UNCERTAINTY QUANTIFICATION FOR MULTIOBJECTIVE STOCHASTIC CONVEX QUADRATIC PROGRAMS

Ziyu Liu[1], Susan R. Hunter[1], Nan Kong[2], and Raghu Pasupathy[3]

[1] *School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA*
[2] *Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN 47907*
[3] *Department of Statistics, Purdue University, West Lafayette, IN 47907*

Abstract

A multiobjective stochastic convex quadratic program (MOSCQP) is a multiobjective optimization problem with convex quadratic objectives that are observed with stochastic error. MOSCQP is a useful problem formulation arising, for example, in model calibration and nonlinear system identification when a single regression model combines data from multiple distinct sources, resulting in a multiobjective least squares problem. We consider structured uncertainty quantification for MOSCQPs, which includes the questions of estimating the efficient and Pareto sets, inference through central limit theorems (CLTs), and constructing asymptotically exact confidence regions on the efficient and Pareto sets. We use parameterization to first write the efficient and Pareto set estimators in closed form, then expand the closed-form expression in a matrix geometric series resulting in a key lemma characterizing the Fréchet derivatives of the efficient and Pareto sets. The key lemma enables a delta theorem analogue for MOSCQPs, resulting in structured uniform CLTs on the estimated efficient and Pareto sets. Finally, we formulate a direct procedure for constructing asymptotically valid confidence regions that retain the efficient and Pareto set shapes endowed by the MOSCQP problem structure. We illustrate the confidence regions through a numerical example.

1 Introduction

We consider the context of uncertainty quantification for multiobjective stochastic (strictly) convex quadratic programs (MOSCQPs). For d objectives, we pose an MOSCQP as

$$\begin{aligned} &\text{minimize} \quad \left[f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_d(x) \end{pmatrix} = \begin{pmatrix} x^\intercal \,\mathrm{E}[H_1(\xi)]x - 2\,\mathrm{E}[V_1(\xi)]^\intercal x + \mathrm{E}[C_1(\xi)] \\ \vdots \\ x^\intercal \,\mathrm{E}[H_d(\xi)]x - 2\,\mathrm{E}[V_d(\xi)]^\intercal x + \mathrm{E}[C_d(\xi)] \end{pmatrix} \right] \\ &\quad\text{s.t.} \;\; x \in \mathbb{R}^{q\times 1}, \end{aligned} \tag{1}$$

where $f\colon \mathbb{R}^q \to \mathbb{R}^d$ is a vector-valued objective function composed of d convex quadratic objectives, ξ is a random object, and for each objective $k \in \{1,\ldots,d\}$, $\mathsf{H}_k \coloneqq \mathrm{E}[H_k(\xi)] \in \mathbb{S}_q^{>}$ is twice the Hessian matrix, $\mathbb{S}_q^{>}$ denotes the set of all real-valued $q \times q$ symmetric positive definite matrices, $\mathsf{V}_k \coloneqq \mathrm{E}[H_k(\xi)] \in \mathbb{R}^{q\times 1}$ is a vector, and $\mathsf{C}_k \coloneqq \mathrm{E}[C_k(\xi)] \in \mathbb{R}$ is a constant. Thus, each objective k can be written in terms of the unknown quantities H_k, V_k, and C_k,

$$f_k(x) = x^\intercal \mathsf{H}_k x - 2\mathsf{V}_k^\intercal x + \mathsf{C}_k. \tag{2}$$

Authors' preprint, Tuesday March 12, 2024 at 6:30pm. Corresponding author: susanhunter@purdue.edu.

Funding: S. R. Hunter gratefully acknowledges that this material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-23-1-0488. R. Pasupathy gratefully acknowledges the Office of Naval Research for support provided by the grants N000141712295 and 13000991.

The solution to (1) is the efficient set. The efficient set and its image, the Pareto set, are defined as

$$\mathcal{E} := \{x^* \in \mathbb{R}^q : \nexists x \in \mathbb{R}^q \text{ such that } f(x) \leqslant f(x^*)\}, \quad \mathcal{P} := \{f(x^*) : x^* \in \mathcal{E}\}, \tag{3}$$

respectively. Here and henceforth, when comparing two vectors $y, y' \in \mathbb{R}^d$, we adopt the convention that $y \leqslant y'$ indicates $y_k \leqslant y'_k$ for all $k \in \{1, \ldots, d\}$ and $y \neq y'$. We use $y \leqq y'$ when equality is allowed; see Subsection 1.5 for notation.

In (2), $\mathsf{H}_k, \mathsf{V}_k$, and C_k are unknown and can only be estimated using data for each objective k. Thus, the efficient set $\mathcal{E}$ and the Pareto set $\mathcal{P}$ are also unknown and can only be estimated. While the MOSCQP in (1) may arise in a variety of contexts, to fix ideas, next, we consider an example that arises in model calibration.

1.1 Least squares regression with multiple data sources

Consider the problem of least squares regression using data from multiple heterogeneous data sources. For example, suppose we have input-output data from a black-box system and input-output data that constitutes auxiliary information to this system [12]. We wish to fit one model that accommodates the information from both data sources. We reformulate the MOSCQP from (1) for multiobjective least-squares as follows. Consider a hypothetical system whose behavior can be modeled using data sourced from d existing systems. For data source $k \in \{1, \ldots, d\}$, the linear model

$$R_k = Q_k x_k^* + \varepsilon_k$$

applies, where $R_k \in \mathbb{R}$ is the response random variable, $Q_k = (Q_{k1}, \ldots, Q_{kq}) \in \mathbb{R}^{1\times q}$ is the random vector of covariates, $\varepsilon_k \in \mathbb{R}$ is a random error term, and $x_k^* \in \mathbb{R}^{q\times 1}$ is the parameter value such that $\mathrm{E}[\varepsilon_k \,|\, x_k^*] = 0$. Let the data sources be distinct, in the sense that $x_k^* \neq x_{k'}^*$ for all $k, k' \in \{1, \ldots, d\}$. Suppose further that the decision-maker does not have a priori knowledge on how best to combine the data sources into a single model. Therefore, calibrating the model is a multiobjective least squares problem,

$$\text{minimize} \quad \left[f(x) = \left(\mathrm{E}[(Q_1 x - R_1)^2], \ldots, \mathrm{E}[(Q_d x - R_d)^2]\right)^\intercal\right] \quad \text{s.t. } x \in \mathbb{R}^{q\times 1} \tag{4}$$

where for each data source k, twice the Hessian matrix is $\mathsf{H}_k = \mathrm{E}[Q_k^\intercal Q_k] \in \mathbb{S}_q^{>}$, the linear term has coefficient $\mathsf{V}_k = \mathrm{E}[Q_k^\intercal R_k] \in \mathbb{R}^{q\times 1}$, the constant is $\mathsf{C}_k = \mathrm{E}[R_k^2] \in \mathbb{R}$, and here, $\xi_k := (Q_k, R_k) = (Q_{k1}, \ldots, Q_{kq}, R_k) \in \mathbb{R}^{1\times(q+1)}$ is a condensed representation of the random variables corresponding to the kth data source. As in (1), the probability measures are unknown, and we cannot evaluate the expected values in (4) directly.

The goal of estimating the solution to (4) is to estimate the parameter values in the decision space that produce the global Pareto set in the objective space. While we could select one data source to use as the objective while posing the other $(d-1)$ data sources as constraints [10], solving the constrained least squares problem estimates only *one* of the points in the globally efficient set $\mathcal{E}$, along with its estimated image in the Pareto set $\mathcal{P}$ [11, 33]. Absent a priori knowledge of how a constrained least squares problem should be formulated, posing the problem with all data sources in the objectives provides the modeler with a broader perspective on the entire set of efficient parameter values when selecting a final model. In the context of nonlinear system identification, [24, 25, 34, 35] discuss the benefits of posing such problems as multiobjective problems; also see [2]. The general multiobjective least squares problem may also arise in the context of a mixed econometric-multiobjective optimization procedure, as suggested by [30].

Despite the potential benefits to the modeler of posing and solving the multiobjective least squares problem in (4), we know of no method for quantifying the error in the efficient and Pareto set estimators, or for constructing confidence regions around the true efficient and Pareto sets, $\mathcal{E}$ and $\mathcal{P}$, that both exploits and preserves their special structure. Answering such questions of uncertainty quantification for MOSCQP can allow the modeler to see how the statistical error *jointly* affects the estimated parameter values before deciding on a final model. Such insights are required to ensure the estimated solution to (4) is evaluated with appropriate and meaningful context, as shown in the numerical example in Section 5.

1.2 Questions answered and overview of main results

The uncertainty quantification questions we consider form the trifecta of classical statistical inference for MOSCQP. That is, we seek answers to the questions of how to

(a) construct estimators $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$ of the efficient and Pareto sets $\mathcal{E}$ and $\mathcal{P}$ in (3), respectively;
(b) characterize the uncertainty in the sets $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$ in relation to the corresponding true counterparts $\mathcal{E}$ and $\mathcal{P}$, respectively, through uniform CLTs; and
(c) construct $(1-\alpha)$ confidence regions on $\mathcal{E}$ and $\mathcal{P}$, that is, identify random classes of sets $\mathcal{X}_{n,1-\alpha}$ and $\mathcal{Y}_{n,1-\alpha}$ such that

$$\lim_{n\to\infty} P\{\mathcal{E} \in \mathcal{X}_{n,1-\alpha}\} = 1-\alpha \text{ and } \lim_{n\to\infty} P\{\mathcal{P} \in \mathcal{Y}_{n,1-\alpha}\} = 1-\alpha.$$

We begin our technical treatment in Section 2 by constructing closed-form estimators for the efficient and Pareto sets. Importantly, we write the closed-form estimators as *deterministic functions* of the estimated parameters of the quadratic objectives. That is, first, we define deterministic maps z_s and g_s which, given an index $s \in \mathbb{R}^d$ arising from the linear weighted sum scalarization [17, 21, 22, 36, 38, 41], take in a symmetric positive definite matrix in $\mathbb{S}_q^{-1}$, a vector in $\mathbb{R}^q$, and a constant, and output a point in the corresponding efficient or Pareto set, respectively. Thus, given sample-path estimators of the quadratic parameters $(\bar{\mathsf{H}}_k, \bar{\mathsf{V}}_k, \bar{\mathsf{C}}_k)$, $k = 1, \dots, d$, with appropriate structure, the estimated efficient and Pareto sets $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$ are random fields expressed as deterministic functions of the estimated parameters and indexed by $s \in \mathbb{R}^d$.

The structure of the efficient and Pareto maps z_s and g_s allow "linearization" through matrix geometric series expansion, leading to the key step of constructing Fréchet derivatives of $\mathcal{E}$ and $\mathcal{P}$ with respect to the quadratic objective parameters $(\mathsf{H}_k, \mathsf{V}_k, \mathsf{C}_k), k = 1, \dots, d$ (Lemma 3). Apart from being important in their own right, the derived Fréchet derivatives, when combined with a joint CLT (Theorem 2) on the estimated parameters of the quadratic objectives $(\bar{\mathsf{H}}_k, \bar{\mathsf{V}}_k, \bar{\mathsf{C}}_k)$, $k = 1, \dots, d$, lead to uncertainty quantification on the estimators $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$ through pointwise and uniform CLTs (Theorems 3 and 4). Uniform CLTs (as opposed to pointwise CLTs) are useful because they provide a more holistic picture of the simultaneous sampling variability of the elements of the set estimators $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$. Theorems 3 and 4 are also correctly seen as delta method [28, Theorem 2.5.2] analogues for MOSCQPs. We do not explicitly treat the questions of variance and bias expansions of $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$, the other classical questions that arise in uncertainty quantification. The method of proof leading to the uniform CLTs, particularly the linearization technique, suggests how variance and bias expansions of $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$ can be obtained.

The uniform CLTs lead naturally to the construction of $(1-\alpha)$ confidence regions on $\mathcal{E}$ and $\mathcal{P}$. However, relying solely on the uniform CLTs implies constructing the confidence regions in infinite-dimensional function space, in part relinquishing the MOSCQP structure. To retain the structure inherent to MOSCQP, we propose a different route which recognizes that all uncertainty in estimating $\mathcal{E}$ and $\mathcal{P}$ comes from uncertainty in the finite-dimensional quadratic parameters $(\mathsf{H}_k, \mathsf{V}_k, \mathsf{C}_k), k = 1, \ldots, d$. First, we use the joint CLT on the estimated parameters of the quadratic objectives (Theorem 2) to construct an ellipsoidal confidence region on these parameters. Then, we use the closed-form expressions for $\mathcal{E}$ and $\mathcal{P}$ as a function of these parameters to arrive at asymptotically valid confidence regions through a push-forward measure (Theorem 5). Thus, each curve that belongs to the constructed confidence regions on $\mathcal{E}$ and $\mathcal{P}$ has the structure of the efficient set in the decision space or the structure of the Pareto set in the objective space, respectively. We illustrate the projected confidence regions in the decision and objective spaces through a numerical example in Section 5.

1.3 Related literature

A general theory of statistical inference for multiobjective stochastic programs (MOSP) is largely unexplored in the literature. Prior work on convergence and algorithms for MOSP in a sample average approximation (SAA) framework [40] includes [9, 15, 26, 44]. To the best of our knowledge, Vogel [43] is the earliest paper to consider a general theory of confidence regions for MOSP. The regions are non-asymptotic, in the sense that they are valid for each n and do not rely on distributional assumptions. The recent work by Hunter and Pasupathy [19] provides CLTs on the efficient and Pareto sets for strictly convex MOSPs. Since the MOSCQP we consider is a special case of strictly convex MOSP, the results presented in [19] apply to the present context. However, the present paper exploits the known quadratic structure inherent in MOSCQP, which is likely to lead to better statistical inference than more general methods.

Given its usefulness, some statistical inference results for MOSPs with special structure have been developed within specific application contexts. For example, [4] consider transportation asset management, and a large body of literature exists for the bi-objective mean-variance portfolio optimization problem [31]; see, for example, [8, 16, 23]. These methods are tailored to their respective applications.

1.4 Remarks on standing assumptions

Throughout the paper, we make several standing assumptions, which we discuss below:

Strict convexity Without strictly convex quadratic objectives, the solution to the linear weighted sum scalarization of (1), formulated in (8), may not be unique. Further, weakly Pareto points may exist [11]. Relaxing strict convexity requires handling these two possible complications, which we postpone to future work.

No constraints Adding constraints in (1), while useful, significantly complicates the technical aspects of uncertainty quantification. We postpone such generalization.

Dimensionality We assume the decision space dimension is greater than or equal to the number of objectives, $q \geq d$. Otherwise, under strict convexity, the MOSCQP can be solved as a sequence of smaller problems using reduction techniques [45, 11].

Enough data Since we are interested in asymptotically valid confidence regions, we assume that $n_k > q \geq d$ for all data sources k (consistent with the multiobjective least squares context discussed in Subsection 1.1) and that $n = \sum_{k=1}^{d} n_k > dq \geq d^2$ is large enough to ensure the naïve sample-path estimators $\bar{\mathsf{H}}_k$ are symmetric and positive definite for each k almost surely. We conjecture that our results hold for other consistent symmetric positive definite estimators under appropriate regularity conditions (e.g., shrinkage estimators; see [37]).

Independent and identically distributed data with known covariance We assume i.i.d. data, in the form of i.i.d. copies of the random object ξ, throughout the paper. Further, our results and implementation assume the covariance structure is known and does not need to be estimated. Given the general lack of literature on statistical inference for MOSPs in Subsection 1.3, we believe the advances in this paper constitute a reasonable first step toward more comprehensive results. We consider the derivation of a CLT and confidence regions for non-i.i.d. data and unknown covariance structure as important future work.

1.5 Notation, terminology, and useful results

For two vectors $y, y' \in \mathbb{R}^d$, we use $y \leqq y'$ to indicate $y_k \leqslant y'_k$ for all $k \in \{1, \ldots, d\}$. The notation $y \leqslant y'$ implies $y \leqq y'$ and $y \neq y'$. For any two quantities x and y, $x \coloneqq y$ denotes that x is defined as y, and $x \equiv y$ denotes that x is equivalent to y. Let $\|y\|_2$ denote the Euclidean norm in $\mathbb{R}^d$, and let $\mathbb{R}^d_{\geqslant} \coloneqq \{y \in \mathbb{R}^d \colon y \geqslant 0\}$.

