CREST
CENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

WORKING PAPER SERIES

UNDER THE NULL OF VALID SPECIFICATION, PRE-TESTS CANNOT MAKE POST-TEST INFERENCE LIBERAL

Clément de Chaisemartin and Xavier D'Haultfoeuille

N°03 /January 2025

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

Under the null of valid specification, pre-tests cannot make post-test inference liberal*

Clément de Chaisemartin[†] Xavier D'Haultfœuille[‡]

Abstract

Consider a parameter of interest, which can be consistently estimated under some conditions. Suppose also that we can at least partly test these conditions with specification tests. We consider the common practice of conducting inference on the parameter of interest conditional on not rejecting these tests. We show that if the tested conditions hold, conditional inference is valid, though possibly conservative. This holds generally, without imposing any assumption on the asymptotic dependence between the estimator of the parameter of interest and the specification test.

1 Introduction

Let β_0 denote a parameter of interest. Let $\widehat{\beta}$ denote an estimator of β_0, which is consistent and asymptotically normal if the probability distribution generating the data, P_U, belongs to some set $\mathcal{P}_0$. Finally, assume that the null hypothesis of valid specification, i.e. $P_U \in \mathcal{P}_0$, is partly testable. In this paper, we consider the common practice of reporting $\widehat{\beta}$ and a confidence interval for β_0 only if a pre-test that $P_U \in \mathcal{P}_0$ is not rejected. One may wonder if pre-testing distorts inference. This note shows that under the null that $P_U \in \mathcal{P}_0$, conditional on not rejecting the pre-test the probability that β_0 belongs to its confidence interval (CI) is at least as large as the CI's nominal level. Therefore, conditional inference is valid, albeit possibly conservative. This result holds under general conditions. In particular, no restriction

*We would like to thank Giuseppe Cavaliere, Hannes Leeb, Fedor Petrov, Benedikt Pötscher and Jonathan Roth for their comments.

[†]Sciences Po Paris, clement.dechaisemartin@sciencespo.fr

[‡]CREST-ENSAE, xavier.dhaultfoeuille@ensae.fr.

on the asymptotic dependence between $\widehat{\beta}$ and the pre-tests is needed. The main condition on the pre-tests is that they need to be convex functionals. This holds for finite-dimensional F-tests, and for infinite-dimensional tests based on empirical processes such as (weighted) Kolmogorov-Smirnov or Crámer-von Mises tests. For finite-dimensional test statistics, our result is a direct application of the Gaussian correlation inequality (Royen, 2014). For infinite-dimensional ones, we rely on an approximation before applying this inequality.

Our result is relevant for statistical methods using observational data to study cause-and-effect relations. All those methods have to rely on an identifying assumption to recover treated units' unobserved counterfactual outcome, and researchers typically start by conducting specification tests to establish the credibility of that assumption. Specification tests can be continuity tests in a regression discontinuity design (RDD) study, balancing tests in an instrumental variable (IV) study, or pre-trends tests in a difference-in-difference (DID) study, to name a few examples. Our result applies to all those tests, and it implies that under the null of valid specification (parallel trends in DID, continuity in RDD, balancedness in IV...), inference conditional on those pre-tests is at worst conservative, but can never be liberal.

Inference conditional on not rejecting a specification pre-test has received some attention in the DID literature. There, our result implies that under the null of parallel trends, inference conditional on not rejecting a pre-trends test is at worst conservative, but can never be liberal, irrespective of the correlation structure between the DID and pre-trends estimators. This complements Propositions 1 and 4 in Roth (2022), which respectively show that under parallel trends, the DID estimator remains conditionally unbiased, and that its variance conditional on passing a centro-symmetric pre-test is smaller than the unconditional variance.[1] Our result also has implications for the comparison of the various DID estimators that have been proposed in binary and staggered designs with heterogeneous treatment effects (see, e.g., Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024). Borusyak et al. (2024) show, under some assumptions, that under the null of parallel trends their estimator leads to exact inference conditional on not rejecting a pre-trends test, which they view as an advantage over other estimators. Actually, our result implies that under the null of no pre-trends, all heterogeneity-robust estimators lead to exact or conservative inference conditional on not rejecting a pre-trends test.

[1] Together, these results do not imply that confidence intervals based on the unconditional variance are liberal, because the estimator's conditional distribution is non-normal.

The consequences of pre-testing under the null is a distinct issue from a few other questions that have been studied in the literature. A first is how to perform inference post model selection (see, e.g. Leeb and Pötscher, 2005, 2006). This is a distinct question because in our setting, there may be no estimation taking place when the pre-test is rejected. To understand this, consider a simple model selection set-up with two nested models. Then, the goal of the pre-test is to determine which of the two models will be used to estimate the coefficients on the variables included in the sparser model. The negative results in that literature are under local alternatives, where the true model is the richer one. Then, pre-testing can lead to liberal inference. Instead, under the null that the true model is the sparser one, our result implies that pre-testing cannot make inference liberal, and we also show in Appendix A that this positive result holds uniformly. Our result complements several other positive results in that literature, which have considered cases, such as nested linear models, where the estimator of the sparse model is asymptotically independent of the pre-test, so that conditional inference is exact under the null (see, e.g. Proposition 3.1 in Leeb and Pötscher, 2003). Instead, our result does not need to impose conditions ensuring asymptotic independence, but it only guarantees conservative inference.

A second distinct question from ours is about the power of pre-tests, which has been notably investigated by Roth (2022) in the context of DID estimation. There, he finds that in many of the applications he revisits in his simulations, such tests could fail to detect sizeable differential trends.

A third distinct issue, also related to inference post model selection, is about the effect of pre-testing on $\widehat{\beta}$ when the pre-test is incorrectly not rejected. In particular, does pre-testing lead to a larger bias and size distortion than not pre-testing and always using $\widehat{\beta}$? In the DID example, Roth (2022) conducts simulations tailored to 12 published articles in economics, assuming that the parallel-trends assumption does not hold due to differential linear trends that have 50% chances of being detected by a pre-trends test. Column 3 of his Table 3 shows that for 10 articles, the bias of $\widehat{\beta}$ conditional on wrongly not-rejecting the pre-trends test is larger than the unconditional bias. However, there are only two articles where the conditional bias is more than 20% larger than the unconditional one, and there are also two articles where the conditional bias is lower. Moreover, the pre-trends test still correctly leads the analyst to not use the DID estimator 50% of the time, which can also lead to a bias reduction if they then use a less-biased estimator instead. Overall, while our result gives a justification for pre-testing under the null, our reading of the simulations in

Roth (2022) is that in DID studies, pre-testing is rarely very harmful under the alternative. To our knowledge, no study has investigated whether pre-testing can be harmful under the alternative in other contexts, such as regression discontinuity designs, instrumental variable studies, or generalized method of moments estimation. Doing so is an interesting avenue for future research.

