Supplementary Appendix to "Generic Inference on Quantile and Quantile Effect Functions for Discrete Outcomes"

Victor Chernozhukov (MIT) Ivan Fernandez-Val (Boston University)
Blaise Melly (University of Bern) Kaspar Wüthrich (UC San Diego)

May 9, 2019

A Imposing Monotonicity and Range Restrictions on Estimates and Confidence Bands for Distribution Functions

In many applications the point estimates $\hat{F}$ and interval estimates $[L', U']$ for the target distribution F do not satisfy logical monotonicity or range restrictions, namely they do not take values in the set $\mathbb{D}$ defined in Section 2 of the main text. Given such an ordered triple $L' \leq \hat{F} \leq U'$, we can always transform it into another ordered triple $L \leq \check{F} \leq U$ that obeys the logical monotonicity and shape restrictions. For example, we can set

$$\check{F} = \mathcal{S}(\hat{F}), \quad L = \mathcal{S}(L'), \quad U = \mathcal{S}(U'), \tag{A.1}$$

where $\mathcal{S}$ is the shaping operator that given a function $y \mapsto f(y)$ yields a mapping $y \mapsto \mathcal{S}(f)(y) \in \mathbb{D}$ with

$$\mathcal{S}(f) = \mathcal{M}(0 \vee f \wedge 1),$$

where the maximum and minimum are taken pointwise, and $\mathcal{M}$ is the rearrangement operator that given a function $f : \mathcal{Y} \mapsto [0,1]$ yields a map $y \mapsto \mathcal{M}(f)(y) \in \mathbb{D}$. Other monotonization operators, such as the projection on the set of weakly increasing functions, can also be used, as we remark further below.

The *rearrangement operator* is defined as follows. Let T be a countable subset of $\mathcal{Y}$. In our leading case where f is the distribution function of a discrete random variable Y, we can choose T as the support of Y and extend f to $\mathcal{Y}$ by constant interpolation, yielding a step function as the distribution of Y on $\mathcal{Y}$. If f is a distribution function of a continuous or mixed random variable Y, we can set T as a grid of values covering the support of Y where we evaluate f and extend f to $\mathcal{Y}$ by linear interpolation. Given a $f : T \mapsto [0,1]$, we first consider $\mathcal{M}f$ as a vector of sorted values of the set $\{f(t) : t \in T\}$, where the sorting is done in a non-decreasing order. Since T is an ordered set of the same cardinality as $\mathcal{M}f$, we can assign the elements of $\mathcal{M}f$ to T in one-to-one manner: to the k-th smallest element of T we assign the k-th smallest element of $\mathcal{M}f$. The resulting mapping $t \mapsto \mathcal{M}f(t)$ is the rearrangement operator. We can extend the rearranged function $\mathcal{M}f$ to $\mathcal{Y}$ by constant or linear interpolation as we describe above.

The following lemma shows that shape restrictions *improve* the finite-sample properties of the estimators and confidence bands.

Lemma 1 (Shaping Improves Point and Interval Estimates)**.** *The shaping operator* $\mathcal{S}$

(a) is weakly contractive under the max distance:

$$\|\mathcal{S}(A) - \mathcal{S}(B)\|_\infty \leq \|A - B\|_\infty, \quad \textit{for any } A,\ B\textit{: } T \to [0,1],$$

(b) is shape-neutral,

$$\mathcal{S}(F) = F \textit{ for any } F \in \mathbb{D},$$

(c) and preserves the partial order:

$$A \leq B \implies \mathcal{S}(A) \leq \mathcal{S}(B), \quad \text{for any } A,\ B\colon T \to [0,1].$$

Consequently,

1. *the re-shaped point estimate constructed via (A.1) is weakly closer to F than the initial estimate under the max distance:*

$$\|\check{F} - F\|_\infty \leq \|\hat{F} - F\|_\infty,$$

2. *the re-shaped confidence band constructed via (A.1) has weakly greater coverage than the initial confidence band:*

$$\mathrm{P}(L' \leq F \leq U') \leq \mathrm{P}(L \leq F \leq U),$$

3. *and the re-shaped confidence band is weakly shorter than the original confidence bands under the max distance,*

$$\|U - L\|_\infty \leq \|U' - L'\|_\infty.$$

Proof. The result follows from Chernozhukov et al. (2009). ∎

The band $[L, U]$ is therefore weakly better than the original band $[L', U']$, in the sense that coverage is preserved while the width of the confidence band is weakly shorter.

