Supplementary Material

Nelly K. Djuazon Emmanuel Selorm Tsyawo

S.1 Proofs of results in the main text

S.1.1 Proof of Theorem 1

First, by the strict monotonicity (invertibility) of $\Phi(\cdot)$ assumed in Assumption 1,

$$\mathrm{F}_{Y_s(0)|G_g=0,G_g+G_h=1}(y) = \Phi\big(\alpha_s^h(y)\big) \iff \alpha_s^h(y) = \Phi^{-1}\big(\mathrm{F}_{hs}(y)\big).$$

Here and below, within the event $G_g + G_h = 1$, the condition $G_g = 0$ selects comparison group h. Hence $\mathrm{F}_{hs}(y) = \mathrm{F}_{Y_s(0)|G_g=0,G_g+G_h=1}(y)$ is identified from the data sampling process, and $\Phi(\cdot)$ is a known strictly increasing function. $\alpha_s^h(y)$ is therefore identified for every $y \in \mathcal{Y}$.

Second, by Assumption 1, $\mathrm{F}_{Y_s(0)|G_g=1,G_g+G_h=1}(y) = \Phi\big(\alpha_s^h(y) + \beta_s^{gh}(y)\big)$. Since $G_g = 1$ selects the treated group in the pairwise comparison sample, $\mathrm{F}_{Y_s(0)|G_g=1,G_g+G_h=1}(y) = \mathrm{F}_{Y_s(0)|G_g=1}(y)$. $\mathrm{F}_{Y_s(0)|G_g=1}(y)$ is not identified from the data sampling process, but thanks to the no-anticipation (Assumption 2), namely, $\mathrm{F}_{Y_s(0)|G_g=1}(y) = \mathrm{F}_{Y_s(1)|G_g=1}(y) = \mathrm{F}_{gs}(y)$ for all $y \in \mathcal{Y}$. Thus, by the invertibility of $\Phi(\cdot)$ assumed in Assumption 1, $\mathrm{F}_{gs}(y) = \Phi\big(\alpha_s^h(y) + \beta_s^{gh}(y)\big) \iff \beta_s^{gh}(y) = \Phi^{-1}(\mathrm{F}_{gs}(y)) - \alpha_s^h(y)$. As $\alpha_s^h(y)$ is identified from the preceding step for every $y \in \mathcal{Y}$ and $\mathrm{F}_{gs}(y)$ is identified from the sampling process for every $y \in \mathcal{Y}$, $\beta_s^{gh}(y)$ is identified for every $y \in \mathcal{Y}$.

Third, it remains to identify $\gamma_{st}^h(y)$ for every $y \in \mathcal{Y}$. By the invertibility assumption in Assumption 1 for comparison group h,

$$\mathrm{F}_{Y_t(0)|G_g=0,G_g+G_h=1}(y) = \mathrm{F}_{ht}(y) = \Phi(\alpha_s^h(y) + \gamma_{st}^h(y)) \iff \gamma_{st}^h(y) = \Phi^{-1}(\mathrm{F}_{ht}(y)) - \alpha_s^h(y).$$

$\mathrm{F}_{ht}(y)$ is identified from the data sampling process for every $y \in \mathcal{Y}$, and $\alpha_s^h(y)$ is identified from the first step for every $y \in \mathcal{Y}$; $\gamma_{st}^h(y)$ is thus identified for every $y \in \mathcal{Y}$.

Putting the above three steps together, $\big(\alpha_s^h(y), \beta_s^{gh}(y), \gamma_{st}^h(y)\big)'$, $y \in \mathcal{Y}$ is identified for $\Phi(\cdot)$ assumed known in Assumption 1. Moreover, $\mathrm{F}_{Y_t(0)|G_g=1}(y) = \mathrm{F}_{Y_t(0)|G_g=1,G_g+G_h=1}(y)$ because $G_g = 1$ selects the treated group in the pairwise comparison sample. Therefore, $\mathrm{F}_{Y_t(0)|G_g=1}(y)$ is identified:

$$\begin{aligned}\mathrm{F}_{Y_t(0)|G_g=1}(y) =& \Phi\big(\alpha_s^h(y) + \beta_s^{gh}(y) + \gamma_{st}^h(y)\big)\\ =& \Phi\Big(\Phi^{-1}\big(\mathrm{F}_{gs}(y)\big) + \Phi^{-1}\big(\mathrm{F}_{ht}(y)\big) - \Phi^{-1}\big(\mathrm{F}_{hs}(y)\big)\Big).\end{aligned}$$

□

S.1.2 An empirical process building block

For any $(g, t) \in \mathcal{G} \times [-\mathcal{T} : T]$, define the empirical process

$$\widetilde{H}_{gt}(y) := \sqrt{N}\big(\widehat{\mathrm{F}}_{gt} - \mathrm{F}_{gt}\big)(y),\ y \in \mathcal{Y}.$$

Empirical processes of the form $\widetilde{H}_{gt}(y)$ constitute the fundamental building blocks of the asymptotic theory in this paper. The following lemma states a standard weak convergence result for such processes.

Lemma 1. *Under Assumptions 3 and 4, the process* $\widetilde{H}_{gt}$ *converges weakly to a tight Gaussian process,*

$$\widetilde{H}_{gt} \rightsquigarrow \mathbb{G}_{gt} \quad in\ \ell^\infty(\mathcal{Y}),$$

for every $(g,t) \in \mathcal{G} \times [-\mathcal{T} : T]$. *The limiting process* $\mathbb{G}_{gt}$ *has mean zero and covariance function* $\Omega_{gt}(y_1, y_2)$ *defined on* $\mathcal{Y} \times \mathcal{Y}$. *An explicit expression for the covariance function is provided in the proof.*

Proof: First, obtain the asymptotically linear expression of $\widetilde{H}_{gt}(y)$.

Recall

$$\mathrm{F}_{gt}(y) := \mathbb{P}(Y_t \le y \mid G = g).$$

Since $S_t \perp\!\!\!\perp (Y_t, G)$ under Assumption 3, one has

$$\mathbb{P}(Y_t \le y \mid G = g) = \mathbb{P}(Y_t \le y \mid G = g, S_t = 1),$$

which is directly identified from the data sampling process. Thus, under the independence condition in Assumption 3 and the regularity conditions of Assumption 4,

$$\mathrm{F}_{gt}(y) = \frac{\mathbb{E}\big[S_t \mathbb{1}\{G = g\}\mathbb{1}\{Y_t \le y\}\big]}{\mathbb{E}\big[S_t \mathbb{1}\{G = g\}\big]} = \frac{1}{\pi_t p_g}\mathbb{E}\big[S_t \mathbb{1}\{G = g\}\mathbb{1}\{Y_t \le y\}\big].$$

The natural sample analogue of F_{gt} has the convenient expression

$$\widehat{\mathrm{F}}_{gt}(y) := \frac{1}{\hat{\pi}_t \hat{p}_g}\frac{1}{N}\sum_{j=1}^{N} S_{jt}\mathbb{1}\{G_j = g\}\mathbb{1}\{Y_{jt} \le y\}$$

using observations in period $t \in [-\mathcal{T} : T]$. Recall $\hat{\pi}_t = N^{-1}\sum_{j=1}^{N} S_{jt}$ and $\hat{p}_g = N^{-1}\sum_{j=1}^{N}\mathbb{1}\{G_j = g\}$. $\pi_t = \mathbb{E}[\hat{\pi}_t]$ and $p_g = \mathbb{E}[\hat{p}_g]$ under the *i.i.d.* sampling condition of Assumption 3.

Then,

$$\widehat{\mathrm{F}}_{gt}(y) - \mathrm{F}_{gt}(y) = \frac{1}{N}\sum_{j=1}^{N}\Big\{\frac{1}{\hat{\pi}_t \hat{p}_g} S_{jt}\mathbb{1}\{G_j = g\}\mathbb{1}\{Y_{jt} \le y\} - \frac{1}{\pi_t p_g}\mathbb{E}\big[S_t \mathbb{1}\{G = g\}\mathbb{1}\{Y_t \le y\}\big]\Big\}.$$

The following decomposition holds:

$$\begin{aligned}\sqrt{N}\big(\widehat{\mathrm{F}}_{gt} - \mathrm{F}_{gt}\big)(y) =& \frac{1}{\pi_t p_g \sqrt{N}}\sum_{j=1}^{N}\Big(S_{jt}\mathbb{1}\{G_j = g\}\mathbb{1}\{Y_{jt} \le y\} - \mathbb{E}\big[S_t \mathbb{1}\{G = g\}\mathbb{1}\{Y_t \le y\}\big]\Big) \\ &- \frac{\mathrm{F}_{gt}(y)}{\pi_t \sqrt{N}}\sum_{j=1}^{N}\big(S_{jt} - \pi_t\big) - \frac{\mathrm{F}_{gt}(y)}{p_g \sqrt{N}}\sum_{j=1}^{N}\big(\mathbb{1}\{G_j = g\} - p_g\big) + \widehat{\mathcal{R}}_{gt}(y)\end{aligned}$$

where the remainder term is given by

$$\begin{aligned}\widehat{\mathcal{R}}_{gt}(y) :=& -\frac{\sqrt{N}(\hat{\pi}_t\hat{p}_g - \pi_t p_g)}{\hat{\pi}_t\hat{p}_g}\cdot\frac{1}{\pi_t p_g}\frac{1}{N}\sum_{j=1}^{N}\Big(S_{jt}\mathbb{1}\{G_j = g\}\mathbb{1}\{Y_{jt} \le y\} - \mathbb{E}\big[S_t \mathbb{1}\{G = g\}\mathbb{1}\{Y_t \le y\}\big]\Big) \\ &+ \frac{\mathrm{F}_{gt}(y)}{\pi_t p_g}\sqrt{N}\frac{(\hat{\pi}_t\hat{p}_g - \pi_t p_g)^2}{\hat{\pi}_t\hat{p}_g} - \frac{\mathrm{F}_{gt}(y)}{\pi_t p_g}\sqrt{N}(\hat{\pi}_t - \pi_t)(\hat{p}_g - p_g) =: \widehat{\mathcal{R}}^{(1)}_{gt}(y) + \widehat{\mathcal{R}}^{(2)}_{gt}(y) + \widehat{\mathcal{R}}^{(3)}_{gt}(y).\end{aligned}$$

It is shown that $\sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}_{gt}(y)| = o_p(1)$. First, by Assumption 3, $\hat{p}_g = \frac{1}{N}\sum_{j=1}^N \mathbb{1}\{G_j = g\}$ is the sample mean of *i.i.d.* Bernoulli random variables with mean p_g. Hence $\hat{p}_g \xrightarrow{p} p_g$, and $\sqrt{N}(\hat{p}_g - p_g) = \mathcal{O}_p(1)$. Since $p_g \in (0,1)$ by Assumption 4, it follows that $1/\hat{p}_g = \mathcal{O}_p(1)$. Likewise, since $\hat{\pi}_t = \frac{1}{N}\sum_{j=1}^N S_{jt}$ and $\pi_t \in (0,1]$, one has $\hat{\pi}_t \xrightarrow{p} \pi_t$, $\sqrt{N}(\hat{\pi}_t - \pi_t) = \mathcal{O}_p(1)$, and $1/\hat{\pi}_t = \mathcal{O}_p(1)$. Therefore, $\hat{\pi}_t\hat{p}_g - \pi_t p_g = \pi_t(\hat{p}_g - p_g) + p_g(\hat{\pi}_t - \pi_t) + (\hat{\pi}_t - \pi_t)(\hat{p}_g - p_g) = \mathcal{O}_p(N^{-1/2})$, and so

$$\frac{\sqrt{N}(\hat{\pi}_t\hat{p}_g - \pi_t p_g)}{\hat{\pi}_t\hat{p}_g} = \mathcal{O}_p(1). \tag{S.1}$$

Next, define $f_y(W_j) := S_{jt}\mathbb{1}\{G_j = g\}\mathbb{1}\{Y_{jt} \le y\}, \quad y \in \mathcal{Y}$. The class $\{f_y : y \in \mathcal{Y}\}$ is uniformly bounded and VC-subgraph. Hence, by the Glivenko–Cantelli theorem,

$$\sup_{y\in\mathcal{Y}}\left|\frac{1}{N}\sum_{j=1}^N \Big(f_y(W_j) - \mathbb{E}[f_y(W_j)]\Big)\right| = o_p(1).$$

Therefore,

$$\begin{aligned}\sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}^{(1)}_{gt}(y)| &\le \left|\frac{\sqrt{N}(\hat{\pi}_t\hat{p}_g - \pi_t p_g)}{\hat{\pi}_t\hat{p}_g}\right| \frac{1}{\pi_t p_g}\sup_{y\in\mathcal{Y}}\left|\frac{1}{N}\sum_{j=1}^N \Big(f_y(W_j) - \mathbb{E}[f_y(W_j)]\Big)\right| \\ &= \mathcal{O}_p(1)\cdot o_p(1) = o_p(1).\end{aligned}$$

Next, since $0 \le \mathrm{F}_{gt}(y) \le 1$, $\sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}^{(2)}_{gt}(y)| \le \frac{1}{\pi_t p_g}\left|\frac{1}{\hat{\pi}_t\hat{p}_g}\right|\sqrt{N}\big(\hat{\pi}_t\hat{p}_g - \pi_t p_g\big)^2$. Because $\hat{\pi}_t\hat{p}_g - \pi_t p_g = \mathcal{O}_p(N^{-1/2})$ and $1/(\hat{\pi}_t\hat{p}_g) = \mathcal{O}_p(1)$, it follows that $\sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}^{(2)}_{gt}(y)| = \mathcal{O}_p(N^{-1/2}) = o_p(1)$.

Finally, again using $0 \le \mathrm{F}_{gt}(y) \le 1$, $\sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}^{(3)}_{gt}(y)| \le \frac{1}{\pi_t p_g}\sqrt{N}\,|\hat{\pi}_t - \pi_t|\,|\hat{p}_g - p_g|$. Since $\hat{\pi}_t - \pi_t = \mathcal{O}_p(N^{-1/2})$ and $\hat{p}_g - p_g = \mathcal{O}_p(N^{-1/2})$, one obtains $\sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}^{(3)}_{gt}(y)| = \mathcal{O}_p(N^{-1/2}) = o_p(1)$.

