A Appendix

A.1 Notation

- Throughout this appendix, "a.s." abbreviates almost surely, "i.p." abbreviates in probability, "i.o.p." abbreviates in outer probability, "w.p.a.1" abbreviates with probability approaching one, "w.o.p.a.1" abbreviates with outer probability approaching one, "SLLN" refers to the strong law of large numbers, "CLT" refers to the central limit theorem and "LIL" refers to the law of iterated logarithm.

- For any $\theta \in \Theta$, we denote $v(m_\theta) = \sqrt{n}\left(\mathbb{E}_n\left(m\left(Z,\theta\right)\right) - \mathbb{E}\left(m\left(Z,\theta\right)\right)\right).$ For any $(\theta, j) \in \Theta \times \{1,...,J\}$, we denote $v_n\left(m_{j,\theta}\right) = \sqrt{n}\left(\mathbb{E}_n\left(m_j\left(Z,\theta\right)\right) - \mathbb{E}\left(m_j\left(Z,\theta\right)\right)\right).$

- We refer to the space of bounded functions that map Θ onto $\mathbb{R}^J$ as $l_J^\infty(\Theta)$ and the space of continuous functions that map Θ onto $\mathbb{R}^J$ as $C_J(\Theta)$. For both spaces, we use the uniform metric, denoted by $\|y\|_\infty$, that is, $\forall y \in l_J^\infty(\Theta)$, $\|y\|_\infty = \sup_{\theta\in\Theta}\|y(\theta)\|$. For matrix spaces, we use the Frobenius norm, that is, $\forall M \in \mathbb{R}^{I\times J}$, $\|M\| = (\sum_{i=1}^{I}\sum_{j=1}^{J} M_{(i,j)}^2)^{1/2}$. Finally, $\mathbf{I}_J$ denotes the $J \times J$ identity matrix and $0_{I\times J}$ denotes the null matrix with I rows and J columns.

- For any $s \in \mathbb{N}$, the space of Borel measurable convex sets in $\mathbb{R}^s$ is denoted by $\mathcal{C}_s$. For any function $H : A_1 \to A_2$, and any set $S \subset A_2$, $H^{-1}(S) = \{x \in A_1 : H(x) \in S\}$. For any $\varepsilon > 0$ and $\forall S \subset \mathbb{R}^s$, $S^\varepsilon = \{x \in \mathbb{R}^s : \exists x' \in S \cap \|x - x'\| \leq \varepsilon\}$ and ∂S denotes the boundary of S.

- For any set of finite elements A, $\{\mathcal{P}^A/\varnothing\}$ denotes the set of all non-empty subsets of A.

- For any square matrix $\Sigma \in \mathbb{R}^{J\times J}$ and any Borel measurable set $A \subseteq \mathbb{R}^J$, $\Phi_\Sigma(A)$ denotes $P(Z \in A)$ where $Z \sim N(0,\Sigma)$. For any square matrix $\Sigma \in \mathbb{R}^{J\times J}$ and any vector $x \in \mathbb{R}^J$, $\Phi_\Sigma(x)$ denotes $P(Z \leq x)$ where $Z \sim N(0,\Sigma)$ and, if $\Sigma \in \mathbb{R}^{J\times J}$ is non-singular, $\phi_\Sigma(x)$ denotes the density of Z, where $Z \sim N(0,\Sigma)$. Finally, if $J = 1$ and $\Sigma = 1$, the reference of the variance covariance matrix may be dropped, so $\Phi = \Phi_1$ and $\phi = \phi_1$.

A.2 Preliminary results

A.2.1 On the assumptions

Lemma A.1 *Assumptions (B1)-(B4) imply assumption (A4).*

Proof. We define $S_Z = S_X \times \mathbb{R}^J$, $Z = (X,Y) : \Omega \to S_Z$, and $m(z,\theta) = \{\{(y_j - M_j(\theta, x_k))\,1[x = x_k]\}_{j=1}^J\}_{k=1}^K : S_Z \times \Theta \to \mathbb{R}^{JK}$. For every $(j,k) \in \{1,2,...,J\} \times \{1,2,...,K\}$, define $m_{j,k}(z,\theta) = (y_j - M_j(\theta, x_k))\,1[x = x_k]$.

For every $(j,k,\theta) \in \{1,2,...,J\} \times \{1,2,...,K\} \times \Theta$, $V(m_{j,k}(Z,\theta)) = V(Y_j|X_k)$, which is finite and positive. Also, $\forall z \in S_Z$, $\{m(z,\theta) - \mathbb{E}(m(Z,\theta))\} : \Theta \to \mathbb{R}^{JK}$ is a continuous function, and the continuous functions defined on a compact space constitute a separable subset of the space of bounded functions. To conclude this proof, we need to show the stochastic equicontinuity property of the empirical process associated to $m(Z,\theta)$. For every $\theta, \theta' \in \Theta$,

$$v_n(m_\theta) - v_n(m_{\theta'}) = \left\{\left\{\sqrt{n}\left(M_j\left(\theta', x_k\right) - M_j\left(\theta, x_k\right)\right)\left(\hat{p}_k - p_k\right)\right\}_{j=1}^J\right\}_{k=1}^K$$

where, $\forall k = 1,2,...,K$, $\hat{p}_k = n^{-1}\sum_{i=1}^n 1[X = x_k]$ and $p_k = P(X = x_k)$.

If the design is fixed, then, $\forall k = 1,...,K$, $\hat{p}_k = p_k$, and so stochastic equicontinuity is trivially satisfied. Thus, we focus the rest of the argument on the random design case. Fix $\varepsilon > 0$ arbitrarily. Let $\delta > 0$

be such that $\sum_{k=1}^{K} 2\Phi(-\varepsilon/(\delta\sqrt{JKp_k(1-p_k)})) < \varepsilon$. For every $k = 1,2,...,K$, $M(\theta,x_k) : \Theta \to \mathbb{R}^J$ is continuous and so, $\exists\eta > 0$ such that,

$$\max_{k=1,...,K}\max_{j=1,...,J}\sup_{\theta\in\Theta}\sup_{\{\theta'\in\Theta:\|\theta'-\theta\|\leq\eta\}}\left\|M_j(\theta,x_k) - M_j(\theta',x_k)\right\| < \delta$$

Therefore, $\sup_{\theta\in\Theta}\sup_{\{\theta':\|\theta'-\theta\|\leq\eta\}}\|v_n(m_\theta) - v_n(m_{\theta'})\| \leq \delta\sqrt{J}||\{\sqrt{n}(\hat{p}_k - p_k)\}_{k=1}^{K}||$, which implies,

$$P^*\left(\sup_{\theta\in\Theta}\sup_{\{\theta'\in\Theta:\|\theta'-\theta\|\leq\eta\}}\|v_n(m_\theta) - v_n(m_{\theta'})\| > \varepsilon\right) \leq \sum\nolimits_{k=1}^{K} P\left(\left|\sqrt{n}(\hat{p}_k - p_k)\right| > \varepsilon/\delta\sqrt{JK}\right)$$

By the CLT,

$$\limsup_{n\to+\infty} P^*\left(\sup_{\theta\in\Theta}\sup_{\{\theta'\in\Theta:\|\theta'-\theta\|\leq\eta\}}\|v_n(m_\theta) - v_n(m_{\theta'})\| > \varepsilon\right) \leq \sum\nolimits_{k=1}^{K} 2\Phi\left(-\varepsilon/\left(\delta\sqrt{JKp_k(1-p_k)}\right)\right)$$

and by the definition of δ, the right hand side is less that ε, completing the proof. ■

A.2.2 Verification of the assumptions in the example

In this section, we complete the specification of the example provided in section 2.1.3 so that the assumptions of the general model are satisfied. Moreover, depending on how we do this, we can also satisfy the assumptions of the conditionally separable model or not.

Suppose that our economic phenomenon corresponds to a binary choice model, where the dependent variable, Y, takes only two values. Without loss of generality, we assume that these values are zero and one and so, $\forall k = 1,2,...,K$, $Y_H(w_k) = 1$ and $Y_L(w_k) = 0$. Moreover, we restrict the values of the exogenous covariate to those for whom the choice is not deterministic, so, $\forall k = 1,2,...,K$, $P(Y = 1|W = w_k) \in (0,1)$. Also, we assume that we have missing data but not all data are missing, so that, $\forall k = 1,2,...,K$, $P(U = 1|W = w_k) \in (0,1)$.

Our econometric model predicts that $\mathbb{E}(Y - f(X,\theta)|W = w) = 0$, where $\theta \in \Theta$. We take Θ to be a convex and a compact subset of $\mathbb{R}^\eta$ for some $\eta < +\infty$. We assume the following properties about the function f: (a) $\forall(\theta,x) \in \Theta\times S_X$, $f(x,\theta) \in [0,1]$, (b) $\exists k \in \{1,2,...,K\}$ such that $\inf_{\theta\in\mathbb{R}^\eta}\mathbb{E}(f(X,\theta)|W = w_k) = 0$ and $\sup_{\theta\in\mathbb{R}^\eta}\mathbb{E}(f(X,\theta)|W = w_k) = 1$, and (c) $\forall(\theta,\theta',x) \in \Theta\times\Theta\times S_X$, $|f(x,\theta) - f(x,\theta')| < B(x)||\theta - \theta'||$ for some function $B(x) : S_X \to \mathbb{R}$ such that $\mathbb{E}(|B(X)|) < +\infty$. All these requirements on f are satisfied in the probit model, where $S_X \subseteq \mathbb{R}^\eta$ and $f(x,\theta) = \Phi(x'\theta)$.

Under these additional conditions, the identified set is given by,

$$\Theta_I = \left\{\theta\in\Theta : \left\{\begin{array}{c}\mathbb{E}((Y(1-U) - f(X,\theta))1[W = w_k]) \leq 0\\ \mathbb{E}(-(Y(1-U) + U - f(X,\theta))1[W = w_k]) \leq 0\end{array}\right\}_{k=1}^{K}\right\}$$

As required by our assumptions, we observe an i.i.d. sample of $\{(Y_i,U_i,X_i,W_i)\}_{i=1}^{n}$.

Notice that the relationship between the explanatory variable X and the exogenous variable W has been left unspecified. We entertain two cases. In the first case, we verify all the assumptions of the general model and we point out that some the assumptions of the conditionally separable model may not be satisfied. In the second case, we verify all the assumptions of the conditionally separable model.

Case 1: Endogenous explanatory variable Suppose that X is an endogenous explanatory random vector. Since W represents the exogenous covariates, this means that $\exists k = 1,2,...,K$, such that $\{X|W = w_k\}$ is a non-deterministic random vector.

We now verify the assumptions of the general model. In the probability space $(\Omega, \mathcal{B}, \mathbf{P})$, we define the random vector $Z = (Y, U, X, W) : \Omega \to S_Z$ where $S_Z = \{\{0,1\} \times \{0,1\} \times S_X \times \{w_k\}_{k=1}^K\}$. We define the function $m(z, \theta) : S_Z \times \Theta \to \mathbb{R}^{2K}$ as follows,

$$m(z,\theta) = m((y,u,x,w),\theta) = \{(y(1-u) - f(x,\theta)) 1[w = w_k], (y(1-u) + u - f(x,\theta)) 1[w = w_k]\}_{k=1}^K$$

Assumptions (A1) and (A2) are explicitly assumed. By definition, $\Theta_I = \{\theta \in \Theta : \{\mathbb{E}(m_j(Z,\theta)) \leq 0\}_{j=1}^{2K}\}$, the function m is measurable and $\mathbb{E}(m(Z,\theta)) : \Theta \to \mathbb{R}^{2K}$ is continuous. Fix $k = \bar{k}$ such that $\inf_{\theta \in \mathbb{R}^\eta} \mathbb{E}(f(X,\theta)|W = w_{\bar{k}}) = 0$ and $\sup_{\theta \in \mathbb{R}^\eta} \mathbb{E}(f(X,\theta)|W = w_{\bar{k}}) = 1$. Since $P(U = 1|W = w_{\bar{k}}) \in (0,1)$ and $P(Y = 1|W = w_{\bar{k}}) \in (0,1)$, either $\mathbb{E}(Y(1-U)|W = w_{\bar{k}}) > 0$ or $\mathbb{E}(Y(1-U) + U|W = w_{\bar{k}}) < 1$. Suppose that $\mathbb{E}(Y(1-U)|W = w_{\bar{k}}) > 0$. Since $\inf_{\theta \in \mathbb{R}^\eta} \mathbb{E}(f(X,\theta)|W = w_{\bar{k}}) = 0$, we can always define Θ large enough so that $\inf_{\theta \in \Theta} \mathbb{E}(f(X,\theta)|W = w_{\bar{k}}) < \mathbb{E}(Y(1-U)|W = w_{\bar{k}})$. Now, suppose that $\mathbb{E}(Y(1-U) + U|W = w_{\bar{k}}) < 1$. Since $\sup_{\theta \in \mathbb{R}^\eta} \mathbb{E}(f(X,\theta)|W = w_{\bar{k}}) = 1$, we can always define Θ large enough so that $\sup_{\theta \in \Theta} \mathbb{E}(f(X,\theta)|W = w_{\bar{k}}) > \mathbb{E}(Y(1-U) + U|W = w_{\bar{k}})$. In either case, Θ_I is a proper subset of Θ. This verifies assumption (A3).

Now we verify assumption (A4). For any $k = 1, 2, ..., K$, $\sup_{\theta \in \Theta} V((Y(1-U) - f(X,\theta)) 1[W = w_k]) = 0$ if and only if $\exists \theta' \in \Theta$ such that,

$$\left\{ \begin{array}{c} P\left(1 = f\left(X, \theta'\right)|W = w_k, Y(1-U) = 1\right) P(Y(1-U) = 1|W = w_k) + \\ P\left(0 = f\left(X, \theta'\right)|W = w_k, Y(1-U) = 0\right)(1 - P(Y(1-U) = 1|W = w_k)) \end{array} \right\} = 1$$

To show that this condition is impossible, it suffices to show that $P(Y(1-U) = 1|W = w_k) \in (0,1)$, which is a consequence of $P(Y = 1|W = w_k) \in (0,1)$ and $P(U = 1|W = w_k) \in (0,1)$. By repeating this argument with $Y(1-U) + U$ instead of $Y(1-U)$, we verify that, $\forall (\theta, j) \in \Theta \times \{1, ..., 2K\}$, $V(m_j(Z,\theta)) > 0$. For every $(\theta, j) \in \Theta \times \{1, ..., 2K\}$, $|m_j(Z,\theta)| \leq 1$ and so, $\forall (\theta, j) \in \Theta \times \{1, ..., 2K\}$, $V(m_j(Z,\theta))$ is bounded. Finally, $\forall \left(\theta, \theta', z\right) \in \Theta \times \Theta \times S_Z$,

$$\begin{aligned} &\left\|(m(z,\theta) - \mathbb{E}(m(Z,\theta))) - \left(m\left(z,\theta'\right) - \mathbb{E}\left(m\left(Z,\theta'\right)\right)\right)\right\| \\ &\leq 2J\left(\left|f(x,\theta) - f\left(x,\theta'\right)\right| + \left|\mathbb{E}\left(f(X,\theta) - f\left(X,\theta'\right)\right)\right|\right) \\ &\leq 2J(B(x) + \mathbb{E}(|B(X)|)) \left\|\theta - \theta'\right\| \end{aligned}$$

By taking $||\theta - \theta'||$ sufficiently small, we can make the left hand side arbitrarily small. Therefore, $\forall z \in S_Z$, $\left\{m(z,\theta) - \mathbb{E}(m(Z,\theta)) : \Theta \to \mathbb{R}^J\right\}$ belongs to $C^J(\Theta)$, which is a separable subset of $l_J^\infty(\Theta)$. Finally, to show stochastic equicontinuity of the empirical process associated to $m(Z,\theta)$, it suffices to show the stochastic equicontinuity of the empirical process associated to $\{f(X,\theta) 1[W = w_k]\}_{k=1}^K$. This can be verified by using arguments in section 2.7.4 in van der Vaart and Wellner [29].

As a final remark, note that it is possible that this example violates some of the assumptions of the conditionally separable model. Since $\exists k = 1, 2, ..., K$ such that $\{X|W = w_k\}$ is non-deterministic, it is possible that $\exists \theta_0 \in \Theta$ such that $\{f(X,\theta_0)|W = w_k\}$ is non-deterministic. In particular, this would be the case in the probit model. If so, the conditional separability required by assumption (B3) would be violated.

Case 2: Exogenous covariates In this case, X is equal to W. Assumption (B1) is implied by random sampling and by $S_X = \{w_k\}_{k=1}^K$. Assumption (B2) has already been verified in the previous case. To verify assumption (B3), in the probability space $(\Omega, \mathcal{B}, \mathbf{P})$, define $Z = \{Y(1-U), Y(1-U) + U\} : \Omega \to S_Z$ where $S_Z = \{0,1\} \times \{0,1\}$ and $M(\theta, x) = \{f(x,\theta), f(x,\theta)\} : \Theta \times S_X \to \mathbb{R}^2$, which implies that,

$$\Theta_I = \left\{\theta \in \Theta : \left\{\{\mathbb{E}(Z_j - M_j(\theta, X)|X = w_k) \leq 0\}_{j=1}^2\right\}_{k=1}^K\right\}$$

Finally, notice that $\forall k=1,2,...,K,\ M(\theta,w_k):\Theta\to\mathbb{R}^2$ is continuous. Conditions under which Θ_I is a proper subset of Θ have been provided in the previous case. This verifies assumption (B3).

By the arguments used in the previous case, if $\forall k=1,2,...,K,\ P(Y=1|W=w_k)\in(0,1)$ and $P(U=1|W=w_k)\in(0,1)$, then, it follows that, $\forall(k,j)=\{1,...,K\}\times\{1,2\}$, $V(Z_j|X=w_k)$ is positive. Since $\|Z\|\leq 1$, it follows that, $\forall(k,j)=\{1,...,K\}\times\{1,2\}$, $\{Z_j|X=w_k\}$ has finite absolute moments of all orders, which verifies assumption (B4).

A.2.3 On the choice of the criterion function

The following lemma characterizes all possible criterion functions.

Lemma A.2 *Under assumption (A3), the function $Q:\Theta\to\mathbb{R}$ is a criterion function if and only if it is given by $Q(\theta)=G_{\mathbf{P}}(\{[\mathbb{E}(m_j(Z,\theta))]_+\}_{i=1}^J)$, where $G_{\mathbf{P}}:\mathbb{R}_+^J\to\mathbb{R}$ is a non-negative function such that $G_{\mathbf{P}}(y)=0$ if and only if $y=0_{J\times 1}$.*

Proof. This proof is elementary and is therefore omitted. ■

Lemma A.2 reveals that there is a wide range of criterion functions. The notation $G_{\mathbf{P}}$ reveals that, in principle, the criterion function could depend on the probability distribution $\mathbf{P}$. In particular, one way in which the probability distribution could enter the specification of the criterion function is through Studentization, that is, by dividing each expectation by its standard deviation. For example, Studentization has been considered by CHT [10] and Andrews and Soares [3]. With Studentization, the criterion function is given by,

$$Q(\theta)=G\left(\left\{\left[\frac{\mathbb{E}(m_j(Z,\theta))}{\sqrt{V(m_j(Z,\theta))}}\right]_+\right\}_{j=1}^J\right)$$

where G is any of the functions admitted by assumption (CF) (or even assumption (CF'), which will be defined below). One benefit of Studentization is that the criterion function is not affected by changes in the scale of any of the moment inequalities. Studentization can be applied to all of the procedures proposed in the paper (bootstrap, asymptotic approximation and subsampling). Similar arguments to the ones used in this paper show that Studentized procedures produce consistent inference in level and have the same rates of convergence of the error in the coverage probability as their non-Studentized counterparts[26]. Unfortunately, in general, Studentization does not generate asymptotically pivotal statistics. We will show this in a more general context in the next paragraph.

A statistic is asymptotically pivotal if its limiting distribution does not depend on unknown parameters. As explained by Hall [14] and Horowitz [16], under certain conditions, the bootstrap approximation of an asymptotically pivotal statistics is more accurate than its asymptotic approximation. This feature is usually referred to as the asymptotic refinement of the bootstrap. We now show that, in general, the criterion functions defined by lemma A.2 cannot be asymptotically pivotal. To see this, consider the following partially identified model, $\Theta_I=\{\theta\in\Theta:\{\mathbb{E}(Y_1)\leq\theta\cap\mathbb{E}(Y_2)\leq\theta\}\}$, where, in particular, $\{Y_{1,i},Y_{2,i}\}_{i=1}^n$ is i.i.d. with $\{Y_1,Y_2\}\sim N((0,0),\Sigma(\rho))$, $\Sigma(\rho)=(1,\rho;\rho,1)$ and $\rho\in(-1,1)$. Given that the diagonal elements of $\Sigma(\rho)$ are equal to one, the non-Studentized and Studentized statistics coincide. Since $G_{\mathbf{P}}$ satisfies the conditions of lemma A.2, we deduce the following,

$$\begin{aligned}&P\left(G_{\mathbf{P}}\left(\left[\sqrt{n}(\mathbb{E}_n(Y_1)-\mathbb{E}(Y_1))\right]_+,\left[\sqrt{n}(\mathbb{E}_n(Y_2)-\mathbb{E}(Y_2))\right]_+\right)\leq 0\right)\\&=P\left(\sqrt{n}(\mathbb{E}_n(Y_1)-\mathbb{E}(Y_1))\leq 0\cap\sqrt{n}(\mathbb{E}_n(Y_2)-\mathbb{E}(Y_2))\leq 0\right)\\&=\int_{-\infty}^0\int_{-\infty}^0\phi_{\Sigma(\rho)}(x_1,x_2)\,dx_1dx_2\end{aligned}$$

[26] We omit these results for the sake of brevity. They are available from the author, upon request.

The last expression is a strictly decreasing function of $|\rho|$, which is an unknown parameter. Thus, in general, it is not possible to define a valid criterion function that produces an asymptotically pivotal statistic of interest.

For the sake of exposition, the main text assumes that the criterion function satisfies assumption (CF). Most of the results of the paper extend to a much larger class of criterion functions, characterized by assumption (CF').

Assumption (CF') The population criterion function is given by $Q(\theta) = G(\{[\mathbb{E}(m_j(Z,\theta))]_+\}_{j=1}^J)$, where $G : \mathbb{R}_+^J \to \mathbb{R}$ is a non-negative function that does not depend on $\mathbf{P}$, is strictly increasing in every coordinate, weakly convex, continuous, homogeneous of degree β and satisfies $G(y) = 0$ if and only if $y = 0$.

Consistency of any of the proposed inferential schemes considered in this paper (bootstrap, asymptotic approximation and subsampling) only requires assumption (CF'). We introduce the assumption (CF) when we are interested in obtaining rates of convergence of the error in the coverage probability. In particular, assumption (CF) is required to show that the error in the coverage probability of our bootstrap procedure converges to zero at a rate of $n^{-1/2}$. In this appendix, we also show that under assumption (CF'), the error in the coverage probability of our bootstrap procedure converges to zero at a rate of $n^{-1/2}\ln n^{1/2}$. Since the criterion function is a choice of the econometrician and since assumption (CF) allows us to prove a better rate of convergence for our bootstrap approximation, we decided to restrict the discussion of the main text to this assumption. Nevertheless, wherever reasonable, this appendix will distinguish between results obtained under these two assumptions.

A.2.4 On the estimation of the identified set

Proof. [**Lemma 2.1**]

Part 1. The definition of Θ_I implies the following sequence of inequalities,

$$
\begin{aligned}
&\sup_{\theta\in\Theta_I}\max_{j=1,\ldots,J}\mathbb{E}_n(m_j(Z,\theta)) \\
&\leq \sup_{\theta\in\Theta_I}\max_{j=1,\ldots,J}(\mathbb{E}_n(m_j(Z,\theta)) - \mathbb{E}(m_j(Z,\theta))) + \sup_{\theta\in\Theta_I}\max_{j=1,\ldots,J}\mathbb{E}(m(Z,\theta)) \\
&\leq n^{-1/2}\sup_{\theta\in\Theta}\max_{j=1,\ldots,J}v_n(m_{j,\theta})
\end{aligned}
$$

Therefore, $\{\sup_{\theta\in\Theta}\max_{j=1,\ldots,J}v_n(m_{j,\theta}) \leq \tau_n\}$ implies $\{\Theta_I \subseteq \hat{\Theta}_I(\tau_n)\}$ and so,

$$
P\left(\liminf\left\{\Theta_I \subseteq \hat{\Theta}_I(\tau_n)\right\}\right) \geq \sum\nolimits_{j=1}^{J} P\left(\liminf\left\{\sup_{\theta\in\Theta}|v_n(m_{j,\theta})| \leq \tau_n\right\}\right) - J + 1
$$

Under the separability assumption and the fact that $\sqrt{\ln\ln n}/\tau_n = o(1)$ a.s., the LIL for empirical processes (see, for example, Kuelbs [19]) implies that the expression on the right hand side is equal to one.

The definition of $\hat{\Theta}_I(\tau_n)$ implies the following sequence of inequalities,

$$
\begin{aligned}
&\sup_{\theta\in\hat{\Theta}_I(\tau_n)}\max_{j=1,\ldots,J}\mathbb{E}(m_j(Z,\theta)) \\
&\leq \sup_{\theta\in\hat{\Theta}_I(\tau_n)}\max_{j=1,\ldots,J}(\mathbb{E}(m_j(Z,\theta)) - \mathbb{E}_n(m_j(Z,\theta))) + \sup_{\theta\in\hat{\Theta}_I(\tau_n)}\max_{j=1,\ldots,J}\mathbb{E}_n(m_j(Z,\theta)) \\
&\leq n^{-1/2}\left(\tau_n - \inf_{\theta\in\Theta}\min_{j=1,\ldots,J}v_n(m_{j,\theta})\right)
\end{aligned}
$$

Therefore, $\{\inf_{\theta\in\Theta}\min_{j=1,...,J} v_n(m_{j,\theta}) \geq -\tau_n\}$ and $(\tau_n/\sqrt{n})/\varepsilon_n = o(1)$ implies $\{\hat{\Theta}_I(\tau_n) \subseteq \Theta_I(\varepsilon_n)\}$ and so,

$$P\left(\liminf\left\{\hat{\Theta}_I(\tau_n) \subseteq \Theta_I(\varepsilon_n)\right\}\right) \geq \sum\nolimits_{j=1}^{J} P\left(\liminf\left\{\sup_{\theta\in\Theta}|v_n(m_{j,\theta})| \leq \tau_n\right\}\right) - J + 1$$

and for the same reasons as before, the expression on the right hand side is equal to one. Elementary properties of lim inf operator complete the proof.

Part 2. Since $\mathbb{E}(m(Z,\theta)) : \Theta \to \mathbb{R}^J$ is lower-semi continuous and Θ is compact, $\max_{j=1,...,J}\mathbb{E}(m_j(Z,\theta))$ achieves a minimum on Θ. Since $\Theta_I = \varnothing$, such minimum is a positive value which we denote by $\varpi > 0$, and so,

$$\begin{aligned}
&\inf_{\theta\in\Theta}\max_{j=1,...,J}\mathbb{E}_n(m_j(Z,\theta)) \\
&\geq \inf_{\theta\in\Theta}\min_{j=1,...,J}(\mathbb{E}_n(m_j(Z,\theta)) - \mathbb{E}(m_j(Z,\theta))) + \inf_{\theta\in\Theta}\max_{j=1,...,J}\mathbb{E}(m_j(Z,\theta)) \\
&\geq n^{-1/2}\inf_{\theta\in\Theta}\min_{j=1,...,J} v_n(m_{j,\theta}) + \varpi
\end{aligned}$$

Therefore, $\{\inf_{\theta\in\Theta}\min_{j=1,...,J} v_n(m_{j,\theta}) \geq -\tau_n\}$ implies $\{\hat{\Theta}_I(\tau_n) = \varnothing\}$ and hence,

$$P\left(\liminf\left\{\hat{\Theta}_I(\tau_n) = \varnothing\right\}\right) \geq \sum\nolimits_{j=1}^{J} P\left(\liminf\left\{\sup_{\theta\in\Theta}|v_n(m_{j,\theta})| \leq \tau_n\right\}\right) - J + 1$$

and for the same reasons as before, the expression on the right hand side is equal to one. ■

A.2.5 Differences with the naive bootstrap

The bootstrap procedure we propose to construct confidence sets differs qualitatively from replacing the subsampling scheme in CHT [10] with the traditional bootstrap.

In order to approximate the quantile of the distribution of interest, the subsampling approximation proposed by CHT [10] would propose the following statistic[27],

$$\Gamma^{SS,CHT}_{b_n,n} = \begin{cases} \sup_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\left\{\left[\sqrt{b_n}\mathbb{E}^{SS}_{b_n,n}(m_j(Z,\theta))\right]_+\right\}_{j=1}^{J}\right) & \text{if } \hat{\Theta}_I(\tau_n) \neq \varnothing \\ 0 & \text{if } \hat{\Theta}_I(\tau_n) = \varnothing \end{cases}$$

where $\{Z^{SS}_i\}_{i=1}^{b_n}$ is a random sample of size b_n extracted without replacement from the data and, $\forall j = 1,2,...,J$, $\mathbb{E}^{SS}_{b_n,n}(m_j(Z,\theta)) = b_n^{-1}\sum_{i=1}^{b_n} m_j(Z^{SS}_i,\theta)$. If we were to (naively) replace their subsampling scheme with a bootstrap scheme, we would propose the following statistic,

$$\Gamma^{Naive}_n = \begin{cases} \sup_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\{[\sqrt{n}\mathbb{E}^*_n(m_j(Z,\theta))]_+\}_{j=1}^{J}\right) & \text{if } \hat{\Theta}_I(\tau_n) \neq \varnothing \\ 0 & \text{if } \hat{\Theta}_I(\tau_n) = \varnothing \end{cases}$$

where $\{Z^*_i\}_{i=1}^{n}$ is a random sample of size n extracted with replacement from the data and, $\forall j = 1,2,...,J$, $\mathbb{E}^*_n(m_j(Z,\theta)) = n^{-1}\sum_{i=1}^{n} m_j(Z^*_i,\theta)$. Since the statistic Γ^{Naive}_n is the consequence of naively replacing one resampling procedure with another one, we will refer to the resulting bootstrap procedure as the *naive bootstrap.* Even though the subsampling approximation proposed by CHT [10] produces consistent inference in level, the naive bootstrap will, in general, result in inconsistent inference in level.

[27] To be precise, the subsampling approximation proposed by CHT [10] would use a different estimator for the identified set. In any case, this difference is asymptotically negligible.

There are two reasons why the naive bootstrap is inconsistent in level. Recall from section 2.2.2 that we estimate the identified set by artificially expanding the sample analogue estimator by a certain amount. The effect of this expansion will be asymptotically negligible for the subsampling procedure in CHT [10] but will generate inconsistencies for the naive bootstrap. We will refer to this problem as *the expansion problem*. The second problem is directly related to the well-known *inconsistency of the bootstrap on the boundary of the parameter space*, studied by Andrews [1].

In order to understand the nature of these problems, we provide two examples. In each of these examples, we show three things. First, we show that the naive bootstrap is inconsistent in level. Second, we show that the bootstrap procedure proposed in this paper corrects these inconsistencies. Third, we show that these inconsistencies are not present in the subsampling scheme proposed by CHT [10].

Problem 1: the expansion problem The objective is to construct a confidence set for the following identified set,

$$\Theta_I = \{\theta \in \Theta : \{\mathbb{E}(Y_1) \leq \theta \leq \mathbb{E}(Y_2)\}\}$$

where $\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = 0$. Suppose that the sample $\{Y_{1,i}, Y_{2,i}\}_{i=1}^n$ is i.i.d. such that, $\forall i = 1, 2, ..., n$, $(Y_{1,i}, Y_{2,i}) \sim N(0, \mathbf{I}_2)$. Notice that all assumptions of the conditionally separable model are satisfied. The distribution of interest is given by $\Gamma_n = \sqrt{n} Q_n(0)$, which is equal to $G([\zeta_1]_+, [\zeta_2]_+)$, where $\zeta \sim N(0, \mathbf{I}_2)$.

Now consider estimation of the identified set. The key feature of this setup is that even though the identified set is non-empty (because $\mathbb{E}(Y_1) \leq \mathbb{E}(Y_2)$), the sample analogue estimator of the identified set, given by $\hat{\Theta}_I^{AP} = \{\theta \in \Theta : \mathbb{E}_n(Y_1) \leq \theta \leq \mathbb{E}_n(Y_2)\}$, is empty with positive probability (in this case, with probability 0.5). Hence, using the estimator $\hat{\Theta}_I^{AP}$ as the domain of the maximization problem in step 3 will not result in consistent inference in level. This illustrates why we need to introduce the sequence $\{\tau_n\}_{n=1}^{+\infty}$ to estimate the identified set. Our estimator for the identified set is given by,

$$\hat{\Theta}_I(\tau_n) = \left\{\theta \in \Theta : \left\{\mathbb{E}_n(Y_1) - \tau_n/\sqrt{n} \leq \theta \leq \mathbb{E}_n(Y_2) + \tau_n/\sqrt{n}\right\}\right\}$$

Consider performing inference with the naive bootstrap. In this setting, it follows that,

$$\Gamma_n^{Naive} = 1\left[\hat{\Theta}_I(\tau_n) \neq \varnothing\right] \max \left\{ \begin{array}{c} G\left([\sqrt{n}(\mathbb{E}_n^*(Y_1) - \mathbb{E}_n(Y_1)) + \tau_n]_+, [\sqrt{n}(\mathbb{E}_n(Y_1) - \mathbb{E}_n^*(Y_2)) - \tau_n]_+\right), \\ G\left([\sqrt{n}(\mathbb{E}_n^*(Y_1) - \mathbb{E}_n(Y_2)) - \tau_n]_+, [\sqrt{n}(\mathbb{E}_n(Y_2) - \mathbb{E}_n^*(Y_2)) + \tau_n]_+\right) \end{array} \right\}$$

For any $\varepsilon > 0$, consider the following events,

$$\begin{aligned} A &= \left\{\left\{\sqrt{n}(\mathbb{E}_n^*(Y_1) - \mathbb{E}_n(Y_1), \mathbb{E}_n^*(Y_2) - \mathbb{E}_n(Y_2)) | \mathcal{X}_n\right\} \xrightarrow{d} N(0, \mathbf{I}_2)\right\} \\ B &= \liminf \left\{\left\{\hat{\Theta}_I(\tau_n) = \varnothing\right\} \cap \left\{\left|\sqrt{n}(\mathbb{E}_n(Y_1) - \mathbb{E}_n(Y_2))\right| \leq \tau_n/2\right\}\right\} \end{aligned}$$

Let $\omega \in \{A \cap B\}$. Since $\omega \in B$, $\exists N \in \mathbb{N}$ such $\forall n \geq N$,

$$\Gamma_n^{Naive} \geq \max \left\{ \begin{array}{c} G\left([\sqrt{n}(\mathbb{E}_n^*(Y_1) - \mathbb{E}_n(Y_1)) + \tau_n]_+, [\sqrt{n}(\mathbb{E}_n(Y_2) - \mathbb{E}_n^*(Y_2)) - 1.5\tau_n]_+\right), \\ G\left([\sqrt{n}(\mathbb{E}_n^*(Y_1) - \mathbb{E}_n(Y_1)) - 1.5\tau_n]_+, [\sqrt{n}(\mathbb{E}_n(Y_2) - \mathbb{E}_n^*(Y_2)) + \tau_n]_+\right) \end{array} \right\}$$

Since $\omega \in A$, the conditional distribution of the right hand side diverges to infinity, a.s.. By the LIL and the requirements on $\{\tau_n\}_{n=1}^{+\infty}$, $P(A) = 1$. By theorem 2.1 in Bickel and Freedman [7], $P(B) = 1$. Hence, the naive bootstrap produces inference that is not consistent in level.

The intuition for this result is as follows. The estimation of the identified set requires the introduction of the sequence $\{\tau_n\}_{n=1}^{+\infty}$, which enters directly into the $[\cdot]_+$ term of the criterion function of the naive bootstrap. Since this sequence diverges to infinity, the distribution of the naive bootstrap approximation also diverges to infinity. As we show next, our bootstrap procedure corrects this problem by removing the sequence from the $[\cdot]_+$ term.

If we choose to perform inference with our proposed bootstrap method, we have the following statistic,

$$\Gamma_n^* = 1\left[\hat{\Theta}_I(\tau_n) \neq \varnothing\right] \max_{\theta \in \hat{\Theta}_I} \left\{ G\left(\begin{array}{c} [\sqrt{n}(\mathbb{E}_n^*(Y_1) - \mathbb{E}_n(Y_1))]_+ 1[|\mathbb{E}_n(Y_1) - \theta| \leq \tau_n/\sqrt{n}], \\ [\sqrt{n}(\mathbb{E}_n(Y_2) - \mathbb{E}_n^*(Y_2))]_+ 1[|\theta - \mathbb{E}_n(Y_2)| \leq \tau_n/\sqrt{n}] \end{array} \right), \right\}$$

Consider $\omega \in \{A \cap B'\}$ where B' is defined by,

$$B' = \liminf \left\{ \left\{ \{0\} \in \hat{\Theta}_I(\tau_n) \right\} \cap \left\{ \left|\sqrt{n}\mathbb{E}_n(Y_1)\right| \leq \tau_n \right\} \cap \left\{ \left|\sqrt{n}\mathbb{E}_n(Y_2)\right| \leq \tau_n \right\} \right\}$$

Since $\omega \in B'$, $\exists N \in \mathbb{N}$, such that, $\forall n \geq N$,

$$\Gamma_n^* = G\left(\left[\sqrt{n}(\mathbb{E}_n^*(Y_1) - \mathbb{E}_n(Y_1))\right]_+, \left[\sqrt{n}(\mathbb{E}_n(Y_2) - \mathbb{E}_n^*(Y_2))\right]_+ \right)$$

Since $\omega \in A$, the conditional distribution of the right hand side converges weakly to $G\left([\zeta_1]_+, [\zeta_2]_+\right)$, where $\zeta \sim N(0, \mathbf{I}_2)$, a.s.. By the same arguments as before, $P(A' \cap B) = 1$ and, thus, our bootstrap procedure leads to consistent inference in level.

