Multivariate Overlapping Batch Means for Valid Confidence Regions

Dashi I. Singham and Yingchieh Yeh

September 25, 2026

Abstract

We establish a valid statistical limit of the multivariate overlapping batched means method for constructing exact confidence regions. Properties of confidence regions generated using non-overlapping multivariate methods are well established. Recent advances in deriving the limiting distributions associated with overlapping batching methods have enabled confidence interval generation for univariate time series. We extend this past work to multivariate overlapping batching to derive the corresponding Student's t-distribution needed for confidence region generation under an arbitrary number of dimensions for the large- and small-batch settings. We show that overlapping confidence regions achieve improved region quality and lower volumes compared to non-overlapping methods. Confidence region quality is measured using multivariate extensions of coverage functions and their generalization, the VAMP1RE metric, which measures validity and mimicry compared to ideal regions generated using the known underlying covariance matrix. Altogether, we show the "something for nothing" advantages of overlapping batching compared to non-overlapping batching in a valid multivariate ellipsoidal confidence region setting.

1 Introduction

Batching methods are a classical approach for generating confidence intervals for the mean of simulation output. Confidence interval procedures (CIPs) are designed to deliver confidence intervals which cover the true mean μ with a desired probability level. Recent efforts have developed confidence region procedures (CRPs) which generate confidence regions for the mean

vector $\boldsymbol{\mu}$ using multivariate simulation output. Batching methods enjoy the simplicity of using cancellation for interval or region construction, rather than requiring a consistent estimator of the variace or covariance matrix.

The notion of cancellation via standardized time series for confidence interval generation was first developed in Schruben (1983). The establishment of batching as a standardized time series method was developed in Glynn and Iglehart (1990), where the authors established function properties that enable cancellation of the variance constant in the univariate setting. Next, Muñoz and Glynn (2001) extend the work of Glynn and Iglehart (1990) to the multivariate case and derive limiting properties of confidence region volumes. These papers employ non-overlapping batching, and give conditions for when limiting random variables converge to the t-distribution in the univariate case, and the F-distribution in the multidimensional case.

Since the development of overlapping batching (Meketon and Schmeiser 1984), there has been interest in developing estimators with potential "something for nothing" benefits of lower estimator variance. However, the lack of tractability around overlapping variance estimators have made confidence interval and confidence region calculations difficult. While the t- and F-distributions can be used for non-overlapping case, approximation methods have been employed in overlapping methods to inflate the degrees of freedom to account for the larger number of batches. Past work in Aktaran-Kalaycı (2006) and Su et al. (2024) yield significant progress towards delivering limiting random variables for variance estimators, while Singham and Yeh (2026) delivers quantiles from these analogous overlapping t-distributions which can be used for confidence interval generation.

We summarize the contributions of this paper. First, we present a limiting distribution for multivariate overlapping batch means (MOB) estimators. This random variable is an overlapping extension of the multivariate limit in Muñoz and Glynn (2001), and a multivariate extension of the univariate overlapping work in Aktaran-Kalaycı (2006) and Su et al. (2024). We provide code that can be used to generate quantiles from this distribution for confidence region generation, analogous to a standard call from a t- or F-distribution.

Next, we establish rigorous methods for evaluation of confidence region quality by extending the prior state-of-the-art methods of coverage functions and their generalization, the VAMP1RE metric, to the multivariate setting. These methods do not rely on an arbitrary choice of a confidence coefficient that is used in most coverage tables. Additionally, while smaller confidence regions are generally preferred, it is important that the regions mimic the behavior of ideal regions generated using an ideal CRP that employs the known underlying covariance estima-

tor. The VAMP1RE metric captures notions of coverage and region quality in a single value that enables clear comparisons between CRPs. We employ the VAMP1RE metric to present comprehensive numerical results comparing the effects of varying the batch size and number of batches. These results suggest that overlapping batching can achieve "something for nothing" relative to non-overlapping batching in the multivariate setting.

Section 2 presents a literature review. Section 3 presents the main multivariate overlapping limit results, while Section 4 presents extensions of coverage functions and the VAMP1RE metric to the multivariate case. Section 5 contains experimental results, while Section 6 concludes. Proofs are included in the Appendix.

2 Literature Review

There exists a large literature on confidence interval procedures, including many batching methods. There have been different approaches for constructing confidence regions in the multivariate setting, for example, using Bonferroni criteria in Raatikainen (1993), and by employing vector autoregressive methods in Charnes and Kelton (1993). Confidence regions (CRs) also can be constructed for stochastic approximation methods (Hsieh and Glynn 2002) and for evaluating initialization bias truncation for multivariate simulation output (Wu et al. 2020). Because our research involves overlapping batching of multivariate data, we will focus our review on multivariate non-overlapping batching methods and overlapping batching methods in Sections 2.1 and 2.2 respectively. Section 2.3 will conclude this section by summarizing past metrics for evaluating CIPs, which will then be extended to evaluating CRs.

2.1 Multivariate Non-Overlapping Batching

First, we review the insights from the literature on non-overlapping batching (NB) methods for multivariate output. Chen and Seila (1987) provided a first extension from univariate batch means to multivariate batch means. Early work by Yang and Nelson (1988) study non-overlapping batching with control variates in the multivariate i.i.d. normal case and find that having more batches decreased the generalized variance, hence more batches are helpful when there are a large number of variates. However, the improvement beyond a certain number of batches is negligible. Furthermore, the number of batches needed may scale up with the sample size n. The expected volume of the CR also decreases with an increasing number of batches b, especially with larger dimension d.

Yang and Nelson (1992) continue this analysis by showing how increasing the number of batches decreases the generalized variance, and the effect is more-significant when d is large. At some point, the marginal improvement of additional batches interferes with the i.i.d. normal assumption due to smaller batch sizes in the non-overlapping case. Yang and Nelson (1992) extends Schmeiser (1982) (which studied the effects of batch sizes in the non-overlapping univariate case) to the case when $d > 1$. While Schmeiser (1982) suggests marginal benefit beyond $b = 30$ in the univariate case, Yang and Nelson (1992) suggest there could be benefits as high as $b = 75$ or 80 in the multivariate case.

Muñoz and Glynn (2001) presents a comprehensive treatment of non-overlapping batching for multivariate time series using standardized time series approaches. They demonstrate how the multivariate batch means method is a standardized time series method, and develop a limit based on Brownian motion for a given class of estimators. This is one of the primary references we will be adapting for the overlapping setting.

Batching is used in the multivariate setting in numerous other cases, for example to construct confidence regions for multiple biased estimators using the Hotelling's T^2 statistic (He and Lam 2022), in managing initialization bias (Schruben 1981), and in multiple comparisons procedures (Goldsman and Nelson 1990). Additionally, non-overlapping batching is used for covariance matrix estimation in Markov Chain Monte Carlo procedures (Liu and Flegal 2018, Vats et al. 2019). Zhu and Dong (2021) develop non-overlapping batching methods for stochastic gradient descent when optimizing multiple model parameters by strategically allowing the batch size to increase over time. Additionally, they explore the effect of the number of batches on confidence region volume. Most recently, Lei et al. (2026) consider non-overlapping batching for constructing confidence regions for multiple quantile values, along with deriving a generalized likelihood ratio for estimating density functions to construct quantile estimates.

2.2 Overlapping Batching

The original benefits of overlapping batch (OB) methods were first proposed in Meketon and Schmeiser (1984), who note that the asymptotical variance of OB estimators is approximately 2/3 that of NB methods. Since then, there have been further attempts to formalize overlapping batch estimators and derive their statistical properties. Politis and Romano (1994) develop a general subsampling approach which allows for estimation of coverage regions of arbitrary statistics across multiple dimensions by taking potentially overlapping batches of data. Damerdji (1991) and Damerdji (1994) critically develop strong consistency of overlapping batch means

estimators.

Recent work has attempted to determine the distribution of overlapping variance estimators so that confidence intervals can be formed, similar to how the Student's t-distribution is used for constructing intervals in non-overlapping batching. Distributional properties of overlapping estimators under FCLT conditions are established in Aktaran-Kalaycı (2006), while Aktaran-Kalaycı et al. (2007) calculates expected values of non-overlapping and overlapping variance estimators. Linear combinations of overlapping variance estimators are developed and optimized in Goldsman et al. (2007) and Aktaran-Kalaycı et al. (2009). Alexopoulos et al. (2007a,b) present overlapping estimators for standardized time series estimators along with efficient computational methods. Yeh and Schmeiser (2015a) develop CIPs for OB which are evaluated using the VAMP1RE criterion, suggesting improved performance compared to NB methods.

Su et al. (2024) rigorously derives the limiting distribution of overlapping estimators associated with batching, sectioning, and standardized time series under a strong invariance assumption, allowing for estimators of all statistical functionals in the univariate case. Singham and Yeh (2026) presents an algorithm and code for producing the actual t-distributions associated with overlapping sectioning and standardized time series allowing for exact confidence interval calculation.

Finally, we highlight key papers that have considered overlapping batching in the multivariate case. Calvin and Nakayama (2013) suggested an extension of overlapping batch methods for multidimensional quantile estimation via standardized time series. Pasupathy et al. (2022) and Pasupathy et al. (2024) employ overlapping batching methods for confidence regions for multiple quantiles estimates, with Pasupathy et al. (2024) focusing on the infinite-dimensional case to establish a CR for the entire quantile field. Finally Liu et al. (2022) develop generalized OB estimators of covariance matrices of multidimensional data to derive MSE optimal batch sizes.