2 Set-up and examples

We are interested in a parameter $\beta_0 \in \mathbb{R}^p$. We use an estimator $\widehat{\beta}$, which is consistent and asymptotically normal if the probability distribution generating the data, P, belongs to some set $\mathcal{P}_0$. Finally, we assume that the null hypothesis of valid specification, i.e. $P \in \mathcal{P}_0$, has testable implications. We now review several instances where those conditions are met.

Example 1 (Difference-in-differences, DID) *We have a panel data set with $T > 2$ periods. Some units (the "treated") receive a binary treatment at a period $t_0 > 2$, whereas other units ("the controls") do not. We are interested in β_0, the average treatment effect on the treated after treatment occurs. β_0 could also be the vector of average treatment effects at t_0, $t_0 + 1$, ..., T. In this example, the null hypothesis of valid specification, i.e. $P \in \mathcal{P}_0$, is that the average counterfactual outcome evolutions without treatment of treated and control units are the same at all periods. Under that parallel-trends assumption, β_0 can be consistently estimated by $\widehat{\beta}$, the usual DID estimator. This parallel trends assumption also has $t_0 - 2$ testable implications: it implies that θ_0, a vector of DID estimands comparing the average outcome evolutions of treated and control units from $t_0 - 1$ to $t_0 - 2$, from $t_0 - 1$ to $t_0 - 3$, ..., and from $t_0 - 1$ to 1, is equal to 0.*

Example 2 (Regression discontinuity designs, RDD) *To simplify the discusssion, we focus on sharp RDD. We are interested in the effect of a binary treatment, where treatment occurs if and only if a running variable X is above a threshold, 0 say. The parameter of interest is $\beta_0 = E[Y(1) - Y(0)|X = 0]$, with $Y(d)$ denoting the potential outcome under treatment value d. If for all d $x \mapsto E[Y(d)|X = x]$ is continuous at 0, $\beta_0 := \lim_{x \downarrow 0} E[Y|X = x] - \lim_{x \uparrow 0} E[Y|X = x]$, which can be estimated by the difference of two nonparametric estimators (Hahn et al., 2001). Two usual specification tests are a test of continuity of the density of X at 0 (McCrary, 2008), and a test that predetermined variables W have continuous means at the threshold. In this example, the null hypothesis* *of valid specification could for instance be Conditions 1b and 2b in Lee (2008), which imply both the RD identifying assumption ($x \mapsto E[Y(d)|X = x]$ continuous at 0) and the null hypotheses of the two specification tests.*

Example 3 (Instrumental variables, IV) *We seek to estimate β_0, the local average treatment effect of an endogenous binary treatment D on an outcome Y, using a binary instrumental variable Z. In this example, the null hypothesis of valid specification is that the instrument is independent of potential outcomes and treatments, is excluded from the outcome equation, and has a monotonic effect on the treatment. Under those assumptions, β_0 is equal to the Wald estimand (Imbens and Angrist, 1994). Kitagawa (2015) shows that those assumptions have a testable implication:*

$$P(Y \leq y, D = d|Z = d) - P(Y \leq y, D = 1 - d|Z = d) \geq 0, \tag{1}$$

for any y in the support of Y and $d \in \{0, 1\}$. If Y is continuously distributed, the null of valid specification has infinitely many testable implications, which we can for instance test using the variance-weighted Kolmogorov-Smirnov test statistic proposed in Kitagawa (2015). In IV studies, researchers sometimes also test that some pre-determined covariates W are mean-independent of the instrument, a testable implication of a stronger null where Z is also assumed to be independent of the covariates.

Example 4 (Generalized method of moments) *We are interested in $\beta_0 \in \mathbb{R}^p$, which satisfies the moment conditions $E[g(U, \beta_0)] = 0$, where U denotes an observed random vector, and $g(u, \beta) \in \mathbb{R}^q$, $q \geq p$. The moment conditions identify β_0, and if $q > p$, the moment conditions are testable, using a standard J-test (Sargan, 1958; Hansen, 1982).*

Example 5 (Linear regressions with a non-binary, uncounfounded treatment) *Let D be a non-binary treatment whose support $\mathcal{D} \subset \mathbb{R}$ includes 0, let $Y(d)$ (resp. Y) denote the potential outcome corresponding to $d \in \mathcal{D}$ (resp. the observed outcome), and let $S = (Y - Y(0))/D$ denote units' potential outcome slope, between a treatment of zero and their actual treatment.*[2] *We are interested in $\beta_0 = E[S]$. Researchers sometimes use $\widehat{\beta}$, the coefficient on D in a linear regression of Y on D, to estimate β_0. In this example, the null hypothesis of valid specification is that $Y(0)$ and S are mean-independent of D. Then, $\widehat{\beta}$ is consistent for β_0. The null hypothesis has the following testable implication:*

$$E[Y|D] - (\alpha_0 + \beta_0 D) = 0, \tag{2}$$

[2] S should be understood as $Y'(0)$ if $D = 0$.

where $\alpha_0 = E[Y] - \beta_0 E[D]$. In settings where the treatment did not exist before the study period, and where one has access to outcome data Y_{-1} for a prior period, one can placebo test that $Y(0)$ is mean-independent of D, for instance by regressing Y_{-1} on D. Furthermore, Theorem 1 in de Chaisemartin et al. (2024) shows that under the maintained assumption that $Y(0)$ is mean-independent of D, (2) implies that S is mean-independent of D. Accordingly, de Chaisemartin et al. (2024) argue that one may report $\widehat{\beta}$ when a placebo test that $Y(0)$ is mean-independent of D and a test of (2) are not rejected. If D is continuously distributed, the null of valid specification has infinitely many testable implications, which we can for instance test using the Kolmogorov-Smirnov or Cramér-von Mises test of linearity of $E[Y|D]$ proposed in Stute (1997) and Stute et al. (1998).[3]

In order to encompass all the examples above, we consider the following two alternative assumptions. In Assumption 2, we allow for test statistics that are functionals of the empirical measure; some of our terminology and notation reflects this.