Combining the three bounds via the triangle inequality yields

$$\sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}_{gt}(y)| \le \sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}^{(1)}_{gt}(y)| + \sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}^{(2)}_{gt}(y)| + \sup_{y\in\mathcal{Y}}|\widehat{\mathcal{R}}^{(3)}_{gt}(y)| = o_p(1).$$

From the foregoing,

$$\sqrt{N}\big(\widehat{\mathrm{F}}_{gt} - \mathrm{F}_{gt}\big)(y) = \frac{1}{\sqrt{N}}\sum_{j=1}^N \psi_{gt}(W_j; y) + o_p(1), \tag{S.2}$$

where

$$\psi_{gt}(W_j; y) = \Big(\frac{S_{jt}\mathbb{1}\{G_j = g\}\mathbb{1}\{Y_{jt} \le y\}}{\pi_t p_g} - \mathrm{F}_{gt}(y)\Big) - \frac{\mathrm{F}_{gt}(y)}{\pi_t}\big(S_{jt} - \pi_t\big) - \frac{\mathrm{F}_{gt}(y)}{p_g}\big(\mathbb{1}\{G_j = g\} - p_g\big),$$

since $\mathbb{E}\big[S_t\mathbb{1}\{G = g\}\mathbb{1}\{Y_t \le y\}\big] = \pi_t p_g \mathrm{F}_{gt}(y)$ under Assumption 3.

Consider the function class $\mathcal{Z}_{gt} := \big\{\psi_{gt}(\cdot; y) : y \in \mathcal{Y}\big\}$. Recall Y_t denotes the generic observed outcome at period t. The class $\{\mathbb{1}\{Y_t \le y\} : y \in \mathcal{Y}\}$ is VC-subgraph (van der Vaart and Wellner, 1996, Example 2.5.4). Since $S_t\mathbb{1}\{G = g\}$ is a bounded measurable multiplier, it follows that $\big\{S_t\mathbb{1}\{G = g\}\mathbb{1}\{Y_t \le y\} : y \in \mathcal{Y}\big\}$ is again VC-subgraph, hence P-Donsker. Moreover, $\mathrm{F}_{gt}(y)$ is a bounded deterministic function of y, and $S_t - \pi_t$ and $\mathbb{1}\{G = g\} - p_g$ are bounded random variables not indexed by y. Therefore, by the preservation of Donsker classes under centring, bounded multipliers, and finite linear combinations, $\mathcal{Z}_{gt}$ is P-Donsker; see, e.g., van der Vaart and Wellner (1996, Sections 2.6.5 and 2.10).

Hence, the P-Donsker property of $\mathcal{Z}_{gt}$ implies that the empirical process

$$\left\{ \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \psi_{gt}(W_j; y) : y \in \mathcal{Y} \right\}$$

is asymptotically equicontinuous and converges weakly in $\ell^\infty(\mathcal{Y})$; see Kosorok (2007, Theorem 2.1). In particular, for any finite collection $y_1, \ldots, y_L \in \mathcal{Y}$, the vector

$$\left(\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \psi_{gt}(W_j; y_\ell) \right)_{\ell=1}^{L}$$

satisfies a multivariate central limit theorem by Assumption 3 and the boundedness of $\psi_{gt}(W_j; y)$, and the Donsker property supplies the required asymptotic equicontinuity. Therefore, $\sqrt{N}\big(\widehat{\mathrm{F}}_{gt} - \mathrm{F}_{gt}\big) \rightsquigarrow \mathbb{G}_{gt}$ in $\ell^\infty(\mathcal{Y})$, for a tight mean-zero Gaussian process $\mathbb{G}_{gt}$. Its covariance function is

$$\Omega_{gt}(y_1, y_2) := \mathbb{E}\big[\psi_{gt}(W; y_1)\psi_{gt}(W; y_2)\big], \qquad (y_1, y_2) \in \mathcal{Y} \times \mathcal{Y}.$$

□

S.1.3 Proof of Theorem 2

Part (a): This part follows directly from Lemma 1 as $\mathrm{F}_{Y_t(1)|G_g=1}(y)$ is identified from the data sampling process: $\mathrm{F}_{Y_t(1)|G_g=1}(y) = \mathrm{F}_{gt}(y)$.

Part (b): Recall the estimator of the counterfactual distribution $\widehat{\mathrm{F}}_{Y_t(0)|G_g=1}(y) = \Phi\big(\widehat{\alpha}_s^h(y) + \widehat{\beta}_s^{gh}(y) + \widehat{\gamma}_{st}^h(y)\big)$. From the identification result, namely Theorem 1,

$$\mathrm{F}_{Y_t(0)|G_g=1}(y) = \Phi\Big(\Phi^{-1}(\mathrm{F}_{gs}(y)) + \Phi^{-1}(\mathrm{F}_{ht}(y)) - \Phi^{-1}(\mathrm{F}_{hs}(y))\Big).$$

Define the function

$$h_{CF}(a) := \Phi\big(\Phi^{-1}(a_1) + \Phi^{-1}(a_2) - \Phi^{-1}(a_3)\big)$$

where $a := (a_1, a_2, a_3)'$. By the continuous differentiability and dominance conditions on $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ in Assumption 5, the induced composition map from $\ell^\infty(\mathcal{Y})^3$ to $\ell^\infty(\mathcal{Y})$ is Hadamard differentiable and the functional delta method (van der Vaart and Wellner (1996, Examples 3.9.2, Lemma 3.9.3, and Theorem 3.9.4)) applies uniformly over $y \in \mathcal{Y}$:

$$\begin{aligned}
\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t(0)|G_g=1} - \mathrm{F}_{Y_t(0)|G_g=1}\big)(y) =& \nabla_{h_{CF}}(\mathrm{F}_{gs}(y), \mathrm{F}_{ht}(y), \mathrm{F}_{hs}(y)) \\
& \times \sqrt{N}\big((\widehat{\mathrm{F}}_{gs} - \mathrm{F}_{gs})(y), (\widehat{\mathrm{F}}_{ht} - \mathrm{F}_{ht})(y), (\widehat{\mathrm{F}}_{hs} - \mathrm{F}_{hs})(y)\big) + o_p(1) \\
=& \nabla_{h_{CF}}\big(\mathrm{F}_{gs}(y), \mathrm{F}_{ht}(y), \mathrm{F}_{hs}(y)\big) \times \big(\widetilde{H}_{gs}(y), \widetilde{H}_{ht}(y), \widetilde{H}_{hs}(y)\big)' + o_p(1) \\
=& \nabla_{gs}(y)\widetilde{H}_{gs}(y) + \nabla_{ht}(y)\widetilde{H}_{ht}(y) - \nabla_{hs}(y)\widetilde{H}_{hs}(y) + o_p(1)
\end{aligned}$$

uniformly in $y \in \mathcal{Y}$ where

$$\begin{aligned}\nabla_{h_{CF}}\big(\mathrm{F}_{gs}(y),\mathrm{F}_{ht}(y),\mathrm{F}_{hs}(y)\big) =&\phi\Big(\Phi^{-1}\big(\mathrm{F}_{gs}(y)\big)+\Phi^{-1}\big(\mathrm{F}_{ht}(y)\big)-\Phi^{-1}\big(\mathrm{F}_{hs}(y)\big)\Big)\\ &\times\left(\frac{1}{\phi\big(\Phi^{-1}(\mathrm{F}_{gs}(y))\big)},\ \frac{1}{\phi\big(\Phi^{-1}(\mathrm{F}_{ht}(y))\big)},\ -\frac{1}{\phi\big(\Phi^{-1}(\mathrm{F}_{hs}(y))\big)}\right)\\ =&:\big(\nabla_{gs}(y),\nabla_{ht}(y),-\nabla_{hs}(y)\big)\end{aligned}$$

is the gradient row vector of h_{CF} and $\phi(b) := d\Phi(b)/db$.

Under the regularity conditions of Assumption 5, $\sup_{y\in\mathcal{Y}}\max\big\{\nabla_{gs}(y),\nabla_{ht}(y),\nabla_{hs}(y)\big\}\le C^2<\infty$. In addition to (S.2), one has

$$\begin{aligned}&\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{hs}(0)|G_g=1}-\mathrm{F}_{Y_t(0)|G_g=1}\big)(y)\\ =&\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Big\{\nabla_{gs}(y)\psi_{gs}(W_j;y)+\nabla_{ht}(y)\psi_{ht}(W_j;y)-\nabla_{hs}(y)\psi_{hs}(W_j;y)\Big\}+o_p(1)\\ =&:\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi_{gt}^{hs}(W_j;y)+o_p(1)\end{aligned}\tag{S.3}$$

with $\mathbb{E}\big[\Psi_{gt}^{hs}(W;y)\big]=0\ \forall y\in\mathcal{Y}$ and $\sup_{y\in\mathcal{Y}}\mathbb{E}|\Psi_{gt}^{hs}(W;y)|^2<\infty$ under the conditions of Lemma 1 and Assumption 5. This implies that under the sampling conditions of Assumption 3, the Multivariate Lindeberg–Lévy Central Limit Theorem applies:

$$\Big(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi_{gt}^{hs}(W_j;y_1),\ldots,\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi_{gt}^{hs}(W_j;y_L)\Big)'$$

converges in distribution to the multivariate normal with the (l,l')'th entry of the covariance matrix: $\mathbb{E}\big[\Psi_{gt}^{hs}(W;y_l)\Psi_{gt}^{hs}(W;y_{l'})\big]$ for every finite collection $\big\{y_l\ :\ 1\le l\le L\big\}\subset\mathcal{Y}$.

Consider the indexing class

$$\mathcal{Z}_{gt}^{hs}:=\big\{\Psi_{gt}^{hs}(\cdot;y):y\in\mathcal{Y}\big\}.$$

By Lemma 1, for each $(g,t)\in\mathcal{G}\times[-\mathcal{T}:T]$, the class $\{\psi_{gt}(\cdot;y):y\in\mathcal{Y}\}$ is P-Donsker. Under the boundedness conditions in Assumption 4 and Assumption 5 (in particular, p_g,π_t bounded away from zero) and $\sup_{y\in\mathcal{Y}}|\nabla_{gs}(y)|+\sup_{y\in\mathcal{Y}}|\nabla_{ht}(y)|+\sup_{y\in\mathcal{Y}}|\nabla_{hs}(y)|<\infty$, it follows that $\mathcal{Z}_{gt}^{hs}$ is obtained from a finite linear combination of P-Donsker classes with bounded multipliers. Hence $\mathcal{Z}_{gt}^{hs}$ is P-Donsker and the corresponding empirical process

$$\left\{\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi_{gt}^{hs}(W_j;y):y\in\mathcal{Y}\right\}$$

is asymptotically equicontinuous and converges weakly in $\ell^\infty(\mathcal{Y})$.

Combining the foregoing yields the weak convergence result in $\ell^\infty(\mathcal{Y})$ (Kosorok, 2007, Theorem 2.1). Therefore, the empirical process $\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{hs}(0)|G_g=1}-\mathrm{F}_{Y_t(0)|G_g=1}\big)$ converges to a tight mean-zero Gaussian process, namely $\mathbb{G}_{gt}^{hs}$. Under Assumption 3, the covariance function of $\mathbb{G}_{gt}^{hs}$ is given by $\Omega_{gt}^{hs}(y_1,y_2):=\mathbb{E}\big[\Psi_{gt}^{hs}(W;y_1)\Psi_{gt}^{hs}(W;y_2)\big]$.

Part (c): From parts (a) and (b) above, in addition to (S.2) and (S.3),

$$\begin{aligned}
\sqrt{N}\big(\widehat{\mathrm{DTT}}_{gt}^{hs} - \mathrm{DTT}_{gt}\big)(y) =& \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t(1)|G_g=1} - \mathrm{F}_{Y_t(1)|G_g=1}\big)(y) - \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{hs}(0)|G_g=1} - \mathrm{F}_{Y_t(0)|G_g=1}\big)(y) \\
=& \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \Big\{ \psi_{gt}(W_j; y) - \Psi_{gt}^{hs}(W_j; y) \Big\} + o_p(1) \\
=: & \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \Upsilon_{gt}^{hs}\big(W_j; y\big) + o_p(1) \\
& \rightsquigarrow \mathbb{H}_{gt}^{hs}
\end{aligned}$$

using the sequence of arguments akin to part (b) above and Lemma 1. The associated function class is P-Donsker by closure under finite sums of P-Donsker classes with bounded coefficients, so the corresponding empirical process is asymptotically equicontinuous and converges weakly in $\ell^\infty(\mathcal{Y})$. Thanks to Assumption 3, the covariance function of $\mathbb{H}_{gt}^{hs}$ is given by $\Sigma_{gt}^{hs}(y_1, y_2) := \mathbb{E}\big[\Upsilon_{gt}^{hs}\big(W; y_1\big)\Upsilon_{gt}^{hs}\big(W; y_2\big)\big]$. □

S.1.4 Proof of Corollary 1

Because $\mathcal{I}$ is finite, the joint expansion in Assumption 6 implies that the estimated-weight contribution is finite-dimensional and tight. Products of the form $\sqrt{N}(\widehat{\omega} - \omega)(\widehat{\mathscr{F}} - \mathscr{F})$ are $o_p(1)$ uniformly over $\mathcal{Y}$, since $\sqrt{N}(\widehat{\omega} - \omega) = O_p(1)$ jointly over $\mathcal{I}$ and the relevant first-stage DF or DTT estimation error is $o_p(1)$ uniformly over $\mathcal{Y}$.