2.3 CIP metrics

Most papers evaluate the success of CIPs through tables reporting observed coverage and halfwidths. However, this only shows a snapshot of performance at particular choices of confidence coefficients. Coverage may be high due to larger intervals at larger confidence coefficient values, and it can be hard to separate the effect of interval size on coverage from the quality of the estimator itself. Additionally, it can be hard to compare different CIPs by looking at coverage values across different tables for each procedure. While we will calculate confidence region volumes to extend the traditional interval half-width metric to the univariate case, we

will also consider more-robust methods than standard coverage and interval-width estimates for evaluating interval quality.

We consider two computational methods for evaluating CIPs that do not depend on the choice of confidence coefficient and allow for straightforward comparisons between procedures. The first method is from the seminal paper Schruben (1980) which established coverage functions that define a distribution of calculated coverage values and produce a graphical representation of coverage quality. The second method, VAMP1RE, established in Yeh and Schmeiser (2015b), extends the concepts of coverage functions to include the effect of precision and interval length into a single value measuring performance of the CIP which can be used to make direct comparisons between procedures. Yeh and Tsai (2025) similarly develop the VAMP1RE metric to evaluate sequential CIPs. We will extend coverage functions and the VAMP1RE metric to the multidimensional setting as metrics to evaluate the effects of multivariate overlapping batching.

We conclude this section by noting the recent work of He and Lam (2024) which determines high-order coverage errors using Edgeworth expansions on t-statistics to enable a better analytical understanding of the finite-sample performance of batching methods in the non-overlapping i.i.d. case. These results suggest difficulty in determining uniform preferences across batching methods (sectioning versus batching) and non-monotonicty with respect to varying the number of batches.

3 Multivariate Overlapping Confidence Limit

3.1 Notation and Definitions

Let d be the dimension of the simulation output, which is the number of multivariate output streams to use to construct a CR for the mean vector $\boldsymbol{\mu}$. Let b be the number of batches, while m is the batch size and n is the total number of samples. Define $\beta := m/n$, which is the proportion of the time series allocated to a single batch. Batches are equally sized and allowed to be overlapping.

Let η be the desired confidence coefficient. This is often denoted as $1-\alpha$ where α is the allowable Type I error. The goal is to design a CRP which constructs a confidence region $R_n \in \mathbb{R}^d$ using n samples such that

$$\lim_{n\to\infty} P(\boldsymbol{\mu} \in R_n) = \eta.$$

The multivariate setting considers confidence regions that are L_p-regions for any chosen

$p \in [1, \infty]$. For the vector $\mathbf{x} \in \mathbb{R}^d$, $\|\mathbf{x}\|_p, p \geq 1$ refers to the L_p norm $\left(\sum_{j=1}^{d} |x_j|^p\right)^{1/p}$. For the special case of the L_2 norm, we use the simplified notation $\|\mathbf{x}\|$, and the L_∞ norm is defined as $||\mathbf{x}||_\infty = \max\{|x_1|, |x_2|, \ldots, |x_d|\}$. As with past work on multivariate non-overlapping batch means, we will employ the L_2 norm which produces ellipsoidal confidence regions.

We have n samples of d-dimensional output $\boldsymbol{Y}_j, j = 1, \ldots, n$, with $\boldsymbol{Y}_j = [Y_{1j}, Y_{2j}, \ldots, Y_{dj}]^T$. Define $\overline{\boldsymbol{X}}$ as the d-length mean vector for the entire time series (n samples) as $\overline{\boldsymbol{X}} = \left[\overline{\boldsymbol{X}}^1, \overline{\boldsymbol{X}}^2, \ldots, \overline{\boldsymbol{X}}^d\right]^T$. "Sectioning" methods use $\overline{\boldsymbol{X}}$ as the point estimate calculated using the mean of the entire series, as opposed to "batching" methods which use the mean of batched mean estimates as the point estimate. This paper will only consider sectioning methods. Next, define a batch mean as an average of m consecutive multivariate observations starting at a particular time point k:

$$\overline{\boldsymbol{X}}_k = \frac{1}{m} \sum_{j=k}^{k+m-1} \boldsymbol{Y}_j.$$

For a given sample size n and batch size m, we can also choose a number of batches b. When $b = n/m$ is an integer this represents the NB case of consecutive non-overlapping batches (or $b = \lfloor n/m \rfloor$ with some samples at the end of the time series omitted). Define the displacement of batches $dp = \lfloor (n-m)/(b-1) \rfloor$ as the number of samples between the starting points of batches. Values of b smaller than $\lfloor n/m \rfloor$ imply spaced batches with unused samples between batches and $dp > m$. For consecutive non-overlapping batches $dp = m$. Alternatively, batches can overlap by choosing $b > n/m$. In the fully overlapping case where $dp = 1$, we have the number of batches $b = n - m + 1$. Partial overlapping occurs when $n/m < b < n - m + 1$ and $1 < dp < m$.

Next, we define the notation related to the dependence within and across time series in $\boldsymbol{Y}$. Define cross-variate covariance as Σ_d and cross-batch covariance as Σ_b. Denote Σ as the asymptotic covariance matrix of the scaled partial sum $\sum_{j=1}^{n} \boldsymbol{Y}_j/\sqrt{n}$, where each entry of Σ is

$$\Sigma(l,k) = \Sigma_d(l,k) + 2\sum_{h=1}^{\infty} \rho_h(l,k)\sqrt{\Sigma_d(l,l)\Sigma_d(k,k)}, \tag{3.1}$$

$$\rho_h(l,k) = \rho(Y_{l,1}, Y_{k,1+h}), \quad l,k = 1, \ldots, d \tag{3.2}$$

where $\rho(Y_{a,b}, Y_{c,d})$ is the autocorrelation between data points Y_{ab} and Y_{cd}.

We will use $\Rightarrow$ to designate convergence in distribution as $n \to \infty$. Finally, let $\boldsymbol{B} = \{\boldsymbol{B}(t) : 0 \leq t \leq 1\}$ be a d-dimensional standard Brownian motion.

3.2 Multivariate Overlapping Batch Means (MOB) Limit

Under a standard CLT, suppose that we can consistently estimate Σ, say using n samples to obtain $\hat{\Sigma}_n$. Then $\hat{\Sigma}_n$ will be positive definite for large enough n, and we have the following limit

$$n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^T \hat{\Sigma}_n^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \Rightarrow \chi_d^2$$

where χ_d^2 is a chi-square distributed random variable with d degrees of freedom. A ellipsoidal confidence interval for $\boldsymbol{\mu}$ can then be constructed using the η-quantiles of the χ_d^2 distribution. Next, we assume a multivariate functional central limit theorem (FCLT) in Assumption 3.1.

Assumption 3.1. (FCLT) Let $\overline{\boldsymbol{Y}}_n(t) = \sum_{j=1}^{\lfloor nt \rfloor} \boldsymbol{Y}_j/n, 0 \leq t \leq 1$. Then, the following functional central limit theorem holds:

$$\sqrt{n}(\overline{\boldsymbol{Y}}_n(t) - t\boldsymbol{\mu}) \Rightarrow G\boldsymbol{B}(t),$$

where $\Sigma = GG^T$.

Then, define a function $g : \mathcal{C}[0,1] \to \mathbb{R}^{d \times d}$ that is applied to a function $x(t), 0 \leq t \leq 1$,

$$g(x) := \frac{1}{b\beta(1-\beta)} \sum_{i=1}^{b} \left(\Delta x(c_i) - \beta x(1)\right) \left(\Delta x(c_i) - \beta x(1)\right)^T \tag{3.3}$$

where $c_i = \frac{(i-1)(1-\beta)}{b-1}$ and $\Delta x(c_i) = x(c_i + \beta) - x(c_i)$. From Muñoz and Glynn (2001), their NB version of $g(x)$ is

$$\hat{g}(x) = \frac{b}{b-1} \sum_{i=1}^{b} \left(\Delta x(i/b) - \frac{x(1)}{b}\right) \left(\Delta x(i/b) - \frac{x(1)}{b}\right)^T$$

with $\Delta x(i/b) = x(i/b) - x((i-1)/b), i = 1, \ldots, b$. In the non-overlapping case $b = 1/\beta$ and we recover $\hat{g} = g$. Thus, our function g in (3.3) is a straightforward extension of Muñoz and Glynn (2001) to the overlapping case. Similarly, g is a multivariate extension of the univariate overlapping function $\chi^2_{\mathrm{OB-I}}$ function in Su et al. (2024), returning a covariance matrix instead of a scalar variance.

Theorem 3.2. MOB Limit. Suppose that Assumption 3.1 holds. Then, if $b \geq (d+1)$, $g(x)$ satisfies the following:

$$(\overline{\boldsymbol{Y}} - \mu)^T g(\overline{\boldsymbol{Y}}_n)^{-1} (\overline{\boldsymbol{Y}} - \mu) \implies \boldsymbol{B}(1)^T g(\boldsymbol{B})^{-1} \boldsymbol{B}(1). \tag{3.4}$$

The proof for Theorem 3.2 is in Appendix A. The proof using $\hat{g}$ from Muñoz and Glynn (2001)

relied on the fact that non-overlapping batch means were multivariate i.i.d. normally distributed so that $\hat{g}(\boldsymbol{B})$ is positive semi-definite with probability 1 via Dykstra (1970). Since we no longer have independence between terms in $g(x)$ due to overlapping batches we need a slightly more-general proof. From Muñoz and Glynn (2001), we also see that $g(\boldsymbol{B})$ is independent of $\boldsymbol{B}(1)$. We note that the right-hand side limit in (3.4) similarly incorporates the dependence across batches so is a different limit (computationally) than the one in Muñoz and Glynn (2001). For non-overlapping batches using $\hat{g}$, it can be shown that

$$\boldsymbol{B}(1)^T g(\boldsymbol{B})^{-1} \boldsymbol{B}(1) \implies \frac{d(b-1)}{b-d} F_{(d,b-d)}$$

where $F_{(d,b-d)}$ is an F-distributed random variable with degrees of freedom $d, b-d$. This is because $\Delta\boldsymbol{B}(i/b)$ terms in the limit are independent and have sample mean $\boldsymbol{B}(1)/b$ and sample covariance $g(\boldsymbol{B})/b$. In the overlapping case, the $\Delta\boldsymbol{B}(x)$ terms are dependent. We can define the covariance across overlapping batches (Σ_b) within the same univariate output series using

$$\text{cov}(D_i, D_j) \longrightarrow_{m,n\to\infty} \beta \left(\max\left\{0, 1 - \frac{|i-j|}{m}\right\} - \beta \right) \tag{3.5}$$

where $D_i = \Delta B(c_i) - \beta B(1)$ and $\text{cov}(\Delta B(c_i), \Delta B(c_j)) = \max\{0, \beta - c_j + c_i\}$ for $c_i \leq c_j$. See Singham and Yeh (2026) for a derivation of this covariance.

