the Poisson($\lambda_{j|t,b}$) distribution.

There are then three possibilities regarding the public insurance caseload. The first is that provider location j does not accept public insurance, in which case we set $c_{jb}^M = c_{jb}^H = 0$. The second and third possibilities are that the provider location accepts exactly one of Medicaid *or* CHP+ and that the provider location accepts both Medicaid *and* CHP+, respectively.

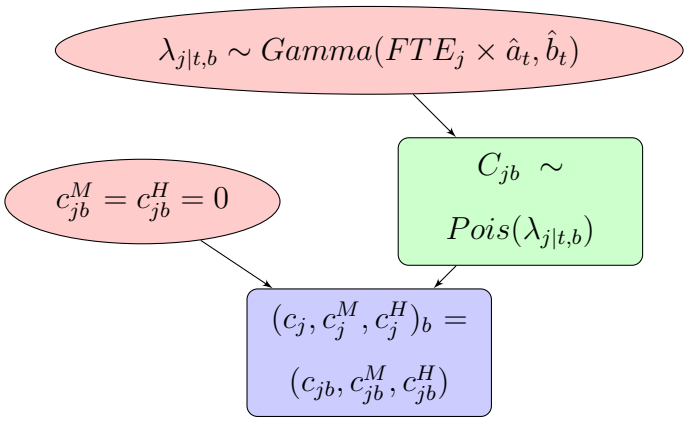


Figure C.4: Sampling of Capacity Parameters: Providers not accepting public insurance

If the providers at locations j accept exactly one of Medicaid *or* CHP+, the next step is to sample the portion of the caseload available to the Medicaid or CHP+ population, setting the caseload available to the unaccepted insurance type to 0. Given the sampled total caseload at location j, c_{jb}, the caseload is split according to a binomial distribution, with the proportions drawn from a Beta distribution. The parameters of the Beta distribution, $\hat{\gamma}_{1t}$ and $\hat{\gamma}_{2t}$, are taken from Table C.6 according the taxonomy of provider location j. Implicit in this procedure is that providers accepting CHP+ but not Medicaid will accept CHP+ patients at the same rate as providers accepting Medicaid but not CHP+ accept Medicaid patients.

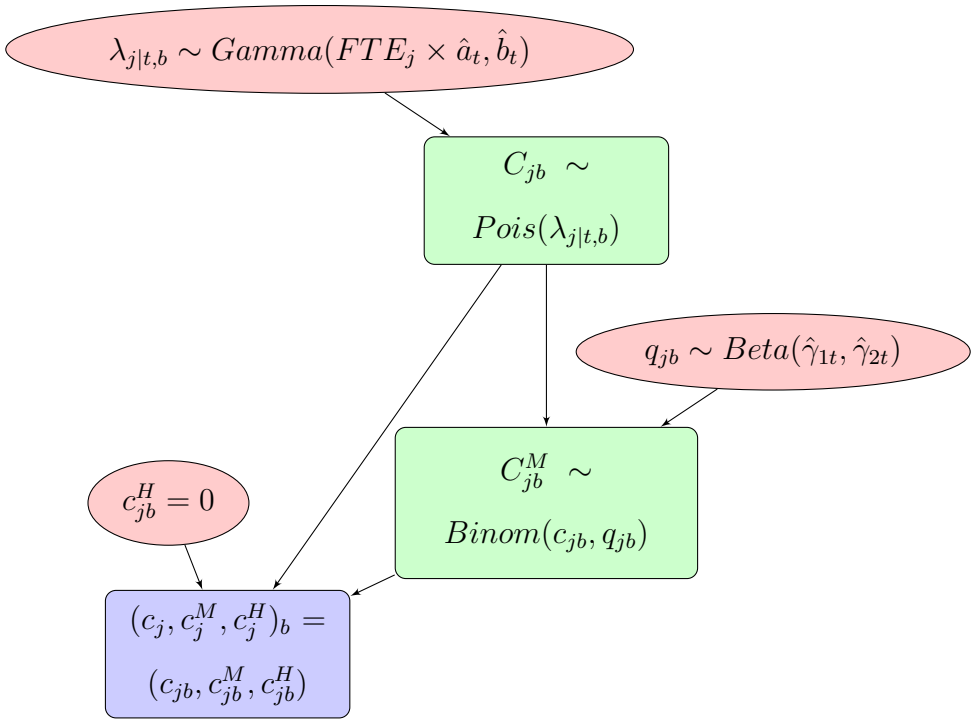


Figure C.5: Sampling of Capacity Parameters: Providers accepting Medicaid *or* CHP+

If provider location j accepts both Medicaid *and* CHP+, the next step is to split the total caseload into caseload available to the Medicaid population, the CHP+ population, and the general population. Recall that in cases where a provider locations accepts both Medicaid and CHP+, we assume that FTE_j to be used in calculating the total caseload is the maximum between the the FTE for Medicaid and CHP+, i.e. $FTE_j = \max\{FTE^M_j, FTE^H_j\}$. We assume that the minimum of the individual FTEs, $FTE^-_J = \min\{FTE^M_j, FTE^H_j\}$, is the number of providers serving both the Medicaid and CHP+ populations, while the remaining $FTE_j - FTE^-_j$ providers server only Medicaid or CHP+. The caseload for these remaining providers is sampled as in Figure C.5. For the providers serving both populations, the caseload is split according to a multinomial distribution, with the proportions drawn from a Dirichlet distribution. Since the Beta distribution with parameters $\hat{\omega}_{1t}$ and $\hat{\omega}_{2t}$ (see Table C.6) may be thought of a marginal distribution of only the proportion of the caseload available to Medicaid, we can simply split the second parameter $\hat{\omega}_{2t}$ in accor-

dance with assumptions about how much is available to CHP+ patients and the privately insured, respectively. Assuming that the caseload available to CHP+ and Medicaid are similar, we assume that the parameters of the Dirichlet distribution are $\hat{\omega}_{1t}$, $\hat{\omega}_{1t}$, and $\hat{\omega}_{2t} - \hat{\omega}_{1t}$, with the first two parameters corresponding to the Medicaid and CHP+ proportions and the third parameter corresponding to the excess caseload available exclusively to the privately insured.