Assumption 1

1. *We observe a sample $(U_i)_{i=1,\dots,n}$ of identically distributed random vectors ($U_i \in \mathbb{R}^k$) with probability distribution $P_U(A) := P(U_1 \in A)$ for all Borel set $A \subset \mathbb{R}^k$.*

2. *If $P_U \in \mathcal{P}_0$, we have,*

$$\left(\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta} - \beta_0\right), \widehat{\Sigma}_\theta^{-1/2}\left(\widehat{\theta} - \theta_0\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} I_p & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}\right),$$

   where I_p is the identity matrix of size p, Σ_{12} is a matrix of size $p \times q$ and $\widehat{\Sigma}_\beta$ (resp. $\widehat{\Sigma}_\theta$ and Σ_{22}) is symmetric positive of size $p \times p$ (resp. $q \times q$).

3. *If $P_U \in \mathcal{P}_0$, we have $\theta_0 = 0$.*

4. *We consider J specification tests of $H_0 : P_U \in \mathcal{P}_0$ based on the statistics $(T_{1,n}, \dots, T_{J,n}) \in \mathbb{R}^J$ that satisfy, under H_0,*

$$T_{j,n} = T_j\left(\widehat{\Sigma}_\theta^{-1/2}\widehat{\theta}\right) + o_P(1). \quad (3)$$

[3] Another example is Angelini et al. (2024), who develop, in the context of proxy-SVARs, a pre-test of the null hypothesis of instrument relevance. This pre-test is asymptotically independent of the main estimators, thus leading to valid conditional inference. Our results show that other pre-tests could also lead to valid albeit possibly conservative inference, even if asymptotic independence does not hold.

Moreover, T_j is convex on $\mathbb{R}^q$ and satisfies $T_j(-x) = T_j(x) \geq 0$ for all $x \in \mathbb{R}^q$ and $T_j(0) = 0$. The critical region of the j-th test is $\{T_{j,n} > q_{j,n}\}$, where $q_{j,n}$ is a random variable satisfying $q_{j,n} \xrightarrow{P} q_j > 0$.

Assumption 2

1. *We observe a sample $(U_i)_{i=1,...,n}$ of i.i.d. random vectors with probability distribution $P_U(A) := P(U_1 \in A)$ for all Borel set A.*

2. *If $P_U \in \mathcal{P}_0$, we have*

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) = n^{-1/2} \sum_{i=1}^{n} \psi(U_i) + o_P(1), \tag{4}$$

   where $E[\psi(U_1)] = 0$ and $E[\|\psi(U_1)\|^2] < \infty$ and $V(\psi(U_1))$ is nonsingular.

3. *If $P_U \in \mathcal{P}_0$, we have $Pf := E[f(U_1)] = 0$ for all $f \in \mathcal{F}$, a Donsker class of real-valued functions.*[4]

4. *We consider J specification tests of $H_0 : P_U \in \mathcal{P}_0$ based on the statistics $(T_{1,n}, ..., T_{J,n}) \in \mathbb{R}^J$ that satisfy, under H_0,*

$$T_{j,n} = T_j(n^{1/2} P_n) + o_P(1), \tag{5}$$

   where P_n denotes the empirical measure of U, indexed by $\mathcal{F}$. Moreover, T_j is convex and satisfies $T_j(-P) = T_j(P) \geq 0$ and $T_j(0) = 0$.[5] *Also, T_j is continuous with respect to the uniform norm $\|P\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |Pf|$. The critical region of the j-th test is $\{T_{j,n} > q_{j,n}\}$, where $q_{j,n}$ is a random variable satisfying $q_{j,n} \xrightarrow{P} q_j > 0$.*

We first discuss the common aspects of Assumptions 1 and 2. First, we do not restrict in either of them the asymptotic dependence of $\widehat{\beta}$ and the test statistics $(T_{1,n}, ..., T_{J,n})$. Second, we assume in both cases that $\widehat{\beta}$ is consistent for β_0 if $P_U \in \mathcal{P}_0$, but we only test implications of $P_U \in \mathcal{P}_0$, rather than $P_U \in \mathcal{P}_0$ itself. Also, it could be that $\widehat{\beta}$ is actually consistent for β_0 under a weaker condition than $P_U \in \mathcal{P}_0$. This flexibility is important because the tested conditions are sometimes stronger than necessary to have that $\widehat{\beta}$ is consistent, see for

[4] For the definition of a Donsker class, see, e.g., Van der Vaart and Wellner (2023), p.130.

[5] Formally, T_j is defined on $\ell^\infty(\mathcal{F})$, the set of bounded functions from $\mathcal{F}$ to $\mathbb{R}$, and in (5), $n^{1/2}P_n$ is seen as an element of $\ell^\infty(\mathcal{F})$.

instance Examples 1 and 2. Third, the J different tests in Point 4 of Assumptions 1 and 2 reflect the fact that one often considers different specification tests, rather than a single test. Fourth, in both assumptions the functions $T_1, ..., T_J$ in the specification tests are convex and symmetric around zero. This condition is important to obtain our main result. It is typically satisfied in practice: all the examples below meet these requirements. Finally, the condition $q_j > 0$ at the end of Assumptions 1 and 2 simply ensures that the threshold of the test is not at the boundary of the support of the asymptotic distribution of $T_{j,n}$ under H_0. Note that we do not require the specification tests to be consistent or that they reach their nominal level asymptotically.

Now, let us discuss the differences between Assumption 1 and Assumption 2 and how they complement each other. On the one hand, Assumption 1 allows for estimators converging at rates lower than $n^{1/2}$, as is the case in the RDD example, and it also allows for non-independent (though identically distributed) variables, thus including the case of stationary time series. Such cases are ruled out in Assumption 2.[6] On the other hand, the specification tests in Assumption 1 are based on a finite-dimensional estimand θ_0. This is not the case in Assumption 2, where the class $\mathcal{F}$ in Point 3 may be infinite and even uncountable. The continuity condition in Point 4 of Assumption 2 is important in this potentially infinite-dimensional setup, to apply the continuous mapping theorem.

We now review five examples of specification tests. The first two fall under Assumption 1, while the last three fall under Assumption 2.

