

Subsampling to Enhance Efficiency in Input Uncertainty Quantification

Henry Lam

Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027,
khl2114@columbia.edu

Huajie Qian

Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027,
h.qian@columbia.edu

In stochastic simulation, input uncertainty refers to the output variability arising from the statistical noise in specifying the input models. This uncertainty can be measured by a variance contribution in the output, which, in the nonparametric setting, is commonly estimated via the bootstrap. However, due to the convolution of the simulation noise and the input noise, the bootstrap consists of a two-layer sampling and typically requires substantial simulation effort. This paper investigates a subsampling framework to reduce the required effort, by leveraging the form of the variance and its estimation error in terms of the data size and the sampling requirement in each layer. We show how the total required effort can be reduced from an order bigger than the data size in the conventional approach to an order independent of the data size in subsampling. We explicitly identify the procedural specifications in our framework that guarantee relative consistency in the estimation, and the corresponding optimal simulation budget allocations. We substantiate our theoretical results with numerical examples.

Key words: bootstrap, subsampling, input uncertainty, variance estimation, nonparametric, nested simulation

1. Introduction

Stochastic simulation is one of the most widely used analytic tools in operations research. It provides a flexible means to approximate complex models and to inform decisions. See, for instance, Law et al. (1991) for applications in manufacturing, revenue management, service and operations systems etc. In practice, the simulation platform relies on input models that are typically observed or calibrated from data. These statistical noises can propagate to the output analysis, leading

to significant errors and suboptimal decision-making. In the literature, this problem is commonly known as input uncertainty or extrinsic uncertainty.

In conventional simulation output analysis where the input model is completely pre-specified, the statistical errors come solely from the Monte Carlo noises, and it suffices to account only for such noises in analyzing the output variability. When input uncertainty is present, such an analysis will undermine the actual variability. One common approach to quantify the additional uncertainty is to estimate the variance in the output that is contributed from the input noises (e.g., Song et al. (2014)); for convenience, we call this the *input variance*. This quantity acts as an uncertainty measure which, when added together with the Monte Carlo variance, gives rise to the overall variance in the outputs. A refined decomposition of input variance across multiple input sources can be used to identify models that are overly ambiguous and flag the need of more data collection (e.g., Song et al. (2014)). Input variance also provides a building block to construct valid output confidence intervals (CIs) that account for combined input and simulation errors (e.g., Cheng and Holland (2004)). Motivated by its central role in quantifying input uncertainty, this paper aims to study the efficient estimation of input variance.

In the literature, bootstrap resampling is a common approach for the above purpose. This applies most prominently in the nonparametric regime, namely when no assumptions are placed on the input parametric family. It could also be used in the parametric case (where more alternatives are available). For example, Cheng and Holland (1997) proposes the variance bootstrap, and Song and Nelson (2015) studies the consistency of this strategy on a random-effect model that describes the uncertainty propagation. A bottleneck with using bootstrap resampling in estimating input variances, however, is the need to “outwash” the simulation noise, which often places substantial burden on the required simulation effort. More precisely, to handle both the input and the simulation noises, the bootstrap procedure typically comprises a two-layer sampling that first resamples the input data (i.e., *outer* sampling), followed by running simulation replications using each resample (i.e., *inner* replications). Due to the reciprocal relation between the magnitude of the input

variance and the input data, the input variance becomes increasingly small as the input data size increases. This deems the control of the relative estimation error increasingly expensive, and requires either a large outer bootstrap size or inner replication size to extinguish the effect of simulation noises.

The main goal of this paper is to investigate *subsampling* as a simulation saver for input variance estimation. This means that, instead of creating distributions by resampling a data set of the full size, we only resample (with or without replacement) a set of smaller size. We show that a judicious use of subsampling can reduce the total simulation effort from an order bigger than the data size in the conventional two-layer bootstrap to an order independent of the data size, while retaining the estimation accuracy. This approach leverages the interplay between the form of the input variance and its estimation error, in terms of the data size and the sampling effort in each layer of the bootstrap. On a high level, the subsample is used to estimate an input variance as if less data are available, followed by a correction of this discrepancy in the data size by properly rescaling the input variance. We call this approach *proportionate subsampled variance bootstrap*. We explicitly identify the procedural specifications in our approach that guarantee estimation consistency, including the minimally required simulation effort in each layer. We also study the theoretical behavior of our estimation error, in relation to the simulation effort allocation in these layers as well as the input data and subsample sizes, which in turn reveals the optimal configurations and provides implementation guidance.

In the statistics literature, subsampling has been used as a remedy for situations where the full-size bootstrap does not apply, due to a lack (or undeterminability) of uniform convergence required for its statistical consistency, which relates to the functional smoothness or regularity of the estimators (e.g., Politis and Romano (1994)). Subsampling has been used in time series and dependent data (e.g., Politis et al. (1999), Hall et al. (1995), Datta and McCormick (1995)), extremal estimation (e.g., Bickel and Sakov (2008)), shape-constrained estimation (e.g., Sen et al. (2010)) and other econometric contexts (e.g., Abadie and Imbens (2008), Andrews and Guggenberger (2009,

2010)). In contrary to these works, our subsampling approach is introduced to reduce the simulation effort faced by the two-layer sampling necessitated from the presence of both the input and simulation noises. In other words, we are not concerned about the issue of uniform convergence, but instead, we aim to distort the relation between the required simulation effort and data size in a way that allows more efficient deconvolution of the effects of the two noises. We also note that, as we will use resampling with replacement (instead of without replacement), our approach is closer to the so-called m out of n bootstrap (Bickel et al. (1997), Bickel and Sakov (2008)). For coherence, throughout the paper we use the term subsampling broadly to indicate a bootstrap with a smaller resample size than the original data size.

We close this introduction with a brief review of other related work in input uncertainty. In the nonparametric regime (the focus of this paper), besides Cheng and Holland (1997) and Song and Nelson (2015) that study bootstrap-based estimation of the input variance, Barton and Schruben (1993) and Barton and Schruben (2001) investigate the percentile bootstrap to construct CIs (i.e., the CI limits are determined from the quantiles of the bootstrap distributions). Like variance bootstrap, percentile bootstrap also encounters two-layer sampling that requires substantial simulation efforts. Yi and Xie (2017) investigates adaptive budget allocation policies based on ranking and selection to reduce simulation cost in the percentile bootstrap, and empirically shows the computational advantage of their approach. On the other hand, contrary to this paper, they do not investigate the required simulation efforts in relation to the input data size. Lam and Qian (2016, 2017) study the use of empirical likelihood as an optimization-based alternative to the percentile bootstrap, which requires simulation efforts to estimate the gradient information that remain substantial. Beyond the frequentist regime considered in this paper, Xie et al. (2019) studies nonparametric Bayesian methods based on Dirichlet process mixtures to estimate the variance contributed from input uncertainty and construct CIs. Glasserman and Xu (2014), Hu et al. (2012), Lam (2016b) and Ghosh and Lam (2019) study input uncertainty from a robust optimization viewpoint, where they compute worst-case bounds subject to constraints or so-called uncertainty sets

that represent partial beliefs on unknown distributions. In the parametric regime, Barton et al. (2013) and Xie et al. (2016) investigate the basic bootstrap with a metamodel built in advance, a technique known as the metamodel-assisted bootstrap. Xie et al. (2016) and Biller and Corlu (2011) study multivariate input uncertainty assuming a parametric dependency structure in the form of product-moment correlations. Cheng and Holland (1997) studies the delta method, and Cheng and Holland (1998, 2004) reduce its computation burden via the so-called two-point method. Lin et al. (2015) and Song and Nelson (2019) study regression approaches to estimate sensitivity coefficients which are used to apply the delta method, generalizing the gradient estimation method in Wieland and Schmeiser (2006). Zhu et al. (2020) studies risk criteria and computation to quantify parametric uncertainty. Finally, Chick (2001), Zouaoui and Wilson (2003), Zouaoui and Wilson (2004) and Xie et al. (2014) study variance estimation and interval construction from a Bayesian perspective. We comment that although the exposition in this paper focuses on the nonparametric setting, the same idea of subsampling can be adapted naturally to the parametric setting, with similar advantages in computational efficiency. For general surveys on input uncertainty, readers are referred to Barton et al. (2002), Henderson (2003), Chick (2006), Barton (2012), Song et al. (2014), Lam (2016a), and Nelson (2013) Chapter 7.

The remainder of the paper is as follows. Section 2 introduces the input uncertainty problem and explains the simulation complexity bottleneck in the existing bootstrap schemes. Section 3 presents our subsampling idea, procedures and the main statistical results. Section 4 discusses the key steps in our theoretical developments. Section 5 reports our numerical experiments. Section 6 concludes the paper. All proofs are relegated to the Appendix.

2. Problem Motivation

This section describes the problem and our motivation. Section 2.1 first describes the input uncertainty problem, Section 2.2 presents the existing bootstrap approach, and Section 2.3 discusses its computational barrier, thus motivating our subsampling investigation. We aim to provide intuitive explanations in this section, and defer mathematical details to later sections.

2.1. The Input Uncertainty Problem

Suppose there are m independent input processes driven by input distributions F_1, F_2, \dots, F_m .

We consider a generic performance measure $\psi(F_1, \dots, F_m)$ that is simulable, i.e., given the input distributions, independent unbiased replications of ψ can be generated in a computer. As a primary example, think of F_1 and F_2 as the interarrival and service time distributions in a queue, and ψ is some output measure such as the mean queue length averaged over a time horizon. Our study also applies when the F_i 's are multivariate distributions.

The input uncertainty problem arises in situations where the input distributions F_1, \dots, F_m are unknown but real-world data are available. One then has to use their estimates $\hat{F}_1, \dots, \hat{F}_m$ to drive the simulation. Denote a point estimate of $\psi(F_1, \dots, F_m)$ as $\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)$, where typically we take

$$\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m) = \frac{1}{q} \sum_{r=1}^q \hat{\psi}_r(\hat{F}_1, \dots, \hat{F}_m)$$

with $\hat{\psi}_r(\hat{F}_1, \dots, \hat{F}_m)$ being a conditionally unbiased simulation replication driven by $\hat{F}_1, \dots, \hat{F}_m$. This point estimate is affected by both the input statistical noises and the simulation noises. By conditioning on the estimated input distributions (or viewing the point estimate as a random effect model with uncorrelated input and simulation noises), the variance of $\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)$ can be expressed as

$$\text{Var}[\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)] = \sigma_I^2 + \sigma_S^2$$

where

$$\sigma_I^2 = \text{Var}[\psi(\hat{F}_1, \dots, \hat{F}_m)] \tag{1}$$

is the input variance, and

$$\sigma_S^2 = \frac{\mathbb{E}[\text{Var}[\hat{\psi}_r(\hat{F}_1, \dots, \hat{F}_m) | \hat{F}_1, \dots, \hat{F}_m]]}{q}$$

is the variance contributed from the simulation noises. Assuming that the estimates \hat{F}_i 's are consistent in estimating F_i 's, then, as input data sizes grow, σ_S^2 is approximately $\text{Var}[\hat{\psi}_r(F_1, \dots, F_m)]/q$ and can be estimated by taking the sample variance of all simulation replications (see, e.g., Cheng

and Holland (1997)). On the other hand, σ_I^2 signifies the output variance contributed solely from the input data noises, assuming a fully accurate evaluation of the performance measure ψ . Estimating σ_I^2 is the key and the challenge in quantifying input uncertainty, which is the focus of this paper.

Before going into details, we discuss two conceptual properties on σ_I^2 that would be relevant in motivating and pinpointing our study. First, suppose further that for each input model i , we have n_i i.i.d. data $\{X_{i,1}, \dots, X_{i,n_i}\}$ generated from the distribution F_i . When n_i 's are large, typically the overall input variance σ_I^2 is decomposable into

$$\sigma_I^2 \approx \sum_{i=1}^m \frac{\sigma_i^2}{n_i} \quad (2)$$

where σ_i^2/n_i is the variance contributed from the data noise for model i , with σ_i^2 being a constant. In the parametric case where \widehat{F}_i comes from a parametric family containing the estimated parameters, this decomposition is well known from the delta method (Asmussen and Glynn (2007), Chapter 3). Here, σ_i^2/n_i is typically $\nabla_i \psi' \Sigma_i \nabla_i \psi$, where $\nabla_i \psi$ is the collection of sensitivity coefficients, i.e., the gradient, with respect to the parameters in model i , and Σ_i is the asymptotic estimation variance of the point estimates of these parameters (scaled reciprocally with n_i). In the nonparametric case where the empirical distribution $\widehat{F}_i(x) := \sum_{j=1}^{n_i} \delta_{X_{i,j}}(x)/n_i$ is used (where $\delta_{X_{i,j}}$ denotes the delta measure at $X_{i,j}$), (2) still holds under mild conditions (e.g., Propositions 1 and 2 in the sequel). In this setting the quantity σ_i^2 is equal to $\text{Var}_{F_i}[g_i(X_i)]$, where $g_i(\cdot)$ is the influence function (Hampel (1974)) of ψ with respect to the distribution F_i , whose domain is the value space of the input variate X_i , and $\text{Var}_{F_i}[\cdot]$ denotes the variance under F_i . The influence function can be viewed as a functional derivative taken with respect to the probability distributions F_i 's (see Serfling (2009), Chapter 6), and dictates the first-order asymptotic behavior of the plug-in estimate of ψ . Although the mathematical form of σ_i^2 's is known, it relies on gradient information that needs to be estimated via simulation itself. Moreover, in the nonparametric case, the gradient dimension in a sense grows with the data size. Thus directly using the delta method in this case could be challenging. In our subsequent developments, we focus on the nonparametric case, both because this is more

challenging, and also that this can be viewed as a generalization of the parametric case by viewing the “parameter” simply as a function of F_i ’s.

Second, under further regularity conditions, a Gaussian approximation holds for $\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)$ so that

$$\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m) \pm z_{1-\alpha/2} \sqrt{\sigma_I^2 + \sigma_S^2} \quad (3)$$

is an asymptotically tight $(1 - \alpha)$ -level CI for $\psi(F_1, \dots, F_m)$, where $z_{1-\alpha/2}$ is the standard normal $1 - \alpha/2$ quantile. This CI, which provides a bound-based alternative to quantify input uncertainty, again requires a statistically valid estimate of σ_I^2 or $\sum_{i=1}^m \sigma_i^2/n_i$ (and σ_S^2). In this paper we primarily focus on the estimation of σ_I^2 and how our proposed approach substantially improves upon previous methods in this regard. Naturally, the improved estimate of σ_I^2 also translates into a better CI when using (3). We caution, however, that an optimal procedural configuration to estimate σ_I^2 does not necessarily correspond to an optimal configuration in constructing the CI, as the performance of the latter is measured by different criteria such as coverage or half-width (such a difference in optimally estimating variance versus CI has also been observed in other contexts such as time series (Sun et al. (2008))). Nonetheless, we will show that a direct plug-in of our new estimator of σ_I^2 into (3) is already enough to significantly outperform conventional bootstrap-based CIs suggested in the literature, both theoretically and also supported by consistent empirical evidence.

Next we will discuss bootstrap resampling, the commonest estimation technique that forms the basis of our comparison.

2.2. Bootstrap Resampling

Let \hat{F}_i^* represent the empirical distribution constructed using a bootstrap resample from the original data $\{X_{i,1}, \dots, X_{i,n_i}\}$ for input F_i , i.e., n_i points drawn by uniformly sampling with replacement from $\{X_{i,1}, \dots, X_{i,n_i}\}$. The bootstrap variance estimator is $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)]$, where $\text{Var}_*[\cdot]$ denotes the variance over the bootstrap resamples from the data, conditional on $\hat{F}_1, \dots, \hat{F}_m$.

The principle of bootstrap entails that $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)] \approx \text{Var}[\psi(\hat{F}_1, \dots, \hat{F}_m)] = \sigma_I^2$. Here $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)]$ is obtained from a (hypothetical) infinite number of bootstrap resamples and

simulation runs per resample. In practice, however, one would need to use a finite bootstrap size and a finite simulation size. This comprises B conditionally independent bootstrap resamples of $\{\widehat{F}_1^*, \dots, \widehat{F}_m^*\}$, and R simulation replications driven by each realization of the resampled input distributions. This generally incurs two layers of Monte Carlo errors.

Denote $\hat{\psi}_r(\widehat{F}_1^b, \dots, \widehat{F}_m^b)$ as the r -th simulation run driven by the b -th bootstrap resample $\{\widehat{F}_1^b, \dots, \widehat{F}_m^b\}$. Denote $\bar{\psi}^b$ as the average of the R simulation runs driven by the b -th resample, and $\bar{\bar{\psi}}$ as the grand sample average from all the BR runs. An unbiased estimator for $\text{Var}_*[\psi(\widehat{F}_1^*, \dots, \widehat{F}_m^*)]$ is given by

$$\frac{1}{B-1} \sum_{b=1}^B (\bar{\psi}^b - \bar{\bar{\psi}})^2 - \frac{V}{R} \quad (4)$$

where

$$V = \frac{1}{B(R-1)} \sum_{b=1}^B \sum_{r=1}^R (\hat{\psi}_r(\widehat{F}_1^b, \dots, \widehat{F}_m^b) - \bar{\psi}^b)^2.$$

To explain, the first term in (4) is an unbiased estimate of the variance of $\bar{\psi}^b$, which is $\text{Var}_*[\psi(\widehat{F}_1^*, \dots, \widehat{F}_m^*)] + (1/R)\mathbb{E}_*[\text{Var}[\hat{\psi}_r(\widehat{F}_1^*, \dots, \widehat{F}_m^*)|\widehat{F}_1^*, \dots, \widehat{F}_m^*]]$ (where $\mathbb{E}_*[\cdot]$ denotes the expectation on \widehat{F}_i^* 's conditional on \widehat{F}_i 's), since $\bar{\psi}^b$ incurs both the bootstrap noise and the simulation noise. In other words, the variance of $\bar{\psi}^b$ is upward biased for $\text{Var}_*[\psi(\widehat{F}_1^*, \dots, \widehat{F}_m^*)]$. The second term in (4), namely V/R , removes this bias. This bias adjustment can be derived by viewing $\text{Var}_*[\psi(\widehat{F}_1^*, \dots, \widehat{F}_m^*)]$ as the variance of a conditional expectation. Alternately, $\hat{\psi}_r(\widehat{F}_1^*, \dots, \widehat{F}_m^*)$ can be viewed as a random effect model where each “group” corresponds to each realization of $\widehat{F}_1^*, \dots, \widehat{F}_m^*$, and (4) estimates the “between-group” variance in an analysis-of-variance (ANOVA). Formula (4) has appeared in the input uncertainty literature, e.g., Cheng and Holland (1997), Song and Nelson (2015), Lin et al. (2015), and also in Zouaoui and Wilson (2004) in the Bayesian context. Algorithm 1 summarizes the procedure.

More generally, to estimate the variance contribution from the data noise of model i only, namely σ_i^2/n_i , one can bootstrap only from $\{X_{i,1}, \dots, X_{i,n_i}\}$ and keep other input distributions $\widehat{F}_j, j \neq i$ fixed. Then \widehat{F}_i^* and $\widehat{F}_j, j \neq i$ are used to drive the simulation runs. With this modification, the same formula (4) or Algorithm 1 is an unbiased estimate for $\text{Var}_*[\psi(\widehat{F}_1, \dots, \widehat{F}_{i-1}, \widehat{F}_i^*, \widehat{F}_{i+1}, \dots, \widehat{F}_m)]$,

Algorithm 1 ANOVA-based Variance Bootstrap

Given: $B \geq 2, R \geq 2$; data = $\{X_{i,j} : i = 1, \dots, m, j = 1, \dots, n_i\}$

for $b = 1$ **to** B **do**

For each i , draw a sample $\{X_{i,1}^b, \dots, X_{i,n_i}^b\}$ uniformly with replacement from the data to obtain a resampled empirical distribution \widehat{F}_i^b

for $r = 1$ **to** R **do**

Simulate $\widehat{\psi}_r(\widehat{F}_1^b, \dots, \widehat{F}_m^b)$

end for

Compute $\bar{\psi}_{BV}^b = \frac{1}{R} \sum_{r=1}^R \widehat{\psi}_r(\widehat{F}_1^b, \dots, \widehat{F}_m^b)$

end for

Compute $V = \frac{1}{B(R-1)} \sum_{b=1}^B \sum_{r=1}^R (\widehat{\psi}_r(\widehat{F}_1^b, \dots, \widehat{F}_m^b) - \bar{\psi}_{BV}^b)^2$ and $\bar{\bar{\psi}}_{BV} = \frac{1}{B} \sum_{b=1}^B \bar{\psi}_{BV}^b$

Output $\hat{\sigma}_{BV}^2 = \frac{1}{B-1} \sum_{b=1}^B (\bar{\psi}_{BV}^b - \bar{\bar{\psi}}_{BV})^2 - \frac{V}{R}$

which is approximately $\text{Var}[\psi(F_1, \dots, F_{i-1}, \widehat{F}_i, F_{i+1}, \dots, F_m)]$ by the bootstrap principle, in turn asymptotically equal to σ_i^2/n_i introduced in (2). This observation appeared in, e.g., Song et al. (2014); in Section 4 we give further justifications.

In subsequent discussions, we use the following notations. For any sequences a and b , both depending on some parameter, say, n , we say that $a = O(b)$ if $|a/b| \leq C$ for some constant $C > 0$ for all sufficiently large n , and $a = o(b)$ if $a/b \rightarrow 0$ as $n \rightarrow \infty$. Alternately, we say $a = \Omega(b)$ if $|a/b| \geq C$ for some constant $C > 0$ for all sufficiently large n , and $a = \omega(b)$ if $|a/b| \rightarrow \infty$ as $n \rightarrow \infty$. We say that $a = \Theta(b)$ if $\underline{C} \leq |a/b| \leq \bar{C}$ as $n \rightarrow \infty$ for some constants $\underline{C}, \bar{C} > 0$. We use $A = O_p(b)$ to represent a random variable A that has stochastic order at least b , i.e., for any $\epsilon > 0$, there exists $M, N > 0$ such that $P(|A/b| \leq M) > 1 - \epsilon$ for $n > N$. We use $A = o_p(b)$ to represent a random variable A that has stochastic order less than b , i.e., $A/b \xrightarrow{p} 0$. We use $A = \Theta_p(b)$ to represent a random variable A that has stochastic order exactly at b , i.e., A satisfies $A = O_p(b)$ but not $A = o_p(b)$.

2.3. A Complexity Barrier

We explain intuitively the total number of simulation runs needed to ensure that the variance bootstrap depicted above can meaningfully estimate the input variance. For convenience, we call this number the *simulation complexity*. This quantity turns out to be of order bigger than the data size. On a high level, it is because the input variance scales reciprocally with the data size (recall (2)). Thus, when the data size increases, the input variance becomes smaller and increasingly difficult to estimate with controlled relative error. This in turn necessitates the use of more simulation runs.

To explain more concretely, denote n as a scaling of the data size, i.e., we assume n_i all grow linearly with n , which in particular implies that σ_I^2 is of order $1/n$. We analyze the error of $\hat{\sigma}_{BV}^2$ from Algorithm 1 in estimating σ_I^2 . Since $\hat{\sigma}_{BV}^2$ is unbiased for $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)]$ which is in turn close to σ_I^2 , roughly speaking it suffices to focus on the variance of $\hat{\sigma}_{BV}^2$. To analyze this latter quantity, we denote a generic simulation run in our procedure, $\hat{\psi}_r(\hat{F}_1^*, \dots, \hat{F}_m^*)$, as

$$\hat{\psi}_r(\hat{F}_1^*, \dots, \hat{F}_m^*) = \psi(\hat{F}_1, \dots, \hat{F}_m) + \delta + \xi$$

where

$$\delta := \psi(\hat{F}_1^*, \dots, \hat{F}_m^*) - \psi(\hat{F}_1, \dots, \hat{F}_m), \quad \xi := \hat{\psi}_r(\hat{F}_1^*, \dots, \hat{F}_m^*) - \psi(\hat{F}_1^*, \dots, \hat{F}_m^*).$$

are the errors arising from the bootstrap of the input distributions and the simulation respectively. If ψ is sufficiently smooth, δ elicits a central limit theorem and is of order $\Theta_p(1/\sqrt{n})$. On the other hand, the simulation noise ξ is of order $\Theta_p(1)$.