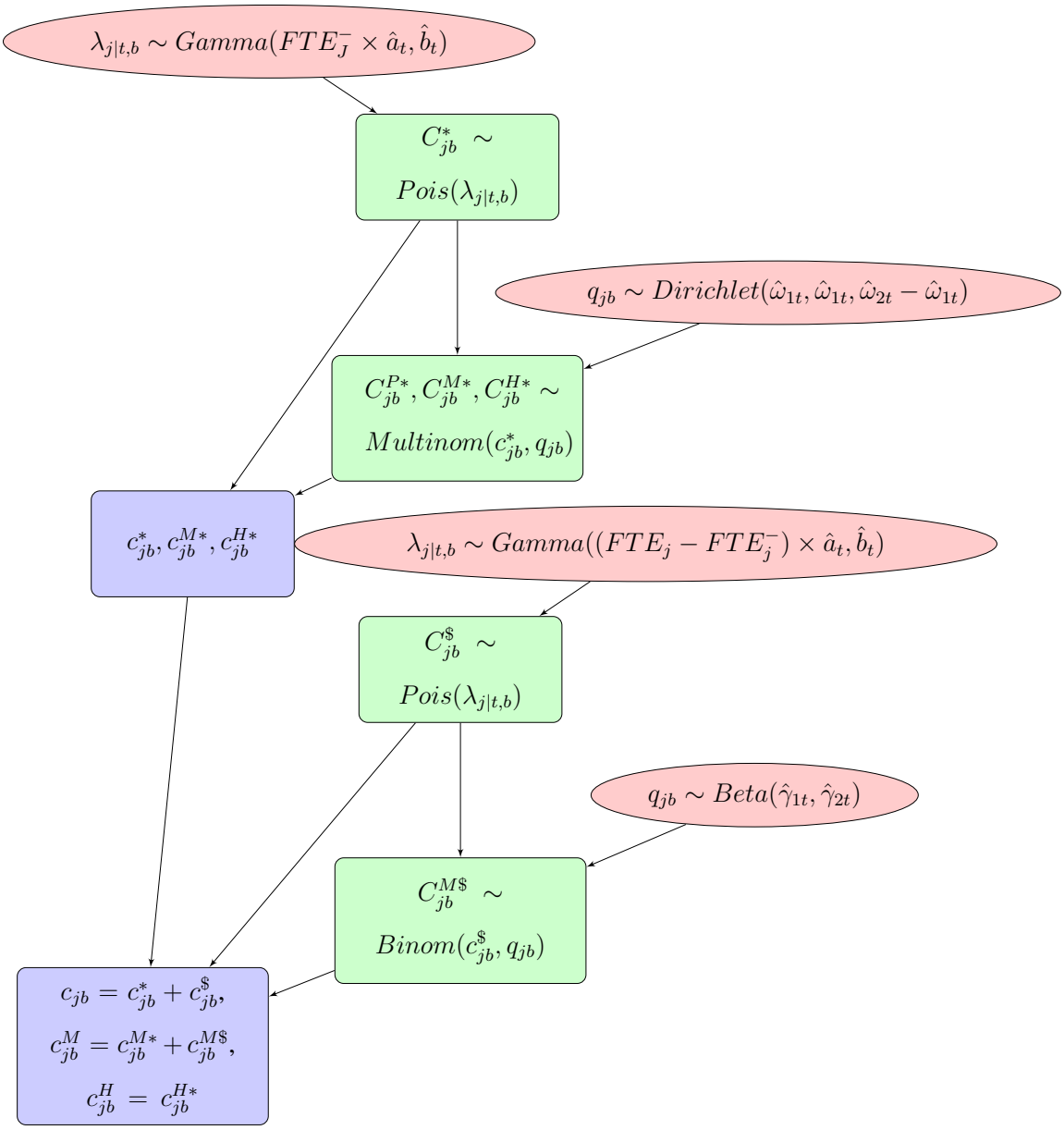


Figure C.6: Sampling of Capacity Parameters: Providers accepting Medicaid *and* CHP+, assuming $FTE_j^M > FTE_j^H$. If $FTE_j^M < FTE_j^H$, swap positions of FTE_j^M, FTE_j^H.

C.1.6 Statistical Inference for Access Measures

Inference Procedures

After obtaining the posterior distributions on the supply/demand parameters of the optimization model, we sample B times, defining B optimization problems. We solve the op-

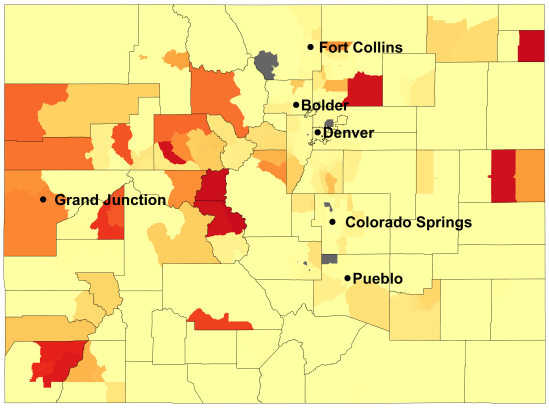
timization for each of the B samples to obtain matchings $\hat{\theta}_1, \ldots, \hat{\theta}_B$ and corresponding access measures $\{(\hat{z}_i, \hat{y}_i, \hat{w}_i, \hat{u}_i)_1, \ldots, (\hat{z}_i, \hat{y}_i, \hat{w}_i, \hat{u}_i)_B\}$ for each census tract $i \in \{1, \ldots, \mathbb{C}\}$. This set of B values of the access measures constitutes an empirical sample from the posterior distribution on the access measures.

We pause here for a slight shift in notation which will be convenience in the subsequent paragraphs. We denote the disparity measures for census tract i, calculated using the bth sample of input data, by $\hat{z}_{ib}$, $\hat{w}_{ib}$, $\hat{y}_{ib}$, and $\hat{u}_{ib}$. Disparities in spatial access to care between publicly and privately insured children can be estimated by the differences $\delta^u_{ib} = \hat{u}_{ib} - \hat{w}_{ib}$. Similarly, disparities between Medicaid insured and privately insured children and between CHP+ insured and privately insured can be estimated by the differences $\delta^z_{ib} = \hat{z}_{ib} - \hat{w}_{ib}$ and $\delta^y_{ib} = \hat{y}_{ib} - \hat{w}_{ib}$, respectively.

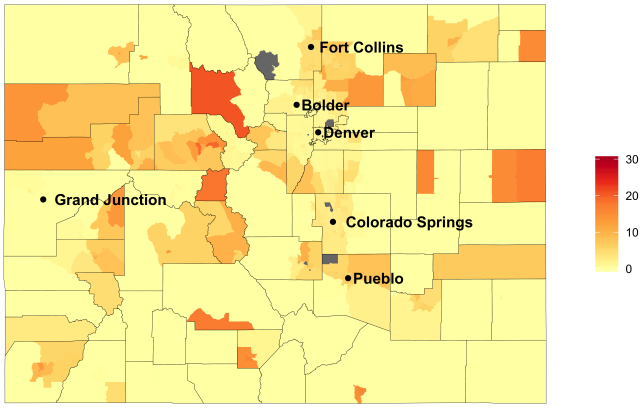
We make inferences on the access measures and disparities in access using simultaneous statistical inference across multiples locations while accounting for spatial dependence using simultaneous confidence bands derived from samples of data points at sample locations form a spatial process. Specifically, we assume that the samples $\chi_{ib}, b = 1, \ldots, B$, are realizations from the random process $\chi(s)$ with mean and covariance functions $\mu(s)$ and $\Sigma(s, s')$, respectively. The samples could be either the samples of the access measures or disparity measures. For locations $s_1, \ldots, s_{\mathbb{C}}$, we have B samples of $\chi_b = (\chi_{1b}, \ldots, \chi_{\mathbb{C}b})$ from the $\mathbb{C}$-variate distribution with mean $\mu_{\mathbb{C}} = (\mu(s_1), \ldots, \mu(s_{\mathbb{C}}))$ and covariance matrix $\Sigma_{\mathbb{C}} = \{\Sigma(s_i, s_j) : i, j \in 1, \ldots, \mathbb{C}\}$. We estimate a simultaneous confidence band for the unknown function $\mu(s)$ over the spatial domain S applying a similar methodology introduced for the estimation of simultaneous credible bands using Markov Chain Monte Carlo approaches.[93] Specifically, we construct simultaneous credible bands by first constructing pointwise credible intervals derived from the $\alpha/2$ and $1 - \alpha/2$ percentiles of the samples $\chi_{i1}, \ldots, \chi_{iB}$ at each location s_i, $i = 1, \ldots, \mathbb{C}$. In the second step, we scale the pointwise intervals by a constant factor until $100(1 - \alpha)\%$ of the samples $\chi_1, \ldots, \chi_B$ are contained in the credible band simultaneously. This approach allows for the credible band

to be more narrow where information is highly concentrated and wider where information is more disparate. In this paper, we use $\alpha = 10\%$, constructing 90% credible intervals.