Specification Test 1 (DID) *In the DID example, for all* $t \in \{1, ..., t_0 - 2\}$ $\widehat{\theta}_t$ *is a DID estimator* t *periods prior to treatment, and* $\widehat{\theta} = (\widehat{\theta}_1, ..., \widehat{\theta}_{t_0-2})$. *Then, a so-called F-test for pre-trends corresponds to the case where* $J = 1$, $\widehat{\Sigma}_\theta$ *is a consistent estimator of the asymptotic variance of* $\widehat{\theta}$, $T_1(x) = x'x$, $T_{1,n} = T_1(\widehat{\Sigma}_\theta^{-1/2}\widehat{\theta})$, *and* $q_{1,n}$ *is the quantile of a chi-squared distribution with* q *degrees of freedom. Similarly, the so-called "Sup-t" test proposed by Montiel Olea and Plagborg-Møller (2019) corresponds to the case where* $J = 1$, $\widehat{\Sigma}_\theta$ *is as before, for all* $(x_1, ..., x_{t_0-2}) \in \mathbb{R}^{t_0-2}$ $T_1(x_1, ..., x_{t_0-2}) = \max(|x_1|, ..., |x_{t_0-2}|)$, $T_{1,n} = T_1\left(\widehat{\theta}_1/\widehat{\Sigma}_{\theta,1,1}, ..., \widehat{\theta}_{t_0-2}/\widehat{\Sigma}_{\theta,t_0-2,t_0-2}\right)$ *with* $\widehat{\Sigma}_{\theta,j,j}$ *denoting the* j*th diagonal element of* $\widehat{\Sigma}_\theta$, *and* $q_{1,n}$ *is the quantile of the max of* $t_0 - 2$ *normally distributed variables with mean 0 and variance* $\widehat{\Sigma}_\theta$.

[6] However, our result under Assumption 2 would go through as long as we can establish the weak convergence of the empirical process $n^{1/2}(P_n - P_U)$, indexed by $\mathcal{F}$, to a Gaussian process.

Specification Test 2 (RDD) *To test the continuity of the density of X at 0, one uses $\widehat{\theta}_1 := \lim_{x\downarrow 0} \widehat{f}_X(x) - \lim_{x\uparrow 0} \widehat{f}_X(z)$, with $\widehat{f}_X$ a nonparametric (e.g., kernel) estimator of f_X. To test that a predetermined scalar variable W has a continuous mean at the threshold, one uses $\widehat{\theta}_2 := \lim_{x\downarrow 0} \widehat{E}[W|X=x] - \lim_{x\uparrow 0} \widehat{E}[W|X=x]$, with $\widehat{E}[W|X=x]$ a nonparametric (e.g., local linear) estimator of $E[W|X=x]$. Then, $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2)'$. One may conduct the two specification tests separately, without adjusting for multiple testing. Then, $J=2$, $\widehat{\Sigma}_\theta$ is diagonal with j-th diagonal term equal to a consistent estimator of the asymptotic variance of $\widehat{\theta}_j$, $T_j(x) = x_j'x_j$, $T_{j,n} = T_j(\widehat{\Sigma}_\theta^{-1/2}\widehat{\theta})$ and $q_{j,n}$ is the quantile of a chi-squared distribution with one degree of freedom.*

Specification Test 3 (Weighted Kolmogorov-Smirnov test statistics) *Such test statistics correspond to $T_1(P) = \sup_{f\in\mathcal{F}} w(f)|Pf|$ for some Donsker clas $\mathcal{F}$. This includes for instance the test statistic of Kitagawa (2015), used in Example 3, and where $\mathcal{F}$ is the set of intervals of the real line.*[7] *If we assume that for all $f \in \mathcal{F}$, $0 < \underline{w} \le w(f) \le \overline{w} < \infty$, then T_1 satisfies the restrictions in Point 4 of Assumption 2. To see that it is continuous, fix P and $\delta > 0$ and let Q be such that $\|P-Q\|_\mathcal{F} < \delta$. Let f^* be such that $w(f^*)|Pf^*| > T_1(P) - \delta$. Then,*

$$\begin{aligned} T_1(P) - T_1(Q) &< w(f^*)|Pf^*| - w(f^*)|Qf^*| + \delta \\ &\le \overline{w} \sup_{f\in\mathcal{F}} |(P-Q)(f)| + \delta \\ &< (\overline{w}+1)\,\delta. \end{aligned}$$

By considering f^{**} be such that $w(f^{**})|Qf^{**}| > T_1(Q) - \delta$, we obtain the same inequality for $T_1(Q) - T_1(P)$. Hence, $|T_1(P) - T_1(Q)| < (\overline{w}+1)\,\delta$, which implies that T_1 is continuous.

Specification Test 4 (Generalized Crámer-von Mises statistic) *Let us assume that $\mathcal{F}$ is in bijection with $\mathcal{X} \subset \mathbb{R}^k$, so that any f can be indexed by $x \in \mathbb{R}^k$; accordingly, we denote it by f_x. Let μ denote a probability measure on $\mathcal{X}$. Then the statistics corresponds to $T_1(P) = \int (Pf_x)^2 d\mu(x)$. T_1 satisfies the restrictions in Point 4 of Assumption 2. To see that it is continuous, fix P and $\delta > 0$ and let Q be such that $\|P-Q\|_\mathcal{F} < \delta$. Then,*

$$\begin{aligned} |(Pf_x)^2 - (QPf_x)^2| &= |(Pf_x + Qf_x) \times (P-Q)f_x| \\ &\le (\|P\|_\mathcal{F} + \|Q\|_\mathcal{F})\,\|P-Q\|_\mathcal{F} \\ &< (2\|P\|_\mathcal{F} + \delta)\,\delta. \end{aligned}$$

[7] In his case, $T_{1,n} = \sup_{f\in\mathcal{F}} w_n(f)|Pf|$ for a random $w_n(\cdot)$, but w_n converges uniformly to some deterministic w, so (5) holds in his context.

As a result,

$$|T_1(P) - T_1(Q)| \leq \int |(Pf_x)^2 - (QPf_x)^2| d\mu(x) < (2\|P\|_{\mathcal{F}} + \delta)\, \delta,$$

which implies that T_1 is continuous.

Specification Test 5 (Kolmogorov-Smirnov test statistics with nuisance parameters) *In some cases, the restrictions $Pf = 0$ hold for $f \in \mathcal{F}_{\eta_0}$, where η_0 is unknown but can be consistently estimated by $\widehat{\eta}$. In Example 5, for instance, $\eta_0 = (\alpha_0, \beta_0)$, $\mathcal{F}_\eta = \{f_{d',\eta} : (y, d) \mapsto \mathbb{1}\{d \leq d'\}(y - \eta_1 - \eta_2 d),\ d' \in \mathcal{D}\}$ and $\widehat{\eta}$ is the OLS estimator in the regression of Y on $X := (1, D)'$. In such cases, Assumption 2 can still hold under regularity conditions, provided that we replace $\mathcal{F}_{\eta_0}$ by another appropriate class. To see this, consider Example 5 again, with a Kolmogorov-Smirnov test statistic (the same reasoning applies to a Cramér-von Mises test statistic). Then $T_{1,n} = n^{1/2} \sup_{d \in \mathcal{D}} |\widehat{\theta}(d)|$, with*

$$\widehat{\theta}(d) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{D_i \leq d\}(Y_i - X_i'\widehat{\eta}).$$