Let $\mathbb{I}_q$ be the $q \times q$ identity matrix. Let $\mathbb{S}_q \subset \mathbb{R}^{q\times q}$ be the space of all real-valued $q \times q$ symmetric matrices, and let $\mathbb{S}_q^{-1} \subset \mathbb{S}_q$ be the subspace of all $q \times q$ symmetric invertible matrices. Let $\mathbb{S}_q^{\geqslant} \subset \mathbb{S}_q$ be the set of all symmetric positive semidefinite matrices and $\mathbb{S}_q^{>} \subset \mathbb{S}_q^{-1}$ be the set of all symmetric positive definite matrices. The *Kronecker product* of matrices $\mathsf{A} = (a_{ij}) \in \mathbb{R}^{m\times n}$ and $\mathsf{B} \in \mathbb{R}^{p\times q}$, denoted $\mathsf{A} \otimes \mathsf{B} \in \mathbb{R}^{mp\times nq}$, is the block matrix defined by entries $[a_{ij}\mathsf{B}]$. For a matrix $\mathsf{A} = (a_1, \ldots, a_n) \in \mathbb{R}^{m\times n}$ where each $a_i \in \mathbb{R}^{m\times 1}$ for $i = 1, \ldots, n$, define $\mathrm{vec}(\mathsf{A}) \coloneqq (a_1^\intercal, \ldots, a_n^\intercal)^\intercal \in \mathbb{R}^{mn\times 1}$ as the vector containing the stacked columns of A. Vectorization is a linear operator [27, p. 82] and can be written as

$$v = \mathrm{vec}(\mathsf{A}) = \textstyle\sum_{i=1}^{n} e_{i,n} \otimes \mathsf{A}e_{i,n} \in \mathbb{R}^{mn\times 1} \tag{5}$$

where $e_{i,n} = (0, \ldots, 0, 1, 0, \ldots, 0)^\intercal \in \mathbb{R}^{n\times 1}$ is a n-dimensional vector of zeros with 1 in the ith place. In (5), $\mathsf{A}e_{i,n}$ represents the ith column of A, and the Kronecker product puts it in the correct position within the vectorization. Further, define the operation that takes an $mn \times 1$ vector back to an $m \times n$ matrix as

$$\mathsf{A} = \mathrm{vec}^{-1}_{m\times n}(v) = \textstyle\sum_{i=1}^{n} (e_{i,n}^\intercal \otimes \mathbb{I}_m) v e_{i,n}^\intercal. \tag{6}$$

Let $\mathcal{C}^q(\mathcal{S})$ be the space of continuous $\mathbb{R}^q$-valued functions on the compact set $\mathcal{S} \subset \mathbb{R}^d$ with the uniform topology $\rho(x, y) \coloneqq \sup_{s\in\mathcal{S}} |x(s) - y(s)|$ for $x, y \in \mathcal{C}^q(\mathcal{S})$.

For a random vector $X \in \mathbb{R}^q$ with $\mathrm{E}[X] = \mu \in \mathbb{R}^q$, its q-by-q covariance matrix is $\mathrm{Var}(X) = \mathrm{E}[(X - \mu)(X - \mu)^\intercal]$ and $\varphi_X(t) \coloneqq \mathrm{E}[\exp\{it^\intercal X\}]$ is its characteristic function. Let $X = (X_1, \ldots, X_q)$ and $Y = (Y_1, \ldots, Y_p)$ be random vectors. Then

$$\mathrm{Cov}(X, Y) \coloneqq \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])^\intercal] = \mathrm{E}[XY^\intercal] - \mathrm{E}[X]\,\mathrm{E}[Y]^\intercal \tag{7}$$

is the $q \times p$ cross-covariance matrix of X and Y. That is, the (i,j)th entry of this matrix contains $\mathrm{Cov}(X_i, Y_j)$, where $i \in \{1,\ldots,q\}, j \in \{1,\ldots,p\}$.

The following Lemma 1 is a special case of the Neumann series.

Lemma 1 (Matrix geometric series [18, p. 351]). *Let M be any square q-by-q matrix with eigenvalues $\lambda_i, i = 1,2,\ldots,q$ satisfying $|\lambda_i| < 1$. Then $M^{-1} = \sum_{n=0}^{\infty}(\mathbb{I}_q - M)^n$. Further, if $S_n := \mathbb{I}_q + M + M^2 + \cdots + M^{n-1} = (\mathbb{I}_q - M)^{-1}(\mathbb{I}_q - M^n)$, the sequence $\{S_n, n \geq 1\}$ converges in operator norm [27, p. 92] to $S := (\mathbb{I}_q - M)^{-1}$ as $n \to \infty$.*

Theorem 1 (Cramér-Wold [3, p. 44]). *If every linear combination of the components of a vector $X \in \mathbb{R}^q$ is normally distributed, that is, $\sum_{i=1}^{q} \lambda_i X_i$ is normally distributed for all vectors $\lambda \in \mathbb{R}^q$, then X is multivariate normal.*

2 Estimators for the efficient and Pareto sets

We define estimators for the efficient and Pareto sets through parameterization with the linear weighted sum scalarization. First, we create two maps, an efficient map z_s and a Pareto map g_s that, given a scalarization parameter s and a (possibly estimated) set of quadratic parameter inputs, produce a point in the corresponding (possibly estimated) efficient and Pareto sets, respectively. Thus, the efficient and Pareto sets and their estimators are written as fields indexed by the scalarization parameter s.

2.1 Parameterization with linear weighted sum

Crucial to our approach is the fact that the efficient and Pareto sets and their estimators can be written in closed form as fields indexed by a parameter due to the weighted sum scalarization [11]. Specifically, let $\mathcal{S} := \{s \in \mathbb{R}^d_{\geqslant} \colon \sum_k s_k = 1\}$ be a normalized weight set containing nonnegative weights, where $\mathbb{R}^d_{\geqslant} := \{y \in \mathbb{R}^d \colon y \geqslant 0\}$. Given $s \in \mathcal{S}$, the linear weighted sum scalarization of (1) is

$$\text{minimize} \;\; \left\{ \sum_{k=1}^{d} s_k f_k(x) = x^\intercal \left[\sum_{k=1}^{d} s_k \mathsf{H}_k \right] x - 2 \left[\sum_{k=1}^{d} s_k \mathsf{V}_k^\intercal \right] x + \sum_{k=1}^{d} s_k \mathsf{C}_k \right\}. \tag{8}$$

Since $\mathsf{H}_k \in \mathbb{S}_q^{>}$ for all k and all $s \in \mathcal{S}$, then $\sum_{k=1}^{d} s_k \mathsf{H}_k \in \mathbb{S}_q^{>}$ for all $s \in \mathcal{S}$. Therefore for each $s \in \mathcal{S}$, the objective function of (8) is strictly convex with unique minimizer

$$x_s^* = [\textstyle\sum_{k=1}^{d} s_k \mathsf{H}_k]^{-1} [\textstyle\sum_{k=1}^{d} s_k \mathsf{V}_k]. \tag{9}$$

By [33, p. 79], [11, p. 71], it follows that (a) $x_s^* \in \mathcal{E}$ for each $s \in \mathcal{S}$, and (b) since each objective f_k from (1) is convex, if $x' \in \mathcal{E}$, then there exists $s \in \mathcal{S}$ such that $x_s^* = x'$. Therefore, the efficient and Pareto sets can be parameterized as

$$\mathcal{E} = \{x_s^* \colon s \in \mathcal{S}\}, \quad \mathcal{P} = \{f(x_s^*) \colon x_s^* \in \mathcal{E}, s \in \mathcal{S}\}. \tag{10}$$

2.1.1 The efficient map

For each $s \in \mathcal{S}$, view x_s^* from (9) as a function of the unknown parameters of the quadratic objectives, $(\mathsf{H}_k, \mathsf{V}_k, \mathsf{C}_k), k = 1,\ldots,d$. Then, given $s \in \mathcal{S}$, we define a map $z_s \colon \mathcal{H} \to \mathbb{R}^q$, $\mathcal{H} \subset \Theta_d$, that takes in quadratic objective parameters of similar form and outputs a point in the decision space, where Θ_d and $\mathcal{H}$ are appropriate normed spaces. For generality, we define the related normed spaces Θ_1 and Θ_d together, in Definition 1. Then, we define $\mathcal{H}$ as an appropriate subset of Θ_d in Definition 2.

Definition 1 (the normed space Θ_j, $j \in \{1, d\}$)**.** Let the space of real-valued $q \times q$ symmetric matrices $\mathbb{S}_q$ be equipped with the spectral norm, which is the (sub-multiplicative) matrix norm $\|\cdot\|_{\mathbb{S}_q}$ induced by the vector norm $\|\cdot\|_2$. Then define the space Θ_j as $\Theta_j \coloneqq (\mathbb{S}_q \times \mathbb{R}^{q\times 1} \times \mathbb{R}) \times \ldots \times (\mathbb{S}_q \times \mathbb{R}^{q\times 1} \times \mathbb{R})$, which is the space $(\mathbb{S}_q \times \mathbb{R}^{q\times 1} \times \mathbb{R})$ crossed with itself j times. Let the point $\theta \coloneqq (h_1, v_1, c_1, \ldots, h_j, v_j, c_j) \in \Theta_j$ and define $\|\theta\|_{\Theta_j} \coloneqq \sum_{k=1}^{j} \|h_k\|_{\mathbb{S}_q} + \|v_k\|_2 + |c_k|$, where it is straightforward to see $\|\cdot\|_{\Theta_j}$ satisfies all required stipulations for being a norm. For two points in $\Theta_j, j \in \{1, d\}$, addition is understood to be elementwise. For example, given $\theta, \tilde{\theta} \in \Theta_d$, $\theta + \tilde{\theta} = (h_1 + \tilde{h}_1, v_1 + \tilde{v}_1, c_1 + \tilde{c}_1, \ldots, h_d + \tilde{h}_d, v_d + \tilde{v}_d, c_d + \tilde{c}_d)$.

Definition 2 (the domain of z_s, $\mathcal{H} \subset \Theta_d$)**.** Let the domain of the map z_s be $\mathcal{H} \coloneqq \{\theta \in \Theta_d \colon h_1, \ldots, h_d \in \mathbb{S}_q^{-1}\}$, which is the set of all elements of Θ_d such that the matrices $h_1, \ldots, h_d$ are real-valued $q \times q$ symmetric invertible matrices.

The normed space Θ_d requires the matrix coefficients of the quadratic terms $h_1, \ldots, h_d$ to be real-valued square symmetric matrices, while $\mathcal{H}$ requires that $h_1, \ldots, h_d$ are also invertible. Since we use Θ_j for $j = d$ often and for $j = 1$ infrequently, henceforth, we omit the subscript when $j = d$. Thus, Θ always refers to Θ_d.

Given $s \in \mathcal{S}$ and $\theta = (h_1, v_1, c_1, \ldots, h_d, v_d, c_d) \in \Theta$, define the efficient map $z_s \colon \mathcal{H} \to \mathbb{R}^q, \mathcal{H} \subset \Theta$ as

$$z_s(\theta) \coloneqq \left[\textstyle\sum_k s_k h_k\right]^{-1} \left[\textstyle\sum_k s_k v_k\right]. \tag{11}$$

(The efficient map ignores the values of $c_1, \ldots, c_d$. However, we include them in θ because they are relevant to the Pareto map, discussed in the next section.) Given the map z_s, we write the efficient set as follows. Let

$$\theta^* \coloneqq (\mathsf{H}_1, \mathsf{V}_1, \mathsf{C}_1, \ldots, \mathsf{H}_d, \mathsf{V}_d, \mathsf{C}_d) \in \mathcal{H}$$

be the true values of the quadratic parameters. Then, write (9) as

$$z_s(\theta^*) = \left[\textstyle\sum_k s_k \mathsf{H}_k\right]^{-1} \left[\textstyle\sum_k s_k \mathsf{V}_k\right]$$

where $z_s(\theta^*) \in \mathcal{E}$. Then, the efficient set in (10) equals $\mathcal{E} = \{z_s(\theta^*) \colon s \in \mathcal{S}\}$.

2.1.2 The Pareto map

To obtain the Pareto set in (10), we take each point in the efficient set x_s^* and apply the vector-valued function f to it. Given $s \in \mathcal{S}$, the points in the efficient set x_s^* are functions of θ^* only. Further, applying f involves applying a function of θ^* only. We exploit this structure to write the following Pareto map. For each $s \in \mathcal{S}$, define the Pareto map $g_s \colon \mathcal{H} \to \mathbb{R}^d, \mathcal{H} \subset \Theta$ as

$$\begin{aligned} g_s(\theta) &\coloneqq \begin{pmatrix} \sum_k s_k v_k^\intercal [\sum_k s_k h_k]^{-1} h_1 [\sum_k s_k h_k]^{-1} \sum_k s_k v_k - 2 v_1^\intercal [\sum_k s_k h_k]^{-1} \sum_k s_k v_k + c_1 \\ \vdots \\ \sum_k s_k v_k^\intercal [\sum_k s_k h_k]^{-1} h_d [\sum_k s_k h_k]^{-1} \sum_k s_k v_k - 2 v_d^\intercal [\sum_k s_k h_k]^{-1} \sum_k s_k v_k + c_d \end{pmatrix} \\ &= \left(z_s(\theta) h_1 z_s(\theta) - 2 v_1^\intercal z_s(\theta) + c_1, \quad \ldots, \quad z_s(\theta) h_d z_s(\theta) - 2 v_d^\intercal z_s(\theta) + c_d\right)^\intercal. \end{aligned} \tag{12}$$

Then the Pareto set in (10) equals $\mathcal{P} = \{g_s(\theta^*)\colon s \in \mathcal{S}\} = \{f \circ z_s(\theta^*)\colon s \in \mathcal{S}\}$, where $f \circ z_s\colon \mathcal{H} \to \mathbb{R}^d$ denotes the composite function of f with the efficient map.

Remark 1. *It is important to notice that while $g_s(\theta^*) = f \circ z_s(\theta^*)$, equality does not hold for general θ. That is, given $\theta \in \Theta$, $g_s(\theta) \neq f \circ z_s(\theta)$ because f always applies θ^* as the quadratic objective parameters. Thus, we cannot simply work with $f \circ z_s(\theta)$, we must work with $g_s(\theta)$ which applies the same θ in all locations.*

2.2 Estimators for the quadratic parameters

The maps z_s and g_s defined in the previous section make clear that all of the error in estimating the efficient and Pareto sets is due solely to the error in estimating the true quadratic parameters $\theta^* = (\mathsf{H}_1, \mathsf{V}_1, \mathsf{C}_1, \ldots, \mathsf{H}_d, \mathsf{V}_d, \mathsf{C}_d)$. Thus, we need only estimate θ^* and use these estimators in the previously defined efficient and Pareto maps. Toward estimating θ^*, suppose we have access to data in the form of i.i.d. copies of the random object $\xi = (\xi_1, \ldots, \xi_d)$, denoted $(\xi_1^{(\ell)}, \ldots, \xi_d^{(\ell)})$ for $\ell = 1, 2, \ldots$. In what follows, we define both general estimators for the context of (1) and specific estimators for use in the multiobjective least squares context of (4).

2.2.1 General estimators

In the general context of (1), for each objective $k = 1, \ldots, d$, we observe values of the random object $\big(H_k(\xi_k^{(\ell)}), V_k(\xi_k^{(\ell)}), C_k(\xi_k^{(\ell)})\big)$ for $\ell = 1, \ldots, n_k$, $n_k > q$ and construct sample-path estimators

$$\bar{\mathsf{H}}_k \coloneqq \tfrac{1}{n_k} \textstyle\sum_{\ell=1}^{n_k} H_k(\xi_k^{(\ell)}), \quad \bar{\mathsf{V}}_k \coloneqq \tfrac{1}{n_k} \textstyle\sum_{\ell=1}^{n_k} V_k(\xi_k^{(\ell)}), \quad \bar{\mathsf{C}}_k \coloneqq \tfrac{1}{n_k} \textstyle\sum_{\ell=1}^{n_k} C_k(\xi_k^{(\ell)}), \tag{13}$$

where $\bar{\mathsf{H}}_k \in \mathbb{S}_q^{>}$ is a real-valued $q \times q$ symmetric positive definite matrix (see Subsection 1.4). Using the notation from (13), denote the estimator for θ^* as

$$\hat{\theta}_n^* \coloneqq (\bar{\mathsf{H}}_1, \bar{\mathsf{V}}_1, \bar{\mathsf{C}}_1, \ldots, \bar{\mathsf{H}}_d, \bar{\mathsf{V}}_d, \bar{\mathsf{C}}_d), \tag{14}$$

where $n \coloneqq \sum_k n_k$ is the total amount of data across all objectives.

2.2.2 Multiobjective least squares estimators

In the context of the multiobjective least squares problem in (4), recall that for each data source $k = 1, \ldots, d$, the random vector of covariates is $Q_k \in \mathbb{R}^{1 \times q}$ and the response random variable is $R_k \in \mathbb{R}$. Linking the notation of (1) with (4), for each data source $k = 1, \ldots, d$,

$$H_k(\xi) = Q_k^\intercal Q_k, \quad V_k(\xi) = Q_k^\intercal R_k, \quad C_k(\xi) = R_k^2. \tag{15}$$

Suppose that for each data source $k = 1, \ldots, d$, we have $n_k > q$ i.i.d. observations of the data vector ξ_k, denoted $\xi_k^{(\ell)} = (Q_k^{(\ell)}, R_k^{(\ell)}) = (Q_{k1}^{(\ell)}, \ldots, Q_{kq}^{(\ell)}, R_k^{(\ell)})$, $\ell = 1, \ldots, n_k$, where $Q_k^{(\ell)} \in \mathbb{R}^{1 \times q}$ is the ℓth random vector of covariates and $R_k^{(\ell)} \in \mathbb{R}$ is the ℓth response random variable. Stack the covariates into the $n_k \times q$ matrix Q_k and the responses into the $n_k \times 1$ matrix R_k. Then the estimators in (13) equal

$$\bar{\mathsf{H}}_k = n_k^{-1}(\mathsf{Q}_k^\intercal \mathsf{Q}_k), \quad \bar{\mathsf{V}}_k = n_k^{-1}(\mathsf{Q}_k^\intercal \mathsf{R}_k), \quad \bar{\mathsf{C}}_k = n_k^{-1}(\mathsf{R}_k^\intercal \mathsf{R}_k). \tag{16}$$

By construction, $\bar{\mathsf{H}}_k$ is at least positive semidefinite for each n_k; by assumption (see Subsection 1.4), it is positive definite. Then, the estimator $\hat{\theta}_n^*$ is identical to (14), except that the individual estimators are calculated using (16).