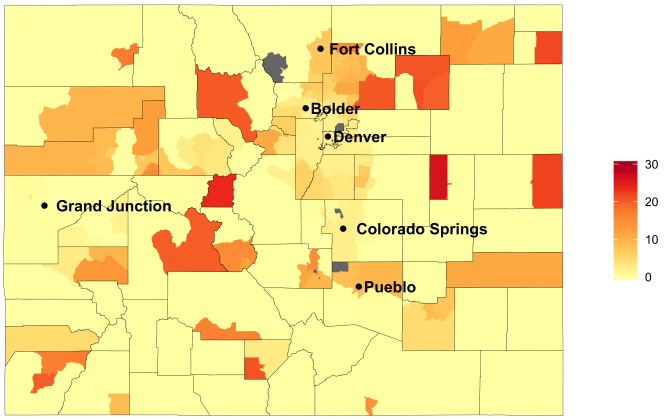
Using simultaneous confidence bands, we can further test whether 1) the access measure is greater than some access standard T and 2) the disparities in access are lower than some intervention threshold value T (e.g., 5 mile in distance traveled). These hypotheses indicate the presence of systemic 1) lack of sufficient access to dental care and 2) the presence of systematic disparities at the threshold level T. This analysis provides statistical inferences on where to intervene and which spatial access measures lead to the most impact in reducing disparities in access.



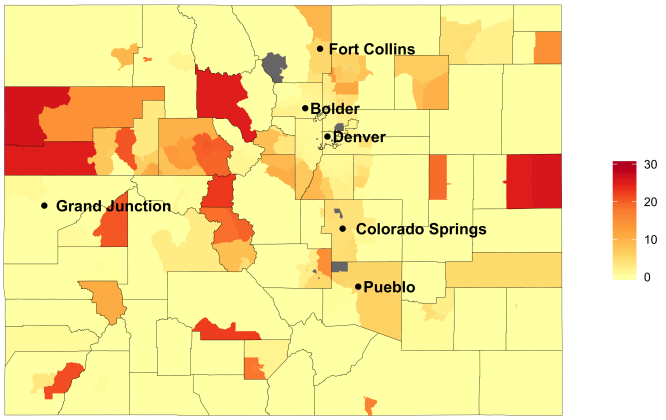
(a) Private Financial Access



(b) All Publicly Insured

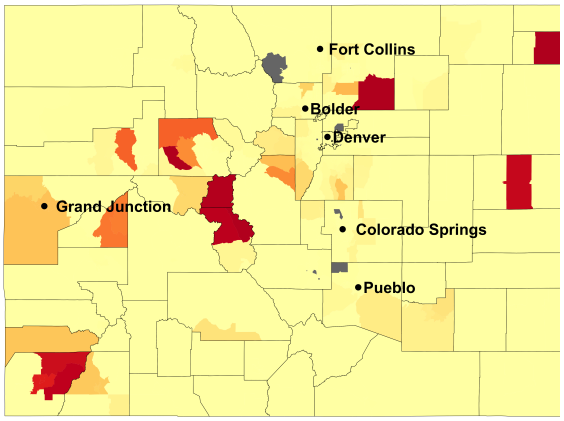


(c) Medicaid Insured

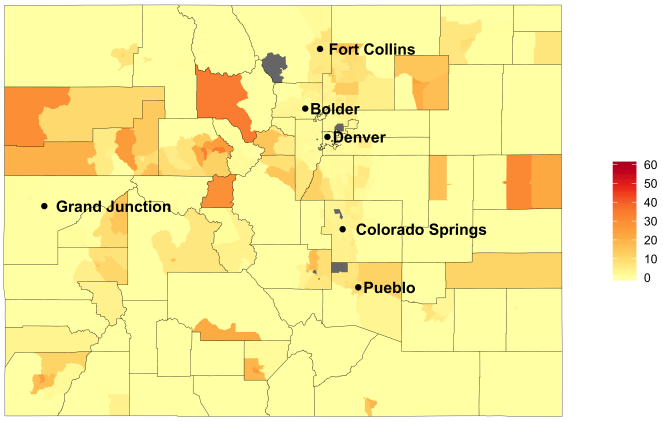


(d) CHP+ Insured

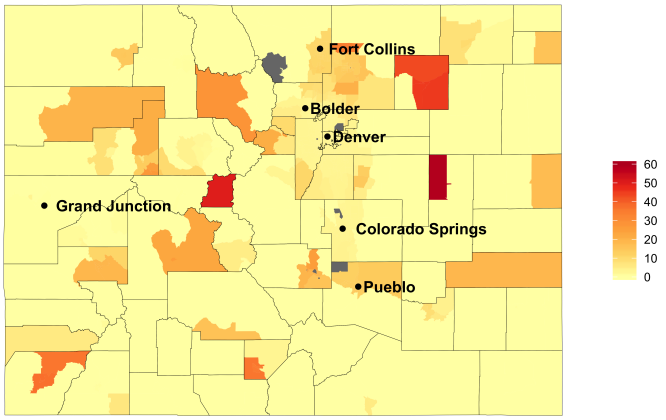
Figure C.7: Standard Deviation of Disparity Measures by Insurance Type