Part (a): The following asymptotically linear representation holds uniformly in $y \in \mathcal{Y}$ thanks to Assumption 6 and Lemma 1:

$$\begin{aligned}
\sqrt{N}(\widehat{\mathrm{F}}_{Y_{\widehat{\omega}}(1)} - \mathrm{F}_{Y_\omega(1)})(y) =& \sum_{(g,t,h,s)\in\mathcal{I}} \Big\{ \mathrm{F}_{gt}(y)\sqrt{N}\big(\widehat{\omega}(g,t,h,s) - \omega(g,t,h,s)\big) + \omega(g,t,h,s) \times \sqrt{N}\big(\widehat{\mathrm{F}}_{gt} - \mathrm{F}_{gt}\big)(y) \\
& + \sqrt{N}\big(\widehat{\omega}(g,t,h,s) - \omega(g,t,h,s)\big)\big(\widehat{\mathrm{F}}_{gt} - \mathrm{F}_{gt}\big)(y) \Big\} \\
=& \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left\{ \sum_{(g,t,h,s)\in\mathcal{I}} \Big\{ \mathrm{F}_{gt}(y)\mathcal{W}_j(g,t,h,s) + \omega(g,t,h,s)\psi_{gt}(W_j; y) \Big\} \right\} + o_p(1) \\
=: & \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \Psi_\omega(W_j; y) + o_p(1).
\end{aligned}$$

Following the sequence of arguments in Lemma 1, together with finite sums over $\mathcal{I}$, the associated class is P-Donsker and the corresponding empirical process is asymptotically equicontinuous. Thus, $\sqrt{N}(\widehat{\mathrm{F}}_{Y_{\widehat{\omega}}(1)} - \mathrm{F}_{Y_\omega(1)})$ converges weakly to a tight Gaussian process with mean 0, namely $\mathbb{G}_\omega$. Under Assumption 3, the covariance function is given by $\Omega_\omega(y_1, y_2) := \mathbb{E}\big[\Psi_\omega(W; y_1)\Psi_\omega(W; y_2)\big]$.

Part (b): The following asymptotically linear representation holds thanks to Assumption 6 and part (b) of the proof of Theorem 2:

$$\begin{aligned}\sqrt{N}(\widehat{\mathrm{F}}_{Y_{\widehat{\omega}}(0)} - \mathrm{F}_{Y_\omega(0)})(y) = &\sum_{(g,t,h,s)\in\mathcal{I}} \Big\{ \mathrm{F}_{Y_t(0)|G_g=1}(y)\sqrt{N}\big(\widehat{\omega}(g,t,h,s) - \omega(g,t,h,s)\big) \\ &\quad + \omega(g,t,h,s) \times \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{hs}(0)|G_g=1} - \mathrm{F}_{Y_t(0)|G_g=1}\big)(y)\Big\} + o_p(1) \\ =&\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Bigg\{\sum_{(g,t,h,s)\in\mathcal{I}}\Big\{\mathrm{F}_{Y_t(0)|G_g=1}(y)\mathcal{W}_j(g,t,h,s) + \omega(g,t,h,s)\Psi_{gt}^{hs}(W_j;y)\Big\}\Bigg\} + o_p(1) \\ =:& \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi_\omega^C(W_j;y) + o_p(1).\end{aligned}$$

Following the sequence of arguments in part (b) of the proof of Theorem 2, together with finite sums over $\mathcal{I}$, the associated class is P-Donsker and the corresponding empirical process is asymptotically equicontinuous. Thus, $\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_{\widehat{\omega}}(0)} - \mathrm{F}_{Y_\omega(0)}\big)$ converges weakly to a tight Gaussian process with mean 0, namely $\mathbb{G}_\omega^C$. Under Assumption 3, the covariance function is given by $\Omega_\omega^C(y_1,y_2) := \mathbb{E}\big[\Psi_\omega^C(W;y_1)\Psi_\omega^C(W;y_2)\big]$.

Part (c): The following asymptotically linear representation holds thanks to Assumption 6 and part (c) of the proof of Theorem 2:

$$\begin{aligned}\sqrt{N}\big(\widehat{\mathrm{DTT}}_{\widehat{\omega}} - \mathrm{DTT}_\omega\big)(y) = &\sum_{(g,t,h,s)\in\mathcal{I}} \Big\{ \mathrm{DTT}_{gt}(y)\sqrt{N}\big(\widehat{\omega}(g,t,h,s) - \omega(g,t,h,s)\big) \\ &\quad + \omega(g,t,h,s) \times \sqrt{N}\big(\widehat{\mathrm{DTT}}_{gt}^{hs} - \mathrm{DTT}_{gt}\big)(y)\Big\} + o_p(1) \\ =&\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Bigg\{\sum_{(g,t,h,s)\in\mathcal{I}}\Big\{\mathrm{DTT}_{gt}(y)\mathcal{W}_j(g,t,h,s) + \omega(g,t,h,s)\Upsilon_{gt}^{hs}(W_j;y)\Big\}\Bigg\} + o_p(1) \\ =:& \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Upsilon_\omega(W_j;y) + o_p(1).\end{aligned}$$

Thus, following the Donsker and asymptotic-equicontinuity arguments in the proof of parts (b) and (c) of Theorem 2, in addition to part (b) above:

$$\sqrt{N}\big(\widehat{\mathrm{DTT}}_{\widehat{\omega}} - \mathrm{DTT}_\omega\big) \rightsquigarrow \mathbb{H}_\omega,$$

where $\mathbb{H}_\omega$ is a tight mean-zero Gaussian process. Thanks to Assumption 3, the covariance function is given by $\Sigma_\omega(y_1,y_2) := \mathbb{E}\Big[\Upsilon_\omega\big(W;y_1\big)\Upsilon_\omega\big(W;y_2\big)\Big]$.

□

S.1.5 Proof of Theorem 3

Let $\widehat{\mathscr{F}}$ denote an estimator of a functional $\mathscr{F}$ defined on $\mathcal{Y}$, e.g., $\mathrm{F}_{Y_t(0)|G_g=1}$, DTT_{gt}, $\mathrm{F}_{Y_\omega(0)}$, or DTT_ω. From the uniform asymptotic linear representations established in Lemma 1, Theorem 2, and Corollary 1, there exists a measurable influence function $\Psi_{\mathscr{F}}(W;y)$ such that

$$\sqrt{N}\big(\widehat{\mathscr{F}} - \mathscr{F}\big)(y) = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi_{\mathscr{F}}(W_j;y) + o_p(1) \quad \text{in } \ell^\infty(\mathcal{Y}),$$

where the associated function class $\{\Psi_{\mathscr{F}}(\cdot;y) : y \in \mathcal{Y}\}$ is P–Donsker, has mean zero, and is square integrable.

Consider the exchangeable bootstrap analogue satisfying the uniform expansion

$$\sqrt{N}\big(\widehat{\mathscr{F}}^{*} - \widehat{\mathscr{F}}\big)(y) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\xi_{Nj} - \bar{\xi}_N)\Psi_{\mathscr{F}}(W_j;y) + o_p(1) \quad \text{in } \ell^\infty(\mathcal{Y}). \tag{S.4}$$

This representation follows from the uniform asymptotic linearity of $\widehat{\mathscr{F}}$ and the exchangeably weighted bootstrap theorem of van der Vaart and Wellner (1996, Theorem 3.6.13), applied under the exchangeability, moment, sample-variance, and sample-mean requirements stated in the bootstrap subsection. Conditional on the data, the bootstrap empirical process

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\xi_{Nj} - \bar{\xi}_N)\Psi_{\mathscr{F}}(W_j;\cdot)$$

converges weakly in probability in $\ell^\infty(\mathcal{Y})$ to the *same* tight Gaussian limit as the original empirical process. Hence, the exchangeable bootstrap consistently estimates the law of $\sqrt{N}(\widehat{\mathscr{F}} - \mathscr{F})$ uniformly over $\mathcal{Y}$.

□

S.1.6 Proof of Theorem 4

Part (a): **(i)** Since the admissible representations satisfy a joint uniform asymptotic linear representation, the stacked process $\sqrt{N}\big(\widehat{\mathbf{F}}_{Y_t(0)|G_g=1} - \mathbf{F}_{Y_t(0)|G_g=1}\big)$ is tight in $\ell^\infty(\mathcal{Y})^{K_{gt}}$, where $\mathbf{F}_{Y_t(0)|G_g=1}(y) := \big(\mathrm{F}_{Y_t^{hs}(0)|G_g=1}(y)\big)'_{(h,s)\in\mathcal{I}_{gt}} \in \mathbb{R}^{K_{gt}}$. The map induced by $\boldsymbol{B}'_{gt}$ is continuous and linear, so under $\mathbb{H}_0$ and for every $y \in \mathcal{Y}$,

$$\begin{aligned}
\widehat{\boldsymbol{Z}}_{gt}(y) &= \sqrt{N}\,\boldsymbol{B}'_{gt}\,\widehat{\mathbf{F}}_{Y_t(0)|G_g=1}(y) \\
&= \sqrt{N}\,\boldsymbol{B}'_{gt}\,\big(\widehat{\mathbf{F}}_{Y_t(0)|G_g=1} - \mathbf{F}_{Y_t(0)|G_g=1}\big)(y) + \sqrt{N}\,\boldsymbol{B}'_{gt}\,\mathbf{F}_{Y_t(0)|G_g=1}(y) \\
&\overset{\mathbb{H}_0}{=} \sqrt{N}\,\boldsymbol{B}'_{gt}\,\big(\widehat{\mathbf{F}}_{Y_t(0)|G_g=1} - \mathbf{F}_{Y_t(0)|G_g=1}\big)(y) + \sqrt{N}\,\boldsymbol{B}'_{gt}\,\mathbf{1}_{K_{gt}}F^0_{gt}(y) \\
&= \sqrt{N}\,\boldsymbol{B}'_{gt}\,\big(\widehat{\mathbf{F}}_{Y_t(0)|G_g=1} - \mathbf{F}_{Y_t(0)|G_g=1}\big)(y),
\end{aligned}$$

where $\mathbf{1}_{K_{gt}}$ is a $K_{gt} \times 1$ vector of ones and $F^0_{gt}(y)$ denotes the common value of the admissible representations under $\mathbb{H}_0$. The third equality holds under $\mathbb{H}_0$ since $\mathbf{F}_{Y_t(0)|G_g=1}(y) = \mathbf{1}_{K_{gt}}F^0_{gt}(y),\ \forall y \in \mathcal{Y}$, and the fourth uses $\boldsymbol{B}'_{gt}\,\mathbf{1}_{K_{gt}} = \mathbf{0}_{L_{gt}}$, since each column of $\boldsymbol{B}_{gt}$ contains one entry equal to 1, one entry equal to -1, and zeroes elsewhere, so $\boldsymbol{B}'_{gt}$ annihilates vectors that are constant across admissible representations. It then follows that $\widehat{\boldsymbol{Z}}_{gt}$ converges weakly in $\ell^\infty(\mathcal{Y})^{L_{gt}}$ to $\boldsymbol{Z}_{gt} := \boldsymbol{B}'_{gt}\mathbb{Z}_{gt}$, where $\mathbb{Z}_{gt}(y) := \big(\mathbb{G}^{hs}_{gt}(y)\big)'_{(h,s)\in\mathcal{I}_{gt}}$ (see Theorem 2(b)) denotes the mean-zero tight Gaussian limit of the stacked process, with covariance kernel $\mathbb{C}ov\big(\mathbb{Z}_{gt}(y_1), \mathbb{Z}_{gt}(y_2)\big) \in \mathbb{R}^{K_{gt}\times K_{gt}}$. Since $\boldsymbol{Z}_{gt}(y) = \boldsymbol{B}'_{gt}\,\mathbb{Z}_{gt}(y)$, the process $\boldsymbol{Z}_{gt}$ is mean-zero tight Gaussian in $\ell^\infty(\mathcal{Y})^{L_{gt}}$ with covariance kernel

$$\mathbb{C}ov(\boldsymbol{Z}_{gt}(y_1), \boldsymbol{Z}_{gt}(y_2)) = \boldsymbol{B}'_{gt}\,\mathbb{C}ov\big(\mathbb{Z}_{gt}(y_1), \mathbb{Z}_{gt}(y_2)\big)\,\boldsymbol{B}_{gt} \in \mathbb{R}^{L_{gt}\times L_{gt}}.$$

Equivalently, the column of $\boldsymbol{B}_{gt}$ associated with distinct elements (h,s) and (h',s') yields the pairwise contrast

$$\widehat{\boldsymbol{Z}}_{gt}(y)_{(h,s),(h',s')} = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \Big(\Psi^{hs}_{gt}(W_j;y) - \Psi^{h's'}_{gt}(W_j;y)\Big) + o_p(1)$$

under $\mathbb{H}_0$.

(ii) Consider the following decomposition:

$$\begin{aligned}\widehat{\mathrm{CvM}}_{gt} &= \int_{\mathcal{Y}} \|\widehat{\boldsymbol{Z}}_{gt}(y)\|_2^2 \, \widehat{H}(dy) \\ &= \int_{\mathcal{Y}} \|\widehat{\boldsymbol{Z}}_{gt}(y)\|_2^2 \, H(dy) + \int_{\mathcal{Y}} \|\widehat{\boldsymbol{Z}}_{gt}(y)\|_2^2 \, (\widehat{H} - H)(dy)\end{aligned}$$

For fixed H, the map $z \mapsto \int_{\mathcal{Y}} \|z(y)\|_2^2 \, H(dy)$ is continuous on $\ell^\infty(\mathcal{Y})^{L_{gt}}$ under the supremum norm, since

$$\big|\|a\|_2^2 - \|b\|_2^2\big| \leq \|a-b\|_2\big(\|a\|_2 + \|b\|_2\big).$$

Thus, the weak convergence result in part (i) and the continuous mapping theorem imply

$$\int_{\mathcal{Y}} \|\widehat{\boldsymbol{Z}}_{gt}(y)\|_2^2 \, H(dy) \xrightarrow{d} \int_{\mathcal{Y}} \|\boldsymbol{Z}_{gt}(y)\|_2^2 \, H(dy).$$

For the empirical weighting measure used here, $\widehat{H}$ is an empirical CDF; hence $\sup_{y\in\mathcal{Y}} \big|\big(\widehat{H} - H\big)(y)\big| = o_p(1)$ by the sampling condition in Assumption 3 and the Glivenko-Cantelli Theorem (van der Vaart, 2000, Theorem 19.1). The condition $\int_{\mathcal{Y}} \|\widehat{\boldsymbol{Z}}_{gt}(y)\|_2^2 \, \big(\widehat{H} - H\big)(dy) = o_p(1)$ follows from the asymptotic equicontinuity and boundedness of the process, using the argument in the concluding part of the proof of Corollary 1 in Sant'Anna and Song (2019). Therefore, $\widehat{\mathrm{CvM}}_{gt} \xrightarrow{d} \int_{\mathcal{Y}} \|\boldsymbol{Z}_{gt}(y)\|_2^2 \, H(dy)$.

