# n° 2019-07

# On the construction of confidence intervals for ratios of expectations

# A.DERUMIGNY[1]
# L. GIRARD[2]
# Y.GUYONVARCH[3]

[1] CREST. E-mail: alexis.derumigny@ensae.fr

[2] CREST. E-mail : lucas.girard@ensae.fr

[3] CREST. E-mail : yannick.guyonvarch@ensae.fr

# On the construction of confidence intervals for ratios of expectations[*]

Alexis Derumigny[†], Lucas Girard[†], Yannick Guyonvarch[†]

April 15, 2019

### Abstract

In econometrics, many parameters of interest can be written as ratios of expectations. The main approach to construct confidence intervals for such parameters is the delta method. However, this asymptotic procedure yields intervals that may not be relevant for small sample sizes or, more generally, in a sequence-of-model framework that allows the expectation in the denominator to decrease to 0 with the sample size. In this setting, we prove a generalization of the delta method for ratios of expectations and the consistency of the nonparametric percentile bootstrap. We also investigate finite-sample inference and show a partial impossibility result: nonasymptotic uniform confidence intervals can be built for ratios of expectations but not at every level. Based on this, we propose an easy-to-compute index to appraise the reliability of the intervals based on the delta method. Simulations and an application illustrate our results and the practical usefulness of our rule of thumb.

**Keywords:** delta method, confidence regions, uniformly valid inference, sequence of models, nonparametric percentile bootstrap.
**JEL:** C18, C19.

# 1 Introduction

In applied econometrics, the prevalent method for constructing confidence intervals (CIs) is asymptotic: the theoretical guarantees for most CIs used in practice hold only when the number of observations tends to infinity. For a large class of parameters, the construction of asymptotic CIs also relies on the delta method. In this paper, we focus on parameters that can be expressed as ratios of expectations for which the delta method is a standard procedure to conduct inference. The objective is twofold: study the behavior of the delta method and other confidence intervals in some difficult settings and provide tools to detect cases in which the delta method may behave poorly.

Many popular parameters in economics take the form of ratios of expectations. Typical examples are conditional expectations since any conditional expectation with a discrete conditioning variable, or a conditioning event, can be written as a ratio of unconditional expectations. For instance, assume that we observe an independent and identically distributed (i.i.d.) sample of individuals indexed by $i \in \{1, \ldots, n\}$ with $W_i$ the wage of an individual and $D_i$ an indicator equal to 1 whenever individual $i$ belongs to some treatment group, say a training program; 0 otherwise. Suppose you are interested in the average wage of participants in the program. We have $\mathbb{E}[W \mid D = 1] = \mathbb{E}[WD] / \mathbb{E}[D]$ as $D$ is binary.

Most confidence intervals used in practice are based on asymptotic justifications, hence possible concerns as regards their finite-sample reliability. For ratios of expectations, we document this issue on simulations (see Section 3.1). One of our findings is that the coverage of the CIs based on the delta method happens to be far below their nominal level, even for large sample sizes, when the expectation in the denominator is close to 0.[1] For some scenarios, these asymptotic CIs require above 100,000 observations to get reasonably close to their nominal level. Yet, denominators close to 0 are not unusual in practice. Coming back to the treatment/wage example, a small denominator would correspond to a binary treatment with a low participation rate.

In order to deal with that issue, we consider sequences of models, namely we authorize the distribution of the observations to change with the sample size. This framework enables to formalize in an asymptotic way the idea of a denominator close to 0. Indeed, in a standard asymptotic viewpoint, with the expectation

---

[1] The definitions of coverage and other fundamental properties of confidence intervals are recalled in Appendix A with the conventions that we use.

in the denominator different from 0, all parameters are fixed and well-defined. Hence, $n$ always grows large enough so that empirical means are close to their expectations and the CIs based on the delta method are valid. In other words, the signal that we want to estimate is constant while the noise goes to 0, and therefore the problem vanishes in this asymptotic perspective. We would like to model more difficult cases, in which the signal can go to 0 as well. This is precisely what the sequence-of-model set-up allows.[2] This is similar to some frameworks that have been developed for weak instrumental variables (IV), see notably Staiger and Stock (1997); Stock and Yogo (2005); Andrews et al. (2019).

In this literature, another approach does not consider sequences of models but designs "robust" procedures that allow to be exactly in the problematic case, namely a null covariance between the instrument and the endogenous regressor (see Anderson and Rubin (1949)). In this case, the parameter of interest is unidentified. In contrast with the weak IV framework, it is worth noting that for ratios in general the parameter of interest is not even defined when the denominator is exactly equal to 0. As a consequence, such an approach seems difficult to extend to our problem.

In our setting, it is unclear, even asymptotically, what the properties of the CIs based on the delta method are when the expectation in the denominator tends to 0. We show that usual CIs can fail and the limiting law of $\widehat{\theta}_n - \theta_n$ may not be Gaussian anymore, denoting by $\theta_n$ the ratio of expectations and $\widehat{\theta}_n$ its empirical counterpart. In some cases, the difference $\widehat{\theta}_n - \theta_n$ may actually have a Cauchy limit, as can be found in the weak IV literature.

We show in this sequence-of-model framework that confidence intervals provided by the nonparametric percentile bootstrap have the same asymptotic properties as the ones obtained with the delta method. Simulations support that claim and even suggest the former have better coverage than the latter in finite samples.

Even in standard settings with a fixed but small denominator, simulations document that asymptotic-based CIs may require very large sample sizes to attain their nominal level. This suggests to study more in details nonasymptotic inference. More precisely, we construct finite-sample CIs, extending old-established concentration inequalities for means to ratios of means. Concentration inequalities for the mean refer to upper bounds on the probability that an empirical mean departs from its expectation more than a given threshold. Such inequalities per-

---

[2]This can also rationalize the practice of applied social researchers (see Example 2.1). The heuristic idea is that researchers can consider narrower effects as the data gets richer.

mit to construct confidence intervals valid for any sample size and for large classes of probability distributions (see in particular Boucheron et al. (2013)). To our knowledge, there is no such result for ratios. We consider distributions within a class characterized by a lower bound on the first moment for the denominator variable, and an upper bound on the second moment for both the numerator and denominator variables.[3]

One additional result highlights there exists a critical confidence level, above which it is not possible to construct nonasymptotic CIs, uniformly valid on such classes, and that are almost surely bounded under every distribution of those classes. More precisely, we exhibit explicit upper and lower bounds on this critical confidence level: the former is a threshold above which we show it is impossible to construct such CIs; the latter is a threshold below which we show how to construct them.

These ideas closely relate to some impossibility results as regards the construction of confidence intervals. A large share of the research effort has concentrated on the problem of constructing confidence intervals for expectations. In an early contribution, Bahadur and Savage (1956) show that, when $\mathcal{P}$ is the set of all distributions on the real line with finite expectation, the parameter of interest $\theta(P)$ is the expectation with respect to a distribution $P \in \mathcal{P}$ and $\Theta = \mathbb{R}$, a confidence interval built from an i.i.d. sample of $n \in \mathbb{N}^*$ observations that has uniform coverage $1 - \alpha$ over $\mathcal{P}$ must contain any real number with probability at least $1 - \alpha$. Broadly speaking, any confidence interval must have infinite length with positive probability for every $P \in \mathcal{P}$ to ensure a coverage of $1 - \alpha$.

Stronger results can be derived when one further restricts $\mathcal{P}$ or $\Theta$. When $\mathcal{P}$ is taken to be the set of all distributions on the real line with variance uniformly bounded by a finite constant, it is possible to show (using the Bienaymé-Chebyshev inequality) that for every $n \in \mathbb{N}^*$ and every $\alpha \in (0, 1)$, there exists a confidence interval that is almost surely bounded under every $P \in \mathcal{P}$ and has coverage $1 - \alpha$. In this case, the obtained CIs have the advantage that their length shrinks to $0$ at the optimal rate $1/\sqrt{n}$. But on the downside, they are not of size $1 - \alpha$, even asymptotically, except for some extreme distributions. This means that they tend to be conservative in practice.

A strand of the literature has also investigated more complex problems in which $\theta(P)$ is not restricted to being an expectation. For general parameters, Dufour

---

[3]We refer to this setting as the "Bienaymé-Chebyshev" (BC) case. In Appendix C, we present similar results for distributions whose supports are bounded ("Hoeffding" case).

(1997) derives a generalization of Bahadur and Savage (1956). An implication of the results in Dufour (1997) is the existence of an impossibility theorem for ratios of expectations. Let $P$ be a distribution on $\mathbb{R}^2$ with marginals $P_X$ and $P_Y$. If $\theta(P) = \mathbb{E}_{P_X}[X] / \mathbb{E}_{P_Y}[Y]$, then for every $\alpha \in (0,1)$, it is impossible to build nontrivial CIs of coverage $1 - \alpha$ when $\mathcal{P}$ is the set of all distributions on $\mathbb{R}^2$ with finite second moments and $\Theta = \{\theta = \mathbb{E}_{P_X}[X] / \mathbb{E}_{P_Y}[Y] : (\mathbb{E}_{P_X}[X], \mathbb{E}_{P_Y}[Y]) \in \mathbb{R} \times \mathbb{R}^*\}$. As will be explained below, this impossibility result disappears as soon as $\mathcal{P}$ is chosen such that $|\mathbb{E}_{P_Y}[Y]|$ is bounded away from 0 uniformly over $\mathcal{P}$. Interestingly, the impossibility breaks down only partly in the sense that there remains an upper bound on confidence levels (that depends on $n$) above which it is impossible to build nontrivial CIs.

Other interesting results can be found in Romano and Wolf (2000) and Pinelis and Molzon (2016). Romano and Wolf (2000) construct nonasymptotic valid confidence intervals that happen to be also asymptotically optimal. However, they only consider expectations. Pinelis and Molzon (2016) study smooth functions of a vector of means and give bounds on the distance between the distribution of the normalized and centered estimator and its Gaussian limiting distribution. Nonetheless, the authors do not link their results to the construction of confidence intervals.

In the light of that existing literature, our nonasymptotic findings can be interpreted as a partial impossibility result. Indeed, even if we assume a known positive lower bound on the expectation in the denominator, the limitation on the attainable coverage of our nonasymptotic CIs remains. That point complements Dufour (1997): for a given sample size $n$, interesting CIs can be built but not at every confidence level. By contrast, provided the expectation in the denominator is not null, the delta method gives CIs at every confidence level, but their coverage is only asymptotic.

To bridge this gap, we suggest a rule of thumb to assess the reliability of the delta method for ratios of expectations in finite samples. The heuristic idea is simply, for a given sample, to compute an estimator of the lower bound on the above-mentioned critical confidence level. This lower bound can be seen as a conservative value for the unknown critical level, which is a necessary criterion to conduct valid inference in finite samples uniformly over a given class of distributions. Hence, for any desired level higher than this bound, the CIs based on the delta method cannot reach this desired uniform level in finite samples. We illustrate the empirical usefulness of that rule of thumb on simulations and with an application to gender

wage disparities in France for the years 2010-2017.

The rest of the paper is organized as follows. Section 2 details our framework and assumptions. In Section 3, we illustrate the weaknesses of the CIs based on the delta method with a denominator "close to 0" on simulations and detail the asymptotic behavior of the delta method and of the nonparametric percentile bootstrap in our sequence-of-model setting. Section 4 is devoted to the construction of nonasymptotic confidence intervals and presents a lower bound on the afore-mentioned critical confidence level. In Section 5, we derive an upper bound on the critical confidence level as well as a lower bound on the length of nonasymptotic CIs. This section also includes the description of a practical index to gauge the soundness of the CIs based on the delta method in finite samples. Section 6 present simulations and an application to a real dataset to illustrate our methods. Section 7 concludes. General definitions about confidence intervals are recalled in Appendix A. The proofs of all results are postponed to Appendix B. Additional results under an alternative set of assumptions ("Hoeffding" case) are detailed in Appendix C. Appendix D presents supplementary simulations.

## 2 Our framework

Throughout the paper, for any random variable $U$ and $n$ i.i.d. replications $(U_{1,n}, \ldots, U_{n,n})$, we denote by $\overline{U}_n$ the empirical mean of $U$, that is $n^{-1} \sum_{i=1}^{n} U_{i,n}$. Assumption 1 defines our sequence-of-model framework and provides the basic requirements to state our asymptotic results.

**Assumption 1.** *For every $n \in \mathbb{N}^*$, we observe a sample $(X_{i,n}, Y_{i,n})_{i=1,\ldots,n} \overset{i.i.d.}{\sim} P_{X,Y,n}$, where $P_{X,Y,n}$ is a given distribution on $\mathbb{R}^2$ that satisfies $\mathbb{E}[Y_{1,n}] > 0$, $\mathbb{E}[X_{1,n}^2] < +\infty$, and $\mathbb{E}[Y_{1,n}^2] < +\infty$.*

Remark that $n$ indexes both the distribution $P_{X,Y,n}$ of the observations in this model and the number of observations $n$. This encompasses the standard i.i.d. set-up if the distribution does not change with $n$: for every $n \in \mathbb{N}^*$, $P_{X,Y,n} = P_{X,Y}$ for some given distribution $P_{X,Y}$. As we assume the existence of a finite expectation, we can consider $\mathbb{E}[Y_{1,n}] \geq 0$ without loss of generality.[4] In order to have properly defined ratios of interest, we need to assume away a null denominator, namely suppose that for every $n \in \mathbb{N}^*$, $\mathbb{E}[Y_{1,n}] > 0$.

---

[4]Otherwise, we simply replace $Y_{i,n}$ by its opposite $-Y_{i,n}$.

**Example 2.1** (Sequences of models and the practice of applied researchers).
*Researcher may look at the average value of a variable $A_{i,n}$ of interest in a subgroup of the data. Subgroups could be defined as the intersections of, say, time, geographical area, gender, age, income brackets and so on. As the number of observations $n$ grows, it is possible to consider subgroups $g_n$ that become thinner and thinner (intersection of more and more variables for instance). This practice could be modelled as estimating $\theta_n := \mathbb{E}[A_{i,n} \mid G_{i,n} = 1] = \mathbb{E}[A_{i,n}G_{i,n}] / \mathbb{P}(G_{i,n} = 1)$ where $G_{i,n}$ is a binary variable that is equal to 1 if an individual $i$ belongs to the subgroup $g_n$. This corresponds to our framework denoting $X_{i,n} := A_{i,n} \times G_{i,n}$ and $Y_{i,n} := G_{i,n}$.*

To derive our nonasymptotic results, Assumption 1 has to be strengthened.

**Assumption 2.** *For every $n \in \mathbb{N}^*$, there exist positive finite constants $l_{Y,n}$, $u_{X,n}$, and $u_{Y,n}$ such that (i) $\mathbb{E}[Y_{1,n}] \geq l_{Y,n} > 0$, (ii) $\mathbb{E}[X_{1,n}^2] \leq u_{X,n}$ and $\mathbb{E}[Y_{1,n}^2] \leq u_{Y,n}$.*

Note that in practice, the value of the constants $l_{Y,n}$, $u_{X,n}$, and $u_{Y,n}$ may not be available for practitioners. This is the reason why, in Section 5.3, we propose heuristic methods that palliate the lack of knowledge of those constants.

The first part of the assumption bounds the expectation of $Y_{1,n}$ away from 0 while the second states that the second moments of $X_{1,n}$ and $Y_{1,n}$ are bounded. These are necessary to derive nonasymptotic CIs with maintained coverage uniformly over a class of distributions and that are not trivial. Otherwise, if $l_{Y,n} = 0$ or in the absence of the upper bounds $u_{X,n}$ and $u_{Y,n}$, the impossibility theorem of Dufour (1997) applies and prevents from constructing nontrivial CIs for any confidence level. In a way, given this result, Assumption 2 can be seen as close to the minimal hypothesis that allows for the possibility of nontrivial confidence intervals with finite-sample guarantees for ratios of expectations. Furthermore, the sequence-of-model framework allows $l_{Y,n}$ to decrease to 0, which enables us to study limiting cases close to but different from the problematic case $l_{Y,n} = 0$.

This set-up, where Assumptions 1 and 2 hold, is named the *BC case* since it is possible under these assumptions to construct nonasymptotic CIs using the Bienaymé-Chebyshev inequality. In Appendix C, we present an adapted version of our results under the assumption that $X_{1,n}$ and $Y_{1,n}$ have a bounded support instead of bounded second moments; a setting we call the *Hoeffding case*.

To sum up, Assumptions 1 and 2 define a set $\mathcal{P}$ of distributions for some constants $l_{Y,n}$, $u_{X,n}$ and $u_{Y,n}$. For a distribution $P_{X,Y,n}$ in $\mathcal{P}$, the parameter of interest

$\theta(P_{X,Y,n})$ is denoted $\theta_n := \mathbb{E}[X_{1,n}] / \mathbb{E}[Y_{1,n}]$ with values in $\mathbb{R}$. To estimate this parameter, we consider its empirical counterpart $\widehat{\theta}_n := \overline{X}_n / \overline{Y}_n$. We seek to construct confidence intervals $C_{n,\alpha}$ for $\theta_n$ with nominal level $1 - \alpha$ based on this estimator.

In practice, it is possible that $\overline{Y}_n = 0$ and it may even happen with a strictly positive probability for non-continuous distributions of $Y$. The estimator $\widehat{\theta}_n$ does not exist for such samples. In such a case, it is difficult to construct meaningful confidence intervals. Different conventions are possible:

- We could choose to define $C_{n,\alpha} = \mathbb{R}$. This entails that $\theta_n$ belongs to $C_{n,\alpha}$ by construction. We believe that such a choice would artificially improve the coverage of $C_{n,\alpha}$ as it induces that the higher $\mathbb{P}(\overline{Y}_n = 0)$, the better the interval in terms of coverage.

- We could choose $C_{n,\alpha} = \emptyset$. The hypothesis $\theta_n = \theta_0$ would then be rejected for every $\theta_0 \in \mathbb{R}$ using the duality between tests and confidence intervals. We would also like to avoid this situation because it may not be reasonable to always reject for the mere reason that $\theta_n$ cannot be estimated in the sample.

- Other choices are possible, for example $C_{n,\alpha} = \{0\}$, but they do not seem sensible either since there is no reason to select only 0 in our confidence interval, especially if $\overline{X}_n \neq 0$.

For these considerations, we choose to let $C_{n,\alpha}$ *undefined* whenever $\overline{Y}_n = 0$, following the convention that ratios $x/0$ are undefined for any real $x$.[5] In practice, when given a realization $\omega \in \Omega$ and a real $a \in \mathbb{R}$, we either know that $a$ belongs to $C_{n,\alpha}(\omega)$, or we know that $a$ does not belong to $C_{n,\alpha}(\omega)$, or $C_{n,\alpha}(\omega)$ is undefined. As a consequence, we have the decomposition $\Omega = \{\omega : a \in C_{n,\alpha}(\omega)\} \sqcup \{\omega : a \notin C_{n,\alpha}(\omega)\} \sqcup \{\omega : C_{n,\alpha}(\omega) \text{ undefined}\}$, where $\sqcup$ denotes the disjoint union of sets. This means that $\mathbb{P}\{a \in C_{n,\alpha}\} + \mathbb{P}\{a \notin C_{n,\alpha}\} + \mathbb{P}\{C_{n,\alpha} \text{ undefined}\} = 1$.

## 3 Limitations of the delta method: when are asymptotic confidence intervals valid?

In practice, for a sample of size $n$, the coverage of asymptotic CIs may be well below their nominal level $1 - \alpha$. Intuitively, this phenomenon should be

---

[5] When facing $\overline{Y}_n = 0$, applied researchers may use other estimators. For instance, one could consider sub-samples (possibly several and combine them in some way) of the data for which the empirical mean in the denominator differs from 0. Nevertheless, the construction of satisfactory estimators in this case lies beyond the scope of this paper.

driven by "problematic" distributions in $\mathcal{P}$ in the following sense: when the true distribution $P$ is close to the boundary of the class $\mathcal{P}$, the probability $c(n, P) := \mathbb{P}_{P^{\otimes n}} (C_{n,\alpha} \ni \theta(P))$ may be much smaller than $1 - \alpha$.[6]

In Section 3.1, with $C_{n,\alpha}$ the confidence interval based on the delta method, we illustrate on simulations that $c(n, P)$ can fail to match $1 - \alpha$ when the expectation in the denominator is fixed close to 0. In other words, it may require a very large number of observations to make reasonable the asymptotic approximation. In Section 3.2, we investigate a more serious issue: in the sequence-of-model framework, we let the expectation in the denominator not only be small but converge to 0 as $n$ increases. We show on simulations that depending on the speed at which the denominator goes to 0, $c(n, P)$ can either converge to the nominal level (more or less quickly) or even not converge at all to this target. This sheds light on a partial failure of the delta method when the denominator goes to 0 that we derive formally in Section 3.3. Finally, in Section 3.4, we show the asymptotic consistency of the nonparametric percentile bootstrap (also known as Efron's percentile bootstrap) in this sequence-of-model framework.

## 3.1  Asymptotic approximation takes time to hold

In this subsection, we consider the i.i.d. case.[7] Under Assumption 1, asymptotic confidence intervals are easily obtained combining the multivariate central limit theorem (CLT) and the delta method:

$$\sqrt{n} \left( \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X]}{\mathbb{E}[Y]} \right) \xrightarrow[n \to +\infty]{d} \mathcal{N}(0, \Sigma), \tag{1}$$

where $\Sigma = \mathbb{V}[X]/\mathbb{E}[Y]^2 + \mathbb{E}[X]^2 \mathbb{V}[Y]/\mathbb{E}[Y]^4 - 2\,\mathbb{C}ov[X, Y]\,\mathbb{E}[X]/\mathbb{E}[Y]^3$ and in practice is replaced by a consistent estimate (Slutsky's lemma).

To assess the quality of the CI based on (1), we compute its $c(n, P)$ using simulations for different sample sizes $n$ and distributions $P$ and compare it to the nominal level. By definition, the pointwise coverage $c(n, P)$ forms an upper bound on the uniform coverage. In our simulations, we choose the level $1 - \alpha = 95\%$. For different sample sizes $n$ and values of $\mathbb{E}[Y]$, we draw $M = 5,000$ i.i.d. samples of size $n$ following $\mathcal{N}(1, 1) \otimes \mathcal{N}(\mathbb{E}[Y], 1)$. We compute $c(n, P)$ for the interval based

---

[6] Recall that in the nonasymptotic approach, the *coverage* of any given confidence interval $C_{n,\alpha}$ is defined as the infimum of $c(n, P)$ for $P$ ranging over the studied class $\mathcal{P}$ of distributions.

[7] For every $n \in \mathbb{N}^*$, $P_{X,Y,n}$ is identical, hence denoted $P_{X,Y}$. To simplify notations, we also denote by $(X, Y)$ a random vector following $P_{X,Y}$.

on the delta method for every pair $(n, \mathbb{E}[Y])$ using the 5,000 replications. The expectation $\mathbb{E}[Y]$ ranges from 0.01 (the denominator is close to 0) to 0.75 (the denominator is far from 0). Figure 1 sums up the results. For every $n$, it turns out that the closer $\mathbb{E}[Y]$ to 0, the smaller the $c(n, P)$ of the delta method. When $\mathbb{E}[Y] = 0.01$, we observe that $c(n, P)$ gets close to the nominal level only for $n$ above 300,000. Additional simulations indicate that the phenomenon is robust across different choices of the distribution $P_{X,Y}$ (see Section D).
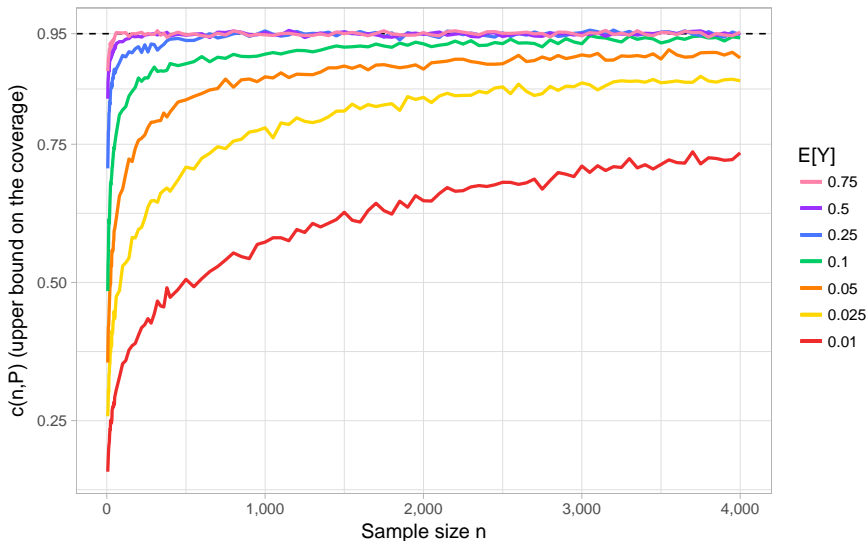


Figure 1: $c(n, P)$ of the asymptotic CIs based on the delta method as a function of the sample size $n$.
Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}(1, 1) \otimes \mathcal{N}(\mathbb{E}[Y], 1)$. The nominal pointwise asymptotic level is set to 0.95. For each pair $(\mathbb{E}[Y], n)$, the coverage is obtained as the mean over 5,000 repetitions.

