Efficient Difference-in-Differences and Event Study Estimators: Supplemental Appendix

Xiaohong Chen[*] Pedro H. C. Sant'Anna[†] Haitian Xie[‡]

June 21, 2025

Appendix A introduces Hausman-type tests for the overidentifying restrictions. Appendix B extends the semiparametric efficiency results to an instrumented DiD setting. Appendix C presents the proofs of the theoretical results.

A Assessing the plausibility of PT assumptions

Another consequence of Lemmas 3.1 and 3.2 is that Assumption PT-All can be *directly* tested, as our DiD model is overidentified. In this section, we describe how to construct a Hausman-type test based on event-study parameters in the context of staggered treatment adoption. We focus on event-study parameters as they are often the main parameter of interest in empirical research, and are also often used to assess the plausibility of the identification assumptions; see, e.g., Roth (2022) and Borusyak, Jaravel and Spiess (2024) for a discussion.

The main idea of our Hausman-type test is to compare $\widehat{ES} = (\widehat{ES}(e), e \in \mathcal{E})$ as defined in (4.5)—which is consistent and semiparametrically efficient under Assumption PT-All—with an event-study estimator that is consistent under Assumption PT-Post but does not require Assumption PT-All. Toward this end, let $\widetilde{ES} = (\widetilde{ES}(e), e \in \mathcal{E})$, where each event-study $\widetilde{ES}(e)$ is given by

$$\widetilde{ES}(e) = \sum_{g \in \mathcal{G}_{\text{trt}}} \frac{\widehat{\pi}_g}{\sum_{g' \in \mathcal{G}_{\text{trt}}} \widehat{\pi}_{g'}} \widetilde{ATT}_{\text{stg}}(g, g+e),$$

with

$$\widetilde{ATT}_{\text{stg}}(g,t) = \mathbb{E}_n \left[\widehat{\widehat{Y}}^{\text{att(g,t)}}_{g,g-1} \right], \tag{A.1}$$

and $\widehat{\widehat{Y}}^{\text{att(g,t)}}_{g',t_{\text{pre}}}$ as in (4.4). Under this construction, $\widetilde{ES}$ is consistent for ES under Assumption PT-Post but less efficient than $\widehat{ES} = (\widehat{ES}(e), e \in \mathcal{E})$.

Based on these two estimators, we can construct a Hausman-type test statistic as

$$\widehat{H} = n \left(\widehat{ES} - \widetilde{ES} \right)' \left(\widehat{\text{aCov}}(\widetilde{ES}) - \widehat{\text{aCov}}(\widehat{ES}) \right)^{-1} \left(\widehat{ES} - \widetilde{ES} \right), \tag{A.2}$$

[*]Yale University. E-mail: xiaohong.chen@yale.edu.
[†]Emory University. E-mail: pedro.santanna@emory.edu
[‡]Peking University. E-mail: xht@gsm.pku.edu.cn

where $\widehat{\text{aCov}}(\cdot)$ denotes the corresponding asymptotic covariance estimator for the asymptotic covariance matrices of $\widehat{ES}$ and $\widetilde{ES}$ given by Theorem 3.2 and Corollary 3.1.[1] The test rejects the parallel trends assumption for all periods and all groups if $\widehat{H}$ exceeds the corresponding critical value of a $\chi^2(|\mathcal{E}|)$ distribution.[2]

Theorem A.1. *Suppose that the estimator $\widehat{ES}$ is constructed under the conditions of Theorem 4.1, and $\widetilde{ES}$ also satisfies the corresponding conditions. Assume that the covariance matrices estimators $\widehat{\text{aCov}}(\widetilde{ES})$ and $\widehat{\text{aCov}}(\widehat{ES})$ are consistent. Then the test statistic $\widehat{H}$ converges in distribution to a $\chi^2(|\mathcal{E}|)$ distributions, where $|\mathcal{E}|$ denotes the number of elements in $\mathcal{E}$. Also, this Hausman test has nontrivial power against all local alternatives.*

Although assessing the plausibility of Assumption PT-All via the Hausman-type test in Theorem A.1 is attractive, it is also worth noting that Assumptions PT-All and PT-Post respectively define the largest and smallest sets of conditional moment restrictions for estimating treatment effects. Thus, in practice, it may be the case that the "plausible" parallel trends condition lies between these two extremes, and one may be interested in approximating this set of conditional moment restrictions. It is possible to do so by coupling our Hausman-type test with a Holm-Bonferroni (Holm, 1979) sequential procedure to select conditional moment restrictions that align with Assumption PT-Post. The main idea is to contrast $\widetilde{ES}$ with event-study estimators that (sequentially) include additional overidentifying restrictions (3.7), and select the event-study estimator with the largest set of restrictions that is not statistically different from $\widetilde{ES}$ (after adjusting for multiple testing). This essentially entails an incremental Sargan Test, as discussed in Chen and Santos (2018) in different contexts.

The specific procedure is described as follows. Let $\mathcal{M}$ denote the set of conditional moment restrictions specified by (2.7) for all $t \geqslant g$, with $g' = g$ and $t_{\text{pre}} = g - 1$, corresponding to the restrictions implied by Assumption PT-Post. For each $g' > t_{\text{pre}}$ (with t'_{pre} fixed at 1), let $\mathcal{M}_{g',t_{\text{pre}}}$ represent $\mathcal{M}$ extended by the additional conditional moment restriction given by (2.7) for g' and t_{pre}. Let L denote the total number of such models $\mathcal{M}_{g',t_{\text{pre}}}$. For each $\mathcal{M}_{g',t_{\text{pre}}}$, we perform a Hausman-type test against the baseline model $\mathcal{M}$ similar to the approach in (A.2) and Theorem A.1. Specifically, we construct event-study estimators $\widehat{ES}$ using the moment restrictions in $\mathcal{M}_{g',t_{\text{pre}}}$ and compute the corresponding p-value, $\text{p}_{g',t_{\text{pre}}}$, of the Hausman statistic in (A.2). These p-values are then ordered from smallest to largest as $\text{p}_{(1)}, \ldots, \text{p}_{(L)}$. Let α denote the family-wise error rate (e.g., $\alpha = 0.05$). The procedure starts with $\ell = 1$ and compares $\text{p}_{(1)}$ with $\frac{\alpha}{L}$. If $\text{p}_{(1)} \geqslant \frac{\alpha}{L}$, the procedure terminates without rejecting any models. Otherwise, the conditional moment corresponding to $\text{p}_{(1)}$ is rejected, and the procedure proceeds to the next step. For each subsequent $\ell = 2, \ldots, L$, $\text{p}_{(\ell)}$ is compared with $\frac{\alpha}{L+1-\ell}$. If $\text{p}_{(\ell)} \geqslant \frac{\alpha}{L+1-\ell}$, the procedure terminates; otherwise, the conditional moment corresponding to $\text{p}_{(\ell)}$ is rejected, and the process continues until either a p-value falls below the threshold or all models are tested.