(iii) By the continuous mapping theorem together with part (i) above, the conclusion follows from van der Vaart and Wellner (1996, Theorem 1.3.6).

Part (b): Under $\mathbb{H}_{an}$, for each pair $(h,s) < (h',s')$,

$$\begin{aligned}&\widehat{\boldsymbol{Z}}_{gt}(y)_{(h,s),(h',s')} = \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{hs}(0)|G_g=1} - \widehat{\mathrm{F}}_{Y_t^{h's'}(0)|G_g=1}\big)(y) \\ &= \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{hs}(0)|G_g=1} - F_{gt}^0\big)(y) - \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{h's'}(0)|G_g=1} - F_{gt}^0\big)(y) \\ &= \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{hs}(0)|G_g=1} - \mathrm{F}_{Y_t^{hs}(0)|G_g=1}\big)(y) - \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{h's'}(0)|G_g=1} - \mathrm{F}_{Y_t^{h's'}(0)|G_g=1}\big)(y) \\ &\quad + \varrho_{gt}^{hs}(y) - \varrho_{gt}^{h's'}(y) \\ &= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Big(\Psi_{gt}^{hs}(W_j;y) - \Psi_{gt}^{h's'}(W_j;y)\Big) + \big(\varrho_{gt}^{hs}(y) - \varrho_{gt}^{h's'}(y)\big) + o_p(1).\end{aligned}$$

Stacking these pairwise drifts over all columns of $\boldsymbol{B}_{gt}$ gives $\boldsymbol{\Delta}_{gt}(y) = \boldsymbol{B}_{gt}'\boldsymbol{R}_{gt}(y)$. The conclusion then follows from the above decomposition and part (a)(i) above.

Part (c): The KS conclusion follows from a direct lower-bound argument. Under a fixed alternative, there exists at least one pair $(h,s) < (h',s') \in \mathcal{I}_{gt}$ and $y^\dagger \in \mathcal{Y}$ such that

$$\Big|\big(\mathrm{F}_{Y_t^{hs}(0)|G_g=1} - \mathrm{F}_{Y_t^{h's'}(0)|G_g=1}\big)(y^\dagger)\Big| > 0.$$

The estimation error in the corresponding pairwise contrast is $\mathcal{O}_p(1)$ after multiplication by $\sqrt{N}$, whereas the fixed discrepancy contributes

$$\sqrt{N}\Big|\big(\mathrm{F}_{Y_t^{hs}(0)|G_g=1} - \mathrm{F}_{Y_t^{h's'}(0)|G_g=1}\big)(y^\dagger)\Big| \to \infty$$

as $N \to \infty$. Therefore, $\widehat{\mathrm{KS}}_{gt} \xrightarrow{p} \infty$.

For the CvM statistic, let $d_{gt}^{hs,h's'}(y) := \mathrm{F}_{Y_t^{hs}(0)|G_g=1}(y) - \mathrm{F}_{Y_t^{h's'}(0)|G_g=1}(y)$. If $\int_{\mathcal{Y}} \big(d_{gt}^{hs,h's'}(y)\big)^2 H(dy) > 0$ for some admissible pair, then empirical-measure consistency gives

$$\int_{\mathcal{Y}} \big(d_{gt}^{hs,h's'}(y)\big)^2 \widehat{H}(dy) = \int_{\mathcal{Y}} \big(d_{gt}^{hs,h's'}(y)\big)^2 H(dy) + o_p(1).$$

The corresponding component of the statistic contains

$$N \int_{\mathcal{Y}} \big(d_{gt}^{hs,h's'}(y)\big)^2 H(dy) + O_p(\sqrt{N}) + O_p(1),$$

up to an additional $o_p(N)$ term from replacing H by $\widehat{H}$. Since the leading deterministic term is of order N and strictly positive, while the remaining terms are of smaller stochastic order, $\widehat{\mathrm{CvM}}_{gt} \xrightarrow{p} \infty$.

Part (d): Using (S.4), the $(h,s) < (h',s')$-th entry of the bootstrapped process has the following representation:

$$\begin{aligned}\widehat{\boldsymbol{Z}}_{gt}^*(y)_{(h,s),(h',s')} =& \sqrt{N}\big(\widehat{\mathrm{F}}^*_{Y_t^{hs}(0)|G_g=1} - \widehat{\mathrm{F}}_{Y_t^{hs}(0)|G_g=1}\big)(y) - \sqrt{N}\big(\widehat{\mathrm{F}}^*_{Y_t^{h's'}(0)|G_g=1} - \widehat{\mathrm{F}}_{Y_t^{h's'}(0)|G_g=1}\big)(y) \\ =& \frac{1}{\sqrt{N}}\sum_{j=1}^{N}(\xi_{Nj} - \bar{\xi}_N)\Big(\Psi_{gt}^{hs}(W_j;y) - \Psi_{gt}^{h's'}(W_j;y)\Big) + o_p(1).\end{aligned}$$

Stacking the pairwise contrasts over all columns of $\boldsymbol{B}_{gt}$ gives

$$\widehat{\boldsymbol{Z}}_{gt}^*(y) = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}(\xi_{Nj} - \bar{\xi}_N)\boldsymbol{B}_{gt}'\boldsymbol{\Psi}_{gt}(W_j;y) + o_p(1) \quad \text{in } \ell^\infty(\mathcal{Y})^{L_{gt}},$$

where $\boldsymbol{\Psi}_{gt}(W_j;y) := \big(\Psi_{gt}^{hs}(W_j;y)\big)_{(h,s)\in\mathcal{I}_{gt}}$. By (S.4) and the exchangeably weighted bootstrap theorem applied to the joint uniform asymptotic linear representation of the admissible counterfactual estimators,

$$\widehat{\boldsymbol{Z}}_{gt}^* \underset{\xi}{\overset{P}{\rightsquigarrow}} \boldsymbol{Z}_{gt} \quad \text{in } \ell^\infty(\mathcal{Y})^{L_{gt}}.$$

The bootstrap KS conclusion follows by the continuous mapping theorem applied to $z \mapsto \sup_{y\in\mathcal{Y}}\|z(y)\|_\infty$. For the CvM statistic, the map $z \mapsto \int_{\mathcal{Y}} \|z(y)\|_2^2\, H(dy)$ is continuous along bounded uniformly continuous sample paths. Replacing H by $\widehat{H}$ is asymptotically negligible by the same asymptotic equicontinuity and boundedness argument used in part (a), now applied conditionally to $\widehat{\boldsymbol{Z}}_{gt}^*$. Hence

$$\widehat{\mathrm{CvM}}_{gt}^* \underset{\xi}{\overset{P}{\rightsquigarrow}} \int_{\mathcal{Y}} \|\boldsymbol{Z}_{gt}(y)\|_2^2\, H(dy) \quad \text{and} \quad \widehat{\mathrm{KS}}_{gt}^* \underset{\xi}{\overset{P}{\rightsquigarrow}} \sup_{y\in\mathcal{Y}}\|\boldsymbol{Z}_{gt}(y)\|_\infty.$$

□

S.2 Extension - Covariates

Kim and Wooldridge (2025) employs an Inverse Probability Weighted (IPW) approach to introducing covariates, where the link function $\Phi(\cdot)$ is the identity function. With possibly non-linear link functions in the current paper, an IPW approach to introducing covariates is less direct. This paper complements the Kim and Wooldridge (2025) approach by using an Outcome Regression approach via Distribution Regression (DR) — see also Fernández-Val, Meier, Vuuren, and Vella (2024). Let $p(X) \in \mathbb{R}^{p_x}$ denote a dictionary of transformations of a set of some elementary time-invariant characteristics X, including the constant term 1, and $p_x < \infty$ fixed.

S.2.1 Identification

The following states versions of Assumptions 1 and 2 that are conditional on pre-treatment covariates X for a fixed group g at a post-treatment period $t \geq g$ using a valid control group $h > t$.

Assumption 1-X (Conditional Distributional Parallel Trends). *For every $y \in \mathcal{Y}$, every $(g, t, h, s) \in \mathcal{I}$, and $(d, \nu) \in (\{0, 1\} \times \{s, t\})$, the following restriction and representation hold on the treated covariate support, i.e., for $\mathbb{P}_{X|G_g=1}$-almost every x:*

$$\mathrm{F}_{Y_\nu(0)|X=x,G_g=d,G_g+G_h=1}(y) = \Phi\Big(p(x)\big(\boldsymbol{\alpha}_s^h(y) + d\boldsymbol{\beta}_s^{gh}(y) + \mathbb{1}\{\nu = t\}\boldsymbol{\gamma}_{st}^h(y)\big)\Big).$$

Observe in this case that $\big(\boldsymbol{\alpha}_s^h(y)', \boldsymbol{\beta}_s^{gh}(y)', \boldsymbol{\gamma}_{st}^h(y)'\big)' \in \mathbb{R}^{3p_x}$. Similarly, the conditional distributional no-anticipation condition is stated as follows.

Assumption 2-X (Conditional Distributional No-Anticipation). *For every $y \in \mathcal{Y}$, every $g \in \mathcal{G} \setminus \{\infty\}$, and any $s \in [-\mathcal{T} : (g - 1)]$,*

$$\mathrm{F}_{Y_s(1)|X=x,G_g=1}(y) = \mathrm{F}_{Y_s(0)|X=x,G_g=1}(y) \text{ for } \mathbb{P}_{X|G_g=1} - a.e.\ x.$$

Assumptions 1-X and 2-X constitute weaker forms of Assumptions 1 and 2 as they only require that the distributional parallel trends and no-anticipation conditions hold in sub-populations characterised by $X = x$ but not necessarily unconditionally. The overlap condition below ensures that the comparison-group conditional distributions are defined on the treated covariate support. Further, by using dictionaries of transformations of X, namely $p(X)$, Assumptions 1-X and 2-X are more likely to hold in given applications.

The following provides a suitable overlap condition that applies to the current setting.

Assumption 1 (Overlap). *For every $(g, t) \in \mathcal{G} \setminus \{\infty\} \times [g : T]$ and any $h \in \big(\mathcal{G} \setminus [1 : t]\big)$, $\mathbb{P}(G_g = 1 \mid X, G_g + G_h = 1) < (1 - c)$ a.s. for some $c \in (0, 1)$.*

The following regularity condition concerns $p(X)$.

Assumption 2 (No Perfect Collinearity). *$p(X) \mid G = g$ has full column rank almost surely on its support for each group $g \in \mathcal{G}$.*

Let $F_X(\cdot)$ and $F_{X|G_g=1}(\cdot)$ denote, respectively, the joint unconditional and conditional CDFs of X. The following extends Theorem 1 to the covariate setting.

Theorem 1 (Identification). *Suppose Assumptions 1-X, 2-X, 1, and 2 hold. Then, for every $y \in \mathcal{Y}$ and every $(g, t, h, s) \in \mathcal{I}$,* *(a) the parameter vector* $\big(\boldsymbol{\alpha}_s^h(y)', \boldsymbol{\beta}_s^{gh}(y)', \boldsymbol{\gamma}_{st}^h(y)'\big)'$ *is identified;*

(b) $\mathrm{F}_{Y_t(0)|X=x,G_g=1}(y)$ *is identified:* $\mathrm{F}_{Y_t(0)|X=x,G_g=1}(y) = \Phi\Big(p(x)\big(\boldsymbol{\alpha}_s^h(y) + \boldsymbol{\beta}_s^{gh}(y) + \boldsymbol{\gamma}_{st}^h(y)\big)\Big)$*; and*

(c) $\mathrm{F}_{Y_t(0)|G_g=1}(y)$ *is identified:* $\mathrm{F}_{Y_t(0)|G_g=1}(y) = \mathbb{E}\Big[\Phi\Big(p(X)\big(\boldsymbol{\alpha}_s^h(y) + \boldsymbol{\beta}_s^{gh}(y) + \boldsymbol{\gamma}_{st}^h(y)\big)\Big)\Big|G_g = 1\Big]$.

S.2.2 Estimation

Unlike the estimators introduced in the main text, which have closed-form expressions, the introduction of covariates X does not lead to closed-form expressions. Estimation is therefore required, for example by quasi-maximum-likelihood Distribution Regression (DR); see Wooldridge (2023) and Chernozhukov, Fernandez-Val, and Melly (2013). One may estimate a binary response model with working CDF $\Phi(\cdot)$. This yields the distribution regression approach of Foresi and Peracchi (1995) and Chernozhukov, Fernandez-Val, and Melly (2013). Under Assumption 2, and because the retained cells are (g,s), (h,s), and (h,t), the regressors $p(X)$, $d\,p(X)$, and $\mathbb{1}\{\nu = t\}p(X)$ are not perfectly collinear.

The following condition relies on the sampling condition in Assumption 3 with covariates included in observed data, namely with the modification $W_j := \big((S_{jt}, Y_{jt})_{t\in[-\mathcal{T}:T]}, G_j, X_j\big)$.

Assumption 3-X (Random Sampling with X)**.** *The vectors* $\big\{W_j := \big((S_{jt}, Y_{jt})_{t\in[-\mathcal{T}:T]}, G_j, X_j\big) : 1 \le j \le N\big\}$ *are independent and identically distributed with* $S_{jt} \perp\!\!\!\perp (Y_{jt}(0), Y_{jt}(1), G_j, X_j)$ *for each* $t \in [-\mathcal{T} : T]$*. In addition,* $\mathbb{E}[\|p(X)\|^4] < \infty$.