## 3.2 Asymptotic results may not hold in the sequence-of-model framework

Unlike the result displayed in (1), it is unclear how $\sqrt{n}\left(\overline{X}_n / \overline{Y}_n - \mathbb{E}[X]/\mathbb{E}[Y]\right)$ behaves asymptotically when we consider sequences of models such that the expectation in the denominator tends to 0 as $n$ increases. For a given specification, Figure 2 shows the $c(n, P)$ of the CIs based on the delta method when $\mathbb{E}[Y_{1,n}] = Cn^{-b}$ where $C$ is set to 0.025 and $b$ varies. For a speed $b \geq 1/2$ (i.e. faster than the usual rate of the CLT), the pointwise coverage $c(n, P)$ of the asymptotic CIs obtained by (1) is not good in the sense that it is far lower than the nominal level $1 - \alpha$ and it does not converge to the latter. Our simulations even suggest that the coverage tends to 0 for $b > 1/2$. For $b < 1/2$, the upper bound $c(n, P)$ on the coverage of

9

the delta method seems to tend to $1 - \alpha$. Yet, in line with Figure 1, the validity of the asymptotic approximation requires very large sample sizes.
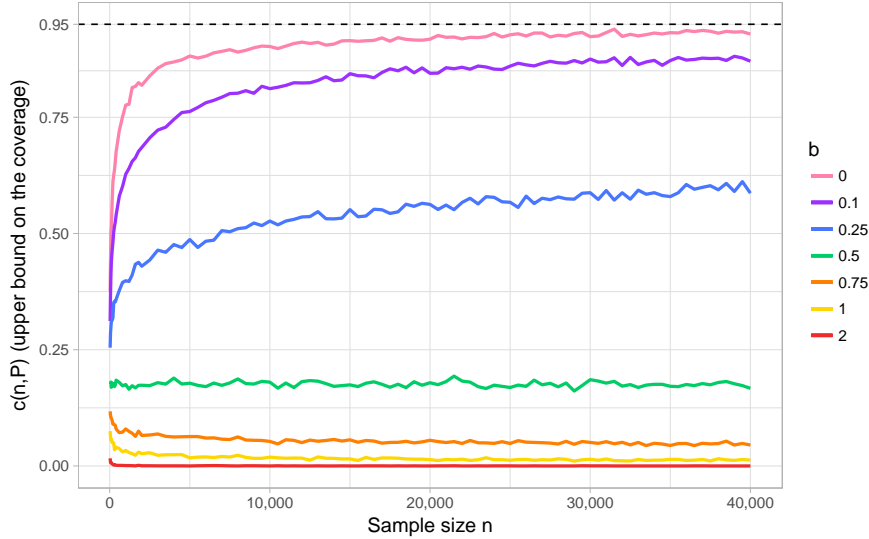


Figure 2: $c(n, P)$ of the asymptotic CIs based on the delta method as a function of the sample size $n$.
Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}(1, 1) \otimes \mathcal{N}(Cn^{-b}, 1)$, with $C = 0.025$. The nominal pointwise asymptotic level is set to 0.95. For each pair $(b, n)$, the coverage is obtained as the mean over 5,000 repetitions.

At this stage, Figure 2 presents some evidence that the CIs based on the delta method need to be adapted for sequences of models and that the rate of decrease toward 0 of the expectation $\mathbb{E}[Y_{1,n}]$ matters. The next subsection details formal results in this set-up.

## 3.3 Extension of the delta method for ratios of expectations in the sequence-of-model framework

We are interested in the asymptotic distribution, as $n$ tends to infinity, of the real random variable $S_n := \sqrt{n} \left( \overline{X}_n / \overline{Y}_n - \mathbb{E}[X_{1,n}] / \mathbb{E}[Y_{1,n}] \right)$. The following theorem states the asymptotic behavior of $S_n$ according to the comparison of $\mathbb{V}[Y_{1,n}] / \sqrt{n}$ and $\mathbb{E}[Y_{1,n}]$ under a multivariate Lyapunov condition. It is proved in Section B.1.

We show that in some cases $|S_n| \xrightarrow[n \to +\infty]{a.s.} +\infty$. It is then impossible to state the limiting distribution $S_n$ in the traditional sense. Despite that, we can still get a more precise result looking at the subsequent terms in the asymptotic expansion of $S_n$. Such an asymptotic expansion is complicated to state, especially in our sequence-of-model framework, since the distributions $P_{X,Y,n}$ change with $n$ without

10

any link from one to the next. To overcome this problem, we consider equivalents in distribution of $S_n$ in the following sense. We say that two sequences of random variables $S_n$ and $T_n$ are *equivalent in distribution* if there exist a probability space $\tilde{\Omega}$ and two sequences of random variables $\tilde{S}_n, \tilde{T}_n$ such that $\forall n \in \mathbb{N}^*$, $S_n \stackrel{d}{=} \tilde{S}_n$ and $T_n \stackrel{d}{=} \tilde{T}_n$, and $\tilde{S}_n$ is equivalent to $\tilde{T}_n$ almost surely as $n \to \infty$. This means that for almost every $\tilde{\omega} \in \tilde{\Omega}$, $\tilde{S}_n(\tilde{\omega})$ is equivalent to $\tilde{T}_n(\tilde{\omega})$ (considered as deterministic sequences of real numbers). This notion enables to formalize the link between $S_n$ and a simpler expression $T_n$.

**Theorem 3.1.** *Let Assumption 1 hold and (i)* $\mathbb{V}[(\gamma_{X,n}X_{1,n}\,,\,\gamma_{Y,n}Y_{1,n})] \to V$ *as* $n \to \infty$ *for some positive sequences* $\{\gamma_{X,n}\}_{n\in\mathbb{N}^*}$ *and* $\{\gamma_{Y,n}\}_{n\in\mathbb{N}^*}$ *where $V$ is a definite positive $2 \times 2$ matrix, (ii)* $\sup_{n\in\mathbb{N}^*} \mathbb{E}\left[|X_{1,n}|^3\gamma_{X,n}^3 + |Y_{1,n}|^3\gamma_{Y,n}^3\right] < +\infty$, *and (iii)* $\mathbb{P}(\overline{Y}_n = 0) \to 0$ *as* $n \to \infty$.
*Denote the signal-to-noise-ratio by* $SNR_n := \mathbb{E}[Y_{1,n}]/(V_{2,2}^{1/2}n^{-1/2}\gamma_{Y,n}^{-1})$.
*    Then, the sequence of random variables* $S_n := \sqrt{n}\left(\overline{X}_n/\overline{Y}_n - \mathbb{E}[X_{1,n}]/\mathbb{E}[Y_{1,n}]\right)$ *satisfies as* $n \to \infty$:

1. *If* $SNR_n \to +\infty$, *then $S_n$ is equivalent in distribution to:*

$$\frac{\sqrt{n}\gamma_{X,n}(\overline{X}_n - \mathbb{E}[X_{1,n}])}{\mathbb{E}[Y_{1,n}]\gamma_{X,n}} - \frac{\sqrt{n}\gamma_{Y,n}(\overline{Y}_n - \mathbb{E}[Y_{1,n}])\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]^2\gamma_{Y,n}}.$$

2. *If there exists a finite constant* $C \neq 0$ *such that* $SNR_n \to C$, *then $S_n$ is equivalent in distribution to:*

$$n\gamma_{Y,n}\mathbb{E}[X_{1,n}]\left(\frac{1}{C + \sqrt{n}\gamma_{Y,n}(\overline{Y}_n - \mathbb{E}[Y_{1,n}])} - \frac{1}{C}\right)$$
$$+ \frac{n\gamma_{X,n}(\overline{X}_n - \mathbb{E}[X_{1,n}]) \times \gamma_{Y,n}}{\left(C + \sqrt{n}\gamma_{Y,n}(\overline{Y}_n - \mathbb{E}[Y_{1,n}])\right) \times \gamma_{X,n}}.$$

3. *If* $SNR_n \to 0$, *then $S_n$ is equivalent in distribution to:*

$$\sqrt{n}\left(\frac{\sqrt{n}\gamma_{X,n}(\overline{X}_n - \mathbb{E}[X_{1,n}])}{\sqrt{n}\gamma_{Y,n}(\overline{Y}_n - \mathbb{E}[Y_{1,n}])} \times \frac{\gamma_{Y,n}}{\gamma_{X,n}} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]}\right).$$

Theorem 3.1 can thus be interpreted as a generalization of the result given by the CLT and the delta method for ratios of expectations. The sequence-of-model framework allows both the expectation and the variance in the denominator to tend to 0. In particular, this happens whenever $Y_{i,n}$ follows a Bernoulli distribution with a parameter $p_n$ tending to 0, as detailed in Example 3.2. For instance, when

Coefficient b

1/2 —
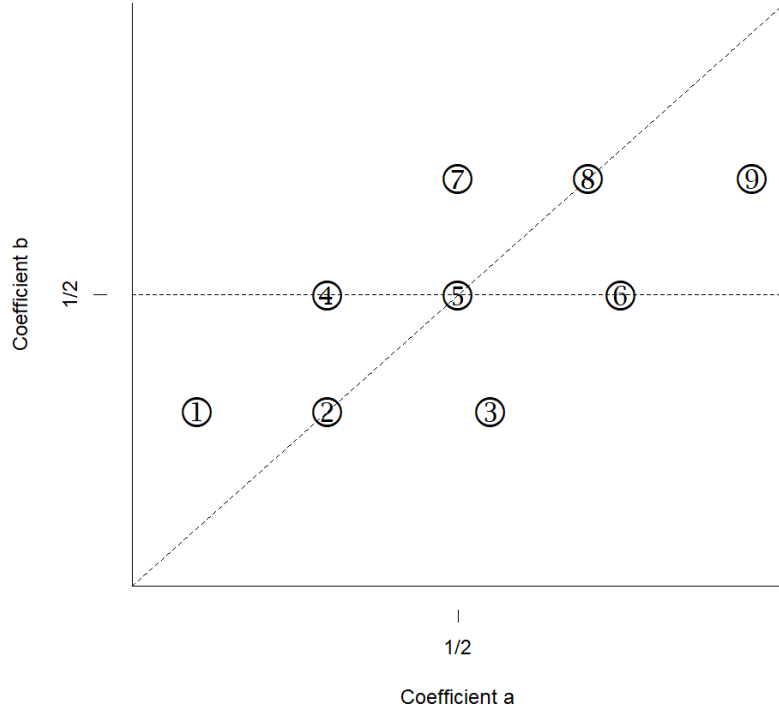
⑦  ⑧  ⑨

④  ⑤  ⑥

① ② ③

|
1/2

Coefficient a

Figure 3: Separation between the different asymptotic regimes as a function of $(a, b)$ for fixed $(a', b') = (0, 0)$, in the case where $\mathbb{E}[X_{1,n}] = C_1/n^a$, $\mathbb{V}[X] = 1/n^{a'}$, $\mathbb{E}[Y_{1,n}] = C_2/n^b$, and $\mathbb{V}[Y] = 1/n^{b'}$, $(a, a', b, b') \in \mathbb{R}_+^4$.

| | $a + b' < b + a'$ | $a + b' = b + a'$ | $a + b' > b + a'$ |
|---|---|---|---|
| $b > 1/2 + b'$ | $n^{1/2 + b' - a'} W_1/W_2$ | $n^{1/2 + b' - a'}\left(W_1/W_2 - C_1/C_2\right)$ | $-n^{1/2 + b - a} C_1/C_2$ |
| $b = 1/2 + b'$ | $n^{1 - a + b'}\left(C_1/(C_2 + W_2) - C_1/C_2\right)$ | $n^{1/2 + b' - a'}\big(C_1/(C_2 + W_2)$ $-C_1/C_2 + W_1/(C_2 + W_2 n^{a'}))$ | $n^{1/2 + b' - a'}\left(W_1/(C_2 + W_2 n^{a'})\right)$ |
| $b < 1/2 + b'$ | $n^{2b - a - b'} C_1 W_2/C_2^2$ | $n^{b - a'}(W_1/C_1 - C_1 W_2/C_2^2)$ | $n^{b - a'} W_1/C_1$ |

Table 1: Limiting law of $S_n := \sqrt{n}\left(\overline{X}_n/\overline{Y}_n - \mathbb{E}[X_{1,n}]/\mathbb{E}[Y_{1,n}]\right)$ in the nine different regimes. The couple of variables $(W_1, W_2)$ follow the distribution $\mathcal{N}(0, V)$, where $V = \lim_{n \to +\infty} \mathbb{V}\left[(n^{a'} X_{1,n}, n^{b'} Y_{1,n})\right]$.

we estimate a conditional expectation with a discrete conditioning variable or a conditioning event, the denominator is an average of indicator variables that follow a Bernoulli distribution. Figure 3 and its companion table highlight the different asymptotic regimes depending on the behaviors of $\{\mathbb{E}[X_{1,n}]\}_{n\in\mathbb{N}^*}$, $\{\mathbb{E}[Y_{1,n}]\}_{n\in\mathbb{N}^*}$, $\{\gamma_{X,n}\}_{n\in\mathbb{N}^*}$ and $\{\gamma_{Y,n}\}_{n\in\mathbb{N}^*}$.

The main takeaway of Theorem 3.1 is that when $\mathbb{E}[X_{1,n}] = C_1/n^a$, $\mathbb{E}[Y_{1,n}] = C_2/n^b$ and $\mathbb{V}[Y] = C_3/n^{b'}$ for some constants $C_1, C_2, C_3 \neq 0$, and $b < 1/2 + b'$, $S_n$ properly renormalized by $n$ to some power still converges in distribution to a Normal random variable. This can be explained using the signal-to-noise ratio (SNR) defined in Theorem 3.1. Indeed, in this first case, the $\text{SNR}_n$ tends to $+\infty$: the signal in the denominator (that is the expectation of $Y_{1,n}$) is asymptotically bigger than the noise (which is $1/(\gamma_{Y,n}n^{1/2})$ up to a constant factor). Asymptotic inference based on the Normal approximation remains valid, even if the length of such confidence intervals may not decrease with the sample size $n$.

In all other cases, when the noise dominates in the denominator, $S_n$ converges weakly to a non-Gaussian distribution, in some cases to a generalized Cauchy distribution with parameters that depend on the data generating process (up to a normalization of some power of $n$). By construction, when the noise dominates, we do not have much information and thus may not be able to conduct inference in these settings. This echoes the impossibility results presented in Section 5, notably Remark 5.3. In the next section, we provide another method for constructing confidence intervals using the nonparametric percentile bootstrap.

**Example 3.2.** *When $Y_{1,n}$ follows a Bernoulli distribution with parameter $p_n$ in $(0,1)$, we are always in the first case of Theorem 3.1, meaning that its expectation $p_n$ is always larger than the noise $\sqrt{p_n(1-p_n)/n}$. This latter formula is obtained by remarking that the standard deviation of $Y_{i,n}$ is $\sqrt{p_n(1-p_n)}$ so that $\gamma_{Y,n} = 1/\sqrt{p_n(1-p_n)}$. However, in order to satisfy the constraint $\mathbb{P}(\overline{Y}_n = 0) \to 0$, we have to impose that $np_n \to +\infty$. Therefore, when $p_n = n^{-b}$, confidence intervals based on the delta method will be pointwise consistent if $b < 1$.*

## 3.4 Validity of the nonparametric bootstrap for sequences of models

In this part, we construct confidence intervals for ratios of expectations using Efron's percentile bootstrap. This technique relies on the nonparametric bootstrap resampling scheme that we now recall. We fix a number $B > 0$ of bootstrap

replications. For a given initial sample $(X_{i,n}, Y_{i,n}), i = 1, \ldots, n$, and a given integer $b$ smaller than $B$, we define the bootstrapped sample $(X_{i,n}^{(b)}, Y_{i,n}^{(b)}), i = 1, \ldots, n$, which is obtained by $n$ i.i.d. resampling from the initial sample, i.e. with replacement. Let $\overline{X}_n^{(b)} := n^{-1} \sum_{i=1}^n X_{i,n}^{(b)}$ be the empirical mean of the numerator in the $b$-th bootstrapped sample (resp. $\overline{Y}_n^{(b)}$ for the denominator).

Then, Efron's percentile bootstrap, also known as the nonparametric percentile bootstrap, consists in using the quantiles of the bootstrapped distribution conditional on the data to conduct inference. More precisely, for every $\tau \in (0, 1)$, let $q_\tau^{boot}$ denote the quantile at level $\tau$ of $\overline{X}_n^{(1)}/\overline{Y}_n^{(1)}$, which is estimated in practice by the empirical quantile at level $\tau$ of the bootstrapped statistics $\left(\overline{X}_n^{(b)}/\overline{Y}_n^{(b)}\right)_{b=1,\ldots,B}$. For a given nominal level $1 - \alpha \in (0, 1)$, the confidence interval we consider is defined as $C_{n,\alpha}^{boot} := \left[q_{\alpha/2}^{boot}, q_{1-\alpha/2}^{boot}\right]$. The following theorem states the consistency of this interval. It is proved in Section B.2.

**Theorem 3.3.** *Let Assumption 1 hold and (i)* $\mathbb{V}[(\gamma_{X,n}X_{1,n}, \gamma_{Y,n}Y_{1,n})] \to V$ *as* $n \to \infty$ *for some positive sequences* $\{\gamma_{X,n}\}_{n\in\mathbb{N}^*}$ *and* $\{\gamma_{Y,n}\}_{n\in\mathbb{N}^*}$ *where* $V$ *is a definite positive* $2 \times 2$ *matrix, (ii)* $\sup_{n\in\mathbb{N}^*} \mathbb{E}\left[(\gamma_{X,n}X_{1,n})^{4+\delta} + (\gamma_{Y,n}Y_{1,n})^{4+\delta}\right] < +\infty$ *for some* $\delta > 0$, *(iii)* $\mathbb{P}(\overline{Y}_n = 0) \to 0$ *as* $n \to \infty$, *and (iv)* $\mathbb{P}(\overline{Y}_n^{(1)} = 0) \to 0$ *as* $n \to \infty$.
*Denote the signal-to-noise-ratio by* $SNR_n := \mathbb{E}[Y_{1,n}]/(V_{2,2}^{1/2}n^{-1/2}\gamma_{Y,n}^{-1})$.
*If* $SNR_n \to +\infty$, *then for every* $\alpha \in (0, 1)$, *the confidence interval* $C_{n,\alpha}^{boot}$ *is pointwise consistent at level* $1 - \alpha$, *viz.* $\mathbb{P}\left(C_{n,\alpha}^{boot} \ni \mathbb{E}[X_{1,n}]/\mathbb{E}[Y_{1,n}]\right) \to 1-\alpha$ *as* $n \to \infty$.

The assumption $\mathbb{P}(\overline{Y}_n^{(1)} = 0) \to 0$ is satisfied for a large set of cases, for instance when the variables $Y_{i,n}$ are continuous or when they follow a Bernoulli distribution with a parameter decreasing to 0 not too fast (see Example 3.4 below).

Note that the moment condition of order $4+\delta$ is nearly sharp. Indeed, the proofs require the strong law of large numbers for $n^{-1} \sum_{i=1}^n X_{1,n}^2$ and $n^{-1} \sum_{i=1}^n Y_{1,n}^2$. As we are dealing with a triangular array of random variables, Theorem 3.1 of Gut (1992) shows that moments of order at least 4 are necessary, even in the simpler case where the distribution $P_{X,Y,n}$ does not depend on $n$.

**Example 3.4** (Example 3.2 continued). *When* $Y_{1,n}$ *follows a Bernoulli distribution with parameter* $p_n = 1/n^b$ *for a given* $b > 0$, *the condition* $\mathbb{P}(\overline{Y}_n^{(1)} = 0) \to 0$ *is satisfied when* $b < 1$. *We refer the reader to Section B.3 for a proof of this claim.*

In practice, even if the theoretical results of the delta method and of the bootstrap are valid under nearly the same set of assumptions, we observe in the sim-
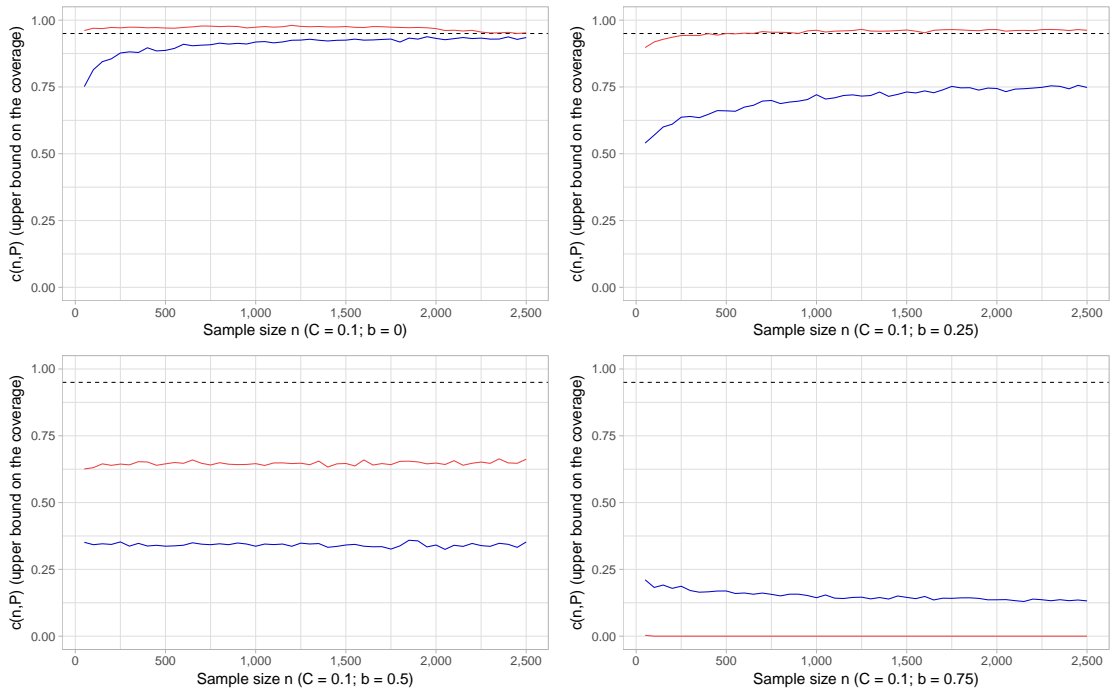
Figure 4: $c(n, P)$ of the asymptotic CIs based on the delta method (blue) and of the CIs constructed with Efron's percentile bootstrap using 2,000 bootstrap replications (red). Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}(1,1) \otimes \mathcal{N}(Cn^{-b}, 1)$, with $C = 0.1$ and $b \in \{0, 0.25, 0.5, 0.75\}$. The nominal pointwise asymptotic level is set to 0.95. For each pair $(b, n)$, the coverage is obtained as the mean over 5,000 repetitions.

ulations in Figure 4 a gap between their pointwise coverage.[8] This fact appears even when $P_{X,Y,n}$ does not depend on $n$ (i.e. $b = 0$). Nonetheless, the coverage gap between these two methods shrinks as $n$ increases provided $b < 0.5$. In the sequence of models where the denominator decreases slowly (i.e. $b = 0.25$) in Figure 4, the bootstrap's coverage is much higher than the one of the delta method. Therefore, the CI provided by the nonparametric percentile bootstrap may be an interesting alternative compared to the delta method when conducting inference with a given sample. This is all the more so as the mean in the denominator is close to 0 (in Figure 4, of the size of $n^{-0.25}/10$ for a variance normalized to 1) and the number of observations is moderately large (a few thousands here).