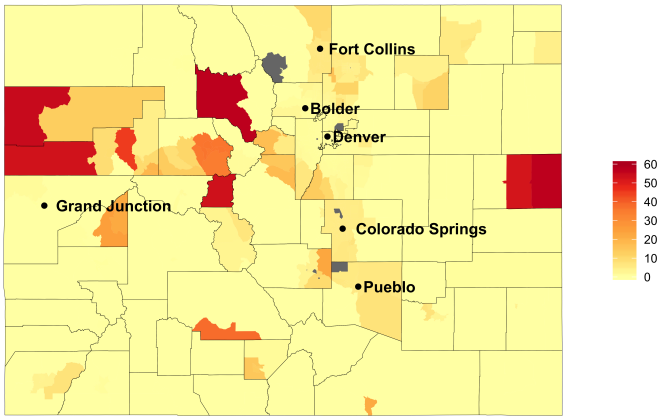
(a) Private Financial Access



(b) All Publicly Insured



(c) Medicaid Insured

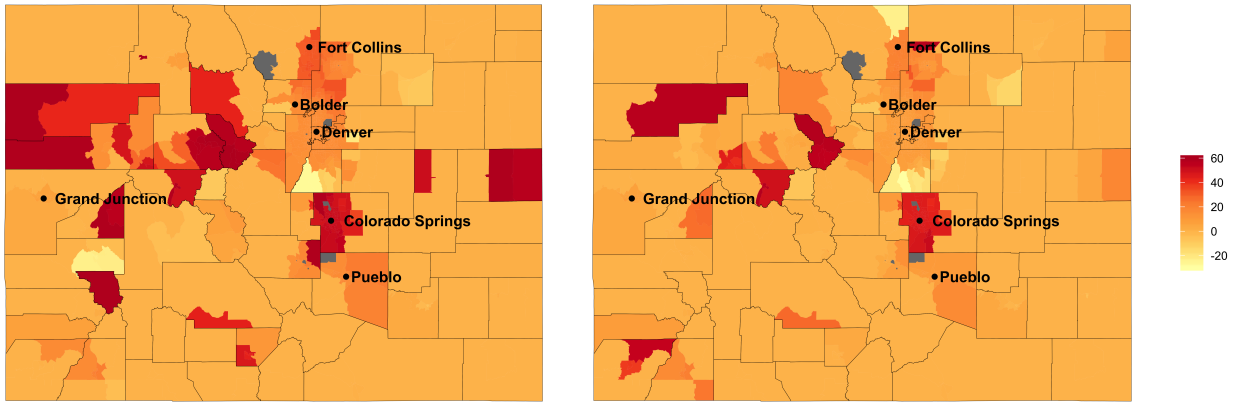


(d) CHP+ Insured

Figure C.8: IQR of Disparity Measures by Insurance Type

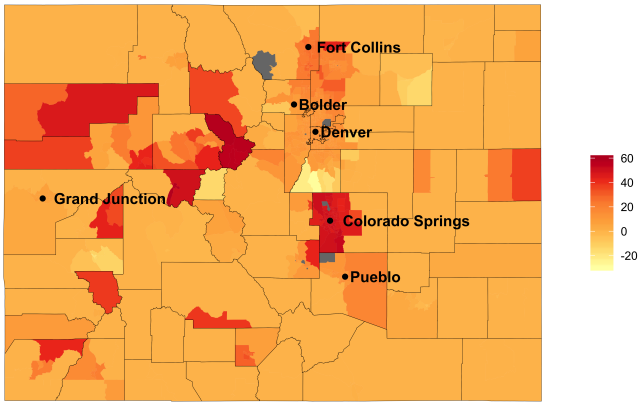
Disparities by Urbanicity and Insurance Type

The median disparity between publicly and privately insured children was greater than 2 miles in 88.5% of census tracts, greater than 5 miles in 87.2% of census tracts, greater than 10 miles in 73.9% of census tracts. The median disparity between the publicly and privately insured was 12.51 (25th percentile = 10.72, 75th percentile = 15.84) miles in urban areas, 5.63 (-0.40, 20.25) miles in suburban areas, and 1.61 (0, 24.11) miles in rural areas. The median disparity between the Medicaid population and privately insured was 15.51 (15.03, 18.85) miles in urban areas, 7.83 (0, 21.30) miles in suburban areas, and 0.79 (0, 39.89) miles in rural areas. The median disparity between the CHP+ population and privately insured was 8.53 (8.21, 11.89) miles in urban areas, 1.00 (0, 7.24) miles in suburban areas, and 0.77 (0, 7.13) miles in rural areas.



(a) Medicaid Insured

(b) CHP+ Insured



(c) Publicly Insured

Figure C.9: Median disparity measure, by financial access

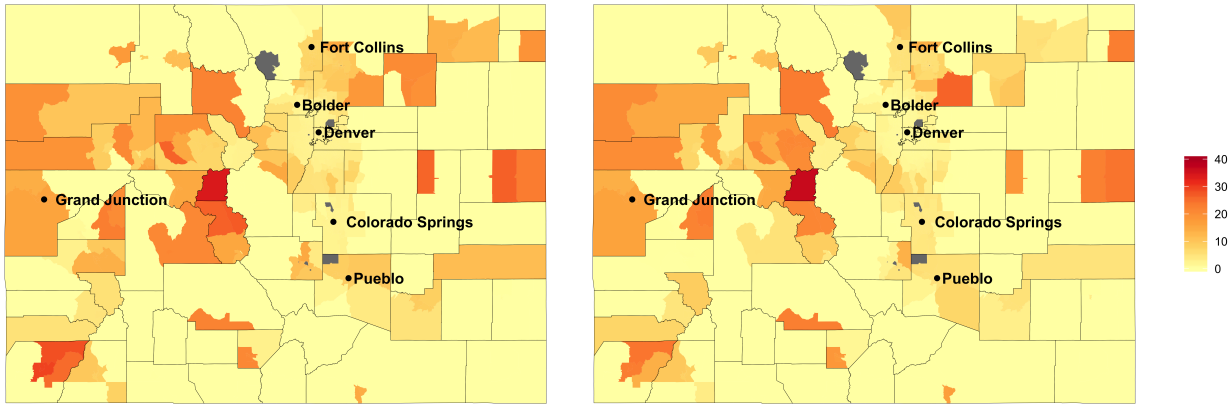
Tables C.8 and C.9 provide a full breakdown of the percentage of census tracts, by urbanicity, with median disparity measures above the threshold $T = 2$ and $T = 10$ miles, respectively, as well as the percentage of census tracts found to have disparities significantly above the T miles.

Table C.8: Percentage of census tracts with disparity measures greater than and *significantly* greater than $T = 2$ miles, by urbanicity and type of financial access

	Urban		Suburban		Rural	
Financial Access	% Above Std	% Sig	% Above Std	% Sig	% Above Std	% Sig
Public	81.1%	71.9%	2.9%	0.8%	2.8%	0.5%
Medicaid	81.2%	69.2%	2.9%	0.6%	2.8%	0.4%
CHP+	80.6%	62.2%	2.5%	0.6%	2.1%	0.3%

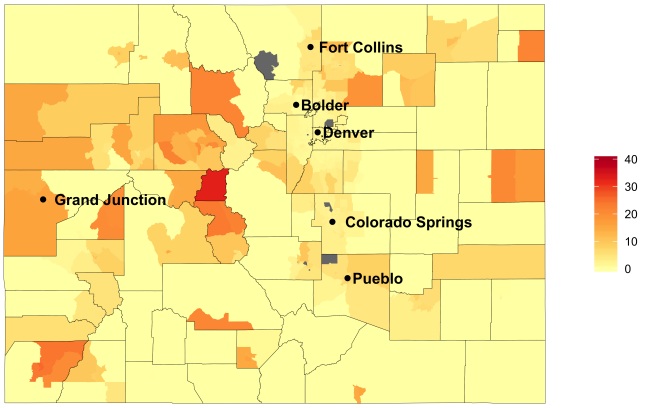
Table C.9: Percentage of census tracts with disparity measures greater than and *significantly* greater than $T = 10$ miles, by urbanicity and type of financial access

	Urban		Suburban		Rural	
Financial Access	% Above Std	% Sig	% Above Std	% Sig	% Above Std	% Sig
Public	68.5%	11.7%	2.3%	0.2%	2%	0.2%
Medicaid	78.6%	12%	2.7%	0.2%	1.9%	0.2%
CHP+	24.6%	10.2%	1.1%	0.2%	1.4%	0.2%