By a uniform law of large numbers on $\sum_{i=1} \mathbb{1}\{D_i \leq d\} X_i/n$ and standard results on regressions, we obtain

$$\sqrt{n}\left(\widehat{\theta}(d) - \theta_0(d)\right) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\mathbb{1}\{D_i \leq d\} + E[\mathbb{1}\{D \leq d\} X'] E[XX']^{-1} X_i\right] (Y_i - X_i'\eta_0) + o_P(1),$$

where the $o_P(1)$ is uniform over $d \in \mathcal{D}$. Let us define

$$\widetilde{\mathcal{F}} := \left\{ \widetilde{f}_{d'} : (y, d) \mapsto \left[\mathbb{1}\{d \leq d'\} + E[\mathbb{1}\{D \leq d'\} X'] E[XX']^{-1}(1, d)'\right] (y - \alpha_0 - \beta_0 d), \right.$$
$$\left. d' \in \mathcal{D} \right\}.$$

Note that we still have $Pf = 0$ for all $f \in \widetilde{\mathcal{F}}$, and like $\mathcal{F}_\eta$, $\widetilde{\mathcal{F}}$ is Donsker. Moreover, by continuity of the supremum functional, $T_{1,n} = T_1(P) + o_P(1)$, with $T_1(P) = \sup_{f \in \widetilde{\mathcal{F}}} |Pf|$. Hence, Point 3 of Assumption 2 and the restrictions on $T_{1,n}$ in Point 4 of that assumption hold with $\mathcal{F} = \widetilde{\mathcal{F}}$. To obtain similar results in other, potentially more complicated setups, see for instance Chapter 13 in Van der Vaart and Wellner (2023).

3 Results

Our first result gives a control on the asymptotic probability that the properly normalized estimator of interest belongs to a convex set, conditional on not rejecting the specification tests.

Theorem 1 *Suppose that $P_U \in \mathcal{P}_0$. Then, for any convex set C that is symmetric around the origin,*

1. *If Assumption 1 holds,*

$$\lim_{n\to\infty} P\left[\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta}-\beta_0\right)\in C \middle| T_{1,n}\leq q_{1,n},...,T_{J,n}\leq q_{J,n}\right]\geq P(Z_1\in C),$$

where $Z_1 \sim \mathcal{N}(0, I_p)$.

2. *If Assumption 2 holds,*

$$\lim_{n\to\infty} P\left[\sqrt{n}\left(\widehat{\beta}-\beta_0\right)\in C \middle| T_{1,n}\leq q_{1,n},...,T_{J,n}\leq q_{J,n}\right]\geq P(Z_2\in C), \tag{6}$$

where $Z_2 \sim \mathcal{N}(0, V(\psi))$.[8]

Heuristically, the proof of the first point goes as follows. First, by convexity and symmetry of the $(T_j)_{j=1,...,J}$, the event $\{T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\}$ is equivalent, if we neglect the remainder term in (5), to $\widehat{\Sigma}_\theta^{-1/2}\widehat{\theta} \in C'$, for some convex set C' that is symmetric around the origin. Second, as

$$\left(\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta}-\beta_0\right), \widehat{\Sigma}_\theta^{-1/2}\left(\widehat{\theta}-\theta_0\right)\right)$$

is asymptotically normal, under the null of valid specification we have that

$$\left(\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta}-\beta_0\right), \widehat{\Sigma}_\theta^{-1/2}\widehat{\theta}\right)$$

is asymptotically normal, with a centered Gaussian distribution μ. Finally, for any convex set C symmetric around the origin,

$$\begin{aligned}
&P\left(\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta}-\beta_0\right)\in C, \widehat{\Sigma}_\theta^{-1/2}\widehat{\theta}\in C'\right)\\
=&P\left(\left(\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta}-\beta_0\right), \widehat{\Sigma}_\theta^{-1/2}\widehat{\theta}\right)\in (C\times\mathbb{R}^q)\cap(\mathbb{R}^p\times C')\right)\\
\approx&\mu\left((C\times\mathbb{R}^q)\cap(\mathbb{R}^p\times C')\right)\\
\geq&\mu\left(C\times\mathbb{R}^q\right)\times\mu\left(\mathbb{R}^p\times C'\right)\\
\approx&P\left(\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta}-\beta_0\right)\in C\right)\times P\left(\widehat{\Sigma}_\theta^{-1/2}\widehat{\theta}\in C'\right).
\end{aligned}$$

[8] Assumption 2 does not guarantee that the test statistics $T_{j,n}$ are measurable. If not, the result holds replacing P on the left-hand side of (6) by the outer probability P^* (see, e.g. Van der Vaart and Wellner, 2023, p.6). The same remark applies to Corollary 1 below.

The key step in the proof, the inequality, follows from the Gaussian correlation inequality (Royen, 2014), which states that for any centered Gaussian distribution μ on $\mathbb{R}^K$ and two convex sets E and F that are symmetric about the origin,

$$\mu(E \cap F) \geq \mu(E)\mu(F).$$

The first point of Theorem 1 follows from dividing both sides of the previous display by $P\left(\widehat{\Sigma}_\theta^{-1/2}\widehat{\theta} \in C'\right)$, the probability that the specification test is not rejected. The second point of Theorem 1 follows from the same tools, but also from the approximation of the limit distribution of $T_j(n^{1/2}P_n)$ by a convex function of a finite-dimensional Gaussian measure.

We now turn to the implications of Theorem 1 for (conditional) inference on β_0. Let $\widehat{V} := \widehat{\Sigma}_\beta$ if Assumption 1 holds; otherwise, let $\widehat{V}$ denote an estimator of $V(\psi(U_1))/n$. We consider the usual F-test of $\beta_0 = b_0$ with test statistic $F_n(b_0) := (\widehat{\beta} - b_0)'\widehat{V}^{-1}(\widehat{\beta} - b_0)$ and critical region $\{F_n(b_0) > q_{1-\alpha}(p)\}$, with $q_{1-\alpha}(p)$ the quantile of order $1 - \alpha$ of a chi-squared distribution with p degrees of freedom. We also consider the standard confidence region

$$\text{CR}_{1-\alpha} = \left\{b \in \mathbb{R}^p : (\widehat{\beta} - b)'\widehat{V}^{-1}(\widehat{\beta} - b) \leq q_{1-\alpha}(p)\right\}.$$

Corollary 1 *Suppose that either Assumption 1 or Assumption 2 holds, $P_U \in \mathcal{P}_0$ and, if Assumption 2 holds, $n\widehat{V} \xrightarrow{P} V(\psi(U_1))$. Then:*

1. *If $\beta_0 = b_0$, $\lim_{n\to\infty} P\left(F_n(b_0) > q_{1-\alpha}(p) | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \leq \alpha$.*
2. *$\lim_{n\to\infty} P\left(\beta_0 \in CR_{1-\alpha} | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \geq 1 - \alpha$.*

Hence, under Assumption 2 and provided that $P_U \in \mathcal{P}_0$, conditional on accepting specification tests the usual F-tests and confidence regions are asymptotically conservative. In the common case of inference on a single coefficient ($p = 1$), this implies that two-sided tests and confidence intervals on β_0 are asymptotically conservative.