Via an ANOVA-type analysis as in Sun et al. (2011), we have

$$\begin{aligned} \text{Var}_*[\hat{\sigma}_{BV}^2] &= \frac{1}{B} (\mathbb{E}_*[\delta^4] - (\mathbb{E}_*[\delta^2])^2) + \frac{2}{B(B-1)} (\mathbb{E}_*[\delta^2])^2 + \frac{2}{B^2 R^2 (B-1)} (\mathbb{E}_*[\xi^2])^2 + \frac{2}{B^2 R^3} \mathbb{E}_*[\xi^4] \\ &\quad + \frac{2(B+1)}{B^2 R (B-1)} \mathbb{E}_*[\delta^2] \mathbb{E}_*[\xi^2] + \frac{2(BR^2 + R^2 - 4R + 3)}{B^2 R^3 (B-1)} \mathbb{E}_*[(\mathbb{E}_*[\xi^2 | \hat{F}_1^*, \dots, \hat{F}_m^*])^2] \\ &\quad + \frac{4B+2}{B^2 R} \mathbb{E}_*[\delta^2 \xi^2] + \frac{4}{B^2 R^2} \mathbb{E}_*[\delta \xi^3]. \end{aligned} \tag{5}$$

Now, putting $\delta = \Theta_p(1/\sqrt{n})$ and $\xi = \Theta_p(1)$ formally into (5), and ignoring constant factors, results in

$$\text{Var}_*[\hat{\sigma}_{BV}^2] = O_p \left(\frac{1}{Bn^2} + \frac{1}{B^2n^2} + \frac{1}{B^3R^2} + \frac{1}{B^2Rn} + \frac{1}{B^2R^3} + \frac{1}{BR^2} + \frac{1}{BRn} + \frac{1}{B^2R^2\sqrt{n}} \right)$$

or simply

$$O_p \left(\frac{1}{Bn^2} + \frac{1}{BR^2} \right) \quad (6)$$

The two terms in (6) correspond to the variances coming from the bootstrap resampling and the simulation runs respectively.

Since σ_I^2 is of order $1/n$, meaningful estimation of σ_I^2 needs measured by the relative error. In other words, we want to achieve $\hat{\sigma}_{BV}^2/\sigma_I^2 \xrightarrow{p} 1$ as the simulation budget grows. This property, which we call relative consistency, requires $\hat{\sigma}_{BV}^2$ to have a variance of order $o(1/n^2)$ in order to compensate for the decreasing order of σ_I^2 .

We argue that this implies unfortunately that the total number of simulation runs, BR , must be $\omega(n)$, i.e., of order higher than the data size. To explain, note that the first term in (6) forces one to use $B = \omega(1)$, i.e., the bootstrap size needs to grow with n , an implication that is quite natural. The second term in (6), on the other hand, dictates also that $BR^2 = \omega(n^2)$. Suppose, for the sake of contradiction, that B and R are chosen such that $BR = O(n)$. Then, because we need $BR \times R = BR^2 = \omega(n^2)$, R must be $\omega(n)$ which, combining with $B = \omega(1)$, implies that $BR = \omega(n)$ and leads to a contradiction.

We summarize the above with the following result. Let N be the total simulation effort, and recall n as the scaling of the data size. We have:

THEOREM 1 (Simulation complexity of the variance bootstrap). *Under Assumptions 1-7 to be stated in Section 4.1, the required simulation budget to achieve relative consistency in estimating σ_I^2 by Algorithm 1, i.e., $\hat{\sigma}_{BV}^2/\sigma_I^2 \xrightarrow{p} 1$, is $N = \omega(n)$.*

Though out of the scope of this paper, there are indications that such a computational barrier occurs in other types of bootstrap. For instance, the percentile bootstrap studied in Barton and

Schruben (1993, 2001) appears to also require an inner replication size large enough compared to the data size in order to obtain valid quantile estimates (the authors actually used one inner replication, but Barton (2012) commented that more is needed). Yi and Xie (2017) provides an interesting approach based on ranking and selection to reduce the simulation effort, though they do not investigate the order of the needed effort relative to the data size. The empirical likelihood framework studied in Lam and Qian (2017) requires a similarly higher order of simulation runs to estimate the influence function. Nonetheless, in this paper we focus only on how to reduce computation load in variance estimation.

3. Procedures and Guarantees in the Subsampling Framework

This section presents our methodologies and results on subsampling. Section 3.1 first explains the rationale and the subsampling procedure. Section 3.2 then presents our main theoretical guarantees, deferring some elaborate developments to Section 4.

3.1. Proportionate Subsampled Variance Bootstrap

As explained before, the reason why the $\hat{\sigma}_{BV}^2$ in Algorithm 1 requires a huge simulation effort, as implied by its variance (6), lies in the small scale of the input variance. In general, in order to estimate a quantity that is of order $1/n$, one must use a sample size more than n so that the estimation error relatively vanishes. This requirement manifests in the inner replication size in constructing $\hat{\sigma}_{BV}^2$.

To reduce the inner replication size, we leverage the relation between the form of the input variance and the estimation variance depicted in (6) as follows. The approximate input variance contributed from model i , with data size n_i , has the form σ_i^2/n_i . If we use the variance bootstrap directly as in Algorithm 1, then we need an order more than n total simulation runs due to (6). Now, pretend that we have fewer data, say s_i , then the input variance will be σ_i^2/s_i , and the required simulation runs is now only of order higher than s_i . An estimate of σ_i^2/s_i , however, already gives us enough information in estimating σ_i^2/n_i , because we can rescale our estimate of σ_i^2/s_i by s_i/n_i .

to get an estimate of σ_i^2/n_i . Estimating σ_i^2/s_i can be done by subsampling the input distribution with size s_i . With this, we can both use fewer simulation runs and also retain correct estimation via multiplying by a s_i/n_i factor.

To make the above argument more transparent, the bootstrap principle and the asymptotic approximation of the input variance imply that

$$\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)] = \sum_{i=1}^m \frac{\sigma_i^2}{n_i} (1 + o_p(1)).$$

The subsampling approach builds on the observation that a similar relation holds for

$$\text{Var}_*[\psi(\hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*)] = \sum_{i=1}^m \frac{\sigma_i^2}{s_i} (1 + o_p(1))$$

where $\hat{F}_{s_i,i}^*$ denotes a bootstrapped input distribution of size s_i (i.e., an empirical distribution of size s_i that is uniformly sampled with replacement from $\{X_{i,1}, \dots, X_{i,n_i}\}$). If we let $s_i = \lfloor \theta n_i \rfloor$ for some $\theta > 0$ so that $s_i \rightarrow \infty$ (where $\lfloor \cdot \rfloor$ is the floor function, i.e. the largest integer less than or equal to \cdot), then we have

$$\text{Var}_*[\psi(\hat{F}_{\lfloor \theta n_1 \rfloor,1}^*, \dots, \hat{F}_{\lfloor \theta n_m \rfloor,m}^*)] = \sum_{i=1}^m \frac{\sigma_i^2}{\theta n_i} (1 + o_p(1)).$$

Multiplying both sides with θ , we get

$$\theta \text{Var}_*[\psi(\hat{F}_{\lfloor \theta n_1 \rfloor,1}^*, \dots, \hat{F}_{\lfloor \theta n_m \rfloor,m}^*)] = \sum_{i=1}^m \frac{\sigma_i^2}{n_i} (1 + o_p(1)).$$

Note that the right hand side above is the original input variance of interest. This leads to our *proportionate subsampled variance bootstrap*: We repeatedly subsample collections of input distributions from the data, with size $\lfloor \theta n_i \rfloor$ for model i , and use them to drive simulation replications. We then apply the ANOVA-based estimator in (4) on these replications, and multiply it by a factor of θ to obtain our final estimate. We summarize this procedure in Algorithm 2. The term “proportionate” refers to the fact that we scale the subsample size for all models with a single factor θ . For convenience, we call θ the *subsample ratio*.

Similar ideas apply to estimating the individual variance contribution from each input model, namely σ_i^2/n_i . Instead of subsampling all input distributions, we only subsample the distribution,

Algorithm 2 Proportionate Subsampled Variance Bootstrap

Parameters: $B \geq 2, R \geq 2, 0 < \theta \leq 1$; data = $\{X_{i,j} : i = 1, \dots, m, j = 1, \dots, n_i\}$

Compute $s_i = \lfloor \theta n_i \rfloor$ for all i

for $b = 1$ **to** B **do**

For each i , draw a subsample $\{X_{i,1}^b, \dots, X_{i,s_i}^b\}$ uniformly with replacement from the data, which

forms the empirical distribution $\widehat{F}_{s_i,i}^b$

for $r = 1$ **to** R **do**

Simulate $\widehat{\psi}_r(\widehat{F}_{s_1,1}^b, \dots, \widehat{F}_{s_m,m}^b)$

end for

Compute $\bar{\psi}^b = \frac{1}{R} \sum_{r=1}^R \widehat{\psi}_r(\widehat{F}_{s_1,1}^b, \dots, \widehat{F}_{s_m,m}^b)$

end for

Compute $V = \frac{1}{B(R-1)} \sum_{b=1}^B \sum_{r=1}^R (\widehat{\psi}_r(\widehat{F}_{s_1,1}^b, \dots, \widehat{F}_{s_m,m}^b) - \bar{\psi}^b)^2$ and $\bar{\bar{\psi}} = \frac{1}{B} \sum_{b=1}^B \bar{\psi}^b$

Output $\widehat{\sigma}_{SVB}^2 = \theta \left(\frac{1}{B-1} \sum_{b=1}^B (\bar{\psi}^b - \bar{\bar{\psi}})^2 - \frac{V}{R} \right)$

say $\widehat{F}_{s_i,i}^*$ whose uncertainty is of interest, while fixing all the other distributions as the original empirical distributions, i.e., $\widehat{F}_j, j \neq i$. All the remaining steps in Algorithm 2 remain the same (thus the “proportionate” part can be dropped). This procedure is depicted in Algorithm 3.

3.2. Statistical Guarantees

Algorithm 2 provides the following guarantees. Recall that $N = BR$ is the total simulation effort, and n is the scaling of the data size. We have the following result:

THEOREM 2 (Procedural configurations to achieve relative consistency). *Under Assumptions 1-7 to be stated in Section 4.1, if the parameters B, R, θ of Algorithm 2 are chosen such that*

$$B = \omega(1), BR^2 = \omega((\theta n)^2), \theta = \omega\left(\frac{1}{n}\right) \quad (7)$$

then the variance estimate $\widehat{\sigma}_{SVB}^2$ is relatively consistent, i.e. $\widehat{\sigma}_{SVB}^2 / \sigma_I^2 \xrightarrow{P} 1$.

Algorithm 3 Subsampled Variance Bootstrap for Variance Contribution from the i -th Input

Model

Parameters: $B \geq 2, R \geq 2, 0 < \theta \leq 1$; data = $\{X_{i,j} : i = 1, \dots, m, j = 1, \dots, n_i\}$

Compute $s_i = \lfloor \theta n_i \rfloor$

for $b = 1$ **to** B **do**

 Draw a subsample $\{X_{i,1}^b, \dots, X_{i,s_i}^b\}$ uniformly with replacement from the i -th input data set,

 which forms the empirical distribution $\widehat{F}_{s_i,i}^b$

for $r = 1$ **to** R **do**

 Simulate $\widehat{\psi}_r(\widehat{F}_1, \dots, \widehat{F}_{i-1}, \widehat{F}_{s_i,i}^b, \widehat{F}_{i+1}, \dots, \widehat{F}_m)$

end for

 Compute $\bar{\psi}^b = \frac{1}{R} \sum_{r=1}^R \widehat{\psi}_r(\widehat{F}_1, \dots, \widehat{F}_{i-1}, \widehat{F}_{s_i,i}^b, \widehat{F}_{i+1}, \dots, \widehat{F}_m)$

end for

 Compute $V = \frac{1}{B(R-1)} \sum_{b=1}^B \sum_{r=1}^R (\widehat{\psi}_r(\widehat{F}_1, \dots, \widehat{F}_{i-1}, \widehat{F}_{s_i,i}^b, \widehat{F}_{i+1}, \dots, \widehat{F}_m) - \bar{\psi}^b)^2$ and $\bar{\bar{\psi}} = \frac{1}{B} \sum_{b=1}^B \bar{\psi}^b$

 Output $\widehat{\sigma}_{SVB,i}^2 = \theta \left(\frac{1}{B-1} \sum_{b=1}^B (\bar{\psi}^b - \bar{\bar{\psi}})^2 - \frac{V}{R} \right)$

Theorem 2 tells us what orders of the bootstrap size B , inner replication size R and subsample ratio θ would guarantee a meaningful estimation of σ_I^2 . Note that $\theta \approx s_i/n_i$ for each i , so that $\theta = \omega(1/n)$ is equivalent to setting the subsample size $s_i = \omega(1)$. In other words, we need the natural requirement that the subsample size grows with the data size, albeit can have an arbitrary rate.

Given a subsample ratio θ specified according to (7), the configurations of B and R under (7) that achieve the minimum overall simulation budget is $B = \omega(1)$ and $R = \Omega(\theta n)$. This is because to minimize $N = BR$ while satisfying the second requirement in (7), it is more economical to allocate as much budget to R instead of B . This is stated precisely as:

COROLLARY 1 (Minimum configurations to achieve relative consistency). *Under the same conditions of Theorem 2, given $\theta = \omega(n^{-1})$, the values of B and R to achieve (7) and hence relative consistency that requires the least order of effort are $B = \omega(1)$ and $R = \Omega(\theta n)$, leading to a total simulation budget $N = \omega(\theta n)$.*

Note that θn is the order of the subsample size. Thus Corollary 1 implies that the required simulation budget must be of higher order than the subsample size. However, since the subsample size can be chosen to grow at an arbitrarily small rate, this implies that the total budget can also grow arbitrarily slowly. Therefore, we have:

COROLLARY 2 (Simulation complexity of proportionate subsampled variance bootstrap).

Under the same conditions of Theorem 2, the minimum required simulation budget to achieve relative consistency in estimating σ_I^2 by Algorithm 2, i.e., $\hat{\sigma}_{SVB}^2/\sigma_I^2 \xrightarrow{P} 1$, is $N = \omega(1)$ by using $\theta = \omega(n^{-1})$.

Compared to Theorem 1, Corollary 2 stipulates that our subsampling approach reduces the required simulation effort from a higher order than n to an arbitrary order, i.e., independent of the data size. This is achieved by using a subsample size that grows with n at an arbitrary order, or equivalently a subsample ratio θ that grows faster than $1/n$.

The following result describes the configurations of our scheme when a certain total simulation effort is given. In particular, it shows, for a given total simulation effort, the range of subsample ratio for which Algorithm 2 can possibly generate valid variance estimates by appropriately choosing B and R :

THEOREM 3 (Valid subsample ratio given total budget). *Assume the same conditions of Theorem 2. Given a total simulation budget $N = \omega(1)$, if the subsample ratio satisfies $\omega(1/n) \leq \theta \leq o(N/n) \wedge 1$, then the bootstrap size B and the inner replication size R can be appropriately chosen according to criterion (7) to achieve relative consistency, i.e., $\hat{\sigma}_{SVB}^2/\sigma_I^2 \xrightarrow{P} 1$.*

The next result is on the optimal configurations of our scheme in minimizing the Monte Carlo error. To proceed, define

$$\sigma_{SVB}^2 = \theta \text{Var}_*[\psi(\hat{F}_{\lfloor \theta n_1 \rfloor, 1}^*, \dots, \hat{F}_{\lfloor \theta n_m \rfloor, m}^*)] \quad (8)$$

as the perfect form of our proportionate subsampled variance bootstrap introduced in Section 3.1, namely without any Monte Carlo noises, and $0 < \theta \leq 1$ is the subsample ratio. We have:

THEOREM 4. *Assume the same conditions of Theorem 2. Given a simulation budget N and a subsample ratio θ such that $N = \omega(\theta n)$ and $\theta = \omega(n^{-1})$, the optimal outer and inner sizes that minimize the order of the conditional mean squared error $\mathbb{E}_*[(\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2)^2]$ are*

$$B^* = \frac{N}{R^*}, \quad R^* = \Theta(\theta n)$$

giving a conditional mean squared error $\mathbb{E}_[(\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2)^2] = \Theta(\theta/(Nn))(1 + o_p(1))$.*

Note that the mean squared error, i.e. $\mathbb{E}_*[(\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2)^2]$, of the Monte Carlo estimate $\hat{\sigma}_{SVB}^2$ is random because the underlying resampling is conditioned on the input data, therefore the bound at the end of Theorem 4 contains a stochastically vanishing term $o_p(1)$.

We next present the optimal tuning of the subsample ratio. This requires a balance of the trade-off between the input statistical error and the Monte Carlo simulation error. To explain, the overall error of $\hat{\sigma}_{SVB}^2$ by Algorithm 2 can be decomposed as

$$\hat{\sigma}_{SVB}^2 - \sigma_I^2 = (\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2) + (\sigma_{SVB}^2 - \sigma_I^2). \quad (9)$$

The first term is the Monte Carlo error for which the optimal outer size B , inner size R and the resulting mean squared error are governed by Theorem 4. In particular, the mean squared error there shows that under a fixed simulation budget N and the optimal allocation $R = \Theta(\theta n)$, the Monte Carlo error gets larger as θ increases. The second term is the statistical errors due to the finiteness of input data and θ . Since θ measures the amount of data contained in the resamples, we expect this second error to become smaller as θ increases. The optimal tuning of θ relies on balancing such a trade-off between the two sources of errors.

We have the following optimal configurations of B , R and θ altogether given a budget N :

THEOREM 5 (Optimal subsample size and budget allocation). *Suppose Assumptions 1, 3-7 in Section 4.1 and Assumptions 10-12 in Section 4.3 hold. For a given simulation budget $N = \omega(1)$, if the subsample ratio θ and outer and inner sizes B, R for Algorithm 2 are set to*

$$\begin{cases} \theta^* = \Theta(N^{1/3}n^{-1}) & \text{if } 1 \ll N \leq n^{3/2} \\ \Theta(n^{-1/2}) \leq \theta^* \leq \Theta(Nn^{-2} \wedge 1) & \text{if } N > n^{3/2} \end{cases} \quad (10)$$

$$R^* = \Theta(\theta^* n), B^* = \frac{N}{R^*} \quad (11)$$

then the gross error $\hat{\sigma}_{SVB}^2 - \sigma_I^2 = \mathcal{E} + o_p(N^{-1/3}n^{-1} + n^{-3/2})$, where the leading term has a mean squared error

$$\mathbb{E}[\mathcal{E}^2] = O\left(\frac{1}{N^{2/3}n^2} + \frac{1}{n^3}\right). \quad (12)$$

Moreover, if $\mathcal{R} = \Theta((ns)^{-1})$ and at least one of the Σ_i 's are positive definite, where \mathcal{R} and Σ_i are as defined in Lemma 2, then (12) holds with an exact order (i.e., $O(\cdot)$ becomes $\Theta(\cdot)$) and the configuration (10), (11) is optimal in the sense that no configuration gives rise to a gross error $\hat{\sigma}_{SVB}^2 - \sigma_I^2 = o_p(N^{-1/3}n^{-1} + n^{-3/2})$.

Note from (12) that, if the budget $N = \omega(1)$, our optimal configurations guarantee the estimation mean squared error decays faster than $1/n^2$. Recall that the input variance is of order $1/n$, and thus an estimation error of order higher than $1/n^2$ ensures that the estimator is relatively consistent in the sense $\hat{\sigma}_{SVB}^2/\sigma_I^2 \xrightarrow{p} 1$. This recovers the result in Corollary 2. We also comment that the algorithmic configuration given in Theorem 5 is chosen to optimize the mean squared error of the input variance estimate, but does not necessarily generates the most accurate CI. There exists evidence (e.g., Sun et al. (2008)) that the optimal choice to minimize the mean squared error of the variance estimate can be different from the one that is optimal for statistical inference, although in our experiments they seem to match closely with each other.

We comment that all the results in this section hold if one estimates the individual variance contribution from each input model i , namely by using Algorithm 3. In this case we are interested in estimating the variance σ_i^2/n_i , and relative consistency means $\hat{\sigma}_{SVB,i}^2/(\sigma_i^2/n_i) \xrightarrow{p} 1$. The data size scaling parameter n can be replaced by n_i in all our results.

Finally, we also comment that the complexity barrier described in Section 2.3 and our framework presented in this section applies in principle to the parametric regime, i.e., when the input distributions are known to lie in parametric families with unknown parameters. The assumptions and mathematical details would need to be catered to that situation, which could be done naturally by viewing the “parameter” as a function of F_i 's.

4. Developments of Theoretical Results

We present our main developments leading to the algorithms and results in Section 3. Section 4.1 first states in detail our assumptions on the performance measure. Section 4.2 presents the theories leading to estimation accuracy, simulation complexity and optimal budget allocation in the proportionate subsampled variance bootstrap. Section 4.3 investigates optimal subsample sizes that lead to overall best configurations.

4.1. Regularity Assumptions

We first assume that the data sets for all input models are of comparable size.

ASSUMPTION 1 (Balanced data). $\limsup_{\text{all } n_i \rightarrow \infty} \frac{\max_i n_i}{\min_i n_i} < \infty$ as all $n_i \rightarrow \infty$.

Recall in Sections 2 and 3 that we have denoted n as a scaling of the data size. More concretely, we take $n = (1/m) \sum_{i=1}^m n_i$ as the average input data size under Assumption 1.

We next state a series of general assumptions on the performance measure ψ . These assumptions hold for common finite-horizon measures, as we will present. For each i let Ξ_i be the support of the i -th true input model F_i , and the collection of distributions \mathcal{P}_i be the convex hull spanned by F_i and all Dirac measures on Ξ_i , i.e.

$$\mathcal{P}_i = \left\{ \nu_1 F_i + \sum_{k=2}^l \nu_k \mathbf{1}_{x_k} : \sum_{k=1}^l \nu_k = 1, \nu_k \geq 0, l < \infty, x_k \in \Xi_i \text{ for all } k \right\}.$$

We assume the following differentiability of the performance measure.

ASSUMPTION 2 (First order differentiability). For any distributions $P_i, Q_i \in \mathcal{P}_i$, denote $P_i^{\nu_i} = (1 - \nu_i)P_i + \nu_i Q_i$ for $\nu_i \in [0, 1]$. Assume there exist functions $g_i(P_1, \dots, P_m; \cdot) : \Xi_i \rightarrow \mathbb{R}$ such that $\mathbb{E}_{P_i}[g_i(P_1, \dots, P_m; X_i)] = 0$ for $i = 1, \dots, m$ and as all ν_i 's approach zero

$$\psi(P_1^{\nu_1}, \dots, P_m^{\nu_m}) - \psi(P_1, \dots, P_m) = \sum_{i=1}^m \nu_i \int g_i(P_1, \dots, P_m; x) d(Q_i - P_i)(x) + o\left(\sqrt{\sum_{i=1}^m \nu_i^2}\right). \quad (13)$$

The differentiability described above is defined with respect to a particular direction, namely $Q_i - P_i$, in the space of probability measures, and is known as Gateaux differentiability or directional

differentiability (e.g., Serfling (2009), Van der Vaart (2000)). Assumption 2 therefore requires the performance measure ψ to be Gateaux differentiable when restricted to the convex set $\mathcal{P}_1 \times \cdots \times \mathcal{P}_m$. The functions g_i 's are also called the influence functions (e.g., Hampel (1974)) that play analogous roles as standard gradients in the Euclidean space. The condition of g_i 's having vanishing means is without loss of generality since such a condition can always be achieved by centering, i.e., subtracting the mean. Note that doing this does not make any difference to the first term of expansion (13) because both Q_i and P_i are probability measures. Taking each $\nu_i = 1$ in (13), one informally obtains the Taylor expansion of ψ around P_i 's

$$\psi(Q_1, \dots, Q_m) - \psi(P_1, \dots, P_m) \approx \sum_{i=1}^m \int g_i(P_1, \dots, P_m; x) d(Q_i - P_i)(x).$$

When each P_i is set to be the true input model F_i and Q_i to be the empirical input model \widehat{F}_i , the above linear expansion is expected to be a reasonably good approximation as the data size grows. The next assumption imposes a moment bound on the error of this approximation:

ASSUMPTION 3 (Smoothness at true input models). *Denote by $g_i(\cdot) := g_i(F_1, \dots, F_m; \cdot)$ the influence functions at the true input distributions $F_i, i = 1, \dots, m$. Assume that the remainder in the Taylor expansion of the performance measure*

$$\psi(\widehat{F}_1, \dots, \widehat{F}_m) = \psi(F_1, \dots, F_m) + \sum_{i=1}^m \int g_i(x) d(\widehat{F}_i - F_i)(x) + \epsilon \quad (14)$$

satisfies $\mathbb{E}[\epsilon^2] = o(n^{-1})$, and the influence functions g_i 's are non-degenerate, i.e. $\sigma_i^2 := \text{Var}_{F_i}[g_i(X_i)] > 0$, and have finite fourth moments, i.e. $\mathbb{E}_{F_i}[g_i^4(X_i)] < \infty$.