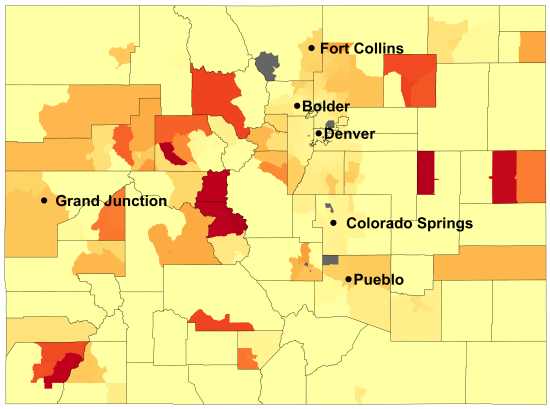
(a) Medicaid Insured

(b) CHP+ Insured

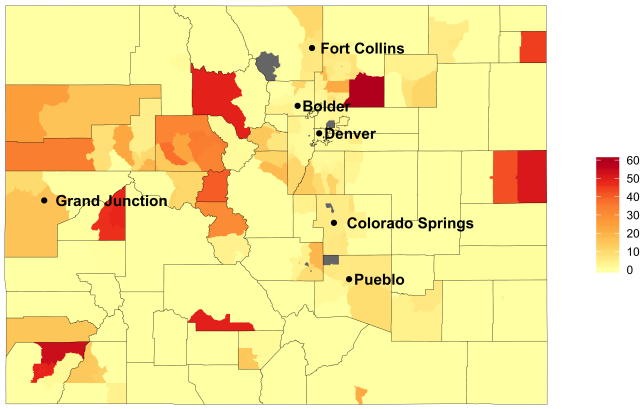


(c) Publicly Insured

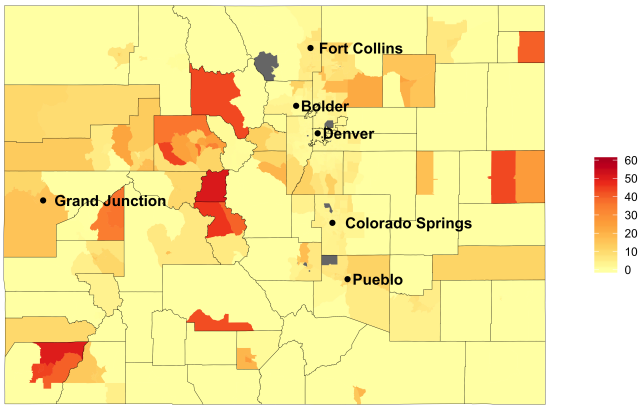
Figure C.10: Standard deviation of disparity measures, by financial access



(a) Medicaid Insured

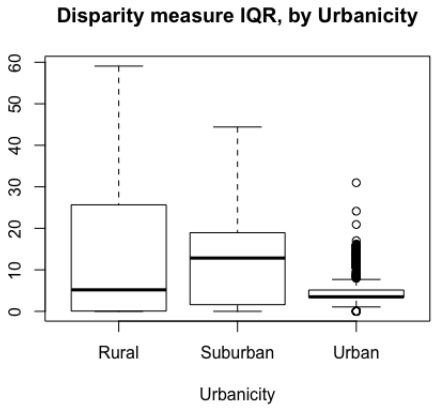


(b) CHP+ Insured

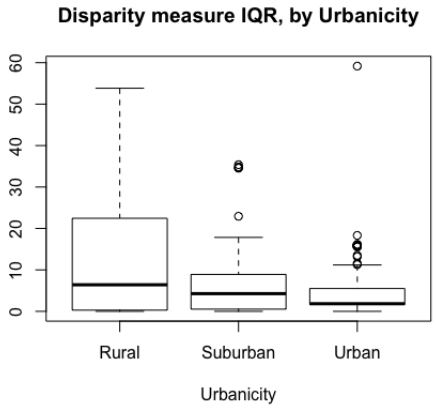


(c) Publicly Insured

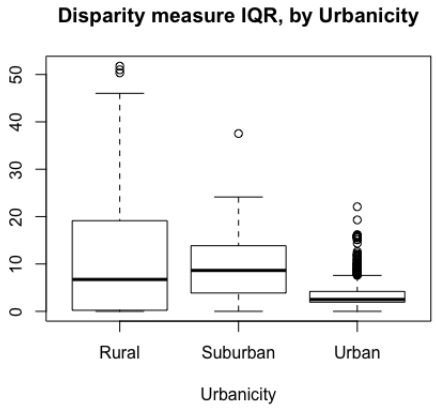
Figure C.11: IQR of disparity measures, by financial access



(a) Medicaid disparity measures

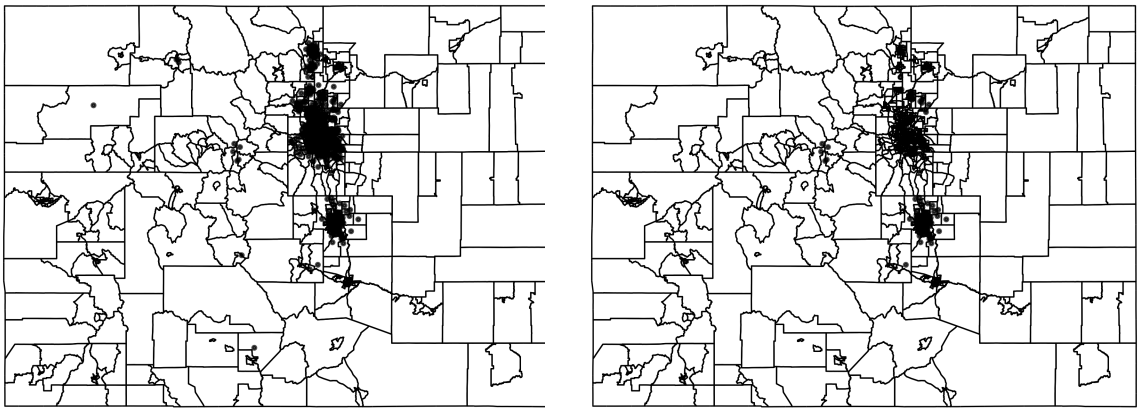


(b) CHP+ disparity measures



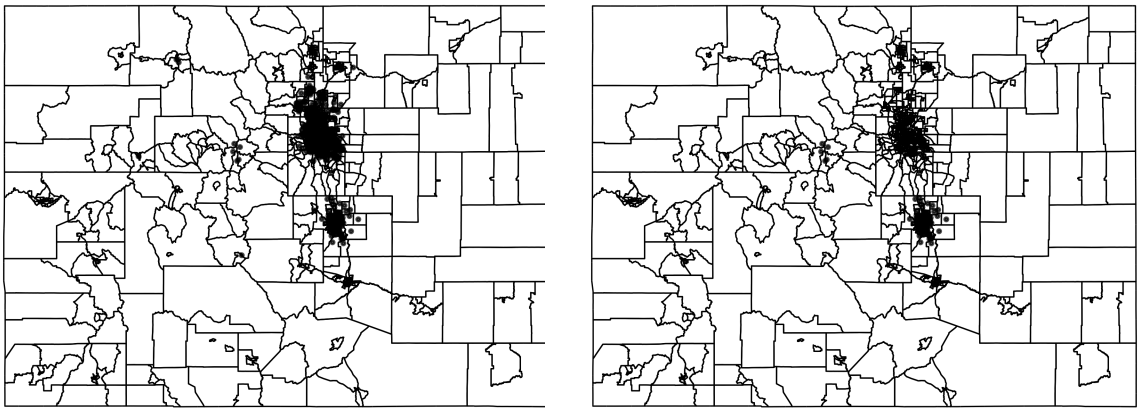
(c) Public insurance disparity measures

Figure C.12: Boxplots of the census tract level IQR of disparity measures for three public insurance types, grouped by urbanicity



