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

\[
\begin{align*}
\text{minimize} \quad & f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_d(x) \end{pmatrix} = \begin{pmatrix} x^\top E[H_1(\xi)]x - 2E[V_1(\xi)]^\top x + E[C_1(\xi)] \\ \vdots \\ x^\top E[H_d(\xi)]x - 2E[V_d(\xi)]^\top x + E[C_d(\xi)] \end{pmatrix} \\
\text{s.t.} \quad & x \in \mathbb{R}^{q \times 1},
\end{align*}
\]  

(1)

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

\[
f_k(x) = x^\top H_k x - 2V_k^\top x + C_k.
\]  

(2)
The solution to (1) is the efficient set. The efficient set and its image, the Pareto set, are defined as
\[ \mathcal{E} := \{ x^* \in \mathbb{R}^q : \exists x \in \mathbb{R}^q \text{ such that } f(x) \leq f(x^*) \}, \quad \mathcal{P} := \{ f(x^*) : x^* \in \mathcal{E} \}, \] respectively. Here and henceforth, when comparing two vectors \( y, y' \in \mathbb{R}^d \), we adopt the convention that \( y \leq y' \) indicates \( y_k \leq y'_k \) for all \( k \in \{1, \ldots, d\} \) and \( y \neq y' \). We use \( y \leq y' \) when equality is allowed; see Subsection 1.5 for notation.

In (2), \( H_k, V_k, \) 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 \( E[\varepsilon_k | x_k^*] = 0 \). 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,
\[
\text{minimize} \quad [ f(x) = (E[(Q_1 x - R_1)^2], \ldots, E[(Q_d x - R_d)^2])^\top ] \quad \text{s.t.} \quad x \in \mathbb{R}^{q \times 1} \quad (4)
\]
where for each data source \( k \), twice the Hessian matrix is \( H_k = E[Q_k^\top Q_k] \in \mathbb{S}_q^+ \), the linear term has coefficient \( V_k = E[Q_k^\top R_k] \in \mathbb{R}^{q \times 1} \), the constant is \( C_k = E[R_k^2] \in \mathbb{R} \), 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 \( k \)th 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 $X_{n,1-\alpha}$ and $Y_{n,1-\alpha}$ such that

$$\lim_{n \to \infty} P\{\mathcal{E} \in X_{n,1-\alpha}\} = 1 - \alpha \quad \text{and} \quad \lim_{n \to \infty} P\{\mathcal{P} \in Y_{n,1-\alpha}\} = 1 - \alpha.$$

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 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 $(H_k, V_k, C_k)$, $k = 1, \ldots, d$, 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 $(H_k, V_k, C_k)$, $k = 1, \ldots, d$ (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 $(H_k, V_k, C_k)$, $k = 1, \ldots, d$, 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 \(E\) and \(P\). 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 \(E\) and \(P\) comes from uncertainty in the finite-dimensional quadratic parameters \((H_k, V_k, C_k), k = 1, \ldots, d\). 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 \(E\) and \(P\) 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 \(E\) and \(P\) 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 naive sample-path estimators \( H_k \) 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 \( \xi \), 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 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 \( y \leq y' \) to indicate \( y_k \leq y'_k \) for all \( k \in \{1, \ldots, d\} \). The notation \( y \leq y' \) implies \( y \leq y' \) and \( y \neq y' \). For any two quantities \( x \) and \( y \), \( x := y \) 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 \( \mathbb{R}_y^d := \{ y \in \mathbb{R}^d : y \geq 0 \} \).

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 \( \mathbb{S}_q^\circ \subset \mathbb{S}_q \) be the set of all symmetric positive semidefinite matrices and \( \mathbb{S}_q^\circ \subset \mathbb{S}_q^{-1} \) be the set of all symmetric positive definite matrices. The Kronecker product of matrices \( A = (a_{ij}) \in \mathbb{R}^{m \times n} \) and \( B \in \mathbb{R}^{p \times q} \), denoted \( A \otimes B \in \mathbb{R}^{mp \times nq} \), is the block matrix defined by entries \( [a_{ij}B] \). For a matrix \( A = (a_1, \ldots, a_n) \in \mathbb{R}^{m \times n} \) where each \( a_i \in \mathbb{R}^{m \times 1} \) for \( i = 1, \ldots, n \), define \( \text{vec}(A) := (a_1^\top, \ldots, a_n^\top)^\top \in \mathbb{R}^{mn \times 1} \) as the vector containing the stacked columns of \( A \). Vectorization is a linear operator [27, p. 82] and can be written as

\[
v = \text{vec}(A) = \sum_{i=1}^n e_{i,n} \otimes A e_{i,n} \in \mathbb{R}^{mn \times 1}
\]

where \( e_{i,n} = (0, \ldots, 0, 1, 0, \ldots, 0)^\top \in \mathbb{R}^{n \times 1} \) is a \( n \)-dimensional vector of zeros with 1 in the \( i \)th place. In (5), \( A e_{i,n} \) represents the \( i \)th 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

\[
A = \text{vec}^{-1}_{m \times n}(v) = \sum_{i=1}^n (e_{i,n}^\top \otimes \mathbb{I}_m) v e_{i,n}^\top.
\]

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

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

\[
\text{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])^\top] = E[XY^\top] - E[X] E[Y]^\top
\]
is the $q \times p$ cross-covariance matrix of $X$ and $Y$. That is, the $(i,j)$th entry of this matrix contains $\text{Cov}(X_i, Y_j)$, 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 $M^{-1} = \sum_{n=0}^{\infty}(I_q - M)^n$. Further, if $S_n := I_q + M + M^2 + \cdots + M^n = (I_q - M)^{-1}(I_q - M^n)$, the sequence $\{S_n, n \geq 1\}$ converges in operator norm [27, p. 92] to $S := (I_q - M)^{-1}$ 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 $S := \{s \in \mathbb{R}_d^q : \sum_k s_k = 1\}$ be a normalized weight set containing nonnegative weights, where $\mathbb{R}_d^q := \{y \in \mathbb{R}^d : y \geq 0\}$. Given $s \in S$, the linear weighted sum scalarization of (1) is

$$\text{minimize } \left\{ \sum_{k=1}^{d} s_k f_k(x) = x^T \left[ \sum_{k=1}^{d} s_k H_k \right] x - 2 \sum_{k=1}^{d} s_k V_k^T x + \sum_{k=1}^{d} s_k C_k \right\}. \quad (8)$$

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

$$x_s^* = \left[ \sum_{k=1}^{d} s_k H_k \right]^{-1} \left[ \sum_{k=1}^{d} s_k V_k \right]. \quad (9)$$

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

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

2.1.1 The efficient map

For each $s \in S$, view $x_s^*$ from (9) as a function of the unknown parameters of the quadratic objectives, $(H_k, V_k, C_k), k = 1, \ldots, d$. Then, given $s \in S$, we define a map $z_s : \mathcal{H} \to \mathbb{R}^q$, $\mathcal{H} \subset \Theta_d$, that takes in quadratic objective parameters of similar form and outputs a point in the decision space, where $\Theta_d$ and $\mathcal{H}$ are appropriate normed spaces. For generality, we
define the related normed spaces $\Theta_1$ and $\Theta_d$ together, in Definition 1. Then, we define $\mathcal{H}$ as an appropriate subset of $\Theta_d$ in Definition 2.

**Definition 1** (the normed space $\Theta_j$, $j \in \{1, d\}$). Let the space of real-valued $q \times q$ symmetric matrices $S_q$ be equipped with the spectral norm, which is the (sub-multiplicative) matrix norm $\|\cdot\|_{S_q}$ induced by the vector norm $\|\cdot\|_2$. Then define the space $\Theta_j$ as $\Theta_j := (S_q \times \mathbb{R}^{q \times 1} \times \mathbb{R}) \times \ldots \times (S_q \times \mathbb{R}^{q \times 1} \times \mathbb{R})$, which is the space $(S_q \times \mathbb{R}^{q \times 1} \times \mathbb{R})$ crossed with itself $j$ times. Let the point $\theta := (h_1, v_1, c_1, \ldots, h_j, v_j, c_j) \in \Theta_j$ and define $\|\theta\|_{\Theta_j} := \sum_{k=1}^j \|h_k\|_{S_q} + \|v_k\|_2 + |c_k|$, 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 $\mathcal{H} := \{\theta \in \Theta_d; h_1, \ldots, h_d \in S_q^{-1}\}$, which is the set of all elements of $\Theta_d$ such that the matrices $h_1, \ldots, h_d$ are real-valued $q \times q$ symmetric invertible matrices.