Assumption 3 entails that the error of the linear approximation formed by influence functions is negligible in the asymptotic sense. Indeed, the linear term in (14) is asymptotically of order $\Theta_p(n^{-1/2})$ by the central limit theorem, whereas the error ϵ is implied by Assumption 3 to be $o_p(n^{-1/2})$. Hence the variance of the linear term contributes dominantly to the overall input variance as n_i 's are large. Then, thanks to the independence among the input models, the input variance can be expressed in the additive form described in (2) together with a negligible error.

PROPOSITION 1. *Under Assumptions 1-3, the input variance σ_I^2 defined in (1) takes the form*

$$\sigma_I^2 = \sum_{i=1}^m \frac{\sigma_i^2}{n_i} + o\left(\frac{1}{n}\right)$$

where each $\sigma_i^2 = \text{Var}_{F_i}[g_i(X_i)]$ is the variance of the i -th influence function.

As the higher order $o(1/n)$ error suggests, the additive decomposition $\sum_{i=1}^m \frac{\sigma_i^2}{n_i}$ is guaranteed to be accurate only in the large-sample regime. Note that this decomposition is used solely as a theoretical vehicle for asymptotic analysis rather than the actual input variance estimator in our procedure, the latter using bootstrapping schemes that could exhibit better finite-sample performances.

As mentioned before, consistent estimation of input variance σ_I^2 relies on the bootstrap principle, for which we make the following additional assumptions. The assumption states that the error of the linear approximation (14) remains small when the underlying distributions F_i are replaced by the empirical input distributions \widehat{F}_i , hence can be viewed as a bootstrapped version of Assumption 3.

ASSUMPTION 4 (**Smoothness at empirical input models**). *Denote by $\hat{g}_i(\cdot) := g_i(\widehat{F}_1, \dots, \widehat{F}_m; \cdot)$ the influence functions at the empirical input distributions $\widehat{F}_i, i = 1, \dots, m$. Assume the empirical influence function converges to the truth in the sense that $\mathbb{E}[(\hat{g}_i - g_i)^4(X_{i,1})] \rightarrow 0$. For each i let \overline{F}_i be either the i -th empirical input model \widehat{F}_i or the resampled model $\widehat{F}_{s_i,i}^*$. For every $(\overline{F}_1, \dots, \overline{F}_m) \in \prod_{i=1}^m \{\widehat{F}_i, \widehat{F}_{s_i,i}^*\}$, assume the remainder in the Taylor expansion*

$$\psi(\overline{F}_1, \dots, \overline{F}_m) = \psi(\widehat{F}_1, \dots, \widehat{F}_m) + \sum_{i=1}^m \int \hat{g}_i(x) d(\overline{F}_i - \widehat{F}_i)(x) + \epsilon^* \quad (15)$$

satisfies $\mathbb{E}_*[(\epsilon^*)^4] = o_p(s^{-2})$.

As the data sizes n_i 's grow, the empirical input distributions \widehat{F}_i converge to the true ones F_i . Hence the empirical influence functions \hat{g}_i 's are expected to approach the influence functions g_i 's associated with the true input distributions, which explains the convergence condition in Assumption 4. The fourth moment condition on the remainder ϵ^* is needed for controlling the variance of our variance estimator. Since the fourth moment is with respect to the resampling measure and thus depends

on the underlying input data, the condition is described in terms of stochastic order. Note that we require (15) to hold not just when $\bar{F}_i = \hat{F}_{s_i, i}^*$ for all i but also when some $\bar{F}_i = \hat{F}_i$. This allows us to estimate the variance contributed from an arbitrary group of input models and in particular an individual input model.

Assumptions 2-4 are on the performance measure ψ itself. Next we impose assumptions on the simulation noise, i.e. the stochastic error $\hat{\psi}_r - \psi$ where $\hat{\psi}_r$ is an unbiased simulation replication for ψ . We denote by $\tau^2(P_1, \dots, P_m)$ the variance of $\hat{\psi}_r$ when simulation is driven by arbitrary input models P_1, \dots, P_m , i.e.

$$\tau^2(P_1, \dots, P_m) = \mathbb{E}_{P_1, \dots, P_m}[(\hat{\psi}_r - \psi(P_1, \dots, P_m))^2].$$

Similarly we denote by $\mu_4(P_1, \dots, P_m)$ the fourth central moment of $\hat{\psi}_r$ under the input models P_1, \dots, P_m

$$\mu_4(P_1, \dots, P_m) = \mathbb{E}_{P_1, \dots, P_m}[(\hat{\psi}_r - \psi(P_1, \dots, P_m))^4].$$

In particular, for convenience we write $\tau^2 = \tau^2(F_1, \dots, F_m)$ for the variance of $\hat{\psi}$ under the true input models, and $\hat{\tau}^2 = \tau^2(\hat{F}_1, \dots, \hat{F}_m)$ for that under the empirical input models.

The assumptions on the simulation noise are:

ASSUMPTION 5 (Convergence of empirical variance). $\hat{\tau}^2 \xrightarrow{p} \tau^2$.

ASSUMPTION 6 (Convergence of bootstrapped variance). For every $(\bar{F}_1, \dots, \bar{F}_m) \in \prod_{i=1}^m \{\hat{F}_i, \hat{F}_{s_i, i}^*\}$, it holds that $\mathbb{E}_*[(\tau^2(\bar{F}_1, \dots, \bar{F}_m) - \hat{\tau}^2)^2] = o_p(1)$.

ASSUMPTION 7 (Boundedness of the fourth moment). For every $(\bar{F}_1, \dots, \bar{F}_m) \in \prod_{i=1}^m \{\hat{F}_i, \hat{F}_{s_i, i}^*\}$, it holds that $\mathbb{E}_*[\mu_4(\bar{F}_1, \dots, \bar{F}_m)] = O_p(1)$.

Assumptions 5 and 6 stipulate that the variance of the simulation replication $\hat{\psi}_r$ as a functional of the underlying input models is smooth enough in the inputs. Conceptually Assumption 5 is in line with Assumption 3 in the sense that both concern smoothness of a functional around the true input models, whereas Assumption 6 is similar to Assumption 4 since both are about smoothness

property around the empirical input models. Assumption 7 is a fourth moment condition like in Assumption 4 used to control the variance of the variance estimator. Similar to Assumption 4, we impose Assumptions 6 and 7 for each $\bar{F}_i = \hat{F}_i$ or $\hat{F}_{s_i, i}^*$ so that the same guarantees remain valid when estimating input variances from individual input models, i.e., Algorithm 3.

Although the above assumptions may look complicated, they can be verified, under minimal conditions, for generic finite-horizon performance measures in the form

$$\psi(F_1, \dots, F_m) = \mathbb{E}_{F_1, \dots, F_m}[h(\mathbf{X}_1, \dots, \mathbf{X}_m)] \quad (16)$$

where $\mathbf{X}_i = (X_i(1), \dots, X_i(T_i))$ represents the i -th input process consisting of T_i i.i.d. variables distributed under F_i , each T_i being a deterministic time, and h is a performance function. An unbiased simulation replication $\hat{\psi}_r$ of the performance measure is $h(\mathbf{X}_1, \dots, \mathbf{X}_m)$.

Suppose we have the following conditions for the performance function h :

ASSUMPTION 8. *For each i , $0 < \text{Var}_{F_i}[\sum_{t=1}^{T_i} \mathbb{E}_{F_1, \dots, F_m}[h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = X_i]] < \infty$.*

ASSUMPTION 9 (**Parameter k**). *For each i let $I_i = (I_i(1), \dots, I_i(T_i))$ be a sequence of indices such that $1 \leq I_i(t) \leq t$, and $\mathbf{X}_{i, I_i} = (X_i(I_i(1)), \dots, X_i(I_i(T_i)))$. Assume*

$$\max_{I_1, \dots, I_m} \mathbb{E}_{F_1, \dots, F_m}[|h(\mathbf{X}_{1, I_1}, \dots, \mathbf{X}_{m, I_m})|^k] < \infty.$$

The conditional expectation in Assumption 8 is in fact the influence function of the performance measure (16) under the true input models. So Assumption 8 is precisely the non-degenerate variance condition in Assumption 3. All other parts of Assumptions 2-7 are consequences of the moment condition in Assumption 9:

THEOREM 6. *Under Assumptions 1, 8 and Assumption 9 with $k = 4$, we have Assumptions 2-7 hold for the finite-horizon performance measure ψ given by (16).*

4.2. Simulation Complexity and Allocation

This section presents theoretical developments on our proportionate subsampled variance bootstrap. We first establish relative consistency assuming infinite computation resources. Recall (8)

as the proportionate subsampled variance bootstrap estimator without any Monte Carlo errors. The following theorem gives a formal statement on the performance of this estimator discussed in Section 3.1.

THEOREM 7. *Under Assumptions 1-4, if the subsample ratio $\theta = \omega(n^{-1})$, then the proportionate subsampled variance bootstrap without Monte Carlo error, namely (8), is relatively consistent as $n_i \rightarrow \infty$, i.e.*

$$\sigma_{SVB}^2 / \sigma_I^2 \xrightarrow{p} 1.$$

The requirement $\theta = \omega(n^{-1})$ implies that $s_i \rightarrow \infty$, which is natural as one needs minimally an increasing subsample size to ensure the consistency of our estimator. It turns out that this minimal requirement is enough to ensure consistency even relative to the magnitude of σ_I^2 .

Now we turn to the discussion of the Monte Carlo estimate of the bootstrap variance generated from Algorithm 2. The following lemma characterizes the amount of Monte Carlo noise in terms of mean squared error.

LEMMA 1. *The output $\hat{\sigma}_{SVB}^2$ of Algorithm 2 is unbiased for the proportionate subsampled variance bootstrap without Monte Carlo errors, namely σ_{SVB}^2 . Furthermore, under Assumptions 1-7, if*

$$B = \omega(1), \theta = \omega\left(\frac{1}{n}\right) \tag{17}$$

and R is arbitrary, then the conditional mean squared error

$$\mathbb{E}_*[(\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2)^2] = \frac{2}{B} \left(\sum_{i=1}^m \frac{\sigma_i^2}{n_i} + \frac{\tau^2 \theta}{R} \right)^2 (1 + o_p(1)). \tag{18}$$

In addition to the condition $\theta = \omega(n^{-1})$ which has appeared in Theorem 7, we also require $B = \omega(1)$ in Lemma 1. As the proof reveals, with such a choice of B , we can extract the leading term of the conditional mean squared error shown in (18), which takes a neat form and is easy to analyze.

Note that σ_I^2 here is of order n^{-1} by Proposition 1. Hence the Monte Carlo noise of the variance estimate output by our algorithm has to vanish faster than n^{-1} in order to achieve relative consistency. Combining Theorem 7 and Lemma 1, we obtain the simulation complexity of $\hat{\sigma}_{SVB}^2$ in

Theorem 2. To establish the theoretical optimal allocation on the outer and inner sizes B , R , for given data sizes n_i , subsample ratio θ , and total simulation budget N , we minimize the conditional mean square error (18) subject to the budget constraint $BR = N$. This gives rise to the following result that gives a more precise (theoretical) statement than Theorem 4.

THEOREM 8. *Suppose Assumptions 1-7 hold. Given a simulation budget N and a subsample ratio θ such that $N = \omega(\theta n)$ and $\theta = \omega(n^{-1})$, the optimal outer and inner sizes that minimize the conditional mean squared error $\mathbb{E}_*[(\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2)^2]$ are*

$$B^* = \frac{N}{R^*}, \quad R^* = \frac{\theta\tau^2}{\sum_{i=1}^m \sigma_i^2/n_i}$$

which gives a conditional mean squared error

$$\mathbb{E}_*[(\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2)^2] = \frac{8\theta\tau^2}{N} \sum_{i=1}^m \frac{\sigma_i^2}{n_i} (1 + o_p(1)). \quad (19)$$

Theorem 8 gives the exact choices of B and R that minimize the Monte Carlo error. However, this is more of theoretical interest because the optimal R^* involves the desired input variance $\sum_{i=1}^m \sigma_i^2/n_i$. Having said that, we can conclude from the theorem that the optimal inner size R is of order $\Theta(\theta n)$, the same as the subsample size, because the input variance is of order $\Theta(1/n)$ by Proposition 1 and τ^2 is a constant. This results in Theorem 4 in Section 3.2.

4.3. Optimal Subsample Ratio

In this section we further establish the optimal subsample ratio θ or equivalently subsample sizes s_i that balance the two sources of errors in (9). For this, we need more regularity conditions on the performance measure. The first assumption we need is third order Gateaux differentiability in the convex set $\mathcal{P}_1 \times \cdots \times \mathcal{P}_m$:

ASSUMPTION 10 (Third order differentiability). *Using the same notations $P_i, Q_i, P_i^{\nu_i}$ as in Assumption 2, assume that there exist second order influence functions $g_{i_1 i_2}(P_1, \dots, P_m; \cdot) : \Xi_{i_1} \times$*

$\Xi_{i_2} \rightarrow \mathbb{R}$ and third order influence functions $g_{i_1 i_2 i_3}(P_1, \dots, P_m; \cdot) : \Xi_{i_1} \times \Xi_{i_2} \times \Xi_{i_3} \rightarrow \mathbb{R}$ for $i_1, i_2, i_3 = 1, \dots, m$ which are symmetric under permutations, namely

$$g_{i_1 i_2}(P_1, \dots, P_m; x_1, x_2) = g_{i_2 i_1}(P_1, \dots, P_m; x_2, x_1)$$

$$g_{i_1 i_2 i_3}(P_1, \dots, P_m; x_1, x_2, x_3) = g_{i_2 i_1 i_3}(P_1, \dots, P_m; x_2, x_1, x_3) = g_{i_1 i_3 i_2}(P_1, \dots, P_m; x_1, x_3, x_2).$$

and for all x, y satisfy

$$\mathbb{E}_{P_{i_2}}[g_{i_1 i_2}(P_1, \dots, P_m; x, X_{i_2})] = 0, \mathbb{E}_{P_{i_3}}[g_{i_1 i_2 i_3}(P_1, \dots, P_m; x, y, X_{i_3})] = 0.$$

Moreover, as all ν_i 's approach zero the following Taylor expansion holds

$$\begin{aligned} & \psi(P_1^{\nu_1}, \dots, P_m^{\nu_m}) - \psi(P_1, \dots, P_m) \\ &= \sum_{i=1}^m \nu_i \int g_i(P_1, \dots, P_m; x) d(Q_i - P_i)(x) + \frac{1}{2} \sum_{i_1, i_2=1}^m \nu_{i_1} \nu_{i_2} \int g_{i_1 i_2}(P_1, \dots, P_m; x_1, x_2) \prod_{k=1}^2 d(Q_{i_k} - P_{i_k})(x_k) \\ &+ \frac{1}{6} \sum_{i_1, i_2, i_3=1}^m \nu_{i_1} \nu_{i_2} \nu_{i_3} \int g_{i_1 i_2 i_3}(P_1, \dots, P_m; x_1, x_2, x_3) \prod_{k=1}^3 d(Q_{i_k} - P_{i_k})(x_k) + o\left(\left(\sum_{i=1}^m \nu_i^2\right)^{\frac{3}{2}}\right). \end{aligned}$$

Assumption 10 complements and strengthens Assumption 2 in that it imposes stronger differentiability property. Similarly, the following two assumptions strengthen Assumptions 3 and 4 respectively by considering cubic expansions.

ASSUMPTION 11 (Third order smoothness at true input models). Denote by $g_{i_1 i_2}(\cdot) := g_{i_1 i_2}(F_1, \dots, F_m; \cdot)$ and $g_{i_1 i_2 i_3}(\cdot) := g_{i_1 i_2 i_3}(F_1, \dots, F_m; \cdot)$ the second and third order influence functions under the true input models. Assume the remainder in the Taylor expansion of the plug-in estimator $\psi(\hat{F}_1, \dots, \hat{F}_m)$

$$\begin{aligned} \psi(\hat{F}_1, \dots, \hat{F}_m) &= \psi(F_1, \dots, F_m) + \sum_{i=1}^m \int g_i(x) d(\hat{F}_i - F_i)(x) + \frac{1}{2} \sum_{i_1, i_2=1}^m \int g_{i_1 i_2}(x_1, x_2) \prod_{k=1}^2 d(\hat{F}_{i_k} - F_{i_k})(x_k) \\ &+ \frac{1}{6} \sum_{i_1, i_2, i_3=1}^m \int g_{i_1 i_2 i_3}(x_1, x_2, x_3) \prod_{k=1}^3 d(\hat{F}_{i_k} - F_{i_k})(x_k) + \epsilon_3 \end{aligned}$$

satisfies $\mathbb{E}[\epsilon_3^2] = o(n^{-3})$, and the high order influence functions satisfy the moment conditions

$$\mathbb{E}[g_{i_1 i_2}^4(X_{i_1,1}, X_{i_2,2})] < \infty, \mathbb{E}[g_{i_1 i_2 i_3}^2(X_{i_1,1}, X_{i_2,2}, X_{i_3,3})] < \infty$$

for all i_1, i_2, i_3 and $j_2 \leq 2, j_3 \leq 3$, where $X_{i,j}$ is the j -th data point from the i -th input model.

Similar to the remainder ϵ in Assumption 3, the moment condition on ϵ_3 here is used to control the error of the cubic approximation of ψ formed by up to third order influence functions. With these additional assumptions, the error term in Proposition 1 can be refined as follows:

PROPOSITION 2. *Under Assumptions 1, 3 and 10-11, the overall input variance, as defined in (1), can be expressed as*

$$\sigma_I^2 = \sum_{i=1}^m \frac{\sigma_i^2}{n_i} + O\left(\frac{1}{n^2}\right).$$

We also need third order differentiability around the empirical input models:

ASSUMPTION 12 (Third order smoothness at empirical input models). *Denote by*

$\hat{g}_{i_1 i_2}(\cdot) := g_{i_1 i_2}(\hat{F}_1, \dots, \hat{F}_m; \cdot)$ and $\hat{g}_{i_1 i_2 i_3}(\cdot) := g_{i_1 i_2 i_3}(\hat{F}_1, \dots, \hat{F}_m; \cdot)$ the second and third order influence functions under the empirical input models. Assume that the remainder in the Taylor expansion of the bootstrapped performance measure $\psi(\hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*)$

$$\begin{aligned} \psi(\hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*) = & \psi(\hat{F}_1, \dots, \hat{F}_m) + \int \hat{g}_i(x) d(\hat{F}_{s_i,i}^* - \hat{F}_i)(x) + \frac{1}{2} \sum_{i_1, i_2=1}^m \int \hat{g}_{i_1 i_2}(x_1, x_2) \prod_{k=1}^2 d(\hat{F}_{s_{i_k}, i_k}^* - \hat{F}_{i_k})(x_k) \\ & + \frac{1}{6} \sum_{i_1, i_2, i_3=1}^m \int \hat{g}_{i_1 i_2 i_3}(x_1, x_2, x_3) \prod_{k=1}^3 d(\hat{F}_{s_{i_k}, i_k}^* - \hat{F}_{i_k})(x_k) + \epsilon_3^* \end{aligned}$$

satisfies $\mathbb{E}_*[(\epsilon_3^*)^2] = o_p(s^{-3})$. In addition, assume the high order empirical influence functions $\hat{g}_{i_1 i_2}$ and $\hat{g}_{i_1 i_2 i_3}$ converge in mean square error, i.e.

$$\mathbb{E}[(\hat{g}_{i_1 i_2} - g_{i_1 i_2})^2(X_{i_1,1}, X_{i_2,2})] \rightarrow 0, \mathbb{E}[(\hat{g}_{i_1 i_2 i_3} - g_{i_1 i_2 i_3})^2(X_{i_1,1}, X_{i_2,2}, X_{i_3,3})] \rightarrow 0$$

for all i_1, i_2, i_3 and $j_2 \leq 2, j_3 \leq 3$, where $X_{i,j}$ is the j -th data point from the i -th input model. For the first order influence function \hat{g}_i , assume the remainder in the Taylor expansion

$$\hat{g}_i(X_{i,1}) = g_i(X_{i,1}) + \sum_{i'=1}^m \int g_{i i'}(X_{i,1}, x) d(\hat{F}_{i'} - F_{i'})(x) - \int g_i(x) d(\hat{F}_i - F_i)(x) + \epsilon_g$$

satisfies $\mathbb{E}[\epsilon_g^2] = o(n^{-1})$.

As for Assumptions 3 and 4, finite-horizon performance measures under mild conditions satisfy the above two assumptions:

THEOREM 9. *Under Assumptions 1, 8 and Assumption 9 with $k=4$, we have Assumptions 10-12 hold for the finite-horizon performance measure ψ given by (16).*

With Assumptions 11 and 12, we can identify the statistical error of our variance estimator assuming infinite computation resources, which we summarize in the following lemma.

LEMMA 2. *Under Assumptions 1, 3-4 and 10-12, the statistical error of the proportionate subsampled bootstrap variance is characterized by*

$$\sigma_{SVB}^2 - \sigma_I^2 = \mathcal{Z} + \mathcal{R} + o_p\left(\frac{1}{n^{3/2}} + \frac{1}{n_s}\right) \quad (20)$$

where \mathcal{Z} is a random variable such that

$$\mathbb{E}[\mathcal{Z}] = 0, \quad \text{Var}[\mathcal{Z}] = \sum_{i=1}^m \frac{\lambda_i^T \Sigma_i \lambda_i}{n_i}$$

with $\lambda_i = (1/n_i, 2/n_1, \dots, 2/n_m)^T$ and

$$\Sigma_i = \text{covariance matrix of } (g_i^2(X_i), \mathbb{E}_{X'_1}[g_1(X'_1)g_{1i}(X'_1, X_i)], \dots, \mathbb{E}_{X'_m}[g_m(X'_m)g_{mi}(X'_m, X_i)]).$$

\mathcal{R} is defined as

$$\begin{aligned} \mathcal{R} = & \sum_{i=1}^m \frac{1}{n_i s_i} \text{Cov}(g_i(X_i), g_{ii}(X_i, X_i)) + \sum_{i,i'=1}^m \frac{1}{n_i s_{i'}} \text{Cov}(g_i(X_i), \mathbb{E}_{X'_{i'}}[g_{ii'}(X_i, X'_{i'}, X'_{i'})]) \\ & + \sum_{i=1}^m \frac{\text{frac}(\theta n_i) \sigma_i^2}{n_i s_i} + \sum_{i,i'=1}^m \frac{\text{Var}[g_{ii'}(X_i, X'_{i'})]}{4 n_i s_{i'}} \end{aligned}$$

where $\text{frac}(x) := x - \lfloor x \rfloor$ denotes the fraction part of $x \in \mathbb{R}$, and for each i , X_i, X'_i are independent copies of the random variable distributed under F_i .

Combining the statistical error (20), and the minimal Monte Carlo error (19) under the optimal budget allocation into the trade-off (9), we obtain the overall error of the output $\hat{\sigma}_{SVB}^2$ of Algorithm 2:

THEOREM 10 (Overall error of the variance estimate). *Suppose Assumptions 1, 3-7 and 10-12 hold. Given a simulation budget N and a subsample ratio θ such that $N = \omega(\theta n)$ and $\theta = \omega(n^{-1})$,*

if outer and inner sizes B, R for Algorithm 2 are chosen to be $R = \Theta(\theta n), B = N/R$, then the gross error of our Monte Carlo estimate $\hat{\sigma}_{SVB}^2 - \sigma_I^2 = \mathcal{E} + o_p(\theta^{1/2}(Nn)^{-1/2} + \theta^{-1}n^{-2} + n^{-3/2})$, where the leading term has a mean squared error

$$\mathbb{E}[\mathcal{E}^2] = \Theta\left(\frac{\theta}{Nn} + \mathcal{R}^2 + \sum_{i=1}^m \frac{\lambda_i^T \Sigma_i \lambda_i}{n_i}\right) \quad (21)$$

where \mathcal{R}, λ_i 's and Σ_i 's are defined in Lemma 2.

It is clear from their definitions in Lemma 2 that $\mathcal{R} = O(\theta^{-1}n^{-2})$ and each $(\lambda_i^T \Sigma_i \lambda_i)/n_i = O(n^{-3})$, hence the mean squared error (21) is in general of order $O(\theta(Nn)^{-1} + \theta^{-2}n^{-4} + n^{-3})$. When \mathcal{R} and at least one of the $\lambda_i^T \Sigma_i \lambda_i$'s satisfy the non-degeneracy condition in Theorem 5, this bound becomes tight in order, and the optimal subsample ratio can be established by minimizing the order of the leading overall error \mathcal{E} .