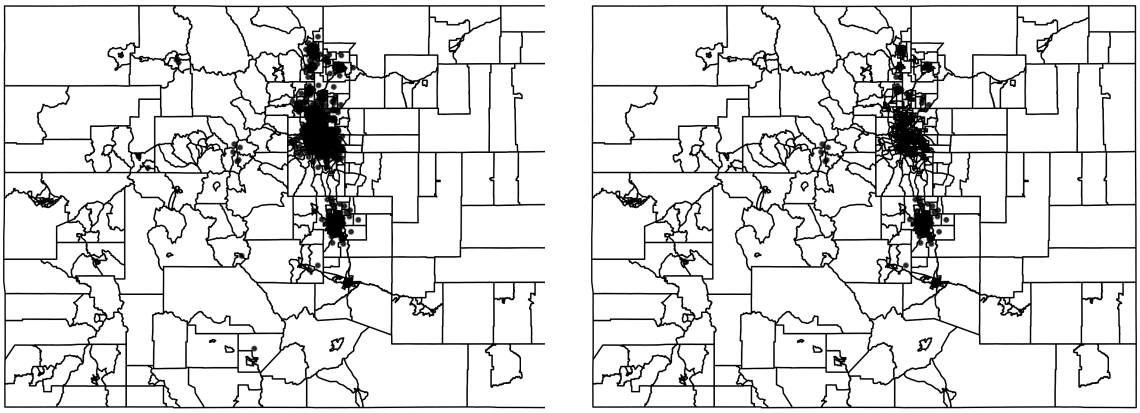
(a) $T = 2$ (b) $T = 10$

Figure C.13: Significant Disparities in Access of T Miles or more: Medicaid and Private Insurance



(a) $T = 2$ (b) $T = 10$

Figure C.14: Significant Disparities in Access of T Miles or more: CHP+ and Private Insurance



(a) $T = 2$ (b) $T = 10$

Figure C.15: Significant Disparities in Access of T Miles or more: Public and Private Insurance

REFERENCES

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers", *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[2] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for machine learning*. Mit Press, 2012.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton University Press, 2009.

[4] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios, "Robust optimization of large-scale systems", *Operations research*, vol. 43, no. 2, pp. 264–281, 1995.

[5] A. Shapiro and T. Homem-de Mello, "A simulation-based approach to two-stage stochastic programming with recourse", *Mathematical Programming*, vol. 81, no. 3, pp. 301–325, 1998.

[6] A. Shapiro, *Stochastic programming by monte carlo simulation methods*, Electronic Article, 2000.

[7] J. L. Higle and S. W. Wallace, "Sensitivity analysis and uncertainty in linear programming", *Interfaces*, vol. 33, no. 4, pp. 53–60, 2003.

[8] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*. Athena Scientific Belmont, MA, 1997, vol. 6.

[9] W. L. Winston and J. B. Goldberg, *Operations research: applications and algorithms*. Thomson/Brooks/Cole Belmont, CA, 2004, vol. 3.

[10] H Arsham and M Oblak, "Perturbation analysis of general lp models: A unified approach to sensitivity, parametric, tolerance, and more-for-less analysis", *Mathematical and Computer Modelling*, vol. 13, no. 8, pp. 79–102, 1990.

[11] T. Gal, *Postoptimal Analyses, Parametric Programming, and Related Topics*, ser. Advanced Book Program - McGraw-Hill Book Company. McGraw-Hill International, 1979, ISBN: 9780070226791.

[12] R. E. Wendell, "The tolerance approach to sensitivity analysis in linear programming", *Management Science*, vol. 31, no. 5, pp. 564–578, 1985.

[13] H. M. Wagner, "Global sensitivity analysis", *Operations Research*, vol. 43, no. 6, pp. 948–969, 1995.

[14] Z. Yu, W. B. Rouse, and N. Serban, "A computational theory of enterprise transformation", *Systems Engineering*, vol. 14, no. 4, pp. 441–454, 2011.

[15] N. R. Council *et al.*, *Assessing the reliability of complex models: mathematical and statistical foundations of verification, validation, and uncertainty quantification*. National Academies Press, 2012.

[16] M. Nobles, N. Serban, and J. Swann, "Spatial accessibility of pediatric primary healthcare: Measurement and inference", *The Annals of Applied Statistics*, pp. 1922–1946, 2014.

[17] E. Garcia, N. Serban, J. Swann, and A. Fitzpatrick, "The effect of geographic access on severe health outcomes for pediatric asthma", *Journal of Allergy and Clinical Immunology*, vol. 136, no. 3, pp. 610–618, 2015.

[18] M. Gentili, K. Isett, N. Serban, and J. Swann, "Small-area estimation of spatial access to care and its implications for policy", *Journal of Urban Health*, vol. 92, no. 5, pp. 864–909, 2015.

[19] M. Gentili, P. Harati, and N. Serban, "Projecting the impact of the affordable care act provisions on accessibility and availability of primary care providers for the adult population in georgia", *American journal of public health*, vol. 106, no. 8, pp. 1470–1476, 2016.

[20] M. Nobles, N. Serban, J. Swann, *et al.*, "Spatial accessibility of pediatric primary healthcare: Measurement and inference", *The Annals of Applied Statistics*, vol. 8, no. 4, pp. 1922–1946, 2014.

[21] T. Gal and H. Greenberg, *Advances in Sensitivity Analysis and Parametric Programming*, ser. International Series in Operations Research & Management Science. Springer US, 1997, ISBN: 9780792399179.

[22] C. Filippi, "Sensitivity analysis in linear programming", *Wiley Encyclopedia of Operations Research and Management Science*, 2011.

[23] J. E. Ward and R. E. Wendell, "Approaches to sensitivity analysis in linear programming", *Annals of Operations Research*, vol. 27, no. 1, pp. 3–38, 1990.

[24] B. Jansen, J. De Jong, C. Roos, and T. Terlaky, "Sensitivity analysis in linear programming: Just be careful!", *European Journal of Operational Research*, vol. 101, no. 1, pp. 15–28, 1997.

[25] S. W. Wallace, “Decision making under uncertainty: Is sensitivity analysis of any use?”, *Operations Research*, vol. 48, no. 1, pp. 20–25, 2000.

[26] S. Bradley, A. Hax, and T. Magnanti, *Applied Mathematical Programming*. Addison Wesley, 1977.

[27] R. E. Wendell, “Using bounds on the data in linear programming: The tolerance approach to sensitivity analysis”, *Mathematical Programming*, vol. 29, no. 3, pp. 304–322, 1984.

[28] M. M. Babbar, “Distributions of solutions of a set of linear equations (with an application to linear programming)”, *Journal of the American Statistical Association*, vol. 50, no. 271, pp. 854–869, 1955.