The work in Su et al. (2024) distinguishes between the large-batch and small-batch settings. Small-batch settings occur when $\beta \to 0$ as $n \to \infty$ and $b \to b_\infty$ for some limiting number of batches b_∞ so that batch sizes become small relative to sample size, and the number of batches is infinite. Large batch settings are when $\beta \to c$ for some constant $c > 0$ and $b \to \infty$. Thus we can define the large-batch limits using

$$g(\boldsymbol{B}) = \begin{cases} \frac{1}{\beta(1-\beta)^2} \int_0^{1-\beta} \left(\Delta\boldsymbol{B}(u) - \beta\boldsymbol{B}(1)\right)\left(\Delta\boldsymbol{B}(u) - \beta\boldsymbol{B}(1)\right)^T du, & \text{for } b = \infty; \\ \frac{1}{b\beta(1-\beta)} \sum_{j=1}^{b} \left(\Delta\boldsymbol{B}(c_j) - \beta\boldsymbol{B}(1)\right)\left(\Delta\boldsymbol{B}(c_j) - \beta\boldsymbol{B}(1)\right)^T, & \text{for } b \in \mathbb{N} \setminus \{1\}, \end{cases} \tag{3.6}$$

where $\Delta\boldsymbol{B}(c_j) = \boldsymbol{B}(c_j + \beta) - \boldsymbol{B}(c_j)$ and $c_j := (j-1)\frac{1-\beta}{b-1}$ which converges to $(j-1)/n$ when $b = n - m + 1$ in the fully overlapping case.

3.3 Confidence Interval Calculation

We now present a CRP which can be used to deliver regions with nominal coverage using this MOB limit. Define the following $d \times d$ sample covariance estimator formed from a d-dimensional

time series with $k(i) = dp(i-1)+1$ with $\overline{\boldsymbol{X}}_{k(i)}$ consisting of batches of size m as

$$\tilde{\Sigma} = \frac{1}{b-1}\sum_{i=1}^{b}\left(\overline{\boldsymbol{X}}_{k(i)} - \overline{\boldsymbol{X}}\right)\left(\overline{\boldsymbol{X}}_{k(i)} - \overline{\boldsymbol{X}}\right)^T.$$

We can write $\tilde{\Sigma} = \frac{b(1-\beta)}{(b-1)\beta} g(\overline{\boldsymbol{Y}}_n)$, and for the specific NB case where $b = 1/\beta$ we have $\tilde{\Sigma} = bg(\overline{\boldsymbol{Y}}_n)$. Thus we can use cancellation from (3.4) to compute confidence intervals using the sample covariance estimate.

Singham and Yeh (2026) synthesizes the literature to bring together the univariate overlapping limits derived in Aktaran-Kalaycı (2006), Alexopoulos et al. (2007b), Su et al. (2024) under FCLT and strong invariance assumptions. They employ an efficient eigenvalue method to construct code which delivers the quantiles of the associated overlapping t-distribution, denoted OB-t, and perform a similar calculation for the OSTS-t distribution. For the MOB case, we present code for a simulation-based approach allowing the user to quickly call quantiles from the distribution of $\boldsymbol{B}(1)^T g(\boldsymbol{B})^{-1}\boldsymbol{B}(1)$ for given choices of n, m, b, d, and η. [Link]

4 Multivariate Confidence Region Evaluation

When testing confidence interval procedures, most researchers display numerical results by choosing some confidence coefficient η to construct confidence intervals (usually 90% or 95%) and computing experimental observed coverage. However, this approach lacks a systematic way of assessing interval quality. If it is known that as the sample size approaches infinity that coverage should approach nominal coverage, it is unclear how much finite-sample coverage loss is acceptable. Furthermore, it is unclear how much the choice of η impacts the numerical results. In many procedures, larger η leads to larger intervals, so as $\eta \to 1$, by definition experimental coverage will approach 100%.

Thus, it is often unclear whether good coverage is the result of a quality procedure or abnormally large confidence intervals. Reporting the average half-width of the calculated confidence intervals can help ensure that intervals are not overly large. But again, it is unclear what an acceptable half-width would be. While smaller intervals are assumed to be better because they are precise, the "ideal" confidence interval would have a half-width that accurately reflects the true underlying variability in the data.

Multidimensional confidence regions can take different forms, as seen in Figure 1 which shows examples for two dimensions for different norms p. A point estimator for the means of two output, $\overline{Y}^1, \overline{Y}^2$ is located in the center of the regions. Then, depending on the choice of

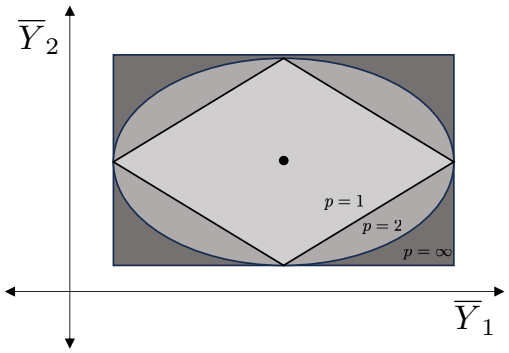

Figure 1: Two-dimensional confidence region example for different choices of the norm, p.

norm p used to define the confidence interval, the confidence regions can take different shapes.

There are two state-of-the-art methods for evaluating confidence interval quality that do not depend on a choice of η and apply to any norm p. The first is the coverage function (Schruben 1980), and the second is the VAMP1RE metric (Yeh and Schmeiser 2015b). We will present easy extensions to the multivariate setting for both of these methods in Sections 4.1 and 4.2. Additionally, we will consider confidence region volume as a metric of quality in Section 4.3.

4.1 Multidimensional Coverage Functions

Schruben (1980) introduced univariate coverage functions as a method of evaluating CIPs. Suppose the CIP is attempting to provide intervals for some parameter θ. Let η^* be a random variable defined as

$$\eta^* = \inf\{\eta \in [0,1]; \theta \in R_\eta\}, \tag{4.1}$$

which is the smallest confidence coefficient that allows the an interval constructed as R_η to cover θ. Then, the coverage function (CF) is the cdf of η^* so that

$$CF(\eta) = Pr\{\eta^* \leq \eta\}. \tag{4.2}$$

Smaller values of η^* imply smaller intervals (using smaller confidence coefficients) are needed to cover θ, which implies better coverage. If the CIP obtains exactly nominal coverage, then for any chosen η, observed coverage will be η and hence the coverage function will be the uniform cdf.

Most CIP results evaluate a procedure by simply recording a binary condition of whether

the estimated CI covers the true parameter of interest or not using a fixed arbitrary confidence coefficient. Performance is reported as the percent of intervals that correctly cover θ. Coverage functions deliver much deeper information, by showing how a procedure performs across the entire confidence coefficient spectrum $\eta \in [0, 1]$. Rather than simply recording whether an interval covered the true parameter, η^* informs how close the sample parameter was to θ. However, calculation of η^* requires knowledge of the cdf function for the t-distribution associated with the estimated limiting distribution. For example, Singham and Yeh (2026) calculate the t-distribution associated with univariate overlapping batching, OB-t and produce coverage functions to compare OB methods. Call the cdf of the OB-t distribution $F_{\text{OB}}(t, n, m, b)$ for particular choices of batching parameter n, m, dp. Then the coverage value is calculated as

$$\eta^* = 2F_{\text{OB}}\left(\frac{|\overline{Y} - \mu|}{\sqrt{\hat{S}^2/n}}, n, m, b\right) - 1$$

where $\hat{S}^2$ is the estimate of the sample variance and two-sided confidence intervals are employed.

The extension of Schruben's coverage functions to the multivariate case is straightforward. Figure 2 shows graphically how η^* can be calculated for an ellipsoidal region. For a set of sample data, we will achieve a point estimate $\overline{\boldsymbol{Y}}$. Suppose the true mean estimate is $\boldsymbol{\mu}$. Then, the coverage value η^* is the confidence coefficient needed to extend the region such that it exactly covers $\boldsymbol{\mu}$. This is a one-dimensional root finding problem just like the univariate coverage case, but using a multivariate limit distribution, in this case,

$$\eta^* = F_{\text{OB}_d}\left((\overline{\boldsymbol{Y}} - \boldsymbol{\mu})g(\overline{\boldsymbol{Y}})^{-1}(\overline{\boldsymbol{Y}} - \boldsymbol{\mu})^T\right)$$

where OB_d has a distribution $\boldsymbol{B}(1)^T g(\boldsymbol{B})^{-1}\boldsymbol{B}(1)$ using the parameters β, b, and d. Then t_{η^*} is the quantile of the OB_d distribution such that the CR exactly covers $\boldsymbol{\mu}$ on its boundary.