5. Numerical Experiments

This section reports our experimental findings. We consider two examples with different scales and complexities:

M/M/1 queue: The first example we consider is an M/M/1 queue that has true arrival rate 0.5 and service rate 1. Suppose the system is empty at time zero. The performance measure of interest is the probability that the waiting time of the 20-th arrival exceeds 2 units of time, whose true value is approximately 0.182. Specifically, the system has two input distributions, i.e., the inter-arrival time distribution $F_1 = \text{Exp}(0.5)$ and the service time distribution $F_2 = \text{Exp}(1)$, for which we have n_1 and n_2 i.i.d. data available respectively. If A_t is the inter-arrival time between the t -th and $(t+1)$ -th arrivals, and S_t is the service time for the t -th arrival, then the system output

$$\psi(F_1, F_2) = \mathbb{E}_{F_1, F_2}[\mathbf{1}\{W_{20} > 2\}]$$

where the waiting time W_{20} is calculated by the Lindley recursion $W_{t+1} = \max\{W_t + S_t - A_t, 0\}$ for $t = 1, \dots, 19$ and $W_1 = 0$. To test the proposed approach under different levels of utilization, we also consider true arrival rate 0.9 and service rate 1, for which case the target performance measure

is taken to be the probability that the waiting time of the 20-th arrival exceeds 6 units of time (true value 0.190). The data sizes n_1, n_2 are chosen so that $n_1 = 2n_2$ in the experiments, so only the minimum $\min_i n_i$ is reported for convenience.

Computer network: We also consider a computer communication network borrowed from Cheng and Holland (1997) and Lin et al. (2015). The structure of the system is characterized by the undirected graph in Figure 1: Four message-processing units, which correspond to the nodes, are connected by four transport channels that are represented by the edges. For every pair i, j

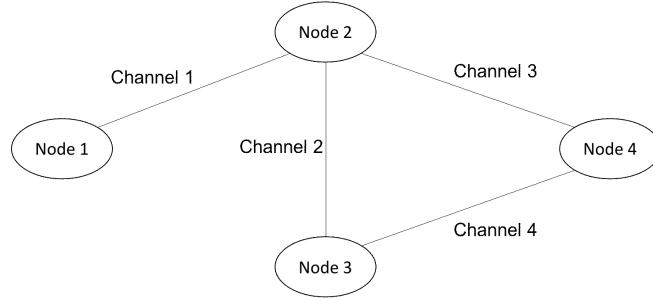


Figure 1 A computer network with four nodes and four channels.

of processing units with $i \neq j$, there are external messages that enter into unit i and are to be transmitted to unit j through a fixed path, and their arrival follows a Poisson process with rate $\lambda_{i,j}$. The specific values for $\lambda_{i,j}$'s are summarized in Table 1. Each unit takes a constant time of

node j node i	1	2	3	4
1	n.a.	40	30	35
2	50	n.a.	45	15
3	60	15	n.a.	20
4	25	30	40	n.a.

Table 1 True arrival rates $\lambda_{i,j}$ of messages to be transmitted from node i to node j .

0.001 seconds to process a message, and has unlimited storage capacity. The messages have lengths that are independent and follow an exponential distribution with mean 300 bits, and each channel

has a capacity of 275000 bits, therefore there are queuing and transmission delays. The messages travel through the channels with a velocity of 150000 miles per second, and the i -th channel has a length of $100 \cdot i$ miles for $i = 1, 2, 3, 4$, leading to a propagation delay of $\frac{100 \cdot i}{150000}$ seconds along the i -th channel. The total time that a message of length l bits occupies the i -th channel is therefore $\frac{l}{275000} + \frac{100 \cdot i}{150000}$ seconds. Suppose the system is empty at time zero. The performance measure of interest is the average delay of the first 30 messages that arrive to the system, or mathematically, $\mathbb{E}[\frac{1}{30} \sum_{k=1}^{30} D_k]$, where D_k is the time for the k -th message to be transmitted from its entering node to destination node. The true value of the performance measure is approximately 6.91×10^{-3} seconds. In the experiment, we assume that the arrival rates of the different types of messages, as well as the distribution of the message length, are unknown, therefore there are 13 input models in total. Like in the example of M/M/1 queue, the data sizes across different input models are kept proportional to each other and only the minimum size is reported.

In the experiments we investigate the simulation efforts needed for our subsampling procedure to generate accurate estimates of the input variance, the impacts of the procedural parameters θ, B, R on the estimation accuracy, and practical guidelines on optimal choices of these parameters. Regarding performance metrics of the method, we primarily focus on the mean squared error of the obtained input variance estimate. In addition, note that our estimated input variance can also be used to construct CIs by plugging into formula (3). We also examine the quality of these CIs, measured by coverage accuracy and width, as impacted by the estimation accuracy of the input variance.

We compare our subsampling approach with the variance bootstrap depicted in Algorithm 1 and the percentile bootstrap suggested by Barton and Schruben (1993, 2001). The percentile bootstrap adopts the same nested simulation structure as in variance bootstrap, but does not estimate the input variance and instead directly outputs order statistics of the resampled performance measures to construct CIs. Specifically, after obtaining B bootstrapped performance measure estimates $\bar{\psi}^b := \frac{1}{R} \sum_{r=1}^R \hat{\psi}_r(\hat{F}_1^b, \dots, \hat{F}_m^b)$, each averaged over R i.i.d. replications, the percentile bootstrap outputs the $\frac{\alpha}{2}(B+1)$ -th and $(1 - \frac{\alpha}{2})(B+1)$ -th order statistics of $\{\bar{\psi}^b : b = 1, \dots, B\}$ as a $(1 - \alpha)$ -level CI.

In converting our subsampled input variance estimate to CI, we also investigate the use of a “splitting” versus a “non-splitting” approach. In most part of this section, we use the splitting approach that divides the budget into two portions with one used to estimate the input variance and the other to compute the point estimator. To describe it in detail, suppose we have a total budget of N simulation runs. We allocate R_v simulation runs to estimate σ_I^2 using either Algorithm 1 or 2, and the remaining $R_e = N - R_v$ simulation runs driven by the empirical input distributions to compute the point estimator $\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)$. When constructing the CI in (3), the simulation variance σ_S^2 is calculated as $\frac{\tilde{\tau}^2}{R_e}$, where $\tilde{\tau}^2$ is the sample variance computed from the R_e simulation replications. The second, “non-splitting”, approach invests all the N simulation runs in estimating σ_I^2 , and constructs the point estimator by averaging all the replications, i.e., $\bar{\psi} = \frac{1}{B} \sum_{b=1}^B \bar{\psi}^b$, where $\bar{\psi}^b$ is the performance measure estimate for the b -th resample from Algorithm 2. The simulation variance σ_S^2 in this case is taken to be the sample variance of all the $\bar{\psi}^b$'s divided by the bootstrap size B . The rationale for this approach is that, when the subsample size θn is large, $\mathbb{E}_*[\bar{\psi}]$ should accurately approximate the plug-in estimator $\psi(\hat{F}_1, \dots, \hat{F}_m)$ with an error that is negligible relative to the input variability. Using the former as a surrogate for the latter avoids splitting the budget; however, we will see later that this may introduce too much bias to maintain the desired coverage level when the subsample size is relatively small.

The rest of this section is organized as follows. Section 5.1 investigates practical guidelines for choosing the algorithmic parameters in our procedure. Using these guidelines, in Section 5.2 we compare the proposed procedure with the variance bootstrap and the percentile bootstrap. Section 5.3 studies further the conversion of input variance estimate into CI, and compares the associated splitting and non-splitting approaches.

5.1. Guidelines for Algorithmic Configuration

We examine the performances using a wide range of parameter choices for θ, B, R . For each of the two considered examples, and input data sizes from 30 to 2000, we test our subsampling approach at various combinations of θ, B, R where the subsample size $\theta \min_i n_i \in \{5, 15, 30, 60, 120\}$

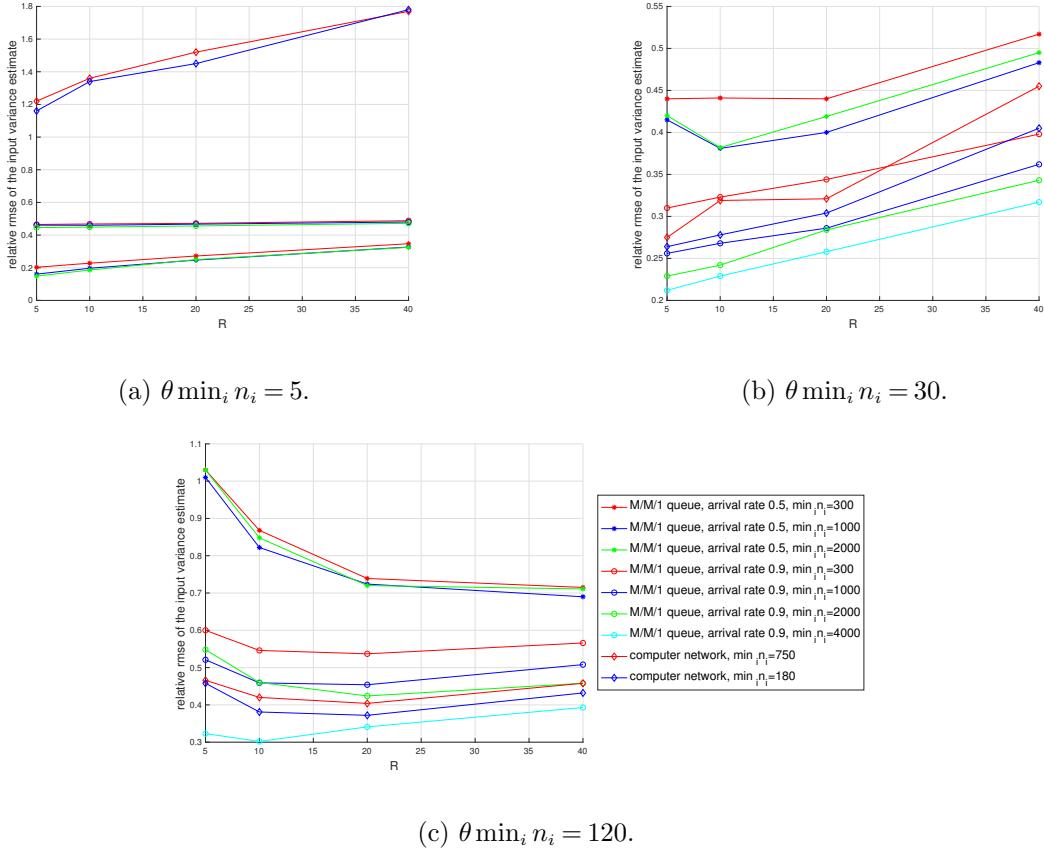


Figure 2 Input variance estimation accuracy under different configurations of B, R such that $BR = 1000$.

and the budget allocation parameters $(B, R) \in \{(25, 40), (50, 20), (100, 10), (200, 5)\}$ (a total of 1000 simulation runs). To calculate the mean square error of the input variance estimate, we perform 1000 independent runs of the procedure, each on an independently generated input data set, and then take the average of the squared errors. The reported error metric is the relative root mean squared error (rmse) which can be expressed as $\frac{\sqrt{\mathbb{E}[(\hat{\sigma}_I^2 - \sigma_I^2)^2]}}{\sigma_I^2}$ where $\hat{\sigma}_I^2$ and σ_I^2 are the estimated and true input variances respectively.

We first study and establish guidelines for the outer size B and inner size R for a given subsample size. Figure 2 shows how the estimation error changes as the inner replication size R grows from 5 to 40 (correspondingly the outer size B drops from 200 to 25) and the subsample size $\theta \min_i n_i$ is fixed at a certain value. Each curve represents the results for one of the considered examples under a particular input data size. Although the precise optimal choice for B, R varies from one example to another even when the subsample size is chosen the same, the estimation error appears robust

to the parameter choices, with a range of values that only slightly underperform the optimal. In particular, compared to the unknown optimal choice, an R between $\frac{1}{6}\theta \min_i n_i$ and $\frac{1}{3}\theta \min_i n_i$ seems to achieve a comparable accuracy level in the variance estimation, hence is recommended as a general choice.

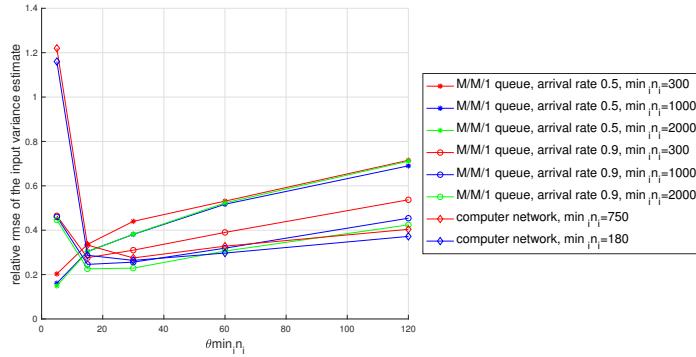


Figure 3 Input variance estimation accuracy under different subsample sizes with B, R optimally tuned.

Now we turn to optimal choices for the subsample size. Provided that B, R is properly chosen as above, we examine the behavior of the variance estimation error as the subsample size varies. As we have discussed in Section 3.1, subsampling is preferred when the input data size is relatively large, and thus we consider input data sizes ≥ 500 for our M/M/1 queue and computer network, and for each considered data size we plot the variance estimation error versus the subsample size in Figure 3. We see that a too large size such as 120 always leads to a larger estimation error than moderate sizes like 30, whereas a too small size around 5 can lift the error by even more in some cases, which is consistent with the theoretical insight from the bound (21). Therefore, in general we recommend the use of a subsample size $\theta \min_i n_i$ between 20 and 40 to optimize the estimation accuracy. Figure 3 shows that, under the suggested subsample size, the relative rmse is as low as 0.2-0.5 across all the cases.

5.2. Comparisons with the Variance Bootstrap and the Percentile Bootstrap

We compare our subsampling method with the standard variance bootstrap and the percentile bootstrap, under the same total budget of 1500 simulation runs. In addition to the relative rmse

of the input variance estimate, we also report the actual coverage probability and width of the CI constructed by plugging in the input variance estimate. To estimate all these performance metrics, we construct 1000 95%-level CIs for the target performance measures, each from an independently generated input data set. The “splitting” approach that splits the total budget into $R_v = 1000$, $R_e = 500$ is adopted for the subsampling approach and the variance bootstrap, whereas for the percentile bootstrap all the 1500 simulation runs are used for the resamples. As suggested in Section 5.1, we use the parameter values $\theta = \frac{30}{\min_i n_i}$, $B = 100$, $R = 10$ in our method in all the cases, whereas for the other two methods we vary the parameter configurations over a reasonable range constrained by the simulation budget and then report the best results generated by these considered configurations. In particular, the parameters for the variance bootstrap are chosen to minimize the mean square error of the input variance estimate from four combinations, “ $B = 25, R = 40$ ”, “ $B = 50, R = 20$ ”, “ $B = 100, R = 10$ ”, “ $B = 200, R = 5$ ”, and those for the percentile bootstrap are chosen to achieve the best the coverage accuracy from four combinations, “ $B = 50, R = 30$ ”, “ $B = 100, R = 15$ ”, “ $B = 300, R = 5$ ”, “ $B = 1500, R = 1$ ”. Note that these give an upper hand to our competing alternatives in the comparisons.

Tables 2 and 3 summarize the experimental results for the M/M/1 queue when the true arrival rate is 0.5 and 0.9 respectively, and Table 4 shows those for the computer network. The shorthand “PSVB” stands for proportionate subsampled variance bootstrap, i.e., our subsampling approach. For each method, the “coverage estimate” column displays estimates of the actual coverage probability based on 1000 independent CIs, and the “CI width” column shows their average width. The second column of each table shows the ratio between the input standard error σ_I and the simulation standard error σ_S for different input data sizes in our “splitting” approach. A ratio close to or greater than 1 means that the input noise is a major source of uncertainty relative to the simulation noise, thus indicating the need to be taken into account in output analysis.

We compare the approaches based on Tables 2-4. Firstly, our subsampling approach significantly outperforms the variance bootstrap in terms of estimation accuracy of the input variance. The

$\min_i n_i$	$\frac{\sigma_I}{\sigma_S}$	PSVB			variance bootstrap			percentile bootstrap	
		relative rmse	coverage estimate	CI width	relative rmse	coverage estimate	CI width	coverage estimate	CI width
30	7.74	0.73	84.3%	0.422	0.73	84.3%	0.422	91.9%	0.467
100	3.77	0.55	92.5%	0.251	0.80	88.6%	0.248	98.8%	0.356
300	2.13	0.44	94.8%	0.156	1.04	85.6%	0.148	99.9%	0.307
1000	1.15	0.38	95.0%	0.103	2.48	89.4%	0.111	100%	0.285
2000	0.79	0.38	95.9%	0.087	5.43	92.8%	0.107	100%	0.280

Table 2 Results for the M/M/1 queue with arrival rate 0.5 and service rate 1.

$\min_i n_i$	$\frac{\sigma_I}{\sigma_S}$	PSVB			variance bootstrap			percentile bootstrap	
		relative rmse	coverage estimate	CI width	relative rmse	coverage estimate	CI width	coverage estimate	CI width
30	11.12	0.59	81.4%	0.609	0.59	81.4%	0.609	94.6%	0.639
100	6.22	0.42	89.9%	0.372	0.63	88.6%	0.386	97.2%	0.446
300	3.46	0.32	92.6%	0.225	0.71	87.0%	0.225	99.3%	0.348
1000	1.86	0.27	93.3%	0.137	1.21	86.3%	0.137	100%	0.307
2000	1.30	0.24	95.0%	0.108	2.19	90.7%	0.119	100%	0.294
4000	0.91	0.23	94.9%	0.089	3.61	91.2%	0.106	100%	0.288

Table 3 Results for the M/M/1 queue with arrival rate 0.9 and service rate 1.

estimates generated by our approach have a smaller relative error than those by the variance bootstrap in all considered cases, and the gap becomes more significant as the data size grows larger. In particular, as the data size grows from 30 to thousands, the estimation error keeps decreasing from 0.7 to 0.25 in our approach, whereas in variance bootstrap it keeps increasing from 0.7 to larger than 1, a level that makes the estimate too crude to be useful. These demonstrate the computational advantage and dictate the use of subsampling especially when the input data size is relatively large. Note that the same budget of 1000 simulation runs are used in input variance

$\min_i n_i$	$\frac{\sigma_L}{\sigma_S}$	PSVB			variance bootstrap			percentile bootstrap	
		relative	coverage	CI width	relative	coverage	CI width	coverage	CI width
		rmse	estimate	($\times 10^{-4}$)	rmse	estimate	($\times 10^{-4}$)	estimate	($\times 10^{-4}$)
30	12.60	0.74	92.0%	19.3	0.74	92.0%	19.3	95.2%	22.0
150	5.36	0.41	94.3%	8.85	0.53	91.3%	8.50	98.3%	11.2
750	2.35	0.32	94.2%	4.27	0.94	86.9%	3.88	100%	7.97
1800	1.53	0.28	95.3%	3.03	1.63	87.1%	3.01	100%	7.34

Table 4 Results for the computer network.

estimation for all considered data sizes and that the estimation accuracy seems much better for large data sizes than for small sizes, and one may wonder whether more simulation runs should be used for small data sizes to further improve the estimation accuracy. It turns out that the estimation errors are mostly due to the inadequacy of the input data rather than the simulation budget, hence a budget of 1000 is already large enough and further increasing the budget does not bring much benefit. For instance, in the case of data size 30 in Table 2, the relative error of the input variance estimate remains as large as 0.69 even if the simulation budget is increased by 10 times.

Secondly, thanks to the high accuracy in the input variance estimates, our subsampling approach generates accurate CIs whose coverage probabilities quickly approach the nominal level 95% as the input data size grows. In contrast, the CIs using the variance bootstrap exhibit under-coverage, and the percentile bootstrap CIs significantly over-cover the truth. We see that the coverage of the variance bootstrap is below 90% in most considered cases, and in the very few cases where the CIs happen to have relatively good coverages, the intervals are much wider than those by our subsampling approach. For example, in the case of data size 2000 in Table 2, the variance bootstrap gives a fairly accurate coverage 92.8%, but on average the interval is 1.23 times as wide as that by our method. This shows that the better estimates of the input variance using subsampling translate to better CIs significantly compared to using the variance bootstrap, in terms of both coverage

accuracy and width. The percentile bootstrap CIs show an overly high coverage probability close to 100% and are 2-3 times wider than those by subsampling for all considered input data sizes except 30. The over-coverage issue of the percentile CIs arises because the order statistics capture only the input noise but not the simulation noise in the resampled performance measures, a phenomenon that has been discussed in Barton et al. (2007, 2018). When one can afford a sufficiently large budget of simulation relative to the input data size, the simulation noise can be made negligible so that the CIs have the correct coverage. However, when simulation resources are relatively limited (e.g., when data size ≥ 100 in Tables 2-4), the CIs are unnecessarily widened by the extra simulation noise that leads to over-coverage. We also notice that the percentile bootstrap CIs do show more accurate coverage than the other two methods when the input data size is 30, which may suggest that the percentile bootstrap is the preferred approach to constructing CIs in small data cases. However, this outperformance is a result of optimally choosing the parameters B, R in hindsight. In our experiments, this best parameter set varies from one case to another, and the actual coverage under different configurations varies in a range of 8%.

Thirdly, results across different input data sizes show that, the advantages of subsampling in both input variance estimation and CI construction are most significant in situations with relatively large input data size. Note that one may argue in such situations input uncertainty is negligible. However, whether this is indeed the case relates to the error tolerance of the decision-maker and the magnitude of the target performance measure itself. For the large data sizes we consider, the input noise appears still relatively substantial. For instance, when the input data size is 2000 in Table 3, the average width of the CIs as a measure of the input uncertainty and simulation uncertainty combined amounts to as much as 57% of the target tail probability, and that the input uncertainty serves as a major component of the total uncertainty (a ratio of 1.3 relative to the simulation uncertainty).

Lastly, in situations with small input data size like 30 the CI coverage clearly falls below 95% in Tables 2 and 3. This under-coverage phenomenon may appear to stem from the nonlinear effect of

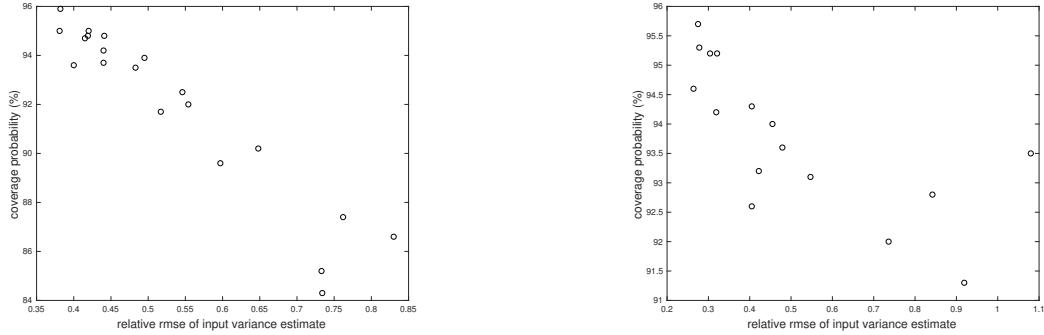
the performance measure that is inadequately captured by the Gaussian-approximation-based CI given in (3). The real reason, as our experiments suggest, turns out to be the insufficient accuracy of the input variance estimates. In fact, if the true input variance σ_I^2 (which can be accurately estimated by repeatedly generating independent input data sets) is plugged into (3) to construct CIs, the coverage probability under the data size 30 rises to 94%-95% for both the M/M/1 queue and the computer network. This indicates a positive impact of an accurate input variance estimate on the CI quality, a point that we will discuss further momentarily.

5.3. Constructing CI via Input Variance and Comparisons of the Splitting and Non-Splitting Approaches

We study in more depth the relation between the input variance estimation accuracy and CI quality, and compare the splitting approach for CI construction that has been used in previous subsections, with the alternate non-splitting approach described at the beginning of this section. Finally, we provide practical budget allocation strategies for the splitting approach.