Remark 7 (Isotonization is Another Option)**.** An alternative to the rearrangement is the isotonization, which projects a given function on the set of weakly increasing functions that map T to $[0, 1]$. This also has the improving properties stated in Lemma 1. In fact any convex combination between isotonization and rearrangement has the improving properties stated in Lemma 1. ∎

Remark 8 (Shape Restrictions on Confidence Bands by Intersection)**.** An alternative way of imposing shape restrictions on the confidence band, is to intersect the initial band $[L', U']$ with $\mathbb{D}$. That is, we simply set

$$[L^I, U^I] = \mathbb{D} \cap [L', U'] = \{w \in \mathbb{D} : L'(y) \leq w(y) \leq U'(y), \quad \forall y \in \mathcal{Y}\}.$$

Thus, U^I is the greatest nondecreasing minorant of $0 \vee U' \wedge 1$ and L^I is the smallest nodecreasing majorant of $0 \vee L' \wedge 1$. This approach gives the tightest confidence bands, in particular

$$[L^I, U^I] \subseteq [L, U].$$

However, this construction might be less robust to misspecification than the rearrangement. For example, imagine that the target function F is not monotone, i.e. $F \notin \mathbb{D}$. This situation might arise when F is the probability limit of some estimator $\hat{F}$ that is inconsistent for the DF due to misspecification. If the confidence band $[L', U']$ is sufficiently tight, then we can end up with an empty intersection band, $[L^I, U^I] = \emptyset$. By contrast $[L, U]$ is non-empty and covers the reshaped target function $F^* = \mathcal{S}(F) \in \mathbb{D}$. ■

B Bootstrap Algorithms for Confidence Bands for Single Quantile Functions

If one is only interested in a single QF $F^{\leftarrow}$, the QF-band constructed based on Algorithm 1 in the main text will generally be conservative. Here, we provide an algorithm that provides asymptotically similar (non-conservative) uniform confidence bands that jointly cover the DF, F, and the corresponding QF, $F^{\leftarrow}$.

Algorithm 2 (Bootstrap Algorithm for Single QF-Band)**.**

1. Obtain many bootstrap draws of the estimator $\hat{F}$,

$$\hat{F}^{*(j)}, \quad j = 1, \ldots, B$$

where the index j enumerates the bootstrap draws and B is the number of bootstrap draws (e.g., $B = 1,000$).

2. For each y in T, compute the robust standard error of $\hat{F}(y)$,

$$\hat{s}(y) = (\hat{Q}(.75, y) - \hat{Q}(.25, y))/(\Phi^{-1}(.75) - \Phi^{-1}(.25)),$$

where $\hat{Q}(\alpha, y)$ denotes the empirical α-quantile of the bootstrap sample $(\hat{F}^{*(j)}(y))_{j=1}^{B}$, and Φ^{-1} denotes the inverse of the standard normal distribution.

3. Compute the critical value

$$c(p) = \ p\text{-quantile of } \left\{ \max_{y \in T} |\hat{F}(y)^{*(j)} - \hat{F}(y)|/\hat{s}(y) \right\}_{j=1}^{B}.$$

4. Construct a preliminary DF-band $[L', U']$ for F of level p via: $[L'(y), U'(y)] = [\hat{F}(y) \pm c(p)\hat{s}(y)]$ for each $y \in T$. Impose the shape restrictions on $\hat{F}$, L' and U' as described in Appendix A. Report $I = [L, U]$ as a p-level DF-band for F.