---

[8]Additional simulations comparing the two types of asymptotic confidence intervals are presented in Appendix D.7.

# 4 Construction of nonasymptotic confidence intervals for ratios of expectations

To construct nonasymptotic confidence intervals, we rely on the possibility to ensure that with large probability (i) $\overline{X}_n$ is close to $\mathbb{E}[X_{1,n}]$, and (ii) $\overline{Y}_n$ is both close to $\mathbb{E}[Y_{1,n}]$ and bounded away from 0. Under Assumptions 1 and 2, the Bienaymé-Chebyshev inequality can be applied to obtain (i) and (ii). On the other hand, without further restrictions, we are only able to build nonasymptotic CIs at nominal levels that are not too close to 1 (see Section 4.2).

This limitation does not arise with nonasymptotic confidence intervals for expectations. In that sense, we can say that building nonasymptotic CIs for ratios of expectations is more demanding. Intuitively, the extra difficulty of the latter task comes from the need to ensure (ii). To stress that point, we show in the next subsection that when $\overline{Y}_n$ is bounded away from 0 and positive almost surely, we can build nonasymptotic CIs at every nominal level.

## 4.1 An easy case: the support of the denominator is well-separated from $0$

We present a simple framework in which it is possible to build nonasymptotic CIs, valid for every $n \in \mathbb{N}^*$, and with coverage $1 - \alpha$ for every $\alpha \in (0, 1)$. To do so, we restrict further the set $\mathcal{P}$ of admissible distributions with the following assumption.

**Assumption 3.** *For every $n \in \mathbb{N}^*$, there exists a positive finite constant $a_{Y,n}$ such that $Y_{1,n} \geq a_{Y,n}$ almost surely.*

Under Assumption 3, for every $n \in \mathbb{N}^*$, $\overline{Y}_n \geq a_{Y,n} > 0$ almost surely under every distribution in $\mathcal{P}$ and $\overline{Y}_n^{-1}$ is bounded from above. This assumption obviously rules out binary $\{0, 1\}$ random variables in the denominator of the ratio, which can be quite restrictive in practice. Under this assumption, the following theorem gives a concentration inequality for our ratio of expectations. It is proved in Section B.4.

**Theorem 4.1.** *Let Assumptions 1, 2 and 3 hold. For every $n \in \mathbb{N}^*$, $\varepsilon > 0$, we have*

$$\sup_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}} \left( \left| \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]} \right| > \frac{\left(\varepsilon + \sqrt{u_{X,n}}\right)\varepsilon}{a_{Y,n} l_{Y,n}} + \frac{\varepsilon}{l_{Y,n}} \right) \leq \frac{u_{X,n}}{n\varepsilon^2} + \frac{u_{Y,n} - l_{Y,n}^2}{n\varepsilon^2}.$$

16

As a consequence, $\inf_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}} \left( \mathbb{E}[X_{1,n}] / \mathbb{E}[Y_{1,n}] \in \left[ \overline{X}_n / \overline{Y}_n \pm t \right] \right) \geq 1 - \alpha$, with the choice

$$t := \frac{1}{l_{Y,n}} \sqrt{\frac{u_{X,n} + u_{Y,n} - l_{Y,n}^2}{n\alpha}} \left( 1 + \frac{1}{a_{Y,n}} \left\{ \sqrt{\frac{u_{X,n} + u_{Y,n} - l_{Y,n}^2}{n\alpha}} + \sqrt{u_{X,n}} \right\} \right),$$

for every $\alpha \in (0, 1)$.

The theorem shows that it is possible to construct nonasymptotic CIs for ratios of expectations, with guaranteed coverage at every confidence level, that are almost surely bounded under every distribution in $\mathcal{P}$ characterized by Assumptions 1, 2 and 3. In Section 4.2, we give an analogous result that only requires Assumptions 1 and 2 to hold, so that it encompasses the case of $\{0, 1\}$-valued denominators. However, the cost to pay will be an upper bound on the achievable coverage of the confidence intervals.

## 4.2 General case: no assumption on the support of the denominator

We seek to build nontrivial nonasymptotic CIs under Assumptions 1 and 2 only. Under Assumption 1, $\mathbb{E}[Y_{1,n}] \neq 0$, so that there is no issue in considering the fraction $\mathbb{E}[X_{1,n}] / \mathbb{E}[Y_{1,n}]$. However, without Assumption 3, $\{\overline{Y}_n = 0\}$ has positive probability in general so that $\overline{X}_n / \overline{Y}_n$ is well-defined with probability less than one. Note that when $P_{Y,n}$ is continuous with respect to Lebesgue's measure, there is no issue in defining $\overline{X}_n / \overline{Y}_n$ anymore since the event $\{\overline{Y}_n = 0\}$ has probability zero. This is not an easier case from a theoretical point of view though since, without more restrictions, $\overline{Y}_n$ can still be arbitrarily close to 0 with positive probability.

**Theorem 4.2.** *Let Assumptions 1 and 2 hold. For every $n \in \mathbb{N}^*$, $\varepsilon > 0, \tilde{\varepsilon} \in (0, 1)$, we have*

$$\sup_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}} \left( \left| \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]} \right| > \left( \frac{(\sqrt{u_{X,n}} + \varepsilon)\tilde{\varepsilon}}{(1 - \tilde{\varepsilon})^2} + \varepsilon \right) \frac{1}{l_{Y,n}} \right) \leq \frac{u_{X,n}}{n\varepsilon^2} + \frac{u_{Y,n} - l_{Y,n}^2}{n\tilde{\varepsilon}^2 l_{Y,n}^2}.$$

As a consequence, $\inf_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}} \left( \mathbb{E}[X_{1,n}] / \mathbb{E}[Y_{1,n}] \in \left[ \overline{X}_n / \overline{Y}_n \pm t \right] \right) \geq 1 - \alpha$, with

17

*the choice*

$$t = \frac{1}{l_{Y,n}} \left( \frac{\left( \sqrt{u_{X,n}} + \sqrt{2u_{X,n}/(n\alpha)} \right) \sqrt{2(u_{Y,n} - l_{Y,n}^2)/(n\alpha l_{Y,n}^2)}}{\left( 1 - \sqrt{2(u_{Y,n} - l_{Y,n}^2)/(n\alpha l_{Y,n}^2)} \right)^2} + \sqrt{\frac{2u_{X,n}}{n\alpha}} \right),$$

*for every* $\alpha > \overline{\alpha}_n := \frac{2(u_{Y,n} - l_{Y,n}^2)}{n l_{Y,n}^2}$.[9]

This theorem is proved in Section B.5. It states that when $l_{Y,n} > 0$, it is possible to build valid nonasymptotic CIs with finite length up to the confidence level $1 - \overline{\alpha}_n$. This is a more positive result than Dufour (1997) which states that it is not possible to build nontrivial nonasymptotic CIs when $l_{Y,n}$ is taken equal to 0, no matter the confidence level. Note that Theorem 4.2 is not an impossibility theorem since it only claims that considering confidence levels smaller than $1 - \overline{\alpha}_n$ is *sufficient* to build nontrivial CIs under Assumptions 1 and 2. The remaining question is to find out whether it is *necessary* to focus on confidence levels that do not exceed a certain threshold under Assumptions 1 and 2. We answer this in Section 5.1.

Theorem 4.2 has two other interesting consequences: for every confidence level up to $1 - \overline{\alpha}_n$, a nonasymptotic interval of the form $\left[ \overline{X}_n / \overline{Y}_n \pm \tilde{t} \right]$ with $\tilde{t} > t$ has coverage $1 - \alpha$ but is unnecessarily conservative. Moreover, if the data generating process does not depend on $n$ (i.e. in the standard i.i.d. set-up), the length of the confidence interval shrinks at the optimal rate $1/\sqrt{n}$ for every fixed $\alpha$. Note that the coefficient 2 in the definition of $\overline{\alpha}_n$ defined above can be reduced to any number $w > 1$, at the expense of increasing the length of the confidence interval (this length actually tends to infinity when $w$ tends to 1).

# 5 Nonasymptotic CIs: impossibility results and practical guidelines

In this section, we prove two impossibility results: a maximum confidence level above which it is impossible to build nontrivial nonasymptotic CIs and a necessary lower bound on the length of nonasymptotic CIs.

---

[9]Equivalently, it means that for a given $\alpha$, the above choice of $t$ is valid for every integer $n > \overline{n}_\alpha := 2(u_{Y,n} - l_{Y,n}^2)/(\alpha l_{Y,n}^2)$.

## 5.1 An upper bound on testable confidence levels

**Proposition 5.1.** *Let $\mathcal{P}$ be the class of all distributions satisfying Assumptions 1 and 2 and $\underline{\alpha}_n := \left(1 - l_{Y,n}^2/u_{Y,n}\right)^n$. For every $n \in \mathbb{N}^*$ and every $\alpha \in (0, \underline{\alpha}_n)$, if $l_{Y,n}^2/u_{Y,n} < 1$, there is no finite $t > 0$ such that $\left[\overline{X}_n/\overline{Y}_n \pm t\right]$ has coverage $1 - \alpha$ over $\mathcal{P}$.*

This theorem asserts that confidence intervals of the form $\left[\overline{X}_n/\overline{Y}_n \pm t\right]$ with coverage higher than $1 - \underline{\alpha}_n$ under Assumptions 1 and 2 are not defined (or are of infinite length) with positive probability for at least one distribution in $\mathcal{P}$. This is due to the fact that $\underline{\alpha}_n$ is a lower bound on $\mathbb{P}(\overline{Y}_n = 0)$ over all distributions in $\mathcal{P}$.

Remark that when $u_{Y,n}/l_{Y,n}^2 = 1$, there is no impossibility result anymore: assume that $u_{Y,n}/l_{Y,n}^2 = 1$ and let $Q$ be a distribution on $\mathbb{R}^2$ that satisfies Assumptions 1 and 2. Let $(X_{i,n}, Y_{i,n})_{i=1}^n \overset{\text{i.i.d.}}{\sim} Q$. We have that $\mathbb{V}[Y_{1,n}] = 0$, which implies that $Y_{1,n} = \mathbb{E}[Y_{1,n}]$ almost surely. Assumption 1 further ensures that $Y_{1,n} \neq 0$ almost surely. Consequently, the results of Section 4.1 apply and allow us to conclude that under Assumptions 1, 2 and $u_{Y,n}/l_{Y,n}^2 = 1$, it is possible to build nontrivial nonasymptotic CIs at every confidence level. Indeed, in that case, we are in fact only estimating a simple mean, and therefore there is no constraint on $\alpha$.

Proposition 5.1 is actually a corollary of the more general Theorem 5.2. It states it is impossible to construct confidence intervals that contain $\overline{X}_n/\overline{Y}_n$ almost surely and are almost surely bounded over $\mathcal{P}$ with coverage greater than $1 - \underline{\alpha}_n$. It is proved in Section B.6.

**Theorem 5.2.** *Let $\mathcal{P}$ be the class of all distributions satisfying Assumptions 1 and 2. Let $n \in \mathbb{N}^*$, and a random set $I_n$ that contains $\overline{X}_n/\overline{Y}_n$ almost surely whenever it is defined and is undefined if $\overline{Y}_n = 0$. Then $\sup_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}}\left(I_n \, undefined\right) \geq \underline{\alpha}_n$.*

Combining Theorems 4.2 and 5.2, we conclude that there exists some critical level $1 - \alpha_n^c$ belonging to the interval $[1 - \overline{\alpha}_n, 1 - \underline{\alpha}_n]$ such that it is impossible to build nontrivial nonasymptotic confidence intervals if and only if their nominal level is above $1 - \alpha_n^c$. Finally, it is worth remarking that with a sample of size $n$, the CIs based on the delta method with a nominal level $1 - \alpha > 1 - \alpha_n^c$ cannot have coverage $1 - \alpha$ uniformly over $\mathcal{P}$ as such CIs verify the condition of Theorem 5.2.

Figure 5 below shows the critical level and its bounds obtained in our nonasymptotic results.
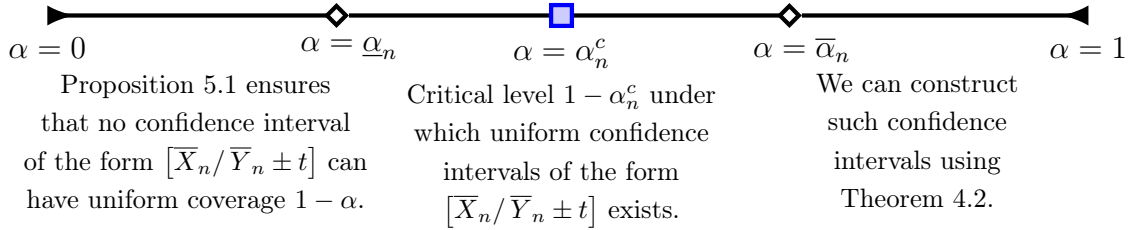
Figure 5: The critical level and its bounds.

**Remark 5.3.** *In the same spirit as in Theorem 3.1, we consider a modified version of the signal-to-noise ratio defined by $\widetilde{SNR}_n := l_{Y,n}/(u_{Y,n}^{1/2}n^{-1/2})$. When $\widetilde{SNR}_n \to +\infty$ (resp. 0) as $n \to \infty$, $\underline{\alpha}_n$ and $\overline{\alpha}_n$ tend to 0 (resp. $+\infty$). When we have enough information ($\widetilde{SNR}_n \to +\infty$), the critical level $1 - \alpha_n^c$ tends to 1. Therefore, for every $\alpha \in (0,1)$, nonasymptotic confidence intervals can be constructed at every level for n large enough. On the contrary, when $\widetilde{SNR}_n \to 0$, the critical level $1 - \alpha_n^c$ tends to 0, which means that it is impossible to construct uniformly valid CIs for n large enough. Finally, when $\widetilde{SNR}_n \to C$ for a positive constant C, a critical level remains as in the nonasymptotic case since $\underline{\alpha}_n \to \exp(-C)$.*

## 5.2 A lower bound on the length of nonasymptotic confidence intervals

The following theorem is an extension of Catoni (2012)[Proposition 6.2] to ratios. It is proved in Section B.7.

**Theorem 5.4.** *For every integer $n \geq 7$, $\alpha \in \left(0, 1 \wedge n/\left(l_{Y,n} + \sqrt{u_{Y,n} - l_{Y,n}^2}\right)^2\right)$, and $\xi < 1$ there exists a distribution $Q$ on $\mathbb{R}^2$ that satisfies Assumptions 1 and 2 such that for $(X_{i,n}, Y_{i,n})_{i=1}^n \overset{i.i.d}{\sim} Q$, we have*

$$\mathbb{P}_{Q^{\otimes n}}\left(\left|\frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]}\right| > \xi\sqrt{\frac{v_n}{3n\alpha}}\right) > \alpha,$$

*where $v_n := u_{X,n}/\left(l_{Y,n} + \sqrt{u_{Y,n} - l_{Y,n}^2}\right)^2$.*

With this theorem, we can claim that CIs of the form $\left[\overline{X}_n/\overline{Y}_n \pm t\right]$ cannot have uniform coverage $1 - \alpha$, for every $\alpha \in \left(0, 1 \wedge n/\left(l_{Y,n} + \sqrt{u_{Y,n} - l_{Y,n}^2}\right)^2\right)$, under Assumptions 1 and 2 if they are shorter than $\sqrt{v_n/(3n\alpha)}$. By a careful inspection of the proof (see Lemma B.6), we can in fact replace the value 3 in the theorem by any number strictly larger than $e = \exp(1)$, at the price of assuming $n \geq n_0$ for

$n_0$ large enough. It is interesting to note that the distributions $Q$ that are built in the proof of the theorem are on the boundary of $\mathcal{P}$ in the sense that they satisfy $\mathbb{E}[X_{1,n}^2] = u_{X,n}$, $\mathbb{E}[Y_{1,n}] = l_{Y,n}$ and $\mathbb{E}[Y_{1,n}^2] = u_{Y,n}$.

## 5.3  Practical methods and plug-in estimators

Nonasymptotic confidence intervals and the thresholds $\overline{\alpha}_n$ and $\overline{n}_\alpha$ based on Theorem 4.2 rely on Assumptions 1 and 2. In practice, building such CIs or computing those thresholds require the knowledge of the constants $l_{Y,n}$, $u_{X,n}$ and $u_{Y,n}$ that determine the class of distributions we consider.[10] Therefore, we need to state some values for those constants. Note that constructing nontrivial and nonasymptotic CIs that overcome the limitations of having to choose some a priori class of distributions is not possible. Indeed, we would get back to Bahadur and Savage (1956) and Dufour (1997) type impossibility results.

How to choose $l_{Y,n}$, $u_{X,n}$ and $u_{Y,n}$ depends on the specific application. Sometimes, stating values can be sensible if researchers do have control or expert knowledge of the variables. Resuming an example started in the introduction, if the variable in the denominator is an indicator of being treated in the setting of a Randomized Controlled Trial, researchers can have intuitions about reasonable values for the lower and upper bounds of the probability of being treated.

The unknown constants are upper and lower bounds on moments that characterize the class $\mathcal{P}$. As such, they can never be recovered from the data since observations are by construction drawn from a single distribution $P \in \mathcal{P}$. Under i.i.d. sampling, sample means converge to their corresponding theoretical moments, provided the latter are finite. Hence, without prior information, a plug-in strategy has to be used which consists in: (i) using the moments of a single distribution instead of the bounds on the class, (ii) estimating those moments with their empirical counterparts. As a consequence, this approach is valid pointwise only and not uniformly over $\mathcal{P}$ anymore. Furthermore, it is only asymptotically justified. On the other hand, for any sample provided $\overline{Y}_n \neq 0$, this plug-in strategy enables us to construct our CIs and the quantity $\overline{n}_\alpha$ (or $\overline{\alpha}_n$), which can be a useful rule of thumb as explained below. We stick to that principle in our simulations and application.

For a given level $1 - \alpha$ and a class of distributions satisfying Assumptions 1 and 2, $\overline{n}_\alpha$ is the minimal sample size required to construct our nonasymptotic CIs.

---

[10]Actually, the computation of $\overline{\alpha}_n$ and $\overline{n}_\alpha$ only require the knowledge of $l_{Y,n}$ and $u_{Y,n}$.

In other words, for a sample size $n < \overline{n}_\alpha$, the data is not rich enough to construct the nonasymptotic CIs of Theorem 4.2 at this level. Heuristically, the comparison of $\overline{n}_\alpha$ and $n$ can be used as a rule of thumb to assess whether the coverage of the CIs based on the delta method matches their nominal level.[11] Several simulations tend to confirm the practical interest of that rule of thumb as $\overline{n}_\alpha$ turns out to be very close to the sample size above which the gap between the coverage of the asymptotic CIs based on the delta method and their nominal level becomes negligible. (see Section 6.1 and Appendix D).

# 6 Numerical applications

## 6.1 Simulations

This section presents simulations that support the use of $\overline{n}_\alpha$, or equivalently $\overline{\alpha}_n$, as a rule of thumb to inspect the reliability of the asymptotic confidence intervals from the delta method.

In Figure 6, a nominal level $1 - \alpha$ is fixed and we show the $c(n, P)$ of the CIs based on the delta method as a function of the sample size $n$, as well as $\overline{n}_\alpha$ derived in Theorem 4.2. It happens that the coverage converges toward its nominal level for sample sizes around $\overline{n}_\alpha$, which supports $\overline{n}_\alpha$ as a rule of thumb of interest in practice.[12] In Figure 7, a sample size is fixed and we show the coverage for different nominal levels, as well as the quantity $\overline{\alpha}_n$. It is the converse of Figure 6 in that sense. In this simulation, $\overline{\alpha}_n$ turns out to fall close to the lowest $\alpha$ (hence highest $1 - \alpha$) for which the coverage of the CIs based on the delta method attains their nominal level.

All in all, Figures 6 and 7 and additional simulations advocate the use of $\overline{n}_\alpha$ derived in Theorem 4.2 (or conversely $\overline{\alpha}_n$) as a rule of thumb to appraise the dependability of the CIs obtained with the delta method for ratios of expectations.

## 6.2 Application to real data

We illustrate our methods with an application related to gender wage disparities. The application resumes our canonical example of conditional expectations since we estimate the proportion of women within wage brackets that are defined as

---

[11]Equivalently, we could compare $\overline{\alpha}_n$ and $\alpha$. As a rule of thumb, $\overline{\alpha}_n$ can be seen as the lowest $\alpha$ (hence the highest nominal level $1-\alpha$) for which the asymptotic CIs based on the delta method are reliable given the sample size $n$.

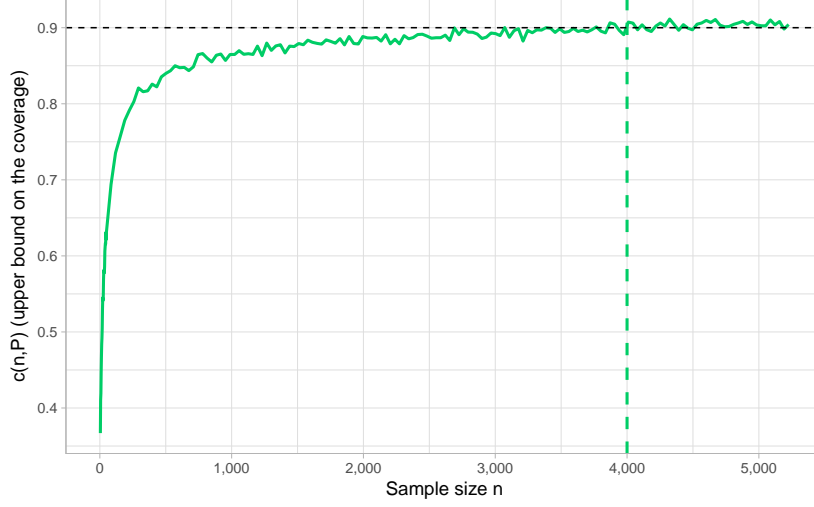[12]This fact holds across various specifications (see additional simulations in Appendix D).