The normed space $\Theta_d$ 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 $\Theta_j$ for $j = d$ often and for $j = 1$ infrequently, henceforth, we omit the subscript when $j = d$. Thus, $\Theta$ always refers to $\Theta_d$.

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

$$z_s(\theta) := \sum_k s_k h_k^{-1} \sum_k s_k v_k.$$  

(The efficient map ignores the values of $c_1, \ldots, c_d$. However, we include them in $\theta$ 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

$$\theta^* := (H_1, V_1, C_1, \ldots, H_d, V_d, C_d) \in \mathcal{H}$$

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

$$z_s(\theta^*) = [\sum_k s_k H_k]^{-1} \sum_k s_k V_k$$

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

### 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 $s \in S$, the points in the efficient set $x^*_s$ are functions of $\theta^*$ only. Further, applying $f$ involves applying a function of $\theta^*$ only. We exploit this structure to write the following Pareto map. For each $s \in S$, define the Pareto map $g_s : \mathcal{H} \rightarrow \mathbb{R}^d, \mathcal{H} \subset \Theta$ as

$$g_s(\theta) := \begin{pmatrix} \sum_k s_k v_1^T [\sum_k s_k h_k]^{-1} h_1 [\sum_k s_k h_k]^{-1} \sum_k s_k v_k - 2 v_1^T [\sum_k s_k h_k]^{-1} \sum_k s_k v_k + c_1 \\ \vdots \\ \sum_k s_k v_d^T [\sum_k s_k h_k]^{-1} h_d [\sum_k s_k h_k]^{-1} \sum_k s_k v_k - 2 v_d^T [\sum_k s_k h_k]^{-1} \sum_k s_k v_k + c_d \end{pmatrix} = (z_s(\theta) h_1 z_s(\theta) - 2 v_1^T z_s(\theta) + c_1, \ldots, z_s(\theta) h_d z_s(\theta) - 2 v_d^T z_s(\theta) + c_d)^T.$$  

(12)
Then the Pareto set in (10) equals \( \mathcal{P} = \{ g_s(\theta^*) : s \in S \} = \{ f \circ z_s(\theta^*) : s \in S \} \), where \( f \circ z_s : \mathcal{H} \to \mathbb{R}^d \) 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 \( \theta \). That is, given \( \theta \in \Theta \), \( g_s(\theta) \neq f \circ z_s(\theta) \) because \( f \) always applies \( \theta^* \) 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 \( \theta \) 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 \( \theta^* = (H_1, V_1, C_1, \ldots, H_d, V_d, C_d) \). Thus, we need only estimate \( \theta^* \) and use these estimators in the previously defined efficient and Pareto maps. Toward estimating \( \theta^* \), 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 \( (\xi^{(\ell)}_1, \ldots, \xi^{(\ell)}_d) \) 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 \( (H_k(\xi_k^{(\ell)}), V_k(\xi_k^{(\ell)}), C_k(\xi_k^{(\ell)})) \) for \( \ell = 1, \ldots, n_k, n_k > q \) and construct sample-path estimators

\[
\bar{H}_k := \frac{1}{n_k} \sum_{\ell=1}^{n_k} H_k(\xi_k^{(\ell)}), \quad \bar{V}_k := \frac{1}{n_k} \sum_{\ell=1}^{n_k} V_k(\xi_k^{(\ell)}), \quad \bar{C}_k := \frac{1}{n_k} \sum_{\ell=1}^{n_k} C_k(\xi_k^{(\ell)}),
\]

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

\[
\hat{\theta}_n^* := (\bar{H}_1, \bar{V}_1, \bar{C}_1, \ldots, \bar{H}_d, \bar{V}_d, \bar{C}_d),
\]

where \( n := \sum_k n_k \) 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 \( k = 1, \ldots, d \),

\[
H_k(\xi) = Q_k^\top Q_k, \quad V_k(\xi) = Q_k^\top R_k, \quad C_k(\xi) = R_k^2.
\]

Suppose that for each data source \( k = 1, \ldots, d \), we have \( n_k > q \) i.i.d. observations of the data vector \( \xi_k \), denoted \( \xi_k(\ell) = (Q_k^{(\ell)}, R_k^{(\ell)}) = (Q_k^{(\ell)}_1, \ldots, Q_k^{(\ell)}_q, R_k^{(\ell)}), \ell = 1, \ldots, n_k \), where \( Q_k^{(\ell)} \in \mathbb{R}^{1 \times q} \) is the \( \ell \)th random vector of covariates and \( R_k^{(\ell)} \in \mathbb{R} \) is the \( \ell \)th 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

\[
\bar{H}_k = n_k^{-1}(Q_k^\top Q_k), \quad \bar{V}_k = n_k^{-1}(Q_k^\top R_k), \quad \bar{C}_k = n_k^{-1}(R_k^\top R_k).
\]
By construction, $\hat{H}_k$ 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 $\hat{H}_k \in S_q^>$ for each $k$ by assumption, we have $\hat{\theta}_n^* \in \mathcal{H}$. Then, we can write efficient and Pareto set estimators, respectively, as

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

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

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

and the sample-path version of (1) is

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

where $\bar{F}_n : \mathbb{R}^q \to \mathbb{R}^d$ 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, $\hat{\mathcal{P}}_n = \{g_s(\hat{\theta}_n^*): s \in S\} = \{\bar{F}_n \circ z_s(\hat{\theta}_n^*): s \in S\}$, where $\bar{F}_n \circ z_s : \mathcal{H} \to \mathbb{R}^d$ 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 $s \in S$. In Subsection 3.3, we derive a corresponding uniform CLT which is uniform in the scalarization parameter $s \in S$. For convenience, we refer to $\Theta$ 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 $\Theta$. To perform these operations, we define relevant versions of the vectorization operation defined in Subsection 1.5 and its inverse for elements of $\Theta$, 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 $\Theta$.

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 \( a_k := (\text{vec}(h_k)^T, v_k^T, c_k)^T \in \mathbb{R}^{(q^2 + q + 1) \times 1} \). Let \( A = [a_1 \ldots a_j] \in \mathbb{R}^{(q^2 + q + 1) \times j} \) be the concatenated matrix where the \( k \)th column equals \( a_k \). Then the vectorization of \( \theta \) is

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

(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), \( \text{vec}^{-1}_\theta(\tilde{v}) = A = [a_1 \ldots a_j] \) is the concatenated matrix. Let \( B_h := [I_{q^2}, 0_{q^2 \times (q+1)}] \) be a \( q^2 \times (q^2 + q + 1) \) matrix, and let \( B_v := [I_{q} \ 0_{q \times 1}] \) be a \( q \times (q^2 + q + 1) \) matrix. Then

\[
\text{vec}^{-1}_\theta(\tilde{v}) := (\text{vec}^{-1}_{q \times q}(B_hAe_{1,j}), B_vAe_{1,j}, e_{q^2+q+1,q^2+q+1}^T A e_{1,j}, \ldots, \\
\text{vec}^{-1}_{q \times q}(B_hAe_{j,j}), B_vAe_{j,j}, e_{q^2+q+1,q^2+q+1}^T A e_{j,j}).
\]

(20)

Equipped with the vectorization operation and its inverse from (19) and (20), respectively, we obtain a CLT on the random vector \( \text{vec}_\theta(\hat{\theta}_n^*) \) in Lemma 2. Then, to obtain a CLT in the native space \( \Theta \) in Theorem 2, we exploit the fact that \( \text{vec}^{-1}_\Theta \) 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 \( \text{vec}_\Theta(\hat{\theta}_n^*) \). First, for the \( k \)th 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 \( \text{vec}_{\Theta_1}(H_k(\xi), V_k(\xi), C_k(\xi)) \). Define the covariance matrix as

\[
\Omega_k := \text{Var}( \text{vec}_{\Theta_1}(H_k(\xi), V_k(\xi), C_k(\xi)))
\]