2.3 Parameterized efficient and Pareto set estimators

Since $\bar{\mathsf{H}}_k \in \mathbb{S}_q^>$ for each k by assumption, we have $\hat{\theta}_n^* \in \mathcal{H}$. Then, we can write efficient and Pareto set estimators, respectively, as

$$\hat{\mathcal{E}}_n = \{z_s(\hat{\theta}_n^*) \colon s \in \mathcal{S}\}, \quad \hat{\mathcal{P}}_n = \{g_s(\hat{\theta}_n^*) \colon s \in \mathcal{S}\}. \tag{17}$$

Remark 2. *The parameterized estimators defined in* (17) *correspond to the SAA estimators for the implicitly-defined sample-path problem. Specifically, using the estimators in* (13)*, the objective function estimators are*

$$\bar{F}_k(x, n) = x^\intercal \bar{\mathsf{H}}_k x - 2\bar{\mathsf{V}}_k^\intercal x + \bar{\mathsf{C}}_k, \quad k = 1, \ldots, d,$$

and the sample-path version of (1) *is*

$$\textit{minimize} \quad \left[\bar{F}_n(x) \coloneqq \left(\bar{F}_1(x, n), \ldots, \bar{F}_d(x, n)\right)^\intercal\right] \quad \textit{s.t.} \quad x \in \mathbb{R}^{q\times 1}, \tag{18}$$

where $\bar{F}_n \colon \mathbb{R}^q \to \mathbb{R}^d$ *is the sample-path function. Then, repeat the analysis in Subsection 2.1 for the sample-path problem in* (18) *to arrive at the same estimators as in* (17)*. Thus,* $\hat{\mathcal{P}}_n = \{g_s(\hat{\theta}_n^*) \colon s \in \mathcal{S}\} = \{\bar{F}_n \circ z_s(\hat{\theta}_n^*) \colon s \in \mathcal{S}\}$*, where* $\bar{F}_n \circ z_s \colon \mathcal{H} \to \mathbb{R}^d$ *is the composite function of the objective estimator from* (18) *with the map* z_s*.*

3 Central Limit Theorems

We derive a uniform CLT on the estimators $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$ from Section 2. We proceed in steps, as follows. First, we derive a joint CLT on the estimated parameters of the quadratic objectives $\hat{\theta}_n^*$ in Subsection 3.1. In Subsection 3.2, we provide expressions for the Fréchet derivatives of the efficient and Pareto maps and use a delta method [6, 28] to obtain a pointwise CLT on the estimators $\hat{\mathcal{E}}_n$ and $\hat{\mathcal{P}}_n$, which is pointwise in the scalarization parameter $s \in \mathcal{S}$. In Subsection 3.3, we derive a corresponding uniform CLT which is uniform in the scalarization parameter $s \in \mathcal{S}$. For convenience, we refer to Θ as the *native space*, $\mathbb{R}^q$ as the *decision space*, and $\mathbb{R}^d$ as the *objective space*.

3.1 Joint CLT on the estimated quadratic parameters

To begin, we derive a joint CLT on the estimated quadratic parameters $\hat{\theta}_n^*$ from (14), which specifies the estimated inputs to the efficient and Pareto maps.

To derive the joint CLT on $\hat{\theta}_n^* \in \mathcal{H} \subset \Theta$ in the native space, we first transform the elements of $\hat{\theta}_n^*$ into a vector, manipulate the vector, and then transform the vector back into an element of Θ. To perform these operations, we define relevant versions of the vectorization operation defined in Subsection 1.5 and its inverse for elements of Θ, as follows. Since we do not require the symmetric matrices in $\hat{\theta}_n^*$ to be invertible during these operations, in this section, we consider $\hat{\theta}_n^*$ as an element of Θ.

Definition 3 (vectorization and its inverse for $\theta \in \Theta_j$, $j \in \{1, d\}$)**.** Let

$$\theta = (h_1, v_1, c_1, \ldots, h_j, v_j, c_j) \in \Theta_j.$$

For each $k \in \{1,\ldots,j\}$, use (5) to define $a_k := \big(\mathrm{vec}(h_k)^\intercal, v_k^\intercal, c_k\big)^\intercal \in \mathbb{R}^{(q^2+q+1)\times 1}$. Let $\mathsf{A} = [a_1 \, \ldots \, a_j] \in \mathbb{R}^{(q^2+q+1)\times j}$ be the concatenated matrix where the kth column equals a_k. Then the vectorization of θ is

$$\tilde{v} = \mathrm{vec}_{\Theta_j}(\theta) := \mathrm{vec}([a_1 \, \ldots \, a_j]) \in \mathbb{R}^{j(q^2+q+1)\times 1}. \tag{19}$$

For some $\tilde{v} \in \mathbb{R}^{j(q^2+q+1)\times 1}$ with appropriate structure, such as $\tilde{v}$ obtained in (19), define the inverse vectorization as follows. Using equation (6), $\mathrm{vec}^{-1}_{(q^2+q+1)\times j}(\tilde{v}) = \mathsf{A} = [a_1 \ldots a_j]$ is the concatenated matrix. Let $\mathsf{B}_h := [\mathbb{I}_{q^2} \;\; 0_{q^2\times(q+1)}]$ be a $q^2 \times (q^2+q+1)$ matrix, and let $\mathsf{B}_v := [0_{q\times q^2} \;\; \mathbb{I}_q \;\; 0_{q\times 1}]$ be a $q \times (q^2+q+1)$ matrix. Then

$$\begin{aligned}\mathrm{vec}^{-1}_{\Theta_j}(\tilde{v}) := \big(&\mathrm{vec}^{-1}_{q\times q}(\mathsf{B}_h\mathsf{A}e_{1,j}),\; \mathsf{B}_v\mathsf{A}e_{1,j},\; e^\intercal_{q^2+q+1,q^2+q+1}\mathsf{A}e_{1,j},\; \ldots,\\ &\mathrm{vec}^{-1}_{q\times q}(\mathsf{B}_h\mathsf{A}e_{j,j}),\; \mathsf{B}_v\mathsf{A}e_{j,j},\; e^\intercal_{q^2+q+1,q^2+q+1}\mathsf{A}e_{j,j}\big).\end{aligned} \tag{20}$$

Equipped with the vectorization operation and its inverse from (19) and (20), respectively, we obtain a CLT on the random vector $\mathrm{vec}_\Theta(\hat{\theta}^*_n)$ in Lemma 2. Then, to obtain a CLT in the native space Θ in Theorem 2, we exploit the fact that vec^{-1}_Θ from (20) is a linear operator.

Before we state the results, we require additional notation which we use to characterize the covariance of the random vector $\mathrm{vec}_\Theta(\hat{\theta}^*_n)$. First, for the kth objective, consider the vectorized version of $(H_k(\xi), V_k(\xi), C_k(\xi)) \in \Theta_1$ (see Definition 1), which we denote as $\mathrm{vec}_{\Theta_1}(H_k(\xi), V_k(\xi), C_k(\xi))$. Define the covariance matrix as

$$\begin{aligned}\Omega_k &:= \mathrm{Var}\left(\mathrm{vec}_{\Theta_1}(H_k(\xi), V_k(\xi), C_k(\xi))\right)\\ &= \mathrm{Var}\left(\mathrm{vec}_{\Theta_1}(Q_k^\intercal Q_k, Q_k^\intercal R_k, R_k^2)\right) \in \mathbb{S}^{\geqslant}_{q^2+q+1},\end{aligned} \tag{21}$$

where the second line provides the equivalent expression under the multiobjective least squares example from Subsection 1.1. Thus, in multiobjective least squares, the Ω_k matrix involves the fourth-order moments of the random variables. Due to the symmetry of the matrices involved, each Ω_k matrix is only positive semidefinite and not positive definite. This structure becomes clearer in the illustrative example from Section 5, where the repeated rows and columns in the Ω_k matrices are explicitly stated. Then across the objectives, the random vector $\mathrm{vec}_\Theta\big(H_1(\xi), V_1(\xi), C_1(\xi), \ldots, H_d(\xi), V_d(\xi), C_d(\xi)\big)$ has mean $\mathrm{vec}_\Theta(\theta^*)$ and covariance matrix

$$\Omega := \begin{bmatrix}\Omega_1 & \ldots & \Omega_{1d}\\ \vdots & \ddots & \vdots\\ \Omega_{d1} & \ldots & \Omega_d\end{bmatrix} \in \mathbb{S}^{\geqslant}_{d(q^2+q+1)}, \tag{22}$$

where for $k, k' \in \{1,\ldots,d\}$, $\Omega_{k',k} = \Omega_{k,k'}$ and

$$\Omega_{k,k'} := \mathrm{Cov}\big(\mathrm{vec}_{\Theta_1}(H_k(\xi), V_k(\xi), C_k(\xi)), \mathrm{vec}_{\Theta_1}(H_{k'}(\xi), V_{k'}(\xi), C_{k'}(\xi))\big) \tag{23}$$

are the cross-covariance matrices defined in (7).

With this notation, in Lemma 2, we provide a CLT on $\mathrm{vec}_\Theta(\hat{\theta}^*_n) \in \mathbb{R}^{d(q^2+q+1)}$.

Lemma 2. *Recall that $n := \sum_k n_k$ is the total amount of data, and let the proportional amount of data for each objective be $\gamma = (\gamma_1, \ldots, \gamma_d)$ where $\gamma_k := n_k/n > 0$, $\sum_k \gamma_k = 1$. Without loss of generality, let γ be such that $\gamma_1 = n_1/n \leq \ldots \leq \gamma_d = n_d/n$, and recall $n\gamma_1 > q$.*

Let $(\xi_1^{(\ell)}, \dots, \xi_d^{(\ell)})$ *be i.i.d. copies of the random object* $(\xi_1, \dots, \xi_d)$ *for all* $\ell = 1, \dots, n\gamma_d$*, so that even though we do not observe all values,*

$$\mathrm{vec}_\Theta\big(H_1(\xi_1^{(\ell)}), V_1(\xi_1^{(\ell)}), C_1(\xi_1^{(\ell)}), \dots, H_d(\xi_d^{(\ell)}), V_d(\xi_d^{(\ell)}), C_d(\xi_1^{(\ell)})\big)$$

are i.i.d. with mean $\mathrm{vec}_\Theta(\theta^*)$ *and covariance matrix* Ω *defined in* (22)*. Then*

$$\sqrt{n}\left(\mathrm{vec}_\Theta(\hat{\theta}_n^*) - \mathrm{vec}_\Theta(\theta^*)\right) \xrightarrow{d} \sqrt{\Omega_\gamma}\, Z, \tag{24}$$

where $Z \in \mathbb{R}^{d(q^2+q+1)}$ *is a standard multivariate normal random vector, and for* Ω_k *and* $\Omega_{k,k'}$ *defined in* (21) *and* (23)*, respectively, the covariance matrix* Ω_γ *is*

$$\Omega_\gamma := \begin{bmatrix} \Omega_1/\gamma_1 & \Omega_{12}/\gamma_2 & \dots & \Omega_{1d}/\gamma_d \\ \Omega_{12}/\gamma_2 & \Omega_2/\gamma_2 & \dots & \Omega_{2d}/\gamma_d \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{1d}/\gamma_d & \Omega_{2d}/\gamma_d & \dots & \Omega_d/\gamma_d \end{bmatrix} \in \mathbb{S}^{\geqslant}_{d(q^2+q+1)}. \tag{25}$$

Proof. Let $\gamma_0 := 0$ and let $\Omega_0 := 0_{(q^2+q+1)\times(q^2+q+1)}$ be a matrix of zeroes. Define the mean-zero random vector $Y_n := \mathrm{vec}_\Theta(\hat{\theta}_n^*) - \mathrm{vec}_\Theta(\theta^*) = \mathrm{vec}_\Theta(\hat{\theta}_n^* - \theta^*)$. By setting unobserved data equal to zero, write Y_n as the sum

$$\frac{1}{n}\sum_{k=1}^{d}\sum_{\ell=n\gamma_{k-1}+1}^{n\gamma_k} \mathrm{vec}_\Theta\Big(\underbrace{0_{q\times q}, 0_{q\times 1}, 0,}_{k-1 \text{ times}} \dots, \frac{H_k(\xi_k^{(\ell)}) - \mathsf{H}_k}{\gamma_k}, \frac{V_k(\xi_k^{(\ell)}) - \mathsf{V}_k}{\gamma_k}, \frac{C_k(\xi_k^{(\ell)}) - \mathsf{C}_k}{\gamma_k}, \\ \dots, \frac{H_d(\xi_d^{(\ell)}) - \mathsf{H}_d}{\gamma_d}, \frac{V_d(\xi_d^{(\ell)}) - \mathsf{V}_d}{\gamma_d}, \frac{C_d(\xi_d^{(\ell)}) - \mathsf{C}_d}{\gamma_d}\Big)$$

and denote the summand as $Y_k^{(\ell)} \in \mathbb{R}^{d(q^2+q+1)\times 1}$. Since $(\xi_1^{(\ell)}, \dots, \xi_d^{(\ell)})$ are i.i.d. for all $\ell = 1, \dots, n\gamma_d$, then for each k, $Y_k^{(\ell)}$ are i.i.d. for all $\ell = n\gamma_{k-1} + 1, \dots, n\gamma_k$. For each k such that $\gamma_k > \gamma_{k-1}$, denote the index for the first in this i.i.d. sequence, $Y_{k,n\gamma_{k-1}+1}$ for $\ell = n\gamma_{k-1}+1$, as $Y_{k,*}$ for $\ell = *$. The $Y_k^{(\ell)}$'s are also independent across k's. The characteristic function of $\sqrt{n}\, Y_n$ is

$$\begin{aligned}\varphi_{\sqrt{n}\,Y_n}(t) &= \mathrm{E}[\exp\{it^\intercal \sqrt{n}\, Y_n\}] = \textstyle\prod_{k=1}^d \mathrm{E}\left[i(t^\intercal/\sqrt{n})\, Y_{k,*}\right]^{n(\gamma_k-\gamma_{k-1})} \\ &= \prod_{k\in\{k' :\, \gamma_{k'} > \gamma_{k'-1}\}} \mathrm{E}\left[i\frac{(t\sqrt{\gamma_k - \gamma_{k-1}})^\intercal}{\sqrt{n(\gamma_k - \gamma_{k-1})}}\, Y_{k,*}\right]^{n(\gamma_k-\gamma_{k-1})}.\end{aligned} \tag{26}$$

Take a limit on both sides of (26) to obtain

$$
\begin{aligned}
\lim_{n\to\infty} \varphi_{\sqrt{n}\,Y_n}(t) &= \prod_{k\in\{k' :\, \gamma_{k'}>\gamma_{k'-1}\}} \lim_{n\to\infty} \left(\varphi_{Y_{k,*}}\left(\frac{t\sqrt{\gamma_k-\gamma_{k-1}}}{\sqrt{n(\gamma_k-\gamma_{k-1})}}\right)\right)^{n(\gamma_k-\gamma_{k-1})} \\
&= \prod_{k\in\{k' :\, \gamma_{k'}>\gamma_{k'-1}\}} \exp\left\{-(1/2)t^{\intercal}(\gamma_k-\gamma_{k-1})\Sigma_{Y_{k,*}}t\right\} \qquad (27)\\
&= \exp\left\{-(1/2)t^{\intercal}\left(\textstyle\sum_{k=1}^{d}(\gamma_k-\gamma_{k-1})\Sigma_{Y_{k,*}}\right)t\right\}, \qquad (28)
\end{aligned}
$$

where (27) follows from the CLT in [13, p. 26]. Continuing from (28), define $\Omega_\gamma := \sum_{k=1}^{d}(\gamma_k-\gamma_{k-1})\Sigma_{Y_{k,*}}$. The relevant covariance matrices,

$$
\Sigma_{Y_{k,*}} = \mathrm{Var}(Y_{k,*}) = \left[\begin{array}{ccc|ccc}
\Omega_0 & \dots & \Omega_0 & \Omega_0 & \dots & \Omega_0 \\
\vdots & \ddots & \vdots & \vdots & & \vdots \\
\Omega_0 & \dots & \Omega_0 & \Omega_0 & \dots & \Omega_0 \\
\hline
\Omega_0 & \dots & \Omega_0 & \Omega_k/\gamma_k^2 & \dots & \Omega_{kd}/(\gamma_k\gamma_d) \\
\vdots & & \vdots & \vdots & \ddots & \vdots \\
\Omega_0 & \dots & \Omega_0 & \Omega_{kd}/(\gamma_k\gamma_d) & \dots & \Omega_d/\gamma_d^2
\end{array}\right],
$$

are positive semidefinite; there are $k-1$ of the Ω_0's along the diagonal of $\Sigma_{Y_{k,*}}$. Noting that $\Sigma_{Y_{1,*}}$ is positive semidefinite when $k=1$, evaluating the sum yields the expression for Ω_γ in (25), which is positive semidefinite. The result follows by applying Lévy's Continuity Theorem [42, p. 14] to (28). □