It is important to understand that the inconsistency problem of the naive bootstrap is not present in the subsampling procedure proposed by CHT [10]. In this case,

$$\Gamma_{b_n,n}^{SS,CHT} = 1\left[\hat{\Theta}_I(\tau_n) \neq \varnothing\right] \max \left\{ \begin{array}{c} G\left(\begin{array}{c} \left[\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(Y_1) - \mathbb{E}_n(Y_1)\right) + \tau_n\sqrt{b_n/n}\right]_+, \\ \left[\sqrt{b_n}\left(\mathbb{E}_n(Y_1) - \mathbb{E}_{b_n,n}^{SS}(Y_2)\right) - \tau_n\sqrt{b_n/n}\right]_+ \end{array} \right), \\ G\left(\begin{array}{c} \left[\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(Y_1) - \mathbb{E}_n(Y_2)\right) - \tau_n\sqrt{b_n/n}\right]_+, \\ \left[\sqrt{b_n}\left(\mathbb{E}_n(Y_2) - \mathbb{E}_{b_n,n}^{SS}(Y_2)\right) + \tau_n\sqrt{b_n/n}\right]_+ \end{array} \right) \end{array} \right\}$$

For any $\varepsilon > 0$, let B'' be defined as,

$$B'' = \liminf \left\{ \left\{ \{0\} \in \hat{\Theta}_I(\tau_n) \right\} \cap \left\{ \left| \sqrt{b_n}(\mathbb{E}_n(Y_1) - \mathbb{E}_n(Y_2)) \right| \leq (1+\varepsilon)\, 2\sqrt{(b_n \ln\ln n)/n} \right\} \right\}$$

Consider $\omega \in \{A \cap B''\}$. If the sequence $\{\tau_n\}_{n=1}^{+\infty}$ is chosen such that $\tau_n\sqrt{b_n/n} = o(1)$ a.s., then, conditionally on the sample, $\Gamma_{b_n,n}^{SS,CHT}$ converges weakly to $G\left([\zeta_1]_+, [\zeta_2]_+\right)$, where $\zeta \sim N(0, \mathbf{I}_2)$. By previous arguments, $P(A'' \cap B) = 1$. Hence, the subsampling scheme proposed by CHT [10] results in consistent inference in level. Just like with the naive bootstrap, the estimation of the identified set introduces a sequence into the $[\cdot]_+$ term of the statistic. The key difference with the naive bootstrap is that this sequence converges to zero (instead of diverging to infinity), so it does not affect the asymptotic distribution.

Problem 2: boundary problem In order to isolate this problem from the previous one, we consider an example of an identified set which can be estimated without the need of introducing any expansion. The identified set is given by,

$$\Theta_I = \{\theta \in \Theta : \{\mathbb{E}(Y_1) \leq \theta\} \cap \{\mathbb{E}(Y_2) \leq \theta\}\}$$

where $\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = 0$. Suppose that the sample $\{Y_{1,i}, Y_{2,i}\}_{i=1}^n$ is i.i.d. such that, $\forall i = 1, 2, ..., n$, $(Y_{1,i}, Y_{2,i}) \sim N(0, \mathbf{I}_2)$. Notice that all assumptions of the conditionally separable model are satisfied. The distribution of interest is given by $\Gamma_n = \sup_{\theta \in \Theta_I} \sqrt{n} Q_n(\theta)$, which is equal to $G\left([\zeta_1]_+, [\zeta_2]_+\right)$, where $\zeta \sim N(0, \mathbf{I}_2)$.

As opposed to the previous example, the identified set has non-empty interior and the sample analogue estimate will always be non-empty. Hence, we can estimate the identified set with the analogy principle estimate, given by $\hat{\Theta}_I(0) = \{\theta \in \Theta : \{\mathbb{E}_n(Y_1) \leq \theta\} \cap \{\mathbb{E}_n(Y_2) \leq \theta\}\}$.

Now consider performing inference with the naive bootstrap. For any constant $c > 0$, consider the following events,

$$\begin{aligned} A &= \left\{ \left\{ \sqrt{n} \left(\mathbb{E}_n^* (Y_1) - \mathbb{E}_n (Y_1), \mathbb{E}_n^* (Y_2) - \mathbb{E}_n (Y_2) \right) | \mathcal{X}_n \right\} \xrightarrow{d} N(0, \mathbf{I}_2) \right\} \\ B &= \limsup \left\{ \left\{ \hat{\Theta}_I (0) \neq \varnothing \right\} \cap \left\{ \sqrt{n} \left(\mathbb{E}_n (Y_1) - \mathbb{E}_n (Y_2) \right) < -c \right\} \right\} \end{aligned}$$

Suppose that $\omega \in \{A \cap B\}$. Since $\omega \in B$, there exists a subsequence $\{n_k\}_{k=1}^{+\infty}$ such that, along this subsequence, $\hat{\Theta}_I (0)$ is non-empty and $\{\sqrt{n_k} (\mathbb{E}_{n_k} (Y_1) - \mathbb{E}_{n_k} (Y_2)) < -c\}$. Along this subsequence,

$$\Gamma_{n_k}^{Naive} \leq G \left(\left[\sqrt{n} \left(\mathbb{E}_{n_k}^* (Y_1) - \mathbb{E}_{n_k} (Y_1) \right) - c \right]_+, \left[\sqrt{n} \left(\mathbb{E}_{n_k}^* (Y_2) - \mathbb{E}_{n_k} (Y_2) \right) \right]_+ \right)$$

and since $\omega \in A$, then the right hand side converges weakly to $G \left([\zeta_1 - c]_+, [\zeta_2]_+ \right)$, where $\zeta \sim N(0, \mathbf{I}_2)$. By using previous arguments, $P(A \cap B) = 1$. This implies that the naive bootstrap produces inference that is not consistent in level.

One may relate this result with the inconsistency of the bootstrap on the boundary of the parameter space. The boundaries of the unknown identified set are determined by the parameters $\mathbb{E}(Y_1)$ and $\mathbb{E}(Y_2)$, which happen to coincide. The boundaries of the sample identified set are determined by $\mathbb{E}_n (Y_1)$ and $\mathbb{E}_n (Y_2)$, which do not coincide, a.s.. As a consequence, the structure of the boundaries of the identified set and the estimator of the identified set do not coincide, a.s., producing inconsistency of the resulting inference.

Now consider doing inference with our proposed bootstrap procedure. Assume that $\omega \in \{A \cap B'\}$ where B' is the following event,

$$B' = \liminf \left\{ \left\{ 0 \in \hat{\Theta}_I (0) \right\} \cap \left\{ \left| \sqrt{n} \mathbb{E}_n (Y_1) \right| \leq \tau_n \right\} \cap \left\{ \left| \sqrt{n} \mathbb{E}_n (Y_2) \right| \leq \tau_n \right\} \right\}$$

Since $\omega \in B'$, $\exists N \in \mathbb{N}$, such that, $\forall n \geq N$, $0 \in \hat{\Theta}_I (0)$, $|\sqrt{n} \mathbb{E}_n (Y_1)| \leq \tau_n$ and $|\sqrt{n} \mathbb{E}_n (Y_2)| \leq \tau_n$. Thus, $\forall n \geq N$,

$$\Gamma_n^* = G \left(\left[\sqrt{n} \left(\mathbb{E}_n^* (Y_1) - \mathbb{E}_n (Y_1) \right) \right]_+, \left[\sqrt{n} \left(\mathbb{E}_n^* (Y_2) - \mathbb{E}_n (Y_2) \right) \right]_+ \right)$$

Since $\omega \in A$, the conditional distribution of the right hand side converges weakly to $G \left([\zeta_1]_+, [\zeta_2]_+ \right)$, where $\zeta \sim N(0, \mathbf{I}_2)$. By using previous arguments $P(A \cap B') = 1$ and so, our bootstrap procedure is consistent in level.

Now consider using the subsampling scheme proposed by CHT [10]. In this case,

$$\Gamma_{b_n,n}^{SS,CHT} = 1 \left[\hat{\Theta}_I (0) \neq \varnothing \right] \left\{ \begin{array}{c} G \left(\begin{array}{c} \left[\sqrt{b_n} \left(\mathbb{E}_{b_n,n}^{SS} (Y_1) - \mathbb{E}_n (Y_1) \right) \right]_+, \\ \left[\sqrt{b_n} \left(\mathbb{E}_{b_n,n}^{SS} (Y_2) - \mathbb{E}_n (Y_1) \right) \right]_+ \end{array} \right) 1 \left[\mathbb{E}_n (Y_1) \geq \mathbb{E}_n (Y_2) \right] + \\ G \left(\begin{array}{c} \left[\sqrt{b_n} \left(\mathbb{E}_{b_n,n}^{SS} (Y_1) - \mathbb{E}_n (Y_2) \right) \right]_+, \\ \left[\sqrt{b_n} \left(\mathbb{E}_{b_n,n}^{SS} (Y_2) - \mathbb{E}_n (Y_2) \right) \right]_+ \end{array} \right) 1 \left[\mathbb{E}_n (Y_1) < \mathbb{E}_n (Y_2) \right] \end{array} \right\}$$

Let $\omega \in \{A \cap B''\}$, where B'' is given by,

$$B'' = \liminf \left\{ \left\{ 0 \in \hat{\Theta}_I (0) \right\} \cap \left\{ \left| \sqrt{b_n} \left(\mathbb{E}_n (Y_1) - \mathbb{E}_n (Y_2) \right) \right| \leq (1 + \varepsilon) 2 \sqrt{(b_n \ln \ln n) / n} \right\} \right\}$$

If $(b_n \ln \ln n)/n = o(1)$ and using previous arguments, $\Gamma_{n,b_n}^{SS,CHT}$ converges weakly to $G \left([\zeta_1]_+, [\zeta_2]_+ \right)$, where $\zeta \sim N(0, \mathbf{I}_2)$. Since $P(A \cap B'') = 1$, the subsampling procedure proposed by CHT [10] is consistent in level.

A.3 Representation results

A.3.1 Representation result for the population test statistic

The following theorem provides an alternative asymptotic representation for the statistic of interest.

Theorem A.1 *Part 1. Assume (A1)-(A4), (CF') and $\Theta_I \neq \varnothing$. Then, $\Gamma_n = H\left(v_n\left(m_\theta\right)\right) + \delta_n$, where,*

1. *for any $\varepsilon > 0$, $\lim_{n\to+\infty} P^*\left(\left|\delta_n\right| > \varepsilon\right) = 0$,*

2. *$v_n\left(m_\theta\right) : \Omega_n \to l_J^\infty\left(\Theta\right)$ is an empirical process that converges weakly to a tight zero-mean Gaussian process, denoted ζ, whose variance covariance function is denoted by Σ. For every $\theta_1, \theta_2 \in \Theta$, $\Sigma\left(\theta_1, \theta_2\right)$ is given by,*

$$\Sigma\left(\theta_1, \theta_2\right) = \mathbb{E}\left[\left(m\left(Z, \theta_1\right) - \mathbb{E}\left(m\left(Z, \theta_1\right)\right)\right)\left(m\left(Z, \theta_2\right) - \mathbb{E}\left(m\left(Z, \theta_2\right)\right)\right)'\right]$$

3. *$H : l_J^\infty\left(\Theta\right) \to \mathbb{R}$ is continuous, non-negative, weakly convex, homogenous of degree $\beta \geq 1$ and $H\left(y\right) = 0$ implies that $\exists\left(\theta_0, j\right) \in \Theta \times \{1, ..., J\}$, $y_j\left(\theta_0\right) \leq 0$.*

Part 2. Let ρ denote the rank of the variance covariance matrix of the vector $\{\{1\left[X = x_k\right] Y_j\}_{j=1}^J\}_{k=1}^K$. If we assume (B1)-(B4), (CF) and $\Theta_I \neq \varnothing$, then, $\Gamma_n = \tilde{H}\left(\sqrt{n}\left(\mathbb{E}_n\left(Z\right) - \mathbb{E}\left(Z\right)\right)\right) + \tilde{\delta}_n$, where,

1. *for any $\varepsilon_n = O(n^{-1/2})$, $P(|\tilde{\delta}_n| > \varepsilon_n) = o(n^{-1/2})$,*

2. *$\{\mathbb{E}_n\left(Z\right) - \mathbb{E}\left(Z\right)\} : \Omega_n \to \mathbb{R}^\rho$ is a zero mean sample average of n i.i.d. observations from a distribution with variance covariance matrix $\mathbf{I}_\rho$. Moreover, this distribution has finite third absolute moments,*

3. *$\tilde{H} : \mathbb{R}^\rho \to \mathbb{R}$ is continuous, non-negative, weakly convex and homogenous of degree one. For all $\mu > 0$, any $|h| \geq \mu > 0$ and any positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\varepsilon_n = o\left(1\right)$, $\tilde{H}^{-1}\left(\left(h - \varepsilon_n, h + \varepsilon_n\right]\right) \subseteq \{\tilde{H}^{-1}\left(\{h\}\right)\}^{\eta_n}$ where $\eta_n = O\left(\varepsilon_n\right)$. Finally, $\tilde{H}\left(y\right) = 0$ implies that for some non-zero vector $b \in \mathbb{R}^\rho$, $b'y \leq 0$.*

Part 3. Assume (A1)-(A4), (CF') and $\Theta_I = \varnothing$. Then, $\Gamma_n = 0$.

Proof. Part 1. Let δ_n be defined as,

$$\delta_n = \sup_{\theta\in\Theta_I} G\left(\left\{\left[\sqrt{n}\mathbb{E}_n\left(m_j\left(Z, \theta\right)\right)\right]_+\right\}_{j=1}^J\right) - \sup_{\theta\in\Theta_I} G\left(\left\{\left[v_n\left(m_{j,\theta}\right)\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z, \theta\right)\right) = 0\right]\right\}_{j=1}^J\right)$$

and set $H\left(y\right) = \sup_{\theta\in\Theta_I} G(\{\left[y_j\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z, \theta\right)\right) = 0\right]\}_{j=1}^J)$.

Point 1. Restrict attention to $\theta \in \Theta_I$ and fix $\varepsilon > 0$ arbitrarily. By definition, $\delta_n \geq 0$ and so, it suffices to show that $P^*\left(\delta_n > \varepsilon\right) = o\left(1\right)$. For any positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\sqrt{\ln\ln n}/\varepsilon_n = o\left(1\right)$ and $\varepsilon_n/\sqrt{n} = o\left(1\right)$, denote $A_n = \left\{\sup_{\theta\in\Theta_I}\left\|v_n\left(m_\theta\right)\right\| \leq \varepsilon_n\right\}$. By the LIL for empirical processes, $P\left(\{A_n\}^c\right) = o\left(1\right)$ and so, it suffices to show that $P^*\left(\delta_n > \varepsilon \cap A_n\right) = o\left(1\right)$.

Denote $G_{n,1}\left(\theta\right) = G(\{[\sqrt{n}\mathbb{E}_n\left(m_j\left(Z, \theta\right)\right)]_+\}_{j=1}^J)$, $G_{n,2}\left(\theta\right) = G(\{\left[v_n\left(m_{j,\theta}\right)\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z, \theta\right)\right) = 0\right]\}_{j=1}^J)$, $\bar{G}_{n,1} = \sup_{\theta\in\Theta_I} G_{n,1}\left(\theta\right)$ and $\bar{G}_{n,2} = \sup_{\theta\in\Theta_I} G_{n,2}\left(\theta\right)$.

By definition of supremum, $\forall\varepsilon > 0$, $\exists\theta \in \Theta_I$ so that $G_{n,1}\left(\theta\right) + \varepsilon/2 \geq \bar{G}_{n,1}$ and so, the event $\{\delta_n > \varepsilon \cap A_n\}$ is equivalent to,

$$\left\{\{\delta_n > \varepsilon\} \cap \left\{\exists\theta \in \Theta_I : \left\{G_{n,1}\left(\theta\right) + \varepsilon/2 \geq \bar{G}_{n,1}\right\} \cap \left\{G_{n,1}\left(\theta\right) - G_{n,2}\left(\theta\right) \geq \varepsilon/2\right\}\right\} \cap A_n\right\}$$

For any $S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$, consider the set $D_n(S)$ given by,

$$D_n(S) = \left\{\Theta_I \cap \left\{\bigcap_{j\in S} \{\mathbb{E}_n(m_j(Z,\theta)) \geq 0\}\right\} \cap \left\{\bigcap_{j\in\{\{1,2,..,J\}/S\}} \{\mathbb{E}_n(m_j(Z,\theta)) < 0\}\right\}\right\}$$

The event $\{\exists\theta \in \Theta_I : \{G_{n,1}(\theta) - G_{n,2}(\theta) \geq \varepsilon/2\}\}$ implies that $\exists j \in \{1,2,...,J\}$ such that $\mathbb{E}_n(m_j(Z,\theta)) \geq 0$ and $\mathbb{E}(m_j(Z,\theta)) < 0$, which, in turn, implies the event $\bigcup_{S\in\{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}} \{\exists\theta \in D_n(S)\}$.

For any $S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$, define the following two sets,

$$\begin{aligned}
\tilde{D}_n(S) &= \left\{\Theta_I \cap \left\{\bigcap_{j\in S} \{\mathbb{E}(m_j(Z,\theta)) \in [-\varepsilon_n/\sqrt{n}, 0]\}\right\}\right\} \\
D(S) &= \left\{\Theta_I \cap \left\{\bigcap_{j\in S} \{\mathbb{E}(m_j(Z,\theta)) = 0\}\right\}\right\}
\end{aligned}$$

For any $S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$, $\{\{\exists\theta \in D_n(S)\} \cap A_n\}$ implies $\{\{\exists\theta \in \tilde{D}_n(S)\} \cap A_n\}$. Also, $\forall S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$, $\lim_{n\to+\infty} \tilde{D}_n(S) = D(S)$, which implies that, $\forall\eta > 0$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\{\exists\theta \in \tilde{D}_n(S)\}$ implies $\{\exists\theta' \in D(S) : \|\theta - \theta'\| < \eta\}$. Then, $\forall\eta > 0$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\{\delta_n > \varepsilon \cap A_n\}$ is equivalent to the event,

$$\bigcup_{S\in\{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}} \left\{\{\delta_n > \varepsilon \cap A_n\} \cap \left\{\begin{array}{c} \exists(\theta,\theta') \in \{D_n(S) \times D(S)\} : \\ \|\theta - \theta'\| \leq \eta \cap \{G_{n,1}(\theta) + \varepsilon/2 \geq \bar{G}_{n,1}\} \end{array}\right\}\right\}$$

Now, $\forall\eta > 0$ and $\forall S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$, the following event,

$$\left\{\{\delta_n > \varepsilon\} \cap \left\{\exists(\theta,\theta') \in \{D_n(S) \times D(S)\} : \|\theta - \theta'\| \leq \eta \cap \{G_{n,1}(\theta) + \varepsilon/2 \geq \bar{G}_{n,1}\}\right\}\right\}$$

leads to the following derivation,

$$\begin{aligned}
& G\left([v_n(m_{j,\theta})]_+ 1[j \in S]\right) + \frac{\varepsilon}{2} \\
& \overset{(1)}{\geq} G\left(\left[\sqrt{n}\mathbb{E}_n(m_j(Z,\theta))\right]_+ 1[j \in S]\right) + \frac{\varepsilon}{2} \\
& \overset{(2)}{\geq} G\left(\left[\sqrt{n}\mathbb{E}_n(m_j(Z,\theta))\right]_+\right) + \frac{\varepsilon}{2} \\
& \overset{(3)}{\geq} \sup_{\tilde{\theta}\in\Theta_I} G\left(\left\{\left[\sqrt{n}\mathbb{E}_n\left(m_j\left(Z,\tilde{\theta}\right)\right)\right]_+\right\}_{j=1}^J\right) \\
& \overset{(4)}{\geq} \sup_{\tilde{\theta}\in\Theta_I} G\left(\left\{\left[v_n\left(m_{j,\tilde{\theta}}\right)\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z,\tilde{\theta}\right)\right) = 0\right]\right\}_{j=1}^J\right) + \varepsilon \\
& \overset{(5)}{\geq} G\left(\left[v_n(m_{j,\theta'})\right]_+ 1[j \in S]\right) + \varepsilon
\end{aligned}$$

where $\overset{(1)}{\geq}$ holds because $\theta \in D_n(S) \subseteq \Theta_I$, $\overset{(2)}{\geq}$ holds because $\theta \in D_n(S)$, and so, $1[j \in S] = 1[\mathbb{E}_n(m_j(Z,\theta)) \geq 0]$, $\overset{(3)}{\geq}$ holds because $\{G_{n,1}(\theta) + \varepsilon/2 \geq \bar{G}_{n,1}\}$, $\overset{(4)}{\geq}$ holds because $\delta_n > \varepsilon$ and $\overset{(5)}{\geq}$ holds because $\theta' \in D(S) \subseteq \Theta_I$, and so, $1[\mathbb{E}(m_j(Z,\theta')) = 0] \geq 1[j \in S]$. As a consequence,

$$\left\{\sup_{\theta\in\Theta} \sup_{\{\theta'\in\Theta:\|\theta'-\theta\|\leq\eta\}} \left|G\left([v_n(m_{j,\theta})]_+ 1[j \in S]\right) - G\left(\left[v_n(m_{j,\theta'})\right]_+ 1[j \in S]\right)\right|\right\} > \frac{\varepsilon}{2}$$

By a continuity argument, $\forall \eta > 0$, $\exists \gamma > 0$ and $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\{\delta_n > \varepsilon \cap A_n\}$ implies that $\{\sup_{\theta \in \Theta_I} \sup_{\|\theta' - \theta\| \leq \eta} \|v_n(m_\theta) - v_n(m_{\theta'})\| > \gamma\}$. As a consequence,

$$\limsup_{n \to +\infty} P^*(\delta_n > \varepsilon \cap A_n) \leq \limsup_{n \to +\infty} P^* \left(\sup_{\theta \in \Theta} \sup_{\{\theta' \in \Theta : \|\theta' - \theta\| \leq \eta\}} \|v_n(m_\theta) - v_n(m_{\theta'})\| > \gamma \right)$$

Taking $\eta \downarrow 0$ and by stochastic equicontinuity, this part is completed.

Point 2. By assumption, the class of functions $\left\{ m(z, \theta) : S_Z \to \mathbb{R}^J \right\}$ indexed by $\theta \in \Theta$ is stochastically equicontinuous for P and the pseudometric $\tau(m_\theta, m_{\theta'}) = \left\| \theta - \theta' \right\|$. Since Θ is assumed to be closed and bounded, Θ is totally bounded for this pseudometric. By theorem 3.7.2 in Dudley [12], this class of functions is P-Donsker and so, $v_n(m_\theta) : \Omega_n \to l_J^\infty(\Theta)$ converges to a tight Borel measurable element in $l_J^\infty(\Theta)$. The nature of the limiting process follows from consideration of its marginals. By the CLT, for every finite collection of elements of Θ, denoted by $\{\theta_l\}_{l=1}^L$, the stochastic process $\{v_n(m_{\theta_l})\}_{l=1}^L$ converges to a zero mean Gaussian random vector with a variance covariance matrix whose (l_1, l_2) element is given by $\Sigma(\theta_{l_1}, \theta_{l_2})$. This completes the proof.

Point 3. The function $H(y) = \sup_{\theta \in \Theta_I} G(\{[y_j(\theta)]_+ 1[\mathbb{E}(m_j(Z, \theta)) = 0]\}_{j=1}^J)$ is trivially continuous and non-negative. Weak convexity can be verified by definition. Homogeneity of degree β can also be verified by definition and it only remains to be shown that $\beta \geq 1$. By weak convexity, for $\alpha \in (0,1)$ and $\forall y, y' \in l_J^\infty(\Theta)$, $H(\alpha y + (1-\alpha) y') \leq \alpha H(y) + (1-\alpha) H(y')$ and if the function y' is chosen so that, $\forall \theta \in \Theta$, $y'(\theta) = 0$, then, this implies that, $H(\alpha y) \leq \alpha H(y)$. By homogeneity of degree β, $H(\alpha y) = \alpha^\beta H(y)$. Now choose $y \in l_J^\infty(\Theta)$ such that $H(y) > 0$ to deduce that $\beta \geq 1$.

Finally, we show that $\exists (\theta_0, j) \in \Theta_I \times \{1, 2, ..., J\}$ such that $\mathbb{E}(m_j(Z, \theta_0)) = 0$. Since the function $\mathbb{E}(m(Z, \theta)) : \Theta \to \mathbb{R}^J$ is lower semi-continuous, Θ_I is closed or, equivalently, $\Theta \cap \{\Theta_I\}^c$ is open. Now proceed by contradiction. That is, suppose that $\forall \theta \in \Theta_I$, $\max_{j=1,...,J} \mathbb{E}(m_j(Z, \theta)) < 0$, which implies that Θ_I is open. Since Θ_I is a proper subset of Θ, $\exists \theta' \in \Theta \cap \{\Theta_I\}^c$. By the case under consideration, $\Theta_I \neq \varnothing$, and so $\exists \theta'' \in \Theta_I$. Consider the set $S = \left\{ \theta \in \Theta : \theta'' \pi + \theta' (1 - \pi), \pi \in [0,1] \right\}$. It then follows that S is a convex set (hence, connected) and it can be expressed as the union of two non-empty open sets (by intersecting it with Θ_I and $\{\Theta_I\}^c$), which is a contradiction. As a corollary, $H(y) = 0$ implies that $y_j(\theta_0) \leq 0$.

Part 2. Let $\tilde{\delta}_n$ be defined as follows,

$$\tilde{\delta}_n = \left\{ \begin{array}{c} \sup_{\theta \in \Theta_I} G \left(\left\{ \left\{ [\sqrt{n} \hat{p}_k (\mathbb{E}_n(Y_j | x_k) - \mathbb{E}(Y_j | x_k)) + \sqrt{n} \hat{p}_k (\mathbb{E}(Y_j | x_k) - M_{j,k}(\theta))]_+ \right\}_{j=1}^J \right\}_{k=1}^K \right) \\ - \sup_{\theta \in \Theta_I} G \left(\left\{ \{ [\sqrt{n} \hat{p}_k (\mathbb{E}_n(Y_j | x_k) - \mathbb{E}(Y_j | x_k))]_+ 1[p_k (M_{j,k}(\theta) - \mathbb{E}(Y_j | x_k)) = 0] \}_{j=1}^J \right\}_{k=1}^K \right) \end{array} \right\}$$

where, $\forall (k, j) \in \{1, ..., K\} \times \{1, ..., J\}$, $p_k = P(X = x_k)$, $\hat{p}_k = n^{-1} \sum_{i=1}^n 1[X_i = x_k]$ and $\mathbb{E}_n(Y_j | x_k) = (\hat{p}_k n)^{-1} \sum_{i=1}^n Y_j 1[X_i = x_k]$.

Point 1. Define $y_n = \{\{\sqrt{n} \hat{p}_k (\mathbb{E}_n(Y_j | x_k) - \mathbb{E}(Y_j | x_k))\}_{j=1}^J\}_{k=1}^K$, $\hat{p} = \{\hat{p}_k\}_{k=1}^K$ and the functions $R_n : \mathbb{R}^K \times \mathbb{R}^{JK} \times \Theta_I \to \mathbb{R}$ and $R : \mathbb{R}^{JK} \times \Theta_I \to \mathbb{R}$ as follows,

$$\begin{aligned} R_n(\pi, y, \theta) &= G \left(\left\{ \left\{ [y_{j,k} + \sqrt{n} \pi_k (\mathbb{E}(Y_j | x_k) - M_{j,k}(\theta))]_+ \right\}_{j=1}^J \right\}_{k=1}^K \right) \\ R(y, \theta) &= G \left(\left\{ \left\{ [y_{j,k}]_+ 1[p_k (M_{j,k}(\theta) - \mathbb{E}(Y_j | x_k)) = 0] \right\}_{j=1}^J \right\}_{k=1}^K \right) \end{aligned}$$

Then, $\tilde{\delta}_n = \sup_{\theta \in \Theta_I} R_n(\hat{p}, y_n, \theta) - \sup_{\theta \in \Theta_I} R(y_n, \theta)$.

Denote $p_L = \min\{p_k\}_{k=1}^K$ and define $\Delta = \{\pi : \sum_{k=1}^K \pi_k = 1,\ \pi_k \geq p_L/2\}$. For any positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ with $\varepsilon_n = o(1)$, consider the following derivation,

$$\begin{aligned}
&\sqrt{n}P\left(\left|\tilde{\delta}_n\right| > \varepsilon_n\right) \\
&= \sqrt{n}\left\{\begin{array}{c} P\left(\left|\sup_{\theta\in\Theta_I} R_n(\hat{p}, y_n, \theta) - \sup_{\theta\in\Theta_I} R(y_n,\theta)\right| > \varepsilon_n \cap \left\{\hat{p}\in\Delta \cap \|y_n\| \leq n^{1/8}\right\}\right) \\ +P\left(\left|\sup_{\theta\in\Theta_I} R_n(\hat{p}, y_n, \theta) - \sup_{\theta\in\Theta_I} R(y_n,\theta)\right| > \varepsilon_n \cap \left\{\hat{p}\notin\Delta \cup \|y_n\| > n^{1/8}\right\}\right) \end{array}\right\} \\
&\leq \left\{\begin{array}{c} \sqrt{n}1\left[\sup_{\pi\in\Delta}\sup_{\|y\|\leq n^{1/8}}\left|\sup_{\theta\in\Theta_I} R_n(\pi,y,\theta) - \sup_{\theta\in\Theta_I} R(y,\theta)\right| > \varepsilon_n\right] \\ +\sqrt{n}P\left(\|y_n\| > n^{1/8}\right) + \sum_{k=1}^K \sqrt{n}P\left(\hat{p}_k \leq p_L/2\right) \end{array}\right\}
\end{aligned}$$

The right hand side is a sum of three terms. We now show that each term is $o(1)$. By Chebyshev's Inequality, $\sqrt{n}P\left(\|y_n\| > n^{1/8}\right) = o(1)$ and $\forall k = 1,2,...,K$, $\sqrt{n}P(\hat{p}_k \leq p_L/2) = o(1)$.

To conclude this point, we show that $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$ and $\forall (y,\pi) \in \left\{\|y\| \leq n^{1/8}\right\} \times \Delta$, $\sup_{\theta\in\Theta_I} R_n(\pi,y,\theta) = \sup_{\theta\in\Theta_I} R(y,\theta)$. By definition, $\forall (\pi,y,\theta) \in \mathbb{R}^K \times \mathbb{R}^{JK} \times \Theta_I$, $R_n(\pi,y,\theta) \geq R(y,\theta)$ and so, $\sup_{\theta\in\Theta_I} R_n(\pi,y,\theta) \geq \sup_{\theta\in\Theta_I} R(y,\theta)$.

For any $S \in \{\mathcal{P}^{\{1,...,J\}\times\{1,...,K\}}/\varnothing\}$, consider the following two sets,

$$\begin{aligned}
D_n(S) &= \left\{\theta\in\Theta_I : \left\{\exists (y,\pi) \in \left\{\|y\| \leq n^{1/8}\right\} \times \Delta : \left\{y_{j,k} + \sqrt{n}\pi_k\left(\mathbb{E}(Y_j|x_k) - M_{j,k}(\theta)\right) > 0\right\}_{(j,k)\in S}\right\}\right\} \\
D(S) &= \left\{\theta\in\Theta_I : \left\{\mathbb{E}(Y_j|x_k) - M_{j,k}(\theta) = 0\right\}_{(j,k)\in S}\right\}
\end{aligned}$$

Fix $(y,\pi) \in \left\{\|y\| \leq n^{1/8}\right\} \times \Delta$ and suppose that $\sup_{\theta\in\Theta_I} R_n(\pi,y,\theta) > \sup_{\theta\in\Theta_I} R(y,\theta)$. We now show that this implies that $\exists \bar{S} \in \{\mathcal{P}^{\{1,...,J\}\times\{1,...,K\}}/\varnothing\}$ such that $D(\bar{S}) = \varnothing$ and $D_n(\bar{S}) \neq \varnothing$. . Since Θ_I is non-empty and compact, and $R_n(\pi,y,\theta) : \Theta \to \mathbb{R}_+$ is upper semi-continuous, then $\exists\theta_0 \in \Theta_I$ such that $R_n(\pi,y,\theta_0) = \sup_{\theta\in\Theta_I} R_n(\pi,y,\theta)$. By definition, $R_n(\pi,y,\theta_0) > \sup_{\theta\in\Theta_I} R(y,\theta)$ implies that $\exists (j,k) \in \{1,...,J\} \times \{1,...,K\}$ such that $\{y_{j,k} + \sqrt{n}\pi_k(\mathbb{E}(Y_j|x_k) - M_{j,k}(\theta_0)) > 0\}$. Let $\bar{S} \in \{\mathcal{P}^{\{1,...,J\}\times\{1,...,K\}}/\varnothing\}$ be defined so that, $\forall (j,k) \in \bar{S}$, $\{y_{j,k} + \sqrt{n}\pi_k(\mathbb{E}(Y_j|x_k) - M_{j,k}(\theta_0)) > 0\}$ and $\forall (j,k) \in \{\{1,...,J\} \times \{1,...,K\}\}\backslash\bar{S}$, $\{y_{j,k} + \sqrt{n}\pi_k(\mathbb{E}(Y_j|x_k) - M_{j,k}(\theta_0)) \leq 0\}$. According to this definition, $\theta_0 \in D_n(\bar{S})$. Furthermore, if $D(\bar{S}) \neq \varnothing$, then $\exists\theta_1 \in \Theta_I$ such that $R(y,\theta_1) \geq R_n(\pi,y,\theta_0)$, which would be a contradiction.

To conclude, it suffices to show that $\exists N \in \mathbb{N}$, such that, $\forall n \geq N$ and $\forall S \in \{\mathcal{P}^{\{1,...,J\}\times\{1,...,K\}}/\varnothing\}$, $D(S) = \varnothing$ implies $D_n(S) = \varnothing$. Fix $S \in \{\mathcal{P}^{\{1,...,J\}\times\{1,...,K\}}/\varnothing\}$ arbitrarily and consider the following argument. The event $D(S) = \varnothing$ implies that $\{\min_{(j,k)\in S}\max_{\theta\in\Theta_I}(\mathbb{E}(Y_j|x_k) - M_{j,k}(\theta)) < 0\}$, which, by continuity of $M_{j,k}$, implies that $\exists\delta > 0$ such that $\{\min_{(j,k)\in S}\max_{\theta\in\Theta_I}(\mathbb{E}(Y_j|x_k) - M_{j,k}(\theta)) < -\delta\}$. If so, $\exists (j,k) \in S$ such that, $\forall (y,\pi,\theta) \in \left\{\|y\| \leq n^{1/8}\right\} \times \Delta \times \Theta_I$, $\{y_{j,k} + \sqrt{n}\pi_k(\mathbb{E}(Y_j|x_k) - M_{j,k}(\theta)) \leq n^{1/8} - \sqrt{n}p_L/2\delta\}$ and thus, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $D_n(S) = \varnothing$.

Points 2 and 3. We now show that $\{\{\hat{p}_k(\mathbb{E}_n(Y_j|x_k) - \mathbb{E}(Y_j|x_k))\}_{j=1}^J\}_{k=1}^K = B(\mathbb{E}_n(Z) - \mathbb{E}(Z))$ where $\{\mathbb{E}_n(Z) - \mathbb{E}(Z)\}$ is an average of i.i.d. vectors denoted by $\{Z_i\}_{i=1}^{+\infty}$, such that $\mathbb{E}(Z_i) = 0_{\rho\times1}$, $V(Z_i) = \mathbf{I}_\rho$ and $\mathbb{E}(|Z_i|^3) < +\infty$.

The random vector $\{\{\hat{p}_k(\mathbb{E}_n(Y_j|x_k) - \mathbb{E}(Y_j|x_k))\}_{j=1}^J\}_{k=1}^K$ is the sample average of an i.i.d. sample of $\{\{1[X = x_k](Y_j - \mathbb{E}(Y_j|x_k))\}_{j=1}^J\}_{k=1}^K$. Denote by Υ the variance covariance matrix of $\{\{1[X = x_k]Y_j\}_{j=1}^J\}_{k=1}^K$, which is also the variance covariance matrix of $\{\{1[X = x_k](Y_j - \mathbb{E}(Y_j|x_k))\}_{j=1}^J\}_{k=1}^K$. Notice that Υ is a block diagonal matrix with K diagonal blocks, whose k^{th} block is given by p_kV_k, where $p_k = P(Y|X = x_k)$ and $V_k = V(Y|X = x_k)$. For every $k = 1,2,...,K$, let ρ_k be the rank of V_k, let B_k be defined as the $J \times \rho_k$ dimensional matrix such that $B_kB_k' = p_kV_k$ and let B be defined as the $JK \times \rho$ dimensional block diagonal matrix that results from using the matrices $\{B_k\}_{k=1}^K$ as the diagonal blocks. For example, for

$K = 3$, B is given by,

$$B = \begin{bmatrix} B_1 & 0_{J\times\rho_2} & 0_{J\times\rho_3} \\ 0_{J\times\rho_1} & B_2 & 0_{J\times\rho_3} \\ 0_{J\times\rho_1} & 0_{J\times\rho_2} & B_3 \end{bmatrix}$$

By construction, $B \in \mathbb{R}^{JK\times\rho}$, has rank ρ and $BB' = \Upsilon$.