[1] It is also straightforward to construct an alternative estimator to $\widehat{\text{aCov}}(\widetilde{ES}) - \widehat{\text{aCov}}(\widehat{ES})$ that leverages the difference of the (efficient) influence function of these two estimators. An advantage of this alternative estimator is that it is positive definite in finite samples. We omit the details as the notation is heavier.

[2] Note that our Hausman-type test is effectively testing if the event-study aggregation of the parallel trends is the same under Assumptions PT-All and PT-Post. It is straightforward to construct a Hausman-type test for Assumptions PT-All and PT-Post based on all the estimators of the $ATT(g,t)$'s. However, we anticipate that this would not be as empirically attractive as the one in (A.2), as event-study parameters often play a more prominent role than the $ATT(g,t)$'s in setups with staggered treatment adoption.

Remark A.1. We caveat that using our Hausman-type test $\widehat{H}$ in Theorem A.1 to decide whether to report $\widehat{ES}$ or $\widetilde{ES}$ can lead to unnecessarily high mean squared errors compared to an oracle selection procedure (Armstrong, Kline and Sun, 2024). Instead of following this approach, one can adopt the adaptive estimation procedure of Armstrong et al. (2024) that can be understood as weighting $\widehat{ES}$ and $\widetilde{ES}$, with weights that are designed to take into account a bias-variance tradeoff. In their context, $\widehat{ES}$ would play the role of the "restricted" estimator while $\widetilde{ES}$ would play the role of the "unrestricted" estimator.

Remark A.2. Another popular technique commonly used to assess the plausibility of parallel trends assumptions is to examine whether event-study coefficients in pre-treatment periods are close to zero, i.e., whether there is some evidence of parallel pre-treatment trends. Although our proposed efficient event-study estimators leverage all the available pre-treatment information to estimate post-treatment average treatment effects, we can also construct "placebo" pre-treatment effects by fixing the estimation procedure in Section 4 but allowing for $t < g$. This shares the same spirit as the pre-treatment event study analysis of Borusyak et al. (2024). Alternatively, one can fix any comparison group and pre-treatment period, and report a pre-treatment event study analysis for these.

B Extension: Instrumented DiD

We extend the instrumented DiD (DiD-IV) setup in Miyaji (2024) to incorporate covariates. There are T time periods: $t = 1, 2, \cdots, T$. Let Z_t denote the instrument in time period t and collect them into the path $Z := (Z_1, \cdots, Z_T)$. The instrument is irreversible: $Z_t \geqslant Z_{t-1}$ for all t. Therefore, the instrument path is uniquely characterized by the initial date of exposure $G^{\mathrm{IV}} := \min\{t : Z_t = 1\}$. The units are grouped based on G^{IV} instead of on the actual treatment. Denote $G_g^{\mathrm{IV}} := \mathbf{1}\{G^{\mathrm{IV}} = g\}$. Let $\mathcal{G}_{\mathrm{trt}}^{\mathrm{IV}}$ be the support of G^{IV} among the units who are eventually exposed to the instrument. Denote $D_t(g)$ as the potential treatment if the unit is first exposed to the instrument in period g. Denote $Y_t(d_t)$ as the potential outcome if the treatment in period t is d_t. This definition already imposes the no carryover assumption that the potential outcomes depend only on the current treatment status (de Chaisemartin and D'Haultfœuille, 2020; Miyaji, 2024) and the exclusion of the instrument. Let X be a set of pretreatment covariates.

The target parameter is the local average treatment effect for the treated (LATT) defined as

$$LATT(g,t) := \mathbb{E}[Y_t(1) - Y_t(0) | G^{\mathrm{IV}} = g, D_t(g) > D_t(\infty)].$$

The following assumptions are imposed.

Assumption DiD-IV.

(1) (Random Sampling) $\{(Y_{i,t=1}, \ldots, Y_{i,t=T}, D_{i,t=1}, \ldots, D_{i,t=T}, X_i', G_i^{\mathrm{IV}})'\}_{i=1}^n$ is a random sample from $(Y_{t=1}, \ldots, Y_{t=T}, D$

(2) (Overlap) For each g, $\mathbb{E}[G_g^{\mathrm{IV}} | X] \in (0,1)$ a.s.

(3) (Monotonicity) $\mathbb{P}(D_t(g) \geqslant D_t(\infty) | X) = 1$ a.s., for $t \geqslant g$.

(4) (No-anticipation in the first stage) $\mathbb{E}[D_t(g) | G^{\mathrm{IV}} = g, X] = \mathbb{E}[D_t(\infty) | G^{\mathrm{IV}} = g, X], t < g$.

(5) (Parallel trends in the treatment) $\mathbb{E}[D_t(\infty) - D_{t-1}(\infty) | G^{\mathrm{IV}} = g, X] = \mathbb{E}[D_t(\infty) - D_{t-1}(\infty) | G^{\mathrm{IV}} = \infty, X]$, for all g, t.

(6) (Parallel trends in the outcome) $\mathbb{E}[Y_t(D_t(\infty)) - Y_{t-1}(D_{t-1}(\infty))|G^{\mathrm{IV}} = g, X] = \mathbb{E}[Y_t(D_t(\infty)) - Y_{t-1}(D_{t-1}(\infty))|G^{\mathrm{IV}} = \infty, X]$, for all g, t.

Lemma B.1. *Under Assumption DiD-IV, $LATT(g,t)$ is identified as*

$$LATT(g,t) = \frac{\mathbb{E}\big[G_g^{IV}(\mathbb{E}[Y_t - Y_{g-1}|G^{IV} = g, X] - \mathbb{E}[Y_t - Y_{g-1}|G^{IV} = \infty, X])\big]}{\mathbb{E}\big[G_g^{IV}(\mathbb{E}[D_t - D_{g-1}|G^{IV} = g, X] - \mathbb{E}[D_t - D_{g-1}|G^{IV} = \infty, X])\big]}.$$

The lemma shows that the LATT parameter is a ratio between two ATT-type parameters. The following moment restrictions define our DiD-IV model under Assumption DiD-IV. For simplicity in exposition, we consider the case with a single date of exposure to the instrument g. The more general staggered case can be derived similarly but with a more complicated expression of the efficient influence function.