5. Report the inverted band $I^{\leftarrow} = [U^{\leftarrow}, L^{\leftarrow}]$ or support restricted inverted band $\tilde{I}^{\leftarrow} = I^{\leftarrow} \cap T$ as a p-level QF-band for $F^{\leftarrow}$

The following corollary of Theorem 1 in the main text provides theoretical justification for Algorithm 2.

Corollary 4 (Validity of Algorithm 2). *Suppose that the rescaled DF estimator $a_n(\hat{F} - F)$ converges in law in $\ell^\infty(\mathcal{Y})$ to a Gaussian process G, having zero mean and a non-degenerate variance function, for some sequence of constants $a_n \to \infty$ as $n \to \infty$, where n is some index (typically the sample size). Suppose that a bootstrap method can consistently approximate the limit law of $a_n(\hat{F} - F)$, namely the distance between the law of $a_n(\hat{F}^* - \hat{F})$ conditional on data, and that of G, converges to zero in probability as $n \to \infty$. The distance is the bounded Lipschitz metric that metrizes weak convergence. Then,*

$$\lim_{n\to\infty} \mathrm{P}(F \in I, F^{\leftarrow} \in \tilde{I}^{\leftarrow}) = p.$$

Proof. Lemma SA.1 of Chernozhukov et al. (2013) implies that $\lim_{n\to\infty} \mathrm{P}(F \in [L', U']) = p$. The result then follows from Lemma 1, Theorem 1 and Corollary 1 in the main text. ∎

C Simulation Study

This section presents simulation evidence on the finite sample performance of our bands. To keep the simulations computationally tractable we analyze a setup without covariates. We generate two independent random samples $\{Y_{1i}\}_{i=1}^n$ and $\{Y_{0i}\}_{i=1}^n$ for the treated and control outcomes, respectively. The estimators of the DFs $\hat{F}_{Y_0}$ and $\hat{F}_{Y_1}$ are simply the empirical distribution functions in the respective samples. We perform 5000 simulations and let the sample size $n \in \{400, 1600, 6400\}$ vary in order to examine the convergence of the coverage rates with respect to the sample size. We consider the problem of constructing uniform confidence bands that cover (i) a single QF: either $F_{Y_1}^{\leftarrow}$ or $F_{Y_0}^{\leftarrow}$, (ii) simultaneously both DFs, both QFs and the QE function: F_{Y_0}, F_{Y_1}, $F_{Y_0}^{\leftarrow}$, $F_{Y_1}^{\leftarrow}$ and $F_{Y_1}^{\leftarrow} - F_{Y_0}^{\leftarrow}$, (iii) only the QE function: $F_{Y_1}^{\leftarrow} - F_{Y_0}^{\leftarrow}$. The confidence bands for a coverage of type (i) are constructed based on Algorithm 2 while the bands for a coverage of type (ii) or (iii) are constructed based on Algorithm 1 in the main text. We consider three confidence levels $p \in \{0.9, 0.95, 0.99\}$.

We consider two families of distributions: a count variable similar to the outcome in the first application and an ordered variable similar to the outcome in the second application. In the first case, Y_{1i} is distributed Poisson with parameter $\lambda = 3$ and Y_{0i} is distributed Poisson with $\lambda \in \{3, 2.75, 2.5\}$. Since the support of the Poisson distribution is unbounded, we estimate the QFs and QE functions for $a \in [0.1, 0.9]$ and invert the bands for the DF over the part of the support that is relevant for the range of quantiles considered. Table 1 displays the empirical coverage rate of the true functions. We report the coverage rate of the DFs and QFs in a single column because they are numerically equal by construction. We also provide the empirical probability to reject the null hypothesis that $F^{\leftarrow}_{Y_1} = F^{\leftarrow}_{Y_0}$. This allows us to measure the empirical size in the first panel (where this hypothesis is satisfied) and the empirical power in the other panels.