Figure 6: $c(n, P)$ of the asymptotic CIs based on the delta method as a function of the sample size $n$ and $\overline{n}_\alpha$.
Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}_2$ (bivariate Gaussian) with $\mathbb{E}[X] = 0.5$, $\mathbb{E}[Y] = 0.1$, $\mathbb{V}[X] = 1$, $\mathbb{V}[Y] = 2$, $\mathbb{C}orr(X, Y) = 0.5$. The nominal pointwise asymptotic level is set to 0.90. For a sample size $n$, the coverage is obtained as the mean over 5,000 repetitions. The dashed vertical line shows $\overline{n}_\alpha := 2\left(u_{Y,n} - l_{Y,n}{}^2\right) / \left(\alpha l_{Y,n}^2\right)$, setting here $\alpha = 0.1$, $l_{Y,n} = \mathbb{E}[Y]$, $u_{Y,n} = \mathbb{E}[Y]^2 + \mathbb{V}[Y]$.
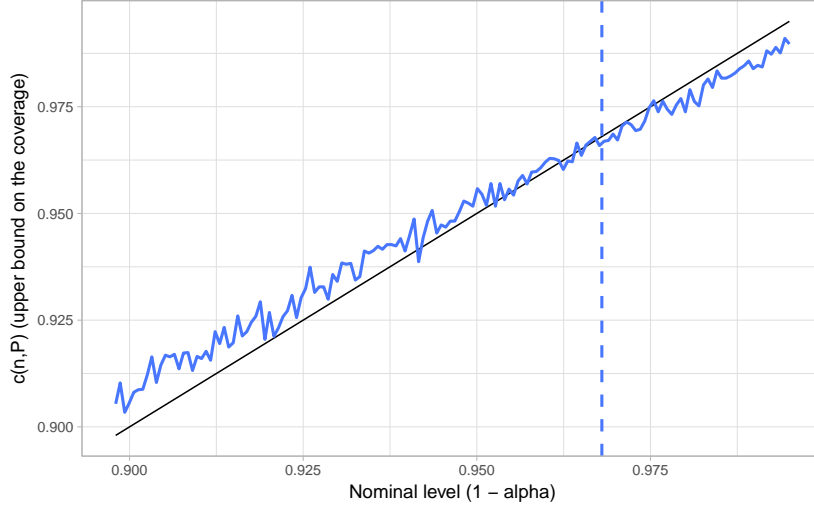


Figure 7: $c(n, P)$ of the asymptotic CIs based on the delta method as a function of the sample size $n$ and $\overline{\alpha}_n$.
Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}_2$ (bivariate Gaussian) with $\mathbb{E}[X] = 0.5$, $\mathbb{E}[Y] = 0.25$, $\mathbb{V}[X] = 2$, $\mathbb{V}[Y] = 1$, $\mathbb{C}orr(X, Y) = 0.5$. The sample size is $n = 1,000$. For each nominal level $1 - \alpha$ in the x-axis, we draw $10,000$ samples, compute the asymptotic CIs and see whether it covers or not the ratio of interest; we report the mean over the $10,000$ repetitions in the y-axis. The solid line is the first bisector $y = x$. The dashed vertical line shows $\overline{\alpha}_n := 2\left(u_{Y,n} - l_{Y,n}{}^2\right) / \left(n l_{Y,n}{}^2\right)$, setting here $l_{Y,n} = \mathbb{E}[Y]$, $u_{Y,n} = \mathbb{E}[Y]^2 + \mathbb{V}[Y]$.

having a wage higher than a given threshold. We use $n = 204{,}246$ observations from the French Labor Survey data between 2010 and 2017.[13]

Let $W$ be a real random variable that indicates the wage of an employee (expressed in euros per month) and $F$ an indicator variable equal to 1 if the employee is a woman and 0 otherwise. For a given threshold wage $w_0$, the parameter of interest is $\mathbb{E}[F \mid W \geq w_0]$. It can be written as a ratio of expectations with $X = F\,\mathbb{1}\{W \geq w_0\} = \mathbb{1}\{F = 1, W \geq w_0\}$ in the numerator and $Y = \mathbb{1}\{W \geq w_0\}$ in the denominator. As we consider higher thresholds $w_0$, the expectation in the denominator gets closer to 0. As an illustration, out of $n = 204{,}246$ observations, 355 individuals have monthly wages higher than 10,000 euros (which corresponds to a mean in the denominator equal to 0.0017); 44 individuals above 20,000 ($\overline{Y}_n = 2.2 \times 10^{-4}$); and only 17 above 30,000 ($\overline{Y}_n = 8.3 \times 10^{-5}$).[14]
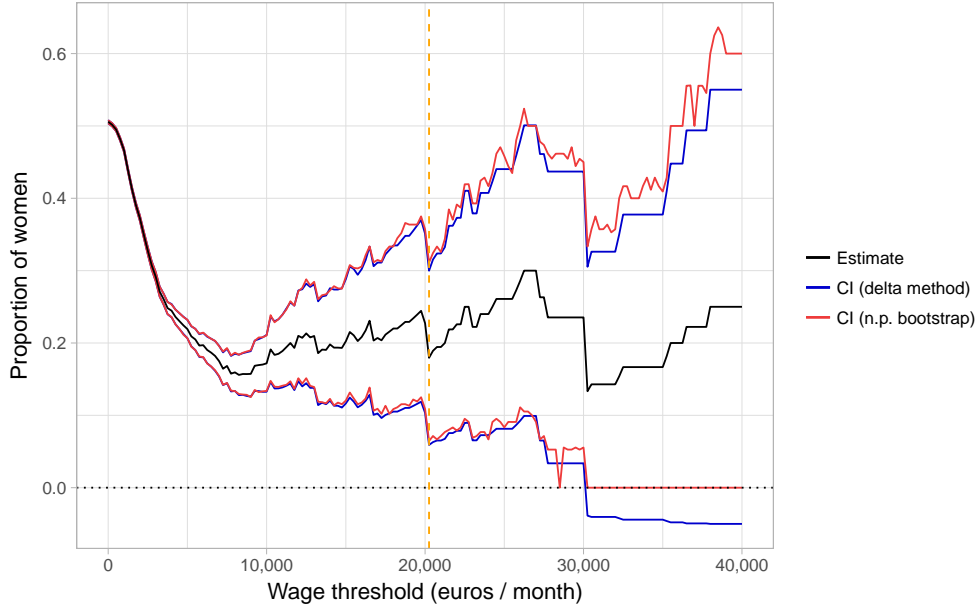


Figure 8: Point estimate and confidence intervals for the parameter $\mathbb{E}[F \mid W \geq w_0]$ as a function of the wage threshold $w_0$. The parameter is the proportion of women within the wage bracket $[w_0, +\infty)$. The nominal level of the CIs is set to 95%. Efron's percentile bootstrap CIs are obtained using 2,000 bootstrap replications. The dashed vertical line represents the lowest wage threshold such that the plug-in counterpart of $\overline{n}_\alpha$ exceeds $n$.

For various thresholds $w_0$, Figure 8 presents the estimate $\widehat{\theta}_n$ and two 95%-nominal-level confidence intervals for the parameter $\mathbb{E}[F \mid W \geq w_0]$: the one based

---

[13]Enquête Emploi en continu (version FPR) – 2010-2017, INSEE [producteur], ADISP [diffuseur].

[14]To give a sense of the wage distribution, note that the empirical quantiles of $W$ at orders 90%; 95%; 99%; and 99.99% are respectively: 2,989; 3,728; 6,000; and 26,024.

on the delta method (see Section 3.1) and the one using Efron's percentile boot-strap (see Section 3.4). With higher thresholds, the expectation in the denominator is closer to 0 which results in wider confidence intervals. For very high thresholds, the CIs become hardly informative. In particular, the lower end of the interval based on the delta method is negative whereas the parameter of interest belongs to $[0, 1]$ by construction.

The dashed vertical line relates to our rule of thumb introduced in Section 5.3. More precisely, given the level $1 - \alpha = 0.95$, for each threshold $w_0$, we compute the plug-in counterpart of $\overline{n}_\alpha$ defined in Theorem 4.2: $2\left(n^{-1}\sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2\right)/\left(\alpha \overline{Y}_n^2\right)$. Given that $Y$ is a binary variable, the latter quantity is increasing with $w_0$ and exceeds $n$ at some threshold represented by the dashed vertical line (here a little above 20,000). Consequently, for higher thresholds, our rule of thumb suggests that the confidence intervals obtained with the delta method might undercover as the expectation in the denominator is "too close to 0" relative to the number of observations. Actually, in the application, it is around this vertical line that the two CIs start to differ. In particular, the upper end of Efron's percentile confidence interval becomes larger than the upper end of the interval based on the delta method.

# 7 Conclusion

This paper studies the construction of confidence intervals for ratios of expectations, which are frequent parameters of interest in applied econometrics.

The most common method to do so is asymptotic and yields CIs based on the asymptotic normality of the empirical means that estimate the numerator and the denominator combined with the delta method. We document on simulations that the coverage of the confidence intervals based on the delta method may fall short of their nominal level when the expectation in the denominator is close to 0, even with fairly large sample size.

To further study the reliability of those CIs, we use a sequence-of-model framework, analogous to what a strand of the weak IV literature does. Indeed, it enables to consider limiting cases, namely here denominators tending to 0. In the weak IV case, the equivalent is to move closer to a null covariance between the endogenous regressor and the instrument. At the limit, the coefficient of interest is not identified. Our problem differs since the parameter is not even defined in

the problematic case of a null denominator. This issue underlies the impossibility type results presented in the paper.

First, in an asymptotic perspective, the possibility of a denominator arbitrarily close to 0 explains why we need a sufficiently slow rate of convergence of the expectation in the denominator to 0 to conduct meaningful inference. More precisely, our main asymptotic results basically show that the CIs based on the delta method are valid, as well as those obtained by Efron's percentile bootstrap, when this speed is lower than $1/\sqrt{n}$ (the standard speed of the CLT). Furthermore, on simulations, Efron's percentile bootstrap CIs reach their nominal level sooner (namely for smaller sample sizes) than the CIs based on the delta method. It suggests that beyond the sequence-of-model rationalization, when confronted in practice to a mean in the denominator close to 0 relative to the size of the sample at hand, Efron's percentile bootstrap CIs may be more trustworthy than the delta method's ones.

Obviously, those cases where the coverage of the CIs based on the delta method can be well below their nominal level do not self-signal to practitioners. This is why the second part of the paper proposes a rule of thumb to detect those cases and thus assess the dependability of the asymptotic CIs based on the delta method on finite samples. This index is based on the construction of nonasymptotic confidence intervals and on impossibility results that stem from the problematic null denominator case.

In substance, even if we bound away from 0 the expectation in the denominator, there remains a partial impossibility result. Indeed, we show that there exists a critical nominal level above which the coverage of any nonasymptotic confidence interval that is undefined when $\overline{Y}_n = 0$ cannot uniformly attain its target level. More precisely, we derive explicit upper and lower bounds on this critical level as a function of the characteristics of the considered class of distributions. Then, the heuristic of our rule of thumb consists in estimating by plug-in a lower bound on this critical level (or equivalently, for a given level, an upper bound on the minimal required sample size). The resulting index can thus be computed immediately on any sample. In addition to its theoretical foundations, various simulations and an application to real data attest the practical usefulness of this rule of thumb.

This paper can be seen as a first step towards nonasymptotic inference in econometric models where the issue of close-to-zero denominators arises. Notable examples may include weak IV, Wald ratios, and difference-in-difference estimands.

# References

ANDERSON, T. W. AND H. RUBIN (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics*, 20, 46–63.

ANDREWS, I., J. STOCK, AND L. SUN (2019): "Weak instruments in IV regression: Theory and practice," *To appear in Annual Review of Economics*.

BAHADUR, R. R. AND L. J. SAVAGE (1956): "The nonexistence of certain statistical procedures in nonparametric problems," *The Annals of Mathematical Statistics*, 27, 1115–1122.

BOUCHERON, S., G. LUGOSI, AND P. MASSART (2013): *Concentration inequalities: A nonasymptotic theory of independence*, Oxford university press.

BÜCHER, A. AND I. KOJADINOVIC (2019): "A note on conditional versus joint unconditional weak convergence in bootstrap consistency results," *To appear in Journal of Theoretical Probability*.

CATONI, O. (2012): "Challenging the empirical mean and empirical variance: a deviation study," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48, 1148–1185.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

GUT, A. (1992): "Complete convergence for arrays," *Periodica Mathematica Hungarica*, 25, 51–75.

PINELIS, I. AND R. MOLZON (2016): "Optimal-order bounds on the rate of convergence to normality in the multivariate delta method," *Electronic Journal of Statistics*, 10, 1001–1063.

ROMANO, J. P. AND M. WOLF (2000): "Finite Sample Nonparametric Inference and Large Sample Efficiency," *The Annals of Statistics*, 28, 756–778.

STAIGER, D. AND J. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. AND M. YOGO (2005): *Testing for Weak Instruments in Linear IV Regression*, New York: Cambridge University Press, 80–108.

VAN DER VAART, A. W. (2000): *Asymptotic statistics*, Cambridge University Press.

# A    General definitions about confidence intervals

A standard situation in statistics or econometrics can be modelled as the observation of a sample of $n \in \mathbb{N}^*$ i.i.d. observations valued in some measurable space $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$. The statistical model is therefore $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}), \mathcal{P})^{\otimes n}$ with $\mathcal{P}$ some specified set of distributions on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$. For every distribution $P \in \mathcal{P}$, let $\theta(P)$ be a parameter of interest and the map $\theta : P \mapsto \theta(P)$ be valued in a metric space $(\Theta, d)$.

We denote by $C_n$ a confidence set for $\theta(P)$. Formally, a confidence set $C_n$ can be defined as a measurable map from $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))^{\otimes n}$ to the measurable space $(\mathcal{F}_\Theta \sqcup \{\text{undefined}\}, \mathcal{B}(\mathcal{F}_\Theta) \sqcup \{\text{undefined}\})$, where $\mathcal{F}_\Theta$ is the family of all closed subsets of $\Theta$ and $\mathcal{B}(\mathcal{F}_\Theta)$ is the sigma-algebra generated by $\{F \in \mathcal{F}_\Theta : F \cap K \neq \emptyset\}$ for $K$ running through the family of compact subsets of $\Theta$.

As the vocabulary may somewhat fluctuate between authors, we define below classical objects to fix the notations and terminology used in this paper. The goal is to build confidence sets for a targeted *confidence level* $1 - \alpha$ (also termed *nominal level* of the confidence set). For $n \in \mathbb{N}^*$, for $\alpha \in (0, 1)$, we say that a confidence set $C_n$ or a sequence of sets $(C_n)_{n \in \mathbb{N}^*}$ has:

i. *coverage* $1 - \alpha$ over $\mathcal{P}$ if:
$$\inf_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}} (C_n \ni \theta(P)) \geq 1 - \alpha$$

ii. *size* $1 - \alpha$ over $\mathcal{P}$ if the inequality is an equality:
$$\inf_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}} (C_n \ni \theta(P)) = 1 - \alpha.$$

iii. *asymptotic coverage* $1 - \alpha$ *pointwise* over $\mathcal{P}$ if:[15]
$$\forall P \in \mathcal{P}, \liminf_{n \to +\infty} \mathbb{P}_{P^{\otimes n}} (C_n \ni \theta(P)) \geq 1 - \alpha.$$

iv. *asymptotic coverage* $1 - \alpha$ *uniformly* over $\mathcal{P}$ if:[16]
$$\liminf_{n \to +\infty} \inf_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}} (C_n \ni \theta(P)) \geq 1 - \alpha.$$

A confidence set with coverage $1 - \alpha$ but size different from $1 - \alpha$ over $\mathcal{P}$ is said to be *conservative* over $\mathcal{P}$[17]. We further define a *nontrivial confidence set* as a confidence set that is almost surely strictly included in $\Theta$ (whenever it is defined) under every distribution in $\mathcal{P}$. For instance, if $\theta(P)$ is the expectation under $P$,

---

[15]Respectively *pointwise asymptotic size* when the inequality is replaced by an equality.

[16]Respectively *uniform asymptotic size* when the inequality is replaced by an equality.

[17]Similarly, a confidence set is said to be *asymptotically conservative pointwise* over $\mathcal{P}$ (respectively *uniformly* over $\mathcal{P}$) if property iii. (resp. property iv.) holds with a strict inequality.

$\Theta = \mathbb{R}$ and $\mathcal{P}$ is the set of all distributions that admit a finite expectation, a nontrivial CI is any CI that is almost surely bounded under every distribution in $\mathcal{P}$. For ratios of expectations, $\Theta = \mathbb{R}$ too and we will use the term *almost surely bounded* as a synonym of nontrivial, without stating "under every distribution in $\mathcal{P}$" when there is no ambiguity as regards the class $\mathcal{P}$ considered.

A family of confidence intervals $(C_{n,\alpha})_{n \in \mathbb{N}^*, \alpha \in (0,1)}$ is said to be *pointwise* (resp. *uniformly*) *consistent* if for every $\alpha \in (0,1)$, the sequence $(C_{n,\alpha})_{n \in \mathbb{N}^*}$ has pointwise (resp. uniformly) asymptotic coverage at level $1 - \alpha$.

# B    Proofs of the results in Sections 3, 4 and 5

## B.1    Proof of Theorem 3.1

Let $\theta_{X,n} := \mathbb{E}[X_{1,n}]$, $\theta_{Y,n} := \mathbb{E}[Y_{1,n}]$. Let $h_{X,n} := \sqrt{n}\gamma_{X,n}(\overline{X}_n - \mathbb{E}[X_{1,n}])$ and $h_{Y,n} := \sqrt{n}\gamma_{Y,n}(\overline{Y}_n - \mathbb{E}[Y_{1,n}])$ be the centered and normalized versions of $\overline{X}_n$ and $\overline{Y}_n$. We first rewrite Theorem 3.1 using this notation.

**Theorem B.1.** *Let Assumption 1 hold. Assume that* $\mathbb{V}[(\gamma_{X,n}X_{1,n}, \gamma_{Y,n}Y_{1,n})] \to V$ *for some positive sequences* $\gamma_{X,n}$ *and* $\gamma_{Y,n}$ *where* $V$ *is a definite positive* $2 \times 2$ *matrix, that* $\mathbb{P}(\overline{Y}_n = 0) \to 0$, *as* $n \to \infty$ *and that*

*Then the sequence of random variables* $A_n := \overline{X}_n/\overline{Y}_n - \theta_{X,n}/\theta_{Y,n}$ *satisfies as* $n \to \infty$:

1. *If* $n^{-1/2} = o(\gamma_{Y,n}\theta_{Y,n})$, *then* $A_n$ *is equivalent to*

$$n^{-1/2}\left(\frac{h_{X,n}}{\theta_{Y,n}\gamma_{X,n}} - \frac{h_{Y,n}\theta_{X,n}}{\gamma_{Y,n}\theta_{Y,n}^2}\right).$$

2. *If there exists a finite constant* $C \neq 0$ *such that* $\sqrt{n}\gamma_{Y,n}\theta_{Y,n} \to C$ *as* $n \to \infty$, *then* $A_n$ *is equivalent to*

$$\sqrt{n}\gamma_{Y,n}\theta_{X,n}\left(\frac{1}{C + h_{Y,n}} - \frac{1}{C}\right) + \frac{h_{X,n}\gamma_{Y,n}}{(C + h_{Y,n})\gamma_{X,n}}.$$

3. *If* $\gamma_{Y,n}\theta_{Y,n} = o(n^{-1/2})$, *then* $A_n$ *is equivalent to*

$$\frac{h_{X,n}\gamma_{Y,n}}{h_{Y,n}\gamma_{X,n}} - \frac{\theta_{X,n}}{\theta_{Y,n}}.$$

30

Let us define $W_n := \mathbb{1}\{\theta_{Y,n} + h_{Y,n}/(\sqrt{n}\gamma_{Y,n}) = 0\}$ and remark that $W_n = 1$ whenever $\overline{Y}_n = 0$. By assumption $\mathbb{P}(\overline{Y}_n = 0) \to 0$, therefore $W_n \xrightarrow[n\to+\infty]{d} \delta_0$. Moreover, by Lyapunov's central limit theorem applied to

$$(h_{X,n}, h_{Y,n}) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_{i,n}\gamma_{X,n}, Y_{i,n}\gamma_{Y,n}) - (\mathbb{E}[X]\gamma_{X,n}, \mathbb{E}[Y]\gamma_{Y,n})\Big),$$

using $V \neq \mathbf{0}$ and the boundedness of $\mathbb{E}\big[|X_{1,n}|^3\big]\gamma_{X,n}^3$ and $\mathbb{E}\big[|Y_{1,n}|^3\big]\gamma_{Y,n}^3$, we obtain $(h_{X,n}, h_{Y,n}) \xrightarrow[n\to+\infty]{d} \mathcal{N}(0,V)$. We also obtain $(h_{X,n}, h_{Y,n}, W_n) \xrightarrow[n\to+\infty]{d} \mathcal{N}(0,V)\otimes\delta_0$ by Slutsky's Lemma. We can therefore apply Skorokhods's almost sure representation theorem, see (Van der Vaart, 2000, Theorem 2.19). It means that there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{U}}, \tilde{\mathbb{P}})$, a sequence of random vectors $(\tilde{h}_{X,n}, \tilde{h}_{Y,n}, \tilde{W}_n)$ such that for every $n \geq 1$, $(\tilde{h}_{X,n}, \tilde{h}_{Y,n}, \tilde{W}_n) \stackrel{d}{=} (h_{X,n}, h_{Y,n}, W_n)$, and a random vector $(\tilde{h}_{X,\infty}, \tilde{h}_{Y,\infty}, \tilde{W}_\infty)$ following the distribution $\mathcal{N}(0,V)\otimes\delta_0$ such that $(\tilde{h}_{X,n}, \tilde{h}_{Y,n}, \tilde{W}_n)$ $\xrightarrow[n\to+\infty]{a.s.} (\tilde{h}_{X,\infty}, \tilde{h}_{Y,\infty}, \tilde{W}_\infty)$, where the convergence is to be seen as of a sequence of random vectors defined on $(\tilde{\Omega}, \tilde{\mathcal{U}}, \tilde{\mathbb{P}})$. Let us define

$$\tilde{A}_n := \frac{\theta_{X,n} + \tilde{h}_{X,n}/(\sqrt{n}\gamma_{X,n})}{\theta_{Y,n} + \tilde{h}_{Y,n}/(\sqrt{n}\gamma_{Y,n})} - \frac{\theta_{X,n}}{\theta_{Y,n}} \stackrel{d}{=} \frac{\theta_{X,n} + h_{X,n}/(\sqrt{n}\gamma_{X,n})}{\theta_{Y,n} + h_{Y,n}/(\sqrt{n}\gamma_{Y,n})} - \frac{\theta_{X,n}}{\theta_{Y,n}}$$

$$= \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\theta_{X,n}}{\theta_{Y,n}} = A_n.$$

Moreover, we have $\tilde{W}_n = \mathbb{1}\{\theta_{Y,n} + \tilde{h}_{Y,n}/(\sqrt{n}\gamma_{Y,n}) = 0\}$ and $\tilde{W}_\infty = 0$ almost surely. We can define

$$\tilde{\Omega}^* = \{\tilde{\omega} \in \tilde{\Omega} : \tilde{W}_n(\tilde{\omega}) \to 0 \text{ and } \exists N > 0, \forall n \geq N, \tilde{h}_{Y,n}(\tilde{\omega}) \neq 0\}.$$

By the almost sure convergence of $(\tilde{h}_{Y,n}, \tilde{W}_n)$, we get $\tilde{\mathbb{P}}(\tilde{\Omega}^*) = 1$, and for every $\tilde{\omega} \in \tilde{\Omega}^*$, $\tilde{W}_n(\tilde{\omega}) = 0$ and $\tilde{h}_{Y,n}(\tilde{\omega}) \neq 0$ for every $n$ large enough. This means that for every given $\tilde{\omega} \in \tilde{\Omega}^*$, and for every $n$ large enough, $\tilde{A}_n$ is well-defined. In the rest of the proof, we will fix such a $\tilde{\omega} \in \tilde{\Omega}^*$, so that all random variables may be considered as deterministic. By the almost sure representation theorem, this means that the equivalents and limits that will be obtained will still be valid in law in the original spaces $\Omega_n$.

**First case:** We have

$$
\begin{aligned}
\tilde{A}_n &= \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\theta_{X,n}}{\theta_{Y,n}} = \frac{\theta_{X,n} + \tilde{h}_{X,n}/(\sqrt{n}\gamma_{X,n})}{\theta_{Y,n} + \tilde{h}_{Y,n}/(\sqrt{n}\gamma_{Y,n})} - \frac{\theta_{X,n}}{\theta_{Y,n}} \\
&= \frac{\theta_{X,n} + \tilde{h}_{X,n}/(\sqrt{n}\gamma_{X,n})}{\theta_{Y,n}} \left(1 - \frac{\tilde{h}_{Y,n}}{\sqrt{n}\gamma_{Y,n}\theta_{Y,n}} + O\big((\sqrt{n}\gamma_{Y,n}\theta_{Y,n})^{-2}\big)\right) - \frac{\theta_{X,n}}{\theta_{Y,n}} \\
&\sim \frac{-\theta_{X,n}\tilde{h}_{Y,n}}{\sqrt{n}\gamma_{Y,n}\theta_{Y,n}^2} + \frac{\tilde{h}_{X,n}}{\sqrt{n}\gamma_{X,n}\theta_{Y,n}},
\end{aligned}
$$

as claimed.

**Second case:** We have

$$
\begin{aligned}
\tilde{A}_n &\sim \frac{\theta_{X,n} + \tilde{h}_{X,n}/(\sqrt{n}\gamma_{X,n})}{C/(\sqrt{n}\gamma_{Y,n}) + \tilde{h}_{Y,n}/(\sqrt{n}\gamma_{Y,n})} - \frac{\theta_{X,n}}{C/(\sqrt{n}\gamma_{Y,n})} \\
&= \frac{\sqrt{n}\gamma_{Y,n}\theta_{X,n} + \tilde{h}_{X,n}\gamma_{Y,n}/\gamma_{X,n}}{C + \tilde{h}_{Y,n}} - \frac{\sqrt{n}\gamma_{Y,n}\theta_{X,n}}{C}.
\end{aligned}
$$

We factorize by $\theta_{X,n}$ in the latter expression, which completes the proof.