(21)

\[
= \text{Var}( \text{vec}_{\Theta_1}(Q_k^T Q_k, Q_k^T R_k, R_k^2)) \in \mathbb{S}_{q^2+q+1},
\]

where the second line provides the equivalent expression under the multiobjective least squares example from Subsection 1.1. Thus, in multiobjective least squares, the \( \Omega_k \) matrix involves the fourth-order moments of the random variables. Due to the symmetry of the matrices involved, each \( \Omega_k \) 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 \( \Omega_k \) matrices are explicitly stated. Then across the objectives, the random vector \( \text{vec}_{\Theta_1}(H_1(\xi), V_1(\xi), C_1(\xi), \ldots, H_d(\xi), V_d(\xi), C_d(\xi)) \) has mean \( \text{vec}_{\Theta_1}(\theta^*) \) and covariance matrix

\[
\Omega := \begin{bmatrix}
\Omega_1 & \ldots & \Omega_{d_1} \\
\vdots & \ddots & \vdots \\
\Omega_{d_1} & \ldots & \Omega_d
\end{bmatrix} \in \mathbb{S}_{d(q^2 + q + 1)},
\]

(22)

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

\[
\Omega_{k,k'} := \text{Cov}(\text{vec}_{\Theta_1}(H_k(\xi), V_k(\xi), C_k(\xi)), \text{vec}_{\Theta_1}(H_{k'}(\xi), V_{k'}(\xi), C_{k'}(\xi)))
\]

(23)

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

With this notation, in Lemma 2, we provide a CLT on \( \text{vec}_{\Theta_1}(\hat{\theta}_n^*) \in \mathbb{R}^{d(q^2 + q + 1)} \).

**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 \( \gamma \) be such that \( \gamma_1 = n_1/n \leq \cdots \leq \gamma_d = n_d/n \), and recall \( n_1 > q \).
Let \((\xi_1^{(\ell)}, \ldots, \xi_d^{(\ell)})\) be i.i.d. copies of the random object \((\xi_1, \ldots, \xi_d)\) for all \(\ell = 1, \ldots, n\gamma_d\), so that even though we do not observe all values,

\[
\text{vec}_\Theta \left( H_1(\xi_1^{(\ell)}), V_1(\xi_1^{(\ell)}), C_1(\xi_1^{(\ell)}), \ldots, H_d(\xi_d^{(\ell)}), V_d(\xi_d^{(\ell)}), C_d(\xi_1^{(\ell)}) \right)
\]

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

\[
\sqrt{n} \left( \text{vec}_\Theta(\hat{\theta}_n^*) - \text{vec}_\Theta(\theta^*) \right) \xrightarrow{d} \sqrt{\Omega} Z,
\]

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

\[
\Omega_\gamma := \begin{bmatrix}
\Omega_{1/\gamma} & \Omega_{12/\gamma} & \ldots & \Omega_{1d/\gamma_d} \\
\Omega_{12/\gamma} & \Omega_{2/\gamma} & \ldots & \Omega_{2d/\gamma_d} \\
\vdots & \vdots & \ddots & \vdots \\
\Omega_{1d/\gamma} & \Omega_{2d/\gamma} & \ldots & \Omega_{d/\gamma_d}
\end{bmatrix} \in \mathbb{S}^{d(q^2+q+1)}.
\]

**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 \(Y_n := \text{vec}_\Theta(\hat{\theta}_n^*) - \text{vec}_\Theta(\theta^*) = \text{vec}_\Theta(\hat{\theta}_n^* - \theta^*)\). By setting unobserved data equal to zero, write \(Y_n\) as the sum

\[
\frac{1}{n} \sum_{k=1}^{d} \sum_{\ell=n\gamma_{k-1}+1}^{n\gamma_k} \text{vec}_\Theta \left( 0_{q \times q}, 0_{q \times 1}, \ldots, \frac{H_k(\xi_k^{(\ell)}) - H_k}{\gamma_k}, \frac{V_k(\xi_k^{(\ell)}) - V_k}{\gamma_k}, \frac{C_k(\xi_k^{(\ell)}) - C_k}{\gamma_k}, \ldots, \frac{H_d(\xi_d^{(\ell)}) - H_d}{\gamma_d}, V_d(\xi_d^{(\ell)}) - V_d, C_d(\xi_d^{(\ell)}) - C_d \right)
\]

and denote the summand as \(Y_k^{(\ell)} \in \mathbb{R}^{d(q^2+q+1) \times 1}\). Since \((\xi_1^{(\ell)}, \ldots, \xi_d^{(\ell)})\) are i.i.d. for all \(\ell = 1, \ldots, n\gamma_d\), then for each \(k\), \(Y_k^{(1)}\) are i.i.d. for all \(\ell = n\gamma_{k-1}+1, \ldots, n\gamma_k\). 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

\[
\varphi_{\sqrt{n} Y_n}(t) = E[\exp(it^\top \sqrt{n} Y_n)] = \prod_{k=1}^{d} \text{E} \left[ \exp \left( \frac{t^\top}{\sqrt{n}} Y_{k,*} \right) \right]^{n(\gamma_k - \gamma_{k-1})}
\]

\[
= \prod_{k \in \{k' : \gamma_{k'} > \gamma_{k'-1}\}} \text{E} \left[ \exp \left( \frac{t^\top (\sqrt{n} \gamma_k - \gamma_{k-1})}{\sqrt{n} (\gamma_k - \gamma_{k-1})} Y_{k,*} \right) \right]^{n(\gamma_k - \gamma_{k-1})}.
\]

(26)
Take a limit on both sides of (26) to obtain
\[
\lim_{n \to \infty} \varphi_n(t) = \prod_{k \in \{k': \gamma_k' > \gamma_k-1\}} \lim_{n \to \infty} \left( \varphi_{Y_{k,*}} \left( \frac{t\sqrt{\gamma_k - \gamma_k-1}}{\sqrt{n(\gamma_k - \gamma_k-1)}} \right) \right)^{n(\gamma_k - \gamma_k-1)}
\]

\[
= \prod_{k \in \{k': \gamma_k' > \gamma_k-1\}} \exp \left\{ -(1/2)t^T(\gamma_k - \gamma_k-1)\Sigma_{Y_{k,*}}t \right\}
\]

(27)

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,

\[
\Sigma_{Y_{k,*}} = \text{Var}(Y_{k,*}) = \begin{bmatrix}
\Omega_0 & \ldots & \Omega_0 \\
\vdots & \ddots & \vdots \\
\Omega_0 & \ldots & \Omega_0
\end{bmatrix}
\]

are positive semidefinite; there are \( k - 1 \) of the \( \Omega_0 \)'s along the diagonal of \( \Sigma_{Y_{k,*}} \). Noting that \( \Sigma_{Y_{k,*}} \) is positive semidefinite when \( k = 1 \), evaluating the sum yields the expression for \( \Omega_\gamma \) 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 \( \bar{\theta}_n^* \). Next, we convert this statement to a statement regarding the convergence in distribution of \( \bar{\theta}_n^* \) 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 \( \tilde{Z} \in \Theta \), defined by

\[
\tilde{Z} = (\tilde{Z}_{h,1}, \tilde{Z}_{v,1}, \tilde{Z}_{c,1}, \ldots, \tilde{Z}_{h,d}, \tilde{Z}_{v,d}, \tilde{Z}_{c,d}) := \text{vec}^{-1}_\Theta \left( \sqrt{\Omega_\gamma} Z \right),
\]

(29)

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

\[
\sqrt{\Omega_\gamma} Z = \begin{bmatrix} w_1^T Z, \ldots, w_{d(q^2+q+1)}^T Z \end{bmatrix}^T,
\]

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

\[
\tilde{Z}_{h,k} = \begin{bmatrix} w_{(k-1)(q^2+q+1)+1}^T Z & \ldots & w_{(k-1)(q^2+q+1)+q^2+q+1}^T Z \\ \vdots & \ddots & \vdots \\ w_{(k-1)(q^2+q+1)+q}^T Z & \ldots & w_{(k-1)(q^2+q+1)+q^2}^T Z \end{bmatrix},
\]

\[
\tilde{Z}_{v,k} = \begin{bmatrix} w_{k(q^2+q+1)-(q+1)+1}^T Z & \ldots & w_{k(q^2+q+1)-1}^T Z \end{bmatrix}^T,
\]

\[
\tilde{Z}_{c,k} = w_{k(q^2+q+1)}^T Z.
\]
Thus, each $\tilde{Z}_{h,k}, \tilde{Z}_{v,k}$, and $\tilde{Z}_{c,k}$ 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 $\Omega_\gamma$ preserves the symmetry of the matrices $\tilde{Z}_{h,k}, k = 1, \ldots, d$. We use this notation to write the CLT in the native space.