Lemma B.2 (Moment-restrictions for overidentified DiD-IV with a single instrument exposure time)**.** *The family of probability distributions of $(Y_{t=1}, ...Y_{t=T}, D_{t=1}, ...D_{t=T}, X', G^{IV})$ satisfying Assumption DiD-IV are observationally equivalent to the family of probability distributions of $(Y_{t=1}, ...Y_{t=T}, D_{t=1}, ...D_{t=T}, X', G^{IV})$ satisfying Assumption DiD-IV(1)-(3) and the following set of moment restrictions: for all post-treatment periods $t \in \{g, \dots, T\}$, with probability one,*

$$\mathbb{E}[G_g^{IV}(LATT(g,t)_{num} - h_1(g,t,X))] = 0,$$
$$\mathbb{E}[G_g^{IV}(LATT(g,t)_{den} - h_2(g,t,X))] = 0,$$
$$\mathbb{E}\left[h_1(g,t,X) - \frac{G_g^{IV}(Y_t - Y_{g-1})}{p_g^{IV}(X)} + \frac{G_\infty^{IV}(Y_t - Y_{g-1})}{p_\infty^{IV}(X)}\Big|X\right] = 0,$$
$$\mathbb{E}\left[h_2(g,t,X) - \frac{G_g^{IV}(D_t - D_{g-1})}{p_g^{IV}(X)} + \frac{G_\infty^{IV}(D_t - D_{g-1})}{p_\infty^{IV}(X)}\Big|X\right] = 0,$$
$$\mathbb{E}\left[\frac{G_g^{IV}(Y_{t_{pre}} - Y_1)}{p_g^{IV}(X)} - \frac{G_\infty^{IV}(Y_{t_{pre}} - Y_1)}{p_\infty^{IV}(X)}\Big|X\right] = 0, \textit{for all } 2 \leqslant t_{pre} \leqslant g-1,$$
$$\mathbb{E}\left[\frac{G_g^{IV}(D_{t_{pre}} - D_1)}{p_g^{IV}(X)} - \frac{G_\infty^{IV}(D_{t_{pre}} - D_1)}{p_\infty^{IV}(X)}\Big|X\right] = 0, \textit{for all } 2 \leqslant t_{pre} \leqslant g-1,$$
$$\mathbb{E}[G_g^{IV} - p_g^{IV}(X)|X] = 0.$$

Using the moment restrictions, the LATT parameter can be written as $LATT(g,t) = \frac{LATT(g,t)_{num}}{LATT(g,t)_{den}}$. The nuisance parameters h_1 and h_2 correspond to the numerator and denominator of the conditional LATT parameter, respectively, while $LATT(g,t)_{num}$ and $LATT(g,t)_{den}$ represent their unconditional counterparts. Compared to the moment restrictions in Lemma 3.1, the ones in Lemma B.2 incorporate the instrument as the treatment and the treatment as the outcome in the restrictions for $LATT(g,t)_{den}$. Nonetheless, this minor distinction does not affect the efficiency calculations. To formally define the efficient influence function, let

$$\mathbb{IF}_{t_{\mathrm{pre}}}^{latt(g,t),num,Y} = \Big(G_g(m_{g,t,t_{\mathrm{pre}}}(X) - m_{\infty,t,t_{\mathrm{pre}}}(X) - LATT(g,t)_{num})\Big)$$
$$+ p_g(X)\Big(\frac{G_g}{p_g(X)}(Y_t - Y_1 - m_{g,t,t_{\mathrm{pre}}}(X))$$

$$- \frac{G_\infty}{p_\infty(X)}(Y_t - Y_1 - m_{\infty,t,t_{\text{pre}}}(X))\Big), 1 \leqslant t_{\text{pre}} \leqslant g-1,$$

and

$$\begin{aligned}\mathbb{IF}^{latt(g,t),num,D}_{t_{\text{pre}}} &= \Big(G_g(\mu_{g,t,t_{\text{pre}}}(X) - \mu_{\infty,t,t_{\text{pre}}}(X) - LATT(g,t)_{num})\Big) \\ &+ p_g(X)\Big(\frac{G_g}{p_g(X)}(Y_t - Y_1 + D_{t_{\text{pre}}} - D_1 - \mu_{g,t,t_{\text{pre}}}(X)) \\ &- \frac{G_\infty}{p_\infty(X)}(Y_t - Y_1 + D_{t_{\text{pre}}} - D_1 - \mu_{\infty,t,t_{\text{pre}}}(X))\Big), 2 \leqslant t_{\text{pre}} \leqslant g-1,\end{aligned}$$

where $\mu_{g,t,t_{\text{pre}}}(X) := \mathbb{E}[Y_t - Y_1 + D_{t_{\text{pre}}} - D_1 | G^{\text{IV}} = g, X]$. Denote the column vector that stacks these $2(g-1)-1$ functions by

$$\mathbb{IF}^{latt(g,t),num} := (\mathbb{IF}^{latt(g,t),num,Y}_1, \cdots, \mathbb{IF}^{latt(g,t),num,Y}_{g-1}, \mathbb{IF}^{latt(g,t),num,D}_2, \cdots, \mathbb{IF}^{latt(g,t),num,D}_{g-1})',$$

and the conditional covariance matrix of $\mathbb{IF}^{latt(g,t),num}$ as $V^{latt(g,t),num}(X)$. Similarly, let

$$\begin{aligned}\mathbb{IF}^{latt(g,t),den,D}_{t_{\text{pre}}} &= \Big(G_g(\tilde{m}_{g,t,t_{\text{pre}}}(X) - \tilde{m}_{\infty,t,t_{\text{pre}}}(X) - LATT(g,t)_{den})\Big) \\ &+ p_g(X)\Big(\frac{G_g}{p_g(X)}(D_t - D_1 - \tilde{m}_{g,t,t_{\text{pre}}}(X)) \\ &- \frac{G_\infty}{p_\infty(X)}(D_t - D_1 - \tilde{m}_{\infty,t,t_{\text{pre}}}(X))\Big), 1 \leqslant t_{\text{pre}} \leqslant g-1,\end{aligned}$$

and

$$\begin{aligned}\mathbb{IF}^{latt(g,t),den,Y}_{t_{\text{pre}}} &= \Big(G_g(\tilde{\mu}_{g,t,t_{\text{pre}}}(X) - \tilde{\mu}_{\infty,t,t_{\text{pre}}}(X) - LATT(g,t)_{num})\Big) \\ &+ p_g(X)\Big(\frac{G_g}{p_g(X)}(D_t - D_1 + Y_{t_{\text{pre}}} - Y_1 - \tilde{\mu}_{g,t,t_{\text{pre}}}(X)) \\ &- \frac{G_\infty}{p_\infty(X)}(D_t - D_1 + Y_{t_{\text{pre}}} - Y_1 - \tilde{\mu}_{\infty,t,t_{\text{pre}}}(X))\Big), 2 \leqslant t_{\text{pre}} \leqslant g-1,\end{aligned}$$

where $\tilde{m}_{g,t,t_{\text{pre}}}(X) := \mathbb{E}[D_t - D_1 | G^{\text{IV}} = g, X]$, and $\tilde{\mu}_{g,t,t_{\text{pre}}}(X) := \mathbb{E}[D_t - D_1 + Y_{t_{\text{pre}}} - Y_1 | G^{\text{IV}} = g, X]$. The vector $\mathbb{IF}^{latt(g,t),den}$ of influence functions and its conditional covariance matrix $V^{latt(g,t),den}(X)$ are defined analogously.