**Third case:** We have

$$
\begin{aligned}
\tilde{A}_n &= \frac{\theta_{X,n} + \tilde{h}_{X,n}/(\sqrt{n}\gamma_{X,n})}{\theta_{Y,n} + \tilde{h}_{Y,n}/(\sqrt{n}\gamma_{Y,n})} - \frac{\theta_{X,n}}{\theta_{Y,n}} = \frac{\theta_{X,n} + \tilde{h}_{X,n}/(\sqrt{n}\gamma_{X,n})}{\big(\tilde{h}_{Y,n} + o(1)\big)/(\sqrt{n}\gamma_{Y,n})} - \frac{\theta_{X,n}}{\theta_{Y,n}} \\
&\sim \frac{\sqrt{n}\theta_{X,n}\gamma_{Y,n}}{\tilde{h}_{Y,n}} + \frac{\tilde{h}_{X,n}\gamma_{Y,n}}{\tilde{h}_{Y,n}\gamma_{X,n}} - \frac{\theta_{X,n}}{\theta_{Y,n}} \\
&\sim \theta_{X,n}\left(\frac{\sqrt{n}\gamma_{X,n}}{\tilde{h}_{Y,n}} - \frac{1}{\theta_{Y,n}}\right) + \frac{\tilde{h}_{X,n}\gamma_{Y,n}}{\tilde{h}_{Y,n}\gamma_{X,n}},
\end{aligned}
$$

and the result follows from the fact that $\sqrt{n}\gamma_{X,n}/\tilde{h}_{Y,n}$ is negligible compared to $1/\theta_{Y,n}$.

$\square$

## B.2 Proof of Theorem 3.3

For $b = 1, 2$, let $h_{X,n} := \sqrt{n}\gamma_{X,n}(\overline{X}_n - \theta_{X,n})$ (resp. $h^{Y,n}$), $S_n := (h_{X,n}, h_{Y,n})'$ and $S_n^{(b)} := (h_{X,n}^{(b)}, h_{Y,n}^{(b)})'$, where $h_{X,n}^{(b)} := \sqrt{n}\gamma_{X,n}(\overline{X}_n^{(b)} - \overline{X}_n)$ is the $b$-th bootstrap replication of $h_{X,n}$ (resp. $h_{Y,n}^{(b)}$).

**Lemma B.2.** *We have* $d_{BL}\left(P_{S_n^{(1)}\,|\,(X_{i,n},Y_{i,n})_{i=1}^n}, \mathcal{N}(0,V)\right) \xrightarrow[n\to+\infty]{a.s.} 0.$

By the Central Limit Theorem, we have $S_n \xrightarrow[n \to +\infty]{d} S$ with $S \sim \mathcal{N}(0, V)$ and by Lemma B.2 (proved in Section B.2.1) and the triangle inequality, we get $d_{BL}\left(P_{S_n^{(1)}|(X_l, Y_l)_{l=1}^n}, P_{S_n}\right) \xrightarrow[n \to +\infty]{P} 0$. Combining both results, Lemma 2.2 in Bücher and Kojadinovic (2019) gives us

$$d_{BL}\left(P_{(S_n, S_n^{(1)}, S_n^{(2)})}, P_S^{\otimes 3}\right) \to 0.$$

Let us define $W_n := \mathbb{1}\{\theta_{Y,n} + h_{Y,n}/(\sqrt{n}\gamma_{Y,n}) = 0\}$ and remark that $W_n = 1$ whenever $\overline{Y}_n = 0$. By assumption $\mathbb{P}(\overline{Y}_n = 0) \to 0$, therefore we have $W_n \xrightarrow[n \to +\infty]{d} \delta_0$. We define also $W_n^{(b)} := \mathbb{1}\{\overline{Y}_n^{(b)}\} = \mathbb{1}\{\overline{Y}_n + h_{Y,n}^{(b)}/(\sqrt{n}\gamma_{Y,n}) = 0\}$, so that $W_n^{(1)} = 1$ whenever $\overline{Y}_n^{(b)} = 0$. In the same way as previously, $W_n^{(b)} \xrightarrow[n \to +\infty]{d} \delta_0$ holds by assumption. Let $Z_n = (S_n, W_n, S_n^{(1)}, W_n^{(1)}, S_n^{(1)}, W_n^{(2)})$ be a random vector of size 9, and let $Z$ be a random vector of size 9 following $(P_S \otimes \delta_0)^{\otimes 3}$.

By Slutsky's lemma, we have $d_{BL}(P_{Z_n}, P_Z) \to 0$ with our new notation. Using Skorokhods's almost sure representation theorem (Van der Vaart, 2000, Theorem 2.19), there exists a probability space $\Omega^+$, a sequence of random vectors $Z_n^+ \in \mathbb{R}^9$ and a vector $Z^+$ defined on $\Omega^+$ such that $Z_n^+ \xrightarrow[n \to +\infty]{a.s.} Z^+$, $Z_n \stackrel{d}{=} Z_n^+$ and $Z \stackrel{d}{=} Z^+$. Let us use the notation

$$
\begin{aligned}
Z_n^+ &= \left(S_n^+, W_n^+, S_n^{(1)+}, W_n^{(1)+}, S_n^{(1)+}, W_n^{(2)+}\right) \\
&= \left(h_{X,n}^+, h_{Y,n}^+, W_n^+, h_{X,n}^{(1)+}, h_{Y,n}^{(1)+}, W_n^{(1)+}, h_{X,n}^{(2)+}, h_{Y,n}^{(2)+}, W_n^{(2)+}\right) \\
\text{and } Z^+ &= \left(Z_1^+, Z_2^+, Z_3^+\right),
\end{aligned}
$$

where $S_n^+, S_n^{(1)+}, S_n^{(2)+}$ are random vectors of dimension 2 and $Z_1^+, Z_2^+, Z_3^+$ are random vectors of dimension 3. We define

$$
\begin{aligned}
A_n &:= \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\theta_{X,n}}{\theta_{Y,n}} = \frac{\theta_{X,n} + h_{X,n}/(\sqrt{n}\gamma_{X,n})}{\theta_{Y,n} + h_{Y,n}/(\sqrt{n}\gamma_{Y,n})} - \frac{\theta_{X,n}}{\theta_{Y,n}} \\
A_n^{(b)} &:= \frac{\overline{X}_n + h_{X,n}^{(b)}/(\sqrt{n}\gamma_{X,n})}{\overline{Y}_n + h_{Y,n}^{(b)}/(\sqrt{n}\gamma_{Y,n})} - \frac{\overline{X}_n}{\overline{Y}_n} \\
&= \frac{\theta_{X,n} + h_{X,n}/(\sqrt{n}\gamma_{X,n}) + h_{X,n}^{(b)}/(\sqrt{n}\gamma_{X,n})}{\theta_{Y,n} + h_{Y,n}/(\sqrt{n}\gamma_{Y,n}) + h_{Y,n}^{(b)}/(\sqrt{n}\gamma_{Y,n})} - \frac{\theta_{X,n} + h_{X,n}/(\sqrt{n}\gamma_{X,n})}{\theta_{Y,n} + h_{Y,n}/(\sqrt{n}\gamma_{Y,n})},
\end{aligned}
$$

and respectively their counterparts $A_n^+$ and $A_n^{(b)+}$ defined on $\Omega^+$. The following lemma, proved in Section B.2.2, ensures the existence of an event of probability 1 on which every quantity is well-defined.

**Lemma B.3.** *There exists an event $\tilde{\Omega} \subset \Omega^+$ such that $\mathbb{P}(\tilde{\Omega}) = 1$ and such that for every $\omega \in \tilde{\Omega}$, and for all $n$ large enough, $h_{Y,n}^+(\omega) \neq 0$, $h_{Y,n}^{(1)+}(\omega) \neq 0$, $h_{Y,n}^{(2)+}(\omega) \neq 0$ and $A_n^+(\omega)$, $A_n^{(1)+}(\omega)$ and $A_n^{(2)+}(\omega)$ are well-defined.*

In the next step, we fix $\omega \in \tilde{\Omega}$ and let $C := \lim_{n \to +\infty} \theta_{X,n}\gamma_{X,n}/\theta_{Y,n}\gamma_{Y,n}$ and

$$\sigma_n := \sqrt{n}\theta_{Y,n}\left(\gamma_{X,n}\mathbb{1}_{\{C \in \mathbb{R}\}} + \gamma_{Y,n}\theta_{Y,n}/\theta_{X,n}\mathbb{1}_{\{|C|=+\infty\}}\right).$$

We restrict ourselves to the case $n^{1/2}\gamma_{Y,n}\theta_{Y,n} \to +\infty$. Theorem 3.1 therefore yields

$$\sigma_n A_n^+(\omega) = \begin{cases} -Ch_{Y,n}^+(\omega) + h_{X,n}^+(\omega) + o(1) \text{ if } C \in \mathbb{R} \\ -h_{Y,n}^+(\omega) + o(1) \text{ else.} \end{cases} \tag{2}$$

Furthermore, the same tools as those used in the proof of Theorem 3.1 plus the fact that $\theta_{Y,n} + h_{Y,n}^+(\omega)/(\sqrt{n}\gamma_{Y,n}) \sim \theta_{Y,n}$ imply

$$\sigma_n A_n^{(b)+}(\omega)$$

$$\sim \sigma_n \left( \frac{-\left(\theta_{X,n} + h_{X,n}^+(\omega)/(\sqrt{n}\gamma_{X,n})\right)}{\sqrt{n}\gamma_{Y,n}\left(\theta_{Y,n} + h_{Y,n}^+(\omega)/(\sqrt{n}\gamma_{Y,n})\right)^2} h_{Y,n}^{(b)+}(\omega) \right.$$

$$\left. + \frac{1}{\sqrt{n}\gamma_{Y,n}\left(\theta_{Y,n} + h_{Y,n}^+(\omega)/(\sqrt{n}\gamma_{X,n})\right)} h_{X,n}^{(b)+}(\omega) \right)$$

$$\sim \sigma_n \left( \frac{-\left(\theta_{X,n} + h_{X,n}^+(\omega)/(\sqrt{n}\gamma_{Y,n})\right)}{\sqrt{n}\gamma_{Y,n}\theta_{Y,n}^2} h_{Y,n}^{(b)+}(\omega) + \frac{1}{\sqrt{n}\gamma_{X,n}\theta_{Y,n}} h_{X,n}^{(b)+}(\omega) \right).$$

We can also remark that when $\theta_{X,n} + h_{X,n}^+(\omega)/(\sqrt{n}\gamma_{X,n}) \sim \theta_{X,n}$

$$\sigma_n A_n^{(b)+}(\omega) = \begin{cases} -Ch_{Y,n}^{(b)+}(\omega) + h_{X,n}^{(b)+}(\omega) + o(1) \text{ if } C \in \mathbb{R} \\ -h_{Y,n}^{(b)+}(\omega) + o(1) \text{ else.} \end{cases} \tag{3}$$

When $\theta_{X,n} + h_{X,n}^+(\omega)/(\sqrt{n}\gamma_{X,n}) = O\left(h_{X,n}^+(\omega)/(\sqrt{n}\gamma_{X,n})\right)$, we have $C = 0$ and we find again that

$$\sigma_n A_n^{(b)+}(\omega) = h_{X,n}^{(b)+}(\omega) + o(1). \tag{4}$$

Let $D_n^+ := \left(-Ch_{Y,n}^+ + h_{X,n}^+\right)\mathbb{1}_{\{|C|<+\infty\}} - h_{Y,n}^+\mathbb{1}_{\{|C|=+\infty\}}$ (resp. $D_n$, $D_n^{(b)}$ and $D_n^{(b)+}$), which corresponds to the dominant terms in Equations (2), (3) and (4) above. By construction of $Z_n^+$ and $Z_n$, we have $Z_n^+ \xrightarrow[n \to +\infty]{a.s.} Z^+$, so that the continuous mapping theorem ensures that $\left(D_n^+, D_n^{(1)+}, D_n^{(2)+}\right) \xrightarrow[n \to +\infty]{a.s.} (U_1, U_2, U_3)$, where

for every $i \in \{1, 2, 3\}$, we define $U_i^+ := \left( -C Z_{i,2}^+ + Z_{i,1}^+ \right) \mathbb{1}_{\{C \in \mathbb{R}\}} - Z_{i,2}^+ \mathbb{1}_{\{|C|=+\infty\}}$ where $Z_{i,1}^+$ (resp. $Z_{i,2}^+$) is the first (resp. second) component of the vector $Z_i^+$. Combining the triangle inequality, Equations (2), (3) and (4), we get

$$\left( \sigma_n A_n^+, \sigma_n A_n^{(1)+}, \sigma_n A_n^{(2)+} \right) \xrightarrow[n \to +\infty]{a.s.} \left( U_1^+, U_2^+, U_3^+ \right).$$

Using the fact that for all $n \in \mathbb{N}$ $\left( A_n, A_n^{(1)}, A_n^{(2)} \right) \overset{d}{=} \left( A_n^+, A_n^{(1)+}, A_n^{(2)+} \right)$, we obtain

$$\left( \sigma_n A_n, \sigma_n A_n^{(1)}, \sigma_n A_n^{(2)} \right) \xrightarrow[n \to +\infty]{d} (U_1^+, U_2^+, U_3^+).$$

Therefore, $d_{BL}\left( P_{\left( \sigma_n A_n, \sigma_n A_n^{(1)}, \sigma_n A_n^{(2)} \right)}, P_{U_1^+}^{\otimes 3} \right) \to 0$ as $n \to +\infty$ and $\sigma_n A_n \xrightarrow[n \to +\infty]{d} U_1^+$. Applying Lemma 2.2 of Bücher and Kojadinovic (2019), we can conclude that

$$d_{BL}\left( P_{\sigma_n A_n^{(1)} \mid (X_{i,n}, Y_{i,n})_{i=1}^n}, P_{U_1^+} \right) \xrightarrow[n \to +\infty]{P} 0.$$

The conclusion follows from Lemma 23.3 in Van der Vaart (2000).

$\square$

### B.2.1 Proof of Lemma B.2

Let $t = (t_X, t_Y)' \in \mathbb{R}^2$, and denote $T_{i,n} = t_X \gamma_{X,n} X_{i,n} + t_Y \gamma_{Y,n} Y_{i,n}$ for $i = 1, \ldots, n$ and $T_{i,n}^{(1)}$ its bootstrap counterpart. Let also $V_{T_{1,n}} := t' \mathbb{V}[(\gamma_{X,n} X_{1,n}, \gamma_{Y,n} Y_{1,n})] t$ and $V_T := t' V t$. We start by showing that for every $t \in \mathbb{R}^2$, $P_{\sqrt{n}\left( \overline{T}_n^{(1)} - \overline{T}_n \right) \mid (X_{i,n}, Y_{i,n})_{i=1}^n}$ converges weakly to $P_T = \mathcal{N}(0, V_T)$ almost surely conditionally on $(X_{i,n}, Y_{i,n})_{i=1}^n$ in the sense of the Lévy criterion for weak convergence, i.e.

$$\left| \mathbb{E}\left[ e^{iu\sqrt{n}\left( \overline{T}_n^{(1)} - \overline{T}_n \right)} \mid (X_{i,n}, Y_{i,n})_{i=1}^n \right] - e^{u^2 V_T / 2} \right| \xrightarrow[n \to +\infty]{a.s.} 0 \quad \forall u \in \mathbb{R}. \qquad (5)$$

To do so, we have to check the steps of the proof of Theorem 23.4 in Van der Vaart (2000). We have

$$\mathbb{E}\left[ \overline{T}_n^{(1)} \mid (X_{i,n}, Y_{i,n})_{i=1}^n \right] = \overline{T}_n \quad \text{and}$$

$$\mathbb{E}\left[ \left( T_{i,n}^{(1)} - \overline{T}_n \right)^2 \mid (X_{i,n}, Y_{i,n})_{i=1}^n \right] = \frac{1}{n} \sum_{i=1}^n T_{i,n}^2 - \overline{T}_n^2.$$

The first requirement is to ensure almost sure convergence to 0 of both quan-

tities $\left|\overline{T}_n - \mathbb{E}\left[T_{1,n}\right]\right|$ and $\left|\frac{1}{n}\sum_{i=1}^n T_{i,n}^2 - \overline{T}_n^2 - V_T\right|$. Under the assumption that $\sup_{n \in \mathbb{N}^*} \mathbb{E}\left[|T_{1,n}|^{4+\delta}\right] < +\infty$, observe that all the conditions of Theorem 2.2 in Gut (1992) are satisfied with $p = 1$. We can thus conclude that $\left|\overline{T}_n - \mathbb{E}\left[T_{1,n}\right]\right| \xrightarrow[n \to +\infty]{a.s.} 0$ and $\left|\frac{1}{n}\sum_{i=1}^n T_{i,n}^2 - \mathbb{E}\left[T_{1,n}^2\right]\right| \xrightarrow[n \to +\infty]{a.s.} 0$. Now using the fact that

$$\left|\frac{1}{n}\sum_{i=1}^n T_{i,n}^2 - \overline{T}_n^2 - V_T\right| \leq \left|\frac{1}{n}\sum_{i=1}^n T_{i,n}^2 - \mathbb{E}\left[T_{1,n}^2\right]\right| + \left|\overline{T}_n - \mathbb{E}\left[T_{1,n}\right]\right|^2$$
$$+ 2\left|\mathbb{E}\left[T_{1,n}\right]\left(\overline{T}_n - \mathbb{E}\left[T_{1,n}\right]\right)\right| + \left|\mathbb{E}\left[T_{1,n}^2\right] - \mathbb{E}\left[T_{1,n}\right]^2 - V_T\right|,$$

as well as $|\mathbb{E}\left[T_{1,n}\right]| = O(1)$ and $\|\mathbb{V}\left[(\gamma_{X,n}X_{1,n}, \gamma_{Y,n}Y_{1,n})\right] - V\| = o(1)$, to conclude that $\left|\frac{1}{n}\sum_{i=1}^n T_{i,n}^2 - \overline{T}_n^2 - V_T\right| \xrightarrow[n \to +\infty]{a.s.} 0$.

The second requirement is to check the Lindeberg condition for the bootstrap which writes

$$\mathbb{E}\left[\left|T_{1,n}^{(1)}\right|^2 \mathbb{1}\left\{\left|T_{1,n}^{(1)}\right|^2 > \epsilon\sqrt{n}\right\} \mid (X_{i,n}, Y_{i,n})_{i=1}^n\right] = \frac{1}{n}\sum_{i=1}^n |T_{i,n}|^2 \mathbb{1}\left\{|T_{i,n}|^2 > \epsilon\sqrt{n}\right\}$$
$$\xrightarrow[n \to +\infty]{a.s.} 0 \quad \forall \epsilon > 0.$$

Let $M : \epsilon \mapsto M(\epsilon)$ be some function of $\epsilon$ to be defined later that does not depend on $n$ and satisfies $0 < M(\epsilon) < +\infty \; \forall \epsilon > 0$. For such a function, there exists for every $\epsilon > 0$, a $n_\epsilon$ such that for every $n > n_\epsilon$,

$$\frac{1}{n}\sum_{i=1}^n |T_{i,n}|^2 \mathbb{1}\left\{|T_{i,n}|^2 > \epsilon\sqrt{n}\right\} \leq \frac{1}{n}\sum_{i=1}^n |T_{i,n}|^2 \mathbb{1}\left\{|T_{i,n}|^2 > M(\epsilon)\right\} \; a.s.$$

By the triangle inequality,

$$\frac{1}{n}\sum_{i=1}^n |T_{i,n}|^2 \mathbb{1}\left\{|T_{i,n}|^2 > M(\epsilon)\right\} \leq$$
$$\left|\frac{1}{n}\sum_{i=1}^n |T_{i,n}|^2 \mathbb{1}\left\{|T_{i,n}|^2 > M(\epsilon)\right\} - \mathbb{E}\left[|T_{1,n}|^2 \mathbb{1}\left\{|T_{1,n}|^2 > M(\epsilon)\right\}\right]\right|$$
$$+ \mathbb{E}\left[|T_{1,n}|^2 \mathbb{1}\left\{|T_{1,n}|^2 > M(\epsilon)\right\}\right].$$

The first term in the upper bound converges to 0 almost surely for every $\epsilon > 0$ under the assumption $\sup_{n \in \mathbb{N}^*} \mathbb{E}\left[|T_{1,n}|^{4+\delta}\right] < +\infty$ thanks to Theorem 2.2 in Gut (1992). The second term in the upper bound can be bounded with the Cauchy-

Schwarz and Markov inequalities

$$\mathbb{E}\left[|T_{1,n}|^2 \mathbb{1}\left\{|T_{1,n}|^2 > M(\epsilon)\right\}\right] \leq \frac{\sup_{n\in\mathbb{N}^*}\sqrt{\mathbb{E}\left[|T_{1,n}|^4\right]\mathbb{E}\left[|T_{1,n}|\right]}.}{\sqrt{M(\epsilon)}}$$

Picking $M(\epsilon) = \epsilon^{-1}\sup_{n\in\mathbb{N}^*}\mathbb{E}\left[|T_{1,n}|^4\right]\mathbb{E}\left[|T_{1,n}|\right]$, we get that for every $\epsilon > 0$

$$\limsup_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}|T_{i,n}|^2 \mathbb{1}\left\{|T_{i,n}|^2 > \epsilon\sqrt{n}\right\} \leq \epsilon \quad a.s.$$

Letting $\epsilon$ go to 0, we see that the Lindeberg condition is satisfied. This entails that (5) is satisfied.

Arguments underpinning the Cramer-Wold device are valid as well so that we can claim that for every $t \in \mathbb{R}^2$

$$\left|\mathbb{E}\left[e^{it'S_n^{(1)}} \mid (X_{i,n}, Y_{i,n})_{i=1}^n\right] - e^{-t'Vt/2}\right| \xrightarrow[n\to+\infty]{a.s.} 0, \tag{6}$$

where $S_n^{(1)} := \sqrt{n}\left(\gamma_{X,n}\left(\overline{X}_n^{(1)} - \overline{X}_n\right), \gamma_{Y,n}\left(\overline{Y}_n^{(1)} - \overline{Y}_n\right)\right)$.

Let $\Omega$ be the set of probability one on which (6) occurs. For every $\omega \in \Omega$,

$$\left(P_{S_n^{(1)}\mid(X_{i,n},Y_{i,n})_{i=1}^n=(X_{i,n}(\omega),Y_{i,n}(\omega))_{i=1}^n}\right)_{n\in\mathbb{N}^*}$$

is a sequence of nonrandom probability measures for which all weak convergence criteria are equivalent. In particular, for every $\omega \in \Omega$, the validity of the Lévy criterion due to (6) ensures that

$$d_{BL}\left(P_{S_n^{(1)}\mid(X_{i,n},Y_{i,n})_{i=1}^n=(X_{i,n}(\omega),Y_{i,n}(\omega))_{i=1}^n}, \mathcal{N}(0,V)\right) = o(1).$$

This is enough to conclude.