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

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

**Proof.** Since $\text{vec}^{-1}_\Theta$ 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,

$$\sqrt{n}(\hat{\theta}_n^* - \theta^*) = \sqrt{n}\left(\text{vec}^{-1}_\Theta(\text{vec}_\Theta(\hat{\theta}_n^*)) - \text{vec}^{-1}_\Theta(\text{vec}_\Theta(\theta^*))\right)$$

$$= \text{vec}^{-1}_\Theta\left(\sqrt{n}\left(\text{vec}(\hat{\theta}_n^*) - \text{vec}(\theta^*)\right)\right) \xrightarrow{d} \text{vec}^{-1}_\Theta(\sqrt{\Omega_\gamma Z}),$$

which implies the result. \qed

### 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 $s \in S$ and $\theta^* \in \mathcal{H}, \mathcal{H} \subset \Theta$ be given, and recall that for $\theta \in \mathcal{H},$

$$\theta - \theta^* = (h_1 - H_1, v_1 - V_1, c_1 - C_1, \ldots, h_d - H_d, v_d - V_d, c_d - C_d) \in \Theta.$$  

Then the following hold:

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

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

Further, $z_s'(\theta^*, \theta - \theta^*) = \left[\sum_k s_k H_k\right]^{-1} \left(\sum_k s_k [(v_k - V_k) - (h_k - H_k)z_s(\theta^*)]\right).$

2. The Pareto map $g_s: \mathcal{H} \to \mathbb{R}^d$ in (12) is Fréchet differentiable at $\theta^*$. That is, there exists a bounded linear operator $g_s'(\theta^*, \cdot): \Theta \to \mathbb{R}^d$ with

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

Further, $g_s'(\theta^*, \theta - \theta^*) = (g_{s,1}'(\theta^*, \theta - \theta^*), g_{s,2}'(\theta^*, \theta - \theta^*), \ldots, g_{s,d}'(\theta^*, \theta - \theta^*))^\top$, where for each $k \in \{1, \ldots, d\}$,

$$g_{s,k}'(\theta^*, \theta - \theta^*) = z_s(\theta^*)^\top(h_k - H_k)z_s(\theta^*) - 2(v_k - V_k)^\top z_s(\theta^*) + (c_k - C_k)$$

$$+ 2[z_s(\theta^*)^\top H_k - V_k^\top] z_s'(\theta^*, \theta - \theta^*).$$
Proof. First, we show Part 1. Recall that \( z_s \) is a map from the normed space \( \Theta \) to \( \mathbb{R}^q \). Use the definition of the Gateaux derivative in [29, p. 171] to obtain

\[
z'_s(\theta^*; \theta - \theta^*) := \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left( z_s(\theta^* + \epsilon(\theta - \theta^*)) - z_s(\theta^*) \right).
\]  

(31)

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

\[
z_s(\theta^* + \epsilon(\theta - \theta^*)) = \left[ \sum_k s_k H_k + \epsilon \sum_k s_k (h_k - H_k) \right]^{-1} \left[ \sum_k s_k V_k + \epsilon \sum_k s_k (v_k - V_k) \right]
\]

\[
= \left[ \frac{i + \epsilon \sum_k s_k H_k}{\sum_k s_k (h_k - H_k)} \right]^{-1} \left[ \sum_k s_k V_k + \epsilon \sum_k s_k (v_k - V_k) \right].
\]

For notational convenience, define the maps \( T: \Theta \to \mathbb{S}_q \) and \( T: \Theta \to \mathbb{R}^q \) by

\[
T(\theta - \theta^*) := \left[ \sum_k s_k H_k \right]^{-1} \sum_k s_k (h_k - H_k),
\]

(32)

\[
\tilde{T}(\theta - \theta^*) := \left[ \sum_k s_k H_k \right]^{-1} \sum_k s_k (v_k - V_k),
\]

(33)

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

\[
z_s(\theta^* + \epsilon(\theta - \theta^*)) = \left[ i + \epsilon T(\theta - \theta^*) \right]^{-1} \left[ z_s(\theta^*) + \epsilon \tilde{T}(\theta - \theta^*) \right]
\]

\[
= \left[ i - \epsilon T(\theta - \theta^*) + (\epsilon T(\theta - \theta^*))^2 + o(\epsilon^2) \right] \left[ z_s(\theta^*) + \epsilon \tilde{T}(\theta - \theta^*) \right]
\]

(34)

\[
= z_s(\theta^*) + \epsilon \tilde{T}(\theta - \theta^*) - \epsilon T(\theta - \theta^*) z_s(\theta^*) - \epsilon^2 T(\theta - \theta^*) \tilde{T}(\theta - \theta^*)
\]

\[
+ (\epsilon T(\theta - \theta^*))^2 z_s(\theta^*) + \epsilon^3 (T(\theta - \theta^*))^2 \tilde{T}(\theta - \theta^*) + o(\epsilon^2),
\]

(35)

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

\[
z'_s(\theta^*; \theta - \theta^*) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left( \epsilon \tilde{T}(\theta - \theta^*) - \epsilon T(\theta - \theta^*) z_s(\theta^*) + o(\epsilon) \right)
\]

\[
= \tilde{T}(\theta - \theta^*) - T(\theta - \theta^*) z_s(\theta^*).
\]

(36)

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

\[
||\theta - \theta^*||_\Theta = \sum_{k=1}^d ||h_k - H_k||_{\mathbb{S}_q} + ||v_k - V_k||_2 + |c_k - C_k| = 1
\]

implies that \( ||h_k - H_k||_{\mathbb{S}_q} \leq 1 \) and \( ||v_k - V_k||_2 \leq 1 \) for all \( k \), and since \( \sum_k s_k = 1, s_k \geq 0 \) for
all $k$, we have that the operator norm of $T$ [27, p. 92] is
\[
\|T\|_{op} = \sup \left\{ \| \sum_k s_k H_k \|^{-1} \sum_k s_k (h_k - H_k) \|_{S_\theta} : \sigma \in \Theta, \| T - \Theta \|_{\Theta} = 1 \right\}
\]
\[
\leq \sup \left\{ \| \sum_k s_k H_k \|^{-1} \sum_k s_k \| h_k - H_k \|_{S_\theta} : \sigma \in \Theta, \| T - \Theta \|_{\Theta} = 1 \right\}
\]
\[
\leq \| \sum_k s_k H_k \|^{-1}_{S_\theta}.
\]
Similar logic follows for $\tilde{T}$. Then since $\| \sum_k s_k H_k \|^{-1}_{S_\theta}$ exists (recall that the inverse of a symmetric matrix is also symmetric), we have
\[
\max \{ \| T \|_{op}, \| \tilde{T} \|_{op} \} \leq \| \sum_k s_k H_k \|^{-1}_{S_\theta} < \infty.
\] (37)
Therefore, both $T$ and $\tilde{T}$ are bounded linear operators of the increment $\sigma - \sigma^*$.