Compared to the previous efficiency results, the difference in this IV setting is the presence of two parallel trends – one in the treatment and one in the outcome. To fully utilize all available information, when estimating the numerator of the LATT, we must account for the overidentifying information from both the parallel trend in the outcome and treatment, and vice versa for the denominator. Consequently, the estimation of either the numerator or the denominator involves twice as many influence functions due to the existence of two parallel trends. As before, the efficient influence function is then obtained by optimally weighting these functions. This result is summarized in the following corollary.

Corollary B.1. *Under Assumption DiD-IV, the efficient influence function for $LATT(g,t)$, $t \geqslant g$, is given* *by*

$$\frac{\mathbb{EIF}^{latt(g,t),num} - LATT(g,t)\mathbb{EIF}^{latt(g,t),den}}{\mathbb{E}\big[G_g^{IV}(\mathbb{E}[D_t - D_{g-1}|G^{IV} = g, X] - \mathbb{E}[D_t - D_{g-1}|G^{IV} = \infty, X])\big]},$$

where

$$\mathbb{EIF}^{latt(g,t),j} = \frac{\mathbf{1}'V^{latt(g,t),j}(X)^{-1}}{\mathbf{1}'V^{latt(g,t),j}(X)^{-1}\mathbf{1}}\mathbb{IF}^{latt(g,t),j}, j = num, den.$$

Assuming the second moment of the efficient influence function is finite, the semiparametric efficiency bound for $LATT(g,t)$ is its second moment.

The efficiency results for staggered exposure to the instrument can be derived analogously to those in Section 3.2, with the incorporation of additional overidentifying restrictions from multiple comparison groups. Building on the derived efficient influence function, semiparametrically efficient estimators for the LATT parameters can be constructed following the approach outlined in Section 4.

C Proofs for theoretical results

Notation: For simplicity, we use t' and t'' to denote pre-treatment periods instead of t_{pre} and t'_{pre}. We write p_{ratio} simply as p. We denote the "generated outcome" $\tilde{Y}^{att(g,t)}_{g',t_{\text{pre}}}$ as $\theta_{g',t''}(W;p,m)$, making explicit its dependence on the nuisance parameters p and m.

Proof of Lemmas 3.1 and 3.2. We prove only the staggered case for Lemma 3.2 as Lemma 3.1 is a more straightforward case. It is easy to see that the moment restrictions are implied by the identification assumptions. Therefore, we only prove the converse implication. It suffices to construct a joint distribution of the potential outcomes that is consistent with the observed outcome and satisfies the identification assumptions. Without loss of generality, we can suppress the covariates, since the analysis can be conducted conditional on each value of the covariates. Denote $Y \equiv (Y_{t=1}, \cdots, Y_{t=T})$ as the vector of observed outcomes and $\Delta Y \equiv (\Delta Y_2, \cdots, \Delta Y_T)$ as the vector of differenced outcomes, where $\Delta Y_t \equiv Y_t - Y_{t-1}$. The notations $Y(g)$, $\Delta Y_t(g)$, and $\Delta Y(g)$ are defined analogously for potential outcomes. Let (Y, G) denote the observed variables that already satisfy random sampling and overlap. Once we have the observed distribution of G, we can define the parameters π_g and p_g accordingly. For each given $g \in \mathcal{G}$, we construct the potential outcomes as follows:

the vectors $Y(g'), g' \in \mathcal{G}$ are jointly independent conditional on $G = g$,

for $g' = g$: $Y(g)|\{G = g\} \overset{d}{=} Y|\{G = g\}$, and

for $g' \neq g$: $Y_1(g')|\{G = g\} \overset{d}{=} Y_1(g)|\{G = g\}, \Delta Y(g')|\{G = g\} \overset{d}{=} \Delta Y|\{G = \infty\}, Y_1(g') \perp \Delta Y(g')|\{G = g\}$,

where $\overset{d}{=}$ denotes equal in distribution. Notice that this construction already ensures that the potential outcome induces the observed outcome because $Y(g)|\{G = g\} \overset{d}{=} Y|\{G = g\}$. The parallel trends condition also holds because $\Delta Y(\infty)|\{G = g\} \overset{d}{=} \Delta Y|\{G = \infty\}$. For the no-anticipation assumption, we have for any

$2 \leqslant t < g$,

$$\begin{aligned}\mathbb{E}[\Delta Y_t(g)|G=g] &= \mathbb{E}[\Delta Y_t|G=g] \\ &= \mathbb{E}[\Delta Y_t|G=\infty] \text{ (by the moment condition } \mathbb{E}[Y_t - Y_{t-1}|G=\infty] = \mathbb{E}[Y_t - Y_{t-1}|G=g]) \\ &= \mathbb{E}[\Delta Y_t(\infty)|G=g] \text{ (by } \Delta Y(\infty)|\{G=g\} \overset{d}{=} \Delta Y|\{G=\infty\}).\end{aligned}$$

Combining the above equality with the fact that $\mathbb{E}[Y_1(\infty)|G=g] = \mathbb{E}[Y_1(g)|G=g]$ (by construction $Y_1(\infty)|\{G=g\} \overset{d}{=} Y_1(g)|\{G=g\}$), we obtain the no-anticipation condition. This completes the proof. □

Proof of Lemma 2.1. Since both t_{pre} and t'_{pre} are pre-treatment for group g', by Assumption NA, the right-hand side of (2.7) is equal to

$$\begin{aligned}&\mathbb{E}[Y_t(g) - Y_{t'_{\text{pre}}}(\infty)|G=g,X] - (\mathbb{E}[Y_t(\infty) - Y_{t_{\text{pre}}}(\infty)|G=\infty,X] + \mathbb{E}[Y_{t_{\text{pre}}}(\infty) - Y_{t'_{\text{pre}}}(\infty)|G=g',X]) \\ =&\mathbb{E}[Y_t(g) - Y_{t'_{\text{pre}}}(\infty)|G=g,X] - \mathbb{E}[Y_t(\infty) - Y_{t'_{\text{pre}}}(\infty)|G=\infty,X] \\ =&CATT(g,t,X),\end{aligned}$$

where the first equality follows from Assumption PT-All, and the second inequality follows from standard DiD calculation. The identification of ATT given CATT follows from taking the expectation conditional on $G = g$. □

Proof of Theorem 3.1. We divide the proof into two parts. In the first part, we focus on a submodel for $ATT(g,t)$ at a single time period t:

$$\begin{aligned}\mathbb{E}[G_g(ATT(g,t) - CATT(g,t,X))] &= 0, \\ \mathbb{E}\left[CATT(g,t,X) - \frac{G_g(Y_t - Y_1)}{p_g(X)} + \frac{G_\infty(Y_t - Y_1)}{p_\infty(X)}\Big|X\right] &= 0, \\ \mathbb{E}\left[\frac{G_g(Y_{t'} - Y_1)}{p_g(X)} - \frac{G_\infty(Y_{t'} - Y_1)}{p_\infty(X)}\Big|X\right] &= 0, 2 \leqslant t' \leqslant g-1, \\ \mathbb{E}[G_g - p_g(X)|X] &= 0,\end{aligned} \tag{C.1}$$

which is an equivalent representation of the restrictions in Lemma 3.1, obtained by replacing Y_{g-1} in the second line with Y_1 for convenience in the proof. Then in the second part of the proof, we show that the EIFs remain the same for the entire set of moment restrictions in Lemma 3.1.