The empirical coverage rates of the bands for a single QF (in the third and fourth columns of Table 1) as well as the coverage rate for both DFs, both QFs and the QE function (in the fifth column) confirms the theoretical results in corollaries 3 and 4. The empirical coverage rates are very close to the intended confidence levels p. The bands for these parameters are not conservative. We know from Theorem 2 in the main text that the bands for the QE function are valid but may be conservative when the goal is to cover only the QE function independently from the other functions. One of the objectives of the simulations is to assess if our QE-bands are narrow enough to be informative. The results in the sixth column of Table 1 show that the coverage rate of the bands is indeed larger than the theoretical coverage rate p when the true QE function is uniformly 0 (design 1) but is very close to p when the distributions of the treated and control outcomes are different. The reason for this result is that the Minkowski difference of two non-conservative confidence sets for two QF is *not* conservative for the difference in the parameters when (at least) one of the confidence set is a *singleton*. While this case is irrelevant for continuous outcomes, it often happens for discrete outcomes. As it can be seen for instance in Figures 4 or 5 in the main text, the confidence bands for the QFs contains a single value at many probability indices. Asymptotically, the bands for the QF of a discrete outcome will contain a single value at all quantiles except in the neighborhoods of the quantiles at which the QF jumps. Thus, asymptotically our bands for the QE function are not conservative except for the case when the QFs of Y_1 and Y_0 are identical, i.e. when $F_1^{\leftarrow} = F_0^{\leftarrow}$ uniformly. The second and third panels of the last column in Table 1 provide the empirical power of our bands to reject the null hypothesis that $F_1^{\leftarrow} = F_0^{\leftarrow}$. Even quite small deviations from the null hypothesis are detected with relatively moderate sample sizes. As expected, the power increases with the sample size and with the deviation from the null hypothesis.

Table 2 presents the results for an ordered outcome. Y_0 and Y_1 are both discretized random Gaussian variables that can take the values $\{0, 1, ..., 5\}$. Y_1 is based on a latent standard Gaussian random variable while we consider three different latent variables for Y_0: $N(0, 1)$, $N(0.2, 1)$ and $N(0.4, 1)$. The cut-off values are the same for both outcomes. They are chosen such that Y_1 takes the values $\{0, 1, ..., 5\}$ with probability $\{0.1, 0.16, 0.24, 0.24, 0.16, 0.1\}$ respectively. The results are extremely similar to the results in Table 1: (i) the coverage rates for a single QF are very close to the intended coverage rate, (ii) the coverage rate for all QFs, DFs and the QE function is also very close to the intended rate, (iii) the coverage rate for the QE function is higher than the intended rate only when the true QE function is uniformly 0, (iv) the power of our bands to reject an incorrect null hypothesis is substantial and increases in the sample size and the deviation from the null hypothesis.

While our bands are–to the best of our knowledge–the only ones that have been proven to cover uniformly the QFs and the QE functions of discrete outcomes, applied researchers may be tempted to use alternative heuristic approaches. For this reason, we compare the performance of our bands for the QE function with four alternative methods.[1] We first experiment with directly bootstrapping the QE function and calculating sup-t bands. However, the pointwise standard errors obtained via bootstrap are numerically equal to zero at many quantiles such that the t-statistic cannot be computed. We tried putting a lower bound on the pointwise standard errors to be able to calculate the t statistics but this resulted in extremely wide bands that always covered the true function. For this reason we do not report these results in the following tables. The second approach that we consider consists in bootstrapping the QE function and calculating constant width bands. This method avoids the need to divide by the estimated pointwise standard errors and could therefore be implemented. The last two approaching are based on jittering (adding random noise) as suggested by Machado and Silva (2005) for count outcomes. We bootstrap the QE function of the smoothed outcomes and construct sup-t bands centered either around the smoothed QE function or around the original, unsmoothed QE function. Machado and Silva (2005) show that standard methods can be used to make inference about the smoothed quantile function. On the contrary, we are interested in covering the original, unsmoothed QE function.