Since $T$ and $\tilde{T}$ are bounded linear operators and (37) holds, (36) implies that $z_s'((\star, \cdot))$ is also a bounded linear operator with respect to the increment. To demonstrate that $z_s'((\star, \cdot))$ is a Fréchet derivative satisfying its definition, it is sufficient to show that $z_s(\sigma) - z_s(\sigma^*) = z_s'(\sigma^*; \sigma - \sigma^*) + o(\| \sigma - \sigma^* \|_{\Theta})$. Use calculations similar to those leading up to (35) to see that for small enough $\sigma - \sigma^*$,
\[
z_s(\sigma) - z_s(\sigma^*) = z_s(\sigma^* + (\sigma - \sigma^*)) - z_s(\sigma^*)
\]
\[
= z_s'(\sigma^*; \sigma - \sigma^*) - T(\sigma - \sigma^*) \tilde{T}(\sigma - \sigma^*) + (T(\sigma - \sigma^*))^2 z_s(\sigma^*)
\]
\[
+ (T(\sigma - \sigma^*))^2 \tilde{T}(\sigma - \sigma^*) + o(\| T(\sigma - \sigma^*) \|_{S_\theta})
\]
\[
= z_s'(\sigma^*; \sigma - \sigma^*) + o(\| \sigma - \sigma^* \|_{\Theta}).
\]

For Part 2, notice again that $g_{s,k} := \lim_{\epsilon \to 0} \frac{1}{\epsilon} (g_{s,k}(\sigma^* + \epsilon(\sigma - \sigma^*)) - g_{s,k}(\sigma^*))$ 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 $\Theta$-valued sequence $\{\hat{\sigma}_n\}$, appropriately scaled, converges weakly to the $\Theta$-valued analog of a normal random object, $\hat{\sigma}$. In analogy with the delta method [6, 28], given $s \in S$, it seems natural that the respective $\mathbb{R}^q$-valued and $\mathbb{R}^d$-valued sequences of estimated efficient and Pareto points $\{z_s(\hat{\sigma}_n^*)\}$ and $\{g_s(\hat{\sigma}_n^*)\}$, 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 $\hat{\sigma} \in \Theta$ is defined in (29). Suppose $\Theta^* \in \mathcal{H}$ and $\hat{\sigma}_n^* \in \mathcal{H}$ for each $n$, where $\mathcal{H} \subset \Theta$. Then for each parameter value $s \in S$,

1. $\sqrt{n} \left[ z_s(\hat{\sigma}_n^*) - z_s(\sigma^*) \right] \overset{d}{\to} z_s'((\sigma^*, \cdot))$, where the right side is a $q$-dimensional mean-zero multivariate normal random vector
\[
z_s'(\sigma^*, \cdot) = \left[ \sum_k s_k H_k \right]^{-1} \left( \sum_k s_k [\hat{\sigma}_{s,k} - \hat{\sigma}_{s,k} z_s(\sigma^*)] \right)
\] (38)

with covariance matrix $\text{Var}(z_s'(\sigma^*, \cdot)) = \text{E} \left[ z_s'(\sigma^*, \cdot)(z_s'(\sigma^*, \cdot))' \right] \in \mathbb{S}^q$.

2. $\sqrt{n} \left[ g_s(\hat{\sigma}_n^*) - g_s(\sigma^*) \right] \overset{d}{\to} g_s'(\sigma^*, \cdot)$, where the right side is a $d$-dimensional multivariate
Then using the upper bound in (37), there exists a constant $T$ such that

$$T \leq \left[ \mathbb{1}_{q} - T(\hat{\theta}_{n}^{*} - \theta^{*}) + (T(\hat{\theta}_{n}^{*} - \theta^{*}))^{2} - (T(\hat{\theta}_{n}^{*} - \theta^{*}))^{3} + \cdots \right] \left[ z_{s}(\theta^{*}) + T(\hat{\theta}_{n}^{*} - \theta^{*}) \right]$$

and covariance matrix $\text{Var}(g'(\theta^{*}, \tilde{Z})) = E [g'(\theta^{*}, \tilde{Z})(g'(\theta^{*}, \tilde{Z})^{T}] \in S_{d}^{\geq}$.

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

$$z_{s}(\hat{\theta}_{n}^{*}) = z_{s}(\theta^{*} + (\hat{\theta}_{n}^{*} - \theta^{*})) = \left[ \mathbb{1}_{q} - T(\hat{\theta}_{n}^{*} - \theta^{*}) + (T(\hat{\theta}_{n}^{*} - \theta^{*}))^{2} - (T(\hat{\theta}_{n}^{*} - \theta^{*}))^{3} + \cdots \right] \left[ z_{s}(\theta^{*}) + T(\hat{\theta}_{n}^{*} - \theta^{*}) \right]$$

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

$$U_{n} := -T(\hat{\theta}_{n}^{*} - \theta^{*})\tilde{T}(\hat{\theta}_{n}^{*} - \theta^{*}) + (T(\hat{\theta}_{n}^{*} - \theta^{*}))^{2}\left[ \mathbb{1}_{q} + T(\hat{\theta}_{n}^{*} - \theta^{*}) \right]^{-1} \left[ z_{s}(\theta^{*}) + T(\hat{\theta}_{n}^{*} - \theta^{*}) \right].$$

Then it follows that

$$\sqrt{n}\left[ z_{s}(\hat{\theta}_{n}^{*}) - z_{s}(\theta^{*}) \right] = \sqrt{n}\left[ z_{s}(\theta^{*}; \hat{\theta}_{n}^{*} - \theta^{*}) + U_{n} \right] = z_{s}(\theta^{*}; \sqrt{n}(\hat{\theta}_{n}^{*} - \theta^{*})) + \sqrt{n} U_{n}.$$  \hfill (39)

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

$$\|T(\hat{\theta}_{n}^{*} - \theta^{*})\tilde{T}(\hat{\theta}_{n}^{*} - \theta^{*})\| \leq \|T\|_{op}\|\tilde{T}\|_{op}\|\hat{\theta}_{n}^{*} - \theta^{*}\|_{S_{q}},$$

$$\|(T(\hat{\theta}_{n}^{*} - \theta^{*}))^{2}\|_{S_{q}} \leq \|T\|_{op}^{2}\|\hat{\theta}_{n}^{*} - \theta^{*}\|_{S_{q}},$$

$$\|z_{s}(\theta^{*}) + T(\hat{\theta}_{n}^{*} - \theta^{*})\| \leq \|z_{s}(\theta^{*})\| + \|\tilde{T}\|_{op}\|\hat{\theta}_{n}^{*} - \theta^{*}\|_{S_{q}}.$$  \hfill (40)

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

$$\|\sqrt{n} U_{n}\| \leq c\sqrt{n}\|\hat{\theta}_{n}^{*} - \theta^{*}\|_{S_{q}} \left( 1 + c\|\mathbb{1}_{q} + T(\hat{\theta}_{n}^{*} - \theta^{*})\|_{S_{q}}^{-1}\right) [1 + \|\hat{\theta}_{n}^{*} - \theta^{*}\|_{S_{q}}] \geq 1.$$  \hfill (41)

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

$$\left[ \mathbb{1}_{q} + T(\hat{\theta}_{n}^{*} - \theta^{*}) \right]^{-1} \overset{p}{\rightarrow} 1,$$

by [39, p. 24], and also $\|\hat{\theta}_{n}^{*} - \theta^{*}\|_{\Theta} \overset{p}{\rightarrow} 0$ and $\sqrt{n}\|\hat{\theta}_{n}^{*} - \theta^{*}\|_{S_{q}} \overset{p}{\rightarrow} 0$. Employing these results and (41) when applying Slutsky’s theorem [39, pp. 19] to $\sqrt{n} U_{n}$, we have

$$\sqrt{n} U_{n} \overset{p}{\rightarrow} 0.$$

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. \qed
3.3 Uniform CLTs on the estimated efficient and Pareto sets

Theorem 3 asserts convergence in distribution for a fixed \( s \in S \). Next, we demonstrate through Theorem 4 that such weak convergence happens uniformly in \( s \in S \), 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 \( C^q(S) \)-valued, \( C^d(S) \)-valued random fields

\[
\tilde{z}'(\theta^*, \tilde{Z}) := \{z'_s(\theta^*, \tilde{Z}), s \in S\}, \quad g'(\theta^*, \tilde{Z}) := \{g'_s(\theta^*, \tilde{Z}), s \in S\},
\]

(42)

are mean-zero Gaussian processes with covariance functions

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

for all \((s, t) \in S \times S\), respectively.