Part 1 We follow the orthogonalization method in Ai and Chen (2012) to derive the efficient influence function and the semiparametric efficiency bound. We adopt the notations in that paper. Denote the available random variables as $W \equiv (Y_{t=1}, \cdots, Y_{t=T}, X', G)'$ and the parameters $\alpha \equiv (\theta, h)$, where $\theta \equiv ATT(g,t)$ is the finite dimensional parameter, and $h \equiv (CATT(g,t,\cdot), p_g)$ contains the first-stage nuisance parameters. Below, we slightly modify the moment conditions to make the derivation easier while preserving the model and the efficiency bound. Let the unconditional and conditional moments be denoted by ρ_1 and ρ_2, respectively:

$$\rho_1(W,\alpha) \equiv p_g(X)ATT(g,t) - G_gCATT(g,t,X),$$

$$\rho_2(W,h) \equiv \begin{pmatrix} CATT(g,t,X) - \frac{G_g(Y_t-Y_1)}{p_g(X)} + \frac{G_\infty(Y_t-Y_1)}{p_\infty(X)} \\ \cdots \\ CATT(g,t,X) - \frac{G_g(Y_t-Y_{g-1})}{p_g(X)} + \frac{G_\infty(Y_t-Y_{g-1})}{p_\infty(X)} \\ G_g - p_g(X) \\ G_\infty - p_\infty(X) \end{pmatrix}.$$

In the unconditional moment, we replace G_g by $p_g(X)$, which makes ρ_1 and ρ_2 orthogonal. This is the orthogonalized unconditional moment one would obtain following the procedure in Ai and Chen (2012). To obtain ρ_2, we simply rotate the original conditional moment restrictions by using the following invertible matrix

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & & & & & & \\ 1 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}.$$

The subsequent analysis shows that the semiparametric efficiency bound remains invariant to such rotations. Whenever necessary, we will use α^*, θ^*, and h^* to denote the true values for the respective parameters, and α, θ, and h to denote generic values. Let

$$\begin{aligned} \Sigma_1 &\equiv \mathbb{E}[\rho_1(W,\alpha^*)\rho_1(W,\alpha^*)'] = \mathbb{E}[p_g(X)^2(CATT(g,t,X) - ATT(g,t))^2], \\ \Sigma_2(X) &\equiv \mathbb{E}[\rho_2(W,h^*)\rho_1(W,h^*)'|X], \\ m_1(\alpha) &\equiv \mathbb{E}[\rho_1(W,\alpha)], \\ m_2(X,\alpha) &\equiv \mathbb{E}[\rho_2(W,h)|X]. \end{aligned}$$

The derivatives of m_1 and m_2 with respect to θ are

$$\begin{aligned} \frac{dm_1(\alpha^*)}{dATT(g,t)} &= \mathbb{E}\left[\frac{d\rho_1(W,\alpha^*)}{dATT(g,t)}\right] = \mathbb{E}[p_g(X)] = \pi_g, \\ \frac{dm_2(X,\alpha^*)}{dATT(g,t)} &= \mathbb{E}\left[\frac{d\rho_2(W,h^*)}{dATT(g,t)}\right] = 0. \end{aligned}$$

Let $h_\tau = h^* + \tau r$ be a smooth path in $\tau \in [0,1]$, where $r \equiv (r_1, r_2, r_3)'$, and $h^* + r$ lies in the nuisance parameter space. The derivative of m_1 with respect to h (in the direction of r) is

$$\frac{dm_1(\alpha^*)}{dh}[r] = \mathbb{E}\left[\frac{d\rho_1(W,\alpha^*)}{dh}[r]\right] = \mathbb{E}[(-p_g(X), ATT(g,t) - CATT(g,t,X), 0)r(X)|X] \equiv \mathbb{E}[A_t(X)r(X)],$$

where $A_t(X) \equiv (-p_g(X), ATT(g,t) - CATT(g,t,X), 0)$. Similarly, the derivative of m_2 with respect to h (in the direction of r) is

$$\frac{dm_2(X,\alpha^*)}{dh}[r] = \begin{pmatrix} r_1(X) + \frac{\mathbb{E}[G_g(Y_t - Y_1)|X]}{p_g(X)^2} r_2(X) - \frac{\mathbb{E}[G_\infty(Y_t - Y_1)|X]}{p_\infty(X)^2} r_3(X) \\ \cdots \\ r_1(X) + \frac{\mathbb{E}[G_g(Y_t - Y_{g-1})|X]}{p_g(X)^2} r_2(X) - \frac{\mathbb{E}[G_\infty(Y_t - Y_{g-1})|X]}{p_\infty(X)^2} r_3(X) \\ -r_2(X) \\ -r_3(X) \end{pmatrix} \equiv L(X)r(X),$$

where the matrix $L(X)$ is defined as

$$L(X) \equiv \begin{pmatrix} 1 & \frac{\mathbb{E}[Y_t - Y_1|G=g,X]}{p_g^*(X)} & -\frac{\mathbb{E}[Y_t - Y_1|G=\infty,X]}{p_\infty^*(X)} \\ \cdots & & \\ 1 & \frac{\mathbb{E}[Y_t - Y_{g-1}|G=g,X]}{p_g^*(X)} & -\frac{\mathbb{E}[Y_t - Y_{g-1}|G=\infty,X]}{p_\infty^*(X)} \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

By Theorem 2.1 of Ai and Chen (2012), the semiparametric efficiency bound for $ATT(g,t)$ is obtained by solving the following minimization problem:

$$\inf_r \quad (\pi_g - \mathbb{E}[A_t(X)r(X)])\Sigma_1^{-1}(\pi_g - \mathbb{E}[A_t(X)r(X)]) + \mathbb{E}[(L(X)r(X))'\Sigma_2(X)^{-1}L(X)r(X)].$$

By the calculus of variations, the optimizer r^* satisfies the following first-order condition:

$$(\pi_g - \mathbb{E}[A_t(X)r^*(X)])\Sigma_1^{-1}\mathbb{E}[A_t(X)r(X)] - \mathbb{E}[(L(X)r^*(X))'\Sigma_2(X)^{-1}L(X)r(X)] = 0,$$

for all r. Denote $B(X) \equiv L(X)'\Sigma_2(X)^{-1}L(X)$. The left-hand side of the above first-order condition can be written as

$$\mathbb{E}\big[\big((\pi_g - \mathbb{E}[A_t(X)r^*(X)])\Sigma_1^{-1}A_t(X) - r^*(X)'B(X)\big)r(X)\big].$$