The results for the count outcomes are provided in Table 3, which compares the coverage probability of our new bands with that of the competing bands as well as the average length of the bands.[2] The constant width bands obtained by bootstrapping directly the QE function are very conservative in all cases. Their average length is two to four times

[1] The results for the QFs are not shown because they are similar.

[2] The computation time of these alternative methods is so high that we decided to not perform the simulations with $6,400$ observations.

higher than the average length of the bands that we have suggested. This bad behavior of the bootstrap for the QF of a discrete outcome comes at no surprise because it is known to be inconsistent for the estimation of the pointwise variance. Huang (1991) finds in simulations that the bootstrap grossly overestimate the variance of the sample median of a discrete outcome, except when the QF jumps exactly at the median. The estimators based on jittering have the opposite problem: their coverage rate is below the intended rate and is even equal to zero for many distributions. The reason is simple: adding noise smoothes the differences over the whole range of quantiles such that the variance is underestimated where the QF jumps but is overestimated where the QF is flat. Note that these results do not contradict the results in Machado and Silva (2005), which consider the smoothed QF as the true function, but show that adding noise to the outcome cannot help covering the unsmoothed QF. Table 4 presents the results of the simulations for the ordered outcomes. The conclusion are similar: bootstrapping the QE function directly leads to very wide bands while bootstrapping the jittered QE function leads to extreme undercoverage of the true function.

To summarize, for both types of discrete outcomes we come to the conclusion that the alternative methods either do not cover the true QE function with at least the chosen coverage rate or are much longer than the suggested bands. As an interesting by-product of these simulations, we note that the average length of our bands converges to zero at the $\sqrt{n}$-rate.

References

Chernozhukov, V., Fernandez-Val, I., Galichon, A., 2009. Improving point and interval estimators of monotone functions by rearrangement. Biometrika 96 (3), 559–575.

Chernozhukov, V., Fernández-Val, I., Melly, B., 2013. Inference on counterfactual distributions. Econometrica 81 (6), 2205–2268.

Huang, J., 1991. Estimating the variance of the sample median, discrete case. Statistics & probability letters 11 (4), 291–298.

Machado, J. A. F., Silva, J. S., 2005. Quantiles for counts. Journal of the American Statistical Association 100 (472), 1226–1237.

Table 1: Performance of the uniform bands for count outcomes

n	p	Empirical coverage probability for				Prob. to reject
		F_0, $F_0^{\leftarrow}$	F_1, $F_1^{\leftarrow}$	all fct.	$F_1^{\leftarrow} - F_0^{\leftarrow}$	$F_1^{\leftarrow} = F_0^{\leftarrow}$
Design 1: $Y_0 \sim Poisson(3)$ and $Y_1 \sim Poisson(3)$						
400	0.99	0.99	0.99	0.99	1.00	0.00
400	0.95	0.96	0.96	0.96	1.00	0.00
400	0.90	0.92	0.92	0.92	1.00	0.00
1600	0.99	0.99	0.99	0.99	1.00	0.00
1600	0.95	0.96	0.96	0.96	1.00	0.00
1600	0.90	0.92	0.91	0.92	1.00	0.00
6400	0.99	0.99	0.99	0.99	1.00	0.00
6400	0.95	0.96	0.95	0.95	1.00	0.00
6400	0.90	0.91	0.91	0.91	1.00	0.00
Design 2: $Y_0 \sim Poisson(3)$ and $Y_1 \sim Poisson(2.75)$						
400	0.99	0.99	0.99	0.99	1.00	0.00
400	0.95	0.96	0.96	0.96	0.98	0.03
400	0.90	0.92	0.92	0.92	0.94	0.07
1600	0.99	0.99	0.99	0.99	0.99	0.19
1600	0.95	0.96	0.95	0.95	0.96	0.49
1600	0.90	0.92	0.91	0.91	0.91	0.65
6400	0.99	0.99	0.99	0.99	0.99	1.00
6400	0.95	0.96	0.96	0.95	0.95	1.00
6400	0.90	0.91	0.92	0.91	0.91	1.00
Design 3: $Y_0 \sim Poisson(3)$ and $Y_1 \sim Poisson(2.5)$						
400	0.99	0.99	0.99	0.99	0.99	0.16
400	0.95	0.96	0.96	0.96	0.97	0.47
400	0.90	0.92	0.91	0.92	0.92	0.64
1600	0.99	0.99	0.99	0.99	0.99	1.00
1600	0.95	0.96	0.96	0.96	0.96	1.00
1600	0.90	0.92	0.92	0.92	0.92	1.00
6400	0.99	0.99	0.99	0.99	0.99	1.00
6400	0.95	0.96	0.96	0.95	0.95	1.00
6400	0.90	0.91	0.92	0.91	0.91	1.00