On the other hand, Corollary 1 does not provide guarantees for one-sided tests and confidence intervals. It turns out, however, that we can obtain such guarantees by the following consequence of the Gaussian correlation inequality.[9]

Lemma 1 *Let μ denote a mean zero Gaussian measure on $\mathbb{R}^{q+1}$, $K = (-\infty, a] \times \mathbb{R}^q$ with $a \geq 0$ and $L = \mathbb{R} \times C$ where $C \subset \mathbb{R}^q$ is convex and symmetric around the origin. Then $\mu(K \cap L) \geq \mu(K)\mu(L)$. The same result holds with $K = [-a, \infty) \times \mathbb{R}^q$.*

[9]We thank Fedor Petrov for giving us the proof of this result.

In this unidimensional case ($p = 1$), we define the test statistic $T_n(b_0) := (\widehat{\beta} - b_0)/\widehat{V}^{1/2}$ and the critical region $\{T_n(b_0) > z_{1-\alpha}\}$, with $z_{1-\alpha}$ the quantile of order $1-\alpha$ of a standard normal distribution (of course, the same result would hold with instead the critical region $\{T_n(b_0) < z_\alpha\}$). Accordingly, we consider the unilateral confidence interval

$$\mathrm{CI}_{1-\alpha} = (-\infty, \widehat{\beta} + \widehat{V}^{1/2} z_{1-\alpha}].$$

Proposition 1 *Suppose that $\alpha \in (0, 1/2]$, $p = 1$ and the assumptions in Corollary 1 hold. Then:*

1. *If $\beta_0 = b_0$, $\lim_{n\to\infty} P\left(T_n(b_0) > z_{1-\alpha}(p) | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \leq \alpha$.*

2. *$\lim_{n\to\infty} P\left(\beta_0 \in CI_{1-\alpha} | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \geq 1 - \alpha$.*

So far, we have considered pointwise results, where the distribution P_U does not depend on n, the sample size. In Appendix A, we consider a setup where P_U may vary with the sample size. We show, under a suitable adaptation of Assumption 1, that the same results as above hold, but now uniformly over $\mathcal{P}_0$.

References

Angelini, G., G. Cavaliere, and L. Fanelli (2024). An identification and testing strategy for proxy-svars with weak proxies. *Journal of Econometrics 238*(2), 105604.

Borusyak, K., X. Jaravel, and J. Spiess (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies forthcoming*.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of econometrics 225*(2), 200–230.

Davydov, Y. A., M. A. Lifshits, et al. (1998). *Local properties of distributions of stochastic functionals*. American Mathematical Soc.

de Chaisemartin, C., D. Ciccia, X. D'Haultfœuille, and F. Knau (2024). Two-way fixed effects and differences-in-differences estimators in heterogeneous adoption designs. arXiv preprint arXiv:2405.04465.

Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society 50*(4), 1029–1054.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*, 467–475.

Kitagawa, T. (2015). A test for instrument validity. *Econometrica 83*(5), 2043–2063.

Lang, R. (1986). A note on the measurability of convex sets. *Archiv der Mathematik 47*, 90–92.

Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics 142*(2), 675–697.

Leeb, H. and B. M. Pötscher (2003). The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory 19*(1), 100–142.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory 21*(1), 21–59.

Leeb, H. and B. M. Pötscher (2006). Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics 34*(1), 2554–2591.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics 142*(2), 698–714.

Montiel Olea, J. L. and M. Plagborg-Møller (2019). Simultaneous confidence bands: Theory, implementation, and an application to svars. *Journal of Applied Econometrics 34*(1), 1–17.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights 4*(3), 305–322.

Royen, T. (2014). A simple proof of the gaussian correlation conjecture extended to multivariate gamma distributions. *Far East Journal of Theoretical Statistics 48*, 139–145.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the econometric society 26*(3), 393–415.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics 25*(2), 613–641.

Stute, W., W. G. Manteiga, and M. P. Quindimil (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association 93*(441), 141–149.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics 225*(2), 175–199.

van der Vaart, A. (2000). *Aymptotic Statistics*. Cambridge University Press.

Van der Vaart, A. W. and J. A. Wellner (2023). *Weak convergence and empirical processes*. Springer.

A Uniform result

We consider here a setting where the probability distribution of the data changes with the sample size; accordingly, we denote it by P_{U_n} instead of P_U. In this setup, Assumption 3 is the exact counterpart of Assumption 1. Note that we emphasize the dependence of the true parameters in P_{U_n} by denoting them now by $\beta(P_{U_n})$ and $\theta(P_{U_n})$.

Assumption 3

1. *For each $n \geq 1$, we observe a sample $(U_{n,i})_{i=1,...,n}$ of identically distributed random vectors ($U_{n,i} \in \mathbb{R}^k$) with probability distribution $P_{U_n}(A) := P(U_{n,1} \in A)$ for all Borel set $A \subset \mathbb{R}^k$.*

2. *For any sequence $(P_{U_n})_{n\geq 1}$ with $P_{U_n} \in \mathcal{P}_0$ for all $n \geq 1$, we have*
$$\left(\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta} - \beta(P_{U_n})\right), \widehat{\Sigma}_\theta^{-1/2}\left(\widehat{\theta} - \theta(P_{U_n})\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} I_p & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}\right),$$
where we use the same notation on matrices as in Assumption 1.

3. *If $P_{U_n} \in \mathcal{P}_0$, we have $\theta(P_{U_n}) = 0$.*

4. *We consider J specification tests of $H_0 : P_{U_n} \in \mathcal{P}_0$ based on the statistics $(T_{1,n}, ..., T_{J,n}) \in \mathbb{R}^J$ that satisfy, under H_0,*
$$T_{j,n} = T_j\left(\widehat{\Sigma}_\theta^{-1/2}\widehat{\theta}\right) + o_P(1),$$
and T_j satisfies the same restrictions as in Assumption 1. The critical region of the j-th test is $\{T_{j,n} > q_{j,n}\}$, where the random variables $q_{j,n}$ satisfy $q_{j,n} \xrightarrow{P} q_j > 0$.