In order for this expectation to be zero for all r, it must hold that

$$(\pi_g - \mathbb{E}[A_t(X)r^*(X)])\Sigma_1^{-1}A_t(X) - r^*(X)'B(X) = 0.$$

We can verify that the solution $r^*(X)$ is given by

$$r^*(X) = \frac{B(X)^{-1}A_t(X)'\pi_g}{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1}.$$

To see this, notice that we have

$$\mathbb{E}[A_t(X)r^*(X)] = \frac{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)']\pi_g}{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1}.$$

Then we have

$$
\begin{aligned}
r^*(X)'B(X) &= \frac{\pi_g A_t(X)B(X)^{-1}B(X)}{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1} = \frac{\pi_g A_t(X)}{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1} \\
&= (\pi_g - \mathbb{E}[A_t(X)r^*(X)])\Sigma_1^{-1}A_t(X).
\end{aligned}
$$

Substituting this expression for r^* into the minimization problem, we obtain the Fisher information:

$$
\begin{aligned}
&(\pi_g - \mathbb{E}[A_t(X)r^*(X)])\Sigma_1^{-1}(\pi_g - \mathbb{E}[A_t(X)r^*(X)]) + \mathbb{E}[r^*(X)'B(X)r^*(X)] \\
=&\frac{\Sigma_1\pi_g^2}{(\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1)^2} + \frac{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)']\pi_g^2}{(\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1)^2} \\
=&\frac{\pi_g^2}{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1}.
\end{aligned}
$$

The semiparametric efficiency bound is the inverse of the Fisher information:

$$
\frac{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1}{\pi_g^2}.
$$

To further simplify the inverse matrix $B(X)^{-1} = (L(X)'\Sigma_2(X)^{-1}L(X))^{-1}$, we aim to show that this matrix has the following block-diagonal form:

$$
(L(X)'\Sigma_2(X)^{-1}L(X))^{-1} = \begin{pmatrix} 1/(\sum_{g',t'}\sum_{j'=1}^{g-1} s_{j,j'}) & 0 & 0 \\ 0 & p_g(X)(1-p_g(X)) & -p_g(X)p_\infty(X) \\ 0 & -p_g(X)p_\infty(X) & p_\infty(X)(1-p_\infty(X)) \end{pmatrix},
$$

where $s_{j,j'}$ denotes the (j,j')-th element of the inverse matrix $S(X) \equiv \Sigma_2(X)^{-1}$. We denote the entries of $\Sigma_2(X)$ as

$$
\Sigma_2(X) = \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{g-1,1} & \sigma_{g,1} & \sigma_{g+1,1} \\ \vdots & & \vdots & \vdots & \vdots \\ \sigma_{g-1,1} & \cdots & \sigma_{g-1,g-1} & \sigma_{g,g-1} & \sigma_{g+1,g-1} \\ \sigma_{g,1} & \cdots & \sigma_{g,g-1} & \sigma_{g,g} & \sigma_{g+1,g} \\ \sigma_{g+1,1} & \cdots & \sigma_{g+1,g-1} & \sigma_{g+1,g} & \sigma_{g+1,g+1} \end{pmatrix}.
$$

Let the bottom-right 2×2 submatrix be denoted by

$$
\Pi \equiv \begin{pmatrix} \sigma_{g,g} & \sigma_{g+1,g} \\ \sigma_{g+1,g} & \sigma_{g+1,g+1} \end{pmatrix} = \begin{pmatrix} p_g(X)(1-p_g(X)) & -p_g(X)p_\infty(X) \\ -p_g(X)p_\infty(X) & p_\infty(X)(1-p_\infty(X)) \end{pmatrix}.
$$

For $1 \leqslant t' \leqslant g-1$, the term $\sigma_{g,t'}$ is

$$
\begin{aligned}
\sigma_{g,t'} &= \mathbb{E}\left[\left(CATT(g,t,X) - \frac{G_g(Y_t - Y_{t'})}{p_g(X)} + \frac{G_\infty(Y_t - Y_{t'})}{p_\infty(X)}\right)(G_g - p_g(X))|X\right] \\
&= p_g(X)CATT(g,t,X) - m_{g,t,t'}(X)
\end{aligned}
$$

$$= -(1 - p_g(X))m_{g,t,t'}(X) - p_g(X)m_{\infty,t,t'}(X).$$

The term $\sigma_{g+1,t'}$ is

$$\begin{aligned}
\sigma_{g+1,t'} &= \mathbb{E}\left[\left(CATT(g,t,X) - \frac{G_g(Y_t - Y_{t'})}{p_g(X)} + \frac{G_\infty(Y_t - Y_{t'})}{p_\infty(X)}\right)(G_\infty - p_\infty(X))|X\right] \\
&= p_\infty(X)CATT(g,t,X) + m_{\infty,t,t'}(X) \\
&= p_\infty(X)m_{g,t,t'}(X) + (1 - p_\infty(X))m_{\infty,t,t'}(X).
\end{aligned}$$

Denote two vectors $\boldsymbol{a} = (a_1, \cdots, a_{g-1})$ and $\boldsymbol{b} = (b_1, \cdots, b_{g-1})$ respectively as $a_{t'} \equiv m_{g,t,t'}(X)/p_g(X)$ and $b_{t'} \equiv -m_{\infty,t,t'}(X)/p_\infty(X)$. Then the matrix $L(X)$ can be written as

$$L(X) = \begin{pmatrix} \mathbf{1}_{g-1} & (\boldsymbol{a}', \boldsymbol{b}') \\ \mathbf{0}_2 & -I_2 \end{pmatrix},$$

where $\mathbf{1}_{g-1}$ is a column vector of ones of length $g-1$, $\mathbf{0}_2$ is a column vector of zeros of length 2, and I_2 is the 2-dimensional identity matrix. Notice that

$$\begin{aligned}
a_{t'}\sigma_{g,g} + b_{t'}\sigma_{g+1,g} = (1 - p_g(X))m_{g,t,t'}(X) + p_g(X)m_{\infty,t,t'}(X) = -\sigma_{g,t'}, \\
a_{t'}\sigma_{g+1,g} + b_{t'}\sigma_{g+1,g+1} = -p_\infty(X)m_{g,t,t'}(X) - (1 - p_\infty(X))m_{\infty,t,t'}(X) = -\sigma_{g+1,t'}.
\end{aligned}$$