Lemma 2 makes statements about the convergence of the vectorized version of $\hat{\theta}_n^*$. Next, we convert this statement to a statement regarding the convergence in distribution of $\hat{\theta}_n^*$ directly, using the inverse vectorization defined in (20) which is a linear map. Before we state the result, we require additional notation. We apply inverse vectorization to the right side of (24) to yield $\tilde{\mathsf{Z}} \in \Theta$, defined by

$$
\tilde{\mathsf{Z}} = (\tilde{\mathsf{Z}}_{h,1}, \tilde{\mathsf{Z}}_{v,1}, \tilde{\mathsf{Z}}_{c,1}, \dots, \tilde{\mathsf{Z}}_{h,d}, \tilde{\mathsf{Z}}_{v,d}, \tilde{\mathsf{Z}}_{c,d}) := \mathrm{vec}_\Theta^{-1}\left(\sqrt{\Omega_\gamma}\,Z\right), \qquad (29)
$$

where for all $k=1,\dots,d$, $\tilde{\mathsf{Z}}_{h,k} \in \mathbb{R}^{q\times q}$, $\tilde{\mathsf{Z}}_{v,k} \in \mathbb{R}^{q}$, $\tilde{\mathsf{Z}}_{c,k} \in \mathbb{R}$. To understand the meaning of (29) more fully, recall that $\sqrt{\Omega_\gamma}$ is symmetric, and let its columns be $w_1, w_2, \dots, w_{d(q^2+q+1)}$. Then write the quantity inside the inverse vectorization as

$$
\sqrt{\Omega_\gamma}\,Z = \left(w_1^{\intercal}Z, \dots, w_{d(q^2+q+1)}^{\intercal}Z\right)^{\intercal},
$$

so that using the inverse vectorization expressions in (6) and (20), for each k,

$$
\begin{aligned}
\tilde{\mathsf{Z}}_{h,k} &= \begin{bmatrix} w_{(k-1)(q^2+q+1)+1}^{\intercal}Z & \dots & w_{(k-1)(q^2+q+1)+q^2-q+1}^{\intercal}Z \\ \vdots & \ddots & \vdots \\ w_{(k-1)(q^2+q+1)+q}^{\intercal}Z & \dots & w_{(k-1)(q^2+q+1)+q^2}^{\intercal}Z \end{bmatrix}, \qquad (30)\\
\tilde{\mathsf{Z}}_{v,k} &= \left[w_{k(q^2+q+1)-(q+1)+1}^{\intercal}Z \dots w_{k(q^2+q+1)-1}^{\intercal}Z\right]^{\intercal}, \quad \tilde{\mathsf{Z}}_{c,k} = w_{k(q^2+q+1)}^{\intercal}Z.
\end{aligned}
$$

Thus, each $\tilde{\mathsf{Z}}_{h,k}$, $\tilde{\mathsf{Z}}_{v,k}$, and $\tilde{\mathsf{Z}}_{c,k}$ is a matrix, a vector, and a scalar, respectively, in which each entry is a scaled version of the same mean-zero normal random variable. Note that if two columns w_i, w_j are such that $w_i = w_j$ for $i \neq j$, the resulting entries are identical. Thus, the structure in Ω_γ preserves the symmetry of the matrices $\tilde{\mathsf{Z}}_{h,k}, k = 1, \ldots, d$. We use this notation to write the CLT in the native space.

Theorem 2 (CLT in the native space Θ)**.** *Let the postulates of Lemma 2 hold. Then for* $\tilde{\mathsf{Z}} \in \Theta$ *defined in* (29),

$$\sqrt{n}\big(\hat{\theta}_n^* - \theta^*\big) \xrightarrow{d} \tilde{\mathsf{Z}}.$$

Proof. Since vec_Θ^{-1} from (20) is a linear operator, the result follows by applying the continuous mapping theorem [7, p. 21] to the result in Lemma 2. That is,

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n^* - \theta^*) &= \sqrt{n}\left(\mathrm{vec}_\Theta^{-1}\big(\mathrm{vec}_\Theta(\hat{\theta}_n^*)\big) - \mathrm{vec}_\Theta^{-1}\left(\mathrm{vec}_\Theta(\theta^*)\right)\right)\\ &= \mathrm{vec}_\Theta^{-1}\left(\sqrt{n}\left(\mathrm{vec}_\Theta(\hat{\theta}_n^*) - \mathrm{vec}_\Theta(\theta^*)\right)\right) \xrightarrow{\mathrm{d}} \mathrm{vec}_\Theta^{-1}(\sqrt{\Omega_\gamma}\, Z),\end{aligned}$$

which implies the result. ☐

3.2 Pointwise CLTs on the estimated efficient and Pareto sets

Given the CLT in the native space, next, we derive pointwise CLTs on the efficient and Pareto sets. First, we present Lemma 3 regarding the Fréchet differentiability of the map z_s and its image [29, p. 172]. This key lemma enables the use of the delta method in deriving the pointwise CLT in this section and, ultimately, the uniform CLT.

Lemma 3 (Fréchet derivatives of the efficient and Pareto maps)**.** *Let* $s \in \mathcal{S}$ *and* $\theta^* \in \mathcal{H}, \mathcal{H} \subset \Theta$ *be given, and recall that for* $\theta \in \mathcal{H}$,

$$\theta - \theta^* = (h_1 - \mathsf{H}_1, v_1 - \mathsf{V}_1, c_1 - \mathsf{C}_1, \ldots, h_d - \mathsf{H}_d, v_d - \mathsf{V}_d, c_d - \mathsf{C}_d) \in \Theta.$$

Then the following hold:

1. *The efficient map* $z_s \colon \mathcal{H} \to \mathbb{R}^q$ *in* (11) *is Fréchet differentiable at* θ^**. That is, there exists a bounded linear operator* $z_s'(\theta^*; \cdot) \colon \Theta \to \mathbb{R}^q$ *with*

$$\left\| z_s(\theta) - \left[z_s(\theta^*) + z_s'(\theta^*; \theta - \theta^*)\right] \right\|_2 = o\big(\, \|\theta - \theta^*\|_\Theta \big).$$

   Further, $z_s'(\theta^*; \theta - \theta^*) = [\sum_k s_k \mathsf{H}_k]^{-1} \left(\sum_k s_k \left[(v_k - \mathsf{V}_k) - (h_k - \mathsf{H}_k) z_s(\theta^*)\right]\right)$.
2. *The Pareto map* $g_s \colon \mathcal{H} \to \mathbb{R}^d$ *in* (12) *is Fréchet differentiable at* θ^**. That is, there exists a bounded linear operator* $g_s'(\theta^*; \cdot) \colon \Theta \to \mathbb{R}^d$ *with*

$$\left\| g_s(\theta) - \left[g_s(\theta^*) + g_s'(\theta^*; \theta - \theta^*)\right] \right\|_2 = o\big(\, \|\theta - \theta^*\|_\Theta \big).$$

   Further, $g_s'(\theta^*; \theta - \theta^*) = (g_{s,1}'(\theta^*, \theta - \theta^*), g_{s,2}'(\theta^*, \theta - \theta^*), \ldots, g_{s,d}'(\theta^*, \theta - \theta^*))^\mathsf{T}$*, where for each* $k \in \{1, \ldots, d\}$,

$$\begin{aligned}g_{s,k}'(\theta^*; \theta - \theta^*) = z_s(\theta^*)^\mathsf{T}(h_k - \mathsf{H}_k) z_s(\theta^*) - 2(v_k - \mathsf{V}_k)^\mathsf{T} z_s(\theta^*) + (c_k - \mathsf{C}_k)\\ +2\left[z_s(\theta^*)^\mathsf{T}\mathsf{H}_k - \mathsf{V}_k^\mathsf{T}\right] z_s'(\theta^*, \theta - \theta^*).\end{aligned}$$

Proof. First, we show Part 1. Recall that z_s is a map from the normed space Θ to $\mathbb{R}^q$. Use the definition of the Gateaux derivative in [29, p. 171] to obtain

$$z_s'\big(\theta^*;\theta-\theta^*\big) \coloneqq \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(z_s\big(\theta^*+\epsilon(\theta-\theta^*)\big)-z_s(\theta^*)\Big). \tag{31}$$

Now, expand $z_s\big(\theta^*+\epsilon(\theta-\theta^*)\big)$ to yield

$$\begin{aligned}
z_s\big(\theta^*+\epsilon(\theta-\theta^*)\big) &= \Big[\sum_k s_k\mathsf{H}_k+\epsilon\sum_k s_k(h_k-\mathsf{H}_k)\Big]^{-1}\Big[\sum_k s_k\mathsf{V}_k+\epsilon\sum_k s_k(v_k-\mathsf{V}_k)\Big]\\
&= \Big[\mathbb{I}_q+\epsilon\Big[\sum_k s_k\mathsf{H}_k\Big]^{-1}\sum_k s_k(h_k-\mathsf{H}_k)\Big]^{-1}\Big[\sum_k s_k\mathsf{H}_k\Big]^{-1}\Big[\sum_k s_k\mathsf{V}_k+\epsilon\sum_k s_k(v_k-\mathsf{V}_k)\Big].
\end{aligned}$$

For notational convenience, define the maps $T\colon\Theta\to\mathbb{S}_q$ and $\widetilde{T}\colon\Theta\to\mathbb{R}^q$ by

$$T(\theta-\theta^*) \coloneqq [\textstyle\sum_k s_k\mathsf{H}_k]^{-1}\sum_k s_k(h_k-\mathsf{H}_k), \tag{32}$$
$$\widetilde{T}(\theta-\theta^*) \coloneqq [\textstyle\sum_k s_k\mathsf{H}_k]^{-1}\sum_k s_k(v_k-\mathsf{V}_k), \tag{33}$$

respectively, and note that in both cases, $[\sum_k s_k\mathsf{H}_k]^{-1}$ is a constant that comes from the fact that we consider the Gateaux derivative at θ^* in (31). Using this notation and continuing from above, it follows that

$$\begin{aligned}
z_s\big(\theta^*+\epsilon(\theta-\theta^*)\big) &= [\mathbb{I}_q+\epsilon T(\theta-\theta^*)]^{-1}[z_s(\theta^*)+\epsilon\widetilde{T}(\theta-\theta^*)]\\
&= \big[\mathbb{I}_q-\epsilon\, T(\theta-\theta^*)+(\epsilon\, T(\theta-\theta^*))^2+o(\epsilon^2)\big]\big[z_s(\theta^*)+\epsilon\widetilde{T}(\theta-\theta^*)\big] && (34)\\
&= z_s(\theta^*)+\epsilon\widetilde{T}(\theta-\theta^*)-\epsilon T(\theta-\theta^*)z_s(\theta^*)-\epsilon^2 T(\theta-\theta^*)\widetilde{T}(\theta-\theta^*) && (35)\\
&\qquad +(\epsilon\, T(\theta-\theta^*))^2 z_s(\theta^*)+\epsilon^3(T(\theta-\theta^*))^2\widetilde{T}(\theta-\theta^*)+o(\epsilon^2),
\end{aligned}$$

where, for small enough ϵ, (34) follows by applying the matrix geometric series from Lemma 1. Using (35) in (31), we have

$$\begin{aligned}
z_s'\big(\theta^*;\theta-\theta^*\big) &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(\epsilon\widetilde{T}(\theta-\theta^*)-\epsilon T(\theta-\theta^*)z_s(\theta^*)+o(\epsilon)\Big)\\
&= \widetilde{T}(\theta-\theta^*)-T(\theta-\theta^*)z_s(\theta^*). && (36)
\end{aligned}$$

Now, to show that $z_s'(\theta^*;\cdot)$ is a bounded linear operator, we first demonstrate that T and $\widetilde{T}$ are bounded linear operators of the increment $\theta-\theta^*$. To see this, note that T and $\widetilde{T}$ obey the conditions for linear operators from [27, p. 82]. Also, since

$$\|\theta-\theta^*\|_\Theta = \textstyle\sum_{k=1}^d\|h_k-\mathsf{H}_k\|_{\mathbb{S}_q}+\|v_k-\mathsf{V}_k\|_2+|c_k-\mathsf{C}_k| = 1$$

implies that $\|h_k-\mathsf{H}_k\|_{\mathbb{S}_q}\le 1$ and $\|v_k-\mathsf{V}_k\|_2\le 1$ for all k, and since $\sum_k s_k=1, s_k\ge 0$ for all k, we have that the operator norm of T [27, p. 92] is

$$\begin{aligned}\|T\|_{op} &= \sup\left\{\left\|[\sum_k s_k \mathsf{H}_k]^{-1}\sum_k s_k(h_k-\mathsf{H}_k)\right\|_{\mathbb{S}_q} : \theta-\theta^*\in\Theta, \|\theta-\theta^*\|_\Theta = 1\right\}\\ &\le \sup\left\{\left\|[\sum_k s_k\mathsf{H}_k]^{-1}\right\|_{\mathbb{S}_q}\sum_k s_k\left\|h_k-\mathsf{H}_k\right\|_{\mathbb{S}_q} : \theta-\theta^*\in\Theta, \|\theta-\theta^*\|_\Theta = 1\right\}\\ &\le \left\|[\sum_k s_k\mathsf{H}_k]^{-1}\right\|_{\mathbb{S}_q}.\end{aligned}$$

Similar logic follows for $\widetilde{T}$. Then since $\left\|[\sum_k s_k\mathsf{H}_k]^{-1}\right\|_{\mathbb{S}_q}$ exists (recall that the inverse of a symmetric matrix is also symmetric), we have

$$\max\{\|T\|_{op}, \|\widetilde{T}\|_{op}\} \le \left\|[\sum_k s_k\mathsf{H}_k]^{-1}\right\|_{\mathbb{S}_q} < \infty. \tag{37}$$

Therefore, both T and $\widetilde{T}$ are bounded linear operators of the increment $\theta-\theta^*$.

Since T and $\widetilde{T}$ are bounded linear operators and (37) holds, (36) implies that $z_s'\big(\theta^*,\cdot\big)$ is also a bounded linear operator with respect to the increment. To demonstrate that $z_s'\big(\theta^*,\cdot\big)$ is a Fréchet derivative satisfying its definition, it is sufficient to show that $z_s(\theta)-z_s(\theta^*) = z_s'(\theta^*;\theta-\theta^*)+o(\|\theta-\theta^*\|_\Theta)$. Use calculations similar to those leading up to (35) to see that for small enough $\theta-\theta^*$,

$$\begin{aligned}z_s(\theta)-z_s(\theta^*) &= z_s(\theta^*+(\theta-\theta^*))-z_s(\theta^*)\\ &= z_s'(\theta^*;\theta-\theta^*) - T(\theta-\theta^*)\widetilde{T}(\theta-\theta^*) + (T(\theta-\theta^*))^2 z_s(\theta^*)\\ &\qquad + (T(\theta-\theta^*))^2\widetilde{T}(\theta-\theta^*) + o\left(\|(T(\theta-\theta^*))^2\|_{\mathbb{S}_q}\right)\\ &= z_s'(\theta^*;\theta-\theta^*) + o\left(\|\theta-\theta^*\|_\Theta\right).\end{aligned}$$

For Part 2, notice again that $g_{s,k}' := \lim_{\epsilon\to 0}\frac{1}{\epsilon}\big(g_{s,k}\big(\theta^*+\epsilon(\theta-\theta^*)\big)-g_{s,k}(\theta^*)\big)$ where $g_{s,k}$ is from (12). The proof follows along lines identical to that of Part 1. □

Now, we are ready to prove pointwise CLTs on the efficient and Pareto set estimators. Specifically, Theorem 2 demonstrates that the Θ-valued sequence $\{\hat{\theta}_n^*\}$, appropriately scaled, converges weakly to the Θ-valued analog of a normal random object, $\tilde{\mathsf{Z}}$. In analogy with the delta method [6, 28], given $s\in\mathcal{S}$, it seems natural that the respective $\mathbb{R}^q$-valued and $\mathbb{R}^d$-valued sequences of estimated efficient and Pareto points $\{z_s(\hat{\theta}_n^*)\}$ and $\{g_s(\hat{\theta}_n^*)\}$, appropriately scaled, should also satisfy a weak convergence theorem. Theorem 3 makes such weak convergence precise.

Theorem 3 (pointwise CLT)**.** *Let the postulates of Lemma 2 hold, and recall that $\tilde{\mathsf{Z}}\in\Theta$ is defined in (29). Suppose $\theta^*\in\mathcal{H}$ and $\hat{\theta}_n^*\in\mathcal{H}$ for each n, where $\mathcal{H}\subset\Theta$. Then for each parameter value $s\in\mathcal{S}$,*

1. $\sqrt{n}\left[z_s(\hat{\theta}_n^*)-z_s(\theta^*)\right]\xrightarrow{d} z_s'(\theta^*,\tilde{\mathsf{Z}})$, *where the right side is a q-dimensional mean-zero multivariate normal random vector*

$$z_s'(\theta^*,\tilde{\mathsf{Z}}) = [\sum_k s_k\mathsf{H}_k]^{-1}\left(\sum_k s_k[\tilde{\mathsf{Z}}_{v,k}-\tilde{\mathsf{Z}}_{h,k}z_s(\theta^*)]\right) \tag{38}$$

with covariance matrix $\mathrm{Var}\big(z_s'(\theta^*,\tilde{\mathsf{Z}})\big) = \mathrm{E}\left[z_s'(\theta^*,\tilde{\mathsf{Z}})\big(z_s'(\theta^*,\tilde{\mathsf{Z}})\big)^\intercal\right]\in\mathbb{S}_q^{\geqslant}$.