Theorem 2 *Suppose that Assumption 3 holds and $P_{U_n} \in \mathcal{P}_0$ for all $n \geq 1$. Then, for any convex set C that is symmetric around the origin,*
$$\lim_{n\to\infty} P\left[\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta} - \beta(P_{U_n})\right) \in C \middle| T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right] \geq P(Z_1 \in C),$$
where $Z_1 \sim \mathcal{N}(0, I_p)$. Moreover,
$$\limsup_{n\to\infty} \sup_{P_{U_n}\in\mathcal{P}_0} P\left(F_n(\beta(P_{U_n})) > q_{1-\alpha}(p) | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \leq \alpha, \tag{7}$$
$$\liminf_{n\to\infty} \inf_{P_{U_n}\in\mathcal{P}_0} P\left(\beta(P_{U_n})) \in CR_{1-\alpha} | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \geq 1 - \alpha. \tag{8}$$
Finally, if $p = 1$ (7) (resp. (8)) also holds true with $T_n(\beta(P_{U_n}))$ instead of $F_n(\beta(P_{U_n}))$ (resp. $CI_{1-\alpha}$ instead of $CR_{1-\alpha}$).

The first part of Theorem 2 establishes that Point 1 of Theorem 1 holds along any sequence $(P_{U_n})_{n\geq 1}$ so long as $P_{U_n} \in \mathcal{P}_0$. The second and third results are similar to those of Corollary 1, but now hold uniformly over $\mathcal{P}_0$.

B Proofs

B.1 Theorem 1

Point 2

We begin by the proof of Point 2 as this is more complicated case. Let us define $Z_n := \sqrt{n}(\widehat{\beta} - \beta_0)$, $T_n := \max_{j=1,...,J}[T_{j,n} - q_{j,n}]$ and $T(P) := \max_{j=1,...,J}[T_j(P) - q_j]$. T is convex and continuous, as the maximum of J convex and continuous functionals. Moreover,

$$P(Z_n \in C, T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}) = P(Z_n \in C, T_n \leq 0).$$

By Points 3 and 4 of Assumption 2,

$$(Z_n, T_n) = (G_n\psi, T(n^{1/2}(P_n - P)) + o_P(1). \tag{9}$$

Let $\mathcal{G} = \mathcal{F} \cup \{\psi^1, ..., \psi^d\}$. As union of Donsker classes, $\mathcal{G}$ is Donsker (see Van der Vaart and Wellner, 2023, p.136). Hence, the empirical process $G_n := n^{1/2}(P_n - P)$ indexed by $\mathcal{G}$ is asymptotically Gaussian; we denote by G its limit. By (9) and the continuous mapping theorem,

$$(Z_n, T_n) \xrightarrow{d} (G\psi, T(G)).$$

Let ∂A denote the boundary of the set A. The boundary of $C \times (-\infty, 0]$ satisfies

$$\partial(C \times (-\infty, 0]) = (\partial C \times (-\infty, 0]) \cup (C \times \{0\}). \tag{10}$$

Moreover, since C is convex, ∂C has Lebesgue measure zero (see, e.g. Lang, 1986, p.90). Since $V(\psi(U_1))$ is nonsingular, the distribution of $G\psi$ is absolutely continuous with respect to the Lebesgue measure. Hence, $P(G\psi \in \partial C) = 0$. By Theorem 11.1 and Problem 11.3 in Davydov et al. (1998) and continuity of T, the cumulative distribution function H of $T(G)$ is strictly increasing on (r_0, ∞), with $r_0 := \inf_{P \in \ell^\infty(\mathcal{F})} T(P) = \max_{j=1,...,J} -q_j < 0$. Hence, H is continuous at 0, $H(0) > 0$ and $P(T(G) = 0) = 0$. Hence, in view of (10),

$$P((G\psi, T(G)) \in \partial(C \times (-\infty, 0])) = 0.$$

Thus, by Portmanteau's lemma (see, e.g., Lemma 18.9 in van der Vaart, 2000),

$$\lim_{n\to\infty} P^*(Z_n \in C, T_n \leq 0) = P(G\psi \in C, T(G) \leq 0),$$

where P^* denotes the outer probability. Similarly, $\lim_{n\to\infty} P^*(T_n \leq 0) = P(T(G) \leq 0) = H(0) > 0$. Hence,

$$\lim_{n\to\infty} P^*(Z_n \in C | T_n \leq 0) = P(G\psi \in C | T(G) \leq 0). \tag{11}$$

Now, fix $\varepsilon > 0$. By continuity of T, there exists $\eta > 0$ such that for any $H \in \ell^\infty(\mathcal{F})$, $\|G - H\|_{\mathcal{F}} < \eta$ implies $|T(G) - T(H)| < \varepsilon$. Let $\rho_P(f) = P[(f - Pf)^2]^{1/2}$. Since $\mathcal{F}$ is totally bounded (see Van der Vaart and Wellner, 2023, pp. 138-139), there exist $K \geq 1$ and $(f_1, ..., f_K) \in \mathcal{F}^K$ such that $\forall f \in \mathcal{F}$, $\min_{k=1,...,K} \rho_P(f, f_k) < \eta$. Let $\widetilde{G}(f) = G(\widetilde{f})$, where $\widetilde{f} = \arg\min_{g\in\{f_1,...,f_K\}} \rho_P(f - g)$. Then,

$$|(G - \widetilde{G})f| = |P(f - \widetilde{f} - P(f - \widetilde{f}))| \leq \rho_P(f - \widetilde{f}) < \eta.$$

Hence, $\|G - \widetilde{G}\|_{\mathcal{F}} < \eta$, which implies $|T(G) - T(\widetilde{G})| < \varepsilon$. As a result,

$$P(G\psi \in C \,|\, T(G) \leq 0) \geq \frac{P(G\psi \in C,\ T(\widetilde{G}) \leq -\varepsilon)}{H(0)}.$$