By the Law of Iterated Expectations (LIE), the conditions of Theorem 1, and Assumption 3-X,

$$\begin{aligned}
\mathrm{F}_{Y_t(0)|G_g=1}(y) =& \mathbb{E}[\mathbb{1}\{Y_t(0) \le y\} \mid G_g = 1] \\
=& \mathbb{E}[\mathbb{1}\{Y_t(0) \le y\} \mid G_g = 1, S_t = 1] \\
=& \mathbb{E}\Big[\mathbb{E}\big[\mathbb{1}\{Y_t(0) \le y\} \mid G_g = 1, S_t = 1, X\big]\Big|G_g = 1, S_t = 1\Big] \\
=& \mathbb{E}\Big[\mathbb{E}\big[\mathbb{1}\{Y_t(0) \le y\} \mid G_g = 1, X\big]\Big|G_g = 1, S_t = 1\Big] \\
=& \int \mathrm{F}_{Y_t(0)|X=x,G_g=1}(y) dF_{X|G_g=1,S_t=1}(x) \\
=& \frac{1}{\pi_t p_g}\mathbb{E}\Big[S_t \mathbb{1}\{G = g\}\Phi\Big(p(X)\big(\boldsymbol{\alpha}_s^h(y) + \boldsymbol{\beta}_s^{gh}(y) + \boldsymbol{\gamma}_{st}^h(y)\big)\Big)\Big]
\end{aligned} \tag{S.5}$$

where the second and fourth equalities follow by the independence of observability S_t from $(Y_t(0), G_g, X)$ (Assumption 3-X). The last equality shows that in practice, $\mathrm{F}_{Y_t(0)|G_g=1}(y)$ is computed by averaging $\mathrm{F}_{Y_t(0)|X=x,G_g=1}(y)$ over x of the treated group g. The estimator of $\mathrm{F}_{Y_t(0)|G_g=1}(y)$ with covariates using control group h and pre-treatment period s is given by

$$\widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1}(y) = \frac{1}{\hat{\pi}_t \hat{p}_g N}\sum_{j=1}^{N} S_{jt}\mathbb{1}\{G_j = g\}\Phi\Big(p(X_j)\big(\widehat{\boldsymbol{\alpha}}_s^h(y) + \widehat{\boldsymbol{\beta}}_s^{gh}(y) + \widehat{\boldsymbol{\gamma}}_{st}^h(y)\big)\Big), \tag{S.6}$$

where the coefficients are obtained from fitting

$$\mathrm{F}_{Y_\nu(0)|X=x,G_g=d,G_g+G_h=1}(y) = \Phi\Big(p(x)\boldsymbol{\alpha}_s^h(y) + dp(x)\boldsymbol{\beta}_s^{gh}(y) + \mathbb{1}\{\nu = t\}p(x)\boldsymbol{\gamma}_{st}^h(y)\Big)$$

via DR on the pooled sample $\{g,h\} \times \{s,t\} \setminus \{(g,t)\}$. $\mathrm{DTT}_{gt}(y)$ can subsequently be estimated as $\widehat{\mathrm{DTT}}_{gt}^{X,hs}(y) = \widehat{\mathrm{F}}_{gt}(y) - \widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1}(y)$.

S.2.3 Asymptotic Theory

Consider the following high-level uniform (in $\mathcal{Y}$) asymptotic linearity condition on the estimators: $\big(\widehat{\boldsymbol{\alpha}}_s^h(y)', \widehat{\boldsymbol{\beta}}_s^{gh}(y)', \widehat{\boldsymbol{\gamma}}_{st}^h(y)'\big)'$. This is a useful step in extending Theorem 2 to the covariate setting.

Assumption 3 (Asymptotic linear representation)**.** *For every $y \in \mathcal{Y}$ and every $(g,t,h,s) \in \mathcal{I}$, the estimators*

$$\big(\widehat{\boldsymbol{\alpha}}_s^h(y)',\, \widehat{\boldsymbol{\beta}}_s^{gh}(y)',\, \widehat{\boldsymbol{\gamma}}_{st}^h(y)'\big)'$$

admit the asymptotically linear representation

$$\sqrt{N}\begin{pmatrix}\widehat{\boldsymbol{\alpha}}_s^h(y) - \boldsymbol{\alpha}_s^h(y)\\ \widehat{\boldsymbol{\beta}}_s^{gh}(y) - \boldsymbol{\beta}_s^{gh}(y)\\ \widehat{\boldsymbol{\gamma}}_{st}^h(y) - \boldsymbol{\gamma}_{st}^h(y)\end{pmatrix} = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\begin{pmatrix}\mathcal{S}_s^h(W_j;y)\\ \mathcal{S}_s^{gh}(W_j;y)\\ \mathcal{S}_{st}^h(W_j;y)\end{pmatrix} + o_p(1),$$

where uniformly in $y \in \mathcal{Y}$, the influence functions are mean zero, have bounded second moments, and the collection $\big\{\mathcal{S}_s^h(\cdot;y),\ \mathcal{S}_s^{gh}(\cdot;y),\ \mathcal{S}_{st}^h(\cdot;y) : y \in \mathcal{Y}\big\}$ forms a P-Donsker class. In the covariate extension, ϕ is continuously differentiable with derivative $\dot{\phi}$ that is bounded and Lipschitz.

Assumption 3 is a high-level regularity condition on the DR coefficient processes. Its validity is established under primitive conditions in Chernozhukov, Fernandez-Val, and Melly (2013, Corollary 5.3 and Lemma E.3).

The following result extends Theorem 2 to the covariate setting.

Theorem 2. *Suppose Assumptions 1-X, 2-X, 3-X, 4, 5, 2, 1, and 3 hold, then for every $y \in \mathcal{Y}$ and every $(g,t,h,s) \in \mathcal{I}$,*

(a) $\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1} - \mathrm{F}_{Y_t(0)|G_g=1}\big) \rightsquigarrow \mathbb{G}_{gt}^{X,hs}$ *in* $\ell^\infty(\mathcal{Y})$ *and*

(b) $\sqrt{N}\big(\widehat{\mathrm{DTT}}_{gt}^{X,hs} - \mathrm{DTT}_{gt}\big) \rightsquigarrow \mathbb{H}_{gt}^{X,hs}$ *in* $\ell^\infty(\mathcal{Y})$

where $\mathbb{G}_{gt}^{X,hs}$ and $\mathbb{H}_{gt}^{X,hs}$ are tight Gaussian processes with mean 0 and respective covariance kernels $\Omega_{gt}^{X,hs}(y_1,y_2)$ and $\Sigma_{gt}^{X,hs}(y_1,y_2)$ defined on $\mathcal{Y}\times\mathcal{Y}$. The expressions of the covariance kernels are provided in the proof.

Proofs of results in Appendix S.2

Proof of Theorem 1

Part (a): The proof of identification in the covariate setting proceeds along the same lines as that of Theorem 1. Recall that $\boldsymbol{\alpha}_s^h(y)$, $\boldsymbol{\beta}_s^{gh}(y)$, and $\boldsymbol{\gamma}_{st}^h(y)$ are each vectors in $\mathbb{R}^{p_x}$. The argument follows the same three-step structure as in the baseline gt-case, with the additional use of covariate variation. In each step, strict monotonicity of $\Phi(\cdot)$ and Assumption 2 ensure that the corresponding coefficient vector is identified; see also the classical identification result of Manski (1988) for binary response models.

Under Assumption 1-X and the overlap condition Assumption 1,

$$\mathrm{F}_{Y_s(0)|X=x,G_g=0,G_g+G_h=1}(y) = \Phi\big(p(x)\boldsymbol{\alpha}_s^h(y)\big),\ \mathbb{P}_{X|G_g=1} - a.e.\ x$$

for control group h and pre-treatment period s. Since $\Phi(\cdot)$ is strictly increasing and $p(X)$ has full column rank, $\boldsymbol{\alpha}_s^h(y)$ is identified for every $y \in \mathcal{Y}$.

Under Assumption 1-X and the overlap condition Assumption 1,

$$\mathrm{F}_{Y_t(0)|X=x,G_g=0,G_g+G_h=1}(y) = \Phi\big(p(x)(\boldsymbol{\alpha}_s^h(y) + \boldsymbol{\gamma}_{st}^h(y))\big) =: \Phi\big(p(x)\boldsymbol{\eta}_{ht}(y)\big),\ \mathbb{P}_{X|G_g=1} - a.e.\ x$$

for all $y \in \mathcal{Y}$ with control group h and post-treatment period t. The identification of $\boldsymbol{\eta}_{ht}(y)$ follows from strict monotonicity of $\Phi(\cdot)$ and Assumption 2. In particular, $\Phi\big(p(x)\boldsymbol{\eta}_{ht}(y)\big) = \Phi\big(p(x)(\boldsymbol{\alpha}_s^h(y) + \boldsymbol{\gamma}_{st}^h(y))\big)$ is equivalent to $p(x)\boldsymbol{\eta}_{ht}(y) = p(x)\big(\boldsymbol{\alpha}_s^h(y) + \boldsymbol{\gamma}_{st}^h(y)\big)$ for $\mathbb{P}_{X|G_g=1}$-a.e. x and all $y \in \mathcal{Y}$. By the full rank condition on $p(X)$ (Assumption 2) and because $\boldsymbol{\alpha}_s^h(y)$ is identified in the first step, $\boldsymbol{\gamma}_{st}^h(y)$ is identified for every $y \in \mathcal{Y}$: $\boldsymbol{\gamma}_{st}^h(y) = \boldsymbol{\eta}_{ht}(y) - \boldsymbol{\alpha}_s^h(y)$.

Under Assumptions 1-X and 2-X,

$$\begin{aligned}\mathrm{F}_{Y_s(0)|X=x,G_g=1}(y) =& \mathrm{F}_{Y_s(1)|X=x,G_g=1}(y) = \mathrm{F}_{Y_s(1)|X=x,G_g=1,G_g+G_h=1}(y) \\ =& \Phi\big(p(x)(\boldsymbol{\alpha}_s^h(y) + \boldsymbol{\beta}_s^{gh}(y))\big) =: \Phi\big(p(x)\boldsymbol{\eta}_{gs}(y)\big),\ \mathbb{P}_{X|G_g=1} - a.e.\ x.\end{aligned}$$

$\boldsymbol{\eta}_{gs}(y)$ is identified by arguments akin to those of $\boldsymbol{\eta}_{ht}(y)$. Thus, $\boldsymbol{\beta}_s^{gh}(y)$ is identified for every $y \in \mathcal{Y}$: $\boldsymbol{\beta}_s^{gh}(y) = \boldsymbol{\eta}_{gs}(y) - \boldsymbol{\alpha}_s^h(y)$.

The conclusion follows by combining the above three parts.

Part (b): This follows directly from the identification of $\big(\boldsymbol{\alpha}_s^h(y)', \boldsymbol{\beta}_s^{gh}(y)', \boldsymbol{\gamma}_{st}^h(y)'\big)'$ in part (a) and Assumption 1-X.

Part (c): In addition to part (b), part (c) follows by integrating the identified conditional counterfactual distribution over the covariate distribution of the treated group. Under Assumption 3-X, observability is independent of $(Y_t(0), G_g, X)$, so the distribution of X among treated units is identified from observed treated units. Hence, $\mathrm{F}_{Y_t(0)|G_g=1}(y)$ is identified for every $y \in \mathcal{Y}$.

□

Proof of Theorem 2

Part (a): The proof of Theorem 2(a) is structured into several steps to facilitate the exposition of the asymptotic results.

Influence function decomposition: Adopting the following short-hand notation: $\widehat{\theta}_{st}^{gh}(y) := \widehat{\boldsymbol{\alpha}}_s^h(y) + \widehat{\boldsymbol{\beta}}_s^{gh}(y) + \widehat{\boldsymbol{\gamma}}_{st}^h(y)$ and $\theta_{st}^{gh}(y) := \boldsymbol{\alpha}_s^h(y) + \boldsymbol{\beta}_s^{gh}(y) + \boldsymbol{\gamma}_{st}^h(y)$, consider the following decomposition using (S.5) and (S.6):

$$
\begin{aligned}
&\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1}-\mathrm{F}_{Y_t(0)|G_g=1}\big)(y)\\
=&\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j=g\}}{\hat{\pi}_t\hat{p}_g}\Phi\Big(p(X_j)\widehat{\theta}^{gh}_{st}(y)\Big)-\mathbb{E}\Big[\frac{S_t\mathbb{1}\{G=g\}}{\pi_t p_g}\Phi\Big(p(X)\theta^{gh}_{st}(y)\Big)\Big]\Big\}\\
=&\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j=g\}}{\pi_t p_g}\Big(\Phi\big(p(X_j)\widehat{\theta}^{gh}_{st}(y)\big)-\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)\Big)\Big\}\\
&-\frac{1}{N}\sum_{j=1}^{N}\Big\{S_{jt}\mathbb{1}\{G_j=g\}\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)\Big\}\times\frac{\sqrt{N}\big(\hat{\pi}_t\hat{p}_g-\pi_t p_g\big)}{\pi_t p_g\hat{\pi}_t\hat{p}_g}\\
&+\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j=g\}\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)-\mathbb{E}\big[S_t\mathbb{1}\{G=g\}\Phi\big(p(X)\theta^{gh}_{st}(y)\big)\big]}{\pi_t p_g}\Big\}\\
&\underbrace{-\frac{\sqrt{N}(\hat{\pi}_t\hat{p}_g-\pi_t p_g)}{\hat{\pi}_t\hat{p}_g}\frac{1}{N}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j=g\}}{\pi_t p_g}\Big(\Phi\big(p(X_j)\widehat{\theta}^{gh}_{st}(y)\big)-\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)\Big)\Big\}}_{\widehat{\mathcal{R}}^{(1)}_{gt,X}(y)}.
\end{aligned}
$$

Observe the following decomposition of the term

$$
\frac{\sqrt{N}\big(\hat{\pi}_t\hat{p}_g-\pi_t p_g\big)}{\pi_t p_g\hat{\pi}_t\hat{p}_g}=\frac{1}{\pi_t p_g\hat{\pi}_t}\sqrt{N}(\hat{\pi}_t-\pi_t)+\frac{1}{p_g\hat{\pi}_t\hat{p}_g}\sqrt{N}(\hat{p}_g-p_g).
$$