4.2 Multidimensional VAMP1RE

A valid CIP delivers a coverage function that is uniformly distributed, i.e. coverage is nominal. However, just because a procedure is valid, does not mean it provides high quality intervals. For example, a procedure that delivers the interval $[-\infty, \infty]$ with probability η and $[0, 0]$ with probability $1 - \eta$ is technically valid. Yeh and Schmeiser (2015b) proposed the VAMP1RE criterion that includes both validity and mimicry in a single evaluation number that allows CIPs to directly be compared. VAMP1RE stands for Validity And Mimicry, employing p-values

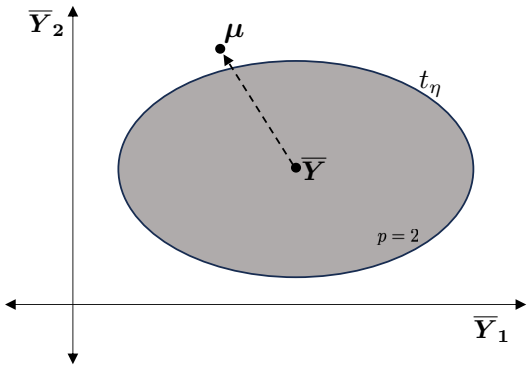

Figure 2: The value of η^* is calculated by determining the value t_{η^*} that exactly covers the true mean $\boldsymbol{\mu}$.

(P) using a single (1) criterion for Rating and ranking CIPs using an Expected vaule.

Mimicry is when a CIP interval closely resembles an interval that would be generated under ideal conditions, for example, when the underlying true variance is known. A good CIP should produce intervals of similar width and stability compared to a procedure with known good properties, or an "ideal" CIP. In our multidimensional interval setting, we can define the ideal CRP as one using the known value of Σ, i.e., using the distribution of $n(\overline{\boldsymbol{Y}} - \boldsymbol{\mu})^T \Sigma^{-1} (\overline{\boldsymbol{Y}} - \boldsymbol{\mu})$ which has a χ^2_d distribution to generate CRs.

Given the same set of simulation output data used in the batching procedure, let Δ be the coverage value η^* from the CIP of interest, and Ψ be the coverage value from the ideal CIP. Both Δ and Ψ are random variables. Then, the VAMP1RE criterion is defined simply as

$$\mathrm{E}[(\Delta - \Psi)^2].$$

Numerous CIPs for which the value of VAMP1RE is known analytically are given in Yeh and Schmeiser (2015b). Smaller values of VAMP1RE are preferred, with the best case value of 0, and the worst case value as 7/12. Valid CIPs where Δ and Ψ are positively correlated have VAMP1RE values less than 1/6. Thus, we expect values of VAMP1RE smaller than 1/6 for good CRPs, with the values decreasing as n increases and confidence region estimates improve. For a given sample of data l, the ideal coverage value can be estimated as

$$\psi_l = F_{\chi^2_d} \left(n(\overline{\boldsymbol{Y}} - \boldsymbol{\mu})^T \Sigma^{-1} (\overline{\boldsymbol{Y}} - \boldsymbol{\mu}) \right)$$

and the actual coverage as

$$\delta_l = F_{\mathrm{OB}_d}\left((\overline{\boldsymbol{Y}} - \boldsymbol{\mu})^T g(\overline{\boldsymbol{Y}})^{-1}(\overline{\boldsymbol{Y}} - \boldsymbol{\mu})\right).$$

Then, we can form a Monte Carlo estimate of VAMP1RE from R macroreplications as

$$\hat{\mathrm{E}}[(\Delta - \Psi)^2] = \frac{\sum_{l=1}^{r}(\delta_l - \psi_l)^2}{R}.$$

The VAMP1RE metric can be decomposed into two terms, validity and mimicry:

$$E[(\Delta - \Psi)^2] = \left[\mathrm{MSE}(\Psi, 1/2) - \frac{1}{12}\right] + 2\left[\frac{1}{12} - \mathrm{cov}(\Delta, \Psi)\right]. \tag{4.3}$$

The first term measures validity (the level by which assumptions of the procedure are violated). Like coverage functions, the idea of the VAMP1RE criterion is easily extended to multidimensional confidence region procedures since we can generate coverage values using the OB_d distribution. Because we derive exact limits for the overlapping batch case, we expect coverage functions to be nominal and the validity measure to be zero for i.i.d. normal data, and for FCLT data as the sample size n increases.

The second term in (4.3) is mimicry, whereby high values imply the CIP or CRP does not produce intervals (or confidence regions) that mimic those produced by the ideal procedure. Low values imply high mimicry. Thus, for valid procedures we can still measure interval quality using mimicry to compare the effects of different batching configurations.

4.3 Confidence Region Volume

In addition to coverage, the volume of the confidence region can also be used as a measure of performance. Traditionally, smaller confidence interval half-widths and smaller CR volumes are desired. We calculate volumes to explore the effects of overlapping batching on region size using function g. Let q_d be a constant $q_d = \frac{\pi^{d/2}}{\Gamma(d/2-1)}$. Then the volume of the CR is defined as

$$V(g) = q_d \det(g(\overline{\boldsymbol{Y}}_n)^{1/2}) t^{d/2}_{\mathrm{OB_d},\eta}$$

where $t_{\mathrm{OB_d},\eta}$ is the η-quantile of the OB_d distribution. When a consistent variance estimator is available, the limiting volume is

$$\lim_{n\to\infty} n^{d/2}\mathrm{E}[V(\hat{\Sigma}_n)] = q_d \det(G)\chi^2_{d,\eta} \tag{4.4}$$

with $\chi^2_{d,\eta}$ as the η-quantile of the χ^2 distribution with d degrees of freedom. Theorem 2 of Muñoz and Glynn (2001) shows that the limit of the consistent estimator is smaller than or equal to the limit using NB batching configuration $\hat{g}$, i.e.,

$$\liminf_{n\to\infty} n^{d/2}\mathrm{E}[V(\hat{g})] \geq \liminf_{n\to\infty} n^{d/2}\mathrm{E}[V(\hat{\Sigma}_n)].$$

This result extends to the OB case using g in (3.3). To reduce the effect of dimensionality and sample size on the volume, we will consider the limiting d-root half-volume which is defined as

$$HV_d = \frac{\sqrt{n}V(g)^{1/d}}{2}. \tag{4.5}$$

For the one-dimensional case, this is the same as the standard confidence interval half-width. Note that while we assume ellipsoidal regions in our volume calculations, the volume of the region could be determined by p as in Figure 1. The volume of the L_p confidence region in dimension d using norm p is

$$V(g) = \frac{(2\Gamma(\frac{1}{p}+1))^d}{\Gamma(\frac{d}{p}+1)} \prod_{j=1}^{d} \frac{t_{p,\eta} S_n^p}{\sqrt{n}} = \frac{t^d_{p,\eta}}{n^{d/2}} \frac{(2\Gamma(\frac{1}{p}+1))^d}{\Gamma(\frac{d}{p}+1)} \prod_{j=1}^{d} S_n^p. \tag{4.6}$$

where $t_{p,\eta}$ is the appropriate quantile associated with the limiting distribution and S_n^p is the variance estimate. The d-root half-volume of (4.6) in the general p case is then

$$HV_d = \frac{t_{p,\eta}}{\sqrt{n}} \frac{(\Gamma(\frac{1}{p}+1))}{\Gamma(\frac{d}{p}+1)^{1/d}} \left[\prod_{j=1}^{d} S_n^p\right]^{1/d}. \tag{4.7}$$

This metric was used in Pasupathy et al. (2024) to evaluate multidimensional confidence intervals on quantile fields. It can be shown that if $d = 1$, (4.7) becomes the standard half-width. If $p = \infty$, then the confidence region is a hyperrectangle and (4.7) is the geometric mean of the dimensions.

5 Numerical Results

In this section, we display the performance of multivariate overlapping regions calculated using the limit derived in (3.4) and compare them to non-overlapping regions. We know from Theorem 3.2 that the coverage will be asymptotically nominal for FCLT data, and exactly nominal for i.i.d. normally distributed data. Thus, we do not display coverage functions but present the more-comprehensive VAMP1RE values. Additionally, we compute the d-root half-volumes as a measure of region size. Each table will demonstrate the effects of increasing batch size m and the number of batches b for a given sample size n. We use 10,000 macroreplications to estimate each table value, and use common random numbers across tables entries and tables.

Each table employs the following color-coding for the various combinations of m and b. The values are red for consecutive NB (where $n = mb$) and magenta for spaced batching. The color cyan signifies partially overlapping batches (where we assume batches are spaced to cover the entire n time series using displacement $dp = (n-m)/(b-1)$) and blue signifies fully overlapping batching where $b = n - m + 1$. The table shows 'NA' values for batching configurations that are infeasible.

5.1 i.i.d. Normal data

We explore a simple estimate of multivariate normal i.i.d. data of dimension $d = 2$ with mean $\boldsymbol{\mu}_2 = [0, 0]$ and covariance matrix $\Sigma_2 = [1, -1.9; -1.9, 6]$. First, we show the VAMP1RE values for the small sample case where $n = 10$ in Table 1. Looking horizontally across the tables we see increasing the number of batches b leads to improved (smaller) VAMP1RE values. Note that for $m = 1$, when there is no batching, the best results are obtained. This is intuitive in the i.i.d. context where there are no benefits to batching and the best results are obtained when as many samples as possible are used ($b = 10$).

Table 1: VAMP1RE values for Normal i.i.d. data of length $n = 10$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2.