For every $(i,k) \in \{1,2,...,n\} \times \{1,2,...,K\}$, repeat the following argument. If $X_i = x_k$, define $W_{k,i} \in \mathbb{R}^{\rho_k}$ such that $\{Y_{j,i} - \mathbb{E}(Y_j|x_k)\}_{j=1}^{J} = B_k W_{k,i}$, and if $X_i \neq x_k$, define $W_{k,i} = 0_{\rho_k\times 1}$. Then, define $W_i = [W_{1,i}, ..., W_{K,i}] \in \mathbb{R}^{\rho}$. By construction, notice that, $\forall i = 1,2,...,n$, $\mathbb{E}(W_i) = 0_{\rho\times 1}$.

Finally, $\forall i = 1,2,...,n$, we define $Z_i = W_i$ and so, $\mathbb{E}(Z) = 0_{\rho\times 1}$. By construction, $\forall i = 1,...,n$, $\{1[X_i = x_k](Y_i - \mathbb{E}(Y|x_k))\}_{k=1}^{K} = B(Z_i - \mathbb{E}(Z))$ and $\mathbb{E}(Z_i - \mathbb{E}(Z)) = 0_{\rho\times 1}$. Since the variance of BZ_i equals BB' and B has rank ρ, then, $V(Z_i - \mathbb{E}(Z)) = \mathbf{I}_\rho$. Finally, if $\{Y_i|X_i = x_k\}_{k=1}^{K}$ is assumed to have finite third absolute moments, then $(Z_i - \mathbb{E}(Z))$ will also have finite third absolute moments. Averaging these observations, we deduce that $\{\{\hat{p}_k(\mathbb{E}_n(Y_j|x_k) - \mathbb{E}(Y_j|x_k))\}_{j=1}^{J}\}_{k=1}^{K} = B(\mathbb{E}_n(Z) - \mathbb{E}(Z))$. By slightly abusing the notation, $\forall (j,k) \in \{1,2,...,J\} \times \{1,2,...,K\}$, let $B_{j,k} \in \mathbb{R}^{1\times\rho}$ denote the $((K-1)j+k)^{th}$ row of B. The function $\tilde{H}(y) : \mathbb{R}^\rho \to \mathbb{R}$ is defined as follows,

$$\tilde{H}(y) = \sup_{\theta\in\Theta_I}\left\{G\left(\left\{\left\{[B_{j,k}y]_+ 1[p_k(M_{j,k}(\theta) - \mathbb{E}(Y_j|x_k)) = 0]\right\}_{j=1}^{J}\right\}_{k=1}^{K}\right)\right\}$$

We show that this function has all the desired properties. This function is continuous, non-negative, and weakly convex by the arguments used in the previous part. Homogeneity of degree one can be verified by definition. Since the matrix B has rank ρ, $\forall (j,k) \in \{1,...,J\} \times \{1,...,K\}$, $B_{j,k} \neq 0_{\rho\times 1}$. By the arguments in part 1, $\exists\ (\theta_0, j, k) \in \Theta_I \times \{1,...,J\} \times \{1,...,K\}$ such that $\mathbb{E}(Y_j|x_k) = M_{j,k}(\theta_0)$, and so, if we define $b' = B_{j,k}$, $\tilde{H}(y) = 0$ implies that for $b \neq 0_{\rho\times 1}$, $b'y \leq 0$.

Finally, consider $y_A \in \tilde{H}^{-1}((h_B - \varepsilon_n, h_B + \varepsilon_n])$. By definition, this means that $\exists h_A$ such that $\|h_A - h_B\| < \varepsilon_n$ and $\tilde{H}(y_A) = h_A$. To conclude the proof, we need to show that $\exists y_B \in \mathbb{R}^\rho$ such that $\|y_A - y_B\| \leq O(\varepsilon_n)$ and $\tilde{H}(y_B) = h_B$.

We consider first the case when $G(x) = \sum_{s=1}^{JK} w_s x_s$ for positive weights $\{w_s\}_{s=1}^{JK}$. For any $z \in \mathbb{R}^\rho$, let $g(z,\theta) = G(\{\{[B_{j,k}z]_+ 1[p_k(M_{j,k}(\theta) - \mathbb{E}(Y_j|x_k)) = 0]\}_{j=1}^{J}\}_{k=1}^{K})$. Since $g(z,\theta)$ depends on θ through indicator functions, then we partition Θ_I into finitely many subsets, according to whether each of the JK indicator functions is turned on or off. From each subset, we can extract one representative. Let $\{\theta_1, \theta_2, ..., \theta_\pi\}$ denote the set of such representatives. By construction, $\forall z \in \mathbb{R}^\rho$, $\max_{\theta\in\Theta_I} g(z,\theta) = \max_{\theta\in\{\theta_1,...,\theta_\pi\}} g(z,\theta)$. For any $(z,\theta) \in \{\mathbb{R}^\rho, \Theta_I\}$, let $\Lambda_+(z,\theta)$ denote the subset of $\{1,...,J\} \times \{1,...,K\}$ such that $M_{j,k}(\theta) = \mathbb{E}(Y_j|x_k)$ and $B_{j,k}z > 0$ and let $\Lambda_0(z,\theta)$ denote the subset of $\{1,...,J\} \times \{1,...,K\}$ such that $M_{j,k}(\theta) = \mathbb{E}(Y_j|x_k)$ and $B_{j,k}z = 0$.

Let $\{\theta_1, ..., \theta_m\}$ denote the subset of the representatives that maximize $g(y_A, \theta)$. Consider any arbitrary $\theta' \in \{\theta_1, ..., \theta_m\}$. By definition, $y_A \in \mathbb{R}^\rho$ satisfies the following equations: $\forall (j,k) \in \Lambda_0(y_A, \theta')$, $B_{j,k}z = 0$ and $\forall (j,k) \in \Lambda_+(y_A, \theta')$, $B_{j,k}z = h_{A,(j,k)} > 0$. By summing the equations for $(j,k) \in \Lambda_+(y_A, \theta')$, we get $\sum_{(j,k)\in\Lambda_+(y_A,\theta')} h_{A,(j,k)} = h_A$. Thus, $y_A \in \mathbb{R}^\rho$ satisfies the following system of equations,

$$\begin{bmatrix} \sum_{(j,k)\in\Lambda_+(y_A,\theta')} B_{(j,k)} \\ \left[B_{(j,k)}\right]_{(j,k)\in\Lambda_0(y_A,\theta')} \end{bmatrix} z = \begin{bmatrix} h_A \\ \left[\vec{0}\right]_{(j,k)\in\Lambda_0(y_A,\theta')} \end{bmatrix}$$

We can repeat this process for the rest of the maximizers, that is, $\forall\theta'' \in \{\theta_2, ..., \theta_m\} \backslash \theta'$. Instead of expressing the information contained in $\Lambda_0(y_A, \theta'')$ as $\sum_{(j,k)\in\Lambda_+(y_A,\theta'')} B_{j,k} = h_A$ we reexpress it as,

$\sum_{(j,k)\in\Lambda_+(y_A,\theta'')} B_{j,k} - \sum_{(j,k)\in\Lambda_+(y_A,\theta')} B_{j,k} = 0$, which gives the following new set of equations,

$$\begin{bmatrix} \sum_{(j,k)\in\Lambda_+(y_A,\theta'')} B_{j,k} - \sum_{(j,k)\in\Lambda_+(y_A,\theta')} B_{j,k} \\ \left[B_{(j,k)}\right]_{(j,k)\in\Lambda_0(y_A,\theta'')} \end{bmatrix} z = \begin{bmatrix} 0 \\ \left[\vec{0}\right]_{(j,k)\in\Lambda_0(y_A,\theta'')} \end{bmatrix}$$

If we put together all the equations from $\theta \in \{\theta_1, \theta_2, ..., \theta_m\}$ in this fashion, we will produce a system of linear equations of the form $[C_1, C_2]' z = [h_A, \vec{0}]'$ where the matrix $[C_1, C_2]'$ does not depend on h_A. Consider the homogenous system $C_2 z = \vec{0}$. The matrix C_2 may or may not have full rank, but can always be reduced to a system $C_3 z = \vec{0}$, where C_3 has full rank. Since $h_A > 0$, $[C_1, C_3]'$ has full rank. If this rank is ρ, then $y_A = \left[[C_1, C_3]'\right]^{-1} [h_A, \vec{0}]'$. If the rank is less than ρ, pick C_4 so that $[C_1, C_3, C_4]$ has rank ρ, set c such that $C_4 y_A = c$ and add the additional (equality) restrictions satisfied by y_A, of the form $C_4 z = c$. Then, $y_A = \left[[C_1, C_3, C_4]'\right]^{-1} [h_A, \vec{0}, c]'$.

Consider $y_B = \left[[C_1, C_3, C_4]'\right]^{-1} [h_B, \vec{0}, c]'$. By construction, $\|y_A - y_B\| = O(\varepsilon_n)$. By construction and a continuity argument, $\forall \theta \in \{\theta_1, ...\theta_m\}$, $\Lambda_+(y_A, \theta) = \Lambda_+(y_B, \theta)$ and if $\sum_{(j,k)\in\Lambda_+(y_A,\theta)} B_{j,k} y_A = h_A$, then $\sum_{(j,k)\in\Lambda_+(y_B,\theta)} B_{j,k} y_B = h_B$. Also by construction, $\forall \theta \in \{\theta_1, ...\theta_m\}$, then $\Lambda_0(y_A, \theta) = \Lambda_0(y_B, \theta)$. By continuity, $\forall (j,k) \in \{1, 2, ..., J\} \times \{1, 2, ..., K\}$ such that $M_{j,k}(\theta) = \mathbb{E}(Y_j | x_k)$ and $B_{j,k} y_A < 0$, then $B_{j,k} y_B < 0$. As a consequence, $\forall \theta \in \{\theta_1, ...\theta_m\}$, $g(y_B, \theta) = h_B$. By continuity, $\forall \theta \in \{\theta_1, ..., \theta_\pi\} \setminus \{\theta_1, ...\theta_m\}$, $g(y_B, \theta) < h_B$. Thus, by construction, $\tilde{H}(y_B) = h_B$.

The arguments for $G(x) = \max_{i=1,...,JK} \{w_i x_i\}$ for positive weights $\{w_i\}_{i=1}^{JK}$ are similar and, therefore, omitted.

Part 3. If $\Theta_I = \varnothing$, then, by definition, $\Gamma_n = 0$. ∎

In part 2 of theorem A.1, assumption (CF) was used to provide certain properties to the function $\tilde{H}$. The following theorem shows how this result changes when assumption (CF) is replaced by assumption (CF').

Theorem A.2 *Let ρ denote the rank of the variance covariance matrix of the vector $\{\{1[X = x_k] Y_j\}_{j=1}^J\}_{k=1}^K$. If we assume (B1)-(B4), (CF') and $\Theta_I \neq \varnothing$, then, $\Gamma_n = \tilde{H}(\sqrt{n}(\mathbb{E}_n(Z) - \mathbb{E}(Z))) + \tilde{\delta}_n$, where,*

1. *for any $\varepsilon_n = O(n^{-1/2})$, $P(|\tilde{\delta}_n| > \varepsilon_n) = o(n^{-1/2})$,*

2. *$\{\mathbb{E}_n(Z) - \mathbb{E}(Z)\} : \Omega_n \to \mathbb{R}^\rho$ is a zero mean sample average of n i.i.d. observations from a distribution with variance covariance matrix $\mathbf{I}_\rho$. Moreover, this distribution has finite third absolute moments,*

3. *$\tilde{H} : \mathbb{R}^\rho \to \mathbb{R}$ is continuous, non-negative, weakly convex and homogeneous of degree $\beta \geq 1$. For any $\mu > 0$, any h such that $|h| \geq \mu > 0$ and any positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\varepsilon_n = o(1)$, $\{\tilde{H}^{-1}(\{h\}^{\varepsilon_n}) \cap \|y\| \leq O(\sqrt{g_n})\} \subseteq \{\tilde{H}^{-1}(\{h\})\}^{\delta_n}$, where $\delta_n = O(\varepsilon_n \sqrt{g_n})$. Finally, $\tilde{H}(y) = 0$ implies that for some non-zero vector $b \in \mathbb{R}^\rho$, $b'y \leq 0$.*

Proof. The definitions of $\{\mathbb{E}_n(Z) - \mathbb{E}(Z)\}$ and $\tilde{H}$ are exactly the same as in part 2 of theorem A.1. To conclude, we only need to show that $\forall \mu > 0$ and $\forall h$ such that $|h| \geq \mu$, $\{\tilde{H}^{-1}(\{h\}^{\varepsilon_n}) \cap \|y\| \leq O(\sqrt{g_n})\} \subseteq \{\tilde{H}^{-1}(\{h\})\}^{\delta_n}$ where $\delta_n = O(\varepsilon_n \sqrt{g_n})$. To this purpose, consider $y' \in \tilde{H}^{-1}(\{h\}^{\varepsilon_n})$ such that $\|y'\| \leq O(\sqrt{g_n})$. We need to show that $\exists y \in \tilde{H}^{-1}(\{h\})$ such that $\|y' - y\| \leq O(\varepsilon_n \sqrt{g_n})$. Consider $y = y'(h/h')^{1/\beta}$. By homogeneity of degree β, $\tilde{H}(y) = h$. By definition,

$$\|y' - y\| \leq \|y'\| \left|1 - (h'/h)^{-1/\beta}\right| \leq O(\sqrt{g_n}) \max\left\{1 - (h'/h)^{-1/\beta}, (h'/h)^{-1/\beta} - 1\right\}$$

where $|h' - h| \leq \varepsilon_n$. For any fixed h such that $|h| \geq \mu > 0$ and $h' \in (h - \varepsilon_n, h + \varepsilon_n]$, a first order Taylor expansion argument implies that $\max\{1 - (h'/h)^{-1/\beta}, (h'/h)^{-1/\beta} - 1\} \leq O(|h' - h|) = O(\varepsilon_n)$. As a consequence, $\|y' - y\| \leq O(\varepsilon_n \sqrt{g_n})$, completing the proof. ∎

A.3.2 Representation result for the bootstrap test statistic

The following theorem shows that the bootstrap test statistic has a representation that is analogous to the one obtained for the population test statistic.

Theorem A.3 *Part 1. Assume (A1)-(A4), (CF') and $\Theta_I \neq \varnothing$. Then, $\Gamma_n^* = H\left(v_n^*\left(m_\theta\right)\right) + \delta_n^*$, where,*

1. *for any $\varepsilon > 0$, $\lim_{n \to +\infty} P^*\left(|\delta_n^*| > \varepsilon | \mathcal{X}_n\right) = 0$, a.s.,*
2. *$\{v_n^*\left(m_\theta\right)|\mathcal{X}_n\} : \Omega_n \to l_J^\infty\left(\Theta\right)$ is an empirical process that converges weakly to the same Gaussian process as in theorem A.1, i.o.p.,*
3. *$H : l_J^\infty\left(\Theta\right) \to \mathbb{R}$ is the same function as in theorem A.1.*

Part 2. Let ρ denote the rank of the variance covariance matrix of the vector $\{\{1\left[X = x_k\right] Y_j\}_{j=1}^J\}_{k=1}^K$. If we assume (B1)-(B4), (CF), $\Theta_I \neq \varnothing$, and we choose the bootstrap procedure to be the one specialized for the conditionally separable model, then, $\Gamma_n^* = \tilde{H}\left(\sqrt{n}\left(\mathbb{E}_n^*\left(Z\right) - \mathbb{E}_n\left(Z\right)\right)\right) + \tilde{\delta}_n^*$, where,

1. *$P(\tilde{\delta}_n^* = 0|\mathcal{X}_n) = 1[\tilde{\delta}_n^* = 0]$ and $\liminf\{\tilde{\delta}_n^* = 0\}$, a.s.,*
2. *$\{\left(\mathbb{E}_n^*\left(Z\right) - \mathbb{E}_n\left(Z\right)\right)|\mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero mean sample average of n independent observations from a distribution with variance covariance matrix $\hat{V}$. Moreover, this distribution has finite third moments, a.s., and $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$,*
3. *$\tilde{H} : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in theorem A.1.*

Part 3. Assume (A1)-(A4), (CF') and $\Theta_I = \varnothing$. Then, $\liminf\{P(\Gamma_n^* = 0|\mathcal{X}_n) = 1\}$, a.s..

Proof. Part 1. By the CLT for bootstrapped empirical processes applied to P-Donsker classes (see, for example, Giné and Zinn [13] or theorem 3.6.13 in van der Vaart and Wellner [29]), $\{v_n^*|\mathcal{X}_n\} : \Omega_n \to l_J^\infty\left(\Theta\right)$ converges weakly to ζ, i.o.p., where ζ is the Gaussian process described in theorem A.1. Let the function $H : l_J^\infty\left(\Theta\right) \to \mathbb{R}$ be defined as in theorem A.1, let $H_n : l_J^\infty\left(\Theta\right) \to \mathbb{R}$ be the following function,

$$H_n\left(y\right) = \sup_{\theta \in \hat{\Theta}_I(\tau_n)} G\left(\left\{\left[y_j\left(\theta\right)\right]_+ 1\left[\left|\mathbb{E}_n\left(m_j\left(Z,\theta\right)\right)\right| \leq \tilde{\tau}_n/\sqrt{n}\right]\right\}_{j=1}^J\right)$$

and let $\delta_n^* = H_n\left(v_n^*\left(m_\theta\right)\right) - H\left(v_n^*\left(m_\theta\right)\right)$. To conclude the proof of this part, it suffices to show that $\forall \varepsilon > 0$, $P\left(|\delta_n^*| > \varepsilon|\mathcal{X}_n\right) = o\left(1\right)$, a.s..

Step 1: We now show that $P\left(\delta_n^* < 0|\mathcal{X}_n\right) = o\left(1\right)$, a.s.. Define the event A_n as follows,

$$A_n = \left\{\left\{\Theta_I \subseteq \hat{\Theta}_I\left(\tau_n\right)\right\} \bigcap \left\{\bigcap_{\theta \in \Theta} \bigcap_{j=1,\ldots,J} \left\{\left\{\mathbb{E}\left(m_j\left(Z,\theta\right)\right) = 0\right\} \Longrightarrow \left\{\left|\mathbb{E}_n\left(m_j\left(Z,\theta\right)\right)\right| \leq \tilde{\tau}_n/\sqrt{n}\right\}\right\}\right\}$$

By definition, A_n implies $\{\delta_n^* \geq 0\}$. Conditional on the sample, A_n is non-random, and so, it suffices to show the $\liminf\{A_n\}$, a.s., which follows from the LIL.

Step 2: We now show that $\forall \varepsilon > 0$, $P\left(\delta_n^* > \varepsilon|\mathcal{X}_n\right) = o\left(1\right)$, a.s..

For any $\epsilon > 0$, let $\Theta_I\left(\epsilon\right) = \{\theta \in \Theta : \{\mathbb{E}\left(m_j\left(Z,\theta\right)\right) \leq \epsilon\}_{j=1}^J\}$ and let $H^\epsilon : l_J^\infty\left(\Theta\right) \to \mathbb{R}$ denote the function $H^\epsilon\left(y\right) = \sup_{\theta \in \Theta_I(\epsilon)} G(\{\left[y_j\left(\theta\right)\right]_+ 1\left[\left|\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\right| < \epsilon\right]\}_{j=1}^J)$. For a positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\varepsilon_n = o\left(1\right)$, $\left(\tau_n/\sqrt{n}\right)\varepsilon_n^{-1} = o\left(1\right)$ and $\left(\tilde{\tau}_n/\sqrt{n}\right)\varepsilon_n^{-1} = o\left(1\right)$, a.s., let A_n' denote the following event,

$$A_n' = \left\{\left\{\hat{\Theta}_I\left(\tau_n\right) \subseteq \Theta_I\left(\varepsilon_n\right)\right\} \bigcap \left\{\bigcap_{\theta \in \Theta} \bigcap_{j=1,\ldots,J} \left\{\left\{\left|\mathbb{E}_n\left(m_j\left(Z,\theta\right)\right)\right| \leq \tilde{\tau}_n/\sqrt{n}\right\} \Longrightarrow \left\{\left|\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\right| < \varepsilon_n\right\}\right\}\right\}$$

and let $\eta_n^{H_1}$ and $\eta_n^{H_2}$ be defined by,

$$\eta_n^{H_1} = \left\{ \begin{array}{c} \sup_{\theta\in\Theta_I(\varepsilon_n)} G\left(\left\{[v_n^*(m_{j,\theta})]_+ 1[|\mathbb{E}(m_j(Z,\theta))| < \varepsilon_n]\right\}_{j=1}^J\right) \\ -\sup_{\theta\in\Theta_I} G\left(\left\{[v_n^*(m_{j,\theta})]_+ 1[|\mathbb{E}(m_j(Z,\theta))| < \varepsilon_n]\right\}_{j=1}^J\right) \end{array} \right\}$$

$$\eta_n^{H_2} = \left\{ \begin{array}{c} \sup_{\theta\in\Theta_I} G\left(\left\{[v_n^*(m_{j,\theta})]_+ 1[|\mathbb{E}(m_j(Z,\theta))| < \varepsilon_n]\right\}_{j=1}^J\right) \\ -\sup_{\theta\in\Theta_I} G\left(\left\{[v_n^*(m_{j,\theta})]_+ 1[\mathbb{E}(m_j(Z,\theta)) = 0]\right\}_{j=1}^J\right) \end{array} \right\}$$

Notice that A_n' implies $\{H_n(v_n^*(m_\theta)) \leq H^{\varepsilon_n}(v_n^*(m_\theta))\}$, which, in turn, implies that $\{\delta_n^* \leq \eta_n^{H_1} + \eta_n^{H_2}\}$. Based on this, consider the following derivation,

$$\begin{aligned} P(\delta_n^* > \varepsilon|\mathcal{X}_n) &= P\left(\{\delta_n^* > \varepsilon\} \cap A_n'|\mathcal{X}_n\right) + P\left(\{\delta_n^* > \varepsilon\} \cap \left\{A_n'\right\}^c|\mathcal{X}_n\right) \\ &\leq P\left(\{\delta_n^* > \varepsilon\} \cap \{H_n(v_n^*(m_\theta)) \leq H^{\varepsilon_n}(v_n^*(m_\theta))\}|\mathcal{X}_n\right) + P\left(\left\{A_n'\right\}^c|\mathcal{X}_n\right) \\ &\leq P\left(\eta_n^{H_1} > \varepsilon/2|\mathcal{X}_n\right) + P\left(\eta_n^{H_2} > \varepsilon/2|\mathcal{X}_n\right) + P\left(\left\{A_n'\right\}^c|\mathcal{X}_n\right) \end{aligned}$$

By the LIL, $\liminf\{A_n'\}$, a.s. and, therefore, $P(\{A_n'\}^c|\mathcal{X}_n) = o(1)$, a.s.. To conclude the proof of this step, it suffices to show that, $\forall\varepsilon > 0$ and $\forall i = 1, 2$, $P(\eta_n^{H_i} > \varepsilon/2|\mathcal{X}_n) = o(1)$ a.s.. We only cover the one for $i = 1$ because the proof for $i = 2$ follows from similar arguments.

Fix $\varepsilon > 0$. Let $G_{n,1}(\theta) = G(\{[v_n^*(m_{j,\theta})]_+ 1[|\mathbb{E}(m_j(Z,\theta))| < \varepsilon_n]\}_{j=1}^J)$, $\bar{G}_{n,1} = \sup_{\theta\in\Theta_I(\varepsilon_n)} G_{n,1}(\theta)$, $G_{n,2} = G(\{[v_n^*(m_{j,\theta})]_+ 1[|\mathbb{E}(m_j(Z,\theta))| < \varepsilon_n]\}_{j=1}^J)$ and $\bar{G}_{n,2} = \sup_{\theta\in\Theta_I} G_{n,2}(\theta)$. By definition, $\eta_n^{H_1} = \bar{G}_{n,1} - \bar{G}_{n,2}$, and so $\{\eta_n^{H_1} > \varepsilon/2|\mathcal{X}_n\}$ implies that $\{\exists\theta \in \{\Theta_I(\varepsilon_n) \cap \{\Theta_I\}^c\} : \{G_{n,1}(\theta) + \varepsilon/4 \geq \bar{G}_{n,1}\}|\mathcal{X}_n\}$.

For any $S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$, consider the sets $D_n(S)$ and $D(S)$ defined as follows,

$$D_n(S) = \left\{\Theta \cap \left\{\left\{\bigcap_{j\in S}\{\mathbb{E}(m_j(Z,\theta)) \in [-\varepsilon_n, \varepsilon_n]\}\right\} \bigcap \left\{\bigcap_{j\in\{1,2,..,J\}\backslash S}\{\mathbb{E}(m_j(Z,\theta)) \leq -\varepsilon_n\}\right\}\right\}\right\}$$

$$D(S) = \left\{\Theta \cap \left\{\bigcap_{j\in S}\{\mathbb{E}(m_j(Z,\theta)) = 0\}\right\} \bigcap \left\{\bigcap_{j\in\{1,2,..,J\}\backslash S}\{\mathbb{E}(m_j(Z,\theta)) \leq 0\}\right\}\right\}$$

By definition, $\{\Theta_I(\varepsilon_n) \cap \{\Theta_I\}^c\} \subseteq \bigcup_{S\in\{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}} D_n(S)$ and so, the event $\{\eta_n^{H_1} > \varepsilon/2|\mathcal{X}_n\}$ implies that $\bigcup_{S\in\{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}}\left\{\exists\theta \in D_n(S) : \{G_{n,1}(\theta) + \varepsilon/4 \geq \bar{G}_{n,1}\}|\mathcal{X}_n\right\}$. For every $S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$ and $\forall\eta > 0$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, the event $\{\exists\theta \in D_n(S)\}$ implies that $\left\{\exists\theta' \in D(S) : \{\left\|\theta - \theta'\right\| < \eta\}\right\}$. Thus, $\forall S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$ and $\forall\eta > 0$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\left\{\exists\theta \in D_n(S) : \{G_{n,1}(\theta) + \varepsilon/4 \geq \bar{G}_{n,1}\}|\mathcal{X}_n\right\}$ is equivalent to $\left\{\exists(\theta,\theta') \in \{D_n(S) \times D(S)\} : \{\{\left\|\theta - \theta'\right\| \leq \eta\} \cap \{G_{n,1}(\theta) + \varepsilon/4 \geq \bar{G}_{n,1}\}\}|\mathcal{X}_n\right\}$. Therefore, $\forall\eta > 0$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, the event $\{\eta_n^{H_1} > \varepsilon/2|\mathcal{X}_n\}$ is equivalent to the following event,

$$\bigcup_{S\in\{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}}\left\{\left\{\eta_n^{H_1} > \varepsilon/2\right\} \cap \left\{ \begin{array}{c} \exists(\theta,\theta') \in \{D_n(S) \times D(S)\} : \\ \left\{\left\{\left\|\theta - \theta'\right\| \leq \eta\right\} \cap \left\{G_{n,1}(\theta) + \varepsilon/4 \geq \bar{G}_{n,1}\right\}\right\} \end{array} \right\} \middle| \mathcal{X}_n\right\}$$

Now, $\forall\eta > 0$ and $\forall S \in \{\mathcal{P}^{\{1,2,..,J\}}/\varnothing\}$, the event,

$$\left\{\left\{\eta_n^{H_1} > \varepsilon/2\right\} \cap \left\{\exists(\theta,\theta') \in \{D_n(S) \times D(S)\} : \left\{\left\{\left\|\theta - \theta'\right\| \leq \eta\right\} \cap \left\{G_{n,1}(\theta) + \varepsilon/4 \geq \bar{G}_{n,1}\right\}\right\}\right\}|\mathcal{X}_n\right\}$$

leads to the following derivation,

$$
\begin{aligned}
& G\left(\left[v_n^*\left(m_{j,\theta}\right)\right]_+ 1\left[j \in S\right]\right)+\frac{\varepsilon}{4} \\
& \overset{(1)}{\geq} G\left(\left[v_n^*\left(m_{j,\theta}\right)\right]_+ 1\left[\left|\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\right|<\varepsilon_n\right]\right)+\frac{\varepsilon}{4} \\
& \overset{(2)}{\geq} \sup_{\tilde{\theta}\in\Theta_I(\varepsilon_n)} G\left(\left\{\left[v_n^*\left(m_{j,\tilde{\theta}}\right)\right]_+ 1\left[\left|\mathbb{E}\left(m_j\left(Z,\tilde{\theta}\right)\right)\right|<\varepsilon_n\right]\right\}_{j=1}^J\right) \\
& \overset{(3)}{\geq} \sup_{\tilde{\theta}\in\Theta_I} G\left(\left\{\left[v_n^*\left(m_{j,\tilde{\theta}}\right)\right]_+ 1\left[\left|\mathbb{E}\left(m_j\left(Z,\tilde{\theta}\right)\right)\right|<\varepsilon_n\right]\right\}_{j=1}^J\right)+\frac{\varepsilon}{2} \\
& \overset{(4)}{\geq} G\left(\left\{\left[v_n^*\left(m_{j,\theta'}\right)\right]_+ 1\left[\left|\mathbb{E}\left(m_j\left(Z,\theta'\right)\right)\right|<\varepsilon_n\right]\right\}_{j=1}^J\right)+\frac{\varepsilon}{2} \\
& \overset{(5)}{\geq} G\left(\left[v_n^*\left(m_{j,\theta'}\right)\right]_+ 1\left[j \in S\right]\right)+\frac{\varepsilon}{2}
\end{aligned}
$$

where $\overset{(1)}{\geq}$ holds because $\theta \in D_n(S)$, and so, $1[j \in S] \geq 1[|\mathbb{E}(m_j(Z,\theta))| < \varepsilon_n]$, $\overset{(2)}{\geq}$ holds by $\{G_{n,1}(\theta)+\varepsilon/4 \geq \bar{G}_{n,1}\}$, $\overset{(3)}{\geq}$ holds because $\{\eta_n^{H_1} > \varepsilon/2\}$, $\overset{(4)}{\geq}$ holds because $\theta' \in D(S) \subseteq \Theta_I$ and $\overset{(5)}{\geq}$ holds because $\theta' \in D(S)$, and thus, $1[|\mathbb{E}(m_j(Z,\theta'))| < \varepsilon_n] \geq 1[j \in S]$. By the arguments used in the proof of theorem A.1 (part 1), $\forall \eta > 0$, $\exists \gamma > 0$ such that,

$$
\limsup_{n\to+\infty} P^*\left(\eta_n^{H_1} > \varepsilon/2|\mathcal{X}_n\right) \leq \limsup_{n\to+\infty} P^*\left(\sup_{\theta\in\Theta}\sup_{\{\theta'\in\Theta:\|\theta'-\theta\|\leq\eta\}} \left\|v_n^*\left(m_\theta\right)-v_n^*\left(m_{\theta'}\right)\right\| > \gamma \middle| \mathcal{X}_n\right)
$$

If we take $\eta \downarrow 0$, theorem 3.6.13 in van der Vaart and Wellner [29] implies that the right hand side is equal to zero i.o.p..

Part 2. Let the matrices $\{B_k\}_{k=1}^K$ and B be defined as in the proof of theorem A.1 (part 2).

Step 1: We now show that, conditionally on $\mathcal{X}_n$, $\{\{\tilde{p}_k^*(\mathbb{E}_n^*(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k))\}_{k=1}^K\}_{j=1}^J$ is the average of n independent observations from a distribution with variance covariance matrix $\hat{\Upsilon} = B\hat{V}B'$ such that $||\hat{V}-\mathbf{I}_\rho|| \leq O_p(n^{-1/2})$ and with finite third absolute moments, a.s.. For every $k = 1,2,...,K$, let $\tilde{p}_k^* = \bar{p}_k$ in the fixed design case and $\tilde{p}_k^* = \hat{p}_k^*$ in the random design case.

We only cover the proof for the fixed design case because the one for the random design case follows from similar arguments. Let $\{n_1,n_2,...,n_K\}$ denote the number of observations in the sample of each covariate value, and so, $\Sigma_{k=1}^K n_k = n$. For each $k = 1,2,...,K$, extract a bootstrap sample of size n_k from the observations in the sample that satisfy $X_i = x_k$ and denote this random sample by $\{Y_{i,k}^*\}_{i=1}^{n_k}$. Next, construct a sample of size n, where that the first n_1 observations are given by $\{Y_{i,1}^* - \mathbb{E}_n(Y|x_1), 0_{1\times J},...,0_{1\times J}\}_{i=1}^{n_1}$, the next n_2 observations are given by $\{0_{1\times J}, Y_{i,2}^* - \mathbb{E}_n(Y|x_2), 0_{1\times J},...,0_{1\times J}\}_{i=1}^{n_2}$, and so on. As a result, we have constructed n observations of JK dimensional vectors, whose average is $\{\{\tilde{p}_k^*(\mathbb{E}_n^*(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k))\}_{k=1}^K\}_{j=1}^J$. Conditional on the sample and the design, these observations are independent, with variance covariance matrix $\hat{\Upsilon}$ and finite third absolute moments, a.s..

Step 2: The next step is to show that $\{\{\tilde{p}_k^*(\mathbb{E}_n^*(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k))\}_{k=1}^K\}_{j=1}^J = B(\mathbb{E}_n^*(Z)-\mathbb{E}_n(Z))$, where $BB' = \Upsilon$ and, conditionally on the sample, $\{\mathbb{E}_n^*(Z)-\mathbb{E}_n(Z)\}$ is the average of a sample of independent observations with mean zero, variance covariance matrix $\hat{V}$ such that $||\hat{V}-\mathbf{I}_\rho|| = O_p(n^{-1/2})$ and finite third absolute moments, a.s..

For every $k = 1,2...,K$, B_k has full rank and so, $\forall i = 1,...,n$, $\exists W_{k,i}^* \in \mathbb{R}^{\rho_k}$ such that $B_k W_{k,i}^* = (Y_i^* - \mathbb{E}_n(Y|x_k))1[X_i^* = x_k]$. For every $i = 1,...,n$, we define $W_i^* = [W_{1,i}^*,...,W_{K,i}^*] \in \mathbb{R}^\rho$. By construction, notice that, $\forall i = 1,2,...,n$, $\mathbb{E}(W_i^*|\mathcal{X}_n) = 0_{\rho\times 1}$.

Finally, $\forall i=1,2,...,n$, we define $Z_i^*=W_i^*$ and, thus, $\mathbb{E}_n(Z)=0_{\rho\times 1}$. By construction, $\{Z_i^*-\mathbb{E}_n(Z)\}_{i=1}^n$ is a sample of random vectors from a distribution with $\mathbb{E}(Z_i^*-\mathbb{E}_n(Z)|\mathcal{X}_n)=0_{\rho\times 1}$ and $V(B(Z_i^*-\mathbb{E}_n(Z))|\mathcal{X}_n)=\hat{\Upsilon}$. The matrix $\hat{\Upsilon}$ satisfies $||\hat{\Upsilon}-\Upsilon||=||B(\hat{V}-\mathbf{I}_\rho)B'||$, where $\hat{V}=V(Z_i^*-\mathbb{E}_n(Z)|\mathcal{X}_n)$. By the CLT, $||\hat{\Upsilon}-\Upsilon||=O_p(n^{-1/2})$ and since $B\in\mathbb{R}^{(JK)\times\rho}$ has rank ρ, it follows that $||\hat{V}-\mathbf{I}_\rho||\leq O_p(n^{-1/2})$. Finally, since $\{(Y_i^*-\mathbb{E}_n(Y|x_k))1[X_i^*=x_k]\}_{k=1}^K$ has finite third moments, a.s., $\{Z_i^*-\mathbb{E}_n(Z)\}$ also has finite third moments, a.s..

Step 3: We now show that $\Gamma_n^*=\tilde{H}(\sqrt{n}(\mathbb{E}_n^*(Z)-\mathbb{E}_n(Z)))+\tilde{\delta}_n^*$ where $\tilde{H}$ is the same function as in theorem A.1 and for any positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\varepsilon_n=O(n^{-1/2})$, $P(|\tilde{\delta}_n^*|>\varepsilon_n|\mathcal{X}_n)=o(n^{-1/2})$, a.s.. By the definitions of Γ_n^* and $\tilde{H}$, it follows that the event $\{\tilde{\delta}_n^*=0\}$ depends exclusively on $\mathcal{X}_n$ and, therefore, $P(\tilde{\delta}_n^*=0|\mathcal{X}_n)=1[\tilde{\delta}_n^*=0]$. Thus, it suffices to show that $\liminf\{\delta_n^*=0\}$, a.s..

Step 3.1: For any arbitrary $S\in\mathcal{P}^{\{\{1,...,J\}\times\{1,...,K\}\}}\backslash\varnothing$, suppose that $\exists\theta_0\in\Theta_I$ that satisfies $\{p_k(\mathbb{E}(Y_j|x_k)-M_j(\theta_0,x_k))=0\}_{(j,k)\in S}$. We show that $\exists N\in\mathbb{N}$ such that, $\forall n\geq N$, $\exists\theta\in\hat{\Theta}_I(\tau_n)$ that satisfies $\{|\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-M_j(\theta,x_k))|\leq\tilde{\tau}_n/\sqrt{n}\}_{(j,k)\in S}$, a.s.. In particular, by the LIL, it follows that,

$$P\left(\liminf\left\{\left\{\theta_0\in\hat{\Theta}_I(\tau_n)\right\}\cap\left\{|\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-M_j(\theta_0,x_k))|\leq\tilde{\tau}_n/\sqrt{n}\right\}_{(j,k)\in S}\right\}\right)=1$$

which is exactly the desired result for $\theta=\theta_0$.