$\square$

### B.2.2 Proof of Lemma B.3

The vector $\left(W_n^+, W_n^{(1)+}, W_n^{(2)+}\right)$ converges almost surely to $(0,0,0)$. As a consequence, there exists an event $\tilde{\Omega}^1$ of probability 1 such that $\forall\omega \in \tilde{\Omega}^1$, $\left(W_n^+(\omega), W_n^{(1)+}(\omega), W_n^{(2)+}(\omega)\right) = (0,0,0)$ for $n$ large enough. As $\left(h_{Y,n}^+, h_{Y,n}^{(1)+}, h_{Y,n}^{(2)+}\right)$ converges almost surely to a continuous vector, there exists an event $\tilde{\Omega}^2$ of probability

1 such that $\forall \omega \in \tilde{\Omega}^2$, the components of $\left(h_{Y,n}^+(\omega), h_{Y,n}^{(1)+}(\omega), h_{Y,n}^{(2)+}(\omega)\right)$ are all non-zero for $n$ large enough. We finally define $\tilde{\Omega} := \tilde{\Omega}^1 \cap \tilde{\Omega}^2$, which is of probability 1 and satisfies the stated conditions. $\square$

## B.3  Proof of Example 3.4

We have

$$
\begin{aligned}
\mathbb{P}(\overline{Y}_n^{(1)} = 0) &= \mathbb{E}\left[\mathbb{P}\big(\overline{Y}_n^{(1)} = 0 \,\big|\, (Y_{i,n})_{i=1}^n\big)\right] \\
&= \mathbb{E}\left[\mathbb{P}\big(Y_{1,n}^{(1)} = 0, \ldots, Y_{n,n}^{(1)} = 0 \,\big|\, (Y_{i,n})_{i=1}^n\big)\right] \\
&= \mathbb{E}\left[\mathbb{P}\big(Y_{1,n}^{(1)} = 0 \,\big|\, (Y_{i,n})_{i=1}^n\big)^n\right] = \mathbb{E}\left[\big(S_n/n\big)^n\right],
\end{aligned}
$$

where $S_n := \sum_{i=1}^n (1 - Y_{i,n}) \sim Bin(n, 1 - p_n)$. Therefore, for any $x > 0$,

$$
\begin{aligned}
\mathbb{P}(\overline{Y}_n^{(1)} = 0) &= \sum_{k=1}^n (k/n)^n \mathbb{P}\big[S_n = k\big] \\
&\leq \sum_{k=1}^{\lfloor n(1-p_n)+x \rfloor} (k/n)^n \mathbb{P}\big[S_n = k\big] + \mathbb{P}\big[S_n \geq n(1-p_n) + x\big] \\
&\leq \left(\frac{n(1-p_n)+x}{n}\right)^n + \mathbb{P}\big[S_n \geq n(1-p_n) + x\big] \\
&\leq \big(1 - p_n + x/n\big)^n + \mathbb{P}\big[S_n - n(1-p_n) \geq x\big].
\end{aligned}
$$

Let $\tilde{S}_n := \big(S_n - n(1-p_n)\big)/\sqrt{np_n(1-p_n)} = O_P(1)$ be the renormalized version of $S_n$ and choose $x = n^a \sqrt{np_n(1-p_n)}$ for $a = (1-b)/3 > 0$. Then

$$
\begin{aligned}
\mathbb{P}(\overline{Y}_n^{(1)} = 0) &\leq \big(1 - p_n + n^a \sqrt{p_n(1-p_n)/n}\big)^n + \mathbb{P}\big[\tilde{S}_n \geq n^a\big] \\
&\leq \exp\Big(n \ln\big(1 - p_n + n^a \sqrt{p_n(1-p_n)/n} + o(p_n)\big)\Big) + o(1) \\
&\leq \exp\Big(n\big(n^{a-b/2-1/2} - n^{-b} + o(n^{-b})\big)\Big) + o(1) \\
&\leq \exp\Big(n^{1/3-b/3-b/2+1/2} - n^{-b+1} + o(n^{-b+1})\Big) + o(1) \\
&\leq \exp\Big(n^{(1-b)5/6} - n^{1-b} + o(n^{-b+1})\Big) + o(1) \\
&\leq \exp\big(-n^{1-b}\big) + o(1) = o(1),
\end{aligned}
$$

which completes the proof. $\square$

## B.4 Proof of Theorem 4.1

We fix arbitrary $n \in \mathbb{N}^*$ and $\varepsilon \in \mathbb{R}_+^*$. Combining the triangle inequality, the bound $|\overline{X}_n| \leq |\overline{X}_n - \mathbb{E}[X_{1,n}]| + |\mathbb{E}[X_{1,n}]|$ and Assumptions 1 to 3, we get

$$
\left| \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]} \right| \leq |\overline{X}_n| \cdot \left| \frac{1}{\overline{Y}_n} - \frac{1}{\mathbb{E}[Y_{1,n}]} \right| + \frac{1}{\mathbb{E}[Y_{1,n}]} \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right|
$$
$$
\leq \frac{\left( \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right| + \sqrt{u_{X,n}} \right) \left| \overline{Y}_n - \mathbb{E}[Y_{1,n}] \right|}{a_{Y,n} l_{Y,n}} + \frac{\left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right|}{l_{Y,n}}.
$$

Consequently, the event considered in Theorem 4.1 is included in the event

$$
\frac{\left( \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right| + \sqrt{u_{X,n}} \right) \left| \overline{Y}_n - \mathbb{E}[Y_{1,n}] \right|}{a_{Y,n} l_{Y,n}} + \frac{\left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right|}{l_{Y,n}}
$$
$$
> \frac{\left( \varepsilon + \sqrt{u_{X,n}} \right) \varepsilon}{a_{Y,n} l_{Y,n}} + \frac{\varepsilon}{l_{Y,n}}. \quad (7)
$$

If both $|\overline{X}_n - \mathbb{E}[X_{1,n}]|$ and $|\overline{Y}_n - \mathbb{E}[Y_{1,n}]|$ are inferior or equal to $\varepsilon$, event (7) cannot happen. By contraposition, we obtain:

$$
\mathbb{P}\left( \frac{\left( \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right| + \sqrt{u_{X,n}} \right) \left| \overline{Y}_n - \mathbb{E}[Y_{1,n}] \right|}{a_{Y,n} l_{Y,n}} + \frac{\left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right|}{l_{Y,n}} \right.
$$
$$
\left. > \frac{\left( \varepsilon + \sqrt{u_{X,n}} \right) \varepsilon}{a_{Y,n} l_{Y,n}} + \frac{\varepsilon}{l_{Y,n}} \right)
$$
$$
\leq \mathbb{P}\left( \left\{ \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right| > \varepsilon \right\} \cup \left\{ \left| \overline{Y}_n - \mathbb{E}[Y_{1,n}] \right| > \varepsilon \right\} \right)
$$
$$
\leq \mathbb{P}\left( \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right| > \varepsilon \right) + \mathbb{P}\left( \left| \overline{Y}_n - \mathbb{E}[Y_{1,n}] \right| > \varepsilon \right),
$$

where we use the union bound for the last inequality. The first conclusion follows from using twice Bienaymé-Chebyshev's inequality applied to the variables $\overline{X}_n$ and $\overline{Y}_n$ and the fact that under Assumptions 1 and 2 and Jensen's inequality, $\mathbb{V}[X_{1,n}] \leq u_{X,n}$ and $\mathbb{V}[Y_{1,n}] \leq u_{Y,n} - l_{Y,n}^2$. The second conclusion follows from solving $(u_{X,n} + u_{Y,n} - l_{Y,n}^2)/(n\varepsilon^2) = \alpha$.

$\square$

## B.5 Proof of Theorem 4.2

We start by introducing and proving an intermediate lemma that is also used to prove Theorem C.2. For a random variable $U$, $\varepsilon > 0$, and $\tilde{\varepsilon} \in (0,1)$ we define

the following events:

$$A_\varepsilon^U := \left\{ \left| \overline{U}_n - \mathbb{E}[U] \right| \leq \varepsilon \right\}, \text{ and } \tilde{A}_{\tilde{\varepsilon}}^U := \left\{ \left| \overline{U}_n - \mathbb{E}[U] \right| \leq \tilde{\varepsilon} \left| \mathbb{E}[U] \right| \right\}.$$

**Lemma B.4.** *Assume that Assumption 1 holds. Then for every $n \in \mathbb{N}^*$, $\varepsilon > 0$ and $\tilde{\varepsilon} \in (0,1)$, we have*

$$\mathbb{P}\left( \left| \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]} \right| > \left( \frac{(|\mathbb{E}[X_{1,n}]| + \varepsilon)\,\tilde{\varepsilon}}{(1 - \tilde{\varepsilon})^2} + \varepsilon \right) \frac{1}{|\mathbb{E}[Y_{1,n}]|} \right)$$
$$\leq 1 - \mathbb{P}\left( A_\varepsilon^{X_{1,n}} \right) + 1 - \mathbb{P}\left( \tilde{A}_{\tilde{\varepsilon}}^{Y_{1,n}} \right).$$

We fix arbitrary $n \in \mathbb{N}^*$, $\varepsilon > 0$ and $\tilde{\varepsilon} \in (0,1)$. By Lemma B.4, we have

$$\mathbb{P}\left( \left| \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]} \right| > \left( \frac{(|\mathbb{E}[X_{1,n}]| + \varepsilon)\,\tilde{\varepsilon}}{(1 - \tilde{\varepsilon})^2} + \varepsilon \right) \frac{1}{|\mathbb{E}[Y_{1,n}]|} \right)$$
$$\leq 1 - \mathbb{P}\left( \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right| \leq \varepsilon \right) + 1 - \mathbb{P}\left( \left| \overline{Y}_n - \mathbb{E}[Y_{1,n}] \right| \leq \tilde{\varepsilon} \left| \mathbb{E}[Y_{1,n}] \right| \right).$$

Using Jensen's inequality and Assumption 2, we have $|\mathbb{E}[X_{1,n}]| \leq (u_{X,n})^{1/2}$, and Assumption 1 entails $1/|\mathbb{E}[Y_{1,n}]| \leq 1/l_{Y,n}$. Consequently, we get

$$\mathbb{P}\left( \left| \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]} \right| > \left( \frac{\left( \sqrt{u_{X,n}} + \varepsilon \right) \tilde{\varepsilon}}{(1 - \tilde{\varepsilon})^2} + \varepsilon \right) \frac{1}{l_{Y,n}} \right)$$
$$\leq 1 - \mathbb{P}\left( \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right| \leq \varepsilon \right) + 1 - \mathbb{P}\left( \left| \overline{Y}_n - \mathbb{E}[Y_{1,n}] \right| \leq \tilde{\varepsilon} \left| \mathbb{E}[Y_{1,n}] \right| \right).$$

Using Bienaymé-Chebyshev's inequality twice gives the bounds

$$1 - \mathbb{P}\left( \left| \overline{X}_n - \mathbb{E}[X_{1,n}] \right| \leq \varepsilon \right) \leq \frac{\mathbb{V}[X_{1,n}]}{n\varepsilon^2}$$
$$1 - \mathbb{P}\left( \left| \overline{Y}_n - \mathbb{E}[Y_{1,n}] \right| \leq \tilde{\varepsilon} \left| \mathbb{E}[Y_{1,n}] \right| \right) \leq \frac{\mathbb{V}[Y_{1,n}]}{n\tilde{\varepsilon}^2 \left( \mathbb{E}[Y_{1,n}] \right)^2}.$$

For the numerator, $\mathbb{V}[X_{1,n}] = \mathbb{E}\left[ X_{1,n}^2 \right] - (\mathbb{E}[X_{1,n}])^2 \leq \mathbb{E}\left[ X_{1,n}^2 \right] \leq u_{X,n}$ using Assumption 2. For the denominator, Assumption 1 immediately entails that $1/l_{Y,n}^2$ is an upper bound on $1/\left( \mathbb{E}[Y_{1,n}] \right)^2$ and $l_{Y,n}^2$ a lower bound on $(\mathbb{E}[Y_{1,n}])^2$. Therefore

$$\frac{\mathbb{V}[Y_{1,n}]}{n\tilde{\varepsilon}^2 \left( \mathbb{E}[Y_{1,n}] \right)^2} \leq \frac{\mathbb{E}\left[ Y_{1,n}^2 \right] - l_{Y,n}^2}{n\tilde{\varepsilon}^2 l_{Y,n}^2} \leq \frac{u_{Y,n} - l_{Y,n}^2}{n\tilde{\varepsilon}^2 l_{Y,n}^2},$$

where the second inequality uses Assumption 2.

40

Combining the two bounds yields the following upper bound on the probability considered in Theorem 4.2

$$\frac{u_{X,n}}{n\varepsilon^2} + \frac{u_{Y,n} - l_{Y,n}{}^2}{n\tilde{\varepsilon}^2 l_{Y,n}{}^2},\tag{8}$$

as claimed.

For the second part of Theorem 4.2, for a fixed $\alpha$, we equalize each of the two terms in (8) to $\alpha/2$ and solve for $\varepsilon$ and $\tilde{\varepsilon}$, which yields:

$$\varepsilon^2 = \frac{2u_{X,n}}{n\alpha} \quad \text{and} \quad \tilde{\varepsilon}^2 = \frac{2\left(u_{Y,n} - l_{Y,n}{}^2\right)}{n\alpha l_{Y,n}^2}.$$

The bound $\overline{\alpha}_n$ comes from the fact that $\tilde{\varepsilon}$ needs to be smaller than 1.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

### B.5.1 Proof of Lemma B.4

We fix arbitrary $\varepsilon > 0$ and $\tilde{\varepsilon} \in (0, 1)$. Without loss of generality, we can assume that $\mathbb{E}[Y_{1,n}] > 0$ and $\mathbb{E}[X_{1,n}] \geq 0$.

First, using the union bound, note that the event $A_\varepsilon^{X_{1,n}} \cap \tilde{A}_{\tilde{\varepsilon}}^{Y_{1,n}}$ holds with a probability bigger than $\mathbb{P}\left(A_\varepsilon^{X_{1,n}}\right) + \mathbb{P}\left(\tilde{A}_{\tilde{\varepsilon}}^{Y_{1,n}}\right) - 1$. Hence, its complement is of probability lower than $1 - \mathbb{P}\left(A_\varepsilon^{X_{1,n}}\right) + 1 - \mathbb{P}\left(\tilde{A}_{\tilde{\varepsilon}}^{Y_{1,n}}\right)$.

Second, we show that the event considered in Lemma B.4 is included in the complement of $A_\varepsilon^{X_{1,n}} \cap \tilde{A}_{\tilde{\varepsilon}}^{Y_{1,n}}$, which concludes the proof. To do so, we reason by contraposition and do the following computations on the event $A_\varepsilon^{X_{1,n}} \cap \tilde{A}_{\tilde{\varepsilon}}^{Y_{1,n}}$. By the triangle inequality, we get

$$\left|\frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]}\right| \leq |\overline{X}_n| \cdot \left|\frac{1}{\overline{Y}_n} - \frac{1}{\mathbb{E}[Y_{1,n}]}\right| + \frac{1}{\mathbb{E}[Y_{1,n}]}\left|\overline{X}_n - \mathbb{E}[X_{1,n}]\right|.$$

We now bound the first term using the mean value theorem applied to the function $f(x) := 1/(x + \mathbb{E}[Y_{1,n}])$

$$\left|\frac{1}{\overline{Y}_n} - \frac{1}{\mathbb{E}[Y_{1,n}]}\right| = \left|f(\overline{Y}_n - \mathbb{E}[Y_{1,n}]) - f(0)\right| \leq \frac{|\overline{Y}_n - \mathbb{E}[Y_{1,n}]|}{(1 - \tilde{\varepsilon})^2 \mathbb{E}[Y_{1,n}]^2}$$
$$\leq \frac{\tilde{\varepsilon}\mathbb{E}[Y_{1,n}]}{(1 - \tilde{\varepsilon})^2 \mathbb{E}[Y_{1,n}]^2},$$

where the first inequality uses the following observation: on the event $\tilde{A}_{\tilde{\varepsilon}}^{Y_{1,n}}$, a lower

bound on $|x + \mathbb{E}[Y_{1,n}]|$ with $x$ varying between 0 and $\overline{Y}_n - \mathbb{E}[Y_{1,n}]$ is $(1-\tilde{\varepsilon})|\mathbb{E}[Y_{1,n}]|$. Therefore, on $A_\varepsilon^{X_{1,n}} \cap \tilde{A}_{\tilde{\varepsilon}}^{Y_{1,n}}$,

$$
\begin{aligned}
\left| \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]} \right| &\leq |\overline{X}_n| \cdot \frac{\tilde{\varepsilon}\mathbb{E}[Y_{1,n}]}{(1-\tilde{\varepsilon})^2 \mathbb{E}[Y_{1,n}]^2} + \frac{\varepsilon}{\mathbb{E}[Y_{1,n}]} \\
&\leq \big( |\mathbb{E}[X_{1,n}]| + |\overline{X}_n - \mathbb{E}[X_{1,n}]| \big) \frac{\tilde{\varepsilon}}{(1-\tilde{\varepsilon})^2 \mathbb{E}[Y_{1,n}]} + \frac{\varepsilon}{\mathbb{E}[Y_{1,n}]} \\
&\leq \frac{\big( |\mathbb{E}[X_{1,n}]| + \varepsilon \big)\tilde{\varepsilon}}{(1-\tilde{\varepsilon})^2 \mathbb{E}[Y_{1,n}]} + \frac{\varepsilon}{\mathbb{E}[Y_{1,n}]} ,
\end{aligned}
$$

where we use the triangle inequality to get the second line. It is indeed the complement of the event considered in the statement of Lemma B.4.

$\square$

## B.6   Proof of Theorem 5.2

This theorem relies crucially on the following lemma.

**Lemma B.5.** *For each $\xi$ in the interval $\big(0, 1 \wedge \big(u_{Y,n}/l_{Y,n}^2 - 1\big)\big)$, there exists a distribution $P_{n,\xi} \in \mathcal{P}$ such that $\mathbb{P}\big(\overline{Y}_n = 0\big) \geq \tilde{\alpha}_n(\xi)$, where $\tilde{\alpha}_n(\xi) := \big(1 - (1 + \xi)l_{Y,n}^2/u_{Y,n}\big)^n$.*

Note that the interval $\big(0, 1 \wedge \big(u_{Y,n}/l_{Y,n}^2 - 1\big)\big)$ is not empty since we have assumed $u_{Y,n}/l_{Y,n}^2 > 1$.

By Lemma B.5, for every $\xi < 1 \wedge \big(u_{Y,n}/l_{Y,n}^2 - 1\big)$, there exists a distribution $P_{n,\xi}$ such that $\mathbb{P}\big(\overline{Y}_n = 0\big) \geq \tilde{\alpha}_n(\xi)$. Taking the supremum over $\xi$, we deduce that

$$
\sup_{P_n \in \mathcal{P}} \mathbb{P}\big(\overline{Y}_n = 0\big) \geq \sup_\xi \tilde{\alpha}_n(\xi) = \underline{\alpha}_n.
$$

Using the assumption that $I_n$ is undefined whenever $\overline{Y}_n = 0$, we deduce that $\mathbb{P}\big(I_n\, undefined\big) \geq \underline{\alpha}_n$.

$\square$

### B.6.1   Proof of Lemma B.5

We consider the following distribution on $\mathbb{R}$

$$
P_{n,l_{Y,n},u_{Y,n},c,\xi} := \left(\frac{c}{n}\right)^{1/n} \delta_{\{0\}} + \frac{1}{2}\left(1 - \left(\frac{c}{n}\right)^{1/n}\right)\delta_{\{y_{c-}\}} + \frac{1}{2}\left(1 - \left(\frac{c}{n}\right)^{1/n}\right)\delta_{\{y_{c+}\}},
$$

where $c \in (0, n)$ is some constant to be chosen later, $y_{c-} := l_{Y,n}(1 - \sqrt{\xi})/(1 - (c/n)^{1/n})$ and $y_{c+} := l_{Y,n}(1 + \sqrt{\xi})/(1 - (c/n)^{1/n})$. Let $Y_{1,n} \sim P_{n,l_{Y,n},u_{Y,n},c,\xi_n}$. Observe that $\mathbb{E}[Y_{1,n}] = l_{Y,n}$ and $\mathbb{E}[Y_{1,n}^2] = l_{Y,n}^2(1 + \xi_n)/(1 - (c/n)^{1/n})$. With the choice

$$c = c_n := n\left(1 - \frac{l_{Y,n}^2}{u_{Y,n}}(1 + \xi)\right)^n,$$

we have $\mathbb{E}[Y_{1,n}^2] = u_{Y,n}$. Note that $C_{n,\alpha}$ is strictly positive, because $1 - \frac{l_{Y,n}^2}{u_{Y,n}}(1 + \xi_n)$ is positive. This is equivalent to $u_{Y,n}/l_{Y,n}^2 > 1 + \xi_n$, which is true by assumption.

Consider now the following product measure on $\mathbb{R}^2$ defined by $P_n := \delta_{\{\sqrt{u_{X,n}}\}} \otimes P_{n,l_{Y,n},u_{Y,n},c_n,\xi}$. Let $(X_{i,n}, Y_{i,n})_{i=1}^n \overset{\text{i.i.d.}}{\sim} P_n$. These random vectors satisfy $\mathbb{E}[X_{1,n}^2] = u_{X,n}$, $\mathbb{E}[Y_{1,n}] = l_{Y,n}$ and $\mathbb{E}[Y_{1,n}^2] = u_{Y,n}$. The next step is to build a lower bound on the event $\{\overline{Y}_n = 0\}$.

The assumption that $(X_{i,n}, Y_{i,n})_{i=1}^n \overset{\text{i.i.d.}}{\sim} P_n$ and the construction of $P_{n,l_{Y,n},u_{Y,n},c_n,\xi}$ imply that

$$\mathbb{P}\left(\overline{Y}_n = 0\right) = \frac{c_n}{n} = \left(1 - \frac{l_{Y,n}^2}{u_{Y,n}}(1 + \xi)\right)^n = \tilde{\alpha}_n(\xi).$$

$\square$

## B.7   Proof of Theorem 5.4

To prove Theorem 5.4, we need the following lemma.

**Lemma B.6.** *For every integer $n \geq 7$ and every $x \in (0, 1)$, $x(1 - x/n)^{n-1} \geq x/3$.*

We start using arguments developed in the proof of Catoni (2012)[Proposition 6.2]. We detail those for the sake of clarity. For every $n \in \mathbb{N}^*$ and $\eta > \sqrt{u_{X,n}}/n$, let us define the following distribution on $\mathbb{R}$, which will be used for the variable in the numerator[18]:

$$P_{n,u_{X,n},\eta} := \frac{u_{X,n}}{2n^2\eta^2}\delta_{\{-n\eta\}} + \left(1 - \frac{u_{X,n}}{n^2\eta^2}\right)\delta_{\{0\}} + \frac{u_{X,n}}{2n^2\eta^2}\delta_{\{n\eta\}}.$$

This distribution is symmetric, centered and has variance $u_{X,n}$. As shown in

---

[18] The notation $\delta$ denotes the Dirac distribution.

Catoni (2012), every i.i.d. sample $(X_{i,n})_{i=1}^n$ drawn from $P_{n,u_{X,n},\eta}$ satisfies

$$\mathbb{P}\left(\overline{X}_n \le -\eta\right) = \mathbb{P}\left(\overline{X}_n \ge \eta\right) \ge \mathbb{P}\left(\overline{X}_n = \eta\right)$$

$$\ge \sum_{i=1}^n \mathbb{P}\left(X_{i,n} = n\eta, X_{j,n} = 0, \ \forall j \ne i\right) = \frac{u_{X,n}}{2n\eta^2}\left(1 - \frac{u_{X,n}}{\eta^2 n^2}\right)^{n-1}.$$

Note further that for every integer $n \ge 2$, $\mathbb{P}\left(\overline{X}_n \ge \eta\right) \ge \mathbb{P}\left(\overline{X}_n = \eta\right)$ becomes a strict inequality strict and for every $\xi \in (0,1)$ $\{|\overline{X}_n| \ge \eta\} \subseteq \{|\overline{X}_n| > \xi\eta\}$. As a result, if $(X_{i,n})_{i=1}^n \overset{\text{i.i.d.}}{\sim} P_{n,u_{X,n},\eta}$, for every $\eta > 0$, we have

$$\mathbb{P}\left(|\overline{X}_n| > \xi\eta\right) > \frac{u_{X,n}}{n\eta^2}\left(1 - \frac{u_{X,n}}{\eta^2 n^2}\right)^{n-1}. \tag{9}$$

The following steps do not show up in Catoni (2012) since they are specific to controlling ratios of expectations and sample averages. For every $n \in \mathbb{N}^*$, let us define the following distribution on $\mathbb{R}$, which will be used for the variable in the denominator

$$P_{n,l_{Y,n},u_{Y,n}} := \frac{1}{2}\delta_{\left\{l_{Y,n} - \sqrt{u_{Y,n} - l_{Y,n}^2}\right\}} + \frac{1}{2}\delta_{\left\{l_{Y,n} + \sqrt{u_{Y,n} - l_{Y,n}^2}\right\}}.$$

Let $(X_{i,n}, Y_{i,n})_{i=1}^n \overset{\text{i.i.d.}}{\sim} P_n := P_{n,u_{X,n},\eta} \otimes P_{n,l_{Y,n},u_{Y,n}}$. Observe that $\mathbb{E}[Y_{1,n}] = l_{Y,n}$ and $\mathbb{E}[Y_{1,n}^2] = u_{Y,n}$. Furthermore, $|\overline{Y}_n| \le l_{Y,n} + \sqrt{u_{Y,n} - l_{Y,n}^2}$ almost surely. This implies that for every $\eta > 0$ and $\xi \in (0,1)$, the following holds

$$\left\{|\overline{X}_n| > \left(l_{Y,n} + \sqrt{u_{Y,n} - l_{Y,n}^2}\right)\xi\eta\right\} \subseteq \left\{\left|\frac{\overline{X}_n}{\overline{Y}_n}\right| > \xi\eta\right\}.$$

For fixed $n \ge 7$ and $\alpha \in \left(0, 1 \wedge n/\left(l_{Y,n} + \sqrt{u_{Y,n} - l_{Y,n}^2}\right)^2\right)$, we choose $\eta = \eta(\alpha) = \sqrt{v_n/3n\alpha}$. Combining the above inclusion with (9), and Lemma B.6 (with the choice $x = 3\alpha$), we conclude that there exists a distribution on $\mathbb{R}^2$, namely $P_n$, that fulfills Assumptions 1 and 2 such that

$$\mathbb{P}\left(\left|\frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]}\right| > \xi\sqrt{\frac{v_n}{3n\alpha}}\right) > \alpha,$$

which completes the proof.