Notes: Based on 5,000 simulations.

Table 2: Performance of the uniform bands for ordered outcomes

n	p	Empirical coverage probability for				Prob. to reject
		F_0, $F_0^{\leftarrow}$	F_1, $F_1^{\leftarrow}$	all fct.	$F_1^{\leftarrow}-F_0^{\leftarrow}$	$F_1^{\leftarrow}=F_0^{\leftarrow}$
Design 1: $Y_0^* \sim N(0,1)$ and $Y_1^* \sim N(0,1)$						
400	0.99	0.98	0.98	0.98	1.00	0.00
400	0.95	0.94	0.94	0.93	1.00	0.00
400	0.90	0.89	0.88	0.88	1.00	0.00
1600	0.99	0.99	0.99	0.99	1.00	0.00
1600	0.95	0.94	0.95	0.94	1.00	0.00
1600	0.90	0.89	0.90	0.89	1.00	0.00
6400	0.99	0.99	0.99	0.99	1.00	0.00
6400	0.95	0.95	0.95	0.94	1.00	0.00
6400	0.90	0.89	0.90	0.90	1.00	0.00
Design 2: $Y_0^* \sim N(0,1)$ and $Y_1^* \sim N(0.2,1)$						
400	0.99	0.98	0.98	0.98	0.99	0.02
400	0.95	0.94	0.93	0.93	0.95	0.11
400	0.90	0.89	0.88	0.88	0.90	0.21
1600	0.99	0.99	0.99	0.98	0.98	0.62
1600	0.95	0.94	0.95	0.94	0.94	0.88
1600	0.90	0.89	0.89	0.89	0.89	0.95
6400	0.99	0.99	0.99	0.99	0.99	1.00
6400	0.95	0.95	0.95	0.95	0.95	1.00
6400	0.90	0.89	0.90	0.89	0.89	1.00
Design 3: $Y_0^* \sim N(0,1)$ and $Y_1^* \sim N(0.4,1)$						
400	0.99	0.98	0.98	0.98	0.99	0.60
400	0.95	0.94	0.94	0.93	0.94	0.88
400	0.90	0.89	0.88	0.88	0.88	0.95
1600	0.99	0.99	0.99	0.98	0.98	1.00
1600	0.95	0.94	0.94	0.94	0.94	1.00
1600	0.90	0.89	0.89	0.89	0.89	1.00
6400	0.99	0.99	0.99	0.99	0.99	1.00
6400	0.95	0.95	0.95	0.95	0.95	1.00
6400	0.90	0.89	0.90	0.90	0.90	1.00

Notes: Based on 5,000 simulations.