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

\[
M_v(s) := \sum_k s_k H_k^{-1}[I_q \otimes s^T] \in \mathbb{R}^{q \times dq},
\]

\[
M_h(s) := \sum_k s_k H_k^{-1}[I_q \otimes z_s(\theta^*)][I_q \otimes s^T] \in \mathbb{R}^{q \times dq^2}.
\]

Further, define the mean-zero random variables

\[
W_v := \text{vec}\left(\begin{array}{cc}
\tilde{Z}^{\top}_{v,1} \\
\vdots \\
\tilde{Z}^{\top}_{v,d}
\end{array}\right) \in \mathbb{R}^{dq \times 1} \quad \text{and} \quad W_h := \text{vec}\left(\begin{array}{c}
\text{vec}(\tilde{Z}^{\top}_{h,1}) \\
\vdots \\
\text{vec}(\tilde{Z}^{\top}_{h,d})
\end{array}\right) \in \mathbb{R}^{dq^2 \times 1}.
\]

Then, \( z'_s(\theta^*, \tilde{Z}) \) from (38) equals \( z'_s(\theta^*, \tilde{Z}) = M_v(s)W_v - M_h(s)W_h \). 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 \( g'(\theta^*, \tilde{Z}) \) is a multivariate Gaussian process. A similar result holds for \( g'(\theta^*, \tilde{Z}) \).

Now, we employ Lemma 4 toward proving the following Theorem 4, which asserts that the weak convergence happens uniformly in \( s \in S \).

**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 \( C^q(S) \)-valued and \( C^d(S) \)-valued sequences

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

satisfy \( \delta_{\varepsilon,n} \overset{d}{\longrightarrow} \tilde{z}'(\theta^*, \tilde{Z}) \) and \( \delta_{\gamma,n} \overset{d}{\longrightarrow} g'(\theta^*, \tilde{Z}) \), respectively, where the limits \( \tilde{z}'(\theta^*, \tilde{Z}) \) and \( g'(\theta^*, \tilde{Z}) \) are mean-zero Gaussian processes identified in Lemma 4.
Proof. From Prokhorov’s theorem [7, p. 57ff.], the proof that \( \delta_{\varepsilon,n} \xrightarrow{d} \mathcal{Y}(\theta^*, \tilde{Z}) \) will be complete if we demonstrate the following: (a) the finite-dimensional distributions of \( \delta_{\varepsilon,n} \) converge to the finite-dimensional distributions of \( \mathcal{Y}(\theta^*, \tilde{Z}) \); and (b) there is no “escape of probability mass” from the sequence \{\( \delta_{\varepsilon,n}, n \geq 1 \); that is, the sequence \{\( \delta_{\varepsilon,n}, n \geq 1 \) is tight.

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

\[
\sum_{j=1}^{r} \lambda_j \sqrt{n} \left[ z_{s^{(j)}}(\hat{\theta}_n^*) - z_{s^{(j)}}(\theta^*) \right] = \sum_{j=1}^{r} \lambda_j \sqrt{n} \left[ z'_{s^{(j)}}(\theta^*; \hat{\theta}_n^* - \theta^*) + U_{s^{(j)},n} \right]
= \sum_{j=1}^{r} \lambda_j z'_{s^{(j)}}(\theta^*; \sqrt{n}(\hat{\theta}_n^* - \theta^*)) + \sqrt{n} U_{s^{(j)},n}. \tag{43}
\]

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

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

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

\[
\sum_{j=1}^{r} \lambda_j \sqrt{n} \left[ z_{s^{(j)}}(\hat{\theta}_n^*) - z_{s^{(j)}}(\theta^*) \right] \xrightarrow{d} \sum_{j=1}^{r} \lambda_j z'_{s^{(j)}}(\theta^*, \mathcal{Z}). \tag{45}
\]

Since choice of \( \tilde{S} \) 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

\[
P\{\delta_{\varepsilon,n} \in K(\epsilon)\} \geq 1 - \epsilon \quad \text{for all } n. \tag{46}
\]

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 \( n \),

\[
P\{\mathcal{Y}(\theta^*, \sqrt{n}(\hat{\theta}_n^* - \theta^*)) \in K(\epsilon)\} \geq 1 - \epsilon. \tag{47}
\]

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 \( n \),

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

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

\[
\mathcal{Y}(\theta^*, \sqrt{n}(\hat{\theta}_n^* - \theta^*)) \subset B(X_{1,n}, L \operatorname{diam}(\mathcal{S})), \tag{49}
\]

where, for convenience, we choose \( s^{(1)} = (s^{(1)}_1, \ldots, s^{(1)}_d) := (1, 0, \ldots, 0) \) and

\[
X_{1,n} := z'_{s^{(1)}}(\theta^*, \sqrt{n}(\hat{\theta}_n^* - \theta^*)) = \sqrt{n} H^{-1}_1 \left[ (\hat{V}_1 - V_1) - (H_1 - H_1)(H_1^{-1}V_1) \right].
\]

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
\[
\mathbf{z}'_s(\theta^*, \sqrt{n}(\hat{\theta}^*_n - \theta^*)) = \left[ \sum_{k} s_k H_k \right]^{-1} \sum_{k} s_k [\sqrt{n}(V_k - V_k) - \sqrt{n}(H_k - H_k)] z_s(\theta^*) .
\]
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
\[
\| \mathbf{z}'_s(\theta^*, \sqrt{n}(\hat{\theta}^*_n - \theta^*)) - \mathbf{z}'_t(\theta^*, \sqrt{n}(\hat{\theta}^*_n - \theta^*)) \| = \| A_s^{-1} B_s - A_t^{-1} B_t \| \\
= \| A_s^{-1} (B_s - B_t) + (A_s^{-1} - A_t^{-1}) B_t \| \\
\leq \| A_s^{-1} \|_{\mathcal{S}_q} \| B_s - B_t \| + \| A_s^{-1} - A_t^{-1} \|_{\mathcal{S}_q} \| B_t \|. \quad (50)
\]
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 \( \lambda_* := \min_k \lambda_1(H_k) \), \( \lambda^* := \max_k \lambda_q(H_k) \), and \( \kappa := \lambda^*/\lambda_* < \infty \).