Step 3.2: For any arbitrary $S\in\mathcal{P}^{\{\{1,...,J\}\times\{1,...,K\}\}}\backslash\varnothing$, suppose that $\nexists\theta\in\Theta_I$ that satisfies $\{p_k(\mathbb{E}(Y_j|x_k)-M_j(\theta,x_k))=0\}_{(j,k)\in S}$. In this step, we show that $\exists N\in\mathbb{N}$ such that, $\forall n\geq N$, $\nexists\theta\in\hat{\Theta}_I(\tau_n)$ that satisfies $\{|\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-M_j(\theta_n,x_k))|\leq\tilde{\tau}_n/\sqrt{n}\}_{(j,k)\in S}$, a.s.. Let $D_n(S)$ be defined as,

$$D_n(S)=\left\{\theta\in\Theta:\left\{\bigcap_{(j,k)\in S}\left\{|\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-M_j(\theta,x_k))|\leq\tilde{\tau}_n/\sqrt{n}\right\}\right\}\right\}$$

It then it suffices to show that $\liminf\{\{\hat{\Theta}_I(\tau_n)\cap D_n(S)\}=\varnothing\}$, a.s..

For any $\epsilon\geq 0$, define $\Theta_I(\epsilon)=\{\theta\in\Theta:\{\{p_k(\mathbb{E}(Y_j|x_k)-M_j(\theta,x_k))\leq\epsilon\}_{j=1}^J\}_{k=1}^K\}$. First, we show that if $\nexists\theta\in\Theta_I$ such that $\{p_k(\mathbb{E}(Y_j|x_k)-M_j(\theta,x_k))=0\}_{(j,k)\in S}$, then, $\exists\eta>0$ and $\exists\varpi>0$ such that,

$$\Theta_I(\eta)\subseteq\left\{\theta\in\Theta:\left\{\max_{(j,k)\in S}|p_k(M_j(\theta,x_k)-\mathbb{E}(Y_j|x_k))|\geq\varpi\right\}\right\}\tag{A.1}$$

To show this, notice that the minimization problem $\inf_{\theta\in\Theta_I}\left\{\max_{(j,k)\in S}|p_k(M_j(\theta,x_k)-\mathbb{E}(Y_j|x_k))|\right\}$ achieves a minimum and, by the case under consideration, the minimum cannot be zero. Assign this minimum to $\varpi>0$. As a consequence, $\Theta_I\subseteq\{\theta\in\Theta:\{\max_{(j,k)\in S}|p_k(M_j(\theta,x_k)-\mathbb{E}(Y_j|x_k))|\geq\varpi\}\}$. By a continuity argument, $\exists\eta>0$ such that satisfies equation (A.1).

By elementary properties, $\liminf\{\{\hat{\Theta}_I(\tau_n)\cap D_n(S)\}=\varnothing\}$, a.s. holds if we show that,

$$P\left(\limsup\left\{\left\{\hat{\Theta}_I(\tau_n)\cap D_n(S)\cap\Theta_I(\eta)\right\}\neq\varnothing\right\}\right)=0\tag{A.2}$$

and

$$P\left(\limsup\left\{\left\{\hat{\Theta}_I(\tau_n)\cap\{\Theta_I(\eta)\}^c\right\}\neq\varnothing\right\}\right)=0\tag{A.3}$$

We begin with equation (A.2). By definition of η, it suffices to show that,

$$P\left(\limsup\left\{\left\{D_n(S)\cap\left\{\theta\in\Theta:\max_{(j,k)\in S}|p_k(M_j(\theta,x_k)-\mathbb{E}(Y_j|x_k))|\geq\varpi\right\}\right\}\neq\varnothing\right\}\right)=0$$

To show this, notice that,

$$\begin{aligned}
&\left\{D_n(S)\cap\left\{\theta\in\Theta:\max_{(j,k)\in S}\left|p_k\left(M_j(\theta,x_k)-\mathbb{E}(Y_j|x_k)\right)\right|\geq\varpi\right\}\right\}\\
&\subseteq\bigcup\nolimits_{(j,k)\in S}\left\{\begin{array}{c}\{|\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-M_j(\theta,x_k))|\leq\tilde{\tau}_n/\sqrt{n}\}\\ \cap\{|p_k(M_j(\theta,x_k)-\mathbb{E}(Y_j|x_k))|\geq\varpi\}\end{array}\right\}\\
&\subseteq\bigcup\nolimits_{(j,k)\in S}\left\{\begin{array}{c}\{|\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-\mathbb{E}(Y_j|x_k))|\geq\varpi/2-\tilde{\tau}_n/\sqrt{n}\}\cup\\ \{|p_k-\hat{p}_k|\sup_{\theta\in\Theta}|M_j(\theta,x_k)-\mathbb{E}(Y_j|x_k)|\geq\varpi/2\}\end{array}\right\}
\end{aligned}$$

and so, the result follows from the SLLN.

To show equation (A.3), notice that,

$$\begin{aligned}
&\left\{\hat{\Theta}_I(\tau_n)\cap\{\Theta_I(\eta)\}^c\right\}\\
&=\left\{\begin{array}{c}\left\{\bigcap_{j=1}^{J}\bigcap_{k=1}^{K}\{\theta\in\Theta:\{\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-M_j(\theta,x_k))\leq\tau_n/\sqrt{n}\}\}\right\}\\ \cap\left\{\bigcup_{j=1}^{J}\bigcup_{k=1}^{K}\{\theta\in\Theta:\{p_k(\mathbb{E}(Y_j|x_k)-M_j(\theta,x_k))>\eta\}\}\right\}\end{array}\right\}\\
&=\left\{\bigcup_{j=1}^{J}\bigcup_{k=1}^{K}\left\{\theta\in\Theta:\left\{\begin{array}{c}\{\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-M_j(\theta,x_k))\leq\tau_n/\sqrt{n}\}\\ \cap\{p_k(\mathbb{E}(Y_j|x_k)-M_j(\theta,x_k))>\eta\}\end{array}\right\}\right\}\right\}\\
&\subseteq\left\{\bigcup_{j=1}^{J}\bigcup_{k=1}^{K}\left\{\begin{array}{c}\{\{\hat{p}_k(\mathbb{E}(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k))>\eta/4\}\}\cup\\ \{|p_k-\hat{p}_k|\sup_{\theta\in\Theta}|(\mathbb{E}(Y_j|x_k)-M_j(\theta,x_k))|>\eta/2\}\end{array}\right\}\right\}
\end{aligned}$$

and so, again, the result follows from the SLLN.

Step 3.3: By step 3.1, $\liminf\{\tilde{\delta}_n^*\geq 0\}$, a.s. and by step 3.2, $\liminf\{\tilde{\delta}_n^*\leq 0\}$. Combining both statements, $\liminf\{\tilde{\delta}_n^*=0\}$, a.s., which completes the proof of this part.

Part 3. By definition, $\{\hat{\Theta}_I(\tau_n)=\varnothing\}$ implies $\{\Gamma_n^*=0\}$, and thus, $P(\Gamma_n^*=0|\mathcal{X}_n)\geq 1[\hat{\Theta}_I(\tau_n)=\varnothing]$. By lemma 2.1, if $\Theta_I=\varnothing$, then, $\liminf\{\hat{\Theta}_I(\tau_n)=\varnothing\}$, a.s., completing the proof. ■

The following theorem shows how the results of theorem A.3 change when assumption (CF) is replaced by assumption (CF').

Theorem A.4 *Let ρ denote the rank of the variance covariance matrix of the vector $\{\{1[X=x_k]Y_j\}_{j=1}^J\}_{k=1}^K$. If we assume (B1)-(B4), (CF') and $\Theta_I\neq\varnothing$, then, $\Gamma_n^*=\tilde{H}(\sqrt{n}(\mathbb{E}_n^*(Z)-\mathbb{E}_n(Z)))+\tilde{\delta}_n^*$, where,*

1. *$P(\tilde{\delta}_n^*=0|\mathcal{X}_n)=1[\tilde{\delta}_n^*=0]$ and $\liminf\{\tilde{\delta}_n^*=0\}$, a.s.,*
2. *$\{(\mathbb{E}_n^*(Z)-\mathbb{E}_n(Z))|\mathcal{X}_n\}:\Omega_n\to\mathbb{R}^\rho$ is a zero mean sample average of n independent observations from a distribution with variance covariance matrix $\hat{V}$. Moreover, this distribution has finite third moments, a.s., and $||\hat{V}-\mathbf{I}_\rho||\leq O_p(n^{-1/2})$,*
3. *$\tilde{H}:\mathbb{R}^\rho\to\mathbb{R}$ is the same function as in theorem A.2.*

Proof. This follows from theorems A.2 and A.3. ■

A.4 Consistency in level

This section collects all the results that take us from the representation theorems to the main theorem of bootstrap consistency, theorem 2.1. We begin with a lemma that characterizes the limiting distribution.

Lemma A.3 *Assume (A1)-(A4) and (CF').*

Part 1. If $\Theta_I \neq \varnothing$, then, $\lim_{m\to+\infty} P(\Gamma_m \leq h) = P(H(\zeta) \leq h)$, *where H and ζ are the function and the stochastic process described in theorem A.1.*

Part 2. If $\Theta_I = \varnothing$, then, $P(\Gamma_n \leq h) = 1[h \geq 0]$.

Part 3. $\lim_{m\to+\infty} P(\Gamma_m \leq h)$ *is continuous for all $h \neq 0$.*

Proof. Parts 1 and 2. Both statements follow directly from theorem A.1.

Part 3. If $\Theta_I = \varnothing$, the statement is trivial from part 2. If $\Theta_I \neq \varnothing$, then, by part 1, $\lim_{n\to+\infty} P(\Gamma_n = h)$ $= P(H(\zeta) = h)$. Since $H \geq 0$, we only need to consider $h > 0$. By theorem A.1, H is weakly convex and lower semicontinuous and, therefore, the result follows from theorem 11.1 in Davydov, Lifshits and Smorodina [11] (part (i)). ■

The traditional definition of bootstrap consistency requires the conditional distribution of the bootstrap approximation to converge *uniformly* to the limiting distribution of the statistic of interest (see, for example, Hall [14] or Horowitz [16]). When $\Theta_I \neq \varnothing$, the limiting distribution has a discontinuity at zero and, given this discontinuity, it is possible that our bootstrap approximation fails to converge (pointwise) at zero. To resolve this issue, our strategy will be to exclude the discontinuity point from our goal. Except on an arbitrarily small neighborhood around zero, we show that the bootstrap approximation is consistent. We refer to this result as *bootstrap consistency on any set excluding zero.*

Theorem A.5 (Bootstrap consistency on any set excluding zero) *Assume (A1)-(A4) and (CF').*

Part 1. If $\Theta_I \neq \varnothing$, then, $\forall \mu > 0$ and $\forall \varepsilon > 0$,

$$\lim_{n\to+\infty} P^*\left(\sup_{|h|\geq\mu}\left|P(\Gamma_n^* \leq h|\mathcal{X}_n) - \lim_{m\to+\infty} P(\Gamma_m \leq h)\right| \leq \varepsilon\right) = 1$$

Part 2. If $\Theta_I = \varnothing$, then,

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}}\left|P(\Gamma_n^* \leq h|\mathcal{X}_n) - \lim_{m\to+\infty} P(\Gamma_m \leq h)\right| = 0\right\}\right) = 1$$

Proof. Part 1. We divide the argument into two steps.

Step 1. We begin by showing the pointwise version of the result. In particular, we now show that $\forall\varepsilon > 0$ and $\forall h \neq 0$, $|P(\Gamma_n^* \leq h|\mathcal{X}_n) - \lim_{m\to+\infty} P(\Gamma_m \leq h)| \leq \varepsilon$, w.o.p.a.1.

By theorem A.3, $\{v_n^*(m_\theta)|\mathcal{X}_n\}$ converges weakly to ζ, i.o.p., $\tilde{H}$ is continuous and $P^*(|\delta_n^*| > \varepsilon|\mathcal{X}_n) =$ $o(1)$, a.s.. Therefore, by the continuous mapping theorem and Slutsky's lemma, $\{\Gamma_n^*|\mathcal{X}_n\}$ converges weakly to $H(\zeta)$, i.o.p.. In other words, if h is a continuity point of $P(H(\zeta) \leq h)$, then, $\forall\varepsilon > 0$,

$$\lim_{n\to+\infty} P^*\left(|P(\Gamma_n^* \leq h|\mathcal{X}_n) - P(H(\zeta) \leq h)| \leq \varepsilon\right) = 1$$

By lemma A.3, $P(H(\zeta) \leq h) = \lim_{m\to+\infty} P(\Gamma_m \leq h)$, which is continuous for all $h \neq 0$. This completes the proof of this step.

Step 2. We now show the uniform version of the result. This result follows from using the pointwise convergence in step 1, the continuity of the function $\lim_{m\to+\infty} P(\Gamma_m \leq h)$ $\forall h$ such that $|h| \geq \mu$ and the defining properties of the CDFs.

Part 2. By theorem A.3, $\liminf\{P(\Gamma_n^* \leq h|\mathcal{X}_n) = 1[h \geq 0]\}$, a.s., and by lemma A.3, $1[h \geq 0] =$ $\lim_{m\to+\infty} P(\Gamma_m \leq h)$. The result follows from combining these two findings. ■

In the case when $\Theta_I \neq \varnothing$, we have dealt with the discontinuity at zero by simply excluding the point from the analysis. This raises the following question: is it important to obtain a bootstrap approximation at zero? The answer is negative. By theorem A.1 and lemma A.3, it follows that $\lim_{m\to+\infty} P(\Gamma_m \leq 0) \leq$

0.5. Since the purpose of the approximation is to conduct hypothesis tests, we are typically interested in approximating the 90, 95 and 99 percentiles of the distribution. For all these quantiles, our consistency result holds w.o.p.a.1. This is the content of the following corollary.

Corollary A.1 *Assume (A1)-(A4), (CF') and $\Theta_I \neq \varnothing$. For any $\alpha \in (0, 0.5)$, define $q_n^B(1-\alpha) = P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right)$. Then, $\left|q_n^B(1-\alpha) - (1-\alpha)\right| = o_{p^*}(1)$.*

Proof. By lemma A.3, $\lim_{n\to+\infty} P(\Gamma_n \leq h) = P(H(\zeta) \leq h)$, *where H is the function and ζ* is the stochastic process described in theorem A.1.

By theorem A.1, $H(\zeta) \leq 0$ implies that $\exists (j, \theta_0) \in \{1, ..., J\} \times \Theta_I$ such that $\zeta(\theta_0) \leq 0$. Since $\zeta(\theta_0) \sim N(0, V(m_j(Z, \theta_0)))$ with $V(m_j(Z, \theta_0)) > 0$, then, $P(H(\zeta) \leq 0) \leq P(\zeta(\theta_0) \leq 0) \leq 0.5 < 1-\alpha$.

Let $c_\infty(1-\alpha)$ denote the $(1-\alpha)$ quantile of the limiting distribution. By lemma A.3, $\forall h > 0$, $P(H(\zeta) \leq h)$ is continuous, and so, $\forall \alpha \in (0, 0.5)$, $P(H(\zeta) \leq c_\infty(1-\alpha)) = (1-\alpha)$, which implies that $c_\infty(1-\alpha) > 0$.

By theorem A.5, $\sup_{|h|\geq\mu} |P(\Gamma_n^* \leq h|\mathcal{X}_n) - P(H(\zeta) \leq h)| = o_{p^*}(1)$. For any $\varepsilon/2 > 0$, choose $\mu > 0$ so that $\{c_\infty(1-\alpha+\varepsilon/2) \geq \mu\}$. By the continuity of $P(H(\zeta) \leq h)$, it follows that,

$$\lim_{n\to+\infty} P^*\left((1-\alpha) \leq P(\Gamma_n^* \leq c_\infty(1-\alpha+\varepsilon/2)|\mathcal{X}_n) \leq (1-\alpha)+\varepsilon\right) = 1$$

By definition, $\{(1-\alpha) \leq P(\Gamma_n^* \leq c_\infty(1-\alpha+\varepsilon/2)|\mathcal{X}_n)\}$ implies $\{\hat{c}_n^B(1-\alpha) \leq c_\infty(1-\alpha+\varepsilon/2)\}$ which, in turn, implies $\{(1-\alpha) \leq P(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n) \leq P(\Gamma_n^* \leq c_\infty(1-\alpha+\varepsilon/2)|\mathcal{X}_n)\}$. Therefore,

$$\begin{aligned}
&\lim_{n\to+\infty} P^*\left(\left|P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right) - (1-\alpha)\right| \leq \varepsilon\right) \\
&\geq \lim_{n\to+\infty} P^*\left(\begin{array}{c} (1-\alpha) \leq P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right) \leq \\ P(\Gamma_n^* \leq c_\infty(1-\alpha+\varepsilon/2)|\mathcal{X}_n) \leq (1-\alpha)+\varepsilon \end{array}\right) = 1
\end{aligned}$$

which completes the proof. ■

The final consequence of these results is the main theorem of this section, theorem 2.1, which is formulated in the main text. Even though the formulation of the theorem in the main text imposes assumption (CF), the proof will only make use of assumption (CF').

Proof. [**Theorem 2.1**] Fix $\alpha \in (0, 0.5)$ and consider the following derivation,

$$\left|P\left(\Theta_I \subseteq \hat{C}_n^B(1-\alpha)\right) - (1-\alpha)\right| \leq \left\{\begin{array}{c} \left|P\left(\Gamma_n \leq \hat{c}_n^B(1-\alpha)\right) - P\left(H(\zeta) \leq \hat{c}_n^B(1-\alpha)\right)\right| + \\ \left|P\left(H(\zeta) \leq \hat{c}_n^B(1-\alpha)\right) - P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right)\right| + \\ \left|P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right) - (1-\alpha)\right| \end{array}\right\} \quad \text{(A.4)}$$

The right hand side of equation (A.4) is the sum of three terms. The second term is $o_{p^*}(1)$ by theorem A.5 and the third term is $o_{p^*}(1)$ by corollary A.1. For any $\mu > 0$, the first term on the right hand side of equation (A.4) satisfies,

$$\begin{aligned}
&\left|P\left(\Gamma_n \leq \hat{c}_n^B(1-\alpha)\right) - P\left(H(\zeta) \leq \hat{c}_n^B(1-\alpha)\right)\right| \\
&\leq \sup_{|h|\geq\mu} |P(\Gamma_n \leq h) - P(H(\zeta) \leq h)| + 1\left[\left|\hat{c}_n^B(1-\alpha)\right| < \mu\right]
\end{aligned} \quad \text{(A.5)}$$

The right hand side of equation (A.5) is the sum of two terms. For any $\mu > 0$, theorem A.1 and the arguments used in the proof of theorem A.5 imply that the first term is $o(1)$. To show that the second term is $o_{p^*}(1)$, it suffices to find $\mu > 0$ such that $\{\hat{c}_n^B(1-\alpha) \geq \mu\}$, w.o.p.a.1. By the arguments in corollary A.1, $\forall \alpha \in (0, 0.5)$, $\{c_\infty(1-\alpha) > 0\}$. By lemma A.3, the limiting distribution evaluated at $c_\infty(1-\alpha)$ is equal to $(1-\alpha)$ and it is continuous on the interval $[0, c_\infty(1-\alpha)]$. Then, by intermediate value theorem, $\exists \eta \in (0, c_\infty (1-\alpha))$ such that $P(H(\zeta) \leq \eta) = ((1-\alpha) - 0.5)/2 + 0.5$. We choose $\mu = \eta$. By theorem A.5,

$$\left|P\left(\Gamma_n^* \leq \mu | \mathcal{X}_n\right) - ((1-\alpha) - 0.5)/2 + 0.5\right| = \left|P\left(\Gamma_n^* \leq \mu | \mathcal{X}_n\right) - \lim_{n\to+\infty} P\left(\Gamma_n \leq \mu\right)\right| = o_{p^*}(1)$$

and thus, $\{P(\Gamma_n^* \leq \mu | \mathcal{X}_n) < (1-\alpha)\}$, w.o.p.a.1. By definition of quantile, $\{(1-\alpha) \leq P\left(\Gamma_n^* \leq \hat{c}_n^B (1-\alpha) | \mathcal{X}_n\right)\}$, and so, by the monotonicity of the CDF, $\{\hat{c}_n^B (1-\alpha) \geq \mu\}$, w.o.p.a.1.

As a consequence of our arguments, we deduce that, $\forall \varepsilon > 0$,

$$\lim_{n\to+\infty} P^* \left(\left|P\left(\Theta_I \subseteq \hat{C}_n^B (1-\alpha)\right) - (1-\alpha)\right| \leq \varepsilon\right) = 1$$

Since the event $\{|P\left(\Theta_I \subseteq \hat{C}_n^B (1-\alpha)\right) - (1-\alpha)| \leq \varepsilon\}$ is non-stochastic, then, the statement of the theorem follows as a conclusion. ■

A.4.1 Stepdown control bootstrap procedure

The inferential schemes described in this paper can be used as an ingredient in a stepdown control procedure like the one described in Romano and Shaikh [24]. In this section, we only describe the stepdown control version of our bootstrap procedure, but the same arguments can be applied to any of the inferential schemes developed in this paper.

As an intermediate step, consider the following auxiliary bootstrap procedure,

1. Choose $\{\tau_n\}_{n=1}^{+\infty}$ to be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln \ln n}/\tau_n = o(1)$, a.s.,

2. Repeat the following for $s = 1, 2, ..., S$. Construct bootstrap samples of size n, by sampling randomly with replacement from the data. Denote the bootstrapped observations by $\{Z_i^*\}_{i=1}^n$ and, $\forall j = 1, 2, ..., J$, let $\mathbb{E}_n^* (m_j (Z, \theta)) = n^{-1} \sum_{i=1}^n m_j (Z_i^*, \theta)$. Compute,

$$\Gamma_n^*(K) = \begin{cases} \sup\limits_{\theta \in K} G\left(\left\{\begin{array}{c} \left[\sqrt{n}\left(\mathbb{E}_n^*(m_j(Z,\theta)) - \mathbb{E}_n(m_j(Z,\theta))\right)\right]_+ \\ *1\left[\left|\mathbb{E}_n(m_j(Z,\theta))\right| \leq \tau_n/\sqrt{n}\right] \end{array}\right\}_{j=1}^J\right) & \text{if } K \neq \varnothing \\ 0 & \text{if } K = \varnothing \end{cases}$$

3. Let $\hat{c}_n^{SB}(K, 1-\alpha)$ be the $(1-\alpha)$ quantile of the bootstrapped distribution of $\Gamma_n^*(K)$, approximated with arbitrary accuracy in the previous step.

Some remarks are in order. The superscript SB of $\hat{c}_n^{SB}(K, 1-\alpha)$ refers to the fact that this quantile will be an ingredient in the stepdown control bootstrap procedure. Also, notice that the quantile is a function of the set K, which constitutes the input of the auxiliary procedure. In particular, if we set $K = \hat{\Theta}_I(\tau_n)$, this auxiliary bootstrap scheme is equal to the one described in section 2.2.3. Finally, if the model under consideration satisfies the assumptions of the conditionally separable model, we can define a auxiliary bootstrap procedure that is specialized for this framework, in the same way as we did in section 2.2.3. We prefer not to do this here to avoid repetition.

Following Romano and Shaikh [24], we can define a $(1-\alpha)$ confidence set for the identified set, denoted $\hat{C}_n^{SB}(1-\alpha)$, by using the following stepdown control bootstrap procedure,

1. Let $K_1 = \hat{\Theta}_I(\tau_n)$. If $\sup_{\theta \in K_1} \sqrt{n} Q_n(\theta) \leq \hat{c}_n^{SB}(K_1, 1-\alpha)$, take $\hat{C}_n^{SB}(1-\alpha) = K_1$ and stop; otherwise, set $K_2 = \left\{\theta \in \Theta : \sqrt{n} Q_n(\theta) \leq \hat{c}_n^{SB}(K_1, 1-\alpha)\right\}$ and continue.

2. If $\sup_{\theta \in K_2} \sqrt{n} Q_n(\theta) \leq \hat{c}_n^{SB}(K_2, 1-\alpha)$, take $\hat{C}_n^{SB}(1-\alpha) = K_2$ and stop; otherwise, set $K_3 = \left\{\theta \in \Theta : \sqrt{n} Q_n(\theta) \leq \hat{c}_n^{SB}(K_2, 1-\alpha)\right\}$ and continue.

$\vdots$

j. If $\sup_{\theta \in K_j} \sqrt{n} Q_n(\theta) \leq \hat{c}_n^{SB}(K_j, 1-\alpha)$, take $\hat{C}_n^{SB}(1-\alpha) = K_j$ and stop; otherwise, set $K_{j+1} = \left\{\theta \in \Theta : \sqrt{n} Q_n(\theta) \leq \hat{c}_n^{SB}(K_j, 1-\alpha)\right\}$ and continue.

$\vdots$

In order to analyze the coverage properties of the stepdown control bootstrap procedure, we first establish the following result.

Lemma A.4 *Assume (A1)-(A4), (CF') and that $\Theta_I \neq \varnothing$. Then, $\forall \alpha \in (0, 0.5)$,*

$$\lim_{n \to +\infty} P\left(\sup_{\theta \in \Theta_I} \sqrt{n} Q_n(\theta) > \hat{c}_n^{SB}(\Theta_I, 1-\alpha)\right) = \alpha$$

Furthermore, $\forall K$ such that $\Theta_I \subseteq K$, $\hat{c}_n^{SB}(\Theta_I, 1-\alpha) \leq \hat{c}_n^{B}(K, 1-\alpha)$.

Proof. The first statement of the theorem follows from replacing the set $\hat{\Theta}_I(\tau_n)$ by the set Θ_I in the proof of theorem 2.1. The second statement follows from the definition of $\hat{c}_n^{SB}(K, 1-\alpha)$. ■

The next lemma shows that the confidence sets constructed using the stepdown control procedure are always contained in the confidence sets constructed using the procedure described in section 2.2.3.

Lemma A.5 *For any $\alpha \in (0,1)$,*

$$\hat{C}_n^{SB}(1-\alpha) \subseteq \hat{C}_n^{B}(1-\alpha)$$

Proof. By definition of the stepdown control bootstrap procedure, $K_1 = \hat{\Theta}_I(\tau_n)$, $\hat{c}_n^{SB}(K_1, 1-\alpha) = \hat{c}_n^{B}(1-\alpha)$ and $K_2 = \hat{C}_n^{B}(1-\alpha)$. Suppose that the stepdown procedure stops in step j^*. If $j^* = 1$, then, by definition, $\hat{C}_n^{SB}(1-\alpha) = K_1 = \hat{\Theta}_I(\tau_n)$ and since the procedure stopped in the first step, $\sup_{\theta \in K_1} \sqrt{n} Q_n(\theta) \leq \hat{c}_n^{SB}(K_1, 1-\alpha)$, which implies that $\hat{C}_n^{SB}(1-\alpha) = \hat{\Theta}_I(\tau_n) \subseteq \hat{C}_n^{B}(1-\alpha)$. If $j^* = 2$, then, $\hat{C}_n^{SB}(1-\alpha) = K_2 = \hat{C}_n^{B}(1-\alpha)$. If $j^* > 2$, then, by definition, $\forall i \in \{2, .., j^*-1\}$, $K_{i+1} = \{\theta \in \Theta : \sqrt{n} Q_n(\theta) \leq \hat{c}_n^{SB}(K_i, 1-\alpha)\}$ and $\sup_{\theta \in K_{i+1}} \sqrt{n} Q_n(\theta) > \hat{c}_n^{SB}(K_{i+1}, 1-\alpha)$. By combining these two, we deduce that, $\forall i \in \{2, .., j^*-1\}$, $\hat{c}_n^{SB}(K_{i+1}, 1-\alpha) < \hat{c}_n^{SB}(K_i, 1-\alpha)$, which, in turn, implies that $K_i \subseteq K_{i-1}$ and, as a consequence, $\hat{C}_n^{SB}(1-\alpha) \subseteq \hat{C}_n^{B}(1-\alpha)$. ■

Based on the previous results, we are ready to establish the consistency in level of the stepdown control bootstrap procedure.

Theorem A.6 *Assume (A1)-(A4), (CF') and that $\Theta_I \neq \varnothing$. Then, $\forall \alpha \in (0, 0.5)$,*

$$\lim_{n \to +\infty} P\left(\Theta_I \subseteq \hat{C}_n^{SB}(1-\alpha)\right) = 1-\alpha$$

Proof. By theorem 2.1 in Romano and Shaikh [24] and lemma A.4, it follows that,

$$\liminf_{n \to +\infty} P\left(\Theta_I \subseteq \hat{C}_n^{SB}(1-\alpha)\right) \geq 1-\alpha$$

By lemma A.5, $\hat{C}_n^{SB}(1-\alpha) \subseteq \hat{C}_n^{B}(1-\alpha)$, and so,

$$\limsup_{n \to +\infty} P\left(\Theta_I \subseteq \hat{C}_n^{SB}(1-\alpha)\right) \leq \limsup_{n \to +\infty} P\left(\Theta_I \subseteq \hat{C}_n^{B}(1-\alpha)\right)$$

By theorem 2.1, the right hand side is equal to $(1-\alpha)$, completing the proof. ■

A.5 Rates of convergence results

This section collects all the results that take us from the representation theorems to the main theorem of rates of convergence of the error in the coverage probability for the bootstrap approximation.

A.5.1 Results under assumption (CF)

The following lemma shows a useful property of the function described in theorem A.1.

Lemma A.6 *Part 1. Let $\tilde{H}$ be the function in theorem A.1, let $\xi \sim N(0, \Xi)$ where $\Xi \in \mathbb{R}^{\rho \times \rho}$ is non-singular and let $\{\varepsilon_n\}_{n=1}^{+\infty}$ be a positive sequence with $\varepsilon_n = o(1)$. Then, $\forall \mu > 0$,*

$$\sup_{|h| \geq \mu} \left| P\left(\tilde{H}(\xi) \in (h - \varepsilon_n, h + \varepsilon_n]\right) \right| \leq O(\varepsilon_n)$$

Part 2. Let $\tilde{H}$ be the function in theorem A.1, let $\{\xi_n | \mathcal{X}_n\} \sim N(0, \Xi_n)$ where $\Xi_n \in \mathbb{R}^{\rho \times \rho}$ is conditionally non-stochastic and non-singular w.p.a.1, and let $\{\varepsilon_n\}_{n=1}^{+\infty}$ be a positive sequence with $\varepsilon_n = o(1)$. Then, $\forall \mu > 0$,

$$\sup_{|h| \geq \mu} \left| P\left(\tilde{H}(\xi_n) \in (h - \varepsilon_n, h + \varepsilon_n] \middle| \mathcal{X}_n\right) \right| \leq O_p(\varepsilon_n)$$

Proof. Part 1. First, consider h such that $h \leq -\mu$. Since $\tilde{H}(\xi_n) \geq 0$ and $\varepsilon_n = o(1)$ then, eventually, $h + \varepsilon_n < 0$ and so $P(\tilde{H}(\xi_n) \leq h + \varepsilon_n) = 0$.

Next, consider h such that $h \geq \mu$. Since $\varepsilon_n = o(1)$, then, eventually, $h - \varepsilon_n > 0$. Since $h > 0$ and $\varepsilon_n = o(1)$ then, by theorem A.1, $\tilde{H}^{-1}((h - \varepsilon_n, h + \varepsilon_n]) \subseteq \{\tilde{H}^{-1}(\{h\})\}^{\gamma_n}$ for $\gamma_n = O(\varepsilon_n)$. By the submultiplicative property of the matrix norm $\Xi^{-1/2}\{\tilde{H}^{-1}(\{h\})\}^{\gamma_n} \subseteq \{\Xi^{-1/2}\tilde{H}^{-1}(\{h\})\}^{\eta_n}$ for $\eta_n = O(\varepsilon_n)$.

By theorem A.1, $\tilde{H}$ is continuous and weakly quasiconvex and so, $\tilde{H}^{-1}(\{h\}) = \partial \tilde{H}^{-1}((-\infty, h])$, where $\tilde{H}^{-1}((-\infty, h]) \in \mathcal{C}_\rho$. Using the submultiplicative property, $\Xi^{-1/2}\partial\tilde{H}^{-1}((-\infty, h]) = \partial\Xi^{-1/2}\tilde{H}^{-1}((-\infty, h])$, where $\Xi^{-1/2}\tilde{H}^{-1}((-\infty, h]) \in \mathcal{C}_\rho$. Combining all these steps, we deduce that,

$$P\left(\tilde{H}(\xi) \in (h - \varepsilon_n, h + \varepsilon_n]\right) \leq \Phi_{\mathbf{I}_\rho}\left(\left\{\partial\left[\Xi^{-1/2}\tilde{H}^{-1}((-\infty, h])\right]\right\}^{\eta_n}\right)$$

where $\Xi^{-1/2}\tilde{H}^{-1}((-\infty, h]) \in \mathcal{C}_\rho$. The right hand side is $O(\varepsilon_n)$ by corollary 3.2 in Bhattacharya and Rao [6] (with $s = 0$).

Part 2. Let $A_n = \{\Xi_n$ is non-singular$\}$ and let $\tilde{\Xi}_n = \Xi_n 1[A_n] + \mathbf{I}_{J \times J} 1[\{A_n\}^c]$. By the arguments in the previous step,

$$P\left(\tilde{H}(\xi_n) \in (h - \varepsilon_n, h + \varepsilon_n] | \mathcal{X}_n\right) \leq \Phi_{\mathbf{I}_\rho}\left(\left\{\partial\left[\tilde{\Xi}_n^{-1/2}\tilde{H}^{-1}((-\infty, h])\right]\right\}^{\eta_n}\right) 1[A_n] + 1[\{A_n\}^c]$$

where $\tilde{\Xi}_n^{-1/2}\tilde{H}^{-1}((-\infty, h]) \in \mathcal{C}_\rho$. The right hand side is a sum of two terms. The first term is $O(\varepsilon_n)$ by corollary 3.2 in Bhattacharya and Rao [6] (with $s = 0$) and the second term is $O_p(\varepsilon_n)$ because Ξ_n is non-singular w.p.a.1. ■

The next theorem provides the rate of convergence of the bootstrap approximation, which is one of the ingredients to obtain the rate of convergence of the error in the coverage probability.

Theorem A.7 (Rate of convergence - bootstrap approximation) *Assume (B1)-(B4) and (CF) and choose the bootstrap procedure to be the one specialized for the conditionally separable model.*

Part 1. *If $\Theta_I \neq \varnothing$, then, $\forall \mu > 0$,*

$$\sup_{|h|\geq\mu} |P(\Gamma_n^* \leq h|\mathcal{X}_n) - P(\Gamma_n \leq h)| = O_p(n^{-1/2})$$

Part 2. *If $\Theta_I = \varnothing$, then,*

$$P\left(\liminf \left\{\sup_{h\in\mathbb{R}} |P(\Gamma_n^* \leq h|\mathcal{X}_n) - P(\Gamma_n \leq h)| = 0\right\}\right) = 1$$

Proof. Part 1. Fix $\mu > 0$ arbitrarily and consider the following argument $\forall h$ such that $|h| \geq \mu$,

$$\sup_{|h|\geq\mu} |P(\Gamma_n^* \leq h|\mathcal{X}_n) - P(\Gamma_n \leq h)| \leq \left\{\begin{array}{c} \sup_{|h|\geq\mu} \left|P(\Gamma_n^* \leq h|\mathcal{X}_n) - P\left(\tilde{H}\left(\hat{\vartheta}\right) \leq h|\mathcal{X}_n\right)\right| \\ +\sup_{|h|\geq\mu} \left|P\left(\tilde{H}\left(\hat{\vartheta}\right) \leq h|\mathcal{X}_n\right) - P\left(\tilde{H}(\vartheta) \leq h\right)\right| \\ +\sup_{|h|\geq\mu} \left|P\left(\tilde{H}(\vartheta) \leq h\right) - P(\Gamma_n \leq h)\right| \end{array}\right\}$$

The right hand side is a sum of three terms. We show that each of them is $O_p(n^{-1/2})$.

Step 1: In this step, we show that $\sup_{|h|\geq\mu} |P(\Gamma_n \leq h) - P(\tilde{H}(\vartheta) \leq h)| \leq O(n^{-1/2})$, where $\vartheta \sim N(0, \mathbf{I}_\rho)$. For $h \leq -\mu$, the statement holds since both Γ_n and $\tilde{H}(\vartheta)$ are non-negative. For $h \geq \mu$ and for any positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\varepsilon_n = O(n^{-1/2})$, theorem A.1 leads to the following derivation,

$$\begin{aligned} &\sup_{|h|\geq\mu} \left\{P(\Gamma_n \leq h) - P\left(\tilde{H}(\vartheta) \leq h\right)\right\} \\ &\leq \sup_{|h|\geq\mu} \left\{\begin{array}{c} P\left(\tilde{H}(\sqrt{n}(\mathbb{E}_n(Z) - \mathbb{E}(Z))) \leq h + \varepsilon_n\right) - P\left(\tilde{H}(\vartheta) \leq h + \varepsilon_n\right) \\ +P\left(\tilde{H}(\vartheta) \in (h - \varepsilon_n, h + \varepsilon_n]\right) + P\left(\left|\tilde{\delta}_n\right| > \varepsilon_n\right) \end{array}\right\} \\ &\leq \left\{\begin{array}{c} \sup_{|h|\geq\mu} \left|P\left(\sqrt{n}(\mathbb{E}_n(Z) - \mathbb{E}(Z)) \in \tilde{H}^{-1}((-\infty, h + \varepsilon_n])\right) - \Phi_{\mathbf{I}_\rho}\left(\tilde{H}^{-1}((-\infty, h + \varepsilon_n])\right)\right| \\ +\sup_{|h|\geq\mu} P\left(\tilde{H}(\vartheta) \in (h - \varepsilon_n, h + \varepsilon_n]\right) + P\left(\left|\tilde{\delta}_n\right| > \varepsilon_n\right) \end{array}\right\} \\ &\leq \left\{\begin{array}{c} \sup_{A\in\mathcal{C}_\rho} \left|P(\sqrt{n}(\mathbb{E}_n(Z) - \mathbb{E}(Z)) \in A) - \Phi_{\mathbf{I}_\rho}(A)\right| + \\ \sup_{|h|\geq\mu} P\left(\tilde{H}(\vartheta) \in (h - \varepsilon_n, h + \varepsilon_n]\right) + P\left(\left|\tilde{\delta}_n\right| > \varepsilon_n\right) \end{array}\right\} \end{aligned}$$

The right hand side is a sum of three terms. By the Berry-Esseén theorem, the first term is $O(n^{-1/2})$, by lemma A.6, the second term is $O(\varepsilon_n) = O(n^{-1/2})$ and by theorem A.1, the last term is $o(n^{-1/2})$. If we combine this result with the analogous argument for $P(\Gamma_n > h)$ (instead of $P(\Gamma_n \leq h)$), we complete this step.