$\square$

### B.7.1 Proof of Lemma B.6

Under our assumptions on $n$ and $x$, $\ln\left(1 - x/n\right)$ is well-defined. Using Taylor-Lagrange formula on the function $[0, x] \ni t \mapsto \ln\left(1 - t/n\right)$ yields:

$$
\begin{aligned}
\left(1 - \frac{x}{n}\right)^{n-1} &= \exp\left((n-1)\ln\left(1 - \frac{x}{n}\right)\right) \\
&= \exp\left(-(n-1)\left(\frac{x}{n} + \frac{1}{2\left(1 - \tau x/n\right)^2}\frac{x^2}{n^2}\right)\right)
\end{aligned}
$$

for some $\tau \in (0, 1)$. Using the fact that $\frac{n-1}{n} \leq 1$, $x \leq 1$ and $\frac{1}{2(1-\tau x/n)^2} \leq \frac{1}{2(1-n^{-1})^2}$, we get that under our assumptions $\left(1 - \frac{x}{n}\right)^{n-1} \geq \exp\left(-\left(1 + \frac{1}{2n(1-n^{-1})^2}\right)\right)$. This bound is actually valid for every $x \in (0, 1)$ and every $n \in \mathbb{N}^*$. The computation of $\exp\left(-\left(1 + \frac{1}{2n(1-n^{-1})^2}\right)\right)$ shows that the latter is larger than $1/4$ whenever $n \geq 3$ and larger than $1/3$ whenever $n \geq 7$.

$\square$

## C  Adapted results for "Hoeffding" framework

**Assumption 4.** *For every $n \in \mathbb{N}^*$, there exist finite constants $a_{X,n}$, $b_{X,n}$, $a_{Y,n}$, $b_{Y,n}$ and $l_{Y,n}$ such that $X_{1,n}$ (respectively $Y_{1,n}$) lies $P_{X,Y,n}$-almost surely in the interval $[a_{X,n}, b_{X,n}]$ (resp. $[a_{Y,n}, b_{Y,n}]$) and $|\mathbb{E}[Y_{1,n}]| \geq l_{Y,n}$.*

The support of $X_{1,n}$ and $Y_{1,n}$ is allowed to change with $n$, even though in many examples of interest, the former can be chosen independent from $n$. Assumptions 1 and 4 together correspond to the *Hoeffding case* because under these two assumptions, we can use the Hoeffding inequality to build nonasymptotic CIs.

### C.1  Concentration inequality in an easy case: the support of the denominator is well-separated from $0$

**Assumption 5.** *For every $n \in \mathbb{N}^*$, the lower bound $a_{Y,n}$ is strictly positive.*

**Theorem C.1.** *Let $u_{X,n} := (b_{X,n} - a_{X,n})^2$ and $u_{Y,n} := (b_{Y,n} - a_{Y,n})^2$. Under Assumptions 1, 4 and 5, we have for every $n \in \mathbb{N}^*$ and $\varepsilon \in \mathbb{R}_+^*$*

$$
\sup_{P \in \mathcal{P}} \mathbb{P}_{P^{\otimes n}}\left(\left|\frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]}\right| > \frac{\varepsilon}{l_{Y,n}}\left\{1 + \frac{1}{a_{Y,n}}\left(|a_{X,n}| \vee |b_{X,n}| + \varepsilon\right)\right\}\right)
$$

$$
\leq 4\exp\left(-\frac{2n\varepsilon^2}{u_{X,n} \vee u_{Y,n}}\right).
$$

As a consequence, $\inf_{P\in\mathcal{P}} \mathbb{P}_{P^{\otimes n}}\left( \mathbb{E}[X_{1,n}]/\mathbb{E}[Y_{1,n}] \in \left[\,\overline{X}_n/\overline{Y}_n \pm t\,\right] \right) \geq 1 - \alpha$, with the following choice for $t$:

$$\sqrt{\frac{(u_{X,n} \vee u_{Y,n})\ln(4/\alpha)}{2nl_{Y,n}{}^2}} \left( 1 + \frac{1}{a_{Y,n}} \left( |a_{X,n}| \vee |b_{X,n}| + \sqrt{\frac{(u_{X,n} \vee u_{Y,n})\ln(4/\alpha)}{2n}} \right) \right),$$

for every $\alpha \in (0,1)$.

The theorem shows that it is possible to construct nonasymptotic CIs for ratios of expectations at every confidence level that are almost surely bounded. However, it requires the additional Assumption 5, that in particular does not allow for binary $\{0,1\}$ random variables in the denominator which may limit its applicability for various applications. In Section C.2, we give an analogous result that only requires Assumptions 1 and 4 to hold, so that it encompasses the case of $\{0,1\}$-valued denominators. However, the cost to pay will be an upper bound on the achievable coverage of the confidence intervals.

## C.2   Concentration inequality in the general case

We seek to build nontrivial nonasymptotic CIs under Assumptions 1 and 4 only. Under Assumption 1, $\mathbb{E}[Y_{1,n}] \neq 0$, so that there is no issue in considering the fraction $\mathbb{E}[X_{1,n}]/\mathbb{E}[Y_{1,n}]$. However, without Assumption 5, $\{\overline{Y}_n = 0\}$ has positive probability in general so that $\overline{X}_n/\overline{Y}_n$ is well-defined with probability less than one and undefined else. Note that when $P_{Y,n}$ is continuous wrt to Lebesgue's measure, there is no issue in defining $\overline{X}_n/\overline{Y}_n$ anymore since the event $\{\overline{Y}_n = 0\}$ has probability zero. This is not an easier case to establish concentration inequalities though, since without more restrictions, $\overline{Y}_n$ can still be arbitrarily close to 0 with positive probability.

**Theorem C.2.** *Assume that Assumptions 1 and 4 hold. For every $n \in \mathbb{N}^*$, $\varepsilon > 0, \tilde{\varepsilon} \in (0,1)$, we have*

$$\sup_{P\in\mathcal{P}} \mathbb{P}_{P^{\otimes n}}\left( \left| \frac{\overline{X}_n}{\overline{Y}_n} - \frac{\mathbb{E}[X_{1,n}]}{\mathbb{E}[Y_{1,n}]} \right| > \left( \frac{(|a_{X,n}| \vee |b_{X,n}| + \varepsilon)\tilde{\varepsilon}}{(1-\tilde{\varepsilon})^2} + \varepsilon \right) \frac{1}{l_{Y,n}} \right)$$
$$\leq 2\exp(-n\varepsilon^2\gamma(X_{1,n})) + 2\exp(-n\tilde{\varepsilon}^2\gamma(Y_{1,n})),$$

*where $\gamma(X_{1,n}) = 2/(b_{X,n} - a_{X,n})^2$ and $\gamma(Y_{1,n}) = 2l_{Y,n}^2/(b_{Y,n} - a_{Y,n})^2$.*

As a consequence, $\inf_{P\in\mathcal{P}} \mathbb{P}_{P^{\otimes n}}\left( \mathbb{E}[X_{1,n}]/\mathbb{E}[Y_{1,n}] \in \left[\,\overline{X}_n/\overline{Y}_n \pm t\,\right] \right) \geq 1 - \alpha,$

*with the choice*

$$t := \sqrt{\frac{\ln(4/\alpha)}{n\gamma\left(X_{1,n}\right) \wedge \gamma\left(Y_{1,n}\right)}} \left(\frac{|a_{X,n}| \vee |b_{X,n}| + \sqrt{\ln(4/\alpha)/\left(n\gamma\left(X_{1,n}\right)\right)}}{\left(1 - \sqrt{\ln(4/\alpha)/\left(n\gamma\left(Y_{1,n}\right)\right)}\right)^2} + 1\right)\frac{1}{l_{Y,n}},$$

*for every* $\alpha > \overline{\alpha}_{n,H} := 4e^{-n\gamma(Y_{1,n})}$.[19]

This theorem is proved in Section C.4. It states that when $l_{Y,n} > 0$, it is possible to build valid nonasymptotic CIs with finite length up to the confidence level $1 - \overline{\alpha}_{n,H}$. This is a more positive result than Dufour (1997) which claims that it is not possible to build nontrivial nonasymptotic CIs when $l_{Y,n}$ is taken equal to 0, no matter the confidence level. Note that Theorem C.2 is not an impossibility theorem since it only claims that considering confidence levels smaller than $1 - \overline{\alpha}_{n,H}$ is *sufficient* to build nontrivial CIs under Assumptions 1 and 4. The remaining question is to find out whether it is *necessary* to focus on confidence levels that do not exceed a certain threshold under Assumptions 1 and 4. We answer this in Section C.3.

Theorem C.2 has two other interesting consequences: for every confidence level up to $1 - \overline{\alpha}_{n,H}$, a nonasymptotic CI of the form $\left[\overline{X}_n/\overline{Y}_n \pm \tilde{t}\right]$ with $\tilde{t} > t$ has good coverage but is too conservative. What is more, if the DGP does not depend on $n$ (*i.e* in the standard i.i.d. set-up), for every fixed $\alpha > \overline{\alpha}_{n,H}$, the length of the confidence interval shrinks at the optimal rate $1/\sqrt{n}$.

## C.3   An upper bound on testable confidence levels

**Theorem C.3.** *For every* $n \in \mathbb{N}^*$, *and every* $\alpha \in \left(0, \underline{\alpha}_{n,H}\right)$, *where* $\underline{\alpha}_{n,H} := \left(1 - l_{Y,n}/(b_{Y,n} - a_{Y,n})\right)^n$, *if* $(b_{Y,n} - a_{Y,n})/l_{Y,n} > 1$, *there is no finite* $t > 0$ *such that* $\left[\overline{X}_n/\overline{Y}_n \pm t\right]$ *has coverage* $1 - \alpha$ *over* $\mathcal{P}_H$, *where* $\mathcal{P}_H$ *is the class of all distributions satisfying Assumptions 1 and 4 for a fixed lower bound* $l_{Y,n}$ *and fixed lengths* $b_{X,n} - a_{X,n}$ *and* $b_{Y,n} - a_{Y,n}$.

This theorem asserts that confidence intervals of the form $\left[\overline{X}_n/\overline{Y}_n \pm t\right]$ with coverage higher than $1 - \underline{\alpha}_{n,H}$ under Assumptions 1 and 4 are not defined (or are of infinite length) with positive probability for at least one distribution in $\mathcal{P}_H$. The additional restriction $(b_{Y,n} - a_{Y,n})/l_{Y,n} > 1$ is rather mild in practice: it is equivalent to $b_{Y,n} - a_{Y,n} > l_{Y,n}$ and is satisfied as soon as $a_{Y,n} \leq 0$ and $b_{Y,n} > l_{Y,n} > 0$. This

---

[19]Equivalently, it means that for a given level $\alpha$, the choice of $t$ is valid for every integer $n > \overline{n}_{\alpha,H} := \ln(4/\alpha)/\gamma\left(Y_{1,n}\right)$.

encompasses all DGPs where the denominator is $\{0, 1\}$-valued and the probability that the denominator equals 1 is bounded from below by $l_{Y,n} \in (0, 1)$.

Note that for Theorems C.1 and C.2, it is required to know not only the length $b_{X,n} - a_{X,n}$ but also the actual endpoints of the support, $a_{X,n}$ and $b_{X,n}$. On the contrary, Theorem C.3 does not require the latter. In that respect, the class of Theorem C.3 is larger than the one of the two preceding theorems.

## C.4  Proof of Theorems C.1 and C.2

The proofs are identical to those of Theorems 4.1 and 4.2, except for the Bienaymé-Chebyshev inequality that has to be replaced with the Hoeffding inequality. The latter can be used under Assumption 4. Note also that $\mathbb{E}[X_{1,n}]$ is now bounded by $|a_{X,n}| \vee |b_{X,n}|$.

$\square$

## C.5  Proof of Theorem C.3

We need the subsequent lemma.

**Lemma C.4.** *For each $\xi$ in the interval $\left(0, 1 \wedge \left((b_{Y,n} - a_{Y,n})/l_{Y,n} - 1\right)\right)$, there exists a distribution $P_{n,\xi} \in \mathcal{P}_H$ such that $\mathbb{P}\left(\overline{Y}_n = 0\right) \geq \tilde{\alpha}_{n,H}(\xi)$, where $\tilde{\alpha}_{n,H}(\xi) := \left(1 - (1 + \xi)l_{Y,n}/(b_{Y,n} - a_{Y,n})\right)^n$.*

Note that the interval $\left(0, 1 \wedge \left((b_{Y,n} - a_{Y,n})/l_{Y,n} - 1\right)\right)$ is not empty since we have assumed $(b_{Y,n} - a_{Y,n})/l_{Y,n} > 1$.

By Lemma C.4, for every $\xi < 1 \wedge \left((b_{Y,n} - a_{Y,n})/l_{Y,n} - 1\right)$, there exists a distribution $P_{n,\xi} \in \mathcal{P}_H$ satisfying Assumptions 1 and 4 such that $\mathbb{P}\left(\overline{Y}_n = 0\right) \geq \tilde{\alpha}_{n,H}(\xi)$. Denote its marginal distributions by $P_{X,n,\xi}$ and $P_{Y,n,\xi}$. Therefore, $P_{n,\xi}$ satisfies Assumptions 1 and 4, and $\overline{X}_n/\overline{Y}_n$ is undefined with probability greater than $\tilde{\alpha}_{n,H}(\xi)$. Taking the supremum over $\xi$, we deduce that

$$\sup_{P_n \in \mathcal{P}_H} \mathbb{P}\left(\overline{Y}_n = 0\right) \geq \sup_{\xi} \tilde{\alpha}_n(\xi) = \underline{\alpha}_{n,H}.$$

This means that the random interval $I_n^* := \left[\overline{X}_n/\overline{Y}_n \pm t\right]$ cannot have coverage higher than $1 - \underline{\alpha}_{n,H}$ since it may be undefined with a probability higher than $\underline{\alpha}_{n,H}$.

$\square$

48

### C.5.1  Proof of Lemma C.4

We consider the following distribution on $\mathbb{R}$

$$P_{n,l_{Y,n},c,\xi} := \left(\frac{c}{n}\right)^{1/n} \delta_{\{0\}} + \frac{1}{2}\left(1 - \left(\frac{c}{n}\right)^{1/n}\right)\delta_{\{y_{c-}\}} + \frac{1}{2}\left(1 - \left(\frac{c}{n}\right)^{1/n}\right)\delta_{\{y_{c+}\}},$$

where $c \in (0,n)$ is some constant to be chosen later, $y_{c-} := l_{Y,n}(1-\xi)/(1-(c/n)^{1/n})$ and $y_{c+} := l_{Y,n}(1+\xi)/(1-(c/n)^{1/n})$. Let $Y_{1,n} \sim P_{n,l_{Y,n},c,\xi_n}$. Observe that $\mathbb{E}[Y_{1,n}] = l_{Y,n}$. With the choice

$$c = c_n := n\left(1 - \frac{l_{Y,n}}{b_{Y,n} - a_{Y,n}}(1+\xi)\right)^n,$$

we have $y_{c+} = b_{Y,n} - a_{Y,n}$. Note that $C_{n,\alpha}$ is strictly positive, because $1 - \frac{l_{Y,n}}{b_{Y,n}-a_{Y,n}}(1+\xi_n) > 0$. This is equivalent to $b_{Y,n} - a_{Y,n}/l_{Y,n} > 1 + \xi_n$, which is true by assumption.

Consider now the following product measure on $\mathbb{R}^2$ defined by $P_n := \left(0.5\delta_{\{0\}} + 0.5\delta_{\{b_{X,n}-a_{X,n}\}}\right) \otimes P_{n,l_{Y,n},c_n,\xi}$. Let $(X_{i,n}, Y_{i,n})_{i=1}^n \overset{\text{i.i.d.}}{\sim} P_n$. These random vectors satisfy $\mathbb{E}[Y_{1,n}] = l_{Y,n}$, $(\max - \min)[Y_{1,n}] = b_{Y,n} - a_{Y,n}$ and $(\max - \min)[X_{1,n}] = b_{X,n} - a_{X,n}$. The next step is to build a lower bound on the event $\{\overline{Y}_n = 0\}$.

The assumption that $(X_{i,n}, Y_{i,n})_{i=1}^n \overset{\text{i.i.d.}}{\sim} P_n$ and the construction of $P_{n,l_{Y,n},c_n,\xi}$ imply that

$$\mathbb{P}\left(\overline{Y}_n = 0\right) = \frac{c_n}{n} = \left(1 - \frac{l_{Y,n}}{b_{X,n} - a_{X,n}}(1+\xi)\right)^n = \tilde{\alpha}_{n,H}(\xi).$$

$\square$

# D  Additional simulations

This section complements the simulations presented in the main body of the article using different distributions for the variables in the numerator and in the denominator.

In this setting of simulations, we use the best bounds by setting the constants $l_{Y,n}$ and $u_{Y,n}$ that define our class of distributions equal to the actual corresponding moments (respectively the expectation for $l_{Y,n}$ and the second moment for $u_{Y,n}$). That is we use $\overline{n}_\alpha = 2\mathbb{V}[Y]/(\alpha\mathbb{E}[Y]^2)$ or $\overline{\alpha}_n = 2\mathbb{V}[Y]/(n\mathbb{E}[Y]^2)$. In practice, our rule-of-thumb uses the plug-in version of those quantities replacing the theoretical unknown moments by their empirical counterparts as explained in Section 5.3.

The following Figures are similar to Figures 6 and 7. They show the $c(n, P)$ of the asymptotic CIs based on the delta method as a function of the sample size $n$ and also reports $\overline{n}_\alpha := 2\left(u_{Y,n} - l_{Y,n}^2\right) / \left(\alpha l_{Y,n}^2\right)$, with $\alpha$ chosen according to the desired nominal level (equal to $1 - \alpha$) and $l_{Y,n} = \mathbb{E}[Y]$, $u_{Y,n} = \mathbb{E}[Y]^2 + \mathbb{V}[Y]$. Consequently, the titles of the figures only indicate the specification used for $P_{X,Y,n}$, the nominal pointwise asymptotic level $1 - \alpha$, and the number of repetitions used to approximate the probability $c(n, P)$.

With discrete distributions for the variable in the denominator, it may happen that $\overline{Y}_n = 0$, all the more so as the expectation and the sample size are low typically. As discussed at the end of Section 2, confidence intervals are said to be undefined when $\overline{Y}_n = 0$. In such cases, for any value $a \in \mathbb{R}$, it is undefined whether $a$ belongs or not to the CIs. Consequently, whenever the sample drawn is such that $\overline{Y}_n = 0$ in the simulations, we count the draw as a no coverage occurrence in the Monte Carlo estimation of $c(n, P)$. In other words, this quantity is approximated as an average over $M$ repetitions and the repetitions for which $\overline{Y}_n = 0$ account for 0 in this average.[20]

## D.1   Gaussian distributions
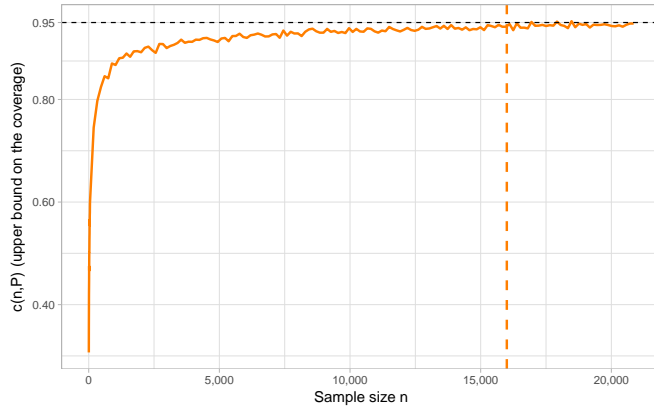


Figure 9: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}(1, 1) \otimes \mathcal{N}(0.05, 1); 1 - \alpha = 0.95; 5,000$ repetitions used.

---

[20]Note that in some specifications, a substantial part of the repetitions yield $\overline{Y}_n = 0$. For instance, with Bernoulli distributions, for $n$ smaller than 10 and the expectation at the denominator equal to 0.01, around 10% only of the repetitions display $\overline{Y}_n \neq 0$.
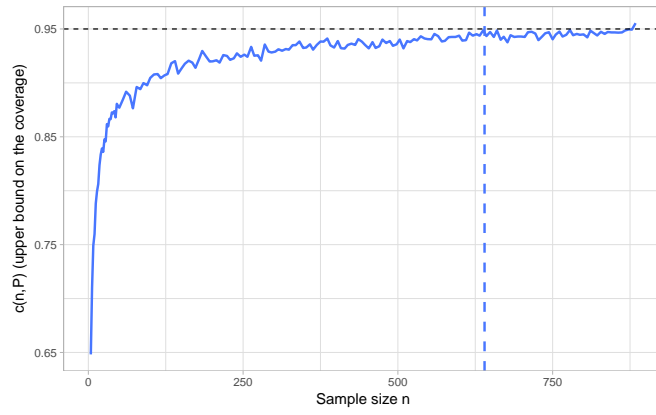
Figure 10: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}(1,1) \otimes \mathcal{N}(0.25,1)$; $1 - \alpha = 0.95$; 5,000 repetitions used.
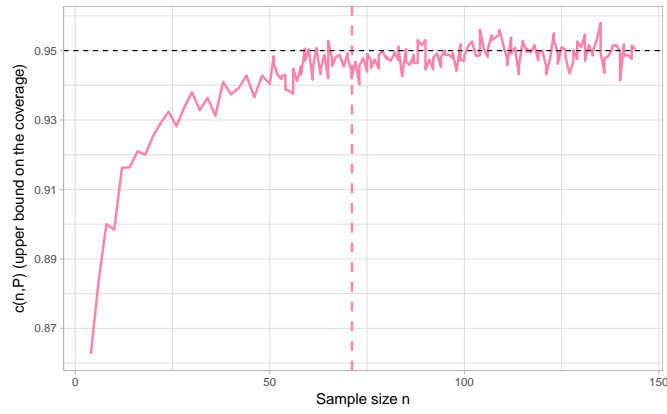


Figure 11: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}(1,1) \otimes \mathcal{N}(0.75,1)$; $1 - \alpha = 0.95$; 5,000 repetitions used.
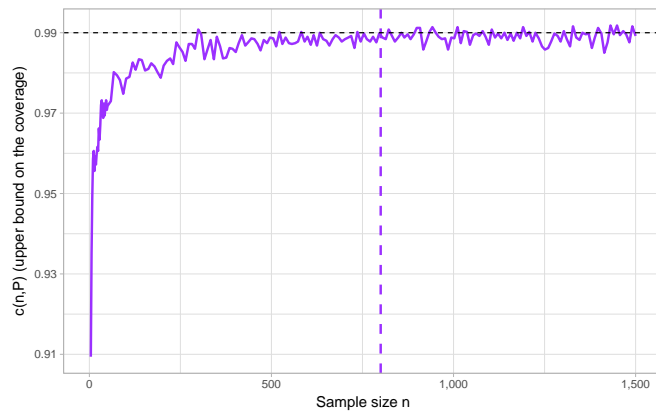


Figure 12: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}_2$ bivariate Gaussian with $\mathbb{E}[X] = 0.5$, $\mathbb{E}[Y] = 0.5$, $\mathbb{V}[X] = 2$, $\mathbb{V}[Y] = 1$, $\mathbb{C}orr(X,Y) = -0.3$; $1 - \alpha = 0.99$; 5,000 repetitions used.