2. $\sqrt{n}\left[g_s(\hat{\theta}_n^*)-g_s(\theta^*)\right]\xrightarrow{d} g_s'(\theta^*,\tilde{\mathsf{Z}})$, *where the right side is a d-dimensional multivariate*

normal random vector with components

$$g'_{s,k}(\theta^*;\tilde{\mathsf{Z}}) = z_s(\theta^*)^\intercal \tilde{\mathsf{Z}}_{h,k} z_s(\theta^*) - 2\tilde{\mathsf{Z}}_{v,k}^\intercal z_s(\theta^*) + \tilde{\mathsf{Z}}_{c,k} + 2\big[z_s(\theta^*)^\intercal \mathsf{H}_k - \mathsf{V}_k^\intercal\big] z'_s(\theta^*,\tilde{\mathsf{Z}})$$

and covariance matrix $\mathrm{Var}\big(g'_s(\theta^*,\tilde{\mathsf{Z}})\big) = \mathrm{E}\left[g'_s(\theta^*,\tilde{\mathsf{Z}})\big(g'_s(\theta^*,\tilde{\mathsf{Z}})\big)^\intercal\right] \in \mathbb{S}_d^{\geqslant}$.

Proof. Let $s \in \mathcal{S}$ be given. Using (34), we have

$$\begin{aligned} z_s(\hat{\theta}_n^*) &= z_s(\theta^* + (\hat{\theta}_n^* - \theta^*)) \\ &= \big[\mathbb{I}_q - T(\hat{\theta}_n^* - \theta^*) + (T(\hat{\theta}_n^* - \theta^*))^2 - (T(\hat{\theta}_n^* - \theta^*))^3 + \ldots\big]\big[z_s(\theta^*) + \widetilde{T}(\hat{\theta}_n^* - \theta^*)\big] \\ &= \big[\mathbb{I}_q - T(\hat{\theta}_n^* - \theta^*) + (T(\hat{\theta}_n^* - \theta^*))^2[\mathbb{I}_q + T(\hat{\theta}_n^* - \theta^*)]^{-1}\big]\big[z_s(\theta^*) + \widetilde{T}(\hat{\theta}_n^* - \theta^*)\big] \\ &= z_s(\theta^*) + z'_s(\theta^*;\hat{\theta}_n^* - \theta^*) + U_n, \end{aligned}$$

where T, $\widetilde{T}$ are defined in (32), (33), respectively, and

$$U_n \coloneqq -T(\hat{\theta}_n^* - \theta^*)\widetilde{T}(\hat{\theta}_n^* - \theta^*) + (T(\hat{\theta}_n^* - \theta^*))^2[\mathbb{I}_q + T(\hat{\theta}_n^* - \theta^*)]^{-1}\big[z_s(\theta^*) + \widetilde{T}(\hat{\theta}_n^* - \theta^*)\big].$$

Then it follows that

$$\begin{aligned} \sqrt{n}\big[z_s(\hat{\theta}_n^*) - z_s(\theta^*)\big] &= \sqrt{n}\big[z'_s(\theta^*;\hat{\theta}_n^* - \theta^*) + U_n\big] \\ &= z'_s\big(\theta^*;\sqrt{n}(\hat{\theta}_n^* - \theta^*)\big) + \sqrt{n}\,U_n. \end{aligned} \tag{39}$$

Consider the term $\sqrt{n}\,U_n$ from (39) first. Since T and $\widetilde{T}$ are bounded linear operators, the following upper bounds apply to terms appearing in U_n:

$$\begin{aligned} \|T(\hat{\theta}_n^* - \theta^*)\widetilde{T}(\hat{\theta}_n^* - \theta^*)\| &\le \|T\|_{op}\|\widetilde{T}\|_{op}\|\hat{\theta}_n^* - \theta^*\|_\Theta^2, \\ \|(T(\hat{\theta}_n^* - \theta^*))^2\|_{\mathbb{S}_q} &\le \|T\|_{op}^2\|\hat{\theta}_n^* - \theta^*\|_\Theta^2, \\ \|z_s(\theta^*) + \widetilde{T}(\hat{\theta}_n^* - \theta^*)\| &\le \|z_s(\theta^*)\| + \|\widetilde{T}\|_{op}\|\hat{\theta}_n^* - \theta^*\|_\Theta. \end{aligned} \tag{40}$$

Then using the upper bound in (37), there exists a constant $c > 0$ such that

$$\|\sqrt{n}\,U_n\| \le c\sqrt{n}\,\|\hat{\theta}_n^* - \theta^*\|_\Theta^2\Big(1 + c\big\|[\mathbb{I}_q + T(\hat{\theta}_n^* - \theta^*)]^{-1}\big\|_{\mathbb{S}_q}[1 + \|\hat{\theta}_n^* - \theta^*\|_\Theta]\Big). \tag{41}$$

From Theorem 2, $\tilde{\mathsf{Z}} = \mathrm{vec}_\Theta^{-1}\left(\sqrt{\Omega_\gamma}\,Z\right)$ for standard multivariate normal Z and $\sqrt{n}\big(\hat{\theta}_n^* - \theta^*\big) \xrightarrow{\mathrm{d}} \tilde{\mathsf{Z}}$. Then as $n \to \infty$, we have $\sqrt{n}\,\|\hat{\theta}_n^* - \theta^*\|_\Theta \xrightarrow{\mathrm{d}} \|\tilde{\mathsf{Z}}\|_\Theta$ and

$$\big\|[\mathbb{I}_q + T(\hat{\theta}_n^* - \theta^*)]^{-1}\big\|_{\mathbb{S}_q} \xrightarrow{\mathrm{p}} 1$$

by [39, p. 24], and also $\|\hat{\theta}_n^* - \theta^*\|_\Theta \xrightarrow{\mathrm{p}} 0$ and $\sqrt{n}\,\|\hat{\theta}_n^* - \theta^*\|_\Theta^2 \xrightarrow{\mathrm{p}} 0$. Employing these results and (41) when applying Slutsky's theorem [39, pp. 19] to $\sqrt{n}\,U_n$, we have

$$\sqrt{n}\,U_n \xrightarrow{\mathrm{p}} 0.$$

Finally, apply the theorem from [39, p. 24] and Slutsky's theorem [39, p. 19] to (39).

Similar arguments demonstrate the limit in Part 2; thus, we omit the proof. $\square$

3.3 Uniform CLTs on the estimated efficient and Pareto sets

Theorem 3 asserts convergence in distribution for a fixed $s \in \mathcal{S}$. Next, we demonstrate through Theorem 4 that such weak convergence happens uniformly in $s \in \mathcal{S}$, where the right-side limits are mean-zero Gaussian processes. First, in Lemma 4, we show that the relevant collections of random vectors are, in fact, Gaussian processes. Recall that a Gaussian process is completely determined by its mean and covariance functions [1, p. 11]. Note that just as a finite collection of normal random variables may not necessarily form a multivariate normal random vector, a collection of Gaussian random vectors may not form a Gaussian process in general.

Lemma 4. *Let $\theta^* \in \mathcal{H}, \mathcal{H} \subset \Theta$. The $C^q(\mathcal{S})$-valued, $C^d(\mathcal{S})$-valued random fields*

$$\mathfrak{z}'(\theta^*, \tilde{\mathsf{Z}}) := \{z'_s(\theta^*, \tilde{\mathsf{Z}}), s \in \mathcal{S}\}, \quad \mathfrak{g}'(\theta^*, \tilde{\mathsf{Z}}) := \{g'_s(\theta^*, \tilde{\mathsf{Z}}), s \in \mathcal{S}\}, \tag{42}$$

are mean-zero Gaussian processes with covariance functions

$$K_{\mathfrak{z}'}(s,t) := \mathrm{E}\left[z'_s(\theta^*, \tilde{\mathsf{Z}})\left[z'_t(\theta^*, \tilde{\mathsf{Z}})\right]^\intercal\right], \quad K_{\mathfrak{g}'}(s,t) := \mathrm{E}\left[g'_s(\theta^*, \tilde{\mathsf{Z}})\left[g'_t(\theta^*, \tilde{\mathsf{Z}})\right]^\intercal\right]$$

for all $(s,t) \in \mathcal{S} \times \mathcal{S}$, respectively.

Proof sketch. Recall that $\tilde{\mathsf{Z}}$ is defined in (29) and explicit expressions of $\tilde{\mathsf{Z}}$ as a function of the standard multivariate normal random vector Z appear in (30). For each $s \in \mathcal{S} \subset \mathbb{R}^d$, define the following matrices

$$\begin{aligned}
\mathsf{M}_v(s) &:= [\textstyle\sum_k s_k \mathsf{H}_k]^{-1}[\mathbb{I}_q \otimes s^\intercal] \in \mathbb{R}^{q \times dq}, \\
\mathsf{M}_h(s) &:= [\textstyle\sum_k s_k \mathsf{H}_k]^{-1}[\mathbb{I}_q \otimes z_s(\theta^*)^\intercal][\mathbb{I}_q \otimes s^\intercal] \in \mathbb{R}^{q \times dq^2}.
\end{aligned}$$

Further, define the mean-zero random variables

$$W_v := \mathrm{vec}\begin{pmatrix} \tilde{\mathsf{Z}}_{v,1}^\intercal \\ \vdots \\ \tilde{\mathsf{Z}}_{v,d}^\intercal \end{pmatrix} \in \mathbb{R}^{dq \times 1} \qquad W_h := \mathrm{vec}\begin{pmatrix} \mathrm{vec}(\tilde{\mathsf{Z}}_{h,1}^\intercal)^\intercal \\ \vdots \\ \mathrm{vec}(\tilde{\mathsf{Z}}_{h,d}^\intercal)^\intercal \end{pmatrix} \in \mathbb{R}^{dq^2 \times 1}.$$

Then, $z'_s(\theta^*, \tilde{\mathsf{Z}})$ from (38) equals $z'_s(\theta^*, \tilde{\mathsf{Z}}) = \mathsf{M}_v(s) W_v - \mathsf{M}_h(s) W_h$. Since all transformations from the standard multivariate normal random vector Z to W_v and W_h are linear, it follows that W_v and W_h are jointly multivariate normal. Now apply Cramér-Wold (Theorem 1) to see that $\mathfrak{z}'(\theta^*, \tilde{\mathsf{Z}})$ is a multivariate Gaussian process. A similar result holds for $\mathfrak{g}'(\theta^*, \tilde{\mathsf{Z}})$. □

Now, we employ Lemma 4 toward proving the following Theorem 4, which asserts that the weak convergence happens uniformly in $s \in \mathcal{S}$.

Theorem 4 (uniform CLT)**.** *Let the postulates of Lemma 2 hold, and suppose $\theta^* \in \mathcal{H}$ and $\hat{\theta}_n^* \in \mathcal{H}$ for each n, $\mathcal{H} \subset \Theta$. The $C^q(\mathcal{S})$-valued and $C^d(\mathcal{S})$-valued sequences*

$$\delta_{\mathcal{E},n} := \left\{\sqrt{n}\left[z_s(\hat{\theta}_n^*) - z_s(\theta^*)\right], s \in \mathcal{S}\right\}, \quad \delta_{\mathcal{P},n} := \left\{\sqrt{n}\left[g_s(\hat{\theta}_n^*) - g_s(\theta^*)\right], s \in \mathcal{S}\right\}$$

satisfy $\delta_{\mathcal{E},n} \xrightarrow{d} \mathfrak{z}'(\theta^*, \tilde{\mathsf{Z}})$ and $\delta_{\mathcal{P},n} \xrightarrow{d} \mathfrak{g}'(\theta^*, \tilde{\mathsf{Z}})$, respectively, where the limits $\mathfrak{z}'(\theta^*, \tilde{\mathsf{Z}})$ and $\mathfrak{g}'(\theta^*, \tilde{\mathsf{Z}})$ are mean-zero Gaussian processes identified in Lemma 4.

Proof. From Prokhorov's theorem [7, p. 57ff.], the proof that $\delta_{\mathcal{E},n} \xrightarrow{\mathrm{d}} \mathfrak{z}'(\theta^*, \tilde{\mathsf{Z}})$ will be complete if we demonstrate the following: (a) the finite-dimensional distributions of $\delta_{\mathcal{E},n}$ converge to the finite-dimensional distributions of $\mathfrak{z}'(\theta^*, \tilde{\mathsf{Z}})$; and (b) there is no "escape of probability mass" from the sequence $\{\delta_{\mathcal{E},n}, n \geq 1\}$; that is, the sequence $\{\delta_{\mathcal{E},n}, n \geq 1\}$ is tight.

First, we prove (a). Fix $\tilde{\mathcal{S}} = \{s^{(1)}, \ldots, s^{(r)}\} \subset \mathcal{S}$. To invoke Cramér-Wold (Theorem 1), for $\lambda_j \in \mathbb{R}$, $j = 1, \ldots, r$, consider the weak convergence of the $\mathbb{R}^q$-valued random variable $\sum_{j=1}^r \lambda_j \sqrt{n}\big[z_{s^{(j)}}(\hat{\theta}_n^*) - z_{s^{(j)}}(\theta^*)\big]$. Retracing the proof of Theorem 3,

$$\begin{aligned}\textstyle\sum_{j=1}^r \lambda_j \sqrt{n}\big[z_{s^{(j)}}(\hat{\theta}_n^*) - z_{s^{(j)}}(\theta^*)\big] &= \textstyle\sum_{j=1}^r \lambda_j \sqrt{n}\big[z'_{s^{(j)}}(\theta^*; \hat{\theta}_n^* - \theta^*) + U_{s^{(j)},n}\big] \\ &= \textstyle\sum_{j=1}^r \lambda_j z'_{s^{(j)}}\big(\theta^*; \sqrt{n}(\hat{\theta}_n^* - \theta^*)\big) + \sqrt{n}\, U_{s^{(j)},n}. \end{aligned} \tag{43}$$

Again following the proof of Theorem 3, for each $j \in \{1, \ldots, r\}$, as $n \to \infty$,

$$\sqrt{n}\, U_{s^{(j)},n} \xrightarrow{\mathrm{p}} 0. \tag{44}$$

Since $\sqrt{n}\big(\hat{\theta}_n^* - \theta^*\big) \xrightarrow{\mathrm{d}} \tilde{\mathsf{Z}}$ from Theorem 2, (43) and (44) along with the theorem from [39, p. 24] and Slutsky's theorem [39, p. 19] imply that

$$\textstyle\sum_{j=1}^r \lambda_j \sqrt{n}\big[z_{s^{(j)}}(\hat{\theta}_n^*) - z_{s^{(j)}}(\theta^*)\big] \xrightarrow{\mathrm{d}} \sum_{j=1}^r \lambda_j z'_{s^{(j)}}(\theta^*, \tilde{\mathsf{Z}}). \tag{45}$$

Since choice of $\tilde{\mathcal{S}}$ and $\lambda_j \in \mathbb{R}, j = 1, \ldots, r$ are arbitrary, (45) implies (a) above.

For (b), we show that for any $\epsilon > 0$, there exists a compact set $K(\epsilon)$ such that

$$P\{\delta_{\mathcal{E},n} \in K(\epsilon)\} \geq 1 - \epsilon \quad \text{for all } n. \tag{46}$$

Since (39) holds with $\sqrt{n}\, U_n = o_p(1)$, using the notation from (42), (46) will hold if we show that there exists a compact set $K(\epsilon)$ such that for large enough n,

$$P\big\{\mathfrak{z}'\big(\theta^*, \sqrt{n}(\hat{\theta}_n^* - \theta^*)\big) \in K(\epsilon)\big\} \geq 1 - \epsilon. \tag{47}$$

To show (47) holds, we demonstrate that $z'_s(\theta^*, \sqrt{n}(\hat{\theta}_n^* - \theta^*))$, seen as a function of $s \in \mathcal{S}$, is L-Lipschitz for all n. That is, for any two points $s, t \in \mathcal{S}$ and all n,

$$\big\|z'_s\big(\theta^*, \sqrt{n}(\hat{\theta}_n^* - \theta^*)\big) - z'_t\big(\theta^*, \sqrt{n}(\hat{\theta}_n^* - \theta^*)\big)\big\| \leq L\|s - t\| \tag{48}$$

for some non-negative random variable L. The inequality in (48) implies that the set $\mathfrak{z}'\big(\theta^*, \sqrt{n}(\hat{\theta}^* - \theta^*)\big)$ can be covered by a ball of radius $L\,\mathrm{diam}(\mathcal{S})$ and centered at any fixed point in the set. For example,

$$\mathfrak{z}'(\theta^*, \sqrt{n}(\hat{\theta}^* - \theta^*)) \subset B(X_{1,n}, L\,\mathrm{diam}(\mathcal{S})), \tag{49}$$

where, for convenience, we choose $s^{(1)} = (s_1^{(1)}, \ldots, s_d^{(1)}) := (1, 0, \ldots, 0)$ and

$$X_{1,n} := z'_{s^{(1)}}(\theta^*, \sqrt{n}(\hat{\theta}_n^* - \theta^*)) = \sqrt{n}\, \mathsf{H}_1^{-1}\big[(\bar{\mathsf{V}}_1 - \mathsf{V}_1) - (\bar{\mathsf{H}}_1 - \mathsf{H}_1)(\mathsf{H}_1^{-1}\mathsf{V}_1)\big].$$

Thus, if (48) holds, then (49) implies (47), and tightness holds.