In addition to the differentiability of $\Phi(\cdot)$ in Assumption 5, obtain the following decomposition:

$$
\begin{aligned}
&\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1}-\mathrm{F}_{Y_t(0)|G_g=1}\big)(y)\\
=&\frac{1}{N}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j=g\}}{\pi_t p_g}\phi\big(p(X_j)\theta^{gh}_{st}(y)\big)p(X_j)\Big\}\sqrt{N}(\widehat{\theta}^{gh}_{st}-\theta^{gh}_{st})(y)\\
&-\frac{1}{N}\sum_{j=1}^{N}\Big\{S_{jt}\mathbb{1}\{G_j=g\}\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)\Big\}\Big(\frac{1}{\pi_t p_g\hat{\pi}_t}\sqrt{N}(\hat{\pi}_t-\pi_t)+\frac{1}{p_g\hat{\pi}_t\hat{p}_g}\sqrt{N}(\hat{p}_g-p_g)\Big)\\
&+\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j=g\}\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)-\mathbb{E}\big[S_t\mathbb{1}\{G=g\}\Phi\big(p(X)\theta^{gh}_{st}(y)\big)\big]}{\pi_t p_g}\Big\}\\
&\underbrace{-\frac{\sqrt{N}\big(\hat{\pi}_t\hat{p}_g-\pi_t p_g\big)}{\hat{\pi}_t\hat{p}_g}\frac{1}{N}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j=g\}}{\pi_t p_g}\Big(\Phi\big(p(X_j)\widehat{\theta}^{gh}_{st}(y)\big)-\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)\Big)\Big\}}_{\widehat{\mathcal{R}}^{(1)}_{gt,X}(y)}\\
&+\underbrace{\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j=g\}}{\pi_t p_g}\Big(\Phi\big(p(X_j)\widehat{\theta}^{gh}_{st}(y)\big)-\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)-\phi\big(p(X_j)\theta^{gh}_{st}(y)\big)p(X_j)(\widehat{\theta}^{gh}_{st}-\theta^{gh}_{st})(y)\Big)\Big\}}_{\widehat{\mathcal{R}}^{(2)}_{gt,X}(y)}.
\end{aligned}
\tag{S.7}
$$

Asymptotic negligibility of remainder terms: Let $\Delta^{gh}_{st}(y):=\big(\widehat{\theta}^{gh}_{st}-\theta^{gh}_{st}\big)(y)$. Observe that $\sup_{y\in\mathcal{Y}}\|\Delta^{gh}_{st}(y)\|=o_p(1)$ and $\sup_{y\in\mathcal{Y}}\sqrt{N}\,\|\Delta^{gh}_{st}(y)\|=\mathcal{O}_p(1)$, under Assumption 3. It is first shown that $\sup_{y\in\mathcal{Y}}\big|\widehat{\mathcal{R}}^{(1)}_{gt,X}(y)\big|=o_p(1)$. By the mean-value theorem and the boundedness of ϕ in Assumption 5,

$$
\Big|\Phi\big(p(X_j)\widehat{\theta}^{gh}_{st}(y)\big)-\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)\Big|\le C\,\|p(X_j)\|\,\|\Delta^{gh}_{st}(y)\|.
$$

In addition to the result $\dfrac{\sqrt{N}\left(\hat{\pi}_t\hat{p}_g - \pi_t p_g\right)}{\pi_t p_g \hat{\pi}_t \hat{p}_g} = \mathcal{O}_p(1)$, which holds under Assumptions 3 and 4 (see (S.1)),

$$\sup_{y\in\mathcal{Y}} \left|\widehat{\mathcal{R}}^{(1)}_{gt,X}(y)\right| \leq \left|\frac{\sqrt{N}(\hat{\pi}_t\hat{p}_g - \pi_t p_g)}{\hat{\pi}_t\hat{p}_g}\right| \frac{1}{\pi_t p_g}\frac{1}{N}\sum_{j=1}^{N} S_{jt}\mathbb{1}\{G_j = g\} \sup_{y\in\mathcal{Y}}\left|\Phi\big(p(X_j)\widehat{\theta}^{gh}_{st}(y)\big) - \Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)\right|$$
$$\leq \left|\frac{\sqrt{N}(\hat{\pi}_t\hat{p}_g - \pi_t p_g)}{\hat{\pi}_t\hat{p}_g}\right| \frac{C}{\pi_t p_g}\left(\frac{1}{N}\sum_{j=1}^{N}\|p(X_j)\|\right)\sup_{y\in\mathcal{Y}}\|\Delta^{gh}_{st}(y)\|.$$

Therefore, in addition to the moment condition on $p(X)$ in Assumption 3-X, $\sup_{y\in\mathcal{Y}}\left|\widehat{\mathcal{R}}^{(1)}_{gt,X}(y)\right| = \mathcal{O}_p(1)\cdot \mathcal{O}_p(1)\cdot o_p(1) = o_p(1)$.

To handle $\widehat{\mathcal{R}}^{(2)}_{gt,X}$, apply Taylor's theorem to $\vartheta \mapsto \Phi(p(X_j)\vartheta)$. Since $\dot{\phi}$ is bounded under Assumption 3, there exists a constant $C < \infty$ such that, uniformly in $y \in \mathcal{Y}$,

$$\left|\Phi\big(p(X_j)\widehat{\theta}^{gh}_{st}(y)\big) - \Phi\big(p(X_j)\theta^{gh}_{st}(y)\big) - \phi\big(p(X_j)\theta^{gh}_{st}(y)\big)p(X_j)\Delta^{gh}_{st}(y)\right| \leq C\|p(X_j)\|^2\|\Delta^{gh}_{st}(y)\|^2.$$

Therefore, using Assumption 3-X and Assumption 3,

$$\sup_{y\in\mathcal{Y}}\left|\widehat{\mathcal{R}}^{(2)}_{gt,X}(y)\right| \leq \frac{C}{\sqrt{N}}\sum_{j=1}^{N}\frac{S_{jt}\mathbb{1}\{G_j = g\}}{\pi_t p_g}\|p(X_j)\|^2 \sup_{y\in\mathcal{Y}}\|\Delta^{gh}_{st}(y)\|^2$$
$$= C\sqrt{N}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{S_{jt}\mathbb{1}\{G_j = g\}}{\pi_t p_g}\|p(X_j)\|^2\right) O_p(N^{-1}) = o_p(1).$$

Since the remainder terms are asymptotically negligible uniformly in $y \in \mathcal{Y}$, i.e., $\sup_{y\in\mathcal{Y}}\left|\widehat{\mathcal{R}}^{(1)}_{gt,X}(y)\right| = o_p(1)$ and $\sup_{y\in\mathcal{Y}}\left|\widehat{\mathcal{R}}^{(2)}_{gt,X}(y)\right| = o_p(1)$, it follows from the decomposition in (S.7) that

$$\begin{aligned}
&\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1} - \mathrm{F}_{Y_t(0)|G_g=1}\big)(y)\\
&= \frac{1}{N}\sum_{j=1}^{N}\left\{\frac{S_{jt}\mathbb{1}\{G_j = g\}}{\pi_t p_g}\phi\big(p(X_j)\theta^{gh}_{st}(y)\big)p(X_j)\right\}\sqrt{N}(\widehat{\theta}^{gh}_{st} - \theta^{gh}_{st})(y)\\
&\quad - \frac{1}{N}\sum_{j=1}^{N}\left\{S_{jt}\mathbb{1}\{G_j = g\}\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big)\right\}\left(\frac{1}{\pi_t p_g\hat{\pi}_t}\sqrt{N}(\hat{\pi}_t - \pi_t) + \frac{1}{p_g\hat{\pi}_t\hat{p}_g}\sqrt{N}(\hat{p}_g - p_g)\right)\\
&\quad + \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left\{\frac{S_{jt}\mathbb{1}\{G_j = g\}\Phi\big(p(X_j)\theta^{gh}_{st}(y)\big) - \mathbb{E}\left[S_t\mathbb{1}\{G = g\}\Phi\big(p(X)\theta^{gh}_{st}(y)\big)\right]}{\pi_t p_g}\right\}\\
&\quad + o_p(1).
\end{aligned}$$

Asymptotic linearity: The next part of the proof studies the non-negligible summands of the above in turn.

First, thanks to Assumption 3-X and Assumption 5, it follows from the Uniform Weak Law of Large Numbers that

$$\sup_{y\in\mathcal{Y}}\left\|\frac{1}{N}\sum_{j=1}^{N}\left\{\frac{S_{jt}\mathbb{1}\{G_j = g\}}{\pi_t p_g}\phi\big(p(X_j)\theta^{gh}_{st}(y)\big)p(X_j)\right\} - \mathbb{E}\left[\phi\big(p(X)\theta^{gh}_{st}(y)\big)p(X) \mid G_g = 1\right]\right\| \xrightarrow{p} 0$$

where $\mathbb{E}\big[\phi\big(p(X)\theta_{st}^{gh}(y)\big)p(X) \mid G_g = 1\big] = \mathbb{E}\Big[\frac{S_t \mathbb{1}\{G = g\}}{\pi_t p_g}\phi\big(p(X)\theta_{st}^{gh}(y)\big)p(X)\Big]$ under Assumption 3-X. Also, under Assumption 3,

$$\sqrt{N}(\widehat{\theta}_{st}^{gh} - \theta_{st}^{gh})(y) = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\big(\mathcal{S}_s^h(W_j; y) + \mathcal{S}_s^{gh}(W_j; y) + \mathcal{S}_{st}^h(W_j; y)\big) + o_p(1).$$

Combining both terms gives uniformly in $y \in \mathcal{Y}$:

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\Big\{\frac{S_{jt}\mathbb{1}\{G_j = g\}}{\pi_t p_g}\phi\big(p(X_j)\theta_{st}^{gh}(y)\big)p(X_j)\Big\}\sqrt{N}(\widehat{\theta}_{st}^{gh} - \theta_{st}^{gh})(y) \\
=&\mathbb{E}\big[\phi\big(p(X)\theta_{st}^{gh}(y)\big)p(X) \mid G_g = 1\big]\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\big(\mathcal{S}_s^h(W_j; y) + \mathcal{S}_s^{gh}(W_j; y) + \mathcal{S}_{st}^h(W_j; y)\big) + o_p(1).
\end{aligned}$$

Second, by Assumption 3-X and Assumption 4, it follows from the UWLLN and the continuous mapping theorem that

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\Big\{S_{jt}\mathbb{1}\{G_j = g\}\Phi\big(p(X_j)\theta_{st}^{gh}(y)\big)\Big\}\Big(\frac{1}{\pi_t p_g \hat{\pi}_t}\sqrt{N}(\hat{\pi}_t - \pi_t) + \frac{1}{p_g \hat{\pi}_t \hat{p}_g}\sqrt{N}(\hat{p}_g - p_g)\Big) \\
=&\mathbb{E}\big[S_t\mathbb{1}\{G = g\}\Phi\big(p(X)\theta_{st}^{gh}(y)\big)\big]\Big(\frac{1}{\pi_t^2 p_g}\sqrt{N}(\hat{\pi}_t - \pi_t) + \frac{1}{\pi_t p_g^2}\sqrt{N}(\hat{p}_g - p_g)\Big) + o_p(1) \\
=&\frac{\mathrm{F}_{Y_t(0)|G_g=1}(y)}{\pi_t}\sqrt{N}(\hat{\pi}_t - \pi_t) + \frac{\mathrm{F}_{Y_t(0)|G_g=1}(y)}{p_g}\sqrt{N}(\hat{p}_g - p_g) + o_p(1)
\end{aligned}$$

Third, since $\frac{\mathbb{E}\big[S_t\mathbb{1}\{G = g\}\Phi\big(p(X)\theta_{st}^{gh}(y)\big)\big]}{\pi_t p_g} = \int \mathrm{F}_{Y_t(0)|X=x,\, G_g=1}(y)\, dF_{X|G_g=1,\, S_t=1}(x)$ under Assumption 1-X and the independence of observability S_t from $(Y_t(0), G_g, X)$ (Assumption 3-X),

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Bigg\{\frac{S_{jt}\mathbb{1}\{G_j = g\}\Phi\big(p(X_j)\theta_{st}^{gh}(y)\big) - \mathbb{E}\big[S_t\mathbb{1}\{G = g\}\Phi\big(p(X)\theta_{st}^{gh}(y)\big)\big]}{\pi_t p_g}\Bigg\} \\
&\qquad = \int \mathrm{F}_{Y_t(0)|X=x,\, G_g=1}(y)\, \sqrt{N}\big(\widehat{F}_{X|G_g=1,S_t=1} - F_{X|G_g=1,S_t=1}\big)(dx)
\end{aligned}$$

where $\widehat{F}_{X|G_g=1,S_t=1}$ denotes the empirical conditional CDF.

It therefore follows that

$$\sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1} - \mathrm{F}_{Y_t(0)|G_g=1}\big)(y)$$

$$= \underbrace{\mathbb{E}\big[\phi\big(p(X)\theta^{gh}_{st}(y)\big)p(X) \mid G_g = 1\big]\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\big(\mathcal{S}^h_s(W_j;y) + \mathcal{S}^{gh}_s(W_j;y) + \mathcal{S}^h_{st}(W_j;y)\big)}_{\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi^A_{F,X}(W_j;y)}$$

$$\underbrace{-\frac{\mathrm{F}_{Y_t(0)|G_g=1}(y)}{\pi_t}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}(S_{jt}-\pi_t) - \frac{\mathrm{F}_{Y_t(0)|G_g=1}(y)}{p_g}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}(\mathbb{1}\{G_j=g\}-p_g)}_{\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi^B_{F,X}(W_j;y)}$$

$$\underbrace{+\int \mathrm{F}_{Y_t(0)|X=x,\,G_g=1}(y)\,\sqrt{N}\big(\widehat{F}_{X|G_g=1,S_t=1} - F_{X|G_g=1,S_t=1}\big)(dx)}_{\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi^C_{F,X}(W_j;y)} + o_p(1)$$

$$=: \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi^{X,hs}_{gt}(W_j;y) + o_p(1).$$

Donsker property of the associated function class: The next step is to investigate the Donsker property of the function class associated with $\Psi^{X,hs}_{gt}(W_j;y)$. The map $c(y) := \mathbb{E}\big[\phi\big(p(X)\theta^{gh}_{st}(y)\big)p(X) \mid G_g = 1\big]$ satisfies $\sup_{y\in\mathcal{Y}}\|c(y)\| < \infty$ under Assumption 5 and the dominance condition on $p(X)$ in Assumption 3. $y \mapsto c(y)$ is bounded, hence multiplication by $c(y)$ preserves Donskerness. In addition to the Donskerness of the function class associated with the DR coefficient process (Assumption 3), the functional class associated with $\Psi^A_{F,X}(W;y)$ is P-Donsker.