In matrix notation, this means that

$$\Pi \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} = -\begin{pmatrix} \sigma_{g,1} & \cdots & \sigma_{g,g-1} \\ \sigma_{g+1,1} & \cdots & \sigma_{g+1,g-1} \end{pmatrix} \equiv -\begin{pmatrix} \tilde{\boldsymbol{\sigma}}_g \\ \tilde{\boldsymbol{\sigma}}_{g+1} \end{pmatrix} \implies \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} = -\Pi^{-1}\begin{pmatrix} \tilde{\boldsymbol{\sigma}}_g \\ \tilde{\boldsymbol{\sigma}}_{g+1} \end{pmatrix},$$

where we denote $\tilde{\boldsymbol{\sigma}}_g \equiv (\sigma_{g,1} \cdots \sigma_{g,g-1})$ and $\tilde{\boldsymbol{\sigma}}_{g+1} \equiv (\sigma_{g+1,1} \cdots \sigma_{g+1,g-1})$, which correspond to the second-to-last row and last row of $\Sigma_2(X)$, respectively, with the last two entries removed. Let $S = \Sigma_2(X)^{-1}$ be denoted by $S = (\boldsymbol{s}_1, \cdots, \boldsymbol{s}_{g+1})$, where $\boldsymbol{s}_j$ are column vectors. Let $\tilde{\boldsymbol{s}}_j$ denote the vector $\boldsymbol{s}_j$ with its last two entries removed. We now compute $L(X)'\Sigma_2(X)^{-1}$:

$$L(X)'S = \begin{pmatrix} \mathbf{1}'_{g-1} & 0 & 0 \\ \boldsymbol{a} & -1 & 0 \\ \boldsymbol{b} & 0 & -1 \end{pmatrix} S = \begin{pmatrix} \mathbf{1}'_{g-1}\tilde{\boldsymbol{s}}_1 & \cdots & \mathbf{1}'_{g-1}\tilde{\boldsymbol{s}}_{g-1} & \mathbf{1}'_{g-1}\tilde{\boldsymbol{s}}_g & \mathbf{1}'_{g-1}\tilde{\boldsymbol{s}}_{g+1} \\ 0 & \cdots & 0 & \boldsymbol{a}\tilde{\boldsymbol{s}}_g - s_{g,g} & \boldsymbol{a}\tilde{\boldsymbol{s}}_{g+1} - s_{g+1,g} \\ 0 & \cdots & 0 & \boldsymbol{b}\tilde{\boldsymbol{s}}_g - s_{g+1,g} & \boldsymbol{b}\tilde{\boldsymbol{s}}_{g+1} - s_{g+1,g+1} \end{pmatrix}.$$

The bottom-left entries are zero because because this submatrix can be written as

$$\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix}(\tilde{\boldsymbol{s}}_1, \cdots, \tilde{\boldsymbol{s}}_{g-1}) - \begin{pmatrix} \boldsymbol{s}_{g,1} & \cdots & \boldsymbol{s}_{g,g-1} \\ \boldsymbol{s}_{g+1,1} & \cdots & \boldsymbol{s}_{g+1,g-1} \end{pmatrix}.$$

Factoring out Π^{-1}, we see that this matrix is equal to Π^{-1} multiplied by the following matrix:

$$\begin{pmatrix} \tilde{\boldsymbol{\sigma}}_g \\ \tilde{\boldsymbol{\sigma}}_{g+1} \end{pmatrix}(\tilde{\boldsymbol{s}}_1, \cdots, \tilde{\boldsymbol{s}}_{g-1}) + \Pi \begin{pmatrix} \boldsymbol{s}_{g,1} & \cdots & \boldsymbol{s}_{g,g-1} \\ \boldsymbol{s}_{g+1,1} & \cdots & \boldsymbol{s}_{g+1,g-1} \end{pmatrix}$$

$$= \begin{pmatrix} \tilde{\boldsymbol{\sigma}}_g \tilde{\boldsymbol{s}}_{t'} + \sigma_{g,g} s_{g,t'} + \sigma_{g+1,g} s_{g+1,t'} \\ \tilde{\boldsymbol{\sigma}}_{g+1} \tilde{\boldsymbol{s}}_{t'} + \sigma_{g+1,g} s_{g,t'} + \sigma_{g+1,g+1} s_{g+1,t'} \end{pmatrix}_{1 \leqslant t' \leqslant g-1}$$

$$= \begin{pmatrix} \boldsymbol{\sigma}_g \boldsymbol{s}_1 & \cdots & \boldsymbol{\sigma}_g \boldsymbol{s}_{g-1} \\ \boldsymbol{\sigma}_{g+1} \boldsymbol{s}_1 & \cdots & \boldsymbol{\sigma}_{g+1} \boldsymbol{s}_{g-1} \end{pmatrix},$$

where $\boldsymbol{\sigma}_g$ and $\boldsymbol{\sigma}_{g+1}$ denote the second-to-last and last rows, respectively, of $\Sigma_2(X)$. The above entries are all zero because, by definition, S is the inverse of $\Sigma_2(X)$. Then $L(X)'SL(X)$ is equal to

$$L(X)'SL(X) = \begin{pmatrix} \sum_{t'=1}^{g-1} \mathbf{1}'_{g-1} \tilde{\boldsymbol{s}}_{t'} & 0 & 0 \\ 0 & -\boldsymbol{a}\tilde{\boldsymbol{s}}_g + s_{g,g} & -\boldsymbol{a}\tilde{\boldsymbol{s}}_{g+1} + s_{g+1,g} \\ 0 & -\boldsymbol{b}\tilde{\boldsymbol{s}}_g + s_{g+1,g} & -\boldsymbol{b}\tilde{\boldsymbol{s}}_{g+1} + s_{g+1,g+1} \end{pmatrix}.$$

Observe that the upper-right block is also zero because $L(X)'SL(X)$ is symmetric by construction. The bottom-right 2×2 matrix

$$\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{s}}_g & \tilde{\boldsymbol{s}}_{g+1} \end{pmatrix} + \begin{pmatrix} s_{g,g} & s_{g+1,g} \\ s_{g+1,g} & s_{g+1,g+1} \end{pmatrix} = \Pi^{-1}.$$