We now show that (48) holds. For given $s = (s_1, \ldots, s_d) \in \mathcal{S}$, write

$$z'_s\big(\theta^*, \sqrt{n}(\hat{\theta}^*_n - \theta^*)\big) = [\underbrace{\textstyle\sum_k s_k \mathsf{H}_k}_{\mathsf{A}_s}]^{-1} \underbrace{\textstyle\sum_k s_k \big[\overbrace{\sqrt{n}\,(\bar{\mathsf{V}}_k - \mathsf{V}_k)}^{\Delta_{\mathsf{V},k}} - \overbrace{\sqrt{n}\,(\bar{\mathsf{H}}_k - \mathsf{H}_k)}^{\Delta_{\mathsf{H},k}}\, z_s(\theta^*)\big]}_{B_s}.$$

The matrix A_s is fixed while the vector B_s is random. For $t = (t_1, \ldots, t_d) \in \mathcal{S}$, A_t and B_t represent analogous quantities with $t = s + \delta$ for $\delta = (\delta_1, \ldots, \delta_d) \in \mathbb{R}^d$. Then

$$\begin{aligned} \big\|z'_s\big(\theta^*, \sqrt{n}(\hat{\theta}^*_n - \theta^*)\big) - z'_t\big(\theta^*, \sqrt{n}(\hat{\theta}^*_n - \theta^*)\big)\big\| &= \big\|\mathsf{A}_s^{-1} B_s - \mathsf{A}_t^{-1} B_t\big\| \\ &= \big\|\mathsf{A}_s^{-1}(B_s - B_t) + (\mathsf{A}_s^{-1} - \mathsf{A}_t^{-1}) B_t\big\| \\ &\le \big\|\mathsf{A}_s^{-1}\big\|_{\mathbb{S}_q} \big\|B_s - B_t\big\| + \big\|\mathsf{A}_s^{-1} - \mathsf{A}_t^{-1}\big\|_{\mathbb{S}_q} \big\|B_t\big\|. \end{aligned} \tag{50}$$

We find upper bounds for each term in (50). In what follows, let $\lambda_1(\cdot)$ and $\lambda_q(\cdot)$ denote the smallest and largest eigenvalues of their arguments, respectively, and define $\lambda_* := \min_k \lambda_1(\mathsf{H}_k)$, $\lambda^* := \max_k \lambda_q(\mathsf{H}_k)$, and $\kappa := \lambda^*/\lambda_* < \infty$.

First, for the terms concerning A_s and A_t in (50), we have

$$\|\mathsf{A}_s^{-1}\|_{\mathbb{S}_q} = \lambda_q(\mathsf{A}_s^{-1}) = \big(\lambda_1(\mathsf{A}_s)\big)^{-1} \le \big(\min_k \lambda_1(\mathsf{H}_k)\big)^{-1} = \lambda_*^{-1} < \infty. \tag{51}$$

Now, consider $\big\|\mathsf{A}_s^{-1} - \mathsf{A}_t^{-1}\big\|$. Rewriting A_t^{-1} and using Lemma 1 for small enough δ,

$$\begin{aligned} \mathsf{A}_s^{-1} - \mathsf{A}_t^{-1} &= \mathsf{A}_s^{-1} - (\textstyle\sum_k s_k \mathsf{H}_k + \delta_k \mathsf{H}_k)^{-1} = [\mathbb{I}_q - (\mathbb{I}_q + \mathsf{A}_s^{-1} \sum_k \delta_k \mathsf{H}_k)^{-1}]\mathsf{A}_s^{-1} \\ &= [\mathsf{A}_s^{-1} \textstyle\sum_k \delta_k \mathsf{H}_k - (\mathsf{A}_s^{-1} \sum_k \delta_k \mathsf{H}_k)^2 + (\mathsf{A}_s^{-1} \sum_k \delta_k \mathsf{H}_k)^3 + \ldots]\mathsf{A}_s^{-1}. \end{aligned}$$

Choose $\|s - t\| < (2\kappa d)^{-1}$. Then using (51), since $|\delta_k| = |s_k - t_k| \le \|s - t\|$,

$$\begin{aligned} \|\mathsf{A}_s^{-1} - \mathsf{A}_t^{-1}\| &\le \lambda_*^{-1}\big[\lambda_*^{-1}\|\textstyle\sum_k \delta_k \mathsf{H}_k\| + (\lambda_*^{-1}\|\sum_k \delta_k \mathsf{H}_k\|)^2 + (\lambda_*^{-1}\|\sum_k \delta_k \mathsf{H}_k\|)^3 + \ldots\big] \\ &\le \lambda_*^{-1}\big[\kappa d\|s - t\| + (\kappa d\|s - t\|)^2 + (\kappa d\|s - t\|)^3 + \ldots\big] \\ &= (\lambda_*^{-1}\kappa d\|s - t\|)\big[1 + (\kappa d\|s - t\|) + (\kappa d\|s - t\|)^2 + \ldots\big] \\ &= (\lambda_*^{-1}\kappa d\|s - t\|)\big[1 - \kappa d\|s - t\|\big]^{-1} \le 2\lambda_*^{-1}\kappa d\|s - t\|. \end{aligned} \tag{52}$$

Similarly, for the terms concerning B_s and B_t in (50), we have

$$\begin{aligned} \|B_t\| \le \sup_{t\in\mathcal{S}}\|B_t\| &= \sup_{t\in\mathcal{S}}\big\|\textstyle\sum_k t_k\big[\Delta_{\mathsf{V},k} - \Delta_{\mathsf{H},k} z_t(\theta^*)\big]\big\| \\ &\le \sup_{t\in\mathcal{S}}\big\|\textstyle\sum_k t_k\big[\Delta_{\mathsf{V},k} - \Delta_{\mathsf{H},k} A_t^{-1} \sum_k t_k \mathsf{V}_k\big]\big\| \\ &\le d\max_k\big\{\big\|\Delta_{\mathsf{V},k}\big\| + \big\|\Delta_{\mathsf{H},k}\big\|\lambda_*^{-1} d(\max_k\|\mathsf{V}_k\|)\big\} =: L_B. \end{aligned} \tag{53}$$

Finally, note that for $|\delta_k| \le \|s - t\|$, we have

$$\begin{aligned} \|z_t(\theta^*) - z_s(\theta^*)\| &= \|(\mathsf{A}_t^{-1} - \mathsf{A}_s^{-1})\textstyle\sum_k s_k \mathsf{V}_k + \mathsf{A}_t^{-1}\sum_k \delta_k \mathsf{V}_k\| \\ &\le (2\lambda_*^{-1}\kappa d\|s - t\|)(d\max_k\|\mathsf{V}_k\|) + \lambda_*^{-1}\|s - t\|(d\max_k\|\mathsf{V}_k\|) \\ &= (2\kappa d + 1)\lambda_*^{-1}(d\max_k\|\mathsf{V}_k\|)\|s - t\|, \end{aligned}$$

which, together with (53), implies

$$\begin{aligned}\|B_s - B_t\| &= \|\textstyle\sum_k \delta_k\big(\Delta_{\mathsf{H},k} z_t(\theta^*) - \Delta_{\mathsf{V},k}\big) + \sum_k s_k \Delta_{\mathsf{H},k}(z_t(\theta^*) - z_s(\theta^*))\| \\ &\le L_B\|s-t\| + (d\max_k\|\Delta_{\mathsf{H},k}\|)(2\kappa d+1)\lambda_*^{-1}(d\max_k\|\mathsf{V}_k\|)\|s-t\|. \end{aligned} \tag{54}$$

Then, use (51), (54), (52), and (53), respectively in (50) to see that the inequality holds for $\|s-t\| < (2\kappa d)^{-1}$. To see that (50) holds for all $s,t \in \mathcal{S}$, apply the upper bound to each small interval in $\mathcal{S}$. We omit the proof for $\delta_{\mathcal{P},n} \xrightarrow{\mathrm{d}} \mathfrak{g}'(\theta^*, \tilde{\mathsf{Z}})$ since it follows along identical lines. □

4 Simultaneous confidence regions

The uniform CLT in Theorem 4 naturally leads to the creation of confidence regions that are simultaneous across values of the scalarization parameter $s \in \mathcal{S}$. However, such confidence regions would be constructed in an infinite-dimensional space, while all error in estimating the efficient and Pareto sets is due solely to error in estimating the finite-dimensional θ^*. Thus, we propose a different method for confidence region construction. First, in Subsection 4.1, we construct a confidence region on θ^*. Then, in Subsection 4.2, we translate this confidence region into confidence regions on the infinite-dimensional objects, the efficient and Pareto sets, through a push-forward measure. Theorem 5 demonstrates that the constructed regions are, in fact, asymptotically valid confidence regions. While the regions could be constructed in other ways, the proposed construction preserves the inherent structure for the accepted curves in the region.

4.1 *Confidence region on the quadratic parameters*

Using Lemma 2 and Theorem 2 directly, we can construct the confidence region in the native space, $\big\{\theta \in \Theta\colon \sqrt{n}\big\|\mathrm{vec}_\Theta(\hat{\theta}_n^*) - \mathrm{vec}_\Theta(\theta)\big\| \le p_{1-\alpha}\big(\big\|\sqrt{\Omega_\gamma}\,Z\big\|\big)\big\}$, where $p_{1-\alpha}(\cdot)$ is a function that returns the $1-\alpha$ quantile of its argument for $\alpha \in (0,1)$. We preserve Ω_γ on the right side because it is positive semidefinite and *not* positive definite, due to repeated rows and columns in the matrix that result from the symmetry of (twice) the Hessian matrices. Thus, Ω_γ cannot be inverted.

However, it is possible to write the confidence region in the native space as a hyperellipsoid in vector space. To see this, notice that we can repeat all the analyses in Section 3 with a vectorized version of the matrices $h_k \in \mathbb{S}_q$ that keeps only the entries in its upper triangular aspect, thereby removing duplicates that arise from symmetry. Specifically, extract the elements of the upper triangular of h_k and vectorize them to yield stacked vector $h_k^{\mathsf{u}} \in \mathbb{R}^{q(q+1)/2}$. Thus, letting h_{kij} be the (i,j)th element of h_k,

$$h_k^{\mathsf{u}\intercal} = (h_{k11}, h_{k12}, h_{k22}, h_{k13}, \ldots, h_{k1d}, \ldots, h_{kdd})^\intercal. \tag{55}$$

Much like the vectorizations defined in (5) and (19), Transforming h_k into h_k^{u} is a linear operation. Next, stack the vector h_k^{u} with v_k and c_k to yield the vector $\beta_k \coloneqq \big(h_k^{\mathsf{u}\intercal}, v_k^\intercal, c_k\big)^\intercal \in \mathbb{R}^{(q(q+1)/2+q+1)\times 1}$. For notational convenience, let

$$b \coloneqq q(q+1)/2 + q + 1, \tag{56}$$

and in analogy to $\mathrm{vec}_\Theta(\hat{\theta}_n^*)$ and $\mathrm{vec}_\Theta(\theta^*)$, define

$$\hat{\beta}_n^* := \left(\bar{\mathsf{H}}_1^{\mathrm{u}\intercal}, \bar{\mathsf{V}}_1^\intercal, \bar{\mathsf{C}}_1, \dots, \bar{\mathsf{H}}_d^{\mathrm{u}\intercal}, \bar{\mathsf{V}}_d^\intercal, \bar{\mathsf{C}}_d\right)^\intercal, \qquad \beta^* := \left(\mathsf{H}_1^{\mathrm{u}\intercal}, \mathsf{V}_1^\intercal, \mathsf{C}_1, \dots, \mathsf{H}_d^{\mathrm{u}\intercal}, \mathsf{V}_d^\intercal, \mathsf{C}_d\right)^\intercal \tag{57}$$

where $\hat{\beta}_n^*, \beta^* \in \mathbb{R}^{db\times 1}$ and, using the same notation as (55),

$$\bar{\mathsf{H}}_k^{\mathrm{u}\intercal} := (\bar{\mathsf{H}}_{k11}, \bar{\mathsf{H}}_{k12}, \bar{\mathsf{H}}_{k22}, \bar{\mathsf{H}}_{k13}, \dots, \bar{\mathsf{H}}_{k1d}, \dots, \bar{\mathsf{H}}_{kdd})^\intercal.$$

Now, construct the positive definite $\Omega_\gamma^{\mathrm{u}}$ matrix in the same way as Ω_γ in (25), except replace each Ω_k with Ω_k^{u} and each cross-covariance matrix $\Omega_{k,k'}$ with $\Omega_{k,k'}^{\mathrm{u}}$, where

$$\begin{aligned}\Omega_k^{\mathrm{u}} &:= \mathrm{Var}\left(\left(H_k^{\mathrm{u}}(\xi)^\intercal, V_k(\xi)^\intercal, C_k(\xi)\right)^\intercal\right) \in \mathbb{S}_b^{>},\\ \Omega_{k,k'}^{\mathrm{u}} &:= \mathrm{Cov}\left(\left(H_k^{\mathrm{u}}(\xi)^\intercal, V_k(\xi)^\intercal, C_k(\xi)\right)^\intercal, \left(H_{k'}^{\mathrm{u}}(\xi)^\intercal, V_{k'}(\xi)^\intercal, C_{k'}(\xi)\right)^\intercal\right).\end{aligned}$$

Then, Lemma 2 implies that

$$\sqrt{n}\left(\Omega_\gamma^{\mathrm{u}}\right)^{-1/2}\left(\hat{\beta}_n^* - \beta^*\right) \xrightarrow{\mathrm{d}} Z \tag{58}$$

where $Z \in \mathbb{R}^{db\times 1}$ on the right side is a standard multivariate normal random vector. Then, the CLT in (58) implies that a $(1-\alpha)$ confidence region on β^* can be constructed as the hyperellipsoid

$$\mathcal{R}_{n,1-\alpha} := \left\{\beta \in \mathbb{R}^{db\times 1} \colon n(\hat{\beta}_n^* - \beta)^\intercal \left(\Omega_\gamma^{\mathrm{u}}\right)^{-1}(\hat{\beta}_n^* - \beta) \le \chi^2_{db,1-\alpha}\right\}, \tag{59}$$

where $\chi^2_{db,1-\alpha}$ denotes the $1-\alpha$ quantile of a χ^2 random variable with db degrees of freedom; recall that b is defined in (56). Thus, $\mathcal{R}_{n,1-\alpha}$ is a hyperellipsoid in $\mathbb{R}^{db\times 1}$.

4.2 Confidence regions in the decision and objective spaces

Next, we transform the region in (59) into asymptotically exact simultaneous confidence regions in the decision and objective spaces. First, we transform a vector in the hyperellipsoid $\beta \in \mathcal{R}_{n,1-\alpha} \subset \mathbb{R}^{db\times 1}$ back into an object in Θ. In analogy to the inverse vectorization defined in (20), let $\mathrm{vec}^{-1}_{\mathbb{R}^{db},\Theta} \colon \mathbb{R}^{db} \to \Theta$ denote the (linear) transformation from β back into the θ space. For each $\beta \in \mathbb{R}^{db}$ such that $\mathrm{vec}^{-1}_{\mathbb{R}^{db},\Theta}(\beta) \in \Theta$, define

$$\mathfrak{e}(\beta) := \left\{z_s(\mathrm{vec}^{-1}_{\mathbb{R}^{db},\Theta}(\beta)), s \in \mathcal{S}\right\}, \quad \mathfrak{p}(\beta) := \left\{g_s(\mathrm{vec}^{-1}_{\mathbb{R}^{db},\Theta}(\beta)), s \in \mathcal{S}\right\}.$$

Then, set

$$\mathcal{X}_{n,1-\alpha} := \left\{\mathfrak{e}(\beta) \colon \beta \in \mathcal{R}_{n,1-\alpha}\right\}, \quad \mathcal{Y}_{n,1-\alpha} := \left\{\mathfrak{p}(\beta) \colon \beta \in \mathcal{R}_{n,1-\alpha}\right\}. \tag{60}$$

The following theorem states the asymptotic validity of the regions in (60).