Let $Z(W) := \Big(\frac{S_t}{\pi_t} - 1\Big) + \Big(\frac{\mathbb{1}\{G=g\}}{p_g} - 1\Big)$. Then $\Psi^B_{F,X}(W;y) = -Z(W)\,\mathrm{F}_{Y_t(0)|G_g=1}(y)$. The index y enters only through the deterministic scalar $\mathrm{F}_{Y_t(0)|G_g=1}(y) \in [0,1]$. Thus, $\{\, w \mapsto -Z(w)\,a : a \in [0,1]\,\}$ is a uniformly bounded one-dimensional linear class, hence P-Donsker.

The third summand $\Psi^C_{F,X}(W_j;y)$ is cast in the same form as the $\widehat{G}_k(f)$ empirical process on page 2224 with the limit in equation 4.2 of Chernozhukov, Fernandez-Val, and Melly (2013) for distribution regression. It follows from Chernozhukov, Fernandez-Val, and Melly (2013, Theorem 4.1) that the associated function class of $\Psi^C_{F,X}(W_j;y)$ is P-Donsker.

Each of the function classes corresponding to $\Psi^A_{F,X}(W_j;y)$, $\Psi^B_{F,X}(W_j;y)$, $\Psi^C_{F,X}(W_j;y)$ is P-Donsker, and P-Donsker classes are closed under finite sums. Therefore the associated class of $\Psi^{X,hs}_{gt}(W_j;y)$ is P-Donsker.

$\sup_{y\in\mathcal{Y}}\mathbb{E}\Big|\Psi^{X,hs}_{gt}(W;y)\Big|^2 < \infty$ under Assumption 5 and Assumption 3-X. Thus, for any arbitrary finite subset $\{y_1,\dots,y_L\} \subset \mathcal{Y}$, it follows from the Multivariate Lindeberg–Lévy Central Limit Theorem that

$$\Big(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi^{X,hs}_{gt}(W_j;y_1),\dots,\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Psi^{X,hs}_{gt}(W_j;y_L)\Big)$$

converges in distribution to the multivariate normal with the (l,l')'th entry of the covariance matrix: $\mathbb{E}\big[\Psi^{X,hs}_{gt}(W;y_l)\Psi^{X,hs}_{gt}(W;y_{l'})\big]$. The P-Donsker property supplies asymptotic equicontinuity, so the empirical process converges weakly in $\ell^\infty(\mathcal{Y})$ to a tight Gaussian process, namely $\mathbb{G}^{X,hs}_{gt}$, with covariance kernel $\Omega^{X,hs}_{gt}(y_1,y_2) := \mathbb{E}\big[\Psi^{X,hs}_{gt}(W;y_1)\Psi^{X,hs}_{gt}(W;y_2)\big]$; see Kosorok (2007, Theorem 2.1).

Part (b): From (S.2) and part (a) above, the following decomposition holds:

$$
\begin{aligned}
\sqrt{N}\big(\widehat{\mathrm{DTT}}_{gt}^{X,hs} - \mathrm{DTT}_{gt}\big)(y)
&= \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t(1)|G_g=1} - \mathrm{F}_{Y_t(1)|G_g=1}\big)(y) - \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1} - \mathrm{F}_{Y_t(0)|G_g=1}\big)(y)\\
&= \sqrt{N}\big(\widehat{\mathrm{F}}_{gt} - \mathrm{F}_{gt}\big)(y) - \sqrt{N}\big(\widehat{\mathrm{F}}_{Y_t^{X,hs}(0)|G_g=1} - \mathrm{F}_{Y_t(0)|G_g=1}\big)(y)\\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Big\{\psi_{gt}(W_j;y) - \Psi_{gt}^{X,hs}(W_j;y)\Big\} + o_p(1)\\
&=: \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\Upsilon_{gt}^{X,hs}\big(W_j;y\big) + o_p(1).
\end{aligned}
$$

From part (a) above and Lemma 1, conclude using the Donsker and asymptotic-equicontinuity arguments in the proof of part (c) of Theorem 2. The covariance kernel of the limiting process is $\Sigma_{gt}^{X,hs}(y_1,y_2) := \mathbb{E}\big[\Upsilon_{gt}^{X,hs}(W;y_1)\Upsilon_{gt}^{X,hs}(W;y_2)\big]$.

□

S.3 Uniform Inference

The results in this section are specialised to $(g,t) \in \mathcal{G}\setminus\{\infty\}\times[g:T]$. Results on convex-weighted DFs, convex-weighted DTTs, and QTTs constructed from convex-weighted DFs follow straightforwardly. Let $\widehat{I}_{gt}^{(1)}(y) = \big[\widehat{L}_{gt}^{(1)}(y),\ \widehat{U}_{gt}^{(1)}(y)\big]$ and $\widehat{I}_{gt}^{(0)}(y) = \big[\widehat{L}_{gt}^{(0)}(y),\ \widehat{U}_{gt}^{(0)}(y)\big]$ denote joint p-level confidence bands for $\mathrm{F}_{Y_t(1)|G_g=1}(y)$ and $\mathrm{F}_{Y_t(0)|G_g=1}(y)$, respectively, and let $\widehat{I}_{gt}^{\mathrm{DTT}}(y) = \big[\widehat{L}_{gt}^{\mathrm{DTT}}(y),\ \widehat{U}_{gt}^{\mathrm{DTT}}(y)\big]$ be the p-level confidence band for DTT_{gt}. The confidence bands can be constructed using steps 1-5 of Algorithm 1 in Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2019) — see also Chernozhukov, Fernandez-Val, and Melly (2013, Algorithm 3). For the band-inversion step used below, the estimated DFs and DF-band endpoints are clipped to $[0,1]$ and monotonised over $\mathcal{Y}$ before taking left inverses. This rearrangement enforces the nondecreasing property required of CDFs for the QF and QTT construction and is first-order equivalent for the estimated DFs and associated endpoints under standard rearrangement arguments; see Chernozhukov, Fernandez-Val, and Galichon (2010) and Chernozhukov, Fernandez-Val, and Melly (2013, p. 2252). The following states the validity of confidence bands on DFs and DTT.

Result 1 (Coverage of Uniform Confidence Bands on DFs and DTT)**.** *Under the conditions of Theorem 3,*

$$
\begin{aligned}
&\liminf_{N\to\infty}\mathbb{P}\Big(\big\{\mathrm{F}_{Y_t(1)|G_g=1}(y) \in \widehat{I}_{gt}^{(1)}(y)\big\}\bigcap\big\{\mathrm{F}_{Y_t(0)|G_g=1}(y) \in \widehat{I}_{gt}^{(0)}(y)\big\}\ \forall y \in \mathcal{Y}\Big) \geq p, \text{ and}\\
&\liminf_{N\to\infty}\mathbb{P}\Big(\mathrm{DTT}_{gt}(y) \in \widehat{I}_{gt}^{\mathrm{DTT}}(y)\ \forall y\in\mathcal{Y}\Big) \geq p
\end{aligned}
$$

by Lemma SA.1(b) in the Supplemental material of Chernozhukov, Fernandez-Val, and Melly (2013).

This paper adopts the band-inversion approach of Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2019) to conduct uniform inference for QFs and QTT. A crucial ingredient for using the authors' results is valid confidence bands on DFs (Result 1). For completeness, theoretical results from Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2019) are adapted to QFs and QTT. The confidence bands on the QFs are given by $\widehat{I}_{gt}^{(1)\leftarrow}(\tau) = \big[\widehat{U}_{gt}^{(1)\leftarrow}(\tau),\ \widehat{L}_{gt}^{(1)\leftarrow}(\tau)\big]$ and $\widehat{I}_{gt}^{(0)\leftarrow}(\tau) = \big[\widehat{U}_{gt}^{(0)\leftarrow}(\tau),\ \widehat{L}_{gt}^{(0)\leftarrow}(\tau)\big]$, $\tau\in(0,1)$, respectively, where $\widehat{L}^{\leftarrow}$ and $\widehat{U}^{\leftarrow}$ denote the left inverses of $\widehat{L}$ and $\widehat{U}$ — see Definition 1. The following definition is essential in deriving the confidence bands of QTT_{gt}.

Definition 1 (Minkowski Difference)**.** *The Minkowski difference between two subsets* I *and* J *of a vector space is* $I \ominus J := \{i - j : i \in I, j \in J\}$. *If* I *and* J *are intervals* $[i_1, i_2]$ *and* $[j_1, j_2]$, *then* $I \ominus J = [i_1, i_2] \ominus [j_1, j_2] = [i_1 - j_2, i_2 - j_1]$.

Following Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2019), the confidence bands on QTT_{gt} are given by $\widehat{I}^{\mathrm{QTT}\leftarrow}_{gt}(\tau) = \widehat{I}^{(1)\leftarrow}_{gt}(\tau) \ominus \widehat{I}^{(0)\leftarrow}_{gt}(\tau) =: \big[\widehat{L}^{\mathrm{QTT}}_{gt}(\tau),\ \widehat{U}^{\mathrm{QTT}}_{gt}(\tau)\big]$. The following states the validity of the band-inverted confidence bands on QFs and QTT à la Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2019).

Result 2 (Coverage of Uniform Confidence Bands on QFs and QTT)**.** *Under the conditions of Theorem 3,*

$$\liminf_{N\to\infty} \mathbb{P}\Bigg(\Big\{\mathrm{F}^{\leftarrow}_{Y_t(1)|G_g=1}(\tau) \in \widehat{I}^{(1)\leftarrow}_{gt}(\tau),\\ \mathrm{F}^{\leftarrow}_{Y_t(0)|G_g=1}(\tau) \in \widehat{I}^{(0)\leftarrow}_{gt}(\tau),\\ \mathrm{QTT}_{gt}(\tau) \in \widehat{I}^{\mathrm{QTT}\leftarrow}_{gt}(\tau),\ \forall \tau \in (0,1)\Big\}\Bigg) \geq p.$$

This follows by Result 1 and Theorem 2 of Chernozhukov, Fernandez-Val, Melly, and Wüthrich (2019).

S.4 Simulations

This section conducts two sets of simulation studies to examine the small sample performance of the distributional DiD estimator and the functional over-identifying restrictions tests.

S.4.1 Estimation and Inference

This simulation study is designed to assess the finite-sample performance of the proposed distributional DiD procedure. Three data-generating processes are considered. The first is an empirical design, labelled DGP0, which is calibrated directly to the empirical application. The remaining two designs are stylised benchmark models, one discrete and one continuous.

Under the benchmark designs, the latent outcome is generated as

$$\dot{Y}_i = \alpha + D_i\beta + t_i\gamma + D_i t_i\delta + U_i,$$

where $D_i \overset{i.i.d.}{\sim} \mathrm{Ber}(0.5)$, $t_i = \mathbb{1}\{i > N/2\}$, U_i is an i.i.d. disturbance, $\alpha = 0.1$, $\beta = 0.2$, $\gamma = -0.1$, and $\delta = 0$. The disturbance U_i is taken to follow either the standard normal distribution or an asymmetric Laplace distribution. The observed outcome is then defined by

$$Y_i = \begin{cases} \max\{\lceil \dot{Y}_i + 1 \rceil, 0\}, & \text{DGP1}, \\ \dot{Y}_i, & \text{DGP2}. \end{cases}$$

Hence, DGP1 yields a discrete and censored outcome, whereas DGP2 yields a continuous outcome.

DGP0 is calibrated to the empirical distribution of the outcome variable in Section 5. Let $\widehat{\mathrm{F}}_{00}$, $\widehat{\mathrm{F}}_{01}$, and $\widehat{\mathrm{F}}_{10}$ denote the empirical CDFs of the outcome for untreated blocks in the pre-treatment and post-treatment periods, and for treated blocks in the pre-treatment period, respectively. Under the null of no treatment effect, the treated post-treatment counterfactual distribution is constructed using the additive DiD relation under the identity link:

$$\widehat{\mathrm{F}}_{11}^{0}(y) = \widehat{\mathrm{F}}_{10}(y) + \widehat{\mathrm{F}}_{01}(y) - \widehat{\mathrm{F}}_{00}(y).$$

$\widehat{\mathrm{F}}_{11}^{0}(y)$ is truncated to lie in $[0, 1]$, monotonised over the empirical outcome grid, and normalised so that its upper endpoint equals one. Simulation draws are then generated from the empirical quantile functions associated with $\widehat{\mathrm{F}}_{00}$, $\widehat{\mathrm{F}}_{01}$, $\widehat{\mathrm{F}}_{10}$, and the projected version of $\widehat{\mathrm{F}}_{11}^{0}$. In this way, DGP0 preserves the truncated empirical support and the principal distributional features of the application data while imposing the null hypothesis $\mathrm{DTT}(y) = 0$ for all $y \in \mathcal{Y}$.

The designs are calibrated so that there is no treatment effect in the underlying data-generating process, and hence provide a finite-sample assessment of estimation error, confidence-band coverage, and sensitivity to the working CDF. DGP0 is useful for assessing performance under an empirically calibrated discrete design, while DGP1 and DGP2 provide complementary evidence under stylised discrete and continuous settings. In the latter designs, the comparison across normal, uniform, and identity links should be interpreted as a robustness exercise for the working-CDF specification rather than as imposing exact size validity for every link by construction.

Tables S.1 to S.3 report simulation results for DGP0, DGP1, and DGP2, respectively. For DGP1 and DGP2, each panel is characterised by the distribution of U, namely the standard normal or the Asymmetric Laplace Distribution $ALD(0, 1, \kappa)$ with $\kappa \in \{0.1, 0.25, 0.5\}$. The inclusion of the ALD cases is useful in examining the performance of the procedure when U is asymmetric. For each table, sample sizes $N \in \{200, 400, 600, 800, 1000\}$ are considered. For DGP2, the evaluation grid is formed from the sample quantiles indexed by $\{\tau_\ell\}_{\ell=1}^{101}$, whereas for the discrete designs it is formed from the realised support after trimming the upper tail as implemented in the simulation code. Results are summarised by $\mathcal{L}_2(\widehat{\mathrm{DTT}})$ where $\mathcal{L}_p(\widehat{\mathrm{F}}) := \left(\frac{1}{K}\sum_{k=1}^{K} |\widehat{\mathrm{F}}(y_k) - F(y_k)|^p\right)^{1/p}$, where K denotes the number of grid-points used in estimating the function $F(\cdot)$, and the empirical rejection rate of the 10% uniform confidence band for DTT. For each design, 499 non-parametric bootstrap samples are used within each of the 500 Monte Carlo replications.