We can write $T(\widetilde{G})$ as $\widetilde{T}_K(Gf_1, ..., Gf_K)$ for some convex function $\widetilde{T}_K$. Let μ denote the Gaussian distribution of $(G\psi, Gf_1, ..., Gf_K)$. μ is a centered Gaussian distribution. Let $E = C \times \mathbb{R}^K$ and $F = \mathbb{R}^p \times \widetilde{T}^{-1}((-\infty, q - \varepsilon])$. Both E and F are convex. Moreover, they are symmetric about the origin since $T(P) = T(-P)$. Then, by the Gaussian correlation inequality (Royen, 2014),

$$\begin{aligned} P(G\psi \in C,\ T(\widetilde{G}) \leq -\varepsilon) &= \mu(E \cap F) \\ &\geq \mu(E)\mu(F) \\ &= P(Z_2 \in C)P(T(\widetilde{G}) \leq -\varepsilon), \end{aligned}$$

where the last equality follows by definition of Z_2. Hence,

$$\begin{aligned} P(G\psi \in C \,|\, T(G) \leq 0) \geq& P(Z_2 \in C)\frac{P(T(\widetilde{G}) \leq -\varepsilon)}{H(0)} \\ \geq& P(Z_2 \in C)\frac{P(T(G) \leq -2\varepsilon)}{H(0)} \\ \geq& P(Z_2 \in C)\frac{H(-2\varepsilon)}{H(0)}. \end{aligned}$$

Since H is continuous at 0 and ε was arbitrary, we finally obtain

$$P(G\psi \in C \,|\, T(G) \leq 0) \geq P(Z_2 \in C).$$

The result follows by combining the last display with (11).

Point 1

The begining of the proof is almost the same as that in Point 2, so we just highlight the differences. First, since T_j is a convex function on $\mathbb{R}^q$, it is continuous. Hence, $T(x) := \max_{j=1,...,J}[T_j(x) - q_j]$ is continuous (and convex). Then, reasoning as above, we obtain

$$\lim_{n\to\infty} P\left[\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta} - \beta_0\right) \in C \,\middle|\, T_n \leq 0\right] = P(Z_1 \in C \,|\, T(Z_1') \leq 0),$$

where $T_n := \max_{j=1,...,J}[T_{j,n} - q_{j,n}]$ and

$$(Z_1, Z_1') \sim \mathcal{N}\left(0, \begin{pmatrix} I_p & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}\right).$$

The result follows directly by the Gaussian correlation inequality.

B.2 Proof of Corollary 1

Point 2 follows directly from Point 1 so we focus on the latter. Suppose first that Assumption 1 holds. Let $C = \{x \in \mathbb{R}^p : x'x \leq q_{1-\alpha}(p)\}$. Then, if $\beta_0 = b_0$,

$$F_n(b_0) \leq q_{1-\alpha}(p) \Leftrightarrow \widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta} - \beta_0\right) \in C.$$

Because C is convex and symmetric around the origin, we have, by Theorem 1,

$$\begin{aligned}
&\lim_{n\to\infty} P\left[F_n(b_0) \leq q_{1-\alpha}(p) \middle| T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right] \\
=&\lim_{n\to\infty} P\left[\widehat{\Sigma}_\beta^{-1/2}\left(\widehat{\beta} - \beta_0\right) \in C \,\middle|\, T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right] \\
\geq& P(Z_1 \in C) = 1 - \alpha.
\end{aligned}$$

The result follows in this case.

Now, suppose that Assumption 2 holds. Because $n\widehat{V} \xrightarrow{P} V(\psi(U_1))$ and $V(\psi(U_1))$ is non-singular, we have, if $\beta_0 = b_0$,

$$F_n(b_0) = n(\widehat{\beta} - \beta_0)'V(\psi(U_1))^{-1}(\widehat{\beta} - \beta_0) + o_P(1).$$

Let $C = \{x : x'V(\psi(U_1))^{-1}x \leq q_{1-\alpha}(p)\}$. C is convex and symmetric around the origin. Then, by the previous display and Theorem 1,

$$\begin{aligned}
&\lim_{n\to\infty} P\left(F_n(b_0) \leq q_{1-\alpha}(p) \,\middle|\, T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \\
=&\lim_{n\to\infty} P\left(\sqrt{n}(\widehat{\beta} - \beta_0) \in C \,\middle|\, T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \\
\geq& P(Z_2 \in C) = 1 - \alpha.
\end{aligned}$$

The result follows.

B.3 Proof of Lemma 1

Let $K_0 = [-a, a] \times \mathbb{R}^{k-1}$, $K_- = (-\infty, -a] \times \mathbb{R}^{k-1}$ and $K_+ = [a, \infty) \times \mathbb{R}^{k-1}$. By the Gaussian correlation inequality and symmetry of μ,

$$\begin{aligned}
\mu(L)\left[\mu(K_0) + 2\mu(K_+)\right] &= \mu(L) \\
&= \mu(L \cap K_0) + \mu(L \cap K_-) + \mu(L \cap K_+) \\
&\geq \mu(L)\mu(K_0) + 2\mu(L \cap K_+).
\end{aligned}$$

Hence, $\mu(L)\mu(K_+) \geq \mu(L \cap K_+)$. As a result,

$$\begin{aligned}
\mu(L \cap K) &= \mu(L) - \mu(L \cap K_+) \\
&\geq \mu(L) - \mu(L)\mu(K_+) \\
&= \mu(L)\mu(K).
\end{aligned}$$

B.4 Proof of Proposition 1

By applying Lemma 1 instead of the initial Gaussian correlation inequality, we obtain the same result as Theorem 1 but with $C = (-\infty, a]$. Then, the same reasoning as in Corollary 1 leads to the result.

B.5 Proof of Theorem 2

The proof the first result is exactly the same as that of Point 1 of Theorem 1. To show (7), fix $\eta > 0$ and pick a sequence $(Q_{U_n})_{n\geq 1}$ in $\mathcal{P}_0$ such that for all $n \geq 1$,

$$\begin{aligned} & P\left(F_n(\beta(Q_{U_n})) > q_{1-\alpha}(p) | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \\ \geq & \sup_{P_{U_n}\in\mathcal{P}_0} P\left(F_n(\beta(P_{U_n})) > q_{1-\alpha}(p) | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) - \eta. \end{aligned} \tag{12}$$

Applying the same reasoning as that used to prove Point 1 of Corollary 1, we obtain

$$\lim_{n\to\infty} P\left(F_n(\beta(Q_{U_n})) > q_{1-\alpha}(p) | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \leq \alpha.$$

Combined with (12), this yields

$$\limsup_{n\to\infty} \sup_{P_{U_n}\in\mathcal{P}_0} P\left(F_n(\beta(P_{U_n})) > q_{1-\alpha}(p) | T_{1,n} \leq q_{1,n}, ..., T_{J,n} \leq q_{J,n}\right) \leq \alpha + \eta.$$

Equation (7) follows since η was arbitrary. The adaptation to $T_n(\beta(P_{U_n}))$, Equation (8) and its adaptation to $\text{CI}_{1-\alpha}$ instead of $\text{CR}_{1-\alpha}$ follow with the same reasoning.

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/