Theorem 5. *Let the postulates of Lemma 2 hold. Suppose $\theta^* \in \mathcal{H}$ and $\hat{\theta}_n^* \in \mathcal{H}$ for each n, $\mathcal{H} \subset \Theta$, and $\alpha \in (0,1)$. Then $\mathcal{X}_{n,1-\alpha}$ and $\mathcal{Y}_{n,1-\alpha}$, defined in (60), are asymptotically valid confidence regions on the efficient set $\mathcal{E}$ and the Pareto set $\mathcal{P}$, respectively; that is,*

$$\lim_{n\to\infty} P\{\mathcal{E} \in \mathcal{X}_{n,1-\alpha}\} = 1-\alpha, \quad \lim_{n\to\infty} P\{\mathcal{P} \in \mathcal{Y}_{n,1-\alpha}\} = 1-\alpha.$$

Proof. First, recall $\mathcal{E} = \mathfrak{e}(\beta^*)$ and $\mathcal{P} = \mathfrak{p}(\beta^*)$ for β^* in (57). From (60), the proof follows trivially through the push-forward measure and the fact that $P\{\beta^* \in \mathcal{R}_{n,1-\alpha}\} \to 1-\alpha$ as $n \to \infty$. More formally, since the CLT on β^* in (58) holds and $\mathfrak{e}(\cdot)$ is measurable, we have $1-\alpha = \lim_{n\to\infty} P\{\beta^* \in \mathcal{R}_{n,1-\alpha}\} = \lim_{n\to\infty} P\{\mathfrak{e}(\beta^*) \in \mathfrak{e}(\mathcal{R}_{n,1-\alpha})\} = \lim_{n\to\infty} P\{\mathcal{E} \in \mathcal{X}_{n,1-\alpha}\}$. An analogous justification holds for $P\{\mathcal{P} \in \mathcal{Y}_{n,1-\alpha}\}$ since $\mathfrak{p}(\cdot)$ is measurable. $\square$

5 Illustrative example

We implement an illustrative example of constructing 95% confidence regions for the least squares problem in (4) for $q = 2$ covariates and $d = 2$ independent data sources with equal sample sizes $n_1 = n_2$ such that $\gamma = (1/2, 1/2)$. For each independent data source $k = 1, 2$, let $\xi_k = (Q_k, R_k) = (Q_{k1}, Q_{k2}, R_k)$ be a multivariate normal random vector with means and covariance matrices specified by

$$\mu_1 = \left[\, \mu_{Q_1} \,\middle|\, \mu_{R_1} \,\right] = \left[\, 1 \;\; 2 \,\middle|\, 6 \,\right], \qquad \Sigma_1 = \left[\begin{array}{c|c} \Sigma_{Q_1} & \Sigma_{Q_1 R_1}^\intercal \\ \hline \Sigma_{Q_1 R_1} & \sigma_{R_1}^2 \end{array}\right] = \left[\begin{array}{cc|c} 6.06 & 2.50 & 2.50 \\ 2.50 & 2.79 & 0.72 \\ \hline 2.50 & 0.72 & 1.84 \end{array}\right],$$

$$\mu_2 = \left[\, \mu_{Q_2} \,\middle|\, \mu_{R_2} \,\right] = \left[\, 2 \;\; 1 \,\middle|\, 15 \,\right], \qquad \Sigma_2 = \left[\begin{array}{c|c} \Sigma_{Q_2} & \Sigma_{Q_2 R_2}^\intercal \\ \hline \Sigma_{Q_2 R_2} & \sigma_{R_2}^2 \end{array}\right] = \left[\begin{array}{cc|c} 1.02 & 0.27 & 0.76 \\ 0.27 & 0.19 & 0.47 \\ \hline 0.76 & 0.47 & 1.45 \end{array}\right].$$

Under this problem structure, for objectives $k = 1, 2$, we calculate the inputs for the true objective function values as $\mathsf{H}_k = \mathrm{E}[Q_k^\intercal Q_k] = \Sigma_{Q_k} + \mu_{Q_k}^\intercal \mu_{Q_k}$, $\mathsf{V}_k = \mathrm{E}[Q_k^\intercal R_k] = \Sigma_{Q_k R_k}^\intercal + \mu_{Q_k}^\intercal \mu_{R_k}$, and $\mathsf{C}_k = \mathrm{E}[R_k^2] = \sigma_{R_k}^2 + \mu_{R_k}^2$, yielding

$$\mathsf{H}_1 = \begin{bmatrix} 7.06 & 4.50 \\ 4.50 & 6.79 \end{bmatrix}, \quad \mathsf{V}_1 = \begin{bmatrix} 8.50 \\ 12.72 \end{bmatrix}, \quad \mathsf{C}_1 = \;\; 37.84, \quad x_1^* = \begin{bmatrix} 0.0171558 \\ 1.8619733 \end{bmatrix},$$
$$\mathsf{H}_2 = \begin{bmatrix} 5.02 & 2.27 \\ 2.27 & 1.19 \end{bmatrix}, \quad \mathsf{V}_2 = \begin{bmatrix} 30.76 \\ 15.47 \end{bmatrix}, \quad \mathsf{C}_2 = 226.45, \quad x_2^* = \begin{bmatrix} 1.8120356 \\ 9.5434279 \end{bmatrix},$$

where the values of x_1^* and x_2^* are reported to seven decimal places; all other values are exact. Figure 1 provides an image of the efficient and Pareto sets.

For i.i.d. observations of $(Q_1^{(\ell)}, R_1^{(\ell)})$ and $(Q_2^{(\ell)}, R_2^{(\ell)})$, $\ell = 1, \ldots, n_1 = n_2$, independently drawn from the multivariate normal distributions defined above and stacked into matrices as described in Subsection 2.2.2, the sample-path version of the bi-objective least-squares problem in (4) is

$$\text{minimize} \;\; \Big[\big(\bar{F}_1(x,n), \bar{F}_2(x,n)\big)^\intercal = \Big(\tfrac{1}{n\gamma_1}\big\|\mathsf{Q}_1 x - \mathsf{R}_1\big\|_2^2, \tfrac{1}{n\gamma_2}\big\|\mathsf{Q}_2 x - \mathsf{R}_2\big\|_2^2\Big)^\intercal\Big],$$

where the relevant estimators are calculated using (15), (16), and (57). These estimators are inputs to the hyperellipsoidal confidence region in (59), which provides inputs to the confidence regions on the efficient and Pareto sets in (60).

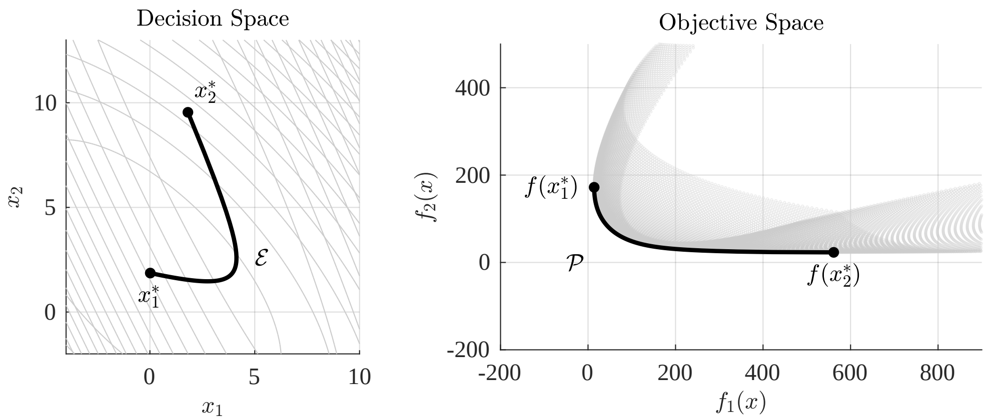

Figure 1: The decision space (left) and the objective space (right) for the bi-objective least squares problem with $q = 2$ covariates and $d = 2$ data sources, plotted using nonbinding box constraints.

5.1 Obtaining the known covariance matrix

We assume the covariance matrix Ω_γ in (59) is known. To obtain the Ω_γ matrix, we first obtain expressions for $\Omega_1, \Omega_2 \in \mathbb{S}_7^{\geqslant}$ from (21) for $k = 1, 2$ using

$$\begin{aligned}\Omega_k = \mathrm{E}\left[\mathrm{vec}_{\Theta_1}(Q_k^\intercal Q_k, Q_k^\intercal R_k, R_k^2)\big(\mathrm{vec}_{\Theta_1}(Q_k^\intercal Q_k, Q_k^\intercal R_k, R_k^2)\big)^\intercal\right]\\ -\mathrm{vec}_{\Theta_1}(\mathsf{H}_k, \mathsf{V}_k, \mathsf{C}_k)\big(\mathrm{vec}_{\Theta_1}(\mathsf{H}_k, \mathsf{V}_k, \mathsf{C}_k)\big)^\intercal.\end{aligned}$$

The term $\mathrm{E}\left[\mathrm{vec}_{\Theta_1}(Q_k^\intercal Q_k, Q_k^\intercal R_k, R_k^2)\big(\mathrm{vec}_{\Theta_1}(Q_k^\intercal Q_k, Q_k^\intercal R_k, R_k^2)\big)^\intercal\right]$ above equals the following (simplified) matrix, with omitted subscript k for compactness:

$$\begin{bmatrix} \mathrm{E}[Q_1^4] & \mathrm{E}[Q_1^3Q_2] & \mathrm{E}[Q_1^3Q_2] & \mathrm{E}[Q_1^2Q_2^2] & \mathrm{E}[Q_1^3R] & \mathrm{E}[Q_1^2Q_2R] & \mathrm{E}[Q_1^2R^2] \\ \mathrm{E}[Q_1^3Q_2] & \mathrm{E}[(Q_1Q_2)^2] & \mathrm{E}[(Q_1Q_2)^2] & \mathrm{E}[Q_1Q_2^3] & \mathrm{E}[Q_1^2Q_2R] & \mathrm{E}[Q_1Q_2^2R] & \mathrm{E}[Q_1Q_2R^2] \\ \mathrm{E}[Q_1^3Q_2] & \mathrm{E}[(Q_1Q_2)^2] & \mathrm{E}[(Q_1Q_2)^2] & \mathrm{E}[Q_1Q_2^3] & \mathrm{E}[Q_1^2Q_2R] & \mathrm{E}[Q_1Q_2^2R] & \mathrm{E}[Q_1Q_2R^2] \\ \mathrm{E}[Q_1^2Q_2^2] & \mathrm{E}[Q_1Q_2^3] & \mathrm{E}[Q_1Q_2^3] & \mathrm{E}[Q_2^4] & \mathrm{E}[Q_2^2Q_1R] & \mathrm{E}[Q_2^3R] & \mathrm{E}[Q_2^2R^2] \\ \mathrm{E}[Q_1^3R] & \mathrm{E}[Q_1^2Q_2R] & \mathrm{E}[Q_1^2Q_2R] & \mathrm{E}[Q_1Q_2^2R] & \mathrm{E}[(Q_1R)^2] & \mathrm{E}[Q_1Q_2R^2] & \mathrm{E}[Q_1R^3] \\ \mathrm{E}[Q_1^2Q_2R] & \mathrm{E}[Q_1Q_2^2R] & \mathrm{E}[Q_1Q_2^2R] & \mathrm{E}[Q_2^3R] & \mathrm{E}[Q_1Q_2R^2] & \mathrm{E}[(Q_2R)^2] & \mathrm{E}[Q_2R^3] \\ \mathrm{E}[Q_1^2R^2] & \mathrm{E}[Q_1Q_2R^2] & \mathrm{E}[Q_1Q_2R^2] & \mathrm{E}[Q_2^2R^2] & \mathrm{E}[Q_1R^3] & \mathrm{E}[Q_2R^3] & \mathrm{E}[R^4] \end{bmatrix}.$$

Then, use Isserlis' Theorem [20, 32] to calculate the exact values

$$\Omega_1 = \begin{bmatrix} 97.6872 & 59.54 & 59.54 & 32.5 & 108.02 & 52.5 & 72.5 \\ 59.54 & 60.1874 & 60.1874 & 45.11 & 104.053 & 66.955 & 72.24 \\ 59.54 & 60.1874 & 60.1874 & 45.11 & 104.053 & 66.955 & 72.24 \\ 32.5 & 45.11 & 45.11 & 60.2082 & 66.48 & 76.7376 & 35.5968 \\ 108.02 & 104.053 & 104.053 & 66.48 & 267.4 & 134.4 & 211.28 \\ 52.5 & 66.955 & 66.955 & 76.7376 & 134.4 & 130.732 & 98.9496 \\ 72.5 & 72.24 & 72.24 & 35.5968 & 211.28 & 98.9496 & 271.731 \end{bmatrix},$$

$$\Omega_2 = \begin{bmatrix} 18.4008 & 6.7908 & 6.7908 & 2.3058 & 68.8304 & 19.6504 & 92.3552 \\ 6.7908 & 3.1267 & 3.1267 & 1.4026 & 27.4846 & 11.7213 & 51.7144 \\ 6.7908 & 3.1267 & 3.1267 & 1.4026 & 27.4846 & 11.7213 & 51.7144 \\ 2.3058 & 1.4026 & 1.4026 & 0.8322 & 10.2338 & 6.8186 & 28.6418 \\ 68.8304 & 27.4846 & 27.4846 & 10.2338 & 282.957 & 89.8987 & 431.204 \\ 19.6504 & 11.7213 & 11.7213 & 6.8186 & 89.8987 & 58.7964 & 256.363 \\ 92.3552 & 51.7144 & 51.7144 & 28.6418 & 431.204 & 256.363 & 1309.21 \end{bmatrix}.$$

Note that approximations to Ω_1 and Ω_2 also can be calculated using Monte Carlo with a very large sample size. The covariance matrices Ω_1 and Ω_2 have identical second and third columns as well as identical second and third rows. Therefore, they are not full rank and the smallest eigenvalues equal zero, demonstrating they are only positive semidefinite and not positive definite. Since the data sources are independent, for $\gamma = (1/2, 1/2)$, construct the $\Omega_\gamma \in \mathbb{S}^{\geqslant}_{14}$ matrix according to (25) with zeroes in place of the cross-covariance matrices.

5.2 Constructing the confidence regions

Constructing the 95% confidence regions according to Section 4 requires obtaining points in the hyperellipsoidal region (59) that can be transformed into curves in the decision and objective spaces. In this example, $\Omega^{\mathrm{u}}_\gamma$ (discussed in Subsection 4.1) equals

$$\Omega^{\mathrm{u}}_\gamma = 2\begin{bmatrix}\Omega^{\mathrm{u}}_1 & 0\\ 0 & \Omega^{\mathrm{u}}_2\end{bmatrix} \in \mathbb{S}^{>}_{12},$$

where Ω^{u}_1 and Ω^{u}_2 are identical to Ω_1 and Ω_2, respectively, except with the second row and second column of each matrix deleted. To obtain a projection of the region $\mathcal{R}_{n,.95}$ in the decision and objective spaces, one can use Monte Carlo to sample m points uniformly in the hyperellipsoid by calculating

$$\beta_j = L\left(\frac{(n^{-1}\chi^2_{12,.95})^{1/2}U^{1/12}Z}{\|Z\|}\right) + \hat{\beta}^*_n, \quad j = 1, 2, \ldots, m$$

where $\hat{\beta}^*_n$ is from (57), L is the lower triangular matrix resulting from the Cholesky decomposition of $\Omega^{\mathrm{u}}_\gamma = LL^\intercal$, U is a uniform$(0,1)$ random variable, and $Z \in \mathbb{R}^{12}$ [14, p. 234f.]. In our numerical experiments, given a fixed m, generating all points *on* the hyperellipsoid rather than *in* the hyperellipsoid appears to create a "larger" (and thus, more accurate) projection. Henceforth, our reported figures omit the uniform random variable in the expression for β_j. Finally, we create the projections by plotting

$$\mathfrak{e}(\beta_j) = \{z_s(\mathrm{vec}^{-1}_{\mathbb{R}^{db},\Theta}(\beta_j)), s \in \tilde{\mathcal{S}}\}, \quad \mathfrak{p}(\beta_j) = \{g_s(\mathrm{vec}^{-1}_{\mathbb{R}^{db},\Theta}(\beta_j)), s \in \tilde{\mathcal{S}}\}, \tag{61}$$

for $j = 1, 2, \ldots, m$ in the decision and objective spaces, respectively, where $\tilde{\mathcal{S}}$ represents an appropriately fine discretization of $\mathcal{S}$.

5.3 Confidence regions with increasing sample sizes

We construct the confidence regions at increasing sample sizes, $n \in \{120, 600, 3000\}$. First, we discuss coverage of the hyperellipsoidal region $\mathcal{R}_{n,.95}$ as the sample size increases. Then, we discuss Figures 2 to 4, which contain two independent (macro) replications of the projected confidence regions at increasing sample sizes.

5.3.1 Coverage

Recall that the hyperellipsoidal region $\mathcal{R}_{n,.95}$ should cover the true value β^* in approximately 95% of macro replications, with coverage improving as n increases. To estimate the coverage, we conduct $10,000,000$ (macro) replications of the procedure that creates $\mathcal{R}_{n,.95}$ and check whether it contains β^*. At each sample size $n \in \{120, 600, 3000\}$, our numerical experiments indicate that the coverage is at least 94.3%, 94.8%, and 94.9%, respectively. The respective

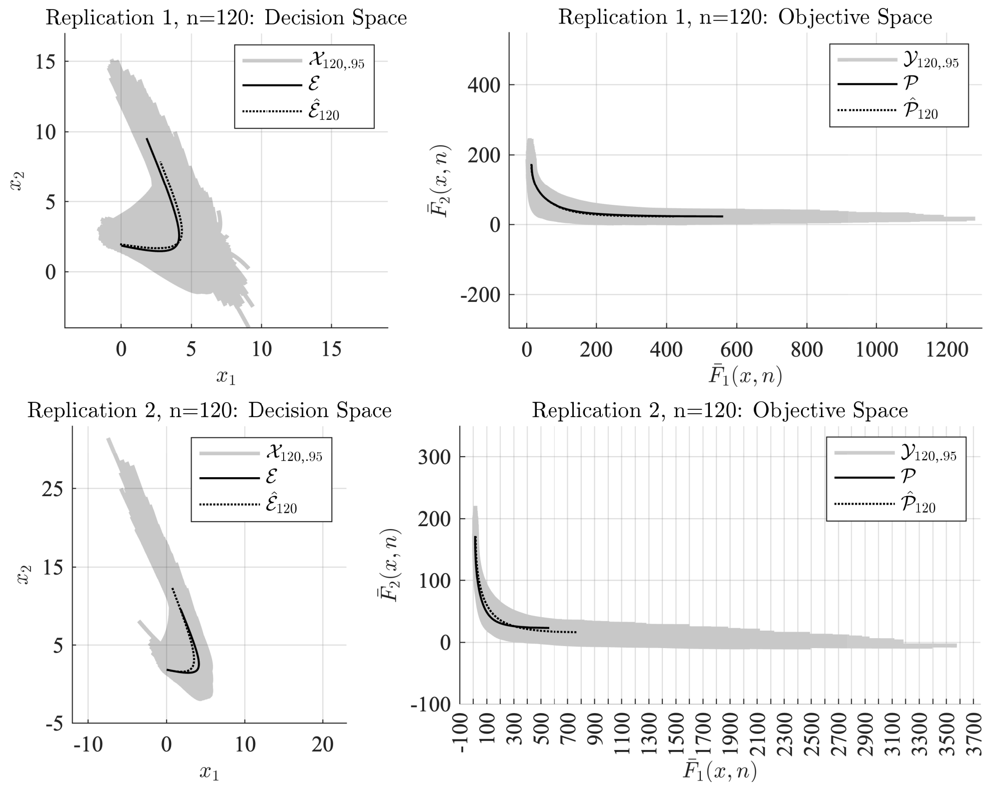

Figure 2: Two independent macro replications of the 95% confidence regions (top, bottom) projected into the decision space (left) and the objective space (right), respectively, for the bi-objective least squares problem with $q = 2$, $d = 2$, $n = 120$ ($n_1 = n_2 = 60$), discretization $\tilde{\mathbb{S}}$ from (62) for plotting, and $m = 80,000$ generated curves per plot. For visibility, each plot uses a different axis scale.

estimated standard errors for these numbers are approximately 7.3×10^{-5}, 7.0×10^{-5}, and 6.9×10^{-5}; therefore, we truncate the coverage estimator to the third decimal place to be conservative.