First, for the terms concerning \( A_s \) and \( A_t \) in (50), we have
\[
\| A_s^{-1} \|_{\mathcal{S}_q} = \lambda_q(A_s^{-1}) = (\lambda_1(A_s))^{-1} \leq (\min_k \lambda_1(H_k))^{-1} = \lambda_*^{-1} < \infty . \quad (51)
\]
Now, consider \( \| A_s^{-1} - A_t^{-1} \| \). Rewriting \( A_t^{-1} \) and using Lemma 1 for small enough \( \delta \),
\[
A_s^{-1} - A_t^{-1} = A_s^{-1} - (\sum_{k} s_k H_k + \delta_k H_k)^{-1} = [I_q - (I_q + A_s^{-1} \sum_k \delta_k H_k)^{-1}] A_s^{-1} \\
= [A_s^{-1} \sum_k \delta_k H_k - (A_s^{-1} \sum_k \delta_k H_k)^2 + (A_s^{-1} \sum_k \delta_k H_k)^3 + \ldots] A_s^{-1} .
\]
Choose \( \| s - t \| < (2\kappa d)^{-1} \). Then using (51), since \( |\delta_k| = |s_k - t_k| \leq \| s - t \| \),
\[
\| A_s^{-1} - A_t^{-1} \| \leq \lambda_*^{-1} \| \sum_k \delta_k H_k \| + (\lambda_*^{-1} \| \sum_k \delta_k H_k \|)^2 + (\lambda_*^{-1} \| \sum_k \delta_k H_k \|)^3 + \ldots \\
\leq \lambda_*^{-1} [\kappa \| s - t \| + (\kappa \| s - t \|)^2 + (\kappa \| s - t \|)^3 + \ldots] \\
= (\lambda_*^{-1} \kappa \| s - t \|) [1 + (\kappa \| s - t \|)^2 + (\kappa \| s - t \|)^3 + \ldots] \\
= (\lambda_*^{-1} \kappa \| s - t \|)^{-1} \leq 2\lambda_*^{-1} \kappa \| s - t \| . \quad (52)
\]
Similarly, for the terms concerning \( B_s \) and \( B_t \) in (50), we have
\[
\| B_t \| \leq \sup_{t \in \mathcal{S}} \| B_t \| = \sup_{t \in \mathcal{S}} \| \sum_k t_k [\Delta V_k - \Delta H_k z_t(\theta^*)] \| \\
\leq \sup_{t \in \mathcal{S}} \| \sum_k t_k [\Delta V_k - \Delta H_k A_t^{-1} \sum_k t_k V_k] \| \\
\leq d \max_k \{ \| \Delta V_k \| + \| \Delta H_k \| \lambda_*^{-1} d(\max_k \| V_k \|) \} =: L_B . \quad (53)
\]
Finally, note that for \( |\delta| \leq \| s - t \| \), we have
\[
\| z_t(\theta^*) - z_s(\theta^*) \| = \| (A_t^{-1} - A_s^{-1}) \sum_k s_k V_k + A_t^{-1} \sum_k \delta_k V_k \| \\
\leq (2\lambda_*^{-1} \kappa \| s - t \|)(d \max_k \| V_k \|) + \lambda_*^{-1} \| s - t \| (d \max_k \| V_k \|) \\
= (2\kappa d + 1)\lambda_*^{-1} (d \max_k \| V_k \|) \| s - t \| ,
\]
which, together with (53), implies
\[
\| B_s - B_t \| = \| \sum_k \delta_k (\Delta_{h,k} z_i(\theta^*) - \Delta_{v,k}) + \sum_k s_k \Delta_{H,k} (z_i(\theta^*) - z_s(\theta^*)) \| \\
\leq L_B \| s - t \| + (d \max_k \| \Delta_{h,k} \|)(2kd + 1) \lambda_k^{-1}(d \max_k \| V_k \|) \| s - t \|. \tag{54}
\]
Then, use (51), (54), (52), and (53), respectively in (50) to see that the inequality holds for \( \| s - t \| < (2kd)^{-1} \). 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 \( \delta_{\gamma,n} \xrightarrow{d} g'(\theta^*, \tilde{Z}) \) since it follows along identical lines. \( \square \)

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 \( \theta^* \). Thus, we propose a different method for confidence region construction. First, in Subsection 4.1, we construct a confidence region on \( \theta^* \). 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, \( \left\{ \theta \in \Theta: \sqrt{n} \| \text{vec}_{\Theta}(\theta^*) - \text{vec}_{\Theta}(\theta) \| \leq p_{1-\alpha} \left( \| \sqrt{\Omega_\gamma} Z \| \right) \right\} \), where \( p_{1-\alpha}(\cdot) \) is a function that returns the \( 1 - \alpha \) quantile of its argument for \( \alpha \in (0,1) \). We preserve \( \Omega_\gamma \) 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, \( \Omega_\gamma \) 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 \( h_k \in \mathcal{S}_q \) 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 \( h_k^u \in \mathbb{R}^{q(q+1)/2} \). Thus, letting \( h_{kij} \) be the \((i,j)\)th element of \( h_k \),
\[
h_{k}^{uT} = (h_{k11}, h_{k12}, h_{k22}, h_{k13}, \ldots, h_{k1d}, \ldots, h_{kdd})^T. \tag{55}
\]
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 \( \beta_k := (h_k^u, v_k^T, c_k)^T \in \mathbb{R}^{(q(q+1)/2+q+1)\times 1} \). For notational convenience, let
\[
b := q(q + 1)/2 + q + 1, \tag{56}
\]
and in analogy to $\text{vec}_d(\hat{\theta}_n^*)$ and $\text{vec}_o(\theta^*)$, define
\[
\hat{\beta}_n^* := (\hat{H}_1^\text{HT}, \hat{V}_1^\text{T}, \hat{C}_1, \ldots, \hat{H}_d^\text{HT}, \hat{V}_d^\text{T}, \hat{C}_d)^\text{T}, \quad \beta^* := (H_1^\text{HT}, V_1^\text{T}, C_1, \ldots, H_d^\text{HT}, V_d^\text{T}, C_d)^\text{T} \tag{57}
\]
where $\hat{\beta}_n^*, \beta^* \in \mathbb{R}^{db \times 1}$ and, using the same notation as (55),
\[
\hat{H}_k^\text{HT} := (\hat{H}_{k11}, \hat{H}_{k12}, \hat{H}_{k22}, \hat{H}_{k13}, \ldots, \hat{H}_{k1d}, \ldots, \hat{H}_{kdd})^\text{T}.
\]

Now, construct the positive definite $\Omega^\gamma_k$ matrix in the same way as $\Omega^\gamma$ in (25), except replace each $\Omega_k$ with $\Omega_k^\alpha$ and each cross-covariance matrix $\Omega_{k,k'}$ with $\Omega_{k,k'}^\alpha$, where
\[
\Omega_k^\alpha := \text{Var} \left( (H_k^\text{H}(\xi)^\text{T}, V_k(\xi)^\text{T}, C_k(\xi))^\text{T} \right) \in \mathbb{S}_+^k,
\]
\[
\Omega_{k,k'}^\alpha := \text{Cov} \left( (H_k^\text{H}(\xi)^\text{T}, V_k(\xi)^\text{T}, C_k(\xi))^\text{T}, (H_{k'}^\text{H}(\xi)^\text{T}, V_{k'}(\xi)^\text{T}, C_{k'}(\xi))^\text{T} \right).
\]

Then, Lemma 2 implies that
\[
\sqrt{n}(\Omega^\gamma_k)^{-1/2}(\hat{\beta}_n^* - \beta^*) \overset{\text{d}}{\rightarrow} Z \tag{58}
\]
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 $\beta^*$ can be constructed as the hyperellipsoid
\[
\mathcal{R}_{n,1-\alpha} := \{ \beta \in \mathbb{R}^{db \times 1} : n(\hat{\beta}_n^* - \beta)^\text{T}(\Omega^\gamma_n)^{-1}(\hat{\beta}_n^* - \beta) \leq \chi^2_{db,1-\alpha} \},
\]
where $\chi^2_{db,1-\alpha}$ denotes the $1 - \alpha$ quantile of a $\chi^2$ 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 $\Theta$. In analogy to the inverse vectorization defined in (20), let $\text{vec}^{-1}_{\mathbb{R}^{db},\Theta} : \mathbb{R}^{db} \rightarrow \Theta$ denote the (linear) transformation from $\beta$ back into the $\theta$ space. For each $\beta \in \mathbb{R}^{db}$ such that $\text{vec}^{-1}_{\mathbb{R}^{db},\Theta}(\beta) \in \Theta$, define
\[
\varepsilon(\beta) := \{ z_s(\text{vec}^{-1}_{\mathbb{R}^{db},\Theta}(\beta)), s \in S \}, \quad p(\beta) := \{ g_s(\text{vec}^{-1}_{\mathbb{R}^{db},\Theta}(\beta)), s \in S \}.
\]

Then, set
\[
\mathcal{X}_{n,1-\alpha} := \{ \varepsilon(\beta) : \beta \in \mathcal{R}_{n,1-\alpha} \}, \quad \mathcal{Y}_{n,1-\alpha} := \{ p(\beta) : \beta \in \mathcal{R}_{n,1-\alpha} \}. \tag{60}
\]

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,
\[
\lim_{n \rightarrow \infty} P\{ \varepsilon \in \mathcal{X}_{n,1-\alpha} \} = 1 - \alpha, \quad \lim_{n \rightarrow \infty} P\{ \mathcal{P} \in \mathcal{Y}_{n,1-\alpha} \} = 1 - \alpha.
\]

**Proof.** First, recall $\mathcal{E} = \varepsilon(\beta^*)$ and $\mathcal{P} = p(\beta^*)$ for $\beta^*$ in (57). From (60), the proof follows
We implement an illustrative example of constructing 95% confidence regions for the least squares problem in (4) as described in Subsection 2.2.2, the sample-path version of the bi-objective least-squares problem in (4) is

\[ \text{minimize } \left( \frac{1}{n_1} \| Q_1 x - R_1 \|_2^2, \frac{1}{n_2} \| Q_2 x - R_2 \|_2^2 \right)^T. \]

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).