	$b=4$	$b=5$	$b=6$	$b=7$	$b=8$	$b=9$	$b=10$
$m=1$	0.039	0.032	0.027	0.024	0.021	0.019	0.017
$m=2$	0.042	0.038	0.032	0.028	0.026	0.024	NA
$m=3$	0.042	0.040	0.034	0.032	0.031	NA	NA
$m=4$	0.049	0.040	0.036	0.034	NA	NA	NA
$m=5$	0.046	0.041	0.041	NA	NA	NA	NA
$m=6$	0.047	0.041	NA	NA	NA	NA	NA
$m=7$	0.046	NA	NA	NA	NA	NA	NA
$m=8$	NA	NA	NA	NA	NA	NA	NA

We break down the VAMP1RE values from Table 1 by validity (in Table 2) and mimicry

(in Table 3). As expected, validity is effectively zero because the CRP is using the exact t-distribution to deliver nominal coverage. The inability to mimic the ideal CRP due to the small sample size represents the bulk of the contribution to the VAMP1RE metric.

Table 2: Validity values for Normal i.i.d. data of length $n = 10$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2.

	$b=4$	$b=5$	$b=6$	$b=7$	$b=8$	$b=9$	$b=10$
$m=1$	0.001	0.000	0.001	0.001	0.001	0.000	0.000
$m=2$	0.000	0.001	0.001	0.001	0.001	0.001	NA
$m=3$	−0.001	0.004	−0.003	0.001	0.001	NA	NA
$m=4$	0.001	−0.001	−0.002	0.000	NA	NA	NA
$m=5$	−0.008	−0.004	0.001	NA	NA	NA	NA
$m=6$	0.000	0.001	NA	NA	NA	NA	NA
$m=7$	0.001	NA	NA	NA	NA	NA	NA
$m=8$	NA	NA	NA	NA	NA	NA	NA

Table 3: Mimicry values for Normal i.i.d. data of length $n = 10$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2.

	$b=4$	$b=5$	$b=6$	$b=7$	$b=8$	$b=9$	$b=10$
$m=1$	0.038	0.031	0.026	0.023	0.020	0.018	0.016
$m=2$	0.042	0.037	0.030	0.027	0.025	0.023	NA
$m=3$	0.043	0.035	0.036	0.031	0.030	NA	NA
$m=4$	0.048	0.041	0.037	0.034	NA	NA	NA
$m=5$	0.053	0.045	0.040	NA	NA	NA	NA
$m=6$	0.046	0.040	NA	NA	NA	NA	NA
$m=7$	0.045	NA	NA	NA	NA	NA	NA
$m=8$	NA	NA	NA	NA	NA	NA	NA

Next, we increase the sample size to $n = 100$ and will explore the effects of large-batch and small-batch limits. Note that for this i.i.d. normal case using Σ_2, the limiting scaled volume for the consistent estimator is calculated using (4.4) as 29.0993 with a d-root half-width of $HV_d^* = \sqrt{29.0993}/2 = 2.697$ where we use HV_d^* to denote the limiting volume for the consistent estimator. The tables will report estimated CRP volume using HV_d calculated as (4.5).

Table 4 displays the VAMP1RE values for $n = 100$. We see that while fully OB values have lower values than consecutive NB values, VAMP1RE is not monotonically decreasing as b increases. Additionally, as the sample size m increases, we see non-monotonic effects, though generally smaller batch sizes are preferred in this i.i.d. normal case.

Table 5 shows the HV_d values, which are all strictly greater than the consistent limiting volume of $HV_d^* = 2.697$. Again, we see benefits of fully OB delivering lower volume values than consecutive NB, however, volume is not decreasing monotonically as b increases. Increasing the batch size also yields non-monotonic effects.

We now show the small-batch limiting effects by keeping the batch sizes m the same as in

Table 4: VAMP1RE values for Normal i.i.d. data of length $n = 100$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2.

	$b = 4$	$b = 5$	$b = 10$	$b = 20$	$b = 50$	$b = 51$	$b = 76$	$b = 81$	$b = 91$	$b = 96$
$m = 2/\beta = 0.02$	0.039	0.031	0.015	0.007	0.003	0.004	0.003	0.003	0.002	0.002
$m = 5/\beta = 0.05$	0.039	0.031	0.015	0.008	0.009	0.009	0.006	0.006	0.005	0.005
$m = 10/\beta = 0.10$	0.039	0.032	0.017	0.012	0.016	0.016	0.011	0.011	0.010	NA
$m = 20/\beta = 0.20$	0.043	0.039	0.022	0.020	0.025	0.025	0.019	0.019	NA	NA
$m = 25/\beta = 0.25$	0.054	0.034	0.026	0.025	0.027	0.027	0.022	NA	NA	NA
$m = 50/\beta = 0.50$	0.051	0.044	0.034	0.034	0.032	0.032	NA	NA	NA	NA
$m = 81$	0.048	0.044	0.039	0.037	NA	NA	NA	NA	NA	NA
$m = 91$	0.047	0.044	0.038	NA	NA	NA	NA	NA	NA	NA
$m = 96$	0.051	0.044	NA	NA	NA	NA	NA	NA	NA	NA

Table 5: HV_d values for Normal i.i.d. data of length $n = 100$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2. The value of $HV_d^* = 2.697$.

	$b = 4$	$b = 5$	$b = 10$	$b = 20$	$b = 50$	$b = 51$	$b = 76$	$b = 81$	$b = 91$	$b = 96$
$m = 2/\beta = 0.02$	4.635	3.984	3.191	2.909	2.775	2.760	2.765	2.761	2.759	2.756
$m = 5/\beta = 0.05$	4.635	3.990	3.197	2.919	2.799	2.798	2.830	2.827	2.824	2.834
$m = 10/\beta = 0.10$	4.656	4.031	3.238	3.018	2.907	2.904	2.967	2.973	2.985	NA
$m = 20/\beta = 0.20$	4.923	4.620	3.465	3.322	3.188	3.198	3.287	3.293	NA	NA
$m = 25/\beta = 0.25$	6.377	4.264	3.674	3.414	3.337	3.348	3.434	NA	NA	NA
$m = 50/\beta = 0.50$	7.187	5.394	4.169	3.922	3.910	3.918	NA	NA	NA	NA
$m = 81$	5.282	4.660	4.246	4.185	NA	NA	NA	NA	NA	NA
$m = 91$	5.351	4.788	4.353	NA	NA	NA	NA	NA	NA	NA
$m = 96$	5.290	4.917	NA	NA	NA	NA	NA	NA	NA	NA

Tables 4 and 5, while increasing n to 1000 so that β approaches zero to calculate VAMP1RE in Table 6 and HV_d in Table 7. We see that the VAMP1RE values in Table 6 are now very close to zero, as we expect a CRP involving a large sample of i.i.d. normal data to perform well. The HV_d values in Table 7 also rapidly approach that of the consistent estimator (2.697), with small batch sizes and high levels of overlap performing particularly well.

Table 6: VAMP1RE values for Normal i.i.d. data of length $n = 1000$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2. (Small-batch)

	$b = 20$	$b = 40$	$b = 50$	$b = 100$	$b = 200$	$b = 500$	$b = 951$	$b = 976$	$b = 981$	$b = 991$
$m = 2/\beta = 0.002$	0.007	0.004	0.003	0.001	0.001	0.000	0.000	0.000	0.000	0.000
$m = 5/\beta = 0.005$	0.008	0.004	0.003	0.002	0.001	0.001	0.001	0.001	0.001	0.001
$m = 10/\beta = 0.010$	0.008	0.004	0.003	0.002	0.001	0.002	0.001	0.001	0.001	0.001
$m = 20/\beta = 0.020$	0.008	0.004	0.003	0.002	0.002	0.004	0.002	0.002	0.002	NA
$m = 25/\beta = 0.025$	0.008	0.004	0.003	0.003	0.003	0.005	0.003	0.003	NA	NA
$m = 50/\beta = 0.050$	0.008	0.006	0.006	0.005	0.006	0.009	0.005	NA	NA	NA
$m = 81$	0.010	0.009	0.009	0.009	0.009	0.014	NA	NA	NA	NA
$m = 91$	0.011	0.010	0.010	0.010	0.010	0.015	NA	NA	NA	NA
$m = 96$	0.012	0.010	0.010	0.010	0.011	0.016	NA	NA	NA	NA

Finally, we examine the large-batch setting by comparing Tables 4 and 5 where $n = 100$ to the case where $n = 1000$ but we scale the batch size m with n so β remains constant. Tables 8 and 9 show the VAMP1RE and volume calculations while keeping β the same as in Tables 4

Table 7: HV_d values for Normal i.i.d. data of length $n = 1000$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2. The value of $HV_d^* = 2.697$. (Small-batch)

	$b = 20$	$b = 40$	$b = 50$	$b = 100$	$b = 200$	$b = 500$	$b = 951$	$b = 976$	$b = 981$	$b = 991$
$m = 2/\beta = 0.002$	2.908	2.800	2.773	2.735	2.716	2.703	2.701	2.704	2.702	2.701
$m = 5/\beta = 0.005$	2.907	2.802	2.778	2.738	2.717	2.705	2.710	2.711	2.711	2.714
$m = 10/\beta = 0.010$	2.906	2.798	2.776	2.736	2.724	2.714	2.724	2.725	2.720	2.720
$m = 20/\beta = 0.020$	2.903	2.798	2.774	2.752	2.745	2.727	2.748	2.750	2.744	NA
$m = 25/\beta = 0.025$	2.908	2.800	2.783	2.768	2.758	2.738	2.761	2.758	NA	NA
$m = 50/\beta = 0.050$	2.923	2.857	2.846	2.832	2.820	2.783	2.833	NA	NA	NA
$m = 81$	2.975	2.930	2.929	2.925	2.912	2.853	NA	NA	NA	NA
$m = 91$	3.016	2.970	2.964	2.958	2.942	2.874	NA	NA	NA	NA
$m = 96$	3.027	2.985	2.976	2.974	2.957	2.888	NA	NA	NA	NA

and 5. We see higher VAMP1RE and HV_d values than the small batch setting due to slower convergence.