Table S.1: DGP0: empirically calibrated discrete design

	DiD, Φ - Normal		DiD, Φ - Unif.		DiD, Φ - Identity		CiC	
N	$\mathcal{L}_2(\widehat{\mathrm{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\mathrm{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\mathrm{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\mathrm{DTT}})$	10% Rej.
200	0.108	0.144	0.082	0.156	0.084	0.120	0.082	0.168
400	0.066	0.076	0.059	0.108	0.059	0.086	0.069	0.190
600	0.053	0.094	0.049	0.110	0.049	0.094	0.064	0.222
800	0.044	0.080	0.042	0.108	0.042	0.100	0.060	0.230
1000	0.037	0.088	0.036	0.088	0.036	0.086	0.061	0.234

Several patterns emerge from Tables S.1 to S.3. First, the proposed distributional DiD estimator performs well across all three designs, with $\mathcal{L}_2(\widehat{\mathrm{DTT}})$ generally declining as the sample size increases. This pattern is evident in the empirically calibrated design as well as in the stylised discrete and continuous designs, which is consistent with satisfactory finite-sample behaviour.

Secondly, the choice of working CDF matters for finite-sample performance, but the differences are moderate between the normal and identity links. In contrast, the uniform link tends to exhibit somewhat higher rejection frequencies, especially in the stylised designs. The identity link performs

Table S.2: DGP1: stylised discrete design

	DiD, Φ - Normal		DiD, Φ - Unif.		DiD, Φ - Identity		CiC	
N	$\mathcal{L}_2(\widehat{\text{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\text{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\text{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\text{DTT}})$	10% Rej.
	$U \sim \text{ALD}(0, 1, 0.10)$							
200	0.124	0.100	0.105	0.132	0.106	0.094	0.107	0.056
400	0.078	0.066	0.075	0.156	0.075	0.112	0.077	0.072
600	0.064	0.082	0.063	0.158	0.063	0.116	0.067	0.094
800	0.054	0.090	0.054	0.138	0.054	0.102	0.058	0.086
1000	0.048	0.078	0.047	0.114	0.047	0.088	0.053	0.106
	$U \sim \text{ALD}(0, 1, 0.25)$							
200	0.111	0.088	0.104	0.162	0.105	0.094	0.113	0.086
400	0.075	0.090	0.075	0.140	0.075	0.080	0.087	0.112
600	0.063	0.108	0.063	0.152	0.063	0.118	0.080	0.180
800	0.053	0.096	0.053	0.130	0.053	0.104	0.076	0.224
1000	0.047	0.094	0.047	0.102	0.047	0.088	0.074	0.260
	$U \sim \text{ALD}(0, 1, 0.50)$							
200	0.112	0.100	0.104	0.132	0.105	0.084	0.124	0.132
400	0.076	0.090	0.075	0.138	0.075	0.102	0.104	0.186
600	0.063	0.100	0.062	0.130	0.062	0.106	0.101	0.226
800	0.054	0.094	0.053	0.108	0.053	0.094	0.100	0.324
1000	0.048	0.084	0.047	0.106	0.047	0.092	0.099	0.366
	$U \sim \mathcal{N}(0, 1)$							
200	0.118	0.106	0.100	0.158	0.102	0.124	0.210	0.372
400	0.086	0.098	0.071	0.106	0.071	0.098	0.202	0.452
600	0.057	0.094	0.056	0.136	0.056	0.110	0.200	0.480
800	0.050	0.094	0.050	0.104	0.050	0.088	0.202	0.542
1000	0.043	0.088	0.043	0.102	0.043	0.084	0.203	0.552

Table S.3: DGP2: stylised continuous design

N	DiD, Φ - Normal		DiD, Φ - Unif.		DiD, Φ - Identity		CiC	
	$\mathcal{L}_2(\widehat{\text{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\text{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\text{DTT}})$	10% Rej.	$\mathcal{L}_2(\widehat{\text{DTT}})$	10% Rej.
	$U \sim \text{ALD}(0, 1, 0.10)$							
200	0.118	0.104	0.109	0.144	0.110	0.104	0.110	0.036
400	0.079	0.044	0.078	0.156	0.078	0.092	0.079	0.058
600	0.065	0.076	0.065	0.180	0.065	0.112	0.066	0.066
800	0.056	0.092	0.056	0.164	0.056	0.106	0.056	0.070
1000	0.050	0.080	0.049	0.130	0.049	0.090	0.050	0.060
	$U \sim \text{ALD}(0, 1, 0.25)$							
200	0.118	0.100	0.109	0.146	0.110	0.108	0.110	0.040
400	0.079	0.048	0.078	0.150	0.078	0.090	0.079	0.058
600	0.065	0.076	0.065	0.174	0.065	0.102	0.066	0.064
800	0.056	0.090	0.056	0.160	0.056	0.102	0.056	0.074
1000	0.049	0.074	0.049	0.114	0.049	0.074	0.050	0.060
	$U \sim \text{ALD}(0, 1, 0.50)$							
200	0.119	0.114	0.109	0.138	0.109	0.100	0.109	0.046
400	0.079	0.046	0.078	0.156	0.078	0.088	0.079	0.060
600	0.066	0.064	0.065	0.176	0.065	0.112	0.065	0.062
800	0.056	0.090	0.056	0.166	0.056	0.112	0.056	0.080
1000	0.050	0.076	0.049	0.126	0.049	0.078	0.049	0.062
	$U \sim \mathcal{N}(0, 1)$							
200	0.120	0.082	0.110	0.168	0.111	0.128	0.111	0.048
400	0.080	0.058	0.079	0.158	0.079	0.124	0.078	0.064
600	0.064	0.060	0.063	0.156	0.063	0.110	0.064	0.052
800	0.056	0.090	0.056	0.156	0.056	0.116	0.056	0.074
1000	0.050	0.074	0.049	0.108	0.049	0.084	0.050	0.058

competitively throughout and often tracks the normal link closely. In the empirically calibrated design, all three versions of the proposed estimator deliver similar bias magnitudes, with the uniform link showing slightly greater over-rejection in some cases.

Finally, the competing Changes-in-Changes (CiC) estimator of Athey and Imbens (2006) performs comparatively less well in the discrete designs, particularly in DGP0 and DGP1, where rejection frequencies are often noticeably larger and the $\mathcal{L}_2$ error does not improve relative to the proposed procedure. In the continuous design, by contrast, CiC is more competitive, as might be expected given its natural suitability for continuous outcomes. Overall, the results suggest that the proposed distributional DiD estimator is robust across a range of designs, while the identity and normal links provide the most stable finite-sample performance.

S.4.2 Over-identifying Restrictions Testing

The data-generating process for the over-identifying restrictions test is calibrated directly to the empirical application and is therefore conditional on the observed data. Let $\widehat{\mathrm{F}}_{00}$ and $\widehat{\mathrm{F}}_{01}$ denote the empirical CDFs of the outcome for blocks without police presence in the pre-treatment and post-treatment periods, respectively, where the pre-treatment period pools months April through June and the post-treatment period pools months August through December. Likewise, let $\widehat{\mathrm{F}}_{10}$ and $\widehat{\mathrm{F}}_{11}$ denote the corresponding empirical CDFs for blocks with increased police presence. The support used in the simulations is the empirical support of the outcome, truncated above at 1.0, and the associated quantile functions are obtained from these empirical CDFs.

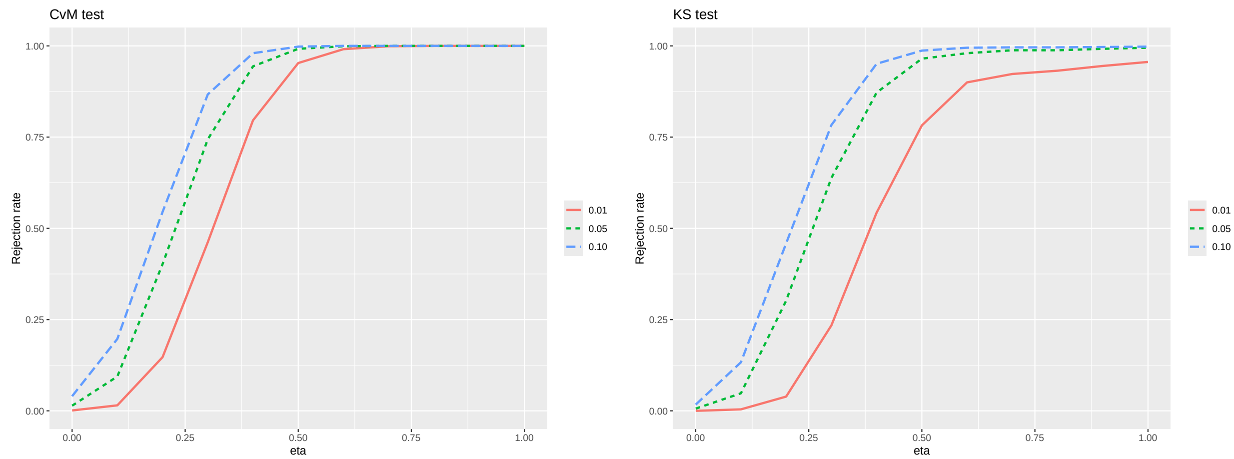
Repeated cross-sections are then simulated for three groups, $G \in \{1, 2, 3\}$, over one pre-treatment period, $t = 0$, and one post-treatment period, $t = 1$. The treated group, $G = 1$, is generated from $(\widehat{\mathrm{F}}_{10}, \widehat{\mathrm{F}}_{11})$. The first control group, $G = 2$, is generated from $(\widehat{\mathrm{F}}_{00}, \widehat{\mathrm{F}}_{01})$ and therefore satisfies the null of valid over-identifying restrictions. The second control group, $G = 3$, is used to introduce a controlled violation of the null. Its pre-treatment distribution is set equal to that of the valid control group, namely $\widehat{\mathrm{F}}_{30} = \widehat{\mathrm{F}}_{00}$, whereas its post-treatment distribution is defined on the empirical grid $\{y_\ell\}_{\ell=1}^{L}$ by

$$\widehat{\mathrm{F}}_{31}^{(\eta)}(y_\ell) = (1-\eta)\widehat{\mathrm{F}}_{01}(y_\ell) + \eta\frac{\ell}{L}, \qquad \eta \in [0, 1].$$

Hence, when $\eta = 0$, one has $\widehat{\mathrm{F}}_{31}^{(0)} = \widehat{\mathrm{F}}_{01}$, so that both control groups satisfy the same untreated transition and the null holds exactly. As η increases, the post-treatment distribution of group $G = 3$ is moved progressively away from the empirical benchmark $\widehat{\mathrm{F}}_{01}$, thereby generating increasingly pronounced violations of the functional over-identifying restrictions.

Figure S.1 shows that both the CvM and KS tests have empirical rejection probabilities close to the nominal level when $\eta = 0$, which is consistent with satisfactory size control under the null. As η increases, the rejection probability rises for both tests, indicating non-trivial power against departures from the over-identifying restrictions. The increase is gradual for small values of η and becomes more pronounced for larger deviations from the null. Overall, both statistics display the expected monotonic power pattern.

Figure S.1: Power Curves of the functional over-identifying restrictions tests



(a) (b)

Notes: Both the CvM and KS tests are conducted at three nominal levels: 0.01, 0.05, and 0.10. 999 empirical bootstrap replications are used for each of the 1000 Monte Carlo replications. The reported figure uses the standard normal link function for both tests.

References

[1] Athey, Susan and Guido Imbens. "Identification and inference in nonlinear difference-in-differences models". *Econometrica* 74.2 (2006), pp. 431–497.

[2] Chernozhukov, Victor, Ivan Fernandez-Val, and Alfred Galichon. "Quantile and probability curves without crossing". *Econometrica* 78.3 (2010), pp. 1093–1125.

[3] Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly. "Inference on counterfactual distributions". *Econometrica* 81.6 (2013), pp. 2205–2268.

[4] Chernozhukov, Victor, Ivan Fernandez-Val, Blaise Melly, and Kaspar Wüthrich. "Generic inference on quantile and quantile effect functions for discrete outcomes". *Journal of the American Statistical Association* (2019).

[5] Fernández-Val, Iván, Jonas Meier, Aico van Vuuren, and Francis Vella. "Distribution Regression Difference-in-Differences". *arXiv preprint arXiv:2409.02311* (2024).

[6] Foresi, Silverio and Franco Peracchi. "The conditional distribution of excess returns: An empirical analysis". *Journal of the American Statistical Association* 90.430 (1995), pp. 451–466.

[7] Kim, Doosoo and Jeffrey M Wooldridge. "Difference-in-differences Estimator of Quantile Treatment Effect on the Treated". *Journal of Business & Economic Statistics* 43.2 (2025), pp. 401–412.

[8] Kosorok, Michael R. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Science & Business Media, 2007.

[9] Manski, Charles F. "Identification of binary response models". *Journal of the American statistical Association* 83.403 (1988), pp. 729–738.

[10] Sant'Anna, Pedro HC and Xiaojun Song. "Specification tests for the propensity score". *Journal of Econometrics* 210.2 (2019), pp. 379–404.

[11] van der Vaart, Aad. *Asymptotic Statistics*. Vol. 3. Cambridge University Press, 2000.

[12] van der Vaart, Aad W and Jon A Wellner. *Weak Convergence and Empirical Processes*. Springer Science & Business Media, 1996.

[13] Wooldridge, Jeffrey M. “Simple approaches to nonlinear difference-in-differences with panel data”. *The Econometrics Journal* 26.3 (2023), pp. C31–C66.