Step 2: We now show that $\sup_{|h|\geq\mu} |P(\Gamma_n^* \leq h|\mathcal{X}_n) - P(\tilde{H}(\hat{\vartheta}) \leq h|\mathcal{X}_n)| \leq O_p(n^{-1/2})$, where $\hat{\vartheta} \sim N(0, \hat{V})$ and $\hat{V}$ is the sample variance of $\{Z_i\}_{i=1}^n$. For $h \leq -\mu$, the statement holds since both Γ_n^* and $\tilde{H}(\hat{\vartheta})$ are non-negative. For $h \geq \mu$ and for any positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\varepsilon_n = O(n^{-1/2})$, theorem A.3 leads to the following derivation,

$$\begin{aligned} &\sup_{|h|\geq\mu} \left\{P(\Gamma_n^* \leq h|\mathcal{X}_n) - P\left(\tilde{H}\left(\hat{\vartheta}\right) \leq h|\mathcal{X}_n\right)\right\} \leq \\ &\leq \left\{\begin{array}{c} \sup_{A\in\mathcal{C}_\rho} \left|P(\sqrt{n}(\mathbb{E}_n^*(Z) - \mathbb{E}_n(Z)) \in A|\mathcal{X}_n) - \Phi_{\hat{V}}(A)\right| + \\ \sup_{|h|\geq\mu} P\left(\tilde{H}\left(\hat{\vartheta}\right) \in (h - \varepsilon_n, h + \varepsilon_n]\middle|\mathcal{X}_n\right) + P\left(\left|\tilde{\delta}_n^*\right| > \varepsilon_n\middle|\mathcal{X}_n\right) \end{array}\right\} \end{aligned}$$

The right hand side is a sum of three terms. Conditional on $\mathcal{X}_n$ and on the design, $\{\mathbb{E}_n^*(Z) - \mathbb{E}_n(Z)\}$ is the average of independent observations with mean zero, variance covariance matrix $\hat{V}$ and finite third moments, w.p.a.1. Thus, the Berry-Esseén theorem implies that the first term is $O_p(n^{-1/2})$. Conditionally on $\mathcal{X}_n$, $\hat{V}$ is non-stochastic and by the SLLN, $\hat{V}$ is non-singular, w.p.a.1. Thus, by lemma A.6, the second term is $O_p(n^{-1/2})$. By theorem A.3, the last term is $o_p(n^{-1/2})$. We combine this with the same argument for $P\left(\Gamma_n^* > h|\mathcal{X}_n\right)$ (instead of $P\left(\Gamma_n^* \leq h|\mathcal{X}_n\right)$) to complete the step.

Step 3: We now show that $\sup_{|h|\geq\mu}|P(\tilde{H}\left(\vartheta\right) \leq h) - P(\tilde{H}(\hat{\vartheta}) \leq h|\mathcal{X}_n)| = O_p(n^{-1/2})$, for $\vartheta \sim N\left(0, \mathbf{I}_\rho\right)$, $\hat{\vartheta} \sim N(0, \hat{V})$ and $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$. It suffices to show that $\int_{\mathbb{R}^\rho}|\phi_{\hat{V}}\left(x\right) - \phi_{\mathbf{I}_\rho}\left(x\right)|dx = O_p(n^{-1/2})$, which follows from simple arguments.

Part 2. By theorem A.1, $\Gamma_n = 0$, or, equivalently, $P\left(\Gamma_n \leq h\right) = 1\left[h \geq 0\right]$. By theorem A.3, $\lim\inf\{P(\Gamma_n^* = 0|\mathcal{X}_n) = 1\}$, a.s., or, equivalently, $\lim\inf\{\sup_{h\in\mathbb{R}}|P(\Gamma_n^* \leq h|\mathcal{X}_n) - 1\left[h \geq 0\right]| = 0\}$, a.s.. The combination of these two statements implies the result. ■

Corollary A.2 *Assume (B1)-(B4), (CF), $\Theta_I \neq \varnothing$ and choose the bootstrap procedure to be the one specialized for the conditionally separable model. For any $\alpha \in (0, 0.5)$, let $q_n^B\left(1-\alpha\right) = P\left(\Gamma_n^* \leq \hat{c}_n^B\left(1-\alpha\right)|\mathcal{X}_n\right)$. Then, $\left|q_n^B\left(1-\alpha\right) - \left(1-\alpha\right)\right| \leq O_p(n^{-1/2})$.*

Proof. Let $c_\infty\left(1-\alpha\right)$ denote the $(1-\alpha)$ quantile of the limiting distribution. By arguments in corollary A.1, $c_\infty\left(1-\alpha\right) > 0$. By theorem A.7, $\forall\mu > 0$ and $\forall\gamma > 0$, $\exists K < +\infty$ such that, $\forall n \in \mathbb{N}$,

$$P\left(\sup_{|h|\geq\mu}|P\left(\Gamma_n^* \leq h|\mathcal{X}_n\right) - P\left(H\left(\vartheta\right) \leq h\right)| \leq Kn^{-1/2}\right) \geq 1-\gamma$$

where $\vartheta \sim N\left(0, \mathbf{I}_\rho\right)$. Choose $\mu > 0$ so that $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\{c_\infty(1-\alpha+Kn^{-1/2}) > \mu\}$. As a consequence, $\forall n \geq N$,

$$P\left(\left|P\left(\Gamma_n^* \leq c_\infty\left(1-\alpha+Kn^{-1/2}\right)\middle|\mathcal{X}_n\right) - P\left(H\left(\vartheta\right) \leq c_\infty\left(1-\alpha+Kn^{-1/2}\right)\right)\right| \leq Kn^{-1/2}\right) \geq 1-\gamma$$

By the continuity of $P\left(H\left(\vartheta\right) \leq h\right)$, $P(H\left(\vartheta\right) \leq c_\infty(1-\alpha+Kn^{-1/2})) = 1-\alpha+Kn^{-1/2}$, so that $\forall n \geq N$,

$$P\left(\left(1-\alpha\right) \leq P\left(\Gamma_n^* \leq c_\infty\left(1-\alpha+Kn^{-1/2}\right)\middle|\mathcal{X}_n\right) \leq \left(1-\alpha\right) + 2Kn^{-1/2}\right) \geq 1-\gamma$$

The event $\{(1-\alpha) \leq P(\Gamma_n^* \leq c_\infty(1-\alpha+Kn^{-1/2})|\mathcal{X}_n)\}$ implies $\{\hat{c}_n^B\left(1-\alpha\right) \leq c_\infty(1-\alpha+Kn^{-1/2})\}$ which, in turn, implies $\{(1-\alpha) \leq P(\Gamma_n^* \leq \hat{c}_n^B\left(1-\alpha\right)|\mathcal{X}_n) \leq P\left(\Gamma_n^* \leq c_\infty(1-\alpha+Kn^{-1/2})|\mathcal{X}_n\right)\}$. Therefore, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$,

$$\begin{aligned}
&P\left(\left|P\left(\Gamma_n^* \leq \hat{c}_n^B\left(1-\alpha\right)|\mathcal{X}_n\right) - \left(1-\alpha\right)\right| \leq 2Kn^{-1/2}\right) \\
&\geq P\left(\left(1-\alpha\right) \leq P\left(\Gamma_n^* \leq \hat{c}_n^B\left(1-\alpha\right)|\mathcal{X}_n\right) \leq \left(1-\alpha\right) + 2Kn^{-1/2}\right) \\
&\geq P\left(\left(1-\alpha\right) \leq P\left(\Gamma_n^* \leq c_\infty\left(1-\alpha+Kn^{-1/2}\right)\middle|\mathcal{X}_n\right) \leq \left(1-\alpha\right) + 2Kn^{-1/2}\right) \geq 1-\gamma
\end{aligned}$$

This conclusion can be extended $\forall n \in \mathbb{N}$ by an appropriate choice of K. ■

The final consequence of these results is the main theorem of this section, theorem 2.2, which is formulated in the main text.

Proof. [**Theorem 2.2**] For any $K > 0$, $\mu > 0$ and $n \in \mathbb{N}$, consider the following derivation,

$$\begin{aligned}
&\left\{\left|P\left(\Theta_I \subseteq \hat{C}_n^B(1-\alpha)\right) - (1-\alpha)\right| > Kn^{-1/2}\right\} \\
&= \left\{\left|P\left(\Gamma_n \leq \hat{c}_n^B(1-\alpha)\right) - (1-\alpha)\right| > Kn^{-1/2}\right\} \\
&\subseteq \left\{\begin{array}{c} \left\{\left|P\left(\Gamma_n \leq \hat{c}_n^B(1-\alpha)\right) - P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right)\right| > (K/2)n^{-1/2}\right\} \\ \cup\left\{\left|P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right) - (1-\alpha)\right| > (K/2)n^{-1/2}\right\} \end{array}\right\} \\
&\subseteq \left\{\begin{array}{c} \sup_{|h|\geq\mu}\left\{\left|P\left(\Gamma_n \leq \mu\right) - P\left(\Gamma_n^* \leq \mu|\mathcal{X}_n\right)\right| > (K/2)n^{-1/2}\right\} \cup \left\{\hat{c}_n^B(1-\alpha) < \mu\right\} \\ \cup\left\{\left|P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right) - (1-\alpha)\right| > (K/2)n^{-1/2}\right\} \end{array}\right\}
\end{aligned}$$

Therefore, $\forall K > 0$, $\forall \mu > 0$ and $\forall n \in \mathbb{N}$,

$$\begin{aligned}
&P\left(\left|P\left(\Theta_I \subseteq \hat{C}_n^B(1-\alpha)\right) - (1-\alpha)\right| > Kn^{-1/2}\right) \\
&\leq \left\{\begin{array}{c} P\left(\sup_{|h|\geq\mu}\left|P\left(\Gamma_n \leq \mu\right) - P\left(\Gamma_n^* \leq \mu|\mathcal{X}_n\right)\right| > (K/2)n^{-1/2}\right) + P\left(\hat{c}_n^B(1-\alpha) < \mu\right) \\ +P\left(\left|P\left(\Gamma_n^* \leq \hat{c}_n^B(1-\alpha)|\mathcal{X}_n\right) - (1-\alpha)\right| > (K/2)n^{-1/2}\right) \end{array}\right\}
\end{aligned}$$

The right hand side is a sum of three terms. For any arbitrary $\varepsilon > 0$, we now show that $\exists K > 0$ and $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, each of the terms in the right hand side is less than $\varepsilon/3$.

By theorem A.7, $\exists K > 0$ such that, $\forall \mu > 0$ and $\forall n \in \mathbb{N}$, the first term is less than $\varepsilon/3$. By arguments in corollary 2.1, $\{\hat{c}_n^B(1-\alpha) \geq \mu\}$, w.p.a.1 and so, $\forall \mu > 0$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, the second term is less than $\varepsilon/3$. By corollary A.2, $\exists K > 0$ such that, $\forall n \in \mathbb{N}$, the third term is less than $\varepsilon/3$.

As a consequence, $\forall \varepsilon > 0$, $\exists K > 0$ and $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$,

$$P\left(\left|P\left(\Theta_I \subseteq \hat{C}_n^B(1-\alpha)\right) - (1-\alpha)\right| \leq Kn^{-1/2}\right) \geq 1-\varepsilon$$

This conclusion can be extended $\forall n \in \mathbb{N}$ by an appropriate choice of K.

To conclude, since the event $\{|P(\Theta_I \subseteq \hat{C}_n^B(1-\alpha)) - (1-\alpha)| \leq Kn^{-1/2}\}$ is non-stochastic, the previous derivation implies that the event must always occur. This completes the proof. ■

Our confidence sets also exhibit desirable coverage properties when $\Theta_I = \varnothing$. By construction, the smallest possible confidence set that could be constructed using the criterion function approach is $\hat{\Theta}_I(0)$. The following lemma shows that if $\Theta_I = \varnothing$, then the confidence set will eventually coincide with $\hat{\Theta}_I(0)$, a.s..

Lemma A.7 *Assume (A1)-(A4), (CF') and $\Theta_I = \varnothing$. Then, $\forall \alpha \in (0,1)$,*

$$P\left(\liminf\left\{\hat{C}_n^B(1-\alpha) = \hat{\Theta}_I(0)\right\}\right) = 1$$

Proof. By theorem A.3, $\liminf\{\Gamma_n^* = 0\}$, a.s., or equivalently, $\forall \alpha \in (0,1)$, $\liminf\left\{\hat{c}_n^B(1-\alpha) = 0\right\}$, a.s., completing the proof. ■

A.5.2 Results under assumption (CF')

In this subsection, we show how the results on the rates of convergence change if we replace assumption (CF) with assumption (CF').

Lemma A.8 *Part 1. Let $\tilde{H}$ be the function in theorem A.1 and assume that $\xi \sim N(0,\Xi)$ with non-singular $\Xi \in \mathbb{R}^{\rho\times\rho}$. Then, $\forall \mu > 0$,*

$$\sup_{|h|\geq\mu} P(\tilde{H}(\xi) \in (h-n^{-1/2}, h+n^{-1/2}]) \leq O(n^{-1/2}(\ln n)^{1/2})$$

Part 2. *Let $\tilde{H}$ be the function in theorem A.1, let $\{\xi_n|\mathcal{X}_n\} \sim N(0,\Xi_n)$ where $\Xi_n \in \mathbb{R}^{\rho\times\rho}$ is conditionally non-stochastic and non-singular w.p.a.1. Then, $\forall\mu > 0$,*

$$\sup_{|h|\geq\mu}\left|P\left(\tilde{H}(\xi_n) \in (h-\varepsilon_n, h+\varepsilon_n]\,|\mathcal{X}_n\right)\right| \leq O_p(n^{-1/2}(\ln n)^{1/2})$$

Proof. Part 1. Consider the following derivation for $\varepsilon_n = n^{-1/2}$,

$$\begin{aligned}\sup_{|h|\geq\mu} P\left(\tilde{H}(\xi) \in (h-\varepsilon_n, h+\varepsilon_n]\right) &= \sup_{|h|\geq\mu} P\left(\vartheta \in \Xi^{-1}\tilde{H}^{-1}\left(\{h\}^{\varepsilon_n}\right)\right) \\ &\leq \left\{\begin{array}{c} \sup_{|h|\geq\mu} P\left(\vartheta \in \left(\Xi^{-1}\tilde{H}^{-1}(\{h\})\right)^{O\left(\varepsilon_n\sqrt{g_n}\right)}\right) \\ +P\left(\|\vartheta\| > O\left(\sqrt{g_n}\right)\right)\end{array}\right\}\end{aligned}$$

where $\vartheta \sim N(0,\mathbf{I}_\rho)$. Choose $g_n = \ln\left(n^{(1+\gamma)}\right)$ for some $\gamma > 0$. By theorem A.2 and corollary 3.2 in Bhattacharya and Rao [6], the first term on the right side is $O\left(\varepsilon_n\sqrt{g_n}\right)$. By theorem 1 in Hüsler, Liu and Singh [15], $P(\|\vartheta\| > O(\sqrt{g_n})) = o(\varepsilon_n\sqrt{g_n})$. Since $\varepsilon_n\sqrt{g_n} = O(n^{-1/2}(\ln n)^{1/2})$, the proof is completed.

Part 2. This follows from part 1 by using the same arguments as in lemma A.3. ■

The next theorem provides rates of convergence of the error in the coverage probability under assumption (CF').

Theorem A.8 *Assume (B1)-(B4) and (CF') and choose the bootstrap procedure to be the one specialized for the conditionally separable model. If $\Theta_I \neq \varnothing$, then, $\forall\alpha \in (0, 0.5)$,*

$$\left|P\left(\Theta_I \subseteq \hat{C}_n^B(1-\alpha)\right) - (1-\alpha)\right| = O\left(n^{-1/2}(\ln n)^{1/2}\right)$$

Proof. Follows from arguments used to prove theorem 2.2 and lemma A.8. ■

A.6 Alternative procedures

A.6.1 Subsampling

We consider two subsampling procedures. The first procedure is, essentially, the subsampling version of the bootstrap procedure proposed in section 2.2.3 and is referred to as subsampling 1. The second procedure is similar to the one proposed by CHT [10] and is referred to as subsampling 2.

Subsampling 1 The procedure is as follows,

1. Choose $\{b_n\}_{n=1}^{+\infty}$ to be a positive sequence such that $b_n \to +\infty$ and $b_n/n = o(1)$ and choose $\{\tau_n\}_{n=1}^{+\infty}$ to be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln\ln n}/\tau_n = o(1)$, a.s.,
2. Estimate the identified set with $\hat{\Theta}_I(\tau_n) = \left\{\theta \in \Theta : \{\mathbb{E}_n(m_j(Z,\theta)) \leq \tau_n/\sqrt{n}\}_{j=1}^J\right\}$,
3. Repeat the following step for $s = 1,2,...,S$. Construct a subsample of size b_n by sampling randomly without replacement from the data. Denote these observations by $\{Z_i^{SS}\}_{i=1}^{b_n}$ and, for every $j = 1,2,...,J$, let $\mathbb{E}_{b_n,n}^{SS}(m_j(Z,\theta)) = b_n^{-1}\sum_{i=1}^{b_n} m_j(Z_i^{SS},\theta)$. Compute,

$$\Gamma_{b_n,n}^{SS_1} = \begin{cases} \sup_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\left\{\begin{array}{c}\left[\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(m_j(Z,\theta)) - \mathbb{E}_n(m_j(Z,\theta))\right)\right]_+ \\ *1\left[|\mathbb{E}_n(m_j(Z,\theta))| \leq \tau_n/\sqrt{n}\right]\end{array}\right\}_{j=1}^J\right) & \text{if } \hat{\Theta}_I(\tau_n) \neq \varnothing \\ 0 & \text{if } \hat{\Theta}_I(\tau_n) = \varnothing\end{cases}$$

4. Let $\hat{c}_{b_n,n}^{SS_1}(1-\alpha)$ be the $(1-\alpha)$ quantile of the distribution of $\Gamma_{b_n,n}^{SS_1}$, simulated with arbitrary accuracy in the previous step. The $(1-\alpha)$ confidence set for the identified set is given by $\hat{C}_{b_n,n}^{SS_1}(1-\alpha)=\{\theta\in\Theta:\sqrt{n}Q_n(\theta)\leq\hat{c}_{b_n,n}^{SS_1}(1-\alpha)\}$.

If the model is conditionally separable, we can consider a subsampling procedure specialized for this framework. In this case, the expression for $\Gamma_{b_n,n}^{SS_1}$ in step 3 would be replaced by,

$$\Gamma_{b_n,n}^{SS_1}=\begin{cases}\sup_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\left\{\left\{\begin{array}{c}\left[\sqrt{b_n}\hat{p}_k^{SS}\left(\mathbb{E}_{b_n,n}^{SS}(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k)\right)\right]_+\\ *1\left[\left|\hat{p}_k\left(\mathbb{E}_n(Y_j|x_k)-M_{j,k}(\theta)\right)\right|\leq\tau_n/\sqrt{n}\right]\end{array}\right\}_{j=1}^J\right\}_{k=1}^K\right) & \text{if } \hat{\Theta}_I(\tau_n)\neq\varnothing\\ 0 & \text{if } \hat{\Theta}_I(\tau_n)=\varnothing\end{cases}$$

where, $\forall(j,k)\in\{1,...,K\}\times\{1,...,J\}$, we define $\hat{p}_k^{SS}=b_n^{-1}\sum_{i=1}^{b_n}1\left[X_i^{SS}=x_k\right]$ and $\mathbb{E}_{b_n,n}^{SS}(Y_j|x_k)=\left(\hat{p}_k^{SS}b_n\right)^{-1}\sum_{i=1}^{b_n}Y_{j,i}^{SS}1\left[X_i^{SS}=x_k\right]$.

The following result is the representation result for this subsampling procedure.

Theorem A.9 *Part 1. Assume (A1)-(A4), (CF') and $\Theta_I\neq\varnothing$. Then, $\Gamma_{b_n,n}^{SS_1}=H(v_{b_n,n}^{SS}(m_\theta))+\delta_{b_n,n}^{SS_1}$, where,*

1. *for any $\varepsilon>0$, $\lim_{n\to+\infty}P^*(|\delta_{b_n,n}^{SS_1}|>\varepsilon|\mathcal{X}_n)=0$, a.s.,*
2. *$\{v_{b_n,n}^{SS}(m_\theta)|\mathcal{X}_n\}:\Omega_n\to l_J^\infty(\Theta)$ is an empirical process that converges weakly to the same Gaussian process as in theorem A.1, i.o.p.,*
3. *$H:l_J^\infty(\Theta)\to\mathbb{R}$ is the same function as in theorem A.1.*

Part 2. Let ρ denote the rank of the variance covariance matrix of the vector $\{\{1[X=x_k]Y_j\}_{j=1}^J\}_{k=1}^K$. If we assume (B1)-(B4), (CF), $\Theta_I\neq\varnothing$, and we choose the subsampling procedure to be the one specialized for the conditionally separable model, then, $\Gamma_{b_n,n}^{SS_1}=\tilde{H}(\sqrt{b_n}(\mathbb{E}_{b_n,n}^{SS}(Z)-\mathbb{E}_n(Z)))+\tilde{\delta}_{b_n,n}^{SS_1}$, where,

1. *$P(\tilde{\delta}_{b_n,n}^{SS_1}=0|\mathcal{X}_n)=1[\tilde{\delta}_{b_n,n}^{SS_1}=0]$ and $\liminf\{\tilde{\delta}_{b_n,n}^{SS_1}=0\}$, a.s.,*
2. *$\{(\mathbb{E}_{b_n,n}^{SS}(Z)-\mathbb{E}_n(Z))|\mathcal{X}_n\}:\Omega_n\to\mathbb{R}^\rho$ is a zero mean sample average of b_n observations sampled without replacement from a distribution with variance covariance matrix $\hat{V}$. Moreover, this distribution has finite third moments, a.s., and $||\hat{V}-\mathbf{I}_\rho||\leq O_p(n^{-1/2})$,*
3. *$\tilde{H}:\mathbb{R}^\rho\to\mathbb{R}$ is the same function as in theorem A.1.*

Part 3. Assume (A1)-(A4), (CF') and $\Theta_I=\varnothing$. Then, $\liminf\{P(\Gamma_{b_n,n}^{SS_1}=0|\mathcal{X}_n)=1\}$, a.s..

Proof. This proof follows the one for theorem A.3 very closely. The only difference that is worthwhile pointing out occurs in part 1.

Part 1. In the proof of theorem A.3, we used the CLT for bootstrapped empirical processes applied to a P-Donsker class. We replace this step with the following one. By theorem 3.6.13 and example 3.6.14 in van der Vaart and Wellner [29], $\{v_{b_n,n}^{SS}(m_\theta)\sqrt{1-b_n/n}|\mathcal{X}_n\}:\Omega_n\to l_J^\infty(\Theta)$ converges weakly to a tight Gaussian process, i.o.p.. Since $b_n/n=o(1)$, Slutsky's lemma implies that the empirical process $\{v_{b_n,n}^{SS}(m_\theta)|\mathcal{X}_n\}:\Omega_n\to l_J^\infty(\Theta)$ also converges weakly to the same tight Gaussian process, i.o.p.. The nature of the limiting process can be characterized by considering its marginal distributions. By theorem 2.2.1 of Politis, Romano and Wolf [23], the tight limiting process is the one characterized in theorem A.1, i.p.. ■

As a consequence of the representation result, we can establish the consistency of subsampling approximation.

Theorem A.10 (Consistency of subsampling 1 excluding zero) *Assume (A1)-(A4), (CF').*

Part 1. *If* $\Theta_I \neq \varnothing$*, then,* $\forall \mu > 0$ *and* $\forall \varepsilon > 0$*,*

$$\lim_{n\to+\infty} P^*\left(\sup_{|h|\geq\mu}\left|P\left(\Gamma^{SS_1}_{b_n,n}\leq h\middle|\mathcal{X}_n\right)-\lim_{m\to+\infty}P\left(\Gamma_m\leq h\right)\right|\leq\varepsilon\right)=1$$

Part 2. *If* $\Theta_I = \varnothing$*, then,*

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}}\left|P\left(\Gamma^{SS_1}_{b_n,n}\leq h\middle|\mathcal{X}_n\right)-\lim_{m\to+\infty}P\left(\Gamma_m\leq h\right)\right|=0\right\}\right)=1$$

Proof. This proof follows the arguments of the proof of theorem A.5. ■

The previous result can be utilized to prove the consistency in level of the subsampling approximation, theorem 2.3, whose formulation is given in the main text.

Proof. [**Theorem 2.3**] This proof follows the arguments used in the proof of theorem 2.1. ■

The remaining results of the subsection have the objective of establishing upper and lower bounds on the rates of convergence of the error in the coverage probability of the subsampling approximation. The next lemma establishes an asymptotic expansion for the distribution of a multidimensional average of subsampled observations.

Lemma A.9 *Assume (B1)-(B4), (CF) and that the distribution of the vector* $\{\{1[X=x_k]Y_j\}_{j=1}^J\}_{k=1}^K$ *is strongly non-lattice. Then, the conditional distribution of* $\{\mathbb{E}^{SS}_{b_n,n}(Z)-\mathbb{E}_n(Z)\}$ *defined as in theorem A.9 satisfies the following representation,*

$$P\left(\sqrt{b_n}\left(\mathbb{E}^{SS}_{b_n,n}(Z)-\mathbb{E}_n(Z)\right)\in S\middle|\mathcal{X}_n\right)=\Phi_{\mathbf{I}_\rho}(S)+K_1(S)\,b_n^{-1/2}+K_2(S)\,b_n/n+o_p\left(b_n^{-1/2}+b_n/n\right)$$

uniformly in $S\in\mathcal{C}_\rho$ *, where* $\sup_{S\in\mathcal{C}_\rho}|K_1(S)|<+\infty$ *and* $\sup_{S\in\mathcal{C}_\rho}|K_2(S)|<+\infty$*.*

Proof. By assumption, the distribution of the vector $\{\{1[X=x_k]Y_j\}_{j=1}^J\}_{k=1}^K$ is strongly non-lattice which implies that the distribution of the vector $\{\{1[X=x_k](Y_j-\mathbb{E}(Y_j|x_k))\}_{j=1}^J\}_{k=1}^K$ is also strongly non-lattice.

According to theorem A.9, $\{\mathbb{E}^{SS}_{b_n,n}(Z)-\mathbb{E}_n(Z)\}$ is the average of a random sample extracted without replacement from observations that satisfy $BZ=\{\{1[X=x_k](Y_j-\mathbb{E}(Y_j|x_k))\}_{j=1}^J\}_{k=1}^K$. As a corollary, $\{\mathbb{E}^{SS}_{b_n,n}(Z)-\mathbb{E}_n(Z)\}$ is the average of a random sample extracted without replacement from i.i.d. observations of a strongly non-lattice distribution.

For any $S\in\mathcal{C}_\rho$, let $S_n(S)=\{x\in\mathbb{R}^\rho:(1-b_n/n)^{-1/2}y\in S\}$. Notice that $S\in\mathcal{C}_\rho$ if and only if $S_n(S)\in\mathcal{C}_\rho$. By definition,

$$P\left(\sqrt{b_n}\left(\mathbb{E}^{SS}_{b_n,n}(Z)-\mathbb{E}_n(Z)\right)\in S\middle|\mathcal{X}_n\right)=P\left(\sqrt{b_n}\,(1-b_n/n)^{-1/2}\left(\mathbb{E}^{SS}_{b_n,n}(Z)-\mathbb{E}_n(Z)\right)\in S_n(S)\middle|\mathcal{X}_n\right)$$

Under the strongly non-lattice assumption, Babu and Singh [4] provide an Edgeworth expansion for averages of samples extracted without replacement from a finite population. Using arguments in Bhattacharya and Rao [6], we show that this expansion has an error term that is $o_p(b_n^{-1/2})$ uniformly for a class of functions. We apply the expansion to the class of indicator functions on the elements of $\mathcal{C}_\rho$. One of the leading terms of the expansion in Babu and Singh [4] is a function of sample moments. If we replace sample moments with population moments, we introduce an error term that is $O_p(n^{-1/2})=o_p(b_n^{-1/2})$, uniformly in $S\in\mathcal{C}_\rho$. After this replacement, the Edgeworth expansion is as follows,

$$P\left(\sqrt{b_n}\left(\mathbb{E}^{SS}_{b_n,n}(Z)-\mathbb{E}_n(Z)\right)\in S\middle|\mathcal{X}_n\right)=\Phi_{\mathbf{I}_\rho}(S_n(S))+b_n^{-1/2}K_1(S_n(S))+o_p\left(b_n^{-1/2}\right)$$

uniformly in $S \in \mathcal{C}_\rho$ and where, $\forall \tilde{S} \in \mathcal{C}_\rho$, $K_1(\tilde{S})$ is given by,

$$K_1\left(\tilde{S}\right) = \sum\nolimits_{l \in \left\{b \in \mathbb{N}^\rho : \sum_{j=1}^{\rho} b_j = 3\right\}} \frac{1}{\prod\limits_{j=1,\ldots,\rho} l_j!} \mathbb{E}\left(\prod_{j=1,\ldots,\rho} \left(Z_j - \mathbb{E}\left(Z_j\right)\right)^{l_j}\right) \int_{y \in \tilde{S}} \left(\prod_{j=1,\ldots,\rho} \frac{\partial^{l_j} \phi_{\mathbf{I}_\rho}(y)}{\partial y_j}\right) dy$$

Using change of variables and a Taylor expansion, we deduce that,

$$b_n^{-1/2} K_1\left(S_n\left(S\right)\right) = b_n^{-1/2} K_1\left(S\right) + o\left(b_n^{-1/2}\right)$$

uniformly in $S \in \mathcal{C}_\rho$. Furthermore, since the normal distribution has finite absolute moments of all orders, it follows that, $\sup_{S \in \mathcal{C}_\rho} \left|K_1\left(S\right)\right| < +\infty$.

By applying change of variables and a Taylor expansion once again, we deduce that,

$$\Phi_{\mathbf{I}_\rho}\left(S_n\left(S\right)\right) = \Phi_{\mathbf{I}_\rho}\left(S\right) + K_2\left(S\right) b_n/n + o\left(b_n/n\right)$$

uniformly in $S \in \mathcal{C}_\rho$, where, $\forall \tilde{S} \in \mathcal{C}_\rho$ and denoting $\vartheta \sim N\left(0, \mathbf{I}_\rho\right)$, $K_2(\tilde{S})$ is given by,

$$K_2\left(\tilde{S}\right) = P\left(\vartheta \in \tilde{S}\right) \mathbb{E}\left(1 - \vartheta'\vartheta | \vartheta \in \tilde{S}\right)$$

Since the normal distribution has finite second moments, it follows $\sup_{S \in \mathcal{C}_\rho} \left|K_2\left(S\right)\right| < +\infty$. ■

The following lemma utilizes the previous result to establish an upper bound in the error in this subsampling approximation.

Lemma A.10 *Assume (B1)-(B4), (CF) and that the distribution of the vector* $\{\{1\left[X = x_k\right] Y_j\}_{j=1}^{J}\}_{k=1}^{K}$ *is strongly non-lattice.*

Part 1. If $\Theta_I \neq \varnothing$*, then,* $\forall \mu > 0$,

$$\sup_{|h| \geq \mu} \left| P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right| \leq O_p\left(b_n^{-1/2} + b_n/n\right)$$

Part 2. If $\Theta_I = \varnothing$*, then,*

$$P\left(\liminf \left\{\sup_{h \in \mathbb{R}} \left| P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right| = 0\right\}\right) = 1$$

Proof. Part 1. Consider following argument,

$$\sup_{|h| \geq \mu} \left| P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right| \leq \left\{ \begin{array}{c} \sup_{|h| \geq \mu} \left| P\left(\Gamma_{b_n,n}^{SS_1} \leq h | \mathcal{X}_n\right) - \lim_{m \to +\infty} P\left(\Gamma_m \leq h\right)\right| \\ + \sup_{|h| \geq \mu} \left|\lim_{m \to +\infty} P\left(\Gamma_m \leq h\right) - P\left(\Gamma_n \leq h\right)\right| \end{array} \right\}$$

The right hand side is a sum of two terms. In the proof of theorem A.7, we showed that the second term is $O_p(n^{-1/2})$. Thus, to complete the proof of this part, it suffices to show that the first term is $O_p(b_n^{-1/2} + b_n/n)$.

By theorem A.1, $\lim_{m \to +\infty} P\left(\Gamma_m \leq h\right) = \Phi_{\mathbf{I}_\rho}(\tilde{H}^{-1}\left(\left(-\infty, h\right]\right))$. For any positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ that satisfies $\varepsilon_n = O(n^{-1/2})$ consider the following derivation,

$$\sup_{|h| \geq \mu} \left(P\left(\Gamma_{b_n,n}^{SS_1} \leq h | \mathcal{X}_n\right) - \Phi_{\mathbf{I}_\rho}\left(\tilde{H}^{-1}\left(\left(-\infty, h\right]\right)\right)\right) \leq$$

$$\left\{ \begin{array}{c} \sup_{|h| \geq \mu} \left| P\left(\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}\left(Z\right) - \mathbb{E}_n\left(Z\right)\right) \in \tilde{H}^{-1}\left(\left(-\infty, h + \varepsilon_n\right]\right) \middle| \mathcal{X}_n\right) - \Phi_{\mathbf{I}_\rho}\left(\tilde{H}^{-1}\left(\left(-\infty, h + \varepsilon_n\right]\right)\right)\right| \\ + P\left(\left|\tilde{\delta}_{b_n,n}^{SS_1}\right| > \varepsilon_n \middle| \mathcal{X}_n\right) + \sup_{|h| \geq \mu} \Phi_{\mathbf{I}_\rho}\left(\tilde{H}^{-1}\left(\left(h - \varepsilon_n, h + \varepsilon_n\right]\right)\right) \end{array} \right\}$$

The upper bound is a sum of three terms. By lemma A.9, the first term is $O_p(b_n^{-1/2}+b_n/n)$, by theorem A.9, the second term is $o_p(n^{-1/2})$ and by lemma A.6, the third term is $O_p(n^{-1/2})$. Thus, the whole expression is $O_p(b_n^{-1/2}+b_n/n)$. The proof of this part is completed by repeating the same argument with $P(\Gamma^{SS_1}_{b_n,n}>h|\mathcal{X}_n)$ (instead of $P(\Gamma^{SS_1}_{b_n,n}\leq h|\mathcal{X}_n)$).

Part 2. This follows from the arguments used in the proof of theorem A.7. ■

Corollary A.3 *Assume (B1)-(B4), (CF), $\Theta_I\neq\varnothing$ and that the distribution of the vector $\{\{1[X=x_k]Y_j\}_{j=1}^J\}_{k=1}^K$ is strongly non-lattice. For any $\alpha\in(0,0.5)$, let $q^{SS_1}_{b_n,n}(1-\alpha)=P(\Gamma^{SS_1}_{b_n,n}\leq\hat{c}^{SS_1}_{b_n,n}(1-\alpha)|\mathcal{X}_n)$. Then, $|q^{SS_1}_{b_n,n}(1-\alpha)-(1-\alpha)|\leq O_p(b_n^{-1/2}+b_n/n)$.*

Proof. This proof follows the arguments used in corollary A.2. ■

These results allow us to show an upper bound on the rate of convergence of the error in the coverage probability, which is formulated in the main text.

Proof. [**Theorem 2.4**] This proof follows the arguments used in theorem 2.2. ■

Theorem 2.4 describes the coverage properties of subsampling 1 when $\Theta_I\neq\varnothing$. In the case when $\Theta_I=\varnothing$, the confidence sets constructed using subsampling 1 present the same coverage properties as the ones shown for the bootstrap in lemma A.7.

As a corollary of theorem 2.4, the subsampling size that minimizes the upper bound on the rate of convergence of the error in the coverage probability is $b_n=O\left(n^{2/3}\right)$, which implies a rate of convergence of order $n^{1/3}$. In the remainder of this section, we show conditions under which this rate is the exact rate of convergence of the error in the coverage probability.

We begin by establishing a useful property of the function associated to one of the leading terms of the asymptotic expansion provided in lemma A.9.

Lemma A.11 *For $\tilde{H}$ defined as in theorem A.1, $\vartheta\sim N(0,\mathbf{I}_\rho)$ and $\forall\gamma>0$, let h_L and h_H be such that $P(\tilde{H}(\vartheta)\leq h_L)=0.72$ and $P(\tilde{H}(\vartheta)\leq h_H)=1-\gamma$ and let $\Lambda(\gamma)=\{S\in\mathcal{C}_\rho:\exists h\in[h_L,h_H]:S=\{y\in\mathbb{R}^\rho:\tilde{H}(y)\leq h\}\}$. Then, the function $K_2:\mathcal{C}_\rho\to\mathbb{R}$ defined in lemma A.9 satisfies $\inf_{S\in\Lambda(\gamma)}|K_2(S)|>0$.*

Proof. Fix $\gamma>0$ arbitrarily and consider any $S\in\Lambda(\gamma)$. Since $\tilde{H}$ is homogenous of degree $\beta\geq1$, if $y\in S$, then $\forall\lambda\in[0,1]$, $\lambda y\in S$.