Table 8: VAMP1RE values for Normal i.i.d. data of length $n = 1000$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2. (Large-batch)

	$b = 4$	$b = 5$	$b = 10$	$b = 20$	$b = 50$	$b = 501$	$b = 751$	$b = 801$	$b = 901$	$b = 951$
$m = 20/\beta = 0.02$	0.039	0.031	0.015	0.008	0.003	0.004	0.003	0.002	0.002	0.002
$m = 50/\beta = 0.05$	0.039	0.032	0.015	0.008	0.006	0.009	0.006	0.006	0.005	0.005
$m = 100/\beta = 0.10$	0.040	0.032	0.017	0.012	0.010	0.016	0.011	0.011	0.010	NA
$m = 200/\beta = 0.20$	0.044	0.039	0.022	0.020	0.019	0.025	0.019	0.019	NA	NA
$m = 250/\beta = 0.25$	0.053	0.036	0.027	0.023	0.022	0.027	0.022	NA	NA	NA
$m = 500/\beta = 0.50$	0.056	0.047	0.035	0.032	0.032	0.031	NA	NA	NA	NA
$m = 981$	0.049	0.046	0.040	0.038	NA	NA	NA	NA	NA	NA
$m = 991$	0.048	0.045	0.039	NA	NA	NA	NA	NA	NA	NA

Table 9: HV_d values for Normal i.i.d. data of length $n = 1000$ with parameters $\boldsymbol{\mu}_2$, Σ_2 and dimension d =2. The value of $HV_d^* = 2.697$. (Large-batch)

	$b = 4$	$b = 5$	$b = 10$	$b = 20$	$b = 50$	$b = 501$	$b = 751$	$b = 801$	$b = 901$	$b = 951$
$m = 20/\beta = 0.02$	4.632	3.996	3.170	2.905	2.774	2.727	2.737	2.742	2.747	2.748
$m = 50/\beta = 0.05$	4.640	3.984	3.184	2.915	2.843	2.784	2.816	2.820	2.830	2.830
$m = 100/\beta = 0.10$	4.663	4.025	3.237	3.034	2.986	2.903	2.965	2.967	2.979	NA
$m = 200/\beta = 0.20$	4.898	4.603	3.467	3.324	3.290	3.192	3.284	3.298	NA	NA
$m = 250/\beta = 0.25$	6.371	4.245	3.679	3.473	3.443	3.337	3.438	NA	NA	NA
$m = 500/\beta = 0.50$	7.001	5.308	4.141	3.956	3.908	3.899	NA	NA	NA	NA
$m = 981$	5.413	4.804	4.413	4.366	NA	NA	NA	NA	NA	NA
$m = 991$	5.449	4.875	4.461	NA	NA	NA	NA	NA	NA	NA

5.2 Vector Auto-regressive Data

This section shows the results for vector auto-regressive (VAR) data to examine the effect of time series dependence. As in the previous example with i.i.d. normal data, we employ $\boldsymbol{\mu}_2 = [0, 0]$ and use a noise covariance matrix $\Sigma_2 = [1, -1.9; -1.9, 6]$. Additionally, the transition matrix we use to simulate VAR data is $\Phi = [0.7, 0.1; 0.2, 0.6]$. For VAR data, it is known that the

underlying variance of the data is

$$\Omega = \Lambda \Sigma_2 \Lambda^T \quad \text{where } \Lambda = (I - \Phi)^{-1}$$

and I is the identity matrix with dimension d. Thus, we use Ω as the ideal consistent variance matrix, and $G = \Omega^{1/2}$. In this setting, the limiting scaled volume from consistent estimation is 290.9925, with a d-root half-volume of $HV_d^* = 8.529$.

Table 10 displays the VAMP1RE values for this VAR data, with Table 11 providing the HW_d values. Note that the VAMP1RE values in Table 10 are extremely similar to those for i.i.d. normal data in Table 4 and 8 when the same β and b values are chosen.

Table 10: VAMP1RE values for Vector autoregressive data of length $n = 10000$ with parameters $\boldsymbol{\mu}_2$, Φ, Ω and dimension d =2.

	$b=4$	$b=5$	$b=10$	$b=20$	$b=50$	$b=5001$	$b=7501$	$b=8001$	$b=9001$	$b=9501$
$m=200/\beta=0.02$	0.039	0.031	0.015	0.008	0.003	0.004	0.003	0.002	0.002	0.002
$m=500/\beta=0.05$	0.039	0.031	0.016	0.008	0.005	0.009	0.006	0.006	0.005	0.005
$m=1000/\beta=0.10$	0.040	0.033	0.017	0.011	0.010	0.016	0.012	0.011	0.010	NA
$m=2000/\beta=0.20$	0.043	0.038	0.022	0.020	0.019	0.025	0.019	0.019	NA	NA
$m=2500/\beta=0.25$	0.053	0.034	0.026	0.023	0.022	0.027	0.022	NA	NA	NA
$m=5000/\beta=0.50$	0.058	0.047	0.035	0.033	0.032	0.032	NA	NA	NA	NA
$m=9981$	0.059	0.065	0.064	0.063	NA	NA	NA	NA	NA	NA
$m=9991$	0.071	0.082	0.082	NA	NA	NA	NA	NA	NA	NA

Table 11: HV_d values for Vector autoregressive data of length $n = 10000$ with parameters $\boldsymbol{\mu}_2$, Φ, Ω and dimension d =2. The value of $HV_d^* = 8.529$.

	$b=4$	$b=5$	$b=10$	$b=20$	$b=50$	$b=5001$	$b=7501$	$b=8001$	$b=9001$	$b=9501$
$m=200/\beta=0.02$	14.479	12.499	9.928	9.124	8.715	8.549	8.586	8.599	8.615	8.623
$m=500/\beta=0.05$	14.615	12.551	10.031	9.185	8.942	8.760	8.852	8.872	8.893	8.898
$m=1000/\beta=0.10$	14.689	12.695	10.204	9.517	9.385	9.125	9.306	9.340	9.381	NA
$m=2000/\beta=0.20$	15.454	14.538	10.924	10.442	10.371	10.044	10.323	10.374	NA	NA
$m=2500/\beta=0.25$	20.081	13.427	11.575	10.907	10.828	10.505	10.808	NA	NA	NA
$m=5000/\beta=0.50$	22.105	16.674	13.027	12.442	12.342	12.294	NA	NA	NA	NA
$m=9981$	14.291	12.321	11.137	10.972	NA	NA	NA	NA	NA	NA
$m=9991$	12.347	10.536	9.469	NA	NA	NA	NA	NA	NA	NA

5.3 VAR, $d = 6$

Here, we explore an example of sparse VAR using $d = 6$, taken from Davis et al. (2016). This example employs the following transition matrix and the noise covariance matrix

$$\Phi_s = \begin{bmatrix} 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.3 & 0 \\ 0.6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8 \end{bmatrix} \quad \Sigma_s = \begin{bmatrix} \delta^2 & \delta/4 & \delta/6 & \delta/8 & \delta/10 & \delta/12 \\ \delta/4 & 1 & 0 & 0 & 0 & 0 \\ \delta/6 & 0 & 1 & 0 & 0 & 0 \\ \delta/8 & 0 & 0 & 1 & 0 & 0 \\ \delta/10 & 0 & 0 & 0 & 1 & 0 \\ \delta/12 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let Ω_s be the corresponding VAR variance for this sparse case. The true mean vector is $\boldsymbol{\mu}_6 = [0, 0, 0, 0, 0, 0]$. We calculate the corresponding value of HV_d value using the consistent variance estimator as $HV_d^* = 3.839$. Table 12 displays the VAMP1RE values and Table 13 displays the volume values. Both tables operate in the large-batch setting where we have kept β constant as n increases. We observe similar trends as before where fully overlapping batching achieve lower VAMP1RE and volume values than consecutive NB values. While the dimension of 6 is much higher than the previous example of dimension 2, the relative sparsity of the covariance and transition matrices yield relatively comparable performance for 10000 samples relative to Tables 10 and 11.

Table 12: VAMP1RE values for Vector autoregressive data of length $n = 10000$ with parameters $\boldsymbol{\mu}_6$, Φ_s, Ω_s and dimension d =6.

	$b = 10$	$b = 20$	$b = 50$	$b = 5001$	$b = 7501$	$b = 8001$	$b = 9001$	$b = 9501$
$m = 200/\beta = 0.02$	0.060	0.027	0.011	0.015	0.009	0.009	0.008	0.007
$m = 500/\beta = 0.05$	0.058	0.029	0.019	0.035	0.021	0.020	0.018	0.017
$m = 1000/\beta = 0.10$	0.066	0.039	0.031	0.056	0.035	0.033	0.031	NA
$m = 2000/\beta = 0.20$	0.069	0.049	0.044	0.060	0.044	0.043	NA	NA
$m = 2500/\beta = 0.25$	0.076	0.051	0.046	0.058	0.045	NA	NA	NA
$m = 5000/\beta = 0.50$	0.075	0.055	0.049	0.048	NA	NA	NA	NA
$m = 9981$	0.110	0.110	NA	NA	NA	NA	NA	NA
$m = 9991$	0.144	NA	NA	NA	NA	NA	NA	NA

Finally, we test another six-dimensional VAR problem that is highly-dependent and has

Table 13: HV_d for Vector autoregressive data of length $n = 10000$ with parameters $\boldsymbol{\mu}_6$, Φ_s, Ω_s and dimension d =6. The value of $HV_d^* = 3.839$.