First, to see how the estimation accuracy of the input variance affects the coverage accuracy of the CIs, we use the splitting approach to compute 95%-level CIs, with 1000 simulation runs assigned to input variance estimation and another 500 runs to point estimator evaluation. Figure 4a plots the coverage probability versus the relative rmse when the subsample size $\theta \min_i n_i$ is chosen 30 in the M/M/1 queue example, where each point corresponds to a particular combination of the data size $\min_i n_i$, the outer replication size B , and the inner replication size R . Figure 4b plots the same for the computer network example with subsample size 30. Both figures clearly show that, the more accurately the input variance is estimated, the closer to the nominal level 95% the coverage probability will be. Accurate estimation of the input variance thus appears to play a crucial role in the construction of accurate CIs.

Next we compare the splitting and non-splitting approaches under the same total budget of 1500 simulation runs. Like in the splitting approach, we use a subsample size $\theta \min_i n_i = 30$ for our non-splitting approach, but use $B = 75, B = 20$ to consume all the 1500 simulation runs. We



(a) M/M/1 queue with arrival rate 0.5, $\theta \min_i n_i = 30$. (b) Computer network, $\theta \min_i n_i = 30$.

Figure 4 Monotonicity between coverage accuracy and input variance estimation accuracy.

find that the CIs generated from the two approaches have similar lengths, but the non-splitting approach underperforms in terms of coverage accuracy. Each plot in Figure 5 shows the coverage probabilities of the non-splitting CIs versus the splitting ones for each of the considered example systems, as the input data size grows from 30 to thousands. We see that when the data size is relatively small (e.g., below 500), the two approaches generate CIs with similar coverage accuracy. When the data size grows larger, however, the coverage probability of the non-splitting CIs keeps dropping in all the three examples, especially in the M/M/1 queue with arrival rate 0.9 where a drop towards 86% is observed, whereas the splitting CIs exhibit almost exact 95% coverage. A possible cause of the undercoverage is the overly small subsample size compared to the input data size, which leads to a high bias in the point estimator. With a subsample size s , the bias of the non-splitting point estimator $\mathbb{E}_*[\bar{\psi}]$ with respect to the truth $\psi(F_1, \dots, F_m)$ can be as large as $O(1/s)$. Given that the input standard error is $\Theta(1/\sqrt{n})$, $\mathbb{E}_*[\bar{\psi}]$ has a negligible bias only when the subsample size is large enough, namely when $s = \omega(\sqrt{n})$, indicating that a small subsample size relative to the data size can corrupt the CI. In our experiment, we find that the (supposedly unobservable) bias can be as large as 25% of the CI width when the input data size is 2000 in the M/M/1 queue with arrival rate 0.9, and that artificially removing the bias from the point estimator can improve the coverage to a similar level achieved by the splitting approach. Because of the bias and the consequent under-coverage issue, we caution the use of the non-splitting approach, that it should only be used when a relatively large subsample size is adopted.

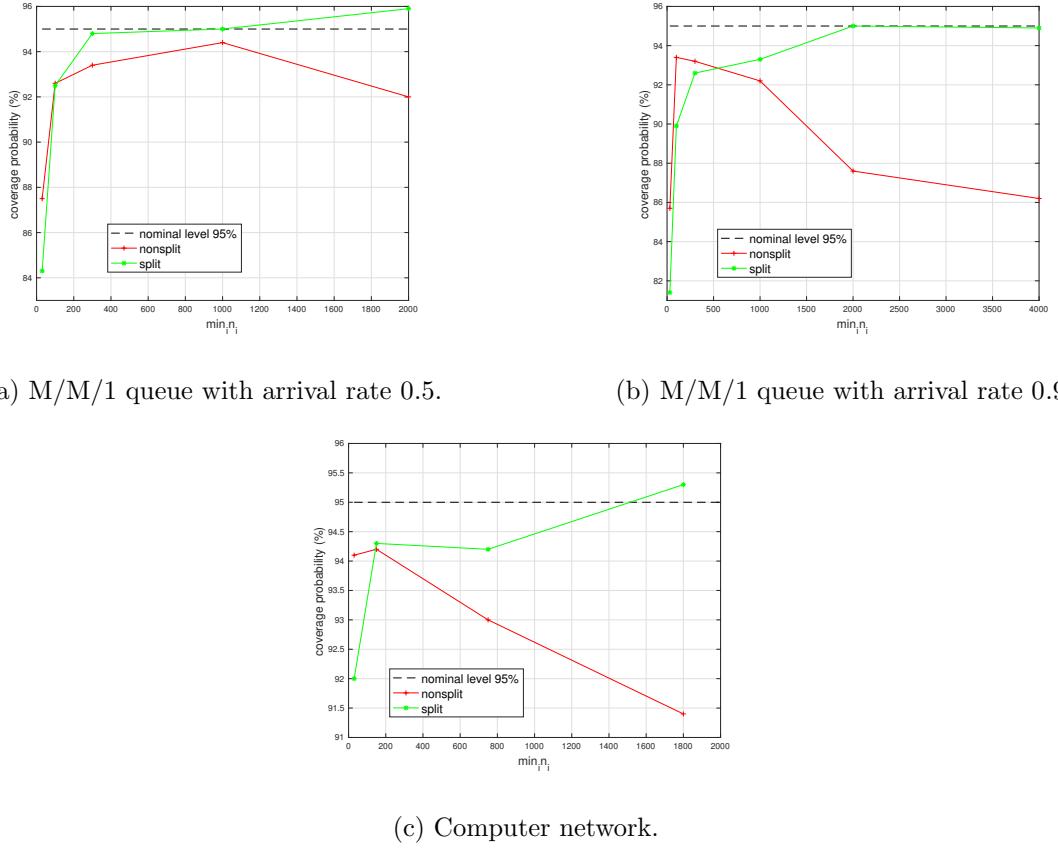


Figure 5 Coverage comparison under the splitting and non-splitting approaches.

Since the splitting approach is recommended, next we explore strategies of splitting a given budget. Our goal is to generate shortest possible CIs that have a sufficiently accurate coverage probability. As in the beginning of the section, denote by R_v the number of simulation runs used to estimate the input variance, and by R_e to construct the point estimator. Under a fixed total budget $R_v + R_e = 1500$, we try four different splits $R_v = 100, 250, 500, 1000$ (accordingly $R_e = 1500 - R_v$), and for each split the subsample size is fixed at $\theta \min_i n_i = 30$ and several choices of B, R are tested among which the one with the best coverage probability is reported. Figure 6 plots the coverage probability versus the CI width for the four considered splits, where the M/M/1 queue with arrival rate 0.9 is considered and input data size is 2000. We notice that the split controls a tradeoff between the coverage accuracy and the CI width. The more simulation runs one allocates to input variance estimation, the more accurate but wider CIs one would obtain, because the input variance is more accurately estimated while the point estimator becomes more noisy. The plot

suggests that allocating 500-1000 replications to variance estimation achieves a good balance of accuracy and width, in the sense that the intervals from the split “500+1000” or “1000 + 500” are only slightly wider than those by other splits and that allocating less (say 250) to variance estimation results in a considerable drop in coverage probability from the nominal level 95%. The results from Tables 2-4, where the split “1000 + 500” is used, also validates the effectiveness of such a strategy. Therefore, for a given simulation budget, we recommend that the user allocate 500-1000 replications to input variance estimation with our subsampling approach and all the remaining budget to the construction of the point estimator.

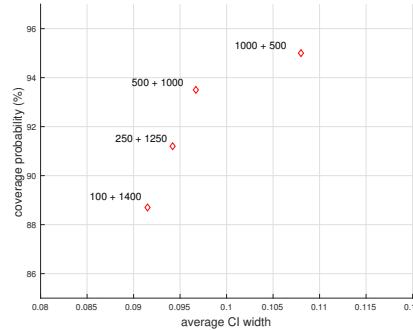


Figure 6 Coverage probability versus CI width, under different budget splits in the form of “ $R_v + R_e$ ”.

Lastly, to validate the various guidelines proposed in this section regarding the choices of the subsample size $\theta \min_i n_i$, outer size B and inner size R , as well as budget allocation strategies for the splitting approach to CI construction, we test their effectiveness and robustness under different configurations of the computer network. Specifically, under a fixed total simulation budget of 1500 runs, we vary the channel capacity, the transmission speed of the channels, and the arrival rates of messages for the computer network (see Appendix EC.5 for these configuration details), otherwise keeping the same setting as stated at the beginning of this section, and apply PSVB with budget split “1000 + 500”, subsample size $\theta \min_i n_i = 30$, and $B = 100, R = 10$ to compute input variance estimates and CIs. The standard variance bootstrap is also tested as a benchmark, with the B and R chosen in hindsight from four candidate combinations, “ $B = 25, R = 40$ ”, “ $B =$

$50, R = 20$ ", " $B = 100, R = 10$ ", and " $B = 200, R = 5$ ", to minimize the mean squared error of the input variance estimate. Plots of relative rmse, CI coverage estimate, and CI width against the

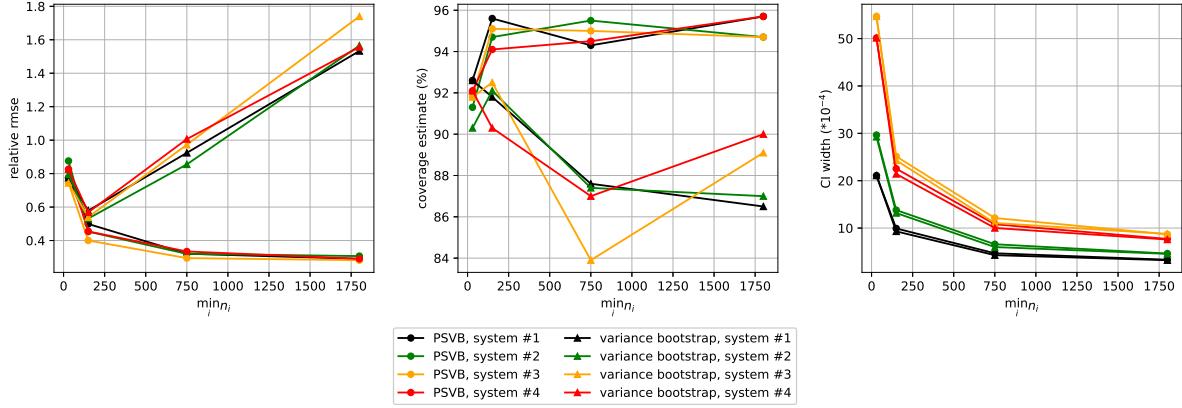


Figure 7 Comparison of PSVB and variance bootstrap under various configurations of the computer network.

input data size $\min_i n_i$ are shown in Figure 7, where each line corresponds to either PSVB or the variance bootstrap applied to one of the four differently configured computer networks. The phenomena that we have observed in Tables 2-4 still persist for all the four computer networks. As the input data size grows, the estimation accuracy of the input variance improves to a level of 0.3 in relative rmse for PSVB, but deteriorates significantly for the variance bootstrap. Accordingly the CI coverage of PSVB stays within a 1% margin around the nominal level 95%, whereas the variance bootstrap shows a significantly lower coverage than the nominal level because of inaccurate input variance estimates. All these demonstrate that the proposed guidelines for using PSVB deliver superior and robust performance across different systems, and therefore can be used as a default algorithmic configuration in practice. However, the relatively low accuracy of the variance estimates and CIs appears again in our approach when the input data size is limited (around 30). The same limitation arises for the variance bootstrap. This reconciles with our observation from Section 5.2 that subsampling is most beneficial for cases with moderately large input data sizes where other approaches like the variance bootstrap start to become computationally demanding.

6. Conclusion

We have explained how estimating input variances in stochastic simulation can require large computation effort when using conventional bootstrapping. This arises as the bootstrap involves a two-layer sampling, which adds up to a total effort of larger order than the data size in order to achieve relative consistency. To alleviate this issue, we have proposed a subsampling method that leverages the relation between the structure of input variance and the estimation error from the two-layer sampling, so that the resulting total effort can be reduced to being independent of the data size. We have presented the theoretical results in this effort reduction, and the optimal choices of the subsample ratio and simulation budget allocation in terms of the data size and the budget. We have also demonstrated numerical results to support our theoretical findings, and provided guidelines in using our proposed methods to estimate input variances and also construct output CIs. Future work comprises a more comprehensive investigation of our subsampling scheme, including its generalization to input processes with serial dependence and potentially non-smooth performance measures such as quantiles and other risk measures.

Acknowledgments

We gratefully acknowledge support from the National Science Foundation under grants CMMI-1542020, CMMI-1523453 and CAREER CMMI-1653339/1834710. A preliminary conference version of this work, Lam and Qian (2018), has appeared in the Proceedings of the Winter Simulation Conference 2018.

References

Abadie A, Imbens GW (2008) On the failure of the bootstrap for matching estimators. *Econometrica* 76(6):1537–1557.

Andrews DW, Guggenberger P (2009) Validity of subsampling and “plug-in asymptotic” inference for parameters defined by moment inequalities. *Econometric Theory* 25(3):669–709.

Andrews DW, Guggenberger P (2010) Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory* 26(2):426–468.

Asmussen S, Glynn PW (2007) *Stochastic Simulation: Algorithms and Analysis*, volume 57 (Springer Science & Business Media).

Barton RR (2012) Tutorial: Input uncertainty in output analysis. Laroque C, Himmelsbach J, Pasupathy R, Rose O, Uhrmacher A, eds., *Proceedings of the 2012 Winter Simulation Conference*, 1–12 (Piscataway, New Jersey: IEEE).

Barton RR, Chick SE, Cheng RC, Henderson SG, Law AM, Schmeiser BW, Leemis LM, Schruben LW, Wilson JR (2002) Panel discussion on current issues in input modeling. Yücesan E, Chen CH, Snowdon JL, Charnes JM, eds., *Proceedings of the 2002 Winter Simulation Conference*, 353–369 (Piscataway, New Jersey: IEEE).

Barton RR, Lam H, Song E (2018) Revisiting direct bootstrap resampling for input model uncertainty. *2018 Winter Simulation Conference (WSC)*, 1635–1645 (IEEE).

Barton RR, Nelson BL, Xie W (2013) Quantifying input uncertainty via simulation confidence intervals. *INFORMS Journal on Computing* 26(1):74–87.

Barton RR, Schruben LW (1993) Uniform and bootstrap resampling of empirical distributions. Evans GW, Mollaghasemi M, Russell E, Biles W, eds., *Proceedings of the 1993 Winter Simulation Conference*, 503–508 (ACM).

Barton RR, Schruben LW (2001) Resampling methods for input modeling. Peters BA, Smith JS, Medeiros DJ, Rohrer MW, eds., *Proceedings of the 2001 Winter Simulation Conference*, volume 1, 372–378 (Piscataway, New Jersey: IEEE).

Barton RR, et al. (2007) Presenting a more complete characterization of uncertainty: Can it be done. *Proceedings of the 2007 INFORMS simulation society research workshop*, 26–60 (INFORMS Simulation Society).

Bickel PJ, Gtze F, van Zwet WR (1997) Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica* 7(1):1–31.

Bickel PJ, Sakov A (2008) On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica* 18(3):967–985.

Biller B, Corlu CG (2011) Accounting for parameter uncertainty in large-scale stochastic simulations with correlated inputs. *Operations Research* 59(3):661–673.

Cheng RC, Holland W (1997) Sensitivity of computer simulation experiments to errors in input data. *Journal of Statistical Computation and Simulation* 57(1-4):219–241.

Cheng RC, Holland W (1998) Two-point methods for assessing variability in simulation output. *Journal of Statistical Computation Simulation* 60(3):183–205.

Cheng RC, Holland W (2004) Calculation of confidence intervals for simulation output. *ACM Transactions on Modeling and Computer Simulation* 14(4):344–362.

Chick SE (2001) Input distribution selection for simulation experiments: Accounting for input uncertainty. *Operations Research* 49(5):744–758.

Chick SE (2006) Bayesian ideas and discrete event simulation: Why, what and how. Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM, Fujimoto RM, eds., *Proceedings of the 2006 Winter Simulation Conference*, 96–106 (Piscataway, New Jersey: IEEE).

Datta S, McCormick WP (1995) Bootstrap inference for a first-order autoregression with positive innovations. *Journal of the American Statistical Association* 90(432):1289–1300.

Efron B, Stein C (1981) The jackknife estimate of variance. *The Annals of Statistics* 586–596.

Ghosh S, Lam H (2019) Robust analysis in stochastic simulation: Computation and performance guarantees. *Operations Research* 67(1):232–249.

Glasserman P, Xu X (2014) Robust risk measurement and model risk. *Quantitative Finance* 14(1):29–58.

Hall P, Horowitz JL, Jing BY (1995) On blocking rules for the bootstrap with dependent data. *Biometrika* 82(3):561–574.

Hampel FR (1974) The influence curve and its role in robust estimation. *Journal of the American Statistical Association* 69(346):383–393.

Henderson SG (2003) Input modeling: Input model uncertainty: Why do we care and what should we do about it? Chick S, Sánchez PJ, Ferrin D, Morrice DJ, eds., *Proceedings of the 2003 Winter Simulation Conference*, 90–100 (Piscataway, New Jersey: IEEE).

Hu Z, Cao J, Hong LJ (2012) Robust simulation of global warming policies using the dice model. *Management science* 58(12):2190–2206.

Lam H (2016a) Advanced tutorial: Input uncertainty and robust analysis in stochastic simulation. *Winter Simulation Conference (WSC), 2016*, 178–192 (IEEE).

Lam H (2016b) Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research* 41(4):1248–1275.

Lam H, Qian H (2016) The empirical likelihood approach to simulation input uncertainty. *Winter Simulation Conference (WSC), 2016*, 791–802 (IEEE).

Lam H, Qian H (2017) Optimization-based quantification of simulation input uncertainty via empirical likelihood. *arXiv preprint arXiv:1707.05917* .

Lam H, Qian H (2018) Subsampling variance for input uncertainty quantification. *2018 Winter Simulation Conference (WSC)*, 1611–1622 (IEEE).

Law AM, Kelton WD, Kelton WD (1991) *Simulation modeling and analysis*, volume 2 (McGraw-Hill New York).

Lin Y, Song E, Nelson B (2015) Single-experiment input uncertainty. *Journal of Simulation* 9(3):249–259.

Nelson B (2013) *Foundations and Methods of Stochastic Simulation: A First Course* (Springer Science & Business Media).

Politis DN, Romano JP (1994) Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics* 2031–2050.

Politis DN, Romano JP, Wolf M (1999) *Subsampling* (Springer).

Searle SR, Casella G, McCulloch CE (2009) *Variance components*, volume 391 (John Wiley & Sons).

Sen B, Banerjee M, Woodroffe M, et al. (2010) Inconsistency of bootstrap: The grenander estimator. *The Annals of Statistics* 38(4):1953–1977.

Serfling RJ (2009) *Approximation Theorems of Mathematical Statistics*, volume 162 (John Wiley & Sons).

Song E, Nelson BL (2015) Quickly assessing contributions to input uncertainty. *IIE Transactions* 47(9):893–909.

Song E, Nelson BL (2019) Input–output uncertainty comparisons for discrete optimization via simulation. *Operations Research* 67(2):562–576.

Song E, Nelson BL, Pegden CD (2014) Advanced tutorial: Input uncertainty quantification. Tolk A, Diallo S, Ryzhov I, Yilmaz L, Buckley S, Miller J, eds., *Proceedings of the 2014 Winter Simulation Conference*, 162–176 (Piscataway, New Jersey: IEEE).

Sun Y, Apley DW, Staum J (2011) Efficient nested simulation for estimating the variance of a conditional expectation. *Operations research* 59(4):998–1007.

Sun Y, Phillips PC, Jin S (2008) Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica* 76(1):175–194.

Van der Vaart AW (2000) *Asymptotic Statistics*, volume 3 (Cambridge University Press).

Wieland JR, Schmeiser BW (2006) Stochastic gradient estimation using a single design point. Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM, Fujimoto RM, eds., *Proceedings of the 2006 Winter Simulation Conference*, 390–397 (Piscataway, New Jersey: IEEE).

Xie W, Li C, Wu Y, Zhang P (2019) A Bayesian nonparametric framework for uncertainty quantification in simulation. *arXiv preprint arXiv:1910.03766* .

Xie W, Nelson BL, Barton RR (2014) A Bayesian framework for quantifying uncertainty in stochastic simulation. *Operations Research* 62(6):1439–1452.

Xie W, Nelson BL, Barton RR (2016) Multivariate input uncertainty in output analysis for stochastic simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 27(1):1–22.

Yi Y, Xie W (2017) An efficient budget allocation approach for quantifying the impact of input uncertainty in stochastic simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 27(4):1–23.

Zhu H, Liu T, Zhou E (2020) Risk quantification in stochastic simulation under input uncertainty. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 30(1):1–24.

Zouaoui F, Wilson JR (2003) Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions* 35(9):781–792.

Zouaoui F, Wilson JR (2004) Accounting for input-model and input-parameter uncertainties in simulation. *IIE Transactions* 36(11):1135–1151.

Proofs of Statements

We first verify the proposed assumptions for the special case of finite-horizon performance measures in Section EC.1. Section EC.2 then proves results on the validity of the input variance decomposition (2). Section EC.3 proves the consistency of the proposed input variance estimate and analyzes its Monte Carlo error in relation to the parameters B, R . Lastly, Section EC.4 further analyzes the statistical error to obtain the overall error of the input variance estimate, and derives the optimal choices for θ, B, R that minimizes the overall error. In all the proofs, we write $a \approx b$ to mean $a/b \xrightarrow{P} 1$.

EC.1. Finite-Horizon Performance Measures

In this section, we show that Assumptions 2-7 and 10-12 hold for the finite-horizon performance measure (16), thereby proving Theorems 6 and 9. We first prove Assumptions 2 and 10, then present the useful Lemma EC.1 which will later be used to prove all other assumptions.