51
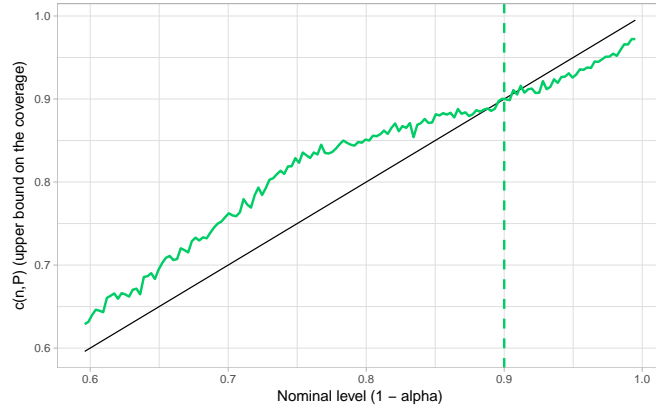
Figure 13: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}(1,1) \otimes \mathcal{N}(0.1,1);\ n = 2,000;\ 5,000$ repetitions used.

## D.2 Student distributions

The specification here is two Student distributions, both in the numerator and in the denominator. Standard Student distributions are centered. We use therefore translated versions by simply adding the expectations in order to avoid a null denominator for the ratio of expectations of interest. Below, $\mathcal{T}(\mu, \nu)$ denotes the distribution of a translated standard Student variable: $\mu + T$ where $T$ is distributed according to a Student distribution with $\nu$ degrees of freedom. To satisfy Assumption 1, we need finite variance: we use degrees of freedom strictly higher than 2 for this purpose.
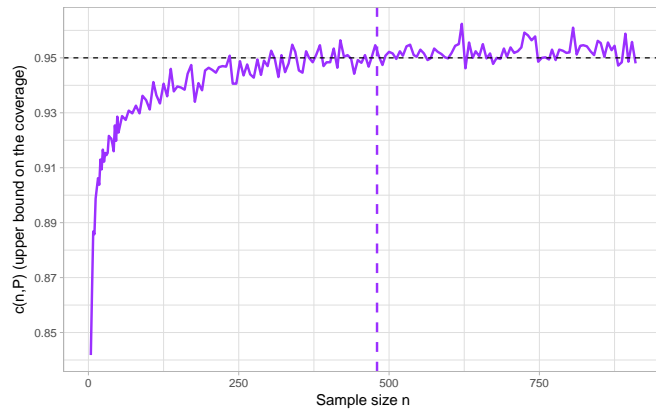


Figure 14: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{T}(0.5,3) \otimes \mathcal{T}(0.5,3);\ 1 - \alpha = 0.95;\ 5,000$ repetitions used.
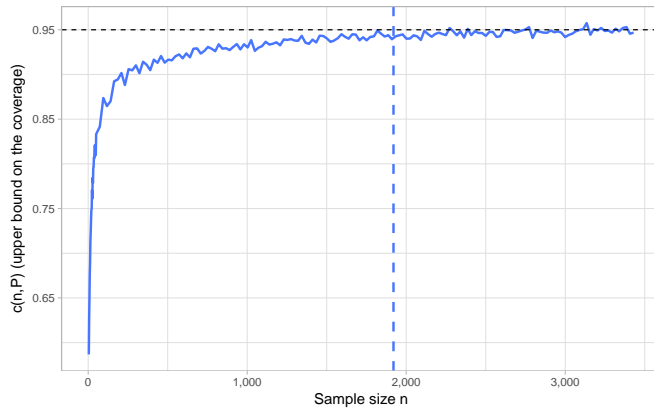
Figure 15: Specification: $\forall n \in \mathbb{N}^*$, the marginal distributions of $X$ and $Y$ are $\mathcal{T}(1,3)$ and $\mathcal{T}(0.25,3)$ respectively and are simulated using a Gaussian copula to have $\mathbb{C}orr(X,Y) \approx 0.5$.; $1 - \alpha = 0.95$; 5,000 repetitions used.

## D.3 Exponential distributions

The specification here is two exponential distributions, both in the numerator and in the denominator. The case of the exponential is specific as a unique parameter determines both the expectation and the variance of the distribution.
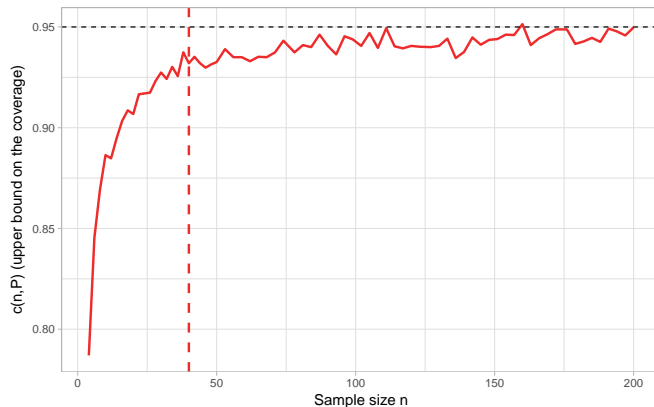


Figure 16: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{E} \otimes \mathcal{E}$ with $\mathbb{E}[X] = 1$ and $\mathbb{E}[Y] = 0.01$; $1 - \alpha = 0.95$; 5,000 repetitions used.

More precisely, the variance is equal to the square of the expectation. Consequently, whatever the parameter of the exponential distribution in the denominator, we have $\overline{n}_\alpha = 4/\alpha$. Previous simulations suggest that the closer the expectation in the denominator to 0, the larger the sample size required for the asymptotic approximation to hold. At first sight, we might thus be worried for the usefulness of our rule-of-thumb to obtain $\overline{n}_\alpha$ independent of $\mathbb{E}[Y]$. Yet, with exponential distributions, the lower the expectation, the lower is the variance too. Intuitively, the lower variance will compensate having an expectation closer to 0. The previ-

ous statement that links the closeness to 0 of the expectation in the denominator and the sample size required to reach the asymptotic approximation presupposes keeping fixed the variance. It cannot be anymore for exponential distributions.
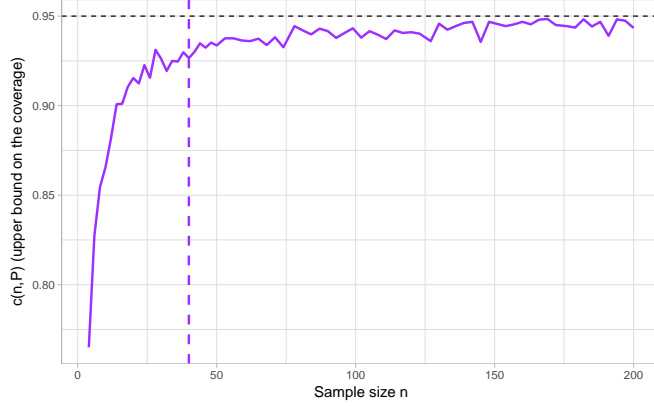


Figure 17: Specification: $\forall n \in \mathbb{N}^*$, the marginal distributions of $X$ and $Y$ are two exponentials with $\mathbb{E}[X] = 1$ and $\mathbb{E}[Y] = 0.5$ and are simulated using a Gaussian copula to have $\mathbb{C}orr(X, Y) \approx 0.75$.; $1 - \alpha = 0.95$; 5,000 repetitions used.

The simulations reveal that the convergence of the coverage of the asymptotic confidence intervals toward their nominal level happens for $n$ around one hundred fifty and has the same pattern whatever the expectation of the exponential distribution in the denominator. Our rule-of-thumb $\overline{n}_\alpha$ appears to be a bit small. Nonetheless, it is coherent that it is constant across the value of $\mathbb{E}[Y]$.

## D.4 Pareto distributions

The specification here is two Pareto distributions, both in the numerator and in the denominator. Pareto distributions have support in $\mathbb{R}_+^*$. They would fall in the easier case when the support of the denominator is well separated from 0. To assess the dependability of our rule-of-thumb in the general case, we use translated Pareto distributions. In what follows, the notation $\mathcal{P}areto(\mathbb{E}[Y], \tau, \gamma)$ denotes the distribution of a random variable that follows a Pareto distribution with shape parameter equal to $\gamma$ translated such that its support is $(\tau, +\infty)$ and its expectation is $\mathbb{E}[Y]$. A variable that is distributed according to $\mathcal{P}areto(\mathbb{E}[Y], \tau, \gamma)$ is equal in distribution to $P + (\mathbb{E}[Y] - \gamma t_Y)/(\gamma - 1)$ with $t_Y = (\mathbb{E}[Y] - \tau) \times (\gamma - 1)$ and $P$ a usual Pareto distribution with support or scale parameter $t_Y$ and shape parameter $\gamma$, that is $P$ has the density $x \mapsto \mathbb{1}\{x \geq t_Y\} \times \gamma t_Y^\gamma / x^{\gamma+1}$ with respect to Lebesgue measure.
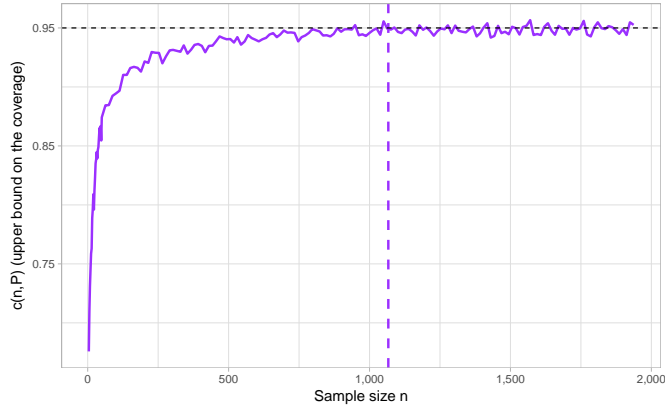
Figure 18: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{P}areto(1, -1.5, 5) \otimes \mathcal{P}areto(\mathbb{E}[Y], -1.5, 5)$, with $\mathbb{E}[Y] = 0.5$; $1 - \alpha = 0.95$; 5,000 repetitions used.
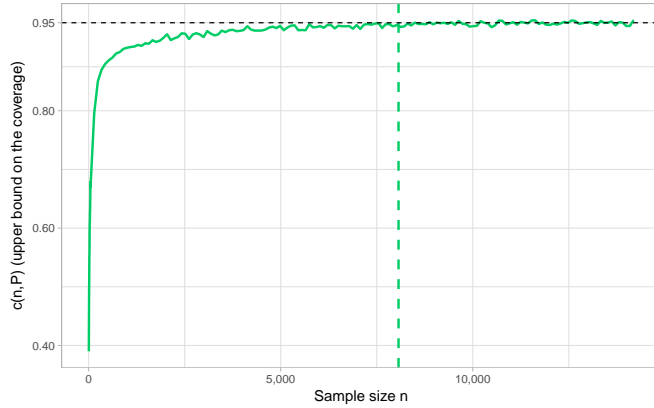


Figure 19: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{P}areto(1, -1.5, 5) \otimes \mathcal{P}areto(\mathbb{E}[Y], -1.5, 5)$, with $\mathbb{E}[Y] = 0.1$; $1 - \alpha = 0.95$; 5,000 repetitions used.

## D.5   Bernoulli distributions

Figure 20 is the equivalent of Figure 1 with Bernoulli distributions. The following graphs illustrate the use of $\overline{n}_\alpha$ to appraise the reliability of the asymptotic confidence based on the delta method. In practice a plug-in strategy has to be used to compute $\overline{n}_\alpha$ and, in the setting of simulations, we simply use the known moments and bounds of the DGP used in the simulation. With two Bernoulli variables in the numerator and the denominator, we are both in the BC and the "Hoeffding" cases. Thus, we show both the one obtained in the BC case $\overline{n}_\alpha := 2\left(u_{Y,n} - l_{Y,n}{}^2\right) / \left(\alpha l_{Y,n}^2\right)$ with a dashed vertical line (Theorem 4.2) and the one obtained in the "Hoeffding" case $\overline{n}_{\alpha,H} := \ln(4/\alpha)/\gamma(Y_{1,n})$, setting here $a_{Y,n} = 0$, $b_{Y,n} = 1$ and $l_{Y,n} = \mathbb{E}[Y]$, with a dotted vertical line (Theorem C.2).
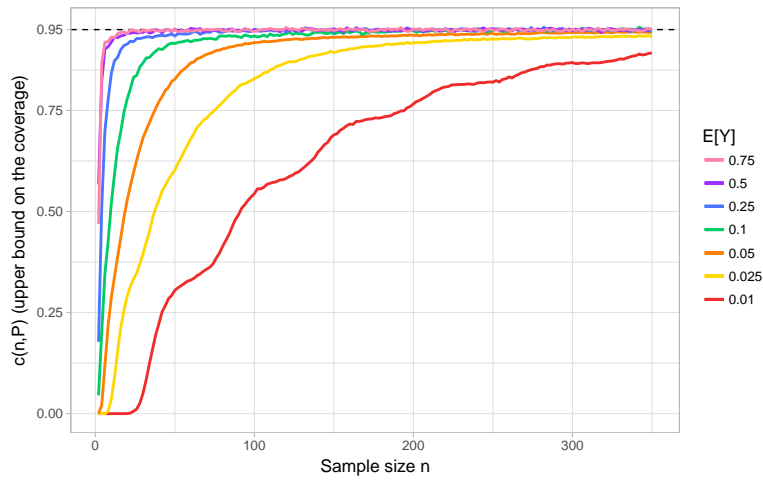
Figure 20: $c(n, P)$ of the CIs based on the delta method as a function of $n$. Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{B}(0.5) \otimes \mathcal{B}(\mathbb{E}[Y])$. The nominal pointwise asymptotic level is set to 0.95. 10,000 repetitions used.
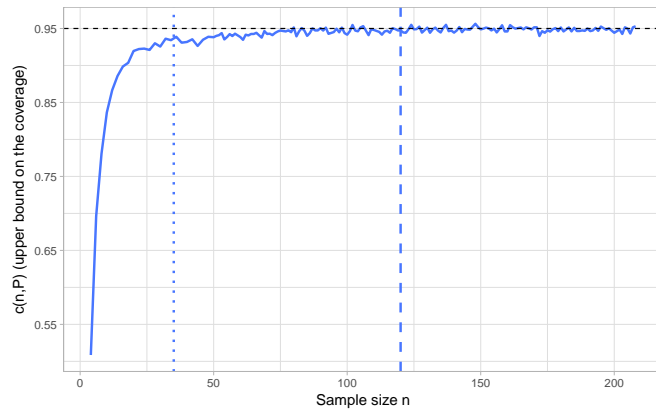


Figure 21: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{B}(0.5) \otimes \mathcal{B}(0.25)$; $1 - \alpha = 0.95$; 5,000 repetitions used.
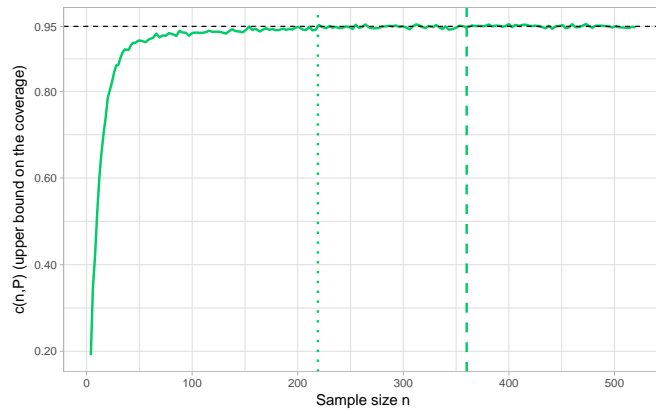


Figure 22: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{B}(0.5) \otimes \mathcal{B}(0.1)$; $1 - \alpha = 0.95$; 5,000 repetitions used.

## D.6    Poisson distributions

The specification here considers two variables distributed according to a Poisson distribution, both in the numerator and in the denominator.
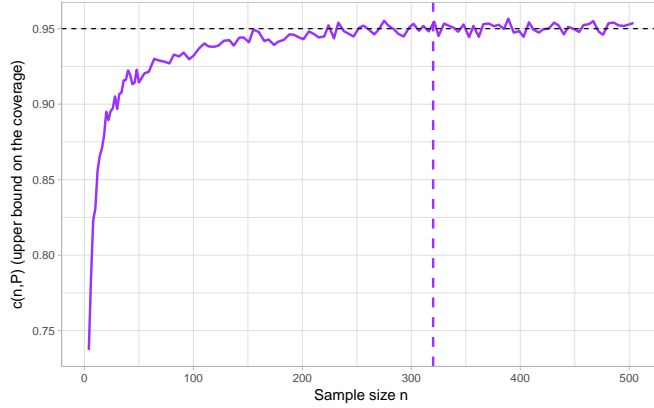


Figure 23: Specification: $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{P}oisson(0.5, 2) \otimes \mathcal{P}oisson(0.5, 2)$; $1 - \alpha = 0.95$; 5,000 repetitions used.
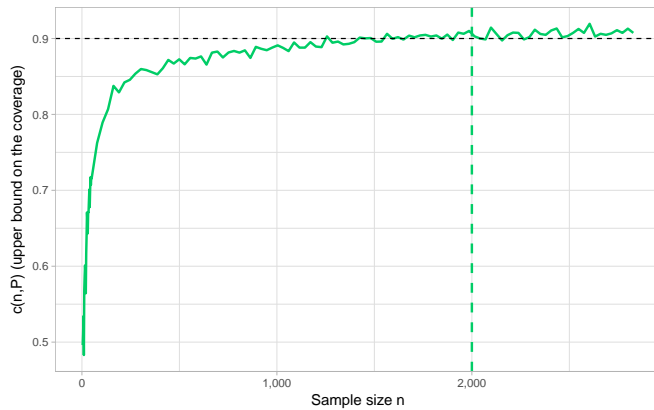


Figure 24: Specification: $\forall n \in \mathbb{N}^*$, the marginal distributions of $X$ and $Y$ are respectively $\mathcal{P}oisson(0.5, 2)$ and $\mathcal{P}oisson(0.1, 1)$ and are simulated using a Gaussian copula to have $\mathbb{C}orr(X, Y) \approx 0.6$; $1 - \alpha = 0.9$; 5,000 repetitions used.

A Poisson distribution is entirely defined by its positive real parameter, which is equal to both its expectation and its variance. Consequently, to have denominator close to 0, we would need small variance too, as in the exponential specification (see Section D.3). In order to disentangle expectation and variance, we use below translated Poisson variables. More precisely, the notation $\mathcal{P}oisson(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^*_+$, denotes a distribution alike to a Poisson, with parameter and variance equal to $\sigma^2$ but translated such that its expectation is $\mu$. That is a variable distributed according to $\mathcal{P}oisson(\mu, \sigma^2)$ is equal in distribution to $P + (\mu - \sigma^2)$

with $P$ a standard Poisson distribution with parameter $\sigma^2$ - that is with density with respect to the counting measure equal to $(\sigma^2)^k \exp(-\sigma^2)/(k!)$ for every $k \in \mathbb{N}$. Thus, a $\mathcal{P}oisson(\mu, \sigma^2)$ has expectation $\mu$ and variance $\sigma^2$.

## D.7 Delta method and nonparametric percentile bootstrap confidence intervals

The two following figures are the equivalent to Figure 4 with different values of $C$. They illustrate that the lower $C$, the lower the signal-to-noise ratio in the denominator, hence the more difficult in some sense is the estimation of $\theta_n$. This is illustrated by the fact that, all other things equal, larger $C$ basically translates $c(n, P)$ upward as revealed by the series of Figures 4, 25, and 26.

These three figures all report the $c(n, P)$ of the CIs based on the delta method (in blue) and of the CIs constructed with Efron's non parametric bootstrap using 2,000 bootstrap replications (in red) with the specification $\forall n \in \mathbb{N}^*, P_{X,Y,n} = \mathcal{N}(1,1) \otimes \mathcal{N}(Cn^{-b}, 1)$, with $b \in \{0, 0.25, 0.5, 0.75\}$. For the three of them, the nominal pointwise asymptotic level is set to 0.95 and for each pair $(b, n)$, the coverage is obtained as the mean over 5,000 repetitions.

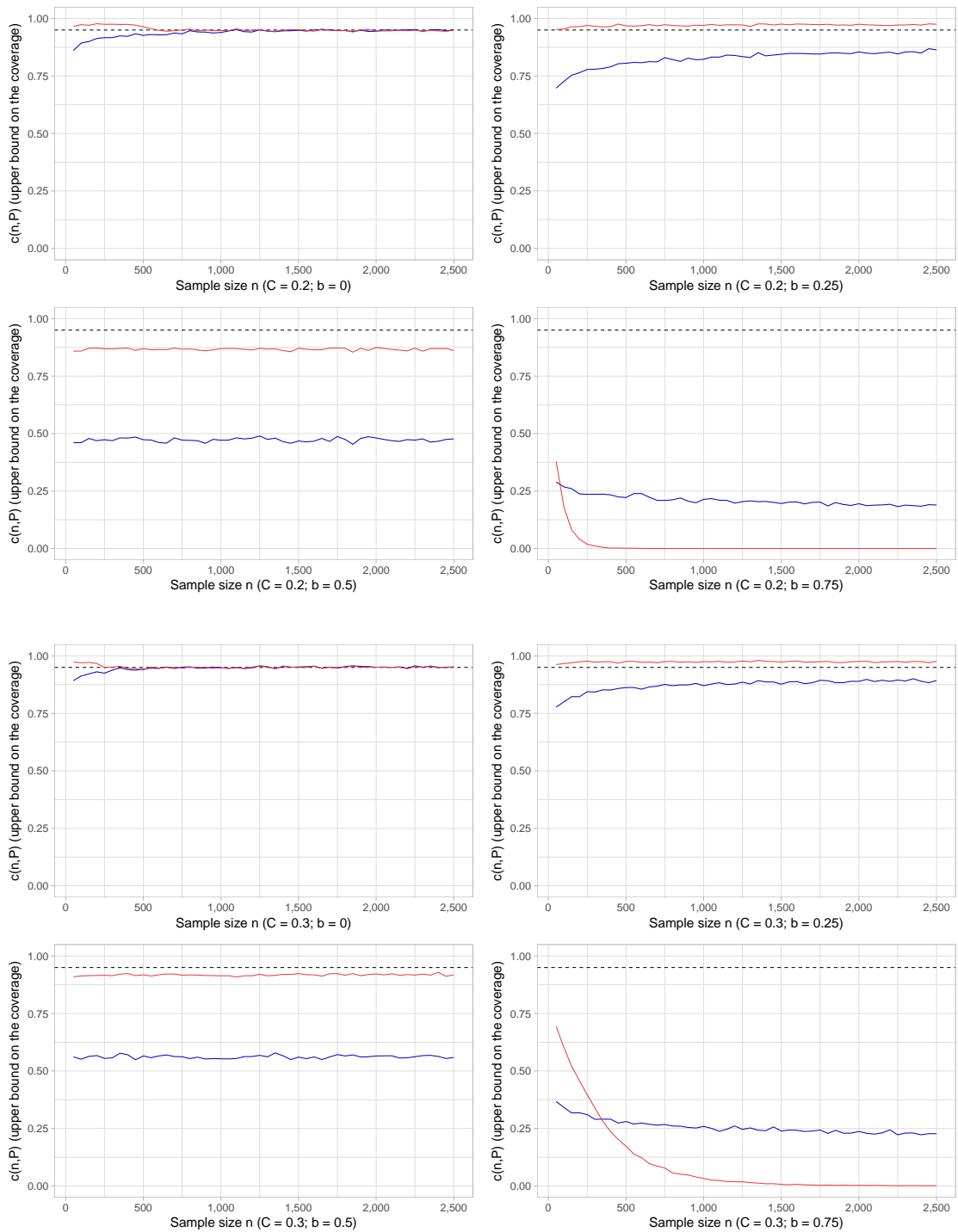Figure 25: delta method in blue; Efron's percentile bootstrap in red; $C = 0.2$.



Figure 26: delta method in blue; Efron's percentile bootstrap in red; $C = 0.3$.