	$b = 10$	$b = 20$	$b = 50$	$b = 5001$	$b = 7501$	$b = 8001$	$b = 9001$	$b = 9501$
$m = 200/\beta = 0.02$	6.653	4.648	4.093	3.903	3.963	3.968	3.982	3.988
$m = 500/\beta = 0.05$	6.707	4.698	4.316	4.057	4.206	4.224	4.257	4.268
$m = 1000/\beta = 0.10$	7.623	5.166	4.818	4.394	4.692	4.728	4.780	NA
$m = 2000/\beta = 0.20$	8.196	6.042	5.668	5.114	5.562	5.611	NA	NA
$m = 2500/\beta = 0.25$	9.790	6.286	5.895	5.388	5.833	NA	NA	NA
$m = 5000/\beta = 0.50$	10.251	6.963	6.495	6.423	NA	NA	NA	NA
$m = 9981$	8.409	6.538	NA	NA	NA	NA	NA	NA
$m = 9991$	7.866	NA	NA	NA	NA	NA	NA	NA

dense covariance and transition matrices. We employ the following parameters:

$$\Phi_d = \begin{bmatrix} 0.17 & 0.27 & 0.33 & 0.22 & 0.21 & 0.15 \\ -0.27 & 0.25 & 0.33 & 0.17 & 0.13 & 0.32 \\ 0.14 & 0.10 & 0.11 & 0.17 & 0.13 & 0.25 \\ 0.11 & 0.19 & -0.23 & 0.24 & 0.11 & 0.11 \\ 0.31 & 0.35 & 0.24 & 0.20 & 0.29 & 0.24 \\ 0.27 & 0.11 & 0.25 & 0.31 & -0.32 & 0.13 \end{bmatrix} \quad \Sigma_d = \begin{bmatrix} 1.47 & 0.07 & -0.15 & 0.29 & -0.01 & 0.06 \\ 0.07 & 1.05 & -0.02 & 0.04 & 0.02 & -0.09 \\ -0.15 & -0.02 & 1.62 & 0.24 & -0.37 & 0.10 \\ 0.29 & 0.04 & 0.24 & 1.99 & -0.30 & 0.26 \\ -0.01 & 0.02 & -0.37 & -0.30 & 1.64 & -0.24 \\ 0.06 & -0.09 & 0.10 & 0.26 & -0.24 & 1.04 \end{bmatrix}.$$

which yields the value of $HV_d^* = 4.031$ and a corresponding Ω_d. Table 14 displays the values of VAMP1RE for this highly-dependent setting, and Table 15 displays the values of HV_d. The values of VAMP1RE appear similar to those in the sparse setting, though the volume values are naturally higher.

Table 14: VAMP1RE values for Vector autoregressive data of length $n = 10000$ with parameters $\boldsymbol{\mu}_6$, Φ_d, Ω_d and dimension d =6.

	$b = 10$	$b = 20$	$b = 50$	$b = 5001$	$b = 7501$	$b = 8001$	$b = 9001$	$b = 9501$
$m = 200/\beta = 0.02$	0.059	0.027	0.011	0.014	0.009	0.008	0.007	0.007
$m = 500/\beta = 0.05$	0.059	0.028	0.018	0.034	0.020	0.019	0.017	0.016
$m = 1000/\beta = 0.10$	0.066	0.039	0.031	0.054	0.034	0.032	0.030	NA
$m = 2000/\beta = 0.20$	0.068	0.050	0.045	0.060	0.045	0.044	NA	NA
$m = 2500/\beta = 0.25$	0.076	0.051	0.046	0.057	0.045	NA	NA	NA
$m = 5000/\beta = 0.50$	0.076	0.055	0.049	0.048	NA	NA	NA	NA
$m = 9981$	0.094	0.088	NA	NA	NA	NA	NA	NA
$m = 9991$	0.122	NA	NA	NA	NA	NA	NA	NA

6 Conclusions

We demonstrate how multivariate overlapping batching methods can be used to derive valid CRPs by extending past work on multivariate non-overlapping methods and univariate overlapping methods. By applying multivariate extensions of coverage values and the VAMP1RE

Table 15: HV_d values for Vector autoregressive data of length $n = 10000$ with parameters $\boldsymbol{\mu}_6$, Φ_d, Ω_d and dimension d =6. The value of $HV_d^* = 4.031$.

	$b = 10$	$b = 20$	$b = 50$	$b = 5001$	$b = 7501$	$b = 8001$	$b = 9001$	$b = 9501$
$m = 200/\beta = 0.02$	7.018	4.903	4.316	4.119	4.183	4.188	4.202	4.209
$m = 500/\beta = 0.05$	7.041	4.943	4.537	4.266	4.422	4.442	4.477	4.488
$m = 1000/\beta = 0.10$	7.988	5.423	5.057	4.618	4.925	4.964	5.019	NA
$m = 2000/\beta = 0.20$	8.605	6.335	5.945	5.368	5.834	5.887	NA	NA
$m = 2500/\beta = 0.25$	10.266	6.594	6.183	5.655	6.120	NA	NA	NA
$m = 5000/\beta = 0.50$	10.768	7.324	6.831	6.757	NA	NA	NA	NA
$m = 9981$	9.807	7.659	NA	NA	NA	NA	NA	NA
$m = 9991$	9.424	NA	NA	NA	NA	NA	NA	NA

criterion, we can assess the quality of CRPs applied to different overlapping batching configurations varying the batch size and number of batches. We find for normally distributed data where the CRP is valid, smaller batch sizes with a high number of batches through overlapping outperforms standard non-overlapping methods both in terms of higher interval quality through better mimicry, and smaller CRP volumes.

References

Tûba Aktaran-Kalaycı. *Steady-state analyses: Variance estimation in simulations and dynamic pricing in service systems*. PhD thesis, 2006.

Tûba Aktaran-Kalaycı, Christos Alexopoulos, Nilay T Argon, David Goldsman, and James R Wilson. Exact expected values of variance estimators for simulation. *Naval Research Logistics*, 54(4):397–410, 2007.

Tûba Aktaran-Kalaycı, Christos Alexopoulos, David Goldsman, and James R Wilson. Optimal linear combinations of overlapping variance estimators for steady-state simulation. *Advancing the Frontiers of Simulation: A Festschrift in Honor of George Samuel Fishman*, pages 291–328, 2009.

Christos Alexopoulos, Nilay T Argon, David Goldsman, Natalie M Steiger, Gamze Tokol, and James R Wilson. Efficient computation of overlapping variance estimators for simulation. *INFORMS Journal on Computing*, 19(3):314–327, 2007a.

Christos Alexopoulos, Nilay T Argon, David Goldsman, Gamze Tokol, and James R Wilson. Overlapping variance estimators for simulation. *Operations Research*, 55(6):1090–1103, 2007b.

James M Calvin and Marvin K Nakayama. Confidence intervals for quantiles with standardized

time series. In *Proceedings of the 2013 Winter Simulations Conference*, pages 601–612, 2013. doi: 10.1109/WSC.2013.6721454.

John M Charnes and W David Kelton. Multivariate autoregressive techniques for constructing confidence regions on the mean vector. *Management science*, 39(9):1112–1129, 1993.

Der-Fa R Chen and Andrew F Seila. Multivariate inference in stationary simulation using batch means. In *Proceedings of the 19th Conference on Winter Simulation*, pages 302–304, 1987. doi: https://doi.org/10.1145/318371.318436.

Halim Damerdji. Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37(11):1424–1440, 1991.

Halim Damerdji. Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research*, 19(2):494–512, 1994.

Richard A Davis, Pengfei Zang, and Tian Zheng. Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, 25(4):1077–1096, 2016.

Richard L Dykstra. Establishing the positive definiteness of the sample covariance matrix. *The Annals of Mathematical Statistics*, 41(6):2153–2154, 1970.

Peter W Glynn and Donald L Iglehart. Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15(1):1–16, 1990.

David Goldsman, Keebom Kang, Seong-Hee Kim, Andrew F Seila, and Gamze Tokol. Combining standardized time series area and Cramér–von Mises variance estimators. *Naval Research Logistics*, 54(4):384–396, 2007.

Lynne Goldsman and Barry L Nelson. Batch-size effects on simulation optimization using multiple comparisons with the best. In *Proceedings of the 1990 Winter Simulation Conference*, pages 288–293, 1990. doi: 10.1109/WSC.1990.129529.

Shengyi He and Henry Lam. Batching on biased estimators. In *Proceedings of the 2022 Winter Simulation Conference*, pages 2606–2616, 2022. doi: 10.1109/WSC57314.2022.10015356.

Shengyi He and Henry Lam. Higher-order coverage errors of batching methods via Edgeworth expansions on t-statistics. *The Annals of Statistics*, 52(4):1360–1383, 2024.

Ming-hua Hsieh and Peter W Glynn. Confidence regions for stochastic approximation algorithms. In *Proceedings of the 2002 Winter Simulation Conference*, volume 1, pages 370–376, 2002. doi: 10.1109/WSC.2002.1172906.

Lei Lei, Christos Alexopoulos, Yijie Peng, and James R Wilson. Estimating confidence intervals

and regions for quantiles in steady-state simulations. *Naval Research Logistics*, 73(5): 853–867, 2026.

Ying Liu and James M Flegal. Weighted batch means estimators in Markov Chain Monte Carlo. *Electronic Journal of Statistics*, 12:3397–3442, 2018.