Proof of Assumptions 2 and 10. The finite horizon structure allows the following expansion of the performance measure $\psi(P_1^{\nu_1}, \dots, P_m^{\nu_m})$ around the input models P_1, \dots, P_m

$$\begin{aligned}
& \psi(P_1^{\nu_1}, \dots, P_m^{\nu_m}) \\
&= \int h(\mathbf{x}_1, \dots, \mathbf{x}_m) \prod_{i=1}^m \prod_{t=1}^{T_i} d(\nu_i(Q_i - P_i) + P_i)(x_{i,t}) \\
&= \psi(P_1, \dots, P_m) + \sum_{d=1}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \prod_{i=1}^m \nu_i^{|\mathcal{T}_i|} \int h(\mathbf{x}_1, \dots, \mathbf{x}_m) \prod_{i=1}^m \prod_{t \notin \mathcal{T}_i} dP_i(x_{i,t}) \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} d(Q_i - P_i)(x_{i,t}) \\
&= \psi(P_1, \dots, P_m) + \sum_{d=1}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \prod_{i=1}^m \nu_i^{|\mathcal{T}_i|} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} d(Q_i - P_i)(x_{i,t})
\end{aligned}$$

where $T = \sum_{i=1}^m T_i$ is the total run length, each $\mathcal{T}_i = \{\mathcal{T}_i(1), \dots, \mathcal{T}_i(|\mathcal{T}_i|)\}$ is an ordered subset of $\{1, 2, \dots, T_i\}$, and

$$h_{\mathcal{T}_1, \dots, \mathcal{T}_m}(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) = \mathbb{E}_{P_1, \dots, P_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = x_{i,t} \text{ for } i, t \in \mathcal{T}_i]. \quad (\text{EC.1})$$

Here each $\mathbf{x}_{i, \mathcal{T}_i} := (x_{i,t})_{t \in \mathcal{T}_i}$. Expressing terms with $d = 1, 2, 3$ in a more explicit form gives

$$\psi(P_1^{\nu_1}, \dots, P_m^{\nu_m})$$

$$\begin{aligned}
&= \psi(P_1, \dots, P_m) + \sum_{i=1}^m \nu_i \int \tilde{g}_i(x) d(Q_i - P_i)(x) + \sum_{i_1 \leq i_2} \nu_{i_1} \nu_{i_2} \int \tilde{g}_{i_1 i_2}(x, y) d(Q_{i_1} - P_{i_1})(x) d(Q_{i_2} - P_{i_2})(y) \\
&+ \sum_{i_1 \leq i_2 \leq i_3} \nu_{i_1} \nu_{i_2} \nu_{i_3} \int \tilde{g}_{i_1 i_2 i_3}(x, y, z) d(Q_{i_1} - P_{i_1})(x) d(Q_{i_2} - P_{i_2})(y) d(Q_{i_3} - P_{i_3})(z) \\
&+ \sum_{d=4}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \prod_{i=1}^m \nu_i^{|\mathcal{T}_i|} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} d(Q_i - P_i)(x_{i,t}). \tag{EC.2}
\end{aligned}$$

where

$$\begin{aligned}
\tilde{g}_i(x) &= \sum_{1 \leq t \leq T_i} \mathbb{E}_{P_1, \dots, P_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = x] \\
\tilde{g}_{i_1 i_2}(x, y) &= \begin{cases} \sum_{1 \leq t_1 < t_2 \leq T_i} \mathbb{E}_{P_1, \dots, P_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t_1) = x, X_i(t_2) = y], & \text{if } i_1 = i_2 = i \\ \sum_{t_1=1}^{T_{i_1}} \sum_{t_2=1}^{T_{i_2}} \mathbb{E}_{P_1, \dots, P_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_{i_1}(t_1) = x, X_{i_2}(t_2) = y], & \text{if } i_1 < i_2 \end{cases} \\
\tilde{g}_{i_1 i_2 i_3}(x, y, z) &= \begin{cases} \sum_{1 \leq t_1 < t_2 < t_3 \leq T_i} \mathbb{E}_{P_1, \dots, P_m} [h | X_i(t_1) = x, X_i(t_2) = y, X_i(t_3) = z], & \text{if } i_1 = i_2 = i_3 = i \\ \sum_{1 \leq t_1 < t_2 \leq T_i} \sum_{t_3=1}^{T_{i_3}} \mathbb{E}_{P_1, \dots, P_m} [h | X_i(t_1) = x, X_i(t_2) = y, X_{i_3}(t_3) = z], & \text{if } i_1 = i_2 = i < i_3 \\ \sum_{t_1=1}^{T_{i_1}} \sum_{1 \leq t_2 < t_3 \leq T_i} \mathbb{E}_{P_1, \dots, P_m} [h | X_{i_1}(t_1) = x, X_i(t_2) = y, X_i(t_3) = z], & \text{if } i_1 < i_2 = i_3 = i \\ \sum_{t_1=1}^{T_{i_1}} \sum_{t_2=1}^{T_{i_2}} \sum_{t_3=1}^{T_{i_3}} \mathbb{E}_{P_1, \dots, P_m} [h | X_{i_1}(t_1) = x, X_{i_2}(t_2) = y, X_{i_3}(t_3) = z], & \text{if } i_1 < i_2 < i_3 \end{cases}.
\end{aligned}$$

Since each signed measure $Q_i - P_i$ in the product measure in (EC.2) has zero total measure, adding to the integrand a function that is independent of at least one of the integration variables does not change the integral value. Hence one can replace \tilde{g} 's by the following centered versions for

$$i_1 \leq i_2 \leq i_3$$

$$\tilde{g}_i^c(x) = \tilde{g}_i(x) - \mathbb{E}[\tilde{g}_i(X_i)]$$

$$\tilde{g}_{i_1 i_2}^c(x, y) = \tilde{g}_{i_1 i_2}(x, y) - \mathbb{E}[\tilde{g}_{i_1 i_2}(X_{i_1}, y)] - \mathbb{E}[\tilde{g}_{i_1 i_2}(x, X_{i_2})] + \mathbb{E}[\tilde{g}_{i_1 i_2}(X_{i_1}, X'_{i_2})]$$

$$\tilde{g}_{i_1 i_2 i_3}^c(x, y, z) = \tilde{g}_{i_1 i_2 i_3}(x, y, z) - \mathbb{E}[\tilde{g}_{i_1 i_2 i_3}(X_{i_1}, y, z)] - \mathbb{E}[\tilde{g}_{i_1 i_2 i_3}(x, X_{i_2}, z)] - \mathbb{E}[\tilde{g}_{i_1 i_2 i_3}(x, y, X_{i_3})]$$

$$+ \mathbb{E}[\tilde{g}_{i_1 i_2 i_3}(X_{i_1}, X'_{i_2}, z)] + \mathbb{E}[\tilde{g}_{i_1 i_2 i_3}(X_{i_1}, y, X'_{i_3})] + \mathbb{E}[\tilde{g}_{i_1 i_2 i_3}(x, X_{i_2}, X'_{i_3})]$$

$$- \mathbb{E}[\tilde{g}_{i_1 i_2 i_3}(X_{i_1}, X'_{i_2}, X''_{i_3})]$$

where X_i, X'_i, X''_i denote independent variables distributed under F_i , and replace the function $h_{\mathcal{T}_1, \dots, \mathcal{T}_m}$ by

$$\begin{aligned} & h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \\ &= h_{\mathcal{T}_1, \dots, \mathcal{T}_m} - \sum_{i, t \in \mathcal{T}_i} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m} dP_i(x_{i, t}) + \sum_{(i_1, t_1) < (i_2, t_2), t_1 \in \mathcal{T}_{i_1}, t_2 \in \mathcal{T}_{i_2}} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m} dP_{i_1}(x_{i_1, t_1}) dP_{i_2}(x_{i_2, t_2}) + \dots \\ &+ (-1)^{\sum_{i=1}^m |\mathcal{T}_i|} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m} \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} dP_i(x_{i, t}) \end{aligned} \quad (\text{EC.3})$$

where the order $(i_1, t_1) < (i_2, t_2)$ is defined as either $i_1 < i_2$, or $i_1 = i_2$ but $t_1 < t_2$. This leads to the new Taylor expansion

$$\begin{aligned} & \psi(P_1^{\nu_1}, \dots, P_m^{\nu_m}) \\ &= \psi(P_1, \dots, P_m) + \sum_{i=1}^m \nu_i \int \tilde{g}_i^c(x) d(Q_i - P_i)(x) + \sum_{i_1 \leq i_2} \nu_{i_1} \nu_{i_2} \int \tilde{g}_{i_1 i_2}^c(x, y) d(Q_{i_1} - P_{i_1})(x) d(Q_{i_2} - P_{i_2})(y) \\ &+ \sum_{i_1 \leq i_2 \leq i_3} \nu_{i_1} \nu_{i_2} \nu_{i_3} \int \tilde{g}_{i_1 i_2 i_3}^c(x, y, z) d(Q_{i_1} - P_{i_1})(x) d(Q_{i_2} - P_{i_2})(y) d(Q_{i_3} - P_{i_3})(z) \\ &+ \sum_{d=4}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i| = d} \prod_{i=1}^m \nu_i^{|\mathcal{T}_i|} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} d(Q_i - P_i)(x_{i, t}). \end{aligned} \quad (\text{EC.4})$$

Note that now all the integrands above have zero marginal means due to centering, e.g.

$$\int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) dP_i(x_{i, t}) = 0 \text{ for all } i \text{ and } t \in \mathcal{T}_i. \quad (\text{EC.5})$$

However, the functions $\tilde{g}_i^c, \tilde{g}_{i_1 i_2}^c, \tilde{g}_{i_1 i_2 i_3}^c$ are not necessarily symmetric under permutations as required in Assumption 10, so we perform the following symmetrization to find the influence functions

$$\begin{aligned} g_i(x) &:= \tilde{g}_i^c(x) \\ g_{ii}(x_1, x_2) &:= \tilde{g}_{ii}^c(x_1, x_2) + \tilde{g}_{ii}^c(x_2, x_1) \\ g_{i_1 i_2}(x_1, x_2) &= g_{i_2 i_1}(x_2, x_1) := \tilde{g}_{i_1 i_2}^c(x_1, x_2) \text{ for } i_1 < i_2 \\ g_{iii}(x_1, x_2, x_3) &:= \sum_{\pi} \tilde{g}_{iii}^c(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)}) \\ g_{i_1 i_1 i_2}(x_1, x_2, x_3) &= g_{i_1 i_2 i_1}(x_1, x_3, x_2) = g_{i_2 i_1 i_1}(x_3, x_1, x_2) := \tilde{g}_{i_1 i_1 i_2}^c(x_1, x_2, x_3) + \tilde{g}_{i_1 i_1 i_2}^c(x_2, x_1, x_3) \text{ for } i_1 < i_2 \\ g_{i_1 i_2 i_2}(x_1, x_2, x_3) &= g_{i_2 i_1 i_2}(x_2, x_1, x_3) = g_{i_2 i_2 i_1}(x_2, x_3, x_1) := \tilde{g}_{i_1 i_2 i_2}^c(x_1, x_2, x_3) + \tilde{g}_{i_1 i_2 i_2}^c(x_1, x_3, x_2) \text{ for } i_1 < i_2 \\ \text{for all } \pi \text{ let } g_{i_{\pi(1)} i_{\pi(2)} i_{\pi(3)}}(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)}) &:= \tilde{g}_{i_1 i_2 i_3}^c(x_1, x_2, x_3) \text{ for } i_1 < i_2 < i_3 \end{aligned}$$

where the dependence on P_1, \dots, P_m is suppressed and π denotes any permutation of $(1, 2, 3)$. Then one can check that $g_{i_1 i_2}$ and $g_{i_1 i_2 i_3}$ not only retain the property of zero marginal means, but also satisfy the symmetry condition in Assumption 10. Permutation symmetry implies that

$$\begin{aligned}
& \psi(P_1^{\nu_1}, \dots, P_m^{\nu_m}) \\
&= \psi(P_1, \dots, P_m) + \sum_{i=1}^m \nu_i \int g_i(x) d(Q_i - P_i)(x) + \frac{1}{2} \sum_{i_1, i_2} \nu_{i_1} \nu_{i_2} \int g_{i_1 i_2}(x, y) d(Q_{i_1} - P_{i_1})(x) d(Q_{i_2} - P_{i_2})(y) \\
&+ \frac{1}{6} \sum_{i_1, i_2, i_3} \nu_{i_1} \nu_{i_2} \nu_{i_3} \int g_{i_1 i_2 i_3}(x, y, z) d(Q_{i_1} - P_{i_1})(x) d(Q_{i_2} - P_{i_2})(y) d(Q_{i_3} - P_{i_3})(z) \\
&+ \sum_{d=4}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \prod_{i=1}^m \nu_i^{|\mathcal{T}_i|} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} d(Q_i - P_i)(x_{i,t}). \tag{EC.6}
\end{aligned}$$

Since the integrals are all finite under Assumption 9, the first-order and third-order remainders of the above expansion are respectively of order $O(\sum_{i=1}^m \nu_i^2)$ and $O((\sum_{i=1}^m \nu_i^2)^2)$, leading to Assumptions 2 and 10. \square

We continue to verify other assumptions, for which we use the following lemma.

LEMMA EC.1. *Suppose Assumption 9 holds with positive and even k . For each i let $\tilde{F}_i \in \{F_i, \hat{F}_i\}$ be either the i -th true or empirical input model. Then the following bounds hold uniformly for every $(\tilde{F}_1, \dots, \tilde{F}_m) \in \prod_{i=1}^m \{F_i, \hat{F}_i\}$ and arbitrary input data size n_i*

$$\max_{I_1, \dots, I_m} \mathbb{E}_{\tilde{F}_1, \dots, \tilde{F}_m} [h^k(\mathbf{X}_{1, I_1}, \dots, \mathbf{X}_{m, I_m})] = O_p(1) \tag{EC.7}$$

$$\mathbb{E}_{F_1, \dots, F_m} [(\psi(\tilde{F}_1, \dots, \tilde{F}_m) - \psi(F_1, \dots, F_m))^k] \leq C_1 \mathcal{M} \left(\sum_{i=1}^m \frac{1}{\sqrt{n_i}} \right)^k \tag{EC.8}$$

$$\begin{aligned}
& \mathbb{E}_{F_1, \dots, F_m} [(\psi(\tilde{F}_1, \dots, \tilde{F}_m) - \psi(F_1, \dots, F_m) - \sum_{i=1}^m \int g_i(x) d(\tilde{F}_i - F_i)(x))^k] \leq C_2 \mathcal{M} \left(\sum_{i=1}^m \frac{1}{\sqrt{n_i}} \right)^{2k} \\
& \tag{EC.9}
\end{aligned}$$

where the influence functions g_i 's are now under the true input models F_1, \dots, F_m . Each empirical influence function \hat{g}_i satisfies

$$\mathbb{E}_{F_1, \dots, F_m} [(g_i(X_{i,1}) - \hat{g}_i(X_{i,1}))^k] \leq C_3 \mathcal{M} \left(\sum_{i=1}^m \frac{1}{\sqrt{n_i}} \right)^k \tag{EC.10}$$

$$\begin{aligned}
& \mathbb{E}_{F_1, \dots, F_m} [(\hat{g}_i(X_{i,1}) - g_i(X_{i,1}) - \sum_{i'=1}^m \int g_{ii'}(X_{i,1}, x) d(\hat{F}_{i'} - F_{i'})(x) + \int g_i(x) d(\hat{F}_i - F_i)(x))^k] \leq C_4 \mathcal{M} \left(\sum_{i=1}^m \frac{1}{\sqrt{n_i}} \right)^{2k} \\
& \tag{EC.11}
\end{aligned}$$

Here C_1, C_2, C_3, C_4 are constants that only depend on k, m and $T := \sum_{i=1}^m T_i$, and

$$\mathcal{M} := \max_{I_1, \dots, I_m} \mathbb{E}_{F_1, \dots, F_m} [h^k(\mathbf{X}_{1, I_1}, \dots, \mathbf{X}_{m, I_m})] < \infty.$$

Proof. The first bound is the most straightforward. By rewriting the expectation $\mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [\cdot]$ as a sum, one can see that for a particular choice of I_1, \dots, I_m

$$\mathbb{E}_{F_1, \dots, F_m} \left[\mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^k(\mathbf{X}_{1, I_1}, \dots, \mathbf{X}_{m, I_m})] \right] \leq \mathcal{M}.$$

Therefore $\mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^k(\mathbf{X}_{1, I_1}, \dots, \mathbf{X}_{m, I_m})] = O_p(1)$ for each I_1, \dots, I_m . Since there are finitely many of them, the maximum is also bounded in probability. This proves the first bound.

To explain the other bounds, we put $\psi(\widetilde{F}_1, \dots, \widetilde{F}_m)$ in the form of the expansion (EC.4) with $\nu_i = 1, P_i = F_i, Q_i = \widetilde{F}_i$ to get

$$\begin{aligned} & \psi(\widetilde{F}_1, \dots, \widetilde{F}_m) \\ &= \psi(F_1, \dots, F_m) + \sum_{d=1}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} d\widetilde{F}_i(x_{i,t}) \\ &= \psi(F_1, \dots, F_m) + \sum_{i=1}^m \int \left(\sum_{t=1}^{T_i} \mathbb{E}_{F_1, \dots, F_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = x] - T_i \psi(F_1, \dots, F_m) \right) d(\widetilde{F}_i - F_i)(x) \\ & \quad + \sum_{d=2}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} d\widetilde{F}_i(x_{i,t}) \end{aligned}$$

where $\int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) dF_i(x_{i,t}) = 0$ for all i and $t \in \mathcal{T}_i$, according to the property of zero marginal means (EC.5). To obtain a moment bound for $h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c$, observe that by Assumption 9 and Jensen's inequality any conditional expectation of the performance function h has a k -th moment at most \mathcal{M} . Since $h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c$ is the sum of several conditional expectations of h , one can apply Minkowski inequality to establish that for any $I_i = (I_i(1), \dots, I_i(|\mathcal{T}_i|)) \in \{1, 2, \dots, |\mathcal{T}_i|\}^{|\mathcal{T}_i|}$, $i = 1, \dots, m$

$$\mathbb{E}_{F_1, \dots, F_m} [(h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{X}_{1, \mathcal{T}_1(I_1)}, \dots, \mathbf{X}_{m, \mathcal{T}_m(I_m)}))^k] \leq 2^{k \sum_{i=1}^m |\mathcal{T}_i|} \mathcal{M}. \quad (\text{EC.12})$$

Again by Minkowski inequality

$$\mathbb{E}_{F_1, \dots, F_m} \left[(\psi(\widetilde{F}_1, \dots, \widetilde{F}_m) - \psi(F_1, \dots, F_m))^k \right]$$

$$\begin{aligned}
&\leq \left(\sum_{d=1}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \left(\mathbb{E}_{F_1, \dots, F_m} \left[\left(\int h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_1, \mathcal{T}_1, \dots, \mathbf{x}_m, \mathcal{T}_m) \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} d\tilde{F}_i(x_{i,t}) \right)^k \right] \right)^{\frac{1}{k}} \right)^k \\
&= \left(\sum_{d=1}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \prod_{i=1}^m \prod_{t \in \mathcal{T}_i} \mathbf{1}(\tilde{F}_{i,t} = \hat{F}_i) \left(\mathbb{E}_{F_1, \dots, F_m} \left[\left(\frac{1}{\prod_{i=1}^m n_i^{|\mathcal{T}_i|}} \sum_{J_1, \dots, J_m} h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{X}_1, J_1, \dots, \mathbf{X}_m, J_m) \right)^k \right] \right)^{\frac{1}{k}} \right)^k \\
&\leq \left(\sum_{d=1}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \left(\mathbb{E}_{F_1, \dots, F_m} \left[\left(\frac{1}{\prod_{i=1}^m n_i^{|\mathcal{T}_i|}} \sum_{J_1, \dots, J_m} h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{X}_1, J_1, \dots, \mathbf{X}_m, J_m) \right)^k \right] \right)^{\frac{1}{k}} \right)^k
\end{aligned}$$

where each $J_i = (J_i(1), \dots, J_i(|\mathcal{T}_i|)) \in \{1, 2, \dots, n_i\}^{|\mathcal{T}_i|}$ and $\mathbf{X}_{i,J_i} = (X_{i,J_i(1)}, \dots, X_{i,J_i(|\mathcal{T}_i|)})$. Note that

$$\begin{aligned}
&\mathbb{E}_{F_1, \dots, F_m} \left[\left(\frac{1}{\prod_{i=1}^m n_i^{|\mathcal{T}_i|}} \sum_{J_1, \dots, J_m} h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{X}_1, J_1, \dots, \mathbf{X}_m, J_m) \right)^k \right] \\
&= \frac{1}{\prod_{i=1}^m n_i^{k|\mathcal{T}_i|}} \sum_{J_1^1, \dots, J_m^1} \dots \sum_{J_1^k, \dots, J_m^k} \mathbb{E}_{F_1, \dots, F_m} [h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{X}_1, J_1^1, \dots, \mathbf{X}_m, J_m^1) \dots h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{X}_1, J_1^k, \dots, \mathbf{X}_m, J_m^k)].
\end{aligned}$$

By (EC.5) the expectation on the right hand side is zero if some data point $X_{i,j}$ appears only once.

Therefore the number of non-zero expectations is bounded above by $C(k, m, \sum_{i=1}^m |\mathcal{T}_i|) \prod_{i=1}^m n_i^{k|\mathcal{T}_i|/2}$,

where $C(k, m, \sum_{i=1}^m |\mathcal{T}_i|)$ is some constant that only depends on $k, m, \sum_{i=1}^m |\mathcal{T}_i|$. Moreover, from (EC.12) each expectation satisfies the following by generalized Hölder's inequality

$$|\mathbb{E}_{F_1, \dots, F_m} [h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{X}_1, J_1^1, \dots, \mathbf{X}_m, J_m^1) \dots h_{\mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{X}_1, J_1^k, \dots, \mathbf{X}_m, J_m^k)]| \leq 2^{k \sum_{i=1}^m |\mathcal{T}_i|} \mathcal{M}.$$

Hence

$$\begin{aligned}
&\mathbb{E}_{F_1, \dots, F_m} \left[(\psi(\hat{F}_1, \dots, \hat{F}_m) - \psi(F_1, \dots, F_m))^k \right] \\
&\leq \left(\sum_{d=1}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} \left(C(k, m, \sum_{i=1}^m |\mathcal{T}_i|) \prod_{i=1}^m n_i^{-k|\mathcal{T}_i|/2} 2^{k \sum_{i=1}^m |\mathcal{T}_i|} \mathcal{M} \right)^{\frac{1}{k}} \right)^k \\
&= \left(\sum_{d=1}^T \sum_{\sum_{i=1}^m |\mathcal{T}_i|=d} C'(k, m, d) \prod_{i=1}^m n_i^{-|\mathcal{T}_i|/2} \mathcal{M}^{\frac{1}{k}} \right)^k \\
&\leq \left(\sum_{d=1}^T C'(k, m, d) \left(\sum_{i=1}^m \frac{T_i}{\sqrt{n_i}} \right)^d \right)^k \mathcal{M} \leq C_1(k, m, T) \mathcal{M} \left(\sum_{i=1}^m \frac{1}{\sqrt{n_i}} \right)^k.
\end{aligned}$$

This gives the second bound.

The third bound can be established by the same argument, but considering only the remainders for which $d \geq 2$.

We then prove the bounds on influence functions. According to the representation of $g_i(P_1, \dots, P_m; \cdot)$ in the proof of Assumptions 2 and 10, the empirical influence function \hat{g}_i is

$$\hat{g}_i(x) = \sum_{t=1}^{T_i} \mathbb{E}_{\hat{F}_1, \dots, \hat{F}_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = x] - T_i \psi(\hat{F}_1, \dots, \hat{F}_m).$$