Case 1: $\rho=1$. In this case, $\forall S\in\Lambda(\gamma)$, $S=[-y_1,y_2]$ for some y_1, $y_2\in\mathbb{R}_+\cap\{+\infty\}$. Define $H_\gamma\in\mathbb{R}$ such that $P(\vartheta\in[-H_\gamma,H_\gamma])=1-\gamma$. It follows immediately that, $\forall[-y_1,y_2]\in\Lambda(\gamma)$, $\min\{y_1,y_2\}\leq H_\gamma$. Define $L\in\mathbb{R}$ such that $P(\vartheta\in[-L,+\infty))=0.72$ (or, equivalently, $P(\vartheta\in(-\infty,L])=0.72$). It follows immediately that, $\forall[-y_1,y_2]\in\Lambda(\gamma)$, $\min\{y_1,y_2\}\geq L$.

By symmetry in the formula, $K_2([-y_1,y_2])=K_2([0,y_1])+K_2([0,y_2])$. Consider the function $f(y)=K_2([0,y]):\mathbb{R}_+\cap\{+\infty\}\to\mathbb{R}$. This function is strictly increasing for $y\leq1$ and strictly decreasing for $y\geq1$. Moreover, $\forall y\in(0,+\infty)$, $f(y)>0$ and $f(0)=f(+\infty)=0$. Therefore, it follows that, $\inf_{L\leq y}K_2([0,y])=0$ and $\inf_{L\leq y\leq H_\gamma}K_2([0,y])=\min\{K_2([0,L]),K_2([0,H_\gamma])\}>0$. Therefore,

$$\begin{aligned}\inf_{S\in\Lambda(\gamma)}K_2(S) &= \inf_{\{\{y_1,y_2\}:[-y_1,y_2]\in\Lambda(\gamma)\}}\{K_2([0,\min\{y_1,y_2\}])+K_2([0,\max\{y_1,y_2\}])\}\\ &\geq \inf_{L\leq y\leq H_\gamma}K_2([0,y])+\inf_{L\leq y}K_2([0,y])\\ &= \min\{K_2([0,L]),K_2([0,H_\gamma])\}\end{aligned}$$

If we set $C_A=\min\{K_2([0,L]),K_2([0,H_\gamma])\}>0$, then $\exists C_A>0$ such that $\inf_{S\in\Lambda(\gamma)}K_2(S)\geq C_A$.

Case 2: $\rho\geq2$. In order to keep track of the dimension, denote $\vartheta_\rho\sim N(0,\mathbf{I}_\rho)$. For every $\rho\geq2$ and $\pi\in[0,1]$, let $c(\pi,\rho)$ be (uniquely) defined by, $P\left(\vartheta_\rho'\vartheta_\rho\leq c(\pi,\rho)\right)=\pi$. Notice that $c(\pi,\rho)$ is an increasing function of π and a decreasing function of ρ. By definition, $\forall S \in \Lambda(\gamma)$ and $\forall \rho \geq 2$, $P\left(\vartheta_\rho' \vartheta_\rho \leq c\left(P\left(\vartheta_\rho \in S\right), \rho\right)\right) = P\left(\vartheta_\rho \in S\right)$, and so, it follows that,

$$P\left(\vartheta_\rho \in \left\{S \cap \left\{x \in \mathbb{R}^\rho : x'x > c\left(P\left(\vartheta_\rho \in S\right), \rho\right)\right\}\right\}\right) = P\left(\vartheta_\rho \in \left\{\{S\}^c \cap \left\{x \in \mathbb{R}^\rho : x'x \leq c\left(P\left(\vartheta_\rho \in S\right), \rho\right)\right\}\right\}\right)$$

Based on this definitions, $\forall S \in \Lambda(\gamma)$ and $\forall \rho \geq 2$, consider the following derivation,

$$\begin{aligned}
K_2(S) &= \int_{x \in S} \left(1 - x'x\right) d\Phi_{\mathbf{I}_\rho}(x) \\
&\leq \left\{ \begin{array}{c} \int_{x \in \{S \cap \{x'x \leq c(P(\vartheta_\rho \in S), \rho)\}\}} \left(1 - x'x\right) d\Phi_{\mathbf{I}_\rho}(x) + \\ \int_{x \in \{S \cap \{x'x > c(P(\vartheta_\rho \in S), \rho)\}\}} \left(1 - c\left(P\left(\vartheta_\rho \in S\right), \rho\right)\right) d\Phi_{\mathbf{I}_\rho}(x) \end{array} \right\} \\
&= \left\{ \begin{array}{c} \int_{x \in \{S \cap \{x'x \leq c(P(\vartheta_\rho \in S), \rho)\}\}} \left(1 - x'x\right) d\Phi_{\mathbf{I}_\rho}(x) + \\ \int_{x \in \{\{S\}^c \cap \{x'x \leq c(P(\vartheta_\rho \in S), \rho)\}\}} \left(1 - c\left(P\left(\vartheta_\rho \in S\right), \rho\right)\right) d\Phi_{\mathbf{I}_\rho}(x) \end{array} \right\} \\
&\leq P\left(\vartheta_\rho \in S\right) \mathbb{E}\left(1 - \vartheta_\rho' \vartheta_\rho | \vartheta_\rho' \vartheta_\rho \leq c\left(P\left(\vartheta_\rho \in S\right), \rho\right)\right)
\end{aligned}$$

Since $c\left(P\left(\vartheta_\rho \in S\right), \rho\right)$ is increasing in the first coordinate and $\forall S \in \Lambda(\gamma)$, $P\left(\vartheta_\rho \in S\right) \geq 0.72$, it follows that, $\forall S \in \Lambda(\gamma)$ and $\forall \rho \geq 2$, $K_2(S) \leq P\left(\vartheta_\rho \in S\right) \mathbb{E}\left(1 - \vartheta_\rho' \vartheta_\rho | \vartheta_\rho' \vartheta_\rho \leq c(0.72, \rho)\right)$. To conclude the proof, it suffices to show that, $\forall \rho \geq 2$, $\inf_{\rho \geq 2} \mathbb{E}\left(\vartheta_\rho' \vartheta_\rho | \vartheta_\rho' \vartheta_\rho \leq c(0.72, \rho)\right) > 1$. It can be verified that $\mathbb{E}\left(\vartheta_\rho' \vartheta_\rho | \vartheta_\rho' \vartheta_\rho \leq c(0.72, \rho)\right)$ is increasing in ρ and it is greater than one of $\rho = 2$. As a consequence, $\forall \rho \geq 2$, $\exists C_B > 0$, $\sup_{S \in \Lambda(\gamma)} K_2(S) \leq -C_B$.

To conclude the proof, define $C = \min\{C_A, C_B\} > 0$ and combine the findings of both cases to deduce that $\inf_{S \in \Lambda(\gamma)} |K_2(S)| \geq C$. ■

The following lemma provides conditions under which the rate of convergence of the subsampling approximation is exactly $b_n^{-1/2} + b_n/n$.

Lemma A.12 *Assume (B1)-(B4), (CF) and that the distribution of the vector* $\{\{1[X = x_k] Y_j\}_{j=1}^J\}_{k=1}^K$ *is strongly non-lattice. Furthermore, assume that* $K_1(\tilde{H}^{-1}((-\infty, c_\infty(1-\alpha)])) > 0$, *where* $K_1 : \mathcal{C}_\rho \to \mathbb{R}$ *is defined as in lemma A.9,* $c_\infty(1-\alpha)$ *is defined by* $P(\tilde{H}(\vartheta) \leq c_\infty(1-\alpha)) = (1-\alpha)$, $\tilde{H}$ *is the function defined in theorem A.1 and* $\vartheta \sim N(0, \mathbf{I}_\rho)$. *If* $\Theta_I \neq \varnothing$, *then,* $\forall \varepsilon > 0$, $\exists \eta > 0$, $\exists C > 0$ *and* $\exists N \in \mathbb{N}$ *such that,* $\forall n \geq N$,

$$P\left(\inf_{h \in [c_\infty(1-\alpha)-\eta, c_\infty(1-\alpha)+\eta]} \left| P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right) \right| \geq C\left(b_n^{-1/2} + b_n/n\right)\right) \geq 1 - \varepsilon$$

Proof. Fix $\mu > 0$ arbitrarily. Consider the following derivation $\forall h$ such that $|h| > \mu$,

$$P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right) = \left\{ \begin{array}{c} \left(P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\tilde{H}\left(\sqrt{b_n}\left(\mathbb{E}_n^{SS}(Z) - \mathbb{E}_n(Z)\right)\right) \leq h \middle| \mathcal{X}_n\right)\right) \\ + \left(P\left(\tilde{H}\left(\sqrt{b_n}\left(\mathbb{E}_n^{SS}(Z) - \mathbb{E}_n(Z)\right)\right) \leq h \middle| \mathcal{X}_n\right) - P\left(\tilde{H}(\vartheta) \leq h\right)\right) \\ + \left(P\left(\tilde{H}(\vartheta) \leq h\right) - P\left(\Gamma_n \leq h\right)\right) \end{array} \right\}$$

where $\vartheta \sim N(0, \mathbf{I}_\rho)$. The right hand side is a sum of three terms. By theorem A.9, the first term is $o_p(n^{-1/2})$, uniformly in $h \in \mathbb{R}$. By arguments used in theorem A.7, the third term is $O_p(n^{-1/2}) = o_p(b_n^{-1/2})$, uniformly in $|h| \geq \mu$. Finally, using lemma A.9, the second term can be expressed as follows,

$$\begin{aligned}
&P\left(\tilde{H}\left(\sqrt{b_n}\left(\mathbb{E}_n^{SS}(Z) - \mathbb{E}_n(Z)\right)\right) \leq h \middle| \mathcal{X}_n\right) - P\left(\tilde{H}(\vartheta) \leq h\right) \\
&= P\left(\sqrt{b_n}\left(\mathbb{E}_n^{SS}(Z) - \mathbb{E}_n(Z)\right) \in \tilde{H}^{-1}((-\infty, h]) \middle| \mathcal{X}_n\right) - \Phi_{\mathbf{I}_\rho}(\tilde{H}^{-1}((-\infty, h]) \\
&= K_1\left(\tilde{H}^{-1}((-\infty, h])\right) b_n^{-1/2} + K_2\left(\tilde{H}^{-1}((-\infty, h])\right) b_n/n + o_p\left(b_n^{-1/2} + b_n/n\right)
\end{aligned}$$

uniformly in $|h| \geq \mu$.

If we combine the information from the three terms, we deduce the following expression,

$$P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right) = K_1\left(\tilde{H}^{-1}\left((-\infty, h]\right)\right) b_n^{-1/2} + K_2\left(\tilde{H}^{-1}\left((-\infty, h]\right)\right) b_n/n + o_p\left(b_n^{-1/2} + b_n/n\right)$$

uniformly in $|h| \geq \mu$.

By arguments in corollary A.1, $c_\infty(1-\alpha) > 0$ and therefore, $\exists \eta' > 0$ such that $[c_\infty(1-\alpha) - \eta', c_\infty(1-\alpha) + \eta'] \subseteq \{h' \in \mathbb{R} : |h'| \geq \mu\}$. By the definition of K_1 and by properties of the function $\tilde{H}$, $K_1(\tilde{H}^{-1}((-\infty, h]))$ is continuous $\forall h \in \{h' \in \mathbb{R} : |h'| \geq \mu\}$. Since $K_1(\tilde{H}^{-1}((-\infty, c_\infty(1-\alpha)])) > 0$, then, by continuity, $\exists C_1 > 0$ and $\exists \eta \in (0, \eta')$ such that, $\inf_{[c_\infty(1-\alpha)-\eta, c_\infty(1-\alpha)+\eta]} K_1(\tilde{H}^{-1}((-\infty, h])) \geq C_1$. By lemma A.11, $\exists C_2 > 0$ such that, $\inf_{[c_\infty(1-\alpha)-\eta, c_\infty(1-\alpha)+\eta]} K_2(\tilde{H}^{-1}((-\infty, h])) \geq C_2$. Finally, let ε_n be defined as follows,

$$\varepsilon_n = \sup_{[c_\infty(1-\alpha)-\eta, c_\infty(1-\alpha)+\eta]} \left| \begin{array}{c} \left(P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right) - \\ \left(K_1\left(\tilde{H}^{-1}\left((-\infty, h]\right)\right) b_n^{-1/2} + K_2\left(\tilde{H}^{-1}\left((-\infty, h]\right)\right) b_n/n\right) \end{array} \right|$$

By definition, $\varepsilon_n = o_p(b_n^{-1/2} + b_n/n)$.

Define $C = \min\{C_1, C_2\}/2 > 0$ and consider the following derivation,

$$\begin{aligned} & P\left(\inf_{h \in [c_\infty(1-\alpha)-\eta, c_\infty(1-\alpha)+\eta]} \left| P\left(\Gamma_{b_n,n}^{SS_1} \leq h \middle| \mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right| \geq C\left(b_n^{-1/2} + b_n/n\right)\right) \\ & \geq P\left(\inf_{\{h \in [c_\infty(1-\alpha)-\eta, c_\infty(1-\alpha)+\eta]\}} \left| \begin{array}{c} K_1\left(\tilde{H}^{-1}\left((-\infty, h]\right)\right) b_n^{-1/2} \\ +K_2\left(\tilde{H}^{-1}\left((-\infty, h]\right)\right) b_n/n \end{array} \right| + \varepsilon_n \geq C\left(b_n^{-1/2} + b_n/n\right)\right) \\ & \geq P\left(\varepsilon_n \geq -C\left(b_n^{-1/2} + b_n/n\right)\right) \end{aligned}$$

Since $\varepsilon_n = o_p(b_n^{-1/2} + b_n/n)$, the right hand side converges to one. This concludes the proof. ■

Lemma A.13 *Assume (B1)-(B4) and (CF). For any* μ_L *and* μ_H *such that* $(\mu_L, \mu_H) \subset (\mu, 1)$, *let* h_L *and* h_H *be such that* $P(\tilde{H}(\vartheta) \leq h_L) = \mu_L$ *and* $P(\tilde{H}(\vartheta) \leq h_H) = \mu_H$, *where* $\tilde{H}$ *is the function defined in theorem A.1 and* $\vartheta \sim N(0, \mathbf{I}_\rho)$. *If* $(1-\alpha) \in (\mu_L, \mu_H)$, *then,* $\lim_{n\to+\infty} P(\hat{c}_{b_n,n}^{SS_1}(1-\alpha) \in (h_L, h_H)) = 1$.

Proof. This follows from lemma A.10 and the arguments used in corollary A.2. ■

The conclusion of this section is the following theorem, which establishes that, under certain conditions, $b_n^{-1/2} + b_n/n$ is the exact rate of convergence of the error in the coverage probability for subsampling approximation 1.

Theorem A.11 *Assume (B1)-(B4), (CF) and that the distribution of the vector* $\{\{1[X = x_k] Y_j\}_{j=1}^J\}_{k=1}^K$ *is absolutely continuous with respect to Lebesgue measure. Moreover, assume that* $K_1(\tilde{H}^{-1}((-\infty, c_\infty(1-\alpha)])) > 0$, *where* K_1 *is the function defined in lemma A.9,* $\tilde{H}$ *is the function defined in theorem A.1 and, for* $\vartheta \sim N(0, \mathbf{I}_\rho)$, $c_\infty(1-\alpha)$ *is defined by* $P(\tilde{H}(\vartheta) \leq c_\infty(1-\alpha)) = (1-\alpha)$. *If* $\Theta_I \neq \varnothing$ *and* $(1-\alpha) \in [0.72, 1)$, *then,* $\exists C > 0$ *and* $\exists N \in \mathbb{N}$ *such that,* $\forall n \geq N$,

$$\left|P\left(\Theta_I \subset C_{b_n,n}^{SS_1}(1-\alpha)\right) - (1-\alpha)\right| > C\left(b_n^{-1/2} + b_n/n\right)$$

Proof. By lemmas A.12 and A.13, $\forall\varepsilon>0$, $\exists\eta>0$, $\exists C>0$ and $\exists N_1\in\mathbb{N}$ such that, $\forall n\geq N_1$,

$$\begin{aligned} P\left(\inf_{h\in[c_\infty(1-\alpha)-\eta,c_\infty(1-\alpha)+\eta]}\left|P\left(\Gamma^{SS_1}_{b_n,n}\leq h|\mathcal{X}_n\right)-P\left(\Gamma_n\leq h\right)\right|\geq 2C\left(b_n^{-1/2}+b_n/n\right)\right) &\geq 1-\varepsilon/2\\ P\left(\hat{c}^{SS_1}_{b_n,n}\left(1-\alpha\right)\notin\left[c_\infty\left(1-\alpha\right)-\eta,c_\infty\left(1-\alpha\right)+\eta\right]\right) &\leq \varepsilon/2 \end{aligned}$$

Therefore, we have the following derivation,

$$\begin{aligned} 1-\varepsilon/2 &\leq P\left(\inf_{h\in[c_\infty(1-\alpha)-\eta,c_\infty(1-\alpha)+\eta]}\left|P\left(\Gamma^{SS_1}_{b_n,n}\leq h|\mathcal{X}_n\right)-P\left(\Gamma_n\leq h\right)\right|\geq 2C\left(b_n^{-1/2}+b_n/n\right)\right)\\ &\leq\left\{\begin{array}{c} P\left(\left|P\left(\Gamma^{SS_1}_{b_n,n}\leq\hat{c}^{SS_1}_{b_n,n}\left(1-\alpha\right)|\mathcal{X}_n\right)-P\left(\Gamma_n\leq\hat{c}^{SS_1}_{b_n,n}\left(1-\alpha\right)\right)\right|\geq 2C\left(b_n^{-1/2}+b_n/n\right)\right)\\ +P\left(\hat{c}^{SS_1}_{b_n,n}\left(1-\alpha\right)\notin\left[c_\infty\left(1-\alpha\right)-\eta,c_\infty\left(1-\alpha\right)+\eta\right]\right)\end{array}\right\}\\ &\leq P\left(\left|q^{SS_1}_{b_n,n}\left(1-\alpha\right)-P\left(\Theta_I\subset\hat{C}^{SS_1}_{b_n,n}\left(1-\alpha\right)\right)\right|\geq 2C\left(b_n^{-1/2}+b_n/n\right)\right)+\varepsilon/2 \end{aligned}$$

where $q^{SS_1}_{b_n,n}\left(1-\alpha\right)=P(\Gamma^{SS_1}_{b_n,n}\leq\hat{c}^{SS_1}_{b_n,n}\left(1-\alpha\right)|\mathcal{X}_n)$ and $P(\Theta_I\subset\hat{C}^{SS_1}_{b_n,n}\left(1-\alpha\right))=P(\Gamma_n\leq\hat{c}^{SS_1}_{b_n,n}\left(1-\alpha\right))$. From this derivation, we deduce that, $\forall\varepsilon>0$, $\exists C>0$ and $\exists N_1\in\mathbb{N}$ such that, $\forall n\geq N_1$,

$$P\left(\left|P\left(\Theta_I\subset\hat{C}^{SS_1}_{b_n,n}\left(1-\alpha\right)\right)-q^{SS_1}_{b_n,n}\left(1-\alpha\right)\right|\geq 2C\left(b_n^{-1/2}+b_n/n\right)\right)\geq 1-\varepsilon$$

By the result in theorem A.9 and since $\alpha\in(0,0.5)$, $\{q^{SS_1}_{b_n,n}\left(1-\alpha\right)>0\}$, w.p.a.1. By properties of the function $\tilde{H}$, $\forall h>0$, the function $P(\Gamma^{SS_1}_{b_n,n}\leq h|\mathcal{X}_n)$ is a piece-wise constant function. Since the distribution of $\{\{1\left[X=x_k\right]Y_j\}_{j=1}^J\}_{k=1}^K$ is absolutely continuous with respect to Lebesgue measure then, for any full rank matrix B, the distribution of Z that satisfies $BZ=\{\{1\left[X=x_k\right]\left(Y_j-\mathbb{E}\left(Y_j|x_k\right)\right)\}_{j=1}^J\}_{k=1}^K$ is also absolutely continuous with respect to the Lebesgue measure. As a consequence, for any two different subsamples of size b_n from the sample, the value of $\mathbb{E}^{SS}_{b_n,n}\left(Z\right)$ will not coincide, a.s., and so, $\forall h>0$, the maximum jump size of the function $P(\Gamma^{SS_1}_{b_n,n}\leq h|\mathcal{X}_n)$ is the mutiplicative inverse of the number of subsamples in the sample. By this argument,

$$P\left(\left|q^{SS_1}_{b_n,n}\left(1-\alpha\right)-\left(1-\alpha\right)\right|\leq n^{-1/2}\right)\geq 1\left[\binom{b_n}{n}^{-1}\leq n^{-1/2}\right]$$

By properties of the combinatorial formula, $\exists N_2$ such that, $\forall n\geq N_2$, the right hand side equals one.

Finally, $\exists N_3$ such that, $\forall n\geq N_3$, $2C(b_n^{-1/2}+b_n/n)-n^{-1/2}\geq C(b_n^{-1/2}+b_n/n)$.

Combining all the findings, consider the following derivation. For every $\varepsilon>0$, $\exists C>0$ and $\exists N=\max\{N_1,N_2,N_3\}$ such that, $\forall n\geq N$,

$$\begin{aligned} 1-\varepsilon &\leq\left\{\begin{array}{c} P\left(\left\{\left|P\left(\Theta_I\subset\hat{C}^{SS_1}_{b_n,n}\left(1-\alpha\right)\right)-q^{SS_1}_{b_n,n}\left(1-\alpha\right)\right|\geq 2C\left(b_n^{-1/2}+b_n/n\right)\right\}\right)\\ +P\left(\left\{\left|q^{SS_1}_{b_n,n}\left(1-\alpha\right)-\left(1-\alpha\right)\right|\leq n^{-1/2}\right\}\right)-1\end{array}\right\}\\ &\leq P\left(\begin{array}{c}\left\{\left|P\left(\Theta_I\subset\hat{C}^{SS_1}_{b_n,n}\left(1-\alpha\right)\right)-q^{SS_1}_{b_n,n}\left(1-\alpha\right)\right|\geq 2C\left(b_n^{-1/2}+b_n/n\right)\right\}\\ \cap\left\{\left|q^{SS_1}_{b_n,n}\left(1-\alpha\right)-\left(1-\alpha\right)\right|\leq n^{-1/2}\right\}\end{array}\right)\\ &\leq P\left(\left\{\begin{array}{c}\left|P\left(\Theta_I\subset\hat{C}^{SS_1}_{b_n,n}\left(1-\alpha\right)\right)-q^{SS_1}_{b_n,n}\left(1-\alpha\right)\right|-\left|q^{SS_1}_{b_n,n}\left(1-\alpha\right)-\left(1-\alpha\right)\right|\\ \geq 2C\left(b_n^{-1/2}+b_n/n\right)-n^{-1/2}\geq C\left(b_n^{-1/2}+b_n/n\right)\end{array}\right\}\right)\\ &\leq P\left(\left|P\left(\Theta_I\subset\hat{C}^{SS_1}_{b_n,n}\left(1-\alpha\right)\right)-\left(1-\alpha\right)\right|\geq C\left(b_n^{-1/2}+b_n/n\right)\right) \end{aligned}$$

Since the event inside the probability is non-random, then it must occur $\forall n \geq N$. ■

Subsampling 2 The procedure is as follows,

1. Choose $\{b_n\}_{n=1}^{+\infty}$ to be a positive sequence such that $b_n \to +\infty$ and $b_n/n = o(1)$ at polynomial rates. Choose $\{\tau_n\}_{n=1}^{+\infty}$ to be a positive sequence such that for some $\gamma > 0$, $(\ln\ln b_n)^{\beta/2+\gamma}\,\tau_n\sqrt{b_n/n} = o(1)$ and $\sqrt{\ln\ln n}/\tau_n = o(1)$, a.s.,

2. Estimate the identified set with $\hat{\Theta}_I(\tau_n) = \{\theta \in \Theta : \{\mathbb{E}_n(m_j(Z,\theta)) \leq \tau_n/\sqrt{n}\}_{j=1}^{J}\}$,

3. Repeat the following step for $s = 1, 2, ..., S$. Construct a subsample of size b_n by sampling randomly without replacement from the data. Denote these observations by $\{Z_i^{SS}\}_{i=1}^{b_n}$ and for every $j = 1, 2, ..., J$, let $\mathbb{E}_{b_n,n}^{SS}(m_j(Z,\theta)) = b_n^{-1}\sum_{i=1}^{b_n} m_j(Z_i^{SS},\theta)$. Compute,

$$\Gamma_{b_n,n}^{SS_2} = \begin{cases} \sup\limits_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\left\{\left[\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(m_j(Z,\theta))\right)\right]_+\right\}_{j=1}^{J}\right) & \text{if } \hat{\Theta}_I(\tau_n) \neq \varnothing \\ 0 & \text{if } \hat{\Theta}_I(\tau_n) = \varnothing \end{cases}$$

4. Let $\hat{c}_{b_n,n}^{SS_2}(1-\alpha)$ be the $(1-\alpha)$ quantile of the distribution of $\Gamma_{b_n,n}^{SS_2}$, simulated with arbitrary accuracy in the previous step. The $(1-\alpha)$ confidence set for the identified set is given by $\hat{C}_{b_n,n}^{SS_2}(1-\alpha) = \{\theta \in \Theta : \sqrt{n}Q_n(\theta) \leq \hat{c}_{b_n,n}^{SS_2}(1-\alpha)\}$.

Some comments are in order. Notice how the conditions over the sequences $\{b_n\}_{n=1}^{+\infty}$ and $\{\tau_n\}_{n=1}^{+\infty}$ in step 1 are stronger than the ones required for the subsampling procedure 1. As we will soon show, under assumptions (A1)-(A4) and (CF')[28], these requirements are sufficient to deduce the consistency in level of subsampling 2. In particular, in order to make the restrictions on the sequence $\{\tau_n\}_{n=1}^{+\infty}$ possible, it is important to satisfy the polynomial requirements on the rate of growth of the sequence $\{b_n\}_{n=1}^{+\infty}$. To see why, note that if $b_n = n/\sqrt{\ln\ln n}$ (so that $b_n/n = o(1)$ at a sub-polynomial rate), there would be a contradiction between $(\ln\ln b_n)^{\beta/2+\gamma}\,\tau_n\sqrt{b_n/n} = o(1)$ and $\sqrt{\ln\ln n}/\tau_n = o(1)$. Furthermore, since this procedure has no recentering term in step 3, it is not possible to define a version of this procedure that is specialized for the conditionally separable model.

The following lemma is an intermediate result regarding empirical processes created from subsampled observations.

Lemma A.14 *For any positive sequence $\{\gamma_n\}_{n=1}^{+\infty}$ such that $\sqrt{\ln\ln b_n}/\gamma_n = o(1)$,*

$$\lim_{n\to+\infty} P^*\left(\sup_{\theta\in\Theta}\max_{j=1,\ldots,J}\left|\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(m_j(Z,\theta)) - \mathbb{E}_n(m_j(Z,\theta))\right)\right| \leq \gamma_n \middle| \mathcal{X}_n\right) = 1, \text{ a.s.}$$

Proof. For any $(\theta, j) \in \Theta \times \{1, ..., J\}$, let $v_{b_n,n}^{SS}(m_{j,\theta}) = \sqrt{b_n}(\mathbb{E}_{b_n,n}^{SS}(m_j(Z,\theta)) - \mathbb{E}_n(m_j(Z,\theta)))$. By elementary derivations, it suffices to show that, $\forall j = 1, ..., J$, $\lim_{n\to+\infty} P^*(\sup_{\theta\in\Theta}|v_{b_n,n}^{SS}(m_{j,\theta})| \leq \gamma_n|\mathcal{X}_n) = 1$, a.s..

For any $\delta > 0$, the compactness of Θ implies that there exists a finite collection of parameters of Θ, denoted by $\{\theta_l\}_{l=1}^{L}$, such that, $\forall\theta \in \Theta$, $\exists l \in \{1, ..., L\}$ that satisfies $||\theta_l - \theta|| < \delta$. Therefore,

$$\begin{aligned} \sup_{\theta\in\Theta}\left|v_{b_n,n}^{SS}(m_{j,\theta})\right| &\leq \max_{l\leq L}\sup_{\{\theta\in\Theta:||\theta_l-\theta||\leq\delta\}}\left|v_{b_n,n}^{SS}(m_{j,\theta}) - v_{b_n,n}^{SS}(m_{j,\theta_l})\right| + \max_{l\leq L}\left|v_{b_n,n}^{SS}(m_{j,\theta_l})\right| \\ &\leq \max_{\theta\in\Theta}\sup_{\{\theta'\in\Theta:||\theta'-\theta||\leq\delta\}}\left|v_{b_n,n}^{SS}(m_{j,\theta}) - v_{b_n,n}^{SS}(m_{j,\theta'})\right| + \max_{l\leq L}\left|v_{b_n,n}^{SS}(m_{j,\theta_l})\right| \end{aligned}$$

[28] If we replace assumption (CF') by assumption (CF), then the conditions on the sequence $\{\tau_n\}_{n=1}^{+\infty}$ can be replaced by $\tau_n\sqrt{b_n/n} = o(1)$ and $\sqrt{\ln\ln n}/\tau_n = o(1)$, a.s..

It then follows that, $\forall\varepsilon > 0$,

$$P^*\left(\sup_{\theta\in\Theta}\left|v_{b_n,n}^{SS}(m_{j,\theta})\right| \leq \gamma_n\right|\mathcal{X}_n\left.\right) \geq$$
$$\left\{\begin{array}{c} P^*\left(\max_{\theta\in\Theta}\sup_{\{\theta'\in\Theta:\|\theta'-\theta\|\leq\delta\}}\left|v_{b_n,n}^{SS}(m_{j,\theta}) - v_{b_n,n}^{SS}(m_{j,\theta'}))\right| \leq \varepsilon\right|\mathcal{X}_n\left.\right) + \\ +\sum_{l=1}^{L} P^*\left(\left|v_{b_n,n}^{SS}(m_{j,\theta_l})\right| \leq \gamma_n/2\right|\mathcal{X}_n\left.\right) - L \end{array}\right\}$$

Thus, it suffices to show that, $\forall j = 1, ..., J$, the following two statements hold,

$$\forall\theta\in\Theta,\ \lim_{n\to+\infty} P^*\left(\left|v_{b_n,n}^{SS}(m_{j,\theta})\right| \leq \gamma_n/2\right|\mathcal{X}_n\left.\right) = 1,\ \text{a.s.}$$
$$\lim_{n\to+\infty} P^*\left(\max_{\theta\in\Theta}\sup_{\{\theta'\in\Theta:\|\theta'-\theta\|\leq\delta\}}\left|v_{b_n,n}^{SS}(m_{j,\theta}) - v_{b_n,n}^{SS}(m_{j,\theta'})\right| \leq \varepsilon\right|\mathcal{X}_n\left.\right) = 1,\ \text{a.s.}$$

We begin with the first statement. For any $(j,\theta) \in \{1,...,J\}\times\Theta$ and conditioning on the sample, $b_n^{-1/2}v_{b_n,n}^{SS}(m_{j,\theta})$ is the zero mean average of random variables sampled without replacement from the observations in the sample. Following Joag-Dev and Proschan [18], random sampling without replacement produces negatively associated random variables. By Shao and Su [27], a random sample of negatively associated random variables satisfies the LIL. By the LIL and $\sqrt{\ln\ln b_n}/\gamma_n = o(1)$, it follows that, $\forall(j,\theta) \in \{1,...,J\}\times\Theta$,

$$\liminf\left\{\left|v_{b_n,n}^{SS}(m_{j,\theta})\right| \leq \gamma_n/2|\mathcal{X}_n\right\},\ \text{a.s.}$$

which implies the first statement.

To show the second statement, we use Markov's inequality and arguments in the proof of theorem 3.6.13 in van der Vaart and Wellner [29]. ■

The following theorem is the representation result for the subsampling procedure under consideration.

Theorem A.12 *Part 1.* *Assume (A1)-(A4), (CF') and $\Theta_I \neq \varnothing$. Then, $\Gamma_{b_n,n}^{SS_2} = H(v_{b_n,n}^{SS}(m_\theta)) + \delta_{b_n,n}^{SS_2}$, where,*

1. *for any $\varepsilon > 0$, $\lim_{n\to+\infty} P^*(|\delta_{b_n,n}^{SS_2}| > \varepsilon|\mathcal{X}_n) = 0$, a.s.,*
2. *$\{v_{b_n,n}^{SS}(m_\theta)|\mathcal{X}_n\} : \Omega_n \to l_J^\infty(\Theta)$ is an empirical process that converges weakly to the same Gaussian process as in theorem A.1, i.o.p.,*
3. *$H : l_J^\infty(\Theta) \to \mathbb{R}$ is the same function as in theorem A.1.*

Part 2. *Let ρ denote the rank of the variance covariance matrix of the vector $\{\{1[X = x_k]Y_j\}_{j=1}^J\}_{k=1}^K$. If we assume (B1)-(B4), (CF) and $\Theta_I \neq \varnothing$, then, $\Gamma_{b_n,n}^{SS_2} = \tilde{H}(\sqrt{b_n}(\mathbb{E}_{b_n,n}^{SS}(Z) - \mathbb{E}_n(Z))) + \tilde{\delta}_{b_n,n}^{SS_2}$, where,*

1. *for some sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\varepsilon_n = O(\tau_n\sqrt{b_n/n})$, $\sqrt{b_n}P(|\tilde{\delta}_{b_n,n}^{SS_2}| > \varepsilon_n|\mathcal{X}_n) = o(1)$, a.s.,*
2. *$\{(\mathbb{E}_{b_n,n}^{SS}(Z) - \mathbb{E}_n(Z))|\mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero mean sample average of b_n observations sampled without replacement from a distribution with variance covariance matrix $\hat{V}$. Moreover, this distribution has finite third moments, a.s., and $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$,*
3. *$\tilde{H} : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in theorem A.1.*

Part 3. *Assume (A1)-(A4), (CF') and $\Theta_I = \varnothing$. Then, $\liminf\{P(\Gamma_{b_n,n}^{SS_2} = 0|\mathcal{X}_n) = 1\}$, a.s..*

Proof. Part 1. Let $\delta_{b_n,n}^{SS_2}$ be defined as,

$$\delta_{b_n,n}^{SS_2} = \sup_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\left\{\left[\sqrt{b_n}\mathbb{E}_{b_n,n}^{SS}\left(m_j\left(Z,\theta\right)\right)\right]_+\right\}_{j=1}^J\right) - H\left(v_{b_n,n}^{SS}\left(m_\theta\right)\right)$$

where H is the function defined in theorem A.1 and, $\forall\left(\theta,j\right)\in\Theta\times\{1,...,J\}$, $\mathbb{E}_{b_n,n}^{SS}\left(m_j\left(Z,\theta\right)\right) = b_n^{-1}\sum_{i=1}^{b_n} m_j(Z_i^{SS},\theta)$, $v_{b_n,n}^{SS}\left(m_{j,\theta}\right) = \sqrt{b_n}(\mathbb{E}_{b_n,n}^{SS}\left(m_j\left(Z,\theta\right)\right) - \mathbb{E}_n\left(m_j\left(Z,\theta\right)\right))$ and $v_{b_n,n}^{SS}\left(m_\theta\right) = \{v_{b_n,n}^{SS}\left(m_{j,\theta}\right)\}_{j=1}^J$. The empirical process $v_{b_n,n}^{SS}\left(m_\theta\right)$ and the function H satisfy all the requirements of the theorem, so it suffices to show that, $\forall\varepsilon>0$, $P^*(|\delta_{b_n,n}^{SS_2}|>\varepsilon|\mathcal{X}_n) = o\left(1\right)$, a.s.. This is the objective of the rest of this part.

For any $\epsilon\geq 0$, let $\Theta_I\left(\epsilon\right) = \{\theta\in\Theta : \{\mathbb{E}_n\left(m_j\left(Z,\theta\right)\right)\leq\epsilon\}_{j=1}^J\}$ and for any $\gamma>0$, let A_n be the following event,

$$A_n = \left\{\begin{array}{c}\left\{\sup_{\theta\in\Theta}\max_{j=1,\ldots,J}\left|v_{b_n,n}^{SS}\left(m_{j,\theta}\right)\right|\leq\left(\ln\ln b_n\right)^{1/2+\gamma/\beta}\right\}\cap \\ \{\sup_{\theta\in\Theta}\max_{j=1,\ldots,J}\left|v_n\left(m_{j,\theta}\right)\right|\leq\tau_n\}\cap\left\{\Theta_I\subseteq\hat{\Theta}_I\left(\tau_n\right)\subseteq\Theta_I\left(\tau_n\sqrt{b_n}\right)\right\}\end{array}\right\}$$

where β is the degree of homogeneity of the function G in assumption (CF').