Ying Liu, Dootika Vats, and James M Flegal. Batch size selection for variance estimators in MCMC. *Methodology and Computing in Applied Probability*, 24(1):65–93, 2022.

Marc S Meketon and Bruce Schmeiser. Overlapping batch means: Something for nothing? In *Proceedings of the 1984 Winter Simulation Conference*, pages 227–230, 1984.

David F Muñoz. On the validity of the batch quantile method for Markov chains. *Operations Research Letters*, 38(3):223–226, 2010.

David F Muñoz and Peter W Glynn. Multivariate standardized time series for steady-state simulation output analysis. *Operations Research*, 49(3):413–422, 2001.

Raghu Pasupathy, Dashi I Singham, and Yingchieh Yeh. Overlapping batch confidence regions on the steady-state quantile vector. In *Proceedings of the 2022 Winter Simulation Conference*, pages 25–36, 2022. doi: 10.1109/WSC57314.2022.10015372.

Raghu Pasupathy, Dashi I Singham, and Yingchieh Yeh. Overlapping batch confidence regions on the steady-state quantile field. *Preprint*, 2024.

Dimitris N Politis and Joseph P Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4):2031–2050, 1994.

Kimmo E E Raatikainen. A sequential procedure for simultaneous estimation of several means. *ACM Transactions on Modeling and Computer Simulation*, 3(2):108–133, 1993.

Bruce Schmeiser. Batch size effects in the analysis of simulation output. *Operations Research*, 30(3):556–568, 1982.

Lee W Schruben. A coverage function for interval estimators of simulation response. *Management Science*, 26(1):18–27, 1980.

Lee W Schruben. Control of initialization bias in multivariate simulation response. *Communications of the ACM*, 24(4):246–252, 1981.

Lee W Schruben. Confidence interval estimation using standardized time series. *Operations Research*, 31(6):1090–1108, 1983.

Dashi I Singham and Yingchieh Yeh. Statistical inference via overlapping batches (OB) and overlapping standardized time series (OSTS) t-distributions. *Under Review*, 2026.

Ziwei Su, Raghu Pasupathy, Yingchieh Yeh, and Peter Glynn. Overlapping batch confidence intervals on statistical functionals constructed from time series: Application to quantiles, optimization, and estimation. *ACM Transactions on Modeling and Computer Simulation*, 34(2):1–43, 2024.

Dootika Vats, James M Flegal, and Galin L Jones. Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*, 106(2):321–337, 2019.

Jianguo Wu, Honglun Xu, Feng Ju, and Tzu-Liang Tseng. Adaptive minimum confidence region rule for multivariate initialization bias truncation in discrete-event simulations. *Technometrics*, 62(4):499–512, 2020.

Wei-Ning Yang and Barry L Nelson. Multivariate estimation and variance reduction in terminating and steady-state simulation. In *Proceedings of the 20th Conference on Winter Simulation*, pages 466–472, 1988. doi: https://doi.org/10.1145/318123.318234.

Wei-Ning Yang and Barry L Nelson. Multivariate batch means and control variates. *Management science*, 38(10):1415–1431, 1992.

Yingchieh Yeh and Bruce Schmeiser. OBM confidence intervals: Something for nothing? In *Proceedings of the 2015 Winter Simulation Conference*, pages 551–561, 2015a. doi: 10.1109/WSC.2015.7408195.

Yingchieh Yeh and Bruce W Schmeiser. VAMP1RE: a single criterion for rating and ranking confidence-interval procedures. *IIE Transactions*, 47(11):1203–1216, 2015b.

Yingchieh Yeh and Shing Chih Tsai. Single criterion for rating confidence interval procedures: a generalization to stochastic stopping rules. *Journal of the Operational Research Society*, 76(7):1481–1489, 2025.

Yi Zhu and Jing Dong. On constructing confidence region for model parameters in stochastic gradient descent via batch means. In *Proceedings of the 2021 Winter Simulation Conference*, pages 1–12, 2021. doi: 10.1109/WSC52266.2021.9715437.

A Proof of Theorem 3.2

Theorem 1 of Muñoz (2010) delivers the main result if Assumption 3.1 holds and the following four conditions hold for function $g(\mathrm{x})$.

1. $g(Gx) = Gg(x)G^T$
2. $g(x - \lambda J) = g(x)$ for $\lambda \in \mathbb{R}^d$ and $J(t) = t, 0 \leq t \leq 1$

3. $g(\boldsymbol{B})$ is positive definite and symmetric almost surely
4. $\mathbb{P}(\boldsymbol{B} \in D(g(\boldsymbol{B}))) = 0$ where $D(g(\boldsymbol{B}))$ is the set of discontinuities of g.

The difference in $g(x)$ from Muñoz and Glynn (2001) and Zhu and Dong (2021) is that the $\Delta x(c_i)$ terms are now dependent across batches in the summation. However, verifying conditions 1, 2, and 4 are still straightforward. Recall

$$g(x) = \frac{1}{b\beta(1-\beta)} \sum_{i=1}^{b} \left(\Delta x(c_i) - \beta x(1)\right) \left(\Delta x(c_i) - \beta x(1)\right)^T \tag{A.1}$$

where $c_i = \frac{(i-1)(1-\beta)}{b-1}$ and $\Delta x(c_i) = x(c_i + \beta) - x(c_i)$. For condition 1, write:

$$\begin{aligned}
g(Gx) &= \frac{1}{b\beta(1-\beta)} \sum_{i=1}^{b} \left(Gx(c_i + \beta) - Gx(c_i) - \beta Gx(1)\right) \left(Gx(c_i + \beta) - Gx(c_i) - \beta Gx(1)\right)^T \\
&= \frac{G}{b\beta(1-\beta)} \sum_{i=1}^{b} \left(\Delta x(c_i) - \beta x(1)\right) \left(\Delta x(c_i) - \beta x(1)\right)^T G^T \\
&= G g(x) G^T.
\end{aligned}$$

For condition 2:

$$\begin{aligned}
g(x - \lambda J) &= \frac{1}{b\beta(1-\beta)} \sum_{i=1}^{b} \big[\left(x(c_i + \beta) - \lambda J(c_i + \beta) - x(c_i) + \lambda J(c_i) - \beta x(1) + \beta\lambda J(1)\right) \times \\
&\qquad \left(x(c_i + \beta) - \lambda J(c_i + \beta) - x(c_i) + \lambda J(c_i) - \beta x(1) + \beta\lambda J(1)\right)^T \big] \\
&= \frac{1}{b\beta(1-\beta)} \sum_{i=1}^{b} \big[\left(x(c_i + \beta) - \lambda(c_i + \beta) - x(c_i) + \lambda(c_i) - \beta x(1) + \beta\lambda\right) \times \\
&\qquad \left(x(c_i + \beta) - \lambda(c_i + \beta) - x(c_i) + \lambda(c_i) - \beta x(1) + \beta\lambda\right)^T \big] \\
&= \frac{1}{b\beta(1-\beta)} \sum_{i=1}^{b} \left(x(c_i + \beta) - x(c_i) - \beta x(1)\right) \left(x(c_i + \beta) - x(c_i) - \beta x(1)\right)^T = g(x)
\end{aligned}$$

For condition 4, because $g(x)$ is a continuous function, this condition is satisfied. Finally, for condition 3, we need to show that $g(\boldsymbol{B})$ is positive definite and symmetric almost surely. According to Dykstra (1970) the sample covariance matrix formed from i.i.d. normal samples is positive definite (w.p. 1) if the underlying variance matrix is positive definite and $b \geq d + 1$ so this condition is easily met for non-overlapping batching when $\Delta\boldsymbol{B}(c_i)$ are independent. We need to show that for any non-zero vector $v \in \mathbb{R}^d$

$$P(v^T g(\boldsymbol{B}) v > 0) = 1.$$

Define

$$\boldsymbol{Z}(t) = \boldsymbol{B}(t+\beta) - \boldsymbol{B}(t) - \beta\boldsymbol{B}(1), \quad t \in [0, 1-\beta]. \tag{A.2}$$

Then the overlapping sample covariance matrix is defined for $b_\infty = \infty$ as

$$g(\boldsymbol{B}) = \frac{1}{\beta(1-\beta)^2} \int_0^{1-\beta} \boldsymbol{Z}(t)\boldsymbol{Z}^T(t)dt. \tag{A.3}$$

Let us assume the opposite effect, that $P(v^T g(\boldsymbol{B})v \leq 0) > 0$. This requires that the integrand of (A.3) is equal to zero so that

$$v^T g(\boldsymbol{B})v = \int_0^{1-\beta} v^T \boldsymbol{Z}(t)\boldsymbol{Z}^T(t)vdt = \int_0^{1-\beta} (v^t \boldsymbol{Z}(t))^2 dt = 0$$

because it cannot be less than zero. This implies $v^T\boldsymbol{Z}(t) = 0$ for all $t \in [0, 1-\beta]$. However, this leads to a contradiction against the properties of Brownian motion. Define $W(t) = v^T\boldsymbol{B}(t)$, and note that for any non-zero v and multidimensional $\boldsymbol{B}(t)$ that $W(t)$ is a scaled one-dimensional standard Brownian motion. Then, rewriting (A.2) yields

$$v^T(\boldsymbol{B}(t+\beta) - \boldsymbol{B}(t)) = \beta v^T\boldsymbol{B}(1) \tag{A.4}$$

$$W(t+\beta) - W(t) = \beta W(1) \tag{A.5}$$

for $t \in [0, 1-\beta]$. However (A.5) cannot hold because it implies the Brownian motion path is a straight line with slope $W(1)$, and this violates the independent increments requirement. □