First we derive the following Taylor expansion for each conditional expectation

$$\begin{aligned} & \mathbb{E}_{\hat{F}_1, \dots, \hat{F}_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = X_{i,1}] \\ &= \int h(\mathbf{x}_1, \dots, \mathbf{x}_m) \prod_{t' \neq t} d\hat{F}_i(x_{i,t'}) \prod_{i' \neq i} \prod_{t'=1}^{T_{i'}} d\hat{F}_{i'}(x_{i',t'}) \Big|_{x_{i,t}=X_{i,1}} \\ &= \mathbb{E}_{F_1, \dots, F_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = X_{i,1}] + \\ & \quad + \sum_{d=1}^{T-1} \sum_{\sum_{i'=1}^m |\mathcal{T}_{i'}| = d, t \notin \mathcal{T}_i} \int h(\mathbf{x}_1, \dots, \mathbf{x}_m) \prod_{t' \notin \mathcal{T}_i, t' \neq t} dF_i(x_{i,t'}) \prod_{i' \neq i} \prod_{t' \notin \mathcal{T}_{i'}} dF_{i'}(x_{i',t'}) \prod_{i'=1}^m \prod_{t' \in \mathcal{T}_{i'}} d(\hat{F}_{i'} - F_{i'})(x_{i',t'}) \Big|_{x_{i,t}=X_{i,1}} \\ &= \mathbb{E}_{F_1, \dots, F_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = X_{i,1}] \\ & \quad + \sum_{d=1}^{T-1} \sum_{\sum_{i'=1}^m |\mathcal{T}_{i'}| = d, t \notin \mathcal{T}_i} \int h_{(i,t), \mathcal{T}_1, \dots, \mathcal{T}_m}(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{i, \mathcal{T}_i \cup \{t\}}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \prod_{i'=1}^m \prod_{t' \in \mathcal{T}_{i'}} d(\hat{F}_{i'} - F_{i'})(x_{i',t'}) \Big|_{x_{i,t}=X_{i,1}} \\ &= \mathbb{E}_{F_1, \dots, F_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = X_{i,1}] \end{aligned} \tag{EC.13}$$

$$\begin{aligned} & + \sum_{t'=1, t' \neq t}^{T_i} \int \mathbb{E}_{F_1, \dots, F_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = X_{i,1}, X_i(t') = x_{i,t'}] d(\hat{F}_i - F_i)(x_{i,t'}) \\ & + \sum_{i' \neq i} \sum_{t'=1}^{T_{i'}} \int \mathbb{E}_{F_1, \dots, F_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = X_{i,1}, X_{i'}(t') = x_{i',t'}] d(\hat{F}_{i'} - F_{i'})(x_{i',t'}) \tag{EC.14} \\ & + \sum_{d=2}^{T-1} \sum_{\sum_{i'=1}^m |\mathcal{T}_{i'}| = d, t \notin \mathcal{T}_i} \int h_{(i,t), \mathcal{T}_1, \dots, \mathcal{T}_m}(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{i, \mathcal{T}_i \cup \{t\}}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \prod_{i'=1}^m \prod_{t' \in \mathcal{T}_{i'}} d(\hat{F}_{i'} - F_{i'})(x_{i',t'}) \Big|_{x_{i,t}=X_{i,1}} \end{aligned}$$

where each $\mathcal{T}_{i'} = \{\mathcal{T}_{i'}(1), \dots, \mathcal{T}_{i'}(|\mathcal{T}_{i'}|)\}$ is still an ordered subset of $\{1, 2, \dots, T_{i'}\}$ but $t \notin \mathcal{T}_i$, and the function $h_{(i,t), \mathcal{T}_1, \dots, \mathcal{T}_m}$ resembles (EC.1) except that the expectation is now further conditioned on $X_i(t) = x_{i,t}$. Introduce the counterpart of (EC.3)

$$\begin{aligned} & h_{(i,t), \mathcal{T}_1, \dots, \mathcal{T}_m}^c(\mathbf{x}_{1, \mathcal{T}_1}, \dots, \mathbf{x}_{i, \mathcal{T}_i \cup \{t\}}, \dots, \mathbf{x}_{m, \mathcal{T}_m}) \\ &= h_{(i,t), \mathcal{T}_1, \dots, \mathcal{T}_m} - \sum_{i', t' \in \mathcal{T}_{i'}} \int h_{(i,t), \mathcal{T}_1, \dots, \mathcal{T}_m} dF_{i'}(x_{i',t'}) + \sum_{(i'_1, t'_1) < (i'_2, t'_2), t'_1 \in \mathcal{T}_{i'_1}, t'_2 \in \mathcal{T}_{i'_2}} \int h_{(i,t), \mathcal{T}_1, \dots, \mathcal{T}_m} dF_{i'_1}(x_{i'_1, t'_1}) dF_{i'_2}(x_{i'_2, t'_2}) \\ & \quad + \dots + (-1)^{\sum_{i'=1}^m |\mathcal{T}_{i'}|} \int h_{(i,t), \mathcal{T}_1, \dots, \mathcal{T}_m} \prod_{i'=1}^m \prod_{t' \in \mathcal{T}_{i'}} dF_{i'}(x_{i',t'}) \end{aligned}$$

then we have the following parallel property of (EC.5)

$$\int h_{(i,t),\mathcal{T}_1,\dots,\mathcal{T}_m}^c(\mathbf{x}_{1,\mathcal{T}_1},\dots,\mathbf{x}_{i,\mathcal{T}_i \cup \{t\}},\dots,\mathbf{x}_{m,\mathcal{T}_m})dF_{i'}(x_{i',t'}) = 0 \text{ for all } i' \text{ and } t' \in \mathcal{T}_{i'}$$

and by comparing the first order remainders (EC.13) and (EC.14) of \hat{g}_i with the second order influence functions $g_{i_1 i_2}$ it is easy to establish that

$$\begin{aligned} & \hat{g}_i(X_{i,1}) - g_i(X_{i,1}) \\ &= \sum_{t=1}^{T_i} (\mathbb{E}_{\hat{F}_1,\dots,\hat{F}_m}[h(\mathbf{X}_1,\dots,\mathbf{X}_m)|X_i(t)=X_{i,1}] - \mathbb{E}_{F_1,\dots,F_m}[h(\mathbf{X}_1,\dots,\mathbf{X}_m)|X_i(t)=X_{i,1}]) \\ & \quad - T_i(\psi(\hat{F}_1,\dots,\hat{F}_m) - \psi(F_1,\dots,F_m)) \\ &= \sum_{i'=1}^m \int g_{ii'}(X_{i,1},x)d(\hat{F}_{i'} - F_{i'})(x) - \int g_i(x)d(\hat{F}_i - F_i)(x) \end{aligned} \tag{EC.15}$$

$$\begin{aligned} & + \sum_{t=1}^{T_i} \sum_{d=2}^{T-1} \sum_{\sum_{i'=1}^m |\mathcal{T}_{i'}|=d} \int h_{(i,t),\mathcal{T}_1,\dots,\mathcal{T}_m}^c(\mathbf{x}_{1,\mathcal{T}_1},\dots,\mathbf{x}_{i,\mathcal{T}_i \cup \{t\}},\dots,\mathbf{x}_{m,\mathcal{T}_m}) \prod_{i'=1}^m \prod_{t' \in \mathcal{T}_{i'}} d\hat{F}_{i'}(x_{i',t'}) \Big|_{x_{i,t}=X_{i,1}} \\ & - T_i \sum_{d=2}^T \sum_{\sum_{i'=1}^m |\mathcal{T}_{i'}|=d} \int h_{\mathcal{T}_1,\dots,\mathcal{T}_m}^c(\mathbf{x}_{1,\mathcal{T}_1},\dots,\mathbf{x}_{m,\mathcal{T}_m}) \prod_{i'=1}^m \prod_{t' \in \mathcal{T}_{i'}} d\hat{F}_{i'}(x_{i',t'}) \end{aligned} \tag{EC.16}$$

By a similar technique used to bound the remainder of $\psi(\tilde{F}_1,\dots,\tilde{F}_m)$, we can establish that the remainder (EC.16) has a k -th moment of order $O(\mathcal{M}(\sum_{i=1}^m \frac{1}{\sqrt{n_i}})^{2k})$, and the first order term (EC.15) has a k -th moment of order $O(\mathcal{M}(\sum_{i=1}^m \frac{1}{\sqrt{n_i}})^k)$. This completes the proof. \square

With Lemma EC.1 we now prove the other assumptions:

Proof of Assumption 3. The moment bound on the remainder, i.e. $\mathbb{E}[\epsilon^2] = o(n^{-1})$, comes from the bound (EC.9) in Lemma EC.1 with $\tilde{F}_i = \hat{F}_i$ for all i and $k = 2$. The non-degeneracy condition on the influence functions is exactly Assumption 8, whereas the finiteness of fourth order moments of g_i easily follows because g_i is simply a sum of T_i conditional expectations of the performance function h and each of the conditional expectations has finite fourth order moment by Assumption 9 and Jensen's inequality. \square

Proof of Assumption 4. The convergence of \hat{g}_i to g_i in fourth order moment is a direct consequence of the bound (EC.10) in Lemma EC.1 with $k = 4$. The moment condition on the remainder ϵ^* can be argued as follows. We treat the empirical distributions $\hat{F}_1,\dots,\hat{F}_m$ as the

truth, and the resampled distributions $\widehat{F}_{s_1,1}^*, \dots, \widehat{F}_{s_m,m}^*$ as the input data, then apply the third bound (EC.9) in Lemma EC.1 with $k = 4$ to get $\mathbb{E}_*[(\epsilon^*)^4] \leq C_2 \widehat{\mathcal{M}} \left(\sum_{i=1}^m \frac{1}{\sqrt{s_i}} \right)^8$, where $\widehat{\mathcal{M}} = \max_{I_1, \dots, I_m} \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^4(\mathbf{X}_{1,I_1}, \dots, \mathbf{X}_{m,I_m})]$ is $O_p(1)$ by the first bound (EC.7) in Lemma EC.1 with $k = 4$. Therefore $\mathbb{E}_*[(\epsilon^*)^4] = O_p((\sum_{i=1}^m \frac{1}{s_i})^4) = o_p(s^{-2})$. \square

Proof of Assumption 5. It suffices to show that $\mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^2] \xrightarrow{p} \mathbb{E}_{F_1, \dots, F_m} [h^2]$ and $\mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h] \xrightarrow{p} \mathbb{E}_{F_1, \dots, F_m} [h]$. The latter convergence follows from the second bound (EC.8) of Lemma EC.1 with $k = 2$ and $\widetilde{F}_i = \widehat{F}_i$ for all i . Since Assumption 9 holds with $k = 4$ for the function h , it also holds with $k = 2$ for the squared function h^2 . One can apply the same bound from Lemma EC.1 with $k = 2$ to h^2 and then conclude the former convergence. \square

Proof of Assumption 6. We write $\bar{\tau}^2 = \tau^2(\bar{F}_1, \dots, \bar{F}_m)$ for short. First rewrite

$$\begin{aligned} (\bar{\tau}^2 - \hat{\tau}^2)^2 &= (\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h^2] - \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^2] - ((\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h])^2 - (\mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h])^2))^2 \\ &\leq 2(\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h^2] - \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^2])^2 + 2((\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h])^2 - (\mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h])^2)^2 \\ &\leq 2(\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h^2] - \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^2])^2 + 4(\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h] - \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h])^4 \\ &\quad + 16(\mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h])^2 (\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h] - \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h])^2. \end{aligned}$$

Applying Lemma EC.1 to h^2 ($k = 2$) with the true distributions being $\widehat{F}_1, \dots, \widehat{F}_m$ we get

$$\mathbb{E}_*[(\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h^2] - \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^2])^2] \leq C_1 \widehat{\mathcal{M}} \left(\sum_{i=1}^m \frac{1}{\sqrt{s_i}} \right)^2 = O_p \left(\sum_{i=1}^m \frac{1}{s_i} \right)$$

where $\widehat{\mathcal{M}} = \max_{I_1, \dots, I_m} \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^4(\mathbf{X}_{1,I_1}, \dots, \mathbf{X}_{m,I_m})] = O_p(1)$. Another application of Lemma EC.1 to h with $k = 4$ gives

$$\mathbb{E}_*[(\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h] - \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h])^4] \leq C_1 \widehat{\mathcal{M}} \left(\sum_{i=1}^m \frac{1}{\sqrt{s_i}} \right)^4 = O_p \left(\sum_{i=1}^m \frac{1}{s_i^2} \right)$$

which implies that $\mathbb{E}_*[(\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h] - \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h])^2] = O_p \left(\sum_{i=1}^m \frac{1}{s_i} \right)$ as a consequence of Cauchy Schwartz inequality. Therefore in sum $\mathbb{E}_*[(\bar{\tau}^2 - \hat{\tau}^2)^2] = O_p \left(\sum_{i=1}^m \frac{1}{s_i} \right) = o_p(1)$. \square

Proof of Assumption 7. Note that $\mu_4(\bar{F}_1, \dots, \bar{F}_m) \leq C \mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h^4]$ for some absolute constant $C > 0$, therefore

$$\mathbb{E}_*[\mu_4(\bar{F}_1, \dots, \bar{F}_m)] \leq C \mathbb{E}_*[\mathbb{E}_{\bar{F}_1, \dots, \bar{F}_m} [h^4]] \leq C \max_{I_1, \dots, I_m} \mathbb{E}_{\widehat{F}_1, \dots, \widehat{F}_m} [h^4(\mathbf{X}_{1,I_1}, \dots, \mathbf{X}_{m,I_m})] = O_p(1)$$

where the last equality is due to the first bound (EC.7) in Lemma EC.1. \square

Proof of Assumption 11. The third order remainder ϵ_3 , or equivalently the sum over $d \geq 4$ in (EC.6) with each $\nu_i = 1$, consists of integrals under the product of at least four signed measures of the form $\widehat{F}_i - F_i$. Therefore, by employing the technique used in proving the second and third bounds (EC.8)(EC.9) in Lemma EC.1, one can show that $\mathbb{E}[\epsilon_3^2] = O(n^{-4})$. The details are omitted since they highly resemble those of Lemma EC.1. The fourth moments of $g_{i_1 i_2}$ and $g_{i_1 i_2 i_3}$ are finite, because each of them is a finite sum of conditional expectations of h which have finite fourth order moments due to Assumption 9 with $k = 4$ and Jensen's inequality. \square

Proof of Assumption 12. For the third order remainder of the resampled performance measure, one can derive the bound $\mathbb{E}_*[(\epsilon_3^*)^2] = O_p(s^{-4})$ in a similar way as in showing the bound (EC.9) in Lemma EC.1. The details are omitted to avoid repetition. Moreover, some straightforward modifications of the proof for the bound (EC.10) in Lemma EC.1 lead to $O(n^{-1})$ upper bounds for the the mean squared errors of second and third order influence functions. The remainder in the Taylor expansion of the first order empirical influence function satisfies $\mathbb{E}[\epsilon_g^2] = O(n^{-2})$ due to the bound (EC.11) in Lemma EC.1 with $k = 2$. \square

EC.2. Proofs of Propositions 1 and 2

This section proves results concerning the validity of the additive decomposition 2 of the input variance. We first prove Proposition 1, and then provide the key Lemma EC.2 that will be used in the proof of Proposition 2 as well as many results in Section EC.4.

Proof of Proposition 1. Following the expansion (14) we can write

$$\begin{aligned} \text{Var}[\psi(\widehat{F}_1, \dots, \widehat{F}_m)] &= \text{Var}\left[\sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j})\right] + \text{Var}[\epsilon] + 2\text{Cov}\left(\sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j}), \epsilon\right) \\ &= \sum_{i=1}^m \frac{\sigma_i^2}{n_i} + o(n^{-1}) + O\left(\sqrt{\text{Var}\left[\sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j})\right] \text{Var}[\epsilon]}\right) \\ &= \sum_{i=1}^m \frac{\sigma_i^2}{n_i} + o(n^{-1}). \end{aligned}$$

This completes the proof. \square

The following important lemma on variance decomposition plays a crucial role in our analysis.

LEMMA EC.2 (ANOVA decomposition, adapted from Efron and Stein (1981)). *Let $Y_i, i = 1, \dots, n$ be independent but not necessarily identically distributed random variables, and $\phi(y_1, \dots, y_n)$ be a function such that $\mathbb{E}[\phi^2(Y_1, \dots, Y_n)] < \infty$, then there exist functions ϕ_{i_1, \dots, i_k} for $1 \leq i_1 < \dots < i_k \leq n$ and $k \leq n$ such that*

$$\begin{aligned} & \phi(Y_1, \dots, Y_n) \\ &= \mu + \sum_{i=1}^n \phi_i(Y_i) + \sum_{i_1 < i_2} \phi_{i_1, i_2}(Y_{i_1}, Y_{i_2}) + \dots + \sum_{i_1 < \dots < i_k} \phi_{i_1, \dots, i_k}(Y_{i_1}, \dots, Y_{i_k}) + \dots + \phi_{1, \dots, n}(Y_1, \dots, Y_n) \end{aligned}$$

where

$$\mu = \mathbb{E}[\phi(Y_1, \dots, Y_n)]$$

$$\phi_i(y) = \mathbb{E}[\phi(Y_1, \dots, Y_n) | Y_i = y] - \mu$$

$$\phi_{i_1, i_2}(y_1, y_2) = \mathbb{E}[\phi(Y_1, \dots, Y_n) | Y_{i_1} = y_1, Y_{i_2} = y_2] - \phi_{i_1}(y_1) - \phi_{i_2}(y_2) - \mu$$

⋮

Moreover, the $2^n - 1$ random variables in the decomposition have mean zero and are mutually uncorrelated.

With this lemma, we can prove Proposition 2:

Proof of Proposition 2. The proof of Proposition 1 derives the following expression for input variance

$$\text{Var}[\psi(\hat{F}_1, \dots, \hat{F}_m)] = \text{Var}\left[\sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j})\right] + \text{Var}[\epsilon] + 2\text{Cov}\left(\sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j}), \epsilon\right)$$

where the covariances can be simplified to

$$\begin{aligned} \text{Cov}\left(\sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j}), \epsilon\right) &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbb{E}[g_i(X_{i,j})(\epsilon - \mathbb{E}[\epsilon])] \\ &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbb{E}[g_i(X_{i,j})(\mathbb{E}[\epsilon | X_{i,j}] - \mathbb{E}[\epsilon])] \\ &= \sum_{i=1}^m \mathbb{E}[g_i(X_{i,1})(\mathbb{E}[\epsilon | X_{i,1}] - \mathbb{E}[\epsilon])]. \end{aligned}$$

Using the cubic expansion in Assumption 11 and the vanishing marginal expectations of influence functions we have

$$\begin{aligned}\mathbb{E}[\epsilon|X_{i,1}] - \mathbb{E}[\epsilon] &= \frac{1}{2n_i^2}(g_{ii}(X_{i,1}, X_{i,1}) - \mathbb{E}[g_{ii}(X_i, X_i)]) + \frac{1}{6n_i^3}(g_{iii}(X_{i,1}, X_{i,1}, X_{i,1}) - \mathbb{E}[g_{iii}(X_i, X_i, X_i)]) \\ &\quad + \frac{n_i - 1}{2n_i^3}\mathbb{E}_{X_i}[g_{iii}(X_{i,1}, X_i, X_i)] + \sum_{i' \neq i} \frac{1}{2n_i n_{i'}} \mathbb{E}_{X_{i'}}[g_{ii' i'}(X_{i,1}, X_{i'}, X_{i'})] \\ &\quad + \mathbb{E}[\epsilon_3|X_{i,1}] - \mathbb{E}[\epsilon_3].\end{aligned}\tag{EC.17}$$

Each term except the last in (EC.17) has a second moment of order $O(n^{-4})$. To argue the last term $\mathbb{E}[\epsilon_3|X_{i,1}] - \mathbb{E}[\epsilon_3]$ also has a second moment of order at most $O(n^{-4})$, note that ϵ_3 is a symmetric statistic hence by Lemma EC.2 $\text{Var}[\mathbb{E}[\epsilon_3|X_{i,1}]] \leq \text{Var}[\epsilon_3]/n_i$ and $\text{Var}[\epsilon_3] = o(n^{-3})$ by assumption, hence $\text{Var}[\mathbb{E}[\epsilon_3|X_{i,1}]] = o(n^{-4})$. This leads to

$$\text{Var}[\mathbb{E}[\epsilon|X_{i,1}]] = O(n^{-4}).$$

Using Cauchy Schwartz inequality we conclude $\text{Cov}(\sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j}), \epsilon) = O(n^{-2})$. On the other hand, one can easily show $\text{Var}[\epsilon] = O(n^{-2})$ by using the same technique in the proof of Lemma EC.1 to bound each term in the cubic expansion. This leads to the desired conclusion. \square

EC.3. Proofs for Results in Section 4.2 and Section 3.2

We now prove the consistency of our proportionate subsampled bootstrap variance σ_{SVB}^2 (Theorem 7), and derive the mean square error of the Monte Carlo estimate $\hat{\sigma}_{SVB}^2$ relative to σ_{SVB}^2 (Lemma 1). These results will then be used to prove Theorems 2 and 8. Theorem 3, Corollaries 1-2 are consequences of Theorem 2. Theorem 4 is a consequence of Theorem 8.

Recall that $\sigma_i^2 = \text{Var}_{F_i}[g_i(X_i)]$ is the variance of the i -th influence function. For its empirical counterpart \hat{g}_i we denote by $\hat{\sigma}_i^2 := \text{Var}_{\hat{F}_i}[\hat{g}_i(X_i)]$ its variance under the empirical input models. Under the convergence condition $\mathbb{E}[(\hat{g}_i - g_i)^4(X_{i,1})] \rightarrow 0$ in Assumption 4, the convergence of $\hat{\sigma}_i^2$ to σ_i^2 follows from

$$\begin{aligned}\left| \hat{\sigma}_i^2 - \frac{1}{n_i} \sum_{j=1}^{n_i} g_i^2(X_{i,j}) \right| &= \left| \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{g}_i^2(X_{i,j}) - \frac{1}{n_i} \sum_{j=1}^{n_i} g_i^2(X_{i,j}) \right| \\ &\leq \frac{2}{n_i} \sqrt{\sum_{j=1}^{n_i} g_i^2(X_{i,j}) \sum_{j=1}^{n_i} (\hat{g}_i - g_i)^2(X_{i,j})} + \frac{1}{n_i} \sum_{j=1}^{n_i} (\hat{g}_i - g_i)^2(X_{i,j}) = o_p(1)\end{aligned}$$

and that $\sum_{j=1}^{n_i} g_i^2(X_{i,j})/n_i \xrightarrow{p} \sigma_i^2$. For convenience we denote by

$$\psi^* = \psi(\hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*), \hat{\psi}^* = \hat{\psi}(\hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*)$$

the expected value and a single simulation replication, respectively, of the performance measure under the resampled input models, and by

$$\hat{\tau}_*^2 = \tau^2(\hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*), \hat{\mu}_4^* = \mu_4(\hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*)$$

the variance and central fourth moment of a single Monte Carlo replication $\hat{\psi}^*$ conditioned on the resampled input models.

Proof of Theorem 7. Let $s_i = \lfloor \theta n_i \rfloor$. Following the expansion (15) with each $\bar{F}_i = \hat{F}_{s_i,i}^*$ we have

$$\begin{aligned} \text{Var}_*[\psi^*] &= \text{Var}_* \left[\sum_{i=1}^m \frac{1}{s_i} \sum_{k=1}^{s_i} \hat{g}_i(X_{i,k}^*) + \epsilon^* \right] \\ &= \text{Var}_* \left[\sum_{i=1}^m \frac{1}{s_i} \sum_{k=1}^{s_i} \hat{g}_i(X_{i,k}^*) \right] + \text{Var}_*[\epsilon^*] + 2\text{Cov}_* \left(\sum_{i=1}^m \frac{1}{s_i} \sum_{k=1}^{s_i} \hat{g}_i(X_{i,k}^*), \epsilon^* \right) \\ &= \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{s_i} + \text{Var}_*[\epsilon^*] + O \left(\sqrt{\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{s_i} \text{Var}_*[\epsilon^*]} \right) \\ &= \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{\lceil \theta n_i \rceil} + \text{Var}_*[\epsilon^*] + O \left(\sqrt{\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{\lceil \theta n_i \rceil} \text{Var}_*[\epsilon^*]} \right). \end{aligned}$$

Hence

$$\begin{aligned} \sigma_{SVB}^2 &= \theta \text{Var}_*[\psi^*] = \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{\lceil \theta n_i \rceil / \theta} + \theta \text{Var}_*[\epsilon^*] + O \left(\sqrt{\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{\lceil \theta n_i \rceil / \theta} \theta \text{Var}_*[\epsilon^*]} \right) \\ &= \sum_{i=1}^m \left(\frac{\hat{\sigma}_i^2}{n_i} + O \left(\frac{\hat{\sigma}_i^2}{n_i^2 \theta} \right) \right) + \theta \text{Var}_*[\epsilon^*] + O \left(\sqrt{\sum_{i=1}^m \left(\frac{\hat{\sigma}_i^2}{n_i} + O \left(\frac{\hat{\sigma}_i^2}{n_i^2 \theta} \right) \right) \theta \text{Var}_*[\epsilon^*]} \right). \end{aligned} \quad (\text{EC.18})$$

The convergence $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$ and that $\theta = \omega(1/n)$ allow us to conclude

$$\frac{1}{\theta} \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{n_i^2} = o_p \left(\sum_{i=1}^m \frac{\sigma_i^2}{n_i} \right), \theta \text{Var}_*[\epsilon^*] = \theta o_p \left(\sum_{i=1}^m \frac{1}{\lceil \theta n_i \rceil} \right) = o_p \left(\sum_{i=1}^m \frac{1}{n_i} \right)$$

therefore $\sigma_{SVB}^2 = \sum_{i=1}^m \frac{\sigma_i^2}{n_i} + o_p \left(\sum_{i=1}^m \frac{\sigma_i^2}{n_i} \right)$. \square

Proof of Lemma 1: Define $w := \hat{\psi}^* - \psi^*$ and $\delta := \psi^* - \mathbb{E}_*[\psi^*]$. Unbiasedness is well known, see e.g. Searle et al. (2009). The variance of $\hat{\sigma}_{SVB}^2/\theta$ has been derived in Sun et al. (2011) as

$$\begin{aligned} \frac{1}{\theta^2} \text{Var}_*[\hat{\sigma}_{SVB}^2] &= \frac{1}{B} (\mathbb{E}_*[\delta^4] - (\mathbb{E}_*[\delta^2])^2) + \frac{2}{B(B-1)} (\mathbb{E}_*[\delta^2])^2 + \frac{2}{B^2 R^2 (B-1)} (\mathbb{E}_*[w^2])^2 \\ &\quad + \frac{2(B+1)}{B^2 R (B-1)} \mathbb{E}_*[\delta^2] \mathbb{E}_*[w^2] + \frac{2}{B^2 R^3} \mathbb{E}_*[w^4] + \frac{4B+2}{B^2 R} \mathbb{E}_*[\delta^2 w^2] \\ &\quad + \frac{2(BR^2 + R^2 - 4R + 3)}{B^2 R^3 (R-1)} \mathbb{E}_*[(\mathbb{E}[w^2 | \hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*])^2] + \frac{4}{B^2 R^2} \mathbb{E}_*[\delta w^3]. \end{aligned}$$

Applying Jensen's inequality (or generalized Holder's inequality) gives

$$\begin{aligned} \mathbb{E}_*[(\mathbb{E}[w^2 | \hat{F}_1^*, \dots, \hat{F}_m^*])^2] &\leq \mathbb{E}_*[w^4] \\ \mathbb{E}_*[\delta^2 w^2] &\leq (\mathbb{E}_*[\delta^4] \mathbb{E}_*[w^4])^{1/2}, \quad |\mathbb{E}_*[\delta w^3]| \leq (\mathbb{E}_*[\delta^4] (\mathbb{E}_*[w^4])^3)^{1/4} \end{aligned}$$