Fix $\varepsilon>0$ and consider the following derivation,

$$\begin{aligned}P^*\left(\left|\delta_{b_n,n}^{SS_2}\right|>\varepsilon\middle|\mathcal{X}_n\right) &= P^*\left(\left\{\left|\delta_{b_n,n}^{SS_2}\right|>\varepsilon\right\}\cap A_n\middle|\mathcal{X}_n\right) + P^*\left(\left\{\left|\delta_{b_n,n}^{SS_2}\right|>\varepsilon\right\}\cap\{A_n\}^c\middle|\mathcal{X}_n\right) \\ &\leq P^*\left(\left\{\left|\delta_{b_n,n}^{SS_2}\right|>\varepsilon\right\}\cap A_n\middle|\mathcal{X}_n\right) + P^*\left(\{A_n\}^c\middle|\mathcal{X}_n\right)\end{aligned}$$

By the LIL and lemmas 2.1 and A.14, it follows that $P^*(\{A_n\}^c|\mathcal{X}_n) = o\left(1\right)$, a.s.. To complete the proof of this part, it suffices to show that $P^*(|\delta_{b_n,n}^{SS_2}|>\varepsilon|\mathcal{X}_n\cap A_n) = o\left(1\right)$, a.s.. The strategy to complete this step is to define two random variables, η_n^L and η_n^H, and show that, conditionally on $\{\mathcal{X}_n\cap A_n\}$, they (eventually) constitute lower and upper bounds of $\delta_{b_n,n}^{SS_2}$, respectively, and that they satisfy $P^*\left(\eta_n^L<-\varepsilon|\mathcal{X}_n\cap A_n\right) = o\left(1\right)$, a.s. and $P^*\left(\eta_n^H>\varepsilon|\mathcal{X}_n\cap A_n\right) = o\left(1\right)$, a.s..

Step 1: Upper bound. Define η_n^H as follows,

$$\eta_n^H = \left\{\begin{array}{c}\sup_{\theta\in\Theta_I\left(\tau_n\sqrt{b_n/n}\right)} G\left(\left\{\left[v_{b_n,n}^{SS}\left(m_{j,\theta}\right)+\tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\geq -1/\ln b_n\right]\right\}_{j=1}^J\right) \\ -\sup_{\theta\in\Theta_I} G\left(\left\{\left[v_{b_n,n}^{SS}\left(m_{j,\theta}\right)\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z,\theta\right)\right)=0\right]\right\}_{j=1}^J\right)\end{array}\right\}$$

We now show that, conditionally on A_n, $\exists N\in\mathbb{N}$ such that, $\forall n\geq N$, $\delta_{b_n,n}^{SS}\leq\eta_n^H$. For any $\theta\in\hat{\Theta}_I\left(\tau_n\right)$, consider the following derivation,

$$\begin{aligned}&\left[\sqrt{b_n}\mathbb{E}_{b_n,n}^{SS}\left(m_j\left(Z,\theta\right)\right)\right]_+ \\ &= \left\{\begin{array}{c}\left[v_{b_n,n}^{SS}\left(m_{j,\theta}\right)+\sqrt{b_n/n}\sqrt{n}\mathbb{E}_n\left(m_j\left(Z,\theta\right)\right)\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\geq -1/\ln b_n\right]+ \\ \left[v_{b_n,n}^{SS}\left(m_{j,\theta}\right)+\sqrt{b_n/n}v_n\left(m_{j,\theta}\right)+\sqrt{b_n}\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\leq -1/\ln b_n\right]\end{array}\right\} \\ &\leq \left\{\begin{array}{c}\left[v_{b_n,n}^{SS}\left(m_{j,\theta}\right)+\tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\geq -1/\ln b_n\right]+ \\ \left[v_{b_n,n}^{SS}\left(m_{j,\theta}\right)+\sqrt{b_n/n}\tau_n-\sqrt{b_n}/\ln b_n\right]_+ 1\left[\mathbb{E}\left(m_j\left(Z,\theta\right)\right)\leq -1/\ln b_n\right]\end{array}\right\}\end{aligned}$$

Conditionally on A_n and since $\tau_n\sqrt{b_n/n} = o(1)$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\{v_{b_n,n}^{SS}(m_{j,\theta}) + \tau_n\sqrt{b_n/n} - \sqrt{b_n}/\ln b_n < 0\}$. Thus, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$,

$$\begin{aligned}
&\sup_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\left\{\left[\sqrt{b_n}\mathbb{E}_{b_n,n}^{SS}(m_j(Z,\theta))\right]_+\right\}_{j=1}^J\right) \\
&\leq \sup_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta}) + \tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) \geq -1/\ln b_n\right]\right\}_{j=1}^J\right)
\end{aligned}$$

Finally, to complete the step, notice that, conditional on A_n, $\hat{\Theta}_I(\tau_n) \subseteq \Theta_I(\tau_n\sqrt{b_n/n})$. From this, it follows that, conditionally on A_n, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\delta_{b_n,n}^{SS} \leq \eta_n^H$.

The next step is to show that $P^*\left(\eta_n^H > \varepsilon|\mathcal{X}_n \cap A_n\right) = o(1)$, a.s.. Notice that $\eta_n^H = \eta_n^{H_1} + \eta_n^{H_2} + \eta_n^{H_3}$, where $\eta_n^{H_1}$, $\eta_n^{H_2}$ and $\eta_n^{H_3}$ are defined as follows,

$$\begin{aligned}
\eta_n^{H_1} &= \left\{\begin{array}{c}
\sup_{\theta\in\Theta_I\left(\tau_n\sqrt{b_n/n}\right)} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta}) + \tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) \geq -1/\ln b_n\right]\right\}_{j=1}^J\right) \\
-\sup_{\theta\in\Theta_I} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta}) + \tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) \geq -1/\ln b_n\right]\right\}_{j=1}^J\right)
\end{array}\right\} \\
\eta_n^{H_2} &= \left\{\begin{array}{c}
\sup_{\theta\in\Theta_I} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta}) + \tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) \geq -1/\ln b_n\right]\right\}_{j=1}^J\right) \\
-\sup_{\theta\in\Theta_I} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta}) + \tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) = 0\right]\right\}_{j=1}^J\right)
\end{array}\right\} \\
\eta_n^{H_3} &= \left\{\begin{array}{c}
\sup_{\theta\in\Theta_I} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta}) + \tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) = 0\right]\right\}_{j=1}^J\right) \\
-\sup_{\theta\in\Theta_I} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta})\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) = 0\right]\right\}_{j=1}^J\right)
\end{array}\right\}
\end{aligned}$$

It is then sufficient to show that, $\forall i = 1,2,3$, $P^*(|\eta_n^{H_i}| > \varepsilon/3|\mathcal{X}_n \cap A_n) = o(1)$, a.s.. We only do this for $i = 3$, because the arguments for $i = 1,2$ are identical to those used in the proof of theorem A.3. To prove the argument for $i = 3$, notice that,

$$\left|\eta_n^{H_3}\right| \leq \sup_{\theta\in\Theta}\sup_{s\in\{0,1\}^J}\left|G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta}) + \tau_n\sqrt{b_n/n}\right]_+ s_j\right\}_{j=1}^J\right) - G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta})\right]_+ s_j\right\}_{j=1}^J\right)\right|$$

Conditionally on A_n, we know that $\{||v_{b_n,n}^{SS}(m_{j,\theta})|| < (\ln\ln b_n)^{1/2+\gamma}\}$ and since $(\ln\ln b_n)^{\beta/2+\gamma}\tau_n\sqrt{b_n/n} = o(1)$, a.s., it is sufficient to show that $\exists K > 0$, such that, $\forall\delta \in (0,1)$ and $\forall B > 1$,

$$\sup_{\{x\in\mathbb{R}^J:\|x\|<B\}}\sup_{\{y\in\mathbb{R}^J:\|x-y\|\leq\delta\}}\sup_{s\in\{0,1\}^J}\left|G\left(\left\{[x_j]_+ s_j\right\}_{j=1}^J\right) - G\left(\left\{[y_j]_+ s_j\right\}_{j=1}^J\right)\right| \leq \delta B^\beta K$$

In particular, we will show the result for $K = 3^\beta G(\{1\}_{j=1}^J)$. To this end, fix $\delta \in (0,1)$ and $B > 1$ and make arbitrary choices of $s \in \{0,1\}^J$, $x \in \{x' \in \mathbb{R}^J : ||x'|| < B\}$ and $y \in \{y' \in \mathbb{R}^J : ||x - y'|| \leq \delta\}$. To complete the proof, it suffices to verify that,

$$\left|G\left(\left\{[x_j]_+ s_j\right\}_{j=1}^J\right) - G\left(\left\{[y_j]_+ s_j\right\}_{j=1}^J\right)\right| \leq \delta B^\beta 3^\beta G(\{1\}_{j=1}^J)$$

If $y = x$, this inequality is trivially satisfied, so we assume that $y \neq x$. In this case, we set $w = y + (y - x) / \|y - x\|$, and so, $\{\|w\| \leq 3B\}$ and $y = x / (1 + \|y - x\|) + w \|y - x\| / (1 + \|y - x\|)$. By properties inherited from G, the function $G(\{[x_j]_+ s_j\}_{j=1}^J) : \mathbb{R}^J \to \mathbb{R}_+$ is non-negative, weakly convex, weakly increasing and homogeneous of degree β. Therefore, consider the following derivation,

$$
\begin{aligned}
& G\left(\left\{[y_j]_+ s_j\right\}_{j=1}^J\right) - G\left(\left\{[x_j]_+ s_j\right\}_{j=1}^J\right) \\
& \leq \frac{1}{1 + \|y - x\|} G\left(\left\{[x_j]_+ s_j\right\}_{j=1}^J\right) + \frac{\|y - x\|}{1 + \|y - x\|} G\left(\left\{[w_j]_+ s_j\right\}_{j=1}^J\right) - G\left(\left\{[x_j]_+ s_j\right\}_{j=1}^J\right) \\
& \leq \|y - x\| \left(G\left(\left\{[w_j]_+ s_j\right\}_{j=1}^J\right) - G\left(\left\{[x_j]_+ s_j\right\}_{j=1}^J\right)\right) \\
& \leq \left(3^\beta G\left(\{1\}_{j=1}^J\right)\right) \delta B^\beta
\end{aligned}
$$

The step is completed by reversing the roles of x and y.

Step 2: Lower bound. Define η_n^L as follows,

$$
\eta_n^L = \left\{ \begin{array}{c} \sup_{\theta \in \Theta_I} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta}) - \tau_n \sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) = 0\right]\right\}_{j=1}^J\right) \\ - \sup_{\theta \in \Theta_I} G\left(\left\{\left[v_{b_n,n}^{SS}(m_{j,\theta})\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) = 0\right]\right\}_{j=1}^J\right) \end{array} \right\}
$$

We now show that, conditionally on A_n, $\delta_{b_n,n}^{SS} \geq \eta_n^L$. First, notice that, conditionally on A_n, $\Theta_I \subseteq \hat{\Theta}_I(\tau_n)$. Second, $\forall \theta \in \Theta_I$, consider the following derivation,

$$
\begin{aligned}
\left[\sqrt{b_n}\mathbb{E}_{b_n,n}^{SS}(m_j(Z,\theta))\right]_+ &= \left[v_{b_n,n}^{SS}(m_{j,\theta}) + v_n(m_j(Z,\theta))\sqrt{b_n/n} + \sqrt{n}\mathbb{E}(m_j(Z,\theta))\right]_+ \\
&\geq \left[v_{b_n,n}^{SS}(m_{j,\theta}) + v_n(m_j(Z,\theta))\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) = 0\right] \\
&\geq \left[v_{b_n,n}^{SS}(m_{j,\theta}) - \tau_n\sqrt{b_n/n}\right]_+ 1\left[\mathbb{E}(m_j(Z,\theta)) = 0\right]
\end{aligned}
$$

The combination of these two implies the result.

To complete this step, we need to show that $P^*\left(\eta_n^L > \varepsilon | \mathcal{X}_n \cap A_n\right) = o(1)$, a.s.. This follows from the arguments used in the previous step.

Part 2. By applying arguments used in the proof of theorem A.3, it follows that $\Gamma_{b_n,n}^{SS_2} = \tilde{H}(\sqrt{b_n}(\mathbb{E}_{b_n,n}^{SS}(Z) - \mathbb{E}_n(Z))) + \tilde{\delta}_{b_n,n}^{SS_2}$ where $\tilde{H}$ and $\{\mathbb{E}_{b_n,n}^{SS}(Z) - \mathbb{E}_n(Z)\}$ are the ones required by the theorem. To complete this part, it suffices to show that $\exists C > 0$ such that, $\sqrt{b_n}P(|\tilde{\delta}_{b_n,n}^{SS_2}| > C\tau_n\sqrt{b_n/n}|\mathcal{X}_n) = o(1)$, a.s..

For every $\epsilon \geq 0$, let $\Theta_I(\epsilon) = \{\theta \in \Theta : \{\{p_k(\mathbb{E}(Y_j|x_k) - M_j(\theta, x_k)) \leq \epsilon\}_{j=1}^J\}_{k=1}^K\}$, let the sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ be such that $(\tau_n/\sqrt{n})\varepsilon_n^{-1} = o(1)$ and $\varepsilon_n = o(1)$ a.s., and let A_n be defined as follows,

$$
A_n = \left\{ \begin{array}{c} \left\{\{|\sqrt{n}\hat{p}_k(\mathbb{E}_n(Y_j|x_k) - \mathbb{E}(Y_j|x_k))| \leq \tau_n\}_{j=1}^J\right\}_{k=1}^K \cap \left\{\Theta_I \subseteq \hat{\Theta}_I(\tau_n) \subseteq \Theta_I(\varepsilon_n)\right\} \\ \cap \{\hat{p}_k > p_L/2\}_{k=1}^K \cap \left\{\left|\sqrt{b_n}\left(\hat{p}_k^{SS} - \hat{p}_k\right)\right| < b_n^{3/8}\right\}_{k=1}^K \cap \\ \{\hat{p}_k^{SS} > p_L/2\}_{k=1}^K \cap \left\{\left\{\left|\sqrt{b_n}\hat{p}_k^{SS}\left(\mathbb{E}_{b_n,n}^{SS}(Y_j|x_k) - \mathbb{E}_n(Y_j|x_k)\right)\right| < b_n^{3/8}\right\}_{j=1}^J\right\}_{k=1}^K \end{array} \right\}
$$

For every $C>0$, consider the following derivation,

$$P\left(\left|\tilde{\delta}_{b_n,n}^{SS_2}\right| > C\tau_n\sqrt{b_n/n}\right| \mathcal{X}_n\right) = \left\{\begin{array}{c} P\left(\left\{\left|\tilde{\delta}_{b_n,n}^{SS_2}\right| > C\tau_n\sqrt{b_n/n}\right\}\cap A_n\right| \mathcal{X}_n\right)+ \\ P\left(\left\{\left|\tilde{\delta}_{b_n,n}^{SS_2}\right| > C\tau_n\sqrt{b_n/n}\right\}\cap \{A_n\}^c\right| \mathcal{X}_n\right) \end{array}\right\}$$
$$\leq P\left(\left|\tilde{\delta}_{b_n,n}^{SS_2}\right| > C\tau_n\sqrt{b_n/n}\right| \mathcal{X}_n\cap A_n\right) + P\left(\{A_n\}^c|\mathcal{X}_n\right)$$

Thus, it suffices to show that the two terms on the right hand side are $o(1)$, a.s..

Step 1: Show that $\sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n)=o(1)$, a.s. By elementary properties, it follows that,

$$\sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n)\leq$$
$$\left\{\begin{array}{c} \sqrt{b_n}P\left(\left\{\begin{array}{c}\{|\sqrt{n}\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-\mathbb{E}(Y_j|x_k))|\leq\tau_n\}\cap \\ \left\{\Theta_I\subseteq\hat{\Theta}_I(\tau_n)\subseteq\Theta_I(\varepsilon_n)\right\}\cap\{\hat{p}_k>p_L/2\}_{k=1}^K\end{array}\right\}^c\right|\mathcal{X}_n\right)+ \\ \sum_{j=1}^J\sum_{k=1}^K P\left(\left|\sqrt{b_n}\hat{p}_k^{SS}\left(\mathbb{E}_{b_n,n}^{SS}(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k)\right)\right|\geq b_n^{3/8}|\mathcal{X}_n\right)+ \\ \sum_{k=1}^K\sqrt{b_n}P\left(\left|\sqrt{b_n}\left(\hat{p}_k^{SS}-\hat{p}_k\right)\right|\geq b_n^{3/8}|\mathcal{X}_n\right)+\sum_{k=1}^K\sqrt{b_n}P\left(\hat{p}_k^{SS}\leq p_L/2|\mathcal{X}_n\right)\end{array}\right\}$$

By the LIL and lemma 2.1, it follows that,

$$\liminf\left\{P\left(\left\{\begin{array}{c}\{|\sqrt{n}\hat{p}_k(\mathbb{E}_n(Y_j|x_k)-\mathbb{E}(Y_j|x_k))|\leq\tau_n\}\cap \\ \left\{\Theta_I\subseteq\hat{\Theta}_I(\tau_n)\subseteq\Theta_I(\varepsilon_n)\right\}\cap\{\hat{p}_k>p_L/2\}_{k=1}^K\end{array}\right\}^c\right|\mathcal{X}_n\right)=0\right\},\text{ a.s.}$$

By Shao and Su [27], $\forall(j,k)\in\{1,...,J\}\times\{1,...,K\}$, $\mathbb{E}(\sqrt{b_n}\hat{p}_k^{SS}(\mathbb{E}_{b_n,n}^{SS}(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k))^2|\mathcal{X}_n)$ is bounded by the same expression except that the sampling is done with replacement (as in the bootstrap) rather than without replacement (as in subsampling). By the SLLN, this alternative bound is finite, a.s.. Therefore, by Markov's inequality, we deduce that, $\forall(j,k)\in\{1,...,J\}\times\{1,...,K\}$, $\sqrt{b_n}P(|\sqrt{b_n}\hat{p}_k^{SS}(\mathbb{E}_{b_n,n}^{SS}(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k))|\geq b_n^{3/8}|\mathcal{X}_n)=o(1)$, a.s.. By a similar argument, $\forall k\in\{1,...,K\}$, it follows that $\sqrt{b_n}P(\sqrt{b_n}|\hat{p}_k^{SS}-\hat{p}_k|\geq b_n^{3/8}|\mathcal{X}_n)=o(1)$, a.s., which, in turn, also implies that $\sqrt{b_n}P(\hat{p}_k^{SS}\leq p_L/2|\mathcal{X}_n)=o(1)$, a.s.. The combination of these findings complete this step.

Step 2: Show that $\sqrt{b_n}P(|\tilde{\delta}_{b_n,n}^{SS}|>C\tau_n\sqrt{b_n/n}|\mathcal{X}_n\cap A_n)=o(1)$, a.s.. The strategy to complete this step is to define two random variables, η_n^L and η_n^H, and show that, conditionally on $\{\mathcal{X}_n\cap A_n\}$, they (eventually) constitute lower and upper bounds of $\Gamma_{b_n,n}^{SS_2}$, respectively, and that they satisfy,

$$\max\left\{\begin{array}{c}\eta_n^H-\tilde{H}\left(\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(Z)-\mathbb{E}_n(Z)\right)\right), \\ \tilde{H}\left(\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(Z)-\mathbb{E}_n(Z)\right)\right)-\eta_n^L\end{array}\right\}<C\tau_n\sqrt{b_n/n}$$

Step 2.1: Upper bound. Define η_n^H as follows,

$$\eta_n^H=\sup_{\theta\in\Theta_I}G\left(\left\{\left\{\begin{array}{c}\left[\sqrt{b_n}\hat{p}_k^{SS}\left(\mathbb{E}_{b_n,n}^{SS}(Y_j|x_k)-\mathbb{E}_n(Y_j|x_k)\right)+\tau_n\sqrt{b_n/n}\right]_+ \\ *1\left[p_k\left(\mathbb{E}(Y_j|x_k)-M_j(\theta,x_k)\right)=0\right]\end{array}\right\}_{j=1}^J\right\}_{k=1}^K\right)$$

We now show that, conditional on $\{\mathcal{X}_n \cap A_n\}$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\Gamma^{SS_2}_{b_n,n} \leq \eta^H_n$, a.s.. To show this, notice that,

$$\Gamma^{SS_2}_{b_n,n} = \sup_{\theta \in \hat{\Theta}_I(\tau_n)} G\left(\left\{\left\{\left[\begin{array}{c} \left[\sqrt{b_n}\hat{p}^{SS}_k \left(\mathbb{E}^{SS}_{b_n,n}(Y_j|x_k) - M_j(\theta,x_k)\right)\right]_+ * \\ 1\left[\hat{p}_k\left(\mathbb{E}_n(Y_j|x_k) - M_j(\theta,x_k)\right) \geq -1/\ln b_n\right] + \\ \left[\sqrt{b_n}\hat{p}^{SS}_k \left(\mathbb{E}^{SS}_{b_n,n}(Y_j|x_k) - M_j(\theta,x_k)\right)\right]_+ * \\ 1\left[\hat{p}_k\left(\mathbb{E}_n(Y_j|x_k) - M_j(\theta,x_k)\right) < -1/\ln b_n\right] \end{array}\right]_+\right\}_{j=1}^J\right\}\right)$$

Conditioning on A_n and on $\{\hat{p}_k(\mathbb{E}_n(Y_j|x_k) - M_j(\theta,x_k)) < -1/\ln b_n\}$, $\{\sqrt{b_n}\hat{p}^{SS}_k(\mathbb{E}^{SS}_{b_n,n}(Y_j|x_k) - M_j(\theta,x_k)) \leq b_n^{3/8} - (p_L/2)\sqrt{b_n}/\ln b_n\}$ and $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\{b_n^{3/8} - (p_L/2)\sqrt{b_n}/\ln b_n < 0\}$. Also, conditionally on A_n, $\{|\sqrt{b_n}\hat{p}^{SS}_k(\mathbb{E}_n(Y_j|x_k) - M_j(\theta,x_k))| \leq \tau_n\sqrt{b_n/n}\}$. Thus, conditionally on A_n, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$,

$$\Gamma^{SS_2}_{b_n,n} \leq \sup_{\theta \in \hat{\Theta}_I(\tau_n)} G\left(\left\{\left\{\left[\begin{array}{c} \left[\sqrt{b_n}\hat{p}^{SS}_k\left(\mathbb{E}^{SS}_{b_n,n}(Y_j|x_k) - \mathbb{E}_n(Y_j|x_k)\right) + \tau_n\sqrt{b_n/n}\right]_+ \\ *1\left[\hat{p}_k\left(\mathbb{E}_n(Y_j|x_k) - M_j(\theta,x_k)\right) \geq -(\ln b_n)^{-1}\right] \end{array}\right]_+\right\}_{j=1}^J\right\}_{k=1}^K\right)$$

By applying arguments used in the proof of theorem A.3 (part 2, step 3) to the right hand side of the previous inequality, it follows that, conditionally on $\{\mathcal{X}_n \cap A_n\}$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\Gamma^{SS_2}_{b_n,n} \leq \eta^H_n$, a.s..

Step 2.2. Lower bound. Define η^L_n as follows,

$$\eta^L_n = \sup_{\theta \in \Theta_I} G\left(\left\{\left\{\begin{array}{c} \left[\sqrt{b_n}\hat{p}^{SS}_k\left(\mathbb{E}^{SS}_{b_n,n}(Y_j|x_k) - \mathbb{E}_n(Y_j|x_k)\right) - \tau_n\sqrt{b_n/n}\right]_+ \\ *1\left[p_k\left(\mathbb{E}(Y_j|x_k) - M_j(\theta,x_k)\right) = 0\right] \end{array}\right\}_{j=1}^J\right\}_{k=1}^K\right)$$

Analogous arguments to those used for the upper bound imply that, conditional on $\{\mathcal{X}_n \cap A_n\}$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\Gamma^{SS_2}_{b_n,n} \geq \eta^L_n$, a.s..

Step 2.3: Use the bounds. Let us denote $\tilde{\Gamma}^{SS_2}_{b_n,n} = \tilde{H}(\sqrt{b_n}(\mathbb{E}^{SS}_{b_n,n}(Z) - \mathbb{E}_n(Z)))$. By steps 2.1 and 2.2, it follows that, conditionally on $\{\mathcal{X}_n \cap A_n\}$, $\exists N \in \mathbb{N}$ such that, $\forall n \geq N$, $\tilde{\delta}^{SS}_{b_n,n} \leq \max\{\eta^H_n - \tilde{\Gamma}^{SS_2}_{b_n,n}, \tilde{\Gamma}^{SS_2}_{b_n,n} - \eta^L_n\}$, a.s.. By assumption (CF), $\forall x \in \mathbb{R}^J$ and $\forall \varepsilon > 0$, $\exists C > 0$, such that $\|G(x+\varepsilon) - G(x)\| \leq C\varepsilon$. Therefore, it follows that, $\max\{\eta^H_n - \tilde{\Gamma}^{SS_2}_{b_n,n}, \tilde{\Gamma}^{SS_2}_{b_n,n} - \eta^L_n\} \leq C\tau_n\sqrt{b_n/n}$. This concludes the part.

Part 3. This follows from the same arguments used in the proof of theorem A.3. ■

As a consequence of the representation result, we can establish the consistency in level of this subsampling approximation.

Theorem A.13 (Consistency in level - subsampling approximation 2) *Assume (A1)-(A4) and (CF). If $\Theta_I \neq \varnothing$, then, $\forall \alpha \in (0, 0.5)$,*

$$\lim_{n \to \infty} P\left(\Theta_I \subseteq \hat{C}^{SS_2}_{b_n,n}(1-\alpha)\right) = (1-\alpha)$$

Proof. This proof follows the arguments used in the proof of theorem 2.1. ■

In the remainder of this section, we use the representation result and lemma A.9 to establish an upper bound in the error of this subsampling approximation.

Lemma A.15 *Assume (B1)-(B4), (CF) and that the distribution of the vector $\{\{1[X = x_k]Y_j\}_{j=1}^J\}_{k=1}^K$ is strongly non-lattice.*

Part 1. *If $\Theta_I \neq \varnothing$, then, $\forall \mu > 0$,*

$$\sup_{|h|\geq\mu} \left| P\left(\Gamma^{SS_2}_{b_n,n} \leq h|\mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right| \leq O_p\left(b_n^{-1/2} + \tau_n\sqrt{b_n/n}\right)$$

Part 2. *If $\Theta_I = \varnothing$, then,*

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}} \left| P\left(\Gamma^{SS_2}_{b_n,n} \leq h|\mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right| = 0\right\}\right) = 1$$

Proof. The argument in this proof is exactly the same as the one used in lemma A.10. The only difference is that the sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ that satisfied $\varepsilon_n = O(n^{-1/2})$ is replaced by a positive sequence $\{\varepsilon_n\}_{n=1}^{+\infty}$ that satisfies $\varepsilon_n = O(\tau_n\sqrt{b_n/n})$ and so, $\sqrt{b_n}P(|\tilde{\delta}^{SS_2}_{b_n,n}| > \varepsilon_n|\mathcal{X}_n) = o(1)$, a.s.. After this substitution, $\sup_{|h|\geq\mu}\Phi_{\mathbf{I}_\rho}(\tilde{H}^{-1}((h-\varepsilon_n, h+\varepsilon_n])) = O_p(\tau_n\sqrt{b_n/n})$ and, thus, this term becomes of the leading terms of the approximation. ■

The next step would be to establish the rate of convergence of the approximation of the quantiles, along the lines of corollary A.3. We skip the formulation because it is identical to the one in that corollary, except for the rate, which is now $O_p(b_n^{-1/2} + \tau_n\sqrt{b_n/n})$. Once this result is obtained, we can provide an upper bound on the rate of convergence of the error in the coverage probability for this subsampling approximation.

Theorem A.14 (ECP - subsampling approximation 2) *Assume (B1)-(B4), (CF) and that the distribution of the vector $\{\{1[X = x_k]Y_j\}_{j=1}^J\}_{k=1}^K$ is strongly non-lattice. If $\Theta_I \neq \varnothing$, then, $\forall\alpha \in (0, 0.5)$,*

$$\left|P\left(\Theta_I \subseteq \hat{C}^{SS_2}_{b_n,n}(1-\alpha)\right) - (1-\alpha)\right| = O\left(b_n^{-1/2} + \tau_n\sqrt{b_n/n}\right)$$

Proof. This proof follows the same arguments used in theorem 2.2. ■

Theorem A.14 describes the coverage properties of subsampling 2 when $\Theta_I \neq \varnothing$. In the case when $\Theta_I = \varnothing$, the confidence sets constructed using subsampling 2 present the same coverage properties as the ones shown for the bootstrap in lemma A.7.

The subsampling size that minimizes the upper bound on the rate of convergence of the error in the coverage probability of subsampling 2 is $b_n = O(\sqrt{n}/\tau_n)$ which produces an upper bound on the rate of convergence of order $\tau_n^{1/2}n^{-1/4}$. As a consequence, the minimum upper bound on the rate of convergence of the error in the coverage probability of subsampling 2 is larger than the minimum upper bound on the rate of convergence of the error in the coverage probability of subsampling 1.

Moreover, the next lemma provides conditions under which the rate obtained in theorem A.14 is not just an upper bound, but the exact rate of convergence of the error in coverage probability of subsampling 2. Since this rate is worse than the upper bound on the rate of convergence of the error in coverage probability of subsampling 1, we deduce that, in certain situations, inference based on subsampling 1 is eventually more precise than inference based on subsampling 2.

Lemma A.16 *Suppose that the identified set is given by $\Theta_I = \{\theta \in \Theta : \mathbb{E}(Y) - \theta \leq 0\}$, where Y has a distribution that is continuous with respect to Lebesgue measure and $\mathbb{E}((Y - \mathbb{E}(Y))^3) < 0$. For any $\alpha \in (0, 0.5)$, $\exists C > 0$ and $\exists N \in \mathbb{N}$, such that, $\forall n \geq N$,*

$$\left|P\left(\Theta_I \subseteq C^{SS_2}_{b_n,n}(1-\alpha)\right) - (1-\alpha)\right| > C\left(b_n^{-1/2} + \tau_n\sqrt{b_n/n}\right)$$

Proof. The statistic of interest is given by $\Gamma_n = [\sqrt{n}(\mathbb{E}_n(Y) - \mathbb{E}(Y))]_+$. Thus, for any $h \geq 0$, $P(\Gamma_n \leq h) = P(\sqrt{n}(\mathbb{E}_n(Y) - \mathbb{E}(Y)) \leq h)$ and, by the Berry-Esseén theorem, $P(\Gamma_n \leq h) = \Phi(h) + o\left({b_n}^{-1/2}\right)$, uniformly in $h \geq 0$. By direct computation, $\Gamma^{SS_2}_{b_n,n} = [\sqrt{b_n}(\mathbb{E}^{SS}_{b_n,n}(Y) - \mathbb{E}_n(Y)) + \tau_n\sqrt{b_n/n}]_+$.

For any $h \geq 0$ and any non-negative sequence $\{a_n\}_{n=1}^{+\infty}$ such that $a_n = o(1)$, lemma A.9 implies the following derivation,

$$
\begin{aligned}
& P\left(\left[\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(Y) - \mathbb{E}_n(Y)\right) + a_n\right]_+ \leq h \middle| \mathcal{X}_n\right) \\
&= P\left(\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(Y) - \mathbb{E}_n(Y)\right) \leq h - a_n | \mathcal{X}_n\right) \\
&= \Phi(h - a_n) + \tilde{K}_1(h - a_n) b_n^{-1/2} + \tilde{K}_2(h - a_n) b_n/n + o_p\left(b_n^{-1/2} + b_n/n\right)
\end{aligned}
$$

uniformly in $h \geq 0$, where $\forall s \in \mathbb{R}$, $\tilde{K}_1(s)$ and $\tilde{K}_2(s)$ are the univariate versions of the functions $K_1(s)$ and $K_2(s)$ defined in lemma A.9, which are given by,

$$
\begin{aligned}
\tilde{K}_1(s) &= \mathbb{E}\left((Y - \mathbb{E}(Y))^3/3!\right)\left(s^2/\sqrt{2\pi}\right)\exp\left(-s^2/2\right) \\
\tilde{K}_2(s) &= \Phi(h) + s/\sqrt{2\pi}\exp\left(-s^2/2\right)
\end{aligned}
$$

Since $a_n = o(1)$, a Taylor expansion combined with properties of the functions $\tilde{K}_1$, $\tilde{K}_2$, Φ and ϕ implies that,

$$
\begin{aligned}
\tilde{K}_1(h - a_n) b_n^{-1/2} + \tilde{K}_2(h - a_n) b_n/n &= \tilde{K}_1(h) b_n^{-1/2} + \tilde{K}_2(h) b_n/n + o(b_n^{-1/2} + b_n/n) \\
\Phi(h - a_n) &= \Phi(h) - a_n \phi(h) + o(a_n)
\end{aligned}
$$

and both hold uniformly in $h \in \mathbb{R}$. Therefore, it follows that,

$$
\begin{aligned}
& P\left(\left[\sqrt{b_n}\left(\mathbb{E}_{b_n,n}^{SS}(Y) - \mathbb{E}_n(Y)\right) + a_n\right]_+ \leq h \middle| \mathcal{X}_n\right) \\
&= P(\Gamma_n \leq h) - a_n \phi(h) + o(a_n) + \tilde{K}_1(h) b_n^{-1/2} + \tilde{K}_2(h) b_n/n + o_p\left(b_n^{-1/2} + b_n/n\right)
\end{aligned}
$$

uniformly in $h \geq 0$. If we apply the previous result to $a_n = \tau_n \sqrt{b_n/n}$ we deduce that,

$$
P\left(\Gamma_{b_n,n}^{SS_2} \leq h \middle| \mathcal{X}_n\right) = P(\Gamma_n \leq h) - \tau_n \sqrt{b_n/n}\phi(h) + \tilde{K}_1(h) b_n^{-1/2} + o_p\left(b_n^{-1/2} + \tau_n \sqrt{b_n/n}\right)
$$

uniformly in $h \geq 0$. With this asymptotic representation, we can follow the arguments used in the proof of theorem A.11 to complete the proof. ■

To conclude, the following lemma shows that the confidence sets constructed using subsampling 2 present very desirable coverage properties when $\Theta_I = \varnothing$.

Lemma A.17 *Assume (A1)-(A4), (CF') and $\Theta_I = \varnothing$. Then, $\forall \alpha \in (0, 1)$,*

$$
P\left(\liminf\left\{C_{b_n,n}^{SS_2}(1 - \alpha) = \hat{\Theta}_I(0)\right\}\right) = 1
$$

Proof. This proof follows the arguments used in lemma A.7. ■

A.6.2 Asymptotic approximation

We consider the following asymptotic approximation to perform inference.

1. Choose $\{\tau_n\}_{n=1}^{+\infty}$ to be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln \ln n}/\tau_n = o(1)$, a.s.,
2. Estimate the identified set with $\hat{\Theta}_I(\tau_n) = \left\{\theta \in \Theta : \{\mathbb{E}_n(m_j(Z, \theta)) \leq \tau_n/\sqrt{n}\}_{j=1}^J\right\}$,

3. Repeat the following step for $s = 1, 2, ..., S$. Draw a random sample of standard normal random of size n, denoted by $\{\zeta_i\}_{i=1}^n$, and construct the following stochastic process $\hat{\vartheta} : \Omega_n \to l_J^\infty(\Theta)$,

$$\hat{\vartheta}(\theta) = n^{-1} \sum\nolimits_{i=1}^n \zeta_i \left(m(Z_i, \theta) - \mathbb{E}_n(m(Z, \theta)) \right)$$

where $\{Z_i\}_{i=1}^n = \mathcal{X}_n$. Compute,

$$\Gamma_n^{AA} = \begin{cases} \sup_{\theta \in \hat{\Theta}_I(\tau_n)} G\left(\left\{ \left[\hat{\vartheta}_j(\theta) \right]_+ * 1\left[|\mathbb{E}_n(m_j(Z, \theta))| \le \tau_n/\sqrt{n} \right] \right\}_{j=1}^J \right) & \text{if } \hat{\Theta}_I(\tau_n) \neq \varnothing \\ 0 & \text{if } \hat{\Theta}_I(\tau_n) = \varnothing \end{cases}$$

4. Let $\hat{c}_n^{AA}(1-\alpha)$ be the $(1-\alpha)$ quantile of the distribution of Γ_n^{AA}, simulated with arbitrary accuracy in the previous step. The $(1-\alpha)$ confidence set for the identified set is given by $\hat{C}_n^{AA}(1-\alpha) = \left\{ \theta \in \Theta : \sqrt{n} Q_n(\theta) \le \hat{c}_n^{AA}(1-\alpha) \right\}$.

If the model is conditionally separable, we can choose to use an asymptotic approximation specialized for this framework. In this case, the Gaussian process in step 3 is replaced by a zero-mean normal vector, denoted by $\hat{\vartheta} : \Omega_n \to \mathbb{R}^J$, with variance covariance matrix $\hat{\Upsilon}$, given by,

$$\hat{\Upsilon} = \mathbb{E}_n \left[\left(\left\{ \{1(X = x_k)[Y_j - \mathbb{E}_n(Y_j|x_k)]\}_{j=1}^J \right\}_{k=1}^K \right) \left(\left\{ \{1(X = x_k)[Y_j - \mathbb{E}_n(Y_j|x_k)]\}_{j=1}^J \right\}_{k=1}^K \right)' \right]$$

As usual, we first establish a representation result for our asymptotic approximation.