5.3.2 *Projected confidence regions*

To illustrate the projected confidence regions, we report two (macro) replications of the procedure described in Subsection 5.2 at increasing sample sizes $n \in \{120, 600, 3000\}$ with equal sample sizes from each data source; that is, $n_1 = n_2 \in \{60, 300, 1500\}$. As the sample sizes increase, the projected regions are constructed with common random numbers; that is, for $n_1 = n_2 = 300$, the first 60 observations from each data source are the same as those reported for $n_1 = n_2 = 60$; likewise, for $n_1 = n_2 = 1500$, the first 300 observations from each data source are the same as those reported for $n_1 = n_2 = 300$. Each projected curve is plotted using discretization

$$\tilde{\mathbb{S}} = \{0, 0.05, \ldots, 0.85, 0.86, 0.87, \ldots, 0.94, 0.942, 0.944, \ldots, 0.998, 0.999, 1\} \tag{62}$$

where the cardinality is $|\tilde{\mathbb{S}}| = 60$ points per plot. Each plot contains $m = 80,000$ projected curves. Figures 2 to 4 show the resulting projections.

Note that in the decision space, each projected curve $\mathfrak{e}(\beta_j)$, $j = 1, \ldots, m$ is an arc of hyperbola [5], and in the objective space, each projected curve $\mathfrak{p}(\beta_j)$, $j = 1, \ldots, m$ is a

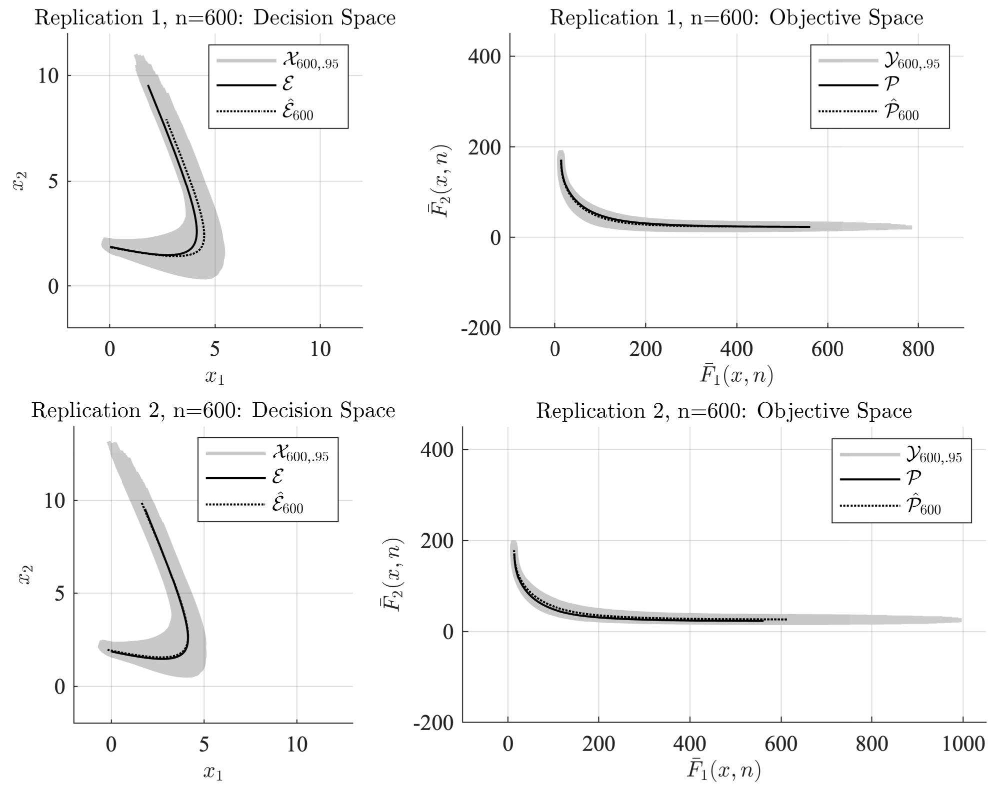

Figure 3: Two independent macro replications of the 95% confidence regions (top, bottom) projected into the decision space (left) and the objective space (right), respectively, for the bi-objective least squares problem with $q = 2$, $d = 2$, $n = 600$ ($n_1 = n_2 = 300$), discretization $\tilde{\mathbb{S}}^2$ from (62) for plotting, and $m = 80,000$ generated curves per plot. For each data source, the first 60 data points are the same as those used to construct Figure 2.

Pareto curve in the sense that no point dominates any other point. Thus, the projected images in the objective space are more constrained in shape than those in the decision space. At $n = 120$ in Figure 2, this lack of structure in the decision space is most apparent: At this sample size, we have not yet determined the direction of curvature for the efficient set in the decision space. In part as a result of joint uncertainty in the location of x_2^* and points near it, replication 1 in Figure 2 (top image) contains arcs of hyperbola that bend in the opposite direction from the true efficient set. The bottom image also demonstrates this uncertainty in the joint locations of x_2^*, points near it, and, consequently, $f(x_2^*)$.

Increasing the sample size to $n = 600$ in Figure 3 appears to resolve some of the uncertainty which is apparent at the lower sample size in Figure 2. Further, as the elliptical region from (59) becomes smaller with increasing n, the projection with a fixed number of $m = 80,000$ curves increases in "quality;" both regions in Figure 3 appear somewhat smoother than those in Figure 2.

In Figure 4, we increase the sample size to $n = 3000$. At this sample size, the projected confidence regions for each replication look quite similar to each other, as one would expect with a large enough sample size.

While additional research is required to make the projected confidence regions useful to practitioners (e.g., by enabling estimation of the covariance matrix $\Omega_\gamma^{\mathrm{u}}$), we remark that at low sample sizes, the projected region in the decision space may resemble a "blob" more

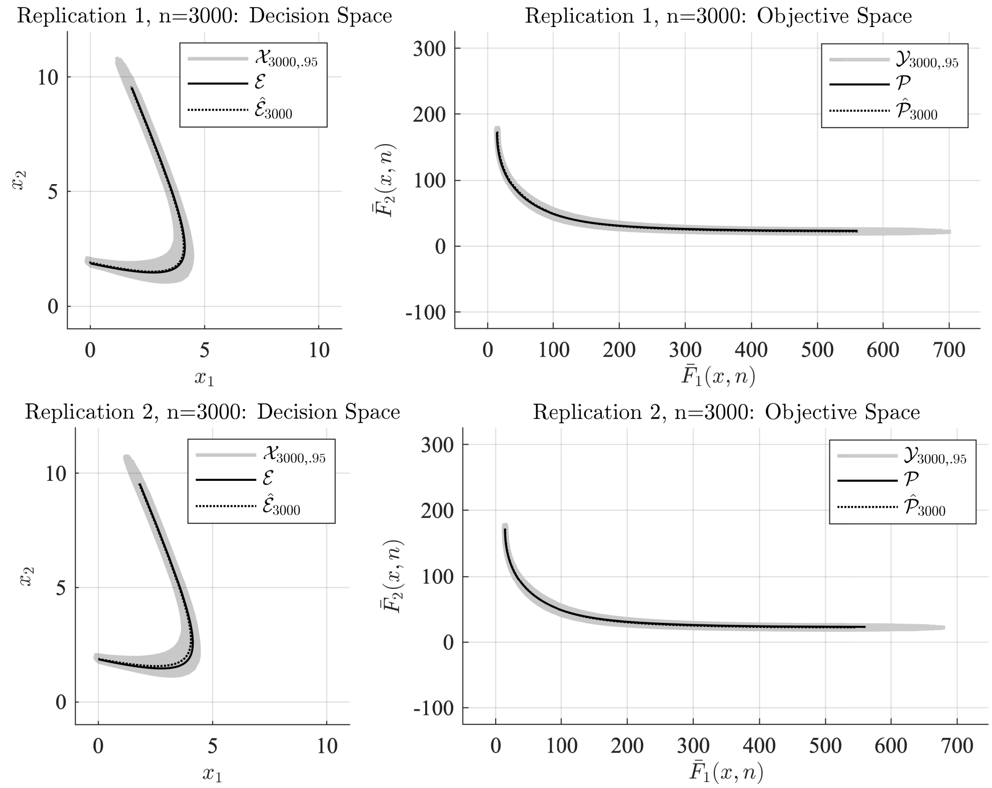

Figure 4: Two independent macro replications of the 95% confidence regions (top, bottom) projected into the decision space (left) and the objective space (right), respectively, for the bi-objective least squares problem with $q = 2$, $d = 2$, $n = 3000$ ($n_1 = n_2 = 1500$), discretization $\tilde{\mathbb{S}}$ from (62) for plotting, and $m = 80,000$ generated curves per plot. For each data source, the first 300 data points are the same as those used to construct Figure 3.

than structured cloud around a curve (see Figure 2 versus Figure 4). Thus, some value in constructing such projections may lie in being able to say whether there is enough data to determine, e.g., the direction of curvature of the efficient set with reasonable confidence. A projected confidence region that does not closely adhere to the estimated efficient set may indicate additional data are required.

6 Concluding remarks

We consider the problem of uncertainty quantification for MOSCQPs. We provide a uniform CLT which crucially relies on the expression for the Fréchet derivative of the points in the efficient and Pareto sets with respect to the matrix and vector parameters of the quadratic objectives. We also provide a direct method for calculating confidence regions on the infinite-dimensional efficient and Pareto sets by first constructing a confidence region in the finite-dimensional space of the matrix and vector parameters of the quadratic objectives, and then passing them through the closed-form expressions for the scalarized efficient and Pareto set estimators. These results form a promising basis for future uncertainty quantification in multiobjective stochastic optimization.

Acknowledgments

S. R. Hunter thanks Prof. Margaret M. Wiecek for discussions surrounding multiobjective optimization and its impact on decision-making.

References

[1] R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer, New York, 2007.

[2] L. A. Aguirre. A bird's eye view of nonlinear system identification. *arXiv*, February 2022.

[3] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, Hoboken, NJ, third edition, 2003.

[4] Q. Bai, M. Miralinaghi, S. Labi, and K. C. Sinha. Methodology for analyzing the trade-offs associated with multi-objective optimization in transportation asset management under uncertainty. *Computer-Aided Civil and Infrastructure Engineering*, 36(4):381–401, 2021.

[5] A. Beato-Moreno, P. Luque-Calvo, R. Osuna-Gómez, and A. Rufián-Lizana. Finding efficient points in multiobjective quadratic programming with strictly convex objective functions. In R. Caballero, F. Ruiz, and R. Steuer, editors, *Advances in Multiple Objective and Goal Programming*, volume 455 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin, Heidelberg, 1997.

[6] A. K. Bera and M. Koley. A history of the delta method and some new results. *Sankhya B: The Indian Journal of Statistics*, 2023.

[7] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley-Interscience, New York, second edition, 1999.

[8] T. Bodnar and W. Schmid. Econometrical analysis of the sample efficient frontier. *The European Journal of Finance*, 15(3):317–335, 2009.

[9] H. Bonnel and C. Collonge. Stochastic optimization over a Pareto set associated with a stochastic multi-objective optimization problem. *Journal of Optimization Theory and Applications*, 162:405–427, 2014.

[10] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, 2004.

[11] M. Ehrgott. *Multicriteria Optimization*, volume 491 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Heidelberg, 2nd edition, 2005.

[12] E. Eskinat, S. H. Johnson, and W. L. Luyben. Use of auxiliary information in system identification. *Industrial & Engineering Chemistry Research*, 32(9):1981–1992, 1993.

[13] T. S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall, London, 1996.

[14] G. S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Series in Operations Research. Springer, New York, 1996.

[15] J. Fliege and H. Xu. Stochastic multiobjective optimization: sample average approximation and applications. *Journal of Optimization Theory and Applications*, 151:135–162, 2011.

[16] B. Font. Bootstrap estimation of the efficient frontier. *Computational Management Science*, 13:541–570, 2016.

[17] T. Glasmachers. Challenges of convex quadratic bi-objective benchmark problems. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '19, pages 559 – 567, New York, NY, USA, 2019. Association for Computing Machinery.

[18] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, New York, 2 edition, 2013.

[19] S. R. Hunter and R. Pasupathy. Central limit theorems for constructing confidence regions in strictly convex multi-objective simulation optimization. In *Proceedings of the 2022 Winter Simulation Conference*, pages 3015–3026, Piscataway, NJ, 2022. IEEE.

[20] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1–2):134–139, 1918.

[21] P. L. W. Jayasekara, N. Adelgren, and M. M. Wiecek. On convex multiobjective programs with application to portfolio optimization. *Journal of Multi-Criteria Decision Analysis*, 27(3-4):189–202, 2020.

[22] P. L. W. Jayasekara, A. C. Pangia, and M. M. Wiecek. On solving parametric multiobjective quadratic programs with parameters in general locations. *Annals of Operations Research*, 320:123–172, 2023.

[23] J. D. Jobson. Confidence regions for the mean-variance efficient set: An alternative approach to estimation risk. *Review of Quantitative Finance and Accounting*, 1:235–257, 1991.

[24] T. A. Johansen. Multi-objective identification of FIR models. In *IFAC Proceedings Volumes*, volume 33, pages 917–922, 2000.

[25] T. A. Johansen and R. Babuška. Multiobjective identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems*, 11(6):847–860, 2003.

[26] S. Kim and J. Ryu. The sample average approximation method for multi-objective stochastic optimization. In S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, editors, *Proceedings of the 2011 Winter Simulation Conference*, pages 4026–4037, Piscataway, NJ, 2011. IEEE.

[27] E. Kreyszig. *Introductory Functional Analysis with Applications.* Wiley Classics Library. John Wiley & Sons, London, 1978.

[28] E. L. Lehmann. *Elements of Large-Sample Theory.* Springer Texts in Statistics. Springer, New York, 1999.

[29] D. G. Luenberger. *Optimization by Vector Space Methods.* Series in Decision and Control. John Wiley & Sons, New York, 1969.

[30] M. Luque, O. D. Marcenaro-Gutierrez, S. González-Gallardo, and A. B. Ruiz. Towards a framework to combine multiobjective optimization and econometrics and an application in economics of education. *RAIRO-Operations Research*, 56:2015–2035, 2022.

[31] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[32] J. V. Michalowicz, J. M. Nichols, F. Bucholtz, and C. C. Olson. A general Isserlis theorem for mixed-Gaussian random variables. *Statistics & Probability Letters*, 81(8):1233–1240, 2011.

[33] K. Miettinen. *Nonlinear Multiobjective Optimization.* Kluwer Academic Publishers, Boston, 1999.

[34] E. G. Nepomuceno, R. H. C. Takahashi, and L. A. Aguirre. Multiobjective parameter estimation for non-linear systems: affine information and least-squares formulation. *International Journal of Control*, 80(6):863–871, 2007.

[35] E. G. Nepomuceno, R. H. C. Takahashi, G. F. V. Amaral, and L. A. Aguirre. Nonlinear identification using prior knowledge of fixed points: A multiobjective approach. *International Journal of Bifurcation and Chaos*, 13(5):1229–1246, 2003.

[36] B. E. Ondes and S. R. Hunter. An upper bound on the Hausdorff distance between a Pareto set and its discretization in bi-objective convex quadratic optimization. *Optimization Letters*, 17:45–74, 2023.

[37] M. Pourahmadi. Covariance estimation: The GLM and regularization perspectives. *Statistical Science*, 26(3):369–387, 2011.

[38] Y. Qi and R. E. Steuer. An analytical derivation of properly efficient sets in multi-objective portfolio selection. *Annals of Operations Research*, 2024.

[39] R. J. Serfling. *Approximation Theorems of Mathematical Statistics.* Wiley Series in Probability and Mathematical Statistics. Wiley & Sons, New York, 1980.

[40] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory.* MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2009.

[41] C. Toure, A. Auger, D. Brockhoff, and N. Hansen. On bi-objective convex-quadratic problems. In K. Deb, E. Goodman, C. A. Coello Coello, K. Klamroth, K. Miettinen, S. Mostaghim, and P. Reed, editors, *Evolutionary Multi-Criterion Optimization*, pages 3–14, Cham, 2019. Springer International Publishing.

[42] A. W. van der Vaart. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

[43] S. Vogel. Random approximations in multiobjective optimization. *Mathematical Programming*, 164:29–53, 2017.

[44] H. Wang. Multi-objective retrospective optimization using stochastic zigzag search. *European Journal of Operational Research*, 263(3):946–960, December 2017.

[45] J. Ward. Structure of efficient sets for convex objectives. *Mathematics of Operations Research*, 14(2):249–257, May 1989.