[29] J. B. Ewbank, B. L. Foote, and H. J. Kumin, “Method for solution of distribution problem of stochastic linear-programming”, *SIAM Journal on Applied Mathematics*, vol. 26, no. 2, pp. 225–238, 1974.

[30] A. Prékopa, “On the probability distribution of the optimum of a random linear program”, *SIAM Journal on Control*, vol. 4, no. 1, pp. 211–222, 1966.

[31] G. Tintner and M. V. R. Sastry, “A note on the use of nonparametric statistics in stochastic linear programming”, *Management Science*, vol. 19, no. 2, pp. 205–210, 1972.

[32] H. M. Wagner, “On the distribution of solutions in linear programming problems”, *Journal of the American Statistical Association*, vol. 53, no. 281, pp. 161–163, 1958.

[33] E. Borgonovo, G. T. Buzzard, and R. E. Wendell, “A global tolerance approach to sensitivity analysis in linear programming”, *European Journal of Operational Research*, vol. 267, no. 1, pp. 321 –337, 2018.

[34] J. E. Oakley and A. O’Hagan, “Probabilistic sensitivity analysis of complex models: A bayesian approach”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 66, no. 3, pp. 751–769, 2004.

[35] P. Hanafizadeh, A. Ghaemi, and M. Tavana, “Local perturbation analysis of linear programming with functional relation among parameters”, *International Journal of Operations Research and Information Systems*, vol. 2, no. 1, pp. 42–65, 2011.

[36] A. Shahin, P. Hanafizadeh, and M. Hladík, “Sensitivity analysis of linear programming in the presence of correlation among right-hand side parameters or objective function coefficients”, *Central European Journal of Operations Research*, vol. 24, no. 3, pp. 563–593, 2016.

[37] I. Lee, S. Curry, and N. Serban, "Solving large batches of linear programs", *INFORMS Journal on Computing*, 2019.

[38] T. Gal, "Rim multiparametric linear programming", *Management Science*, vol. 21, no. 5, pp. 567–575, 1975.

[39] C. N. Jones, M. Barić, and M. Morari, "Multiparametric linear programming with applications to control", *European Journal of Control*, vol. 13, no. 2-3, pp. 152–170, 2007.

[40] D. Q. Mayne, "Model predictive control: Recent developments and future promise", *Automatica*, vol. 50, no. 12, pp. 2967–2986, 2014.

[41] A. Alessio and A. Bemporad, "A survey on explicit model predictive control", in *Nonlinear model predictive control*, Springer, 2009, pp. 345–369.

[42] Y. Ji, R. J. Thomas, and L. Tong, "Probabilistic forecasting of real-time lmp and network congestion", *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 831–841, 2017.

[43] C. Filippi, "A fresh view on the tolerance approach to sensitivity analysis in linear programming", *European Journal of Operational Research*, vol. 167, no. 1, pp. 1–19, 2005.

[44] R. E. Wendell, "Sensitivity analysis revisited and extended", *Decision Sciences*, vol. 23, no. 5, pp. 1127–1142, 1992.

[45] F. R. Wondolowski, "A generalization of wendell's tolerance approach to sensitivity analysis in linear programming", *Decision Sciences*, vol. 22, no. 4, pp. 792–811, 1991.

[46] R. E. Wendell and W. Chen, "Tolerance sensitivity analysis: Thirty years later", *Croatian Operational Research Review*, vol. 1, no. 1, pp. 12–21, 2010.

[47] M. Hladík, "Tolerance analysis in linear systems and linear programming", *Optimization Methods and Software*, vol. 26, no. 3, pp. 381–396, 2011.

[48] H. Arsham and M. Oblak, "Perturbation analysis of general lp models - a unified approach to sensitivity, parametric, tolerance, and more-for-less analysis", *Mathematical and Computer Modelling*, vol. 13, no. 8, pp. 79–102, 1990.

[49] G. B. Dantzig, *Linear Programming and Extensions*. Princeton University Press, 1963.

[50] F. Borrelli, A. Bemporad, and M. Morari, "Geometric algorithm for multiparametric linear programming", *Journal of optimization theory and applications*, vol. 118, no. 3, pp. 515–540, 2003.

[51] M. Hladík, "Multiparametric linear programming: Support set and optimal partition invariancy", *European Journal of Operational Research*, vol. 202, no. 1, pp. 25–31, 2010.

[52] T. Gal, "Linear parametric programming - a brief survey", *Mathematical Programming Study*, vol. 21, no. Jun, pp. 43–68, 1982.

[53] R. E. Wendell, "Tolerance sensitivity and optimality bounds in linear programming", *Management Science*, vol. 50, no. 6, pp. 797–803, 2004.

[54] C. Filippi and G. Romanin-Jacur, "Multiparametric demand transportation problem", *European Journal of Operational Research*, vol. 139, no. 2, pp. 206–219, 2002.

[55] A. Bemporad, F. Borrelli, and M. Morari, "Model predictive control based on linear programming - the explicit solution", *IEEE Transactions on Automatic Control*, vol. 47, no. 12, pp. 1974–1985, 2002.

[56] P. Doganis, E. Aggelogiannaki, and H. Sarimveis, "A model predictive control and time series forecasting framework for supply chain management", *World Academy of Science, Engineering and Technology*, 2006.

[57] J. Niu, Z.-h. Xu, J. Zhao, Z.-j. Shao, and J.-x. Qian, "Model predictive control with an on-line identification model of a supply chain unit", *Journal of Zhejiang University SCIENCE C*, vol. 11, no. 5, pp. 394–400, 2010.

[58] P. Doganis, E. Aggelogiannaki, and H. Sarimveis, "A combined model predictive control and time series forecasting framework for production-inventory systems", *International Journal of Production Research*, vol. 46, no. 24, pp. 6841–6853, 2008.

[59] P. Ignaciuk and A. Bartoszewicz, "Linear-quadratic optimal control of periodic-review perishable inventory systems", *IEEE Transactions on Control Systems Technology*, vol. 20, no. 5, pp. 1400–1407, 2012.

[60] R. M. Van Slyke and R. Wets, "L-shaped linear programs with applications to optimal control and stochastic programming", *SIAM Journal on Applied Mathematics*, vol. 17, no. 4, pp. 638–663, 1969.

[61] A. Shapiro and T. Homem-de Mello, "A simulation-based approach to two-stage stochastic programming with recourse", *Mathematical Programming*, vol. 81, no. 3, pp. 301–325, 1998.

[62] L. Hong and H. Lam, "A statistical perspective on linear programs with uncertain parameters", in *The 2015 Winter Simulation Conference*, L. Yilmaz, W. Chan, I. Moon, T. Roeder, C. Macal, and M. Rossetti, Eds., 2015.

[63] A. Ruszczynski, "Regularized decomposition of stochastic programs: Algorithmic techniques and numerical results", Tech. Rep., 1993.

[64] A. Ruszczynski and A. Swietanowski, "Accelerating the regularized decomposition method for two stage stochastic linear problems", *European Journal of Operational Research*, vol. 101, no. 2, pp. 328–342, 1997.

[65] S. J. Garstka and D. P. Rutenberg, "Computation in discrete stochastic programs with recourse", *Operations Research*, vol. 21, no. 1, pp. 112–122, 1973.

[66] T. Gal and J. Nedoma, "Multiparametric linear programming", *Management Science Series a-Theory*, vol. 18, no. 7, pp. 406–422, 1972.