Table 3: Comparison with alternative bands for the QE fct.: count outcomes

n	p	Coverage probability of the band:				Average length of the band:			
		new	boot.	jitter1	jitter2	new	boot.	jitter1	jitter2
Design 1: $Y_0 \sim Poisson(3)$ and $Y_1 \sim Poisson(3)$									
400	0.99	1.00	1.00	0.99	0.00	1.61	4.00	1.46	1.46
400	0.95	1.00	1.00	0.96	0.00	1.32	3.99	1.20	1.20
400	0.90	1.00	1.00	0.93	0.00	1.18	3.96	1.08	1.08
1600	0.99	1.00	1.00	0.99	0.01	0.75	3.98	0.67	0.67
1600	0.95	1.00	1.00	0.96	0.01	0.63	3.93	0.56	0.56
1600	0.90	1.00	1.00	0.92	0.01	0.57	3.75	0.51	0.51
Design 2: $Y_0 \sim Poisson(3)$ and $Y_1 \sim Poisson(2.75)$									
400	0.99	1.00	1.00	0.01	0.00	1.56	3.99	1.43	1.43
400	0.95	0.98	1.00	0.00	0.00	1.28	3.82	1.18	1.18
400	0.90	0.94	1.00	0.00	0.00	1.15	3.46	1.05	1.05
1600	0.99	0.99	1.00	0.00	0.00	0.73	2.77	0.66	0.66
1600	0.95	0.96	1.00	0.00	0.00	0.61	2.25	0.55	0.55
1600	0.90	0.91	1.00	0.00	0.00	0.55	2.08	0.50	0.50
Design 3: $Y_0 \sim Poisson(3)$ and $Y_1 \sim Poisson(2.5)$									
400	0.99	0.99	1.00	0.02	0.00	1.52	3.91	1.41	1.41
400	0.95	0.97	1.00	0.00	0.00	1.25	3.30	1.16	1.16
400	0.90	0.92	1.00	0.00	0.00	1.13	2.75	1.04	1.04
1600	0.99	0.99	1.00	0.00	0.00	0.72	2.35	0.65	0.65
1600	0.95	0.96	1.00	0.00	0.00	0.61	2.07	0.54	0.54
1600	0.90	0.92	1.00	0.00	0.00	0.55	2.02	0.49	0.49

Notes: Based on $5,000$ simulations.

Table 4: Comparison with alternative bands for QE: ordered outcomes

n	p	Coverage probability of the band:				Average length of the band:			
		new	boot.	jitter1	jitter2	new	boot.	jitter1	jitter2
Design 2: $Y_0^* \sim N(0,1)$ and $Y_1^* \sim N(0,1)$									
400	0.99	1.00	1.00	0.99	0.01	1.34	4.00	1.46	1.46
400	0.95	1.00	1.00	0.96	0.00	1.12	4.00	1.19	1.19
400	0.90	1.00	1.00	0.91	0.00	1.01	3.98	1.07	1.07
1600	0.99	1.00	1.00	0.99	0.00	0.66	3.99	0.64	0.64
1600	0.95	1.00	1.00	0.95	0.00	0.56	3.95	0.54	0.54
1600	0.90	1.00	1.00	0.91	0.00	0.51	3.82	0.49	0.49
Design 2: $Y_0^* \sim N(0,1)$ and $Y_1^* \sim N(0.2,1)$									
400	0.99	0.99	1.00	0.07	0.01	1.33	3.94	1.49	1.49
400	0.95	0.95	1.00	0.01	0.00	1.11	3.58	1.21	1.21
400	0.90	0.90	1.00	0.00	0.00	1.01	3.17	1.08	1.08
1600	0.99	0.98	1.00	0.00	0.00	0.66	2.34	0.65	0.65
1600	0.95	0.94	1.00	0.00	0.00	0.56	2.06	0.54	0.54
1600	0.90	0.89	1.00	0.00	0.00	0.50	2.01	0.49	0.49
Design 3: $Y_0^* \sim N(0,1)$ and $Y_1^* \sim N(0.4,1)$									
400	0.99	0.99	1.00	0.12	0.01	1.33	3.87	1.57	1.57
400	0.95	0.94	1.00	0.01	0.00	1.10	3.10	1.25	1.25
400	0.90	0.88	1.00	0.00	0.00	1.00	2.61	1.11	1.11
1600	0.99	0.98	1.00	0.00	0.00	0.65	2.03	0.67	0.67
1600	0.95	0.94	1.00	0.00	0.00	0.55	2.00	0.56	0.56
1600	0.90	0.89	1.00	0.00	0.00	0.50	2.00	0.50	0.50

Notes: Based on 5,000 simulations.