The convergence condition $\mathbb{E}[(\hat{g}_i - g_i)^4(X_{i,1})] \rightarrow 0$ implies that $\frac{1}{n_i} \sum_{j=1}^{n_i} \hat{g}_i^4(X_{i,j}) = \frac{1}{n_i} \sum_{j=1}^{n_i} g_i^4(X_{i,j}) + o_p(1) = O_p(1)$. Together with the moment condition $\mathbb{E}_*[(\epsilon^* - \mathbb{E}_*[\epsilon^*])^4] = o_p(s^{-2})$, we get

$$\mathbb{E}_*[\delta^4] = 3 \left(\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{s_i} \right)^2 + o_p \left(\left(\sum_{i=1}^m \frac{1}{s_i} \right)^2 \right), \quad \mathbb{E}_*[\delta^2] = \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{s_i} + o_p \left(\sum_{i=1}^m \frac{1}{s_i} \right), \quad \mathbb{E}_*[w^4] = \mathbb{E}_*[\mu_4^*] = O_p(1).$$

Hence the leading terms of the mean squared error can be identified as

$$\begin{aligned} \frac{1}{\theta^2} \text{Var}_*[\hat{\sigma}_{SVB}^2] &\approx \frac{1}{B} (\mathbb{E}_*[\delta^4] - (\mathbb{E}_*[\delta^2])^2) + \frac{4}{BR} \mathbb{E}_*[\delta^2 w^2] + \frac{2}{BR^2} \mathbb{E}_*[(\mathbb{E}[w^2 | \hat{F}_{s_1,1}^*, \dots, \hat{F}_{s_m,m}^*])^2] \\ &\approx \frac{2}{B} \left(\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{s_i} \right)^2 + \frac{4\hat{\tau}^2}{BR} \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{s_i} + \frac{2\hat{\tau}^4}{BR^2} = \frac{2}{B} \left(\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{s_i} + \frac{\hat{\tau}^2}{R} \right)^2. \end{aligned}$$

Here $a \approx b$ means $a/b \xrightarrow{p} 1$ as aforementioned. Therefore the variance can be expressed as

$$\text{Var}_*[\hat{\sigma}_{SVB}^2] = \frac{2}{B} \left(\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{n_i} + \frac{\hat{\tau}^2 \theta}{R} \right)^2 (1 + o_p(1)) = \frac{2}{B} \left(\sum_{i=1}^m \frac{\sigma_i^2}{n_i} + \frac{\tau^2 \theta}{R} \right)^2 (1 + o_p(1))$$

where the second equality holds because of the convergence of $\hat{\sigma}_i^2, \hat{\tau}^2$ to σ_i^2, τ^2 . \square

Proof of Theorem 2. Under the choice of B, R, θ stated in the theorem, we have $\text{Var}_*[\hat{\sigma}_{SVB}^2] = o_p(1/n^2)$ hence $\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2 = o_p(1/n)$ on one hand. On the other hand we know the subsampling bootstrap variance estimate $\hat{\sigma}_{SVB}^2$ is consistent for σ_I^2 and $\sigma_I^2 = \Theta(1/n)$ hence $\sigma_{SVB}^2 - \sigma_I^2 = o_p(1/n)$. Now $\hat{\sigma}_{SVB}^2 - \sigma_I^2 = \hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2 + \sigma_{SVB}^2 - \sigma_I^2 = o_p(1/n)$ from which consistency immediately follows. \square

Proof of Theorem 8. One can easily verify that such B^* and R^* minimize the mean squared error (18) under the constraint that $BR = N$ and $B = \omega(1)$. The mean square error (19) then follows from evaluating (18) at B^*, R^* . \square

Proof of Corollary 1. It is obvious that when $B = \omega(1)$ and $R = \Omega(\theta n)$ the configuration (7) is satisfied hence the estimate $\hat{\sigma}_{SVB}^2$ is relatively consistent under such allocation. To show that a simulation budget $N = \omega(\theta n)$ is necessary for (7) to hold, note that multiplying the first two requirements in (7) gives that $B^2 R^2 = \omega((\theta n)^2)$, hence $BR = \omega(\theta n)$ must hold true. \square

Proof of Corollary 2. This follows from letting $\theta = \omega(1/n)$ in Corollary 1 so that the required simulation budget $N = \omega(\theta n) = \omega(\omega(1)) = \omega(1)$. \square

Proof of Theorem 3. The requirement $\omega(1/n) \leq \theta$ is stipulated by (7). If $\theta \leq o(N/n) \wedge 1$, then we have $\theta n = o(N)$, or equivalently $N/(\theta n) = \omega(1)$, so that we can afford a $B = \omega(1)$ when $R = \Omega(\theta n)$ to satisfy the first two requirements of (7). Theorem 2 then guarantees consistent variance estimation. \square

Proof of Theorem 4. It follows from Theorem 8 by observing that $\tau^2 = \Theta(1)$ and $\sum_{i=1}^m \sigma_i^2/n_i = \Theta(1/n)$. \square

EC.4. Proofs for Results in Section 4.3 and Theorem 5

In this section we analyze the statistical error of $\hat{\sigma}_{SVB}^2$ relative to the true input variance σ_I^2 , therefore, combined with the Monte Carlo error $\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2$ given in Lemma 1, provide the overall error of the estimate $\hat{\sigma}_{SVB}^2$, and then minimize the overall error to obtain the optimal choices for the parameters θ, B, R . We first prove Lemma 2 using Lemma EC.1 and Proposition 2 which have been presented in Section EC.2, then use Lemma 2 to conclude Theorem 10. Lastly, Theorem 5 is derived from Theorem 10.

Proof of Lemma 2. The proof of Theorem 7 derives the following expression for the proportionate subsampled bootstrap variance

$$\frac{\sigma_{SVB}^2}{\theta} = \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{s_i} + \text{Var}_*[\epsilon^*] + 2\text{Cov}_* \left(\sum_{i=1}^m \frac{1}{s_i} \sum_{k=1}^{s_i} \hat{g}_i(X_{i,k}^*, \epsilon^*) \right).$$

As is the case in the proof of Proposition 2, the covariances can be simplified to

$$\text{Cov}_*(\sum_{i=1}^m \frac{1}{s_i} \sum_{k=1}^{s_i} \hat{g}_i(X_{i,k}^*), \epsilon^*) = \sum_{i=1}^m \mathbb{E}_*[\hat{g}_i(X_{i,1}^*) (\mathbb{E}_*[\epsilon^*|X_{i,1}^*] - \mathbb{E}_*[\epsilon^*])].$$

This leads to

$$\sigma_{SVB}^2 = \sum_{i=1}^m \frac{\theta \hat{\sigma}_i^2}{[\theta n_i]} + \theta \mathbb{E}_*[(\epsilon^* - \mathbb{E}_*[\epsilon^*])^2] + 2\theta \sum_{i=1}^m \mathbb{E}_*[\hat{g}_i(X_{i,1}^*) (\mathbb{E}_*[\epsilon^*|X_{i,1}^*] - \mathbb{E}_*[\epsilon^*])].$$

From the above expression of the variance estimator one can verify that it suffices to show the following three results

$$\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{n_i} = \sigma_I^2 + \mathcal{Z} + o_p\left(\frac{1}{n^{3/2}}\right) \quad (\text{EC.19})$$

$$\mathbb{E}_*[(\epsilon^* - \mathbb{E}_*[\epsilon^*])^2] = \sum_{i,i'=1}^m \frac{1}{4s_i s_{i'}} \text{Var}[g_{ii'}(X_i, X'_{i'})] + o_p\left(\frac{1}{s^2}\right) \quad (\text{EC.20})$$

$$\begin{aligned} & \mathbb{E}_*[\hat{g}_i(X_{i,1}^*) (\mathbb{E}_*[\epsilon^*|X_{i,1}^*] - \mathbb{E}_*[\epsilon^*])] \\ &= \frac{1}{2s_i^2} \text{Cov}(g_i(X_i), g_{ii}(X_i, X_i)) + \sum_{i'=1}^m \frac{1}{2s_i s_{i'}} \text{Cov}(g_i(X_i), \mathbb{E}_{X'_{i'}}[g_{ii'i'}(X_i, X'_{i'}, X'_{i'})]) + o_p\left(\frac{1}{s^2}\right). \end{aligned} \quad (\text{EC.21})$$

To see this, if the three equations hold then

$$\begin{aligned} \sigma_{SVB}^2 &= \sum_{i=1}^m \frac{\theta \hat{\sigma}_i^2}{\theta n_i - \text{frac}(\theta n_i)} + \sum_{i,i'=1}^m \frac{1}{4n_i s_{i'}} \text{Var}[g_{ii'}(X_i, X'_{i'})] + o_p\left(\frac{\theta}{s^2}\right) \\ &\quad + \sum_{i=1}^m \frac{1}{n_i s_i} \text{Cov}(g_i(X_i), g_{ii}(X_i, X_i)) + \sum_{i,i'=1}^m \frac{1}{n_i s_{i'}} \text{Cov}(g_i(X_i), \mathbb{E}_{X'_{i'}}[g_{ii'i'}(X_i, X'_{i'}, X'_{i'})]) + o_p\left(\frac{\theta}{s^2}\right) \\ &= \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{n_i} + \sum_{i=1}^m \frac{\text{frac}(\theta n_i) \sigma_i^2}{n_i s_i} + o_p\left(\frac{1}{ns}\right) + \sum_{i,i'=1}^m \frac{1}{4n_i s_{i'}} \text{Var}[g_{ii'}(X_i, X'_{i'})] \\ &\quad + \sum_{i=1}^m \frac{1}{n_i s_i} \text{Cov}(g_i(X_i), g_{ii}(X_i, X_i)) + \sum_{i,i'=1}^m \frac{1}{n_i s_{i'}} \text{Cov}(g_i(X_i), \mathbb{E}_{X'_{i'}}[g_{ii'i'}(X_i, X'_{i'}, X'_{i'})]) + o_p\left(\frac{1}{ns}\right) \\ &= \sum_{i=1}^m \frac{\hat{\sigma}_i^2}{n_i} + \mathcal{R} + o_p\left(\frac{1}{ns}\right) \\ &= \sigma_I^2 + \mathcal{Z} + \mathcal{R} + o_p\left(\frac{1}{ns}\right) + o_p\left(\frac{1}{n^{3/2}}\right) \end{aligned}$$

where (EC.20) and (EC.21) are used in the first equality and (EC.19) used in the last equality.

Now we prove the above three equations (EC.19)-(EC.21). By the expansion of \hat{g}_i from Assumption 12 and the vanishing moment condition on the remainder ϵ_g , we write

$$\begin{aligned}\hat{\sigma}_i^2 &= \frac{1}{n_i} \sum_{j=1}^{n_i} g_i^2(X_{i,j}) + \frac{2}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j}) \left(\sum_{i'=1}^m \frac{1}{n_{i'}} \sum_{j'=1}^{n_{i'}} g_{ii'}(X_{i,j}, X_{i',j'}) + \frac{1}{n_i} \sum_{j'=1}^{n_i} g_i(X_{i,j'}) \right) \\ &\quad + \frac{1}{n_i} \sum_{j=1}^{n_i} \left(\sum_{i'=1}^m \frac{1}{n_{i'}} \sum_{j'=1}^{n_{i'}} g_{ii'}(X_{i,j}, X_{i',j'}) + \frac{1}{n_i} \sum_{j'=1}^{n_i} g_i(X_{i,j'}) \right)^2 + o_p\left(\frac{1}{\sqrt{n}}\right) \quad (\text{EC.22})\end{aligned}$$

$$\begin{aligned}&= \frac{1}{n_i} \sum_{j=1}^{n_i} g_i^2(X_{i,j}) + \frac{2}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j}) \sum_{i'=1}^m \frac{1}{n_{i'}} \sum_{j'=1}^{n_{i'}} g_{ii'}(X_{i,j}, X_{i',j'}) + 2 \left(\frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j}) \right)^2 + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{1}{n_i} \sum_{j=1}^{n_i} g_i^2(X_{i,j}) + 2 \sum_{i'=1}^m \frac{1}{n_i n_{i'}} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} g_i(X_{i,j}) g_{ii'}(X_{i,j}, X_{i',j'}) + o_p\left(\frac{1}{\sqrt{n}}\right). \quad (\text{EC.23})\end{aligned}$$

Note that the first term in line (EC.22) has an expectation of order $O(1/n)$ hence can be absorbed into the $o_p(1/\sqrt{n})$ term. Similarly the fourth line (EC.23) holds because $(\sum_{j=1}^{n_i} g_i(X_{i,j})/n_i)^2$ has an expectation of order $O(1/n)$. The second term in (EC.23) is a sum of m V-statistics, each of which by standard results is well approximated by the Hajek projection

$$\frac{1}{n_i n_{i'}} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} g_i(X_{i,j}) g_{ii'}(X_{i,j}, X_{i',j'}) = \frac{1}{n_{i'}} \sum_{j'=1}^{n_{i'}} \mathbb{E}_{X_i} [g_i(X_i) g_{ii'}(X_i, X_{i',j'})] + O_p\left(\frac{1}{n}\right).$$

The finite fourth moment condition of g_i and $g_{i_1 i_2}$ are used to ensure that the product $g_i(X_{i,j}) g_{ii'}(X_{i,j}, X_{i',j'})$ has a finite second moment so that the above approximation holds. Denoting

$$\mu_1^i = \frac{1}{n_i} \sum_{j=1}^{n_i} g_i^2(X_{i,j}), \quad \mu_2^{ii'} = \frac{1}{n_{i'}} \sum_{j'=1}^{n_{i'}} \mathbb{E}_{X_i} [g_i(X_i) g_{ii'}(X_i, X_{i',j'})]$$

we have

$$\sum_{i=1}^m \frac{\hat{\sigma}_i^2}{n_i} = \sum_{i=1}^m \frac{\mu_1^i}{n_i} + 2 \sum_{i=1}^m \sum_{i'=1}^m \frac{\mu_2^{ii'}}{n_i} + o_p\left(\frac{1}{n^{3/2}}\right).$$

Because of independence among input models the variance of the leading term takes the additive form $\sum_{i=1}^m \lambda_i^T \Sigma_i \lambda_i / n_i$ as described in the theorem. By Proposition 2 $\sigma_I^2 = \sum_{i=1}^m \sigma_i^2 / n_i + O(n^{-2})$ hence equation (EC.19) follows. To show (EC.20), we note that in the cubic expansion of Assumption 12 the cubic term and the remainder ϵ_3^* both have a second moment of order $O_p(s^{-3})$. Therefore

it suffices to consider the quadratic term. Since the second order influence function $\hat{g}_{i_1 i_2}$ has vanishing marginal expected value, one can verify that

$$\begin{aligned} \text{Var}_* \left[\sum_{i,i'=1}^m \frac{1}{s_i s_{i'}} \sum_{j=1}^{s_i} \sum_{j'=1}^{s_{i'}} \hat{g}_{ii'}(X_{i,j}^*, X_{i',j'}) \right] &= \sum_{i,i'=1}^m \frac{1}{s_i s_{i'} n_i n_{i'}} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} \hat{g}_{ii'}^2(X_{i,j}, X_{i',j'}) + O_p\left(\frac{1}{s^3}\right) \\ &= \sum_{i,i'=1}^m \frac{1}{s_i s_{i'} n_i n_{i'}} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} g_{ii'}^2(X_{i,j}, X_{i',j'}) + o_p\left(\frac{1}{s^2}\right) \end{aligned}$$

where the second equality follows from the convergence of $\hat{g}_{ii'}$ to $g_{ii'}$ as imposed in Assumption 12. Equation (EC.20) then follows from consistency of the V-statistic $\sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} g_{ii'}^2(X_{i,j}, X_{i',j'}) / (n_i n_{i'})$.

Let's continue to prove equation (EC.21). Denote by X_i^* a generic resampled data point from the i -th input data set. Then one can check that

$$\begin{aligned} &\mathbb{E}_*[\epsilon^* | X_{i,1}^* = X_{i,j}] - \mathbb{E}_*[\epsilon^*] \\ &= \frac{1}{2s_i^2} (\hat{g}_{ii}(X_{i,j}, X_{i,j}) - \mathbb{E}_*[\hat{g}_{ii}(X_i^*, X_i^*)]) + \frac{1}{6s_i^3} (\hat{g}_{iii}(X_{i,j}, X_{i,j}, X_{i,j}) - \mathbb{E}_*[\hat{g}_{iii}(X_i^*, X_i^*, X_i^*)]) \\ &\quad + \frac{s_i - 1}{2s_i^3} \mathbb{E}_*[\hat{g}_{iii}(X_{i,j}, X_i^*, X_i^*)] + \sum_{i' \neq i} \frac{1}{2s_i s_{i'}} \mathbb{E}_*[\hat{g}_{ii'i'}(X_{i,j}, X_{i'}, X_{i'})] + \mathbb{E}_*[\epsilon_3^* | X_{i,1}^* = X_{i,j}] - \mathbb{E}_*[\epsilon_3^*]. \end{aligned}$$

Note that $\text{Var}_*[\epsilon_3^* | X_{i,1}^*] = o_p(s^{-4})$ because of Assumption 12 and Lemma EC.2. Hence

$$\begin{aligned} &\mathbb{E}_*[\hat{g}_i(X_{i,1}^*) (\mathbb{E}_*[\epsilon^* | X_{i,1}^*] - \mathbb{E}_*[\epsilon^*])] \\ &= \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{g}_i(X_{i,j}) \left(\frac{1}{2s_i^2} (\hat{g}_{ii}(X_{i,j}, X_{i,j}) - \mathbb{E}_*[\hat{g}_{ii}(X_i^*, X_i^*)]) + \sum_{i'=1}^m \frac{1}{2s_i s_{i'}} \mathbb{E}_*[\hat{g}_{ii'i'}(X_{i,j}, X_{i'}, X_{i'})] \right) + o_p\left(\frac{1}{s^2}\right) \\ &= \frac{1}{2s_i^2} \text{Cov}_*(\hat{g}_i(X_i^*), \hat{g}_{ii}(X_i^*, X_i^*)) + \sum_{i'=1}^m \frac{1}{2s_i s_{i'}} \text{Cov}_*(\hat{g}_i(X_i^*), \mathbb{E}_{X_{i'}^*}[\hat{g}_{ii'i'}(X_i^*, X_{i'}, X_{i'})]) + o_p\left(\frac{1}{s^2}\right) \\ &= \frac{1}{2s_i^2} \text{Cov}(g_i(X_i), g_{ii}(X_i, X_i)) + \sum_{i'=1}^m \frac{1}{2s_i s_{i'}} \text{Cov}(g_i(X_i), \mathbb{E}_{X_{i'}^*}[g_{ii'i'}(X_i, X_{i'}, X_{i'})]) + o_p\left(\frac{1}{s^2}\right) \end{aligned}$$

where the $o_p(1/s^2)$ term in the first equality comes from applying Cauchy Schwartz inequality, and the last equality holds since convergence of $\hat{g}_i, \hat{g}_{i_1 i_2}, \hat{g}_{i_1 i_2 i_3}$ to $g_i, g_{i_1 i_2}, g_{i_1 i_2 i_3}$ in mean squared error implies

$$\begin{aligned} \text{Cov}_*(\hat{g}_i(X_i^*), \hat{g}_{ii}(X_i^*, X_i^*)) &\xrightarrow{p} \text{Cov}(g_i(X_i), g_{ii}(X_i, X_i)) \\ \text{Cov}_*(\hat{g}_i(X_i^*), \mathbb{E}_{X_{i'}^*}[\hat{g}_{ii'i'}(X_i^*, X_{i'}, X_{i'})]) &\xrightarrow{p} \text{Cov}(g_i(X_i), \mathbb{E}_{X_{i'}^*}[g_{ii'i'}(X_i, X_{i'}, X_{i'})]). \end{aligned}$$

This gives rise to the equation (EC.21). \square

Proof of Theorems 5 and 10. We first show Theorem 10. Under a given subsampling ratio θ , we know from Lemma 2 and Theorem 4 that under the optimal allocation $B^* = N/R^*$ and $R^* = \Theta(\theta n)$

$$\begin{aligned}\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2 &= \mathcal{E}_1 + o_p\left(\sqrt{\frac{\theta}{Nn}}\right) \\ \sigma_{SVB}^2 - \sigma_I^2 &= \mathcal{E}_2 + o_p\left(\frac{1}{n^{3/2}} + \frac{1}{\theta n^2}\right)\end{aligned}$$

where the errors $\mathcal{E}_1, \mathcal{E}_2$ satisfy $\mathbb{E}_*[\mathcal{E}_1] = 0, \mathbb{E}[\mathcal{E}_1^2] = \Theta(\theta/(Nn))$ and $\mathbb{E}[\mathcal{E}_2^2] = \mathcal{R}^2 + \sum_{i=1}^m \lambda_i^T \Sigma_i \lambda_i / n_i$.

Letting $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$, we have $\mathbb{E}[\mathcal{E}^2] = \mathbb{E}[\mathcal{E}_1^2] + \mathbb{E}[\mathcal{E}_2^2]$ because

$$\mathbb{E}[\mathcal{E}_1 \mathcal{E}_2] = \mathbb{E}_{\text{data}}[\mathbb{E}_*[\mathcal{E}_1 \mathcal{E}_2]] = \mathbb{E}_{\text{data}}[\mathcal{E}_2 \mathbb{E}_*[\mathcal{E}_1]] = 0.$$

This gives Theorem 10.

To prove Theorem 5, note that if $\mathcal{R} = \Theta((ns)^{-1})$, and at least one of the Σ_i 's are positive definite, then $\sum_{i=1}^m \lambda_i^T \Sigma_i \lambda_i / n_i = \Theta(1/n^3)$ hence $\mathbb{E}[\mathcal{E}_2^2] = \Theta(1/n^3 + 1/(\theta^2 n^4))$. We have

$$\hat{\sigma}_{SVB}^2 - \sigma_I^2 = \mathcal{E} + o_p\left(\sqrt{\frac{\theta}{Nn}} + \frac{1}{n^{3/2}} + \frac{1}{\theta n^2}\right)$$

where $\mathbb{E}[\mathcal{E}^2] = \Theta(\theta/(Nn) + 1/n^3 + 1/(\theta^2 n^4))$. To minimize the leading term \mathcal{E} , just note that $\theta/(Nn) + 1/(\theta^2 n^4)$ is minimized at $\theta^* = (2N)^{1/3}/n$ resulting in $\mathbb{E}[\mathcal{E}^2] = \Theta(1/(N^{2/3} n^2) + 1/n^3)$. When $N > n^{3/2}$, we have $1/(N^{2/3} n^2) < 1/n^3$, hence as long as θ^* is chosen such that $\theta^*/(Nn) \leq 1/n^3$ and $1/(\theta^{*2} n^4) \leq 1/n^3$, or equivalently $1/\sqrt{n} \leq \theta^* \leq N/n^2 \wedge 1$, then the error $\mathbb{E}[\mathcal{E}^2] = \Theta(1/n^3)$.

This leads to the optimal subsample size (10). If the depicted conditions do not hold, we have $\mathbb{E}[\mathcal{E}_2^2] \leq \Theta(1/n^3 + 1/(\theta^2 n^4))$ in general, hence all upper bounds we just obtained for $\mathbb{E}[\mathcal{E}^2]$ could be loose in order, leading to (12). \square

EC.5. Computer Network Configuration Details

	channel capacity (bits)	transmission speed (miles per second)
system #1	275000	200000
system #2	200000	125000
system #3	150000	100000
system #4	200000	125000

Table EC.1 Channel capacity and transmission speed configurations.

node i \ node j	1	2	3	4	
node i	1	n.a.	50	40	45
node j	1	60	n.a.	55	25
node i	2	70	25	n.a.	30
node j	2	35	40	50	n.a.

Table EC.2 True arrival rates $\lambda_{i,j}$ for system #1.

node i \ node j	1	2	3	4	
node i	1	n.a.	35	25	30
node j	1	45	n.a.	40	10
node i	2	55	10	n.a.	15
node j	2	20	25	35	n.a.

Table EC.3 True arrival rates $\lambda_{i,j}$ for system #2.

node i \ node j	1	2	3	4	
node i	1	n.a.	20	50	15
node j	1	30	n.a.	25	35
node i	2	40	35	n.a.	40
node j	2	45	50	20	n.a.

Table EC.4 True arrival rates $\lambda_{i,j}$ for system #3.

		node <i>j</i>	1	2	3	4
		node <i>i</i>	1	2	3	4
1		n.a.	70	60	20	
2		25	n.a.	25	30	
3		80	10	n.a.	10	
4		50	60	20	n.a.	

Table EC.5 True arrival rates $\lambda_{i,j}$ for system #4.