[67] S. I. Gass and T. L. Saaty, "Parametric objective function .2. generalization", *Journal of the Operations Research Society of America*, vol. 3, no. 4, pp. 395–401, 1955.

[68] B Bereanu, "On stochastic linear programming i. distribution problems: A single random variable", *Rev. Roumaine Math. Pures Appl*, vol. 8, no. 4, pp. 683–697, 1963.

[69] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*. Athena Scientific, 1997, xv, 587 p.

[70] Y. T. Lee and A. Sidford, "Path finding 1: Solving linear programs with õ(sqrt(rank)) linear system solves", *arXiv preprint arXiv:1312.6677*, 2015.

[71] T. Koch, T. Achterberg, E. Andersen, O. Bastert, T. Berthold, R. E. Bixby, E. Danna, G. Gamrath, A. M. Gleixner, S. Heinz, A. Lodi, H. Mittelmann, T. Ralphs, D. Salvagnin, D. E. Steffy, and K. Wolter, "MIPLIB 2010", *Mathematical Programming Computation*, vol. 3, no. 2, pp. 103–163, 2011.

[72] T. Gal, "Rim multiparametric linear programming", *Management Science*, vol. 21, no. 5, pp. 567–575, 1975.

[73] C. Barnhart and G. Laporte, *Handbooks in operations research and management science: transportation*. Elsevier, 2006, vol. 14.

[74] S. Dempe, *Wiley encyclopaedia of operations research and management science, by james j. cochran*, 2013.

[75] P. M. Griffin, C. R. Scherrer, and J. L. Swann, "Optimization of community health center locations and service offerings with statistical need estimation", *IIE transactions*, vol. 40, no. 9, pp. 880–892, 2008.

[76] M. S. Daskin and L. K. Dean, "Location of health care facilities", in *Operations research and health care*, Springer, 2005, pp. 43–76.

[77] L. Brotcorne, G. Laporte, and F. Semet, "Ambulance location and relocation models", *European journal of operational research*, vol. 147, no. 3, pp. 451–463, 2003.

[78] S. V. Begur, D. M. Miller, and J. R. Weaver, "An integrated spatial dss for scheduling and routing home-health-care nurses", *Interfaces*, vol. 27, no. 4, pp. 35–48, 1997.

[79] A. Ekici, P. Keskinocak, and J. L. Swann, "Modeling influenza pandemic and planning food distribution", *Manufacturing & Service Operations Management*, vol. 16, no. 1, pp. 11–27, 2013.

[80] F. Wang, "Measurement, optimization, and impact of health care accessibility: A methodological review", *Annals of the Association of American Geographers*, vol. 102, no. 5, pp. 1104–1112, 2012.

[81] N. R. Council *et al.*, *Assessing the reliability of complex models: mathematical and statistical foundations of verification, validation, and uncertainty quantification*. National Academies Press, 2012.

[82] S. Cao, M. Gentili, P. M. Griffin, S. O. Griffin, and N. Serban, "Disparities in preventive dental care among children in georgia.", *Preventing chronic disease*, vol. 14, E104–E104, 2017.

[83] M. Gentili, P. Harati, N. Serban, J. O'connor, and J. Swann, "Quantifying disparities in accessibility and availability of pediatric primary care across multiple states with implications for targeted interventions", *Health services research*, vol. 53, no. 3, pp. 1458–1477, 2018.

[84] N. Serban and S. L. Tomar, "Ada health policy institute's methodology overestimates spatial access to dental care for publicly insured children", *Journal of public health dentistry*, vol. 78, no. 4, pp. 291–295, 2018.

[85] U. G. A. Office, *Efforts under way to improve childrens access to dental services, but sustained attention needed to address ongoing concerns*, 2010.

[86] T. Nguyen, "Risk factors for treatment of gum disease among asian and hispanic adult groups", 2018.

[87] I. of Medicine, *Advancing Oral Health in America*. Washington, DC: The National Academies Press, 2011, ISBN: 978-0-309-18630-8.

[88] C. for Medicare, M. S. (CMS), *et al.*, *Improving access to and utilization of oral health services for children in medicaid and chip programs: Cms oral health strategy*, 2011.

[89] U. S.G. A. Office, *Oral Health: Efforts Under Way to Improve Children's Access to Dental Services, But Sustained Attention Needed to Address Ongoing Concerns: Report to Congressional Committees*. US Government Accountability Office, 2010.

[90] Y. Zheng, I. Lee, and N. Serban, "Regularized optimization with spatial coupling for robust decision making", *European Journal of Operational Research*, vol. 270, no. 3, pp. 898–906, 2018.

[91] J. L. H. Stamm, N. Serban, J. Swann, and P. Wortley, "Quantifying and explaining accessibility with application to the 2009 h1n1 vaccination campaign", *Health care management science*, vol. 20, no. 1, pp. 76–93, 2017.

[92] N. Serban, C. Bush, and S. L. Tomar, "Medicaid caseload for pediatric oral health care", *The Journal of the American Dental Association*, vol. 150, no. 4, pp. 294–304, 2019.

[93] T. Krivobokova, T. Kneib, and G. Claeskens, "Simultaneous confidence bands for penalized spline estimators", *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 852–863, 2010.

[94] E. B. Erhardt, B. Nandram, and J. W. Choi, "Bayesian simultaneous intervals for small areas: An application to variation in maps", *International Journal of Statistics and Probability*, vol. 1, no. 2, p. 229, 2012.

[95] *39 percent of u.s. dentists participate in medicaid or chip for child dental services*.

[96] *Professionally active dentists*.

[97] "Nadp provider networks and accessibility requirements by state", Tech. Rep., 2017.

[98] H. Sarper, "Monte carlo simulation for analysis of the optimum value distribution in stochastic mathematical programs", *Mathematics and Computers in Simulation*, vol. 35, no. 6, pp. 469 –480, 1993.

[99] A. D. Association *et al.*, "Characteristics of private dental practices: Selected 2017 results from the survey of dental practice (tables in excel) chicago", *IL: Health Policy Institute*, 2017.

[100] B. Sen, J. Blackburn, M. A. Morrisey, M. L. Kilgore, D. J. Becker, C. Caldwell, and N. Menachemi, "Effectiveness of preventive dental visits in reducing nonpreventive dental visits and expenditures", *Pediatrics*, vol. 131, no. 6, pp. 1107–1113, 2013.

[101] I. Lee, S. Monahan, N. Serban, P. M. Griffin, and S. L. Tomar, "Estimating the cost savings of preventive dental services delivered to medicaid-enrolled children in six southeastern states", *Health services research*, vol. 53, no. 5, pp. 3592–3616, 2018.

[102] CMS, *National provider identifier standard (npi)*, Accessed: 2016-05-05, 2015.

[103] OECD, *Organization for economic and co-operative development, health care resources: Physicians by age and gender*, Accessed: 2016-05-05, 2016.

[104] HRSA, *Projecting the supply and demand for primary care practitioners through 2020*, 2013.

[105] *Texas am geocoding services*.

[106] *American community survey, 2017 american community survey 5-year estimates, table b17024 [database on the internet]*.

[107] *Colorado and the acas medicaid expansion*.

[108] *Using american community survey estimates and margins of error webinar*, Webinar, 2017.