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

\[
\begin{align*}
\mu_1 &= \left[ \begin{array}{c} \mu_{Q1} \\ \mu_{R1} \end{array} \right] = \left[ \begin{array}{c} 1 \\ 2 \\ 6 \end{array} \right], \quad \Sigma_1 = \left[ \begin{array}{cc} \Sigma_{Q1} & \Sigma_{Q1,R1} \\ \Sigma_{Q1,R1} & \sigma_{R1}^2 \end{array} \right] = \left[ \begin{array}{cc} 6.06 & 2.50 \\ 2.50 & 2.79 \end{array} \right], \\
\mu_2 &= \left[ \begin{array}{c} \mu_{Q2} \\ \mu_{R2} \end{array} \right] = \left[ \begin{array}{c} 2 \\ 1 \\ 15 \end{array} \right], \quad \Sigma_2 = \left[ \begin{array}{cc} \Sigma_{Q2} & \Sigma_{Q2,R2} \\ \Sigma_{Q2,R2} & \sigma_{R2}^2 \end{array} \right] = \left[ \begin{array}{cc} 1.02 & 0.27 \\ 0.27 & 0.19 \end{array} \right].
\end{align*}
\]

Under this problem structure, for objectives \( k = 1, 2 \), we calculate the inputs for the true objective function values as

\[
H_k = E[Q_k^\top Q_k] = \Sigma_{Q_k} + \mu_{Q_k} \mu_{Q_k}, \quad V_k = E[Q_k^\top R_k] = \Sigma_{Q_k,R_k} + \mu_{Q_k} \mu_{R_k}, \quad C_k = E[R_k^2] = \sigma_{R_k}^2 + \mu_{R_k}^2,
\]

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

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 \( \beta^* \) in (58) holds and \( \mathbb{E}(\cdot) \) is measurable, we have \( 1 - \alpha = \lim_{n \to \infty} P\{ \beta^* \in \mathcal{R}_{n,1-\alpha} \} = \lim_{n \to \infty} P\{ \mathbb{E}(\beta^*) \in \mathcal{E}(\mathcal{R}_{n,1-\alpha}) \} = \lim_{n \to \infty} P\{ \mathbb{E} \in \mathcal{X}_{n,1-\alpha} \} \). An analogous justification holds for \( P\{ \beta \in \mathcal{Y}_{n,1-\alpha} \} \) since \( \mathbb{P}(\cdot) \) is measurable. □
5.1 Obtaining the known covariance matrix

We assume the covariance matrix $\Omega_\gamma$ in (59) is known. To obtain the $\Omega_\gamma$ matrix, we first obtain expressions for $\Omega_1, \Omega_2 \in \mathbb{S}_2^q$ from (21) for $k = 1, 2$

$$
\Omega_k = E \left[ \text{vec}_{\Theta_1}(Q_k^\top Q_k, Q_k^\top R_k, R_k^2) \right] \left( \text{vec}_{\Theta_1}(Q_k^\top Q_k, Q_k^\top R_k, R_k^2) \right)^\top - \text{vec}_{\Theta_1}(H_k, V_k, C_k) \left( \text{vec}_{\Theta_1}(H_k, V_k, C_k) \right)^\top.
$$

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

$$
\begin{bmatrix}
E[Q_1^2] & E[Q_1^2 Q_2] & E[Q_1^2 Q_2^2] & E[Q_1^2 R_1] & E[Q_1^2 Q_2 R_1] & E[Q_1^2 R_1^2] \\
E[Q_1^2 Q_2] & E[(Q_1 Q_2)^2] & E[Q_1 Q_2 Q_2^2] & E[Q_1 Q_2^2 R_1] & E[Q_1 Q_2 Q_2 R_1] & E[Q_1 Q_2 R_1^2] \\
E[Q_1^2 Q_2^2] & E[(Q_2 Q_2^2)^2] & E[Q_2^2 Q_2 R_1] & E[Q_2^2 Q_2 Q_2 R_1] & E[Q_2^2 Q_2 R_1^2] & E[Q_2^2 R_1^2] \\
E[Q_1^2 R_1] & E[Q_1 Q_2 R_1] & E[Q_1 Q_2 Q_2 R_1] & E[Q_1 Q_2 R_1^2] & E[Q_1 Q_2^2 R_1^2] & E[Q_1 R_1^2] \\
E[Q_1^2 Q_2 R_1] & E[Q_1 Q_2^2 R_1] & E[Q_1 Q_2 Q_2 R_1] & E[Q_1 Q_2 R_1^2] & E[Q_1 Q_2^2 R_1^2] & E[Q_1 R_1^2] \\
E[Q_1^2 R_1^2] & E[Q_1 Q_2 R_1^2] & E[Q_1 Q_2 Q_2 R_1^2] & E[Q_1 Q_2 R_1^2] & E[Q_1 Q_2^2 R_1^2] & E[Q_1 R_1^2] \\
\end{bmatrix}.
$$

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

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

$$
\Omega_2 = \begin{bmatrix}
2.3058 & 1.4026 & 1.4026 & 0.8322 & 10.2338 & 6.8186 & 28.6418 \\
68.8304 & 27.4846 & 27.4846 & 10.2338 & 282.957 & 89.8987 & 431.204 \\
92.3552 & 51.7144 & 51.7144 & 28.6418 & 431.204 & 256.363 & 1309.21 \\
\end{bmatrix}.
$$
Note that approximations to $\Omega_1$ and $\Omega_2$ also can be calculated using Monte Carlo with a very large sample size. The covariance matrices $\Omega_1$ and $\Omega_2$ 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 $\Omega_\gamma \in S^\geq_{14}$ 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, $\Omega^u_\gamma$ (discussed in Subsection 4.1) equals $\Omega^u_\gamma = 2 \Omega^u_1 \Omega^u_2 \in S^>_{12}$, where $\Omega^u_1$ and $\Omega^u_2$ are identical to $\Omega_1$ and $\Omega_2$, respectively, except with the second row and second column of each matrix deleted. To obtain a projection of the region $R_{n,.95}$ 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\|} \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 $\Omega^u_\gamma = LL^\top$, $U$ 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 $\beta_j$. Finally, we create the projections by plotting$
e(\beta_j) = \{ z_s(\text{vec}^{-1}_{\mathbb{R}^{12\times 2}}(\beta_j)), s \in \tilde{S} \}, \quad \mathbf{p}(\beta_j) = \{ g_s(\text{vec}^{-1}_{\mathbb{R}^{12\times 2}}(\beta_j)), s \in \tilde{S} \}, (61)$for $j = 1, 2, \ldots, m$ in the decision and objective spaces, respectively, where $\tilde{S}$ represents an appropriately fine discretization of $S$.

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 $R_{n,.95}$ 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 $R_{n,.95}$ should cover the true value $\beta^*$ 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 $R_{n,.95}$ and check whether it contains $\beta^*$. 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
estimated standard errors for these numbers are approximately $7.3 \times 10^{-5}$, $7.0 \times 10^{-5}$, and $6.9 \times 10^{-5}$; 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

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

where the cardinality is $|\tilde{S}| = 60$ 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 $e(\beta_j)$, $j = 1, \ldots, m$ is an arc of hyperbola [5], and in the objective space, each projected curve $p(\beta_j)$, $j = 1, \ldots, m$ 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 \( \tilde{S} \) 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 \( \Omega^u \)), 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 $\tilde{S}$ 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