Theorem A.15 *Part 1. Assume (A1)-(A4), (CF') and $\Theta_I \neq \varnothing$. Then, $\Gamma_n^{AA} = H(\hat{\vartheta}) + \delta_n^{AA}$, where,*

1. *for any $\varepsilon > 0$, $\lim_{n \to +\infty} P^*\left(\left| \delta_n^{AA} \right| > \varepsilon | \mathcal{X}_n \right) = 0$, a.s.,*

2. *$\{\hat{\vartheta}(\theta) | \mathcal{X}_n\} : \Omega_n \to l_J^\infty(\Theta)$ is an stochastic process that converges weakly to the same Gaussian process as in theorem A.1, i.p.,*

3. *$H : l_J^\infty(\Theta) \to \mathbb{R}$ is the same function as in theorem A.1.*

Part 2. Let ρ denote the rank of the variance covariance matrix of the vector $\{\{1[X = x_k] Y_j\}_{j=1}^J\}_{k=1}^K$. If we assume (B1)-(B4), (CF), $\Theta_I \neq \varnothing$, and we choose the asymptotic approximation procedure to be the one specialized for the conditionally separable model, then, $\Gamma_n^{AA} = \tilde{H}(\tilde{\vartheta}) + \tilde{\delta}_n^{AA}$, where,

1. *$P(\tilde{\delta}_n^{AA} = 0 | \mathcal{X}_n) = 1[\tilde{\delta}_n^{AA} = 0]$ and $\liminf\{\tilde{\delta}_n^{AA} = 0\}$, a.s.,*

2. *$\{\tilde{\vartheta} | \mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero mean normally distributed vector with variance covariance matrix $\hat{V}$. Moreover, this distribution has finite third moments, a.s., and $||\hat{V} - \mathbf{I}_\rho|| \le O_p(n^{-1/2})$,*

3. *$\tilde{H} : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in theorem A.1.*

Part 3. Assume (A1)-(A4), (CF') and $\Theta_I = \varnothing$. Then, $\liminf\{P(\Gamma_n^{AA} = 0 | \mathcal{X}_n) = 1\}$, a.s..

Proof. This proof follows the proof of theorem A.3 very closely. The only main difference to point out occurs in the proof of part 1..

Part 1. In the proof of theorem A.3, we used the CLT for bootstrapped empirical processes applied to a P-Donsker class. In this proof, this step is replaced with the argument in remark 4.2 of CHT [10]. ■

We now establish the consistency of the asymptotic approximation.

Theorem A.16 (Consistency of asymptotic approximation excluding zero) *Assume (A1)-(A4) and (CF').*

Part 1. If $\Theta_I \neq \varnothing$, then, $\forall \mu > 0$ and $\forall \varepsilon > 0$,

$$\lim_{n\to+\infty} P^*\left(\sup_{|h|\geq\mu}\left|P\left(\Gamma_n^{AA}\leq h|\mathcal{X}_n\right) - \lim_{m\to+\infty} P\left(\Gamma_m \leq h\right)\right| \leq \varepsilon\right) = 1$$

Part 2. If $\Theta_I = \varnothing$, then,

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}}\left|P\left(\Gamma_n^{AA}\leq h|\mathcal{X}_n\right) - \lim_{m\to+\infty} P\left(\Gamma_m \leq h\right)\right| = 0\right\}\right) = 1$$

Proof. This proof follows the arguments used in the proof of theorem A.5. ■

With the consistency of the approximation, we can establish the consistency in level of the inference based on the asymptotic approximation. The theorem is formulated in the main text.

Proof. [**Theorem 2.5**] This proof follows the arguments in the proof of theorem 2.1. ■

We now deduce the rate of convergence of the asymptotic approximation.

Theorem A.17 (Rate of convergence - asymptotic approximation) *Assume (B1)-(B4) and (CF).*

Part 1. If $\Theta_I \neq \varnothing$, then,

$$\sup_{|h|\geq\mu}\left|P\left(\Gamma_n^{AA}\leq h|\mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right| \leq O_p(n^{-1/2})$$

Part 2. If $\Theta_I = \varnothing$, then,

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}}\left|P\left(\Gamma_n^{AA}\leq h|\mathcal{X}_n\right) - P\left(\Gamma_n \leq h\right)\right| = 0\right\}\right) = 1$$

Proof. This proof follows the arguments of the proof of theorem A.7. ■

Based on the rate of convergence, we can establish the upper bound on the rate of convergence of the error in the coverage probability. This theorem is formulated in the main text.

Proof. [**Theorem 2.6**] This proof follows the arguments used in the proof of theorem 2.2. ■

Theorem 2.6 describes the coverage properties of the asymptotic approximation when $\Theta_I \neq \varnothing$. In the case when $\Theta_I = \varnothing$, the confidence sets constructed using the asymptotic approximation present the same coverage properties as the ones shown for the bootstrap in lemma A.7.

A.7 Monte Carlo simulations

In order to evaluate the finite sample behavior of the different inferential methods, we consider two sets of Monte Carlo simulations.

In the first set of simulations, we propose two abstract partially identified models that represent relatively simple subsets of the real line. These designs are purposely chosen so that the proposed bootstrap procedure provides consistent inference in level and the naive bootstrap described in section A.2.5 does not. Within these designs, we perform an extensive study to understand how changes in the numerous parameters of the simulations affect the results.

The second set of Monte Carlo simulations are performed in a well known econometric model, namely, a probit model with missing data. This constitutes a more realistic framework where we can compare the performance of the inferential procedures considered in the paper.

Parameter	Values used
F	$F_1 : (Y_1, Y_2) \sim N(0, \mathbf{I}_2)$, $F_2 : (Y_1, Y_2) \sim N(0, (1, 0.5; 0.5, 1))$ $F_3 : (Y_1, Y_2) \sim N(0, (1, -0.5; -0.5, 1))$ $F_4 : Y_1 \sim t_3,\ Y_2 \sim t_3,\ Y_1 \perp Y_2$
n	100, 1000
τ_n	0, $\ln\ln n$, $\ln n$, $n^{1/8}$, $n^{1/4}$
b_n	$\{20,\ 33,\ 50\}$ for $n = 100$ $\{200,\ 333,\ 500\}$ for $n = 1000$

Table 1: Monte Carlo designs

A.7.1 Abstract partially identified models

We consider the following two designs,

$$\begin{array}{ll} \text{design 1} & \Theta_I = \{\theta \in \Theta : \mathbb{E}(Y_1) \leq \theta \leq \mathbb{E}(Y_2)\} \\ \text{design 2} & \Theta_I = \{\theta \in \Theta : \{\mathbb{E}(Y_1) \leq \theta \cap \mathbb{E}(Y_2) \leq \theta\}\} \end{array}$$

where, in particular, $\mathbb{E}(Y_1) = 0$ and $\mathbb{E}(Y_2) = 0$.

The data are an i.i.d. sample of size n from a distribution that is denoted by F. To implement our inference, we use the criterion function $Q(\theta) = \sum_{j=1}^{2} [\mathbb{E}(m_j(Z, \theta))]_+$, which satisfies assumption (CF). Each number presented in the tables is the result of 10,000 Monte Carlo simulations. In each simulation, the distribution of the bootstrap, subsampling and asymptotic approximation are approximated from (the same) 1,000 Monte Carlo draws[29]. In order to implement any of our inferential procedures, we need to specify the distribution F, the sample size n and the sequence $\{\tau_n\}_{n=1}^{+\infty}$. Finally, the subsampling procedures also require the choice of the subsampling size sequence $\{b_n\}_{n=1}^{+\infty}$. Table 1 shows all the values used for each of these parameters.

We briefly comment on some of the choices for our parameters. We consider four different bivariate distributions F. In the first three distributions, we use normal random vectors, with zero, positive and negative correlation. This distribution has finite moments of all orders and thus satisfies the moment assumptions required by the conditionally separable model. The fourth bivariate distribution produces independent pairs of Student t-distributed random variables with three degrees of freedom. This distribution has infinite fourth absolute moment, violating assumption (B4). In this case, we are interested in understanding the effect of fat tails on our coverage results.

We conduct simulations with a relatively small sample size ($n = 100$) and relatively big sample size ($n = 1,000$). For each value of the sample size, we let the sequence τ_n vary among five different values. According to our theoretical results, choosing $\tau_n = 0$ will, in general, not produce consistent inference in level. All other proposed choices for τ_n, that is, $\ln\ln n$, $\ln n$, $n^{1/8}$ and $n^{1/4}$ should result in consistent inference in level.

Our tables present the percentage of times of our confidence sets cover the identified set when the desired coverage level is 95%. The columns in the table represent each of the approximation schemes: B denotes bootstrap, AA denotes asymptotic approximation, $\mathrm{SS}_1(b_n)$ denotes subsampling 1 with subsampling size equal to b_n and $\mathrm{SS}_2(b_n)$ denotes subsampling 2 with subsampling size equal to b_n.

Design 1 In this design, the identified set is given by $\Theta_I = \{\theta \in \Theta : \{\mathbb{E}(Y_1) \leq \theta \leq \mathbb{E}(Y_2)\}\}$, where $\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = 0$, and so, $\Theta_I = \{0\}$. Conducting inference in this design is challenging because the

[29] We have also produced simulations with $S = 100$ and $S = 200$ but they produced very similar results.

Empirical coverage for $(1-\alpha) = 95\%$

F	τ_n	B	AA	$\text{SS}_1\,(20)$	$\text{SS}_1\,(33)$	$\text{SS}_1\,(50)$	$\text{SS}_2\,(20)$	$\text{SS}_2\,(33)$	$\text{SS}_2\,(50)$
F_1	0	48.9%	48.5%	47.4%	47.4%	48.0%	48.0%	49.1%	49.1%
F_1	$\ln\ln n$	93.8%	93.8%	81.8%	86.2%	90.4%	98.1%	98.4%	98.4%
F_1	$\ln n$	93.8%	93.8%	81.8%	86.2%	90.4%	100%	100%	99.8%
F_1	$n^{1/4}$	93.8%	93.8%	81.8%	86.2%	90.4%	99.0%	99.0%	99.0%
F_1	$n^{1/8}$	93.8%	93.8%	81.8%	86.2%	90.4%	100%	100%	99.4%
F_2	0	44.4%	45.1%	41.9%	43.4%	43.6%	43.7%	45.8%	46.1%
F_2	$\ln\ln n$	91.0%	91.0%	78.5%	84.4%	85.8%	96.3%	96.4%	96.9%
F_2	$\ln n$	91.0%	91.0%	78.5%	84.4%	85.8%	100%	100%	99.9%
F_2	$n^{1/4}$	91.0%	91.0%	78.5%	84.4%	85.8%	98.0%	97.4%	97.0%
F_2	$n^{1/8}$	91.0%	91.0%	78.5%	84.4%	85.8%	99.9%	99.9%	99.4%
F_3	0	47.4%	47.3%	47.3%	47.4%	47.4%	47.4%	47.4%	47.4%
F_3	$\ln\ln n$	95.1%	95.3%	86.0%	90.2%	92.7%	89.9%	90.1%	88.8%
F_3	$\ln n$	95.1%	95.3%	86.0%	90.2%	92.7%	99.7%	98.8%	98.0%
F_3	$n^{1/4}$	95.1%	95.3%	86.0%	90.2%	92.7%	91.4%	91.3%	90.5%
F_3	$n^{1/8}$	95.1%	95.3%	86.0%	90.2%	92.7%	98.0%	97.6%	95.6%
F_4	0	47.7%	48.0%	46.1%	47.1%	47.7%	48.3%	48.5%	48.8%
F_4	$\ln\ln n$	87.8%	87.8%	80.0%	83.9%	86.2%	89.1%	89.3%	89.3%
F_4	$\ln n$	94.1%	94.1%	80.6%	86.8%	91.1%	100%	100%	99.4%
F_4	$n^{1/4}$	90.9%	90.9%	80.2%	84.9%	88.2%	92.6%	92.8%	92.6%
F_4	$n^{1/8}$	94.1%	94.1%	80.6%	86.8%	91.1%	99.9%	99.9%	99.3%

Table 2: Results of first Monte Carlo design with $n = 100$

identified set is non-empty but has an empty interior.

Table 2 provides the simulation results with the smaller sample size. Before analyzing each approximation scheme separately, we note that all of them suffer from severe undercoverage when $\tau_n = 0$. The reason for this undercoverage lies in the peculiar structure of the design. Even though the identified set is non-empty, the estimator of the identified set with $\tau_n = 0$, that is, $\hat{\Theta}_I(0)$, is empty with positive probability. Thus, any inferential procedure that uses $\tau_n = 0$ will severely undercover the identified set.

We now turn to the analysis of each approximation method for positive levels of τ_n. We begin with subsampling schemes. Relative to the rest of the methods, subsampling 1 seems to produce undercoverage and subsampling 2 seems to produce overcoverage. The undercoverage of subsampling 1 holds for all of the distributions and seems to become worse as we decrease the subsampling size. The empirical coverage for subsampling 1 is relatively insensitive to the particular value of τ_n. This is expected, because the effect of the choice of τ_n in this subsampling procedure is limited to the estimation of the identified set and to the indicator functions in the criterion function. As long as the value of τ_n is such that the estimated identified set is non-empty and the appropriate indicator functions are turned on or off, the particular value of the statistic is insensitive to this parameter.

Except for the relatively low values of τ_n, subsampling 2 suffers from a severe overcoverage of the identified set, which gets worse as τ_n increases. This can be explained by our analysis in section A.2.5. Given that subsampling 2 has no recentering term, the expression $\tau_n\sqrt{b_n/n}$ will appear in the subsampling criterion function. This term has no asymptotic effect but, in small samples, it causes the criterion function of subsampling 2 to be larger than desired.

Relative to the subsampling approximations, the bootstrap and the asymptotic approximation produce a coverage frequency that is closer to the desired coverage level. Moreover, both procedures are

Empirical coverage for $(1-\alpha) = 95\%$

F	τ_n	B	AA	$\text{SS}_1\,(200)$	$\text{SS}_1\,(333)$	$\text{SS}_1\,(500)$	$\text{SS}_2\,(200)$	$\text{SS}_2\,(333)$	$\text{SS}_2\,(500)$
F_1	0	50.6%	50.6%	49.7%	49.7%	50.5%	50.5%	50.8%	50.8%
F_1	$\ln\ln n$	92.9%	93.5%	82.9%	87.7%	90.9%	95.5%	94.1%	92.9%
F_1	$\ln n$	92.9%	93.5%	82.9%	87.7%	90.9%	100%	100%	99.9%
F_1	$n^{1/4}$	92.9%	93.5%	82.9%	87.7%	90.9%	96.6%	96.4%	95.5%
F_1	$n^{1/8}$	92.9%	93.5%	82.9%	87.7%	90.9%	100%	99.9%	99.9%
F_2	0	50.6%	52.0%	50.3%	51.0%	51.6%	51.9%	52.7%	52.9%
F_2	$\ln\ln n$	93.1%	92.8%	80.7%	85.0%	89.5%	97.3%	96.8%	96.2%
F_2	$\ln n$	93.1%	92.8%	80.7%	85.0%	89.5%	100%	100%	100%
F_2	$n^{1/4}$	93.1%	92.8%	80.7%	85.0%	89.5%	98.0%	97.9%	97.4%
F_2	$n^{1/8}$	93.1%	92.8%	80.7%	85.0%	89.5%	100%	100%	99.9%
F_3	0	50.3%	50.3%	50.3%	50.3%	50.3%	50.3%	50.3%	50.3%
F_3	$\ln\ln n$	93.0%	92.8%	85.1%	88.1%	91.0%	91.0%	89.4%	89.0%
F_3	$\ln n$	93.0%	92.8%	85.1%	88.1%	91.0%	100%	99.6%	99.2%
F_3	$n^{1/8}$	93.0%	92.8%	85.1%	88.1%	91.0%	92.7%	91.7%	91.4%
F_3	$n^{1/4}$	93.0%	92.8%	85.1%	88.1%	91.0%	99.6%	99.4%	98.9%
F_4	0	51.6%	51.6%	49.3%	50.1%	50.5%	51.2%	51.6%	51.7%
F_4	$\ln\ln n$	91.9%	91.3%	82.5%	86.2%	88.9%	92.9%	92.9%	92.8%
F_4	$\ln n$	92.9%	92.9%	82.6%	86.5%	89.5%	100%	100%	100%
F_4	$n^{1/4}$	92.7%	92.7%	82.5%	86.4%	89.2%	97.4%	97.4%	97.3%
F_4	$n^{1/8}$	92.9%	92.9%	82.6%	86.5%	89.5%	100%	100%	99.8%

Table 3: Results of first Monte Carlo design with $n = 1,000$

producing an approximation of similar quality, which seems to be in line with our analysis regarding the rates of convergence. In this design, both procedures seem to be slightly undercovering the identified set. We rationalize this in the following way. The identified set is defined by two moment inequalities which are binding. Our inferential procedures need to learn this structure from the sample. If sampling error introduces a mistake in the number of sample moment inequalities that are considered to be binding, this can only result in underestimation of this number. As a result, there will be a tendency to undercoverage of the identified set. Also, as expected, the coverage results of both procedures are relatively insensitive to the value of τ_n. This has the same explanation as in the case of subsampling 1.

Table 3 presents the results of the first design with the larger sample size. The results of these simulations are similar to the ones with smaller sample size. Once again, all inferential procedures with $\tau_n = 0$ have undercoverage problems, by exactly the same reasons as before. For positive levels of τ_n, subsampling 1 produces undercoverage, subsampling 2 produces overcoverage and our bootstrap and our asymptotic approximation still produce better results than any of the subsampling schemes, with a slight tendency to undercover. According to our results, it appears that increasing the sample size from 100 to 1,000 does not change the quality of any of the approximations.

Design 2 In this design, the identified set is given by $\Theta_I = \{\theta \in \Theta : \{\mathbb{E}(Y_1) \leq \theta\} \cap \{\mathbb{E}(Y_2) \leq \theta\}\}$, where $\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = 0$, and so, $\Theta_I = [0, +\infty)$. Therefore, the identified set is non-empty and has non-empty interior. This design imposes a challenge in the sense that two moment inequalities are binding in the population but, for all of the sampling distributions we consider, at most one these inequalities will be binding in the sample, a.s..

Table 4 provides the simulation results for our smaller sample size. The results are similar to the

Empirical coverage for $(1-\alpha) = 95\%$

F	τ_n	B	AA	SS_1 (20)	SS_1 (33)	SS_1 (50)	SS_2 (20)	SS_2 (33)	SS_2 (50)
F_1	0	83.9%	83.3%	72.3%	77.3%	79.4%	76.0%	81.2%	85.1%
F_1	$\ln\ln n$	94.1%	93.5%	84.8%	89.1%	91.7%	97.3%	97.4%	97.0%
F_1	$\ln n$	94.9%	94.8%	85.1%	90.4%	92.6%	100%	100%	100%
F_1	$n^{1/4}$	94.4%	94.2%	85.0%	89.8%	92.0%	98.6%	98.3%	98.1%
F_1	$n^{1/8}$	94.9%	94.8%	85.1%	90.4%	92.6%	100%	100%	100%
F_2	0	81.1%	82.2%	72.2%	76.5%	78.4%	79.5%	85.2%	88.6%
F_2	$\ln\ln n$	93.9%	95.0%	86.6%	90.9%	93.0%	97.9%	97.9%	97.9%
F_2	$\ln n$	93.9%	95.0%	86.8%	91.0%	93.0%	100%	100%	100%
F_2	$n^{1/4}$	93.9%	95.0%	86.8%	91.0%	93.0%	99.1%	99.0%	98.2%
F_2	$n^{1/8}$	93.9%	95.0%	86.8%	91.0%	93.0%	100%	100%	100%
F_3	0	87.2%	87.6%	70.3%	78.5%	82.2%	72.0%	80.9%	84.6%
F_3	$\ln\ln n$	92.6%	93.2%	81.5%	87.1%	89.6%	97.5%	97.6%	96.7%
F_3	$\ln n$	95.1%	95.5%	82.7%	89.1%	92.1%	100%	100%	100%
F_3	$n^{1/4}$	94.3%	94.8%	82.5%	88.4%	91.5%	98.2%	98.4%	98.0%
F_3	$n^{1/8}$	95.1%	95.5%	82.7%	89.1%	92.1%	100%	100%	99.9%
F_4	0	83.7%	84.6%	71.4%	77.1%	81.3%	74.1%	80.9%	85.6%
F_4	$\ln\ln n$	89.7%	90.6%	78.7%	83.4%	87.6%	91.6%	92.8%	93.2%
F_4	$\ln n$	93.5%	95.1%	84.0%	88.6%	92.0%	99.9%	100%	100%
F_4	$n^{1/4}$	90.4%	91.6%	80.2%	84.3%	87.9%	93.3%	93.5%	94.1%
F_4	$n^{1/8}$	93.2%	95.0%	83.8%	88.2%	91.6%	99.0%	99.2%	98.8%

Table 4: Results of second Monte Carlo design with $n = 100$

ones obtained in the first design. When $\tau_n = 0$, all of the inferential procedures produce extreme undercoverage. This is expected, because if we set $\tau_n = 0$, a moment inequality will be considered to be binding if and only if it is satisfied with equality in the sample. Therefore, with probability one, the sample moment inequalities will never be simultaneously binding even though there is a point in the identified set for which their population analogues are simultaneously binding. As explained in section A.2.5, this problem is related to the inconsistency of the bootstrap in the boundary of the parameter space, studied by Andrews [1].

For positive values of τ_n, subsampling 1 tends to undercover the identified set and subsampling 2 tends to overcover the identified set. In the case of subsampling 1, the undercoverage error is smaller than the one found in the previous design. Our bootstrap and our asymptotic approximation present a very good finite sample performance, which is much better than the performance obtained with any of the subsampling procedures.

Table 5 presents the results of the second design with the larger sample size. Once again, the results of this simulations are similar to those obtained with the smaller sample size. By the same reasons as before, all simulations with $\tau_n = 0$ have undercoverage problems. The coverage results for positive levels of τ_n are similar to the ones obtained in table 4. Subsampling 1 produces undercoverage and subsampling 2 produces overcoverage. Increasing the sample size does not seem to improve the quality of the subsampling approximations. Our bootstrap and our asymptotic approximation still produce very accurate results that are better than any of the subsampling schemes.

Empirical coverage for $(1-\alpha) = 95\%$

F	τ_n	B	AA	SS_1 (200)	SS_1 (333)	SS_1 (500)	SS_2 (200)	SS_2 (333)	SS_2 (500)
F_1	0	82.8%	82.7%	71.5%	76.2%	79.5%	74.8%	80.2%	84.1%
F_1	$\ln\ln n$	93.2%	93.6%	83.6%	88.4%	90.7%	99.4%	98.8%	99.0%
F_1	$\ln n$	93.5%	93.9%	83.6%	88.6%	91.0%	100%	100%	100%
F_1	$n^{1/4}$	93.5%	93.9%	83.6%	88.6%	91.0%	99.6%	99.6%	99.0%
F_1	$n^{1/8}$	93.5%	93.9%	83.6%	88.6%	91.0%	100%	100%	100%
F_2	0	81.8%	81.3%	72.6%	76.2%	79.0%	78.2%	82.7%	87.8%
F_2	$\ln\ln n$	93.8%	94.1%	87.2%	90.4%	91.8%	99.4%	99.0%	98.9%
F_2	$\ln n$	93.8%	94.1%	87.2%	90.4%	91.8%	100%	100%	100%
F_2	$n^{1/4}$	93.8%	94.1%	87.2%	90.4%	91.8%	99.7%	99.6%	99.1%
F_2	$n^{1/8}$	93.8%	94.1%	87.2%	90.4%	91.8%	100%	100%	100%
F_3	0	87.0%	87.5%	72.0%	78.8%	82.8%	74.3%	80.1%	84.1%
F_3	$\ln\ln n$	94.9%	94.8%	82.3%	88.3%	91.8%	99.1%	99.0%	98.6%
F_3	$\ln n$	95.1%	95.1%	82.6%	88.6%	92.3%	100%	100%	100%
F_3	$n^{1/4}$	95.1%	95.1%	82.6%	88.6%	92.3%	99.5%	99.4%	99.3%
F_3	$n^{1/8}$	95.1%	95.1%	82.6%	88.6%	92.3%	100%	100%	100%
F_4	0	81.8%	80.5%	69.7%	74.1%	78.7%	72.1%	78.9%	83.0%
F_4	$\ln\ln n$	91.1%	90.6%	79.8%	85.2%	88.4%	93.2%	94.6%	95.2%
F_4	$\ln n$	95.2%	94.9%	81.9%	87.0%	91.1%	100%	100%	100%
F_4	$n^{1/4}$	92.3%	91.6%	81.0%	85.4%	88.9%	95.8%	96.4%	97.2%
F_4	$n^{1/8}$	95.2%	94.9%	81.9%	87.0%	91.1%	100%	100%	100%

Table 5: Results of second Monte Carlo design with $n = 1,000$

		Covariate values		
		$x_1 = (1, 0)$	$x_2 = (0, 1)$	$x_3 = (1, 1)$
Design 1	$\mathbb{E}(YW\|x)$	$\Phi(-0.5)$	$\Phi(-0.5)$	$\Phi(-0.5)$
	$\mathbb{E}(W\|x)$	$2\Phi(-0.5)$	$2\Phi(-0.5)$	$2\Phi(-0.5)$
		$x_1 = (1, 0)$	$x_2 = (0, 1)$	$x_3 = (1, 1)$
Design 2	$\mathbb{E}(YW\|x)$	$\Phi(-0.5)$	$\Phi(-0.5)$	$\Phi(-1)$
	$\mathbb{E}(W\|x)$	$2\Phi(-0.5)$	$2\Phi(-0.5)$	$\Phi(-1) + \Phi(-0.5)$
		$x_1 = (1, 0)$	$x_2 = (0, 1)$	$x_3 = (-1, 0)$
Design 3	$\mathbb{E}(YW\|x)$	$\Phi(-0.5)$	$\Phi(-0.5)$	$\Phi(-0.5)$
	$\mathbb{E}(W\|x)$	$\Phi(-0.5) + \Phi(0)$	$2\Phi(-0.5)$	$\Phi(-0.5) + \Phi(0)$
		$x_1 = (1, 0)$	$x_2 = (0, 1)$	$x_3 = (-1, 0)$
Design 4	$\mathbb{E}(YW\|x)$	$\Phi(-0.5)$	$\Phi(0)$	$\Phi(-0.5)$
	$\mathbb{E}(W\|x)$	$\Phi(-0.5) + \Phi(0.1)$	$\Phi(0) + \Phi(-1)$	$\Phi(-0.5) + \Phi(0.1)$

Table 6: Monte Carlo designs

A.7.2 Probit model with missing data

For our second set of simulations, consider a binary choice model with missing data. Suppose that we are interested in the decision of individuals between two mutually exclusive and exhaustive choices: choice 0 or choice 1. Let Y denote this choice, which is assumed to be generated by $Y = 1[X\beta \geq \varepsilon]$, where X is a vector of observable explanatory variables with support denoted by S_X, ε is an unobservable explanatory variable and β denotes the parameters of interest. Assume that $\varepsilon \sim N(0, 1)$ independent of X, which implies that we adopt the probit model. Therefore, $P(Y = 1|X = x) = \mathbb{E}(Y|X = x) = \Phi(x\beta)$.

Suppose that the covariates are observed for every respondent but, for some respondents, we do not observe the choice. Denote by W the variable that takes value one if the choice is observed and zero otherwise. The identified set is given by,

$$\Theta_I = \left\{\beta \in \Theta : \{\mathbb{E}(YW|x) \leq \Phi(x\beta) \leq \mathbb{E}(YW + (1 - W)|x)\}_{x \in S_X}\right\}$$

We consider four Monte Carlo designs which differ in the definition of S_X and in the value of $\{\mathbb{E}(YW|x), \mathbb{E}(W|x)\}_{x \in S_X}$. These designs are described in table 6.

For all simulations we will sample $n = 600$ observations, with 100 observations for the first covariate, 200 observations for the second covariate and 300 observations for the third covariate. For each value of the covariate, we sample $\{Y|X\}$ and $\{W|X\}$ independently from a Bernoulli distribution with the mean specified by table 6.

To implement our inference, we use the criterion function $Q(\theta) = \sum_{j=1}^{J}[\mathbb{E}(m_j(Z, \theta))]_+$, which satisfies assumption (CF). Each number presented in the tables is the result of 1,000 Monte Carlo simulations. In each simulation, the distribution of the bootstrap, subsampling and asymptotic approximation are approximated from (the same) 200 Monte Carlo draws.

In order to implement any of the inferential procedures, we need to specify the sequence $\{\tau_n\}_{n=1}^{+\infty}$. For all of the procedures, we conducted simulations with $\tau_n = \ln\ln n$ and $\tau_n = \ln n$ and we obtained similar results. From this experience, we conjecture that the results are relatively robust to the choice of the sequence $\{\tau_n\}_{n=1}^{+\infty}$. For the sake of brevity, our tables only report results for $\tau_n = \ln\ln n$. The subsampling procedures also require specifying the subsampling size sequence $\{b_n\}_{n=1}^{+\infty}$. For the sake of

Empirical coverage for different values of $(1-\alpha)$				
Procedure	75%	90%	95%	99%
Subsampling 1 ($b_n = 300$)	47.5%	66.3%	75.9%	87.9%
Subsampling 1 ($b_n = 200$)	57.7%	77.5%	85.9%	94.7%
Subsampling 2 ($b_n = 300$)	100%	100%	100%	100%
Subsampling 2 ($b_n = 200$)	100%	100%	100%	100%
Our bootstrap	74.9%	89.8%	95.4%	99.0%
Our asymptotic approximation	74.2%	89.5%	95.0%	99.6%

Table 7: Results of first Monte Carlo design

brevity, we show the results for $b_n = 300$ and $b_n = 200$ but the results for other choices of subsampling size produced qualitatively similar results.

Design 1 The identified set is characterized by a pair of moment inequalities for each of the three covariate values. Combining these restrictions, the identified set is the one depicted in figure 1.

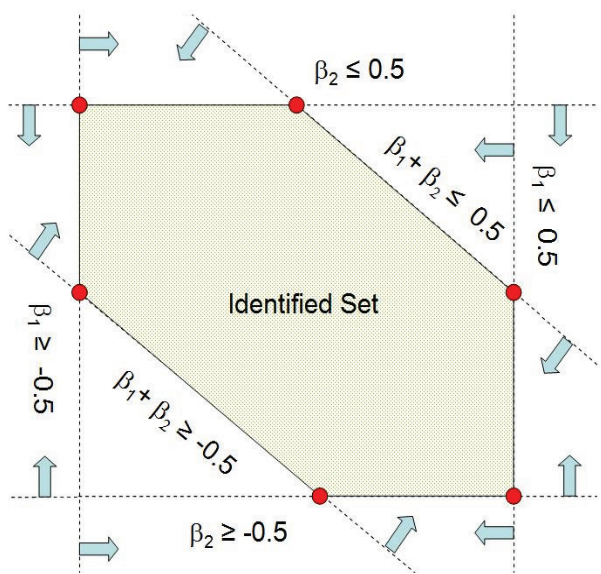

Figure 1: Identified set for first Monte Carlo design

The distinctive characteristic of this design is that the identified set has non-empty interior everywhere and that the boundaries of the identified set are defined by, at most, two constrains satisfied with equality. As a consequence, in this particular case, we can obtain consistent inference using bootstrap, subsampling or asymptotic approximation even if we set $\tau_n = 0$.

Table 7 presents the empirical coverage for each inferential procedure. All of the subsampling procedures exhibit a mediocre finite sample behavior. Subsampling 1 undercovers the identified set and subsampling 2 overcovers the identified set. The analysis of section A.2.5 explains that the overcoverage of subsampling 2 could be a consequence of what we refer as the expansion problem. The bootstrap and the asymptotic approximation proposed in this paper achieve a very satisfactory performance.

Empirical coverage for different values of $(1-\alpha)$				
Procedure	75%	90%	95%	99%
Subsampling 1 ($b_n = 200$)	43.4%	64.3%	73.3%	88.3%
Subsampling 1 ($b_n = 300$)	55.6%	74.7%	84.3%	93.8%
Subsampling 2 ($b_n = 200$)	100%	100%	100%	100%
Subsampling 2 ($b_n = 300$)	100%	100%	100%	100%
Our bootstrap	75.5%	91.6%	95.9%	99.0%
Our asymptotic approximation	75.0%	91.8%	95.4%	99.0%

Table 8: Results of second Monte Carlo design

Design 2 The identified set in this design is described in figure 2. As in the previous design, the identified set has non-empty interior everywhere. The difference with respect to the previous design is that there is one point in the identified set, namely the point $(\beta_1, \beta_2) = (-0.5, -0.5)$, where one of the restrictions, $\beta_1 + \beta_2 \geq -1$, is both irrelevant and satisfied with equality.

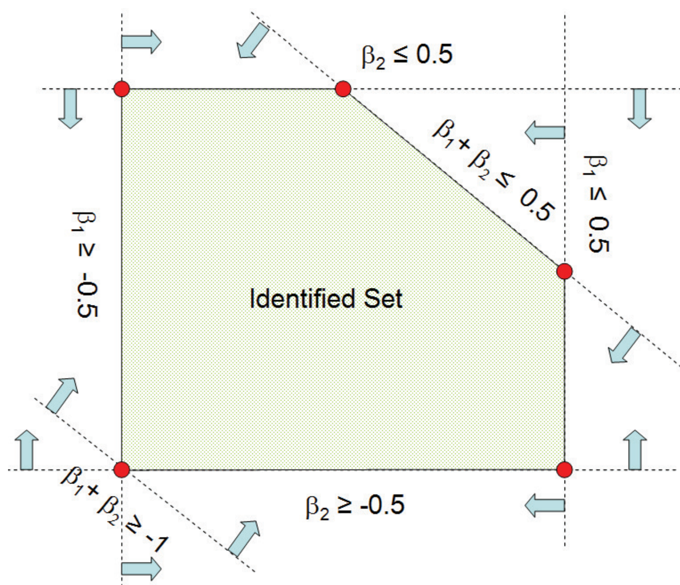

Figure 2: Identified set for the second Monte Carlo design

The results are presented in table 8. The subsampling procedures have a mediocre finite sample behavior: subsampling 1 suffers from undercoverage and subsampling 2 suffers from overcoverage. Our bootstrap and our asymptotic approximation exhibit a satisfactory performance.

Design 3 Figure 3 describes the identified set in this design. This design differs from the previous two in that the identified set has empty interior and the analogy principle estimator of the identified set is empty with positive probability. This illustrates why we need to artificially expand the analogy principle estimator in order to generate an estimator of the identified set that is adequate for the purpose of inference.

The results are given in table 9. As usual, the subsampling procedures have a mediocre finite sample behavior: subsampling 1 suffers from undercoverage and subsampling 2 suffers from overcoverage. Our bootstrap and our asymptotic approximation procedures produce a satisfactory finite sample performance.

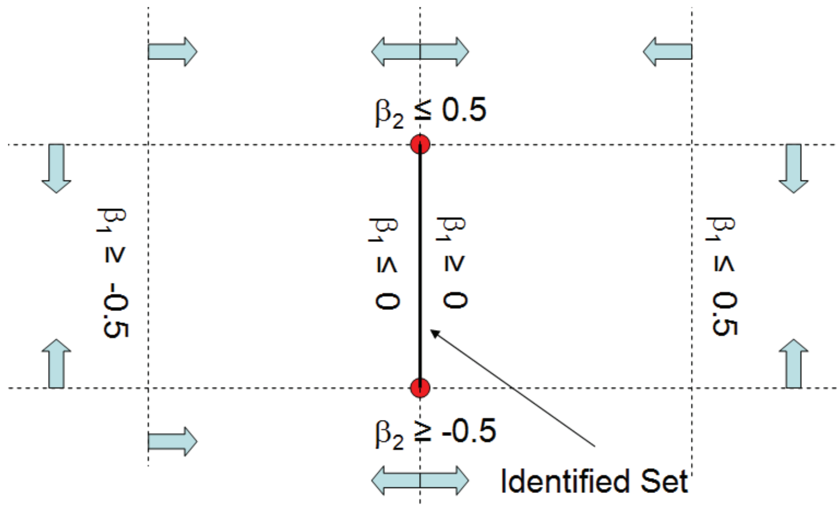

Figure 3: Identified set for third Monte Carlo design

Empirical coverage for different values of $(1-\alpha)$				
Procedure	75%	90%	95%	99%
Subsampling 1 ($b_n = 200$)	45.6%	45.6%	45.6%	45.7%
Subsampling 1 ($b_n = 300$)	45.6%	45.6%	45.7%	45.7%
Subsampling 2 ($b_n = 200$)	100%	100%	100%	100%
Subsampling 2 ($b_n = 300$)	100%	100%	100%	100%
Our bootstrap	76.2%	89.9%	95.7%	98.8%
Our asymptotic approximation	76.4%	90.5%	95.8%	98.9%

Table 9: Results of third Monte Carlo design

Design 4 In this case, the identified set is empty or, equivalently, the model is misspecified. Since the identified set is empty, the empirical coverage is trivially 100%. Therefore, in this design, we will compare the relative sizes of the confidence sets for different inferential methods. In order to achieve this task, we need to define a measure of size of the confidence sets generated by the different inferential methods. For any confidence set $C_n \subseteq \Theta$, we consider the following function,

$$\Pi\left(C_n\right) = \begin{cases} \sup_{\theta \in C_n} \left\{\sqrt{n} Q_n\left(\theta\right)\right\} & \text{if } C_n \neq \varnothing \\ 0 & \text{if } C_n = \varnothing \end{cases}$$

It is not hard to show that the function Π constitutes a metric for confidence sets generated by the criterion function approach.

Table 10 presents the average value of Π for each of the inferential procedures. Not surprisingly, the relative sizes of these confidence sets are in line with the results obtained in the previous designs. Subsampling 1 produces confidence sets that are relatively small and subsampling 2 produces confidence sets that are relatively big. Our bootstrap procedure and our asymptotic approximation generate confidence sets in between these two.

Average Π-size of confidence set for different values of $(1-\alpha)$				
Procedure	75%	90%	95%	99%
Subsampling 1 ($b_n = 200$)	0.14	0.16	0.17	0.19
Subsampling 1 ($b_n = 300$)	0.13	0.15	0.16	0.18
Subsampling 2 ($b_n = 200$)	1.98	2.12	2.22	2.38
Subsampling 2 ($b_n = 300$)	1.81	1.98	2.09	2.30
Our bootstrap	0.54	0.74	0.87	1.11
Our asymptotic approximation	0.54	0.75	0.88	1.11

Table 10: Results of fourth Monte Carlo design