This holds because, after factoring out Π^{-1}, this matrix equals Π^{-1} multiplied by the following matrix

$$\begin{pmatrix} \tilde{\boldsymbol{\sigma}}_g \\ \tilde{\boldsymbol{\sigma}}_{g+1} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{s}}_g & \tilde{\boldsymbol{s}}_{g+1} \end{pmatrix} + \Pi \begin{pmatrix} s_{g,g} & s_{g+1,g} \\ s_{g+1,g} & s_{g+1,g+1} \end{pmatrix}$$
$$= \begin{pmatrix} \tilde{\boldsymbol{\sigma}}_g \tilde{\boldsymbol{s}}_g + \sigma_{g,g} s_{g,g} + \sigma_{g+1,g} s_{g+1,g} & \tilde{\boldsymbol{\sigma}}_g \tilde{\boldsymbol{s}}_{g+1} + \sigma_{g,g} s_{g+1,g} + \sigma_{g+1,g} s_{g+1,g+1} \\ \tilde{\boldsymbol{\sigma}}_{g+1} \tilde{\boldsymbol{s}}_g + \sigma_{g+1,g} s_{g,g} + \sigma_{g+1,g+1} s_{g+1,g} & \tilde{\boldsymbol{\sigma}}_{g+1} \tilde{\boldsymbol{s}}_{g+1} + \sigma_{g+1,g} s_{g+1,g} + \sigma_{g+1,g+1} s_{g+1,g+1} \end{pmatrix}$$
$$= \begin{pmatrix} \boldsymbol{\sigma}_g \boldsymbol{s}_g & \boldsymbol{\sigma}_g \boldsymbol{s}_{g+1} \\ \boldsymbol{\sigma}_{g+1} \boldsymbol{s}_g & \boldsymbol{\sigma}_{g+1} \boldsymbol{s}_{g+1} \end{pmatrix} = I_2,$$

where the last equality follows because S is the inverse of $\Sigma_2(X)$. Therefore, by the property of the block-diagonal matrix, the inverse of $L(X)'SL(X)$ is again a block-diagonal matrix:

$$(L(X)'SL(X))^{-1} = \operatorname{diag}\left(\left(\sum_{j,j'=1}^{g-1} s_{j,j'} \right)^{-1}, \Pi \right).$$

Applying this block-diagonal structure, we obtain that

$$A_t(X)(L(X)'SL(X))^{-1} A_t(X)' = \frac{p_g(X)^2}{\sum_{j,j'=1}^{g-1} s_{j,j'}} + p_g(X)(1 - p_g(X))(CATT(g,t,X) - ATT(g,t))^2. \quad \text{(C.2)}$$

The term $\frac{1}{\sum_{j,j'=1}^{g-1} s_{j,j'}} = |\Sigma_2(X)| / (\sum_{j,j'=1}^{g-1} (-1)^{j+j'} |\Sigma_{2,jj'}|)$, where $\Sigma_{2,jj'}$ denotes the submatrix of S obtained by removing the jth row and j'th column. We decompose Σ_2 into 4 blocks:

$$\sigma_2 = \begin{pmatrix} V_1 & V_2' \\ V_2 & \Pi \end{pmatrix},$$

where V_1 and V_2 are defined as

$$V_1 \equiv \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{g-1,1} \\ \vdots & & \vdots \\ \sigma_{g-1,1} & \cdots & \sigma_{g-1,g-1} \end{pmatrix}, V_2 \equiv \begin{pmatrix} \sigma_{g,1} & \cdots & \sigma_{g,g-1} \\ \sigma_{g+1,1} & \cdots & \sigma_{g+1,g-1} \end{pmatrix}.$$

The determinant of Σ_2 can be computed using the Schur complement as

$$|\Sigma_2| = |\Pi||V_1 - V_2'\Pi^{-1}V_2|.$$

The term $|\Pi|$ equals $\sigma_{g,g}\sigma_{g+1,g+1} - \sigma_{g+1,g}^2$. The Schur complement of Π is the $(g-1)\times(g-1)$ symmetric matrix $V_1 - V_2'\Pi^{-1}V_2$, whose (j,j')-th elements is

$$\sigma_{j,j'} - \frac{\sigma_{g,j}\sigma_{g,j'}\sigma_{g+1,g+1} - \sigma_{g,j}\sigma_{g+1,j'}\sigma_{g+1,g} - \sigma_{g,j'}\sigma_{g+1,j}\sigma_{g+1,g} - \sigma_{g+1,j}\sigma_{g+1,j'}\sigma_{g,g}}{\sigma_{g,g}\sigma_{g+1,g+1} - \sigma_{g+1,g}^2}.$$

A direct calculation verifies that the above term is equal to the (j,j')-th entry of $V_{gt}^*(X)$, yielding the identity $V_1 - V_2'\Pi^{-1}V_2 = V_{gt}^*(X)$, where recall that $V_{gt}^*(X)$ is the $(g-1)\times(g-1)$ matrix whose (j,k)-th element is

$$\frac{1}{p_g(X)}\operatorname{Cov}(Y_t - Y_j, Y_t - Y_k|G = g, X) + \frac{1}{1-p_g(X)}\operatorname{Cov}(Y_t - Y_j, Y_t - Y_k|G = \infty, X). \tag{C.3}$$

The same procedure can be applied to each $\Sigma_{2,jj'}$ and show that the determinant of each block $\Sigma_{2,jj'}$ is equal to $|\Pi|$ multiplied by $|V_{gt,jj'}^*(X)|$, where $V_{gt,jj'}^*(X)$ is the minor of $V_{gt}^*(X)$ formed by deleting its j-th row and j'-th column. Substituting these into (C.2), we find that the semiparametric efficiency bound equals:

$$\begin{aligned}
&\frac{\mathbb{E}[A_t(X)B(X)^{-1}A_t(X)'] + \Sigma_1}{\pi_g^2} \\
=&\frac{1}{\pi_g^2}\mathbb{E}\left[\frac{p_g(X)^2}{\sum_{j,j'=1}^{g-1} s_{j,j'}} + p_g(X)(CATT(g,t,X) - ATT(g,t))^2\right] \\
=&\frac{1}{\pi_g^2}\mathbb{E}\left[\frac{p_g(X)^2|V_{gt}^*(X)|}{\sum_{j,j'=1}^{g-1}(-1)^{j+j'}|V_{gt,jj'}^*(X)|} + p_g(X)(CATT(g,t,X) - ATT(g,t))^2\right].
\end{aligned}$$

The efficient influence function is the efficiency bound multiplied by the efficient score function. The efficient score function is given by the proof of Theorem 2.1 in Ai and Chen (2012):

$$-\left(\pi_g - \mathbb{E}[A_t(X)r^*(X)]\right)\Sigma_1^{-1}(p_g(X)(ATT(g,t) - CATT(g,t,X))) + (L(X)r^*(X))'\Sigma_2(X)^{-1}\rho_2(W,h^*).$$

The first part of the efficient score is equal to

$$\left(\pi_g - \mathbb{E}[A_t(X)r^*(X)]\right)\Sigma_1^{-1}(p_g(X)(ATT(g,t) - CATT(g,t,X))) = \frac{p_g(X)(ATT(g,t) - CATT(g,t,X))}{\pi_g\Omega_g(ATT(g,t))}.$$

In the second part, $L(X)r^*(X)$ is equal to

$$L(X)r^*(X) = \frac{L(X)B(X)^{-1}A_t(X)'}{\pi_g \Omega_g(ATT(g,t))}.$$

We calculate the term $A_t(X)'$ in the above numerator. Denote $\bar{s} \equiv \sum_{j,j'=1}^{g-1} s_{j,j'}$. The product $L(X)B(X)^{-1}$ is

$$\begin{aligned}
&L(X)B(X)^{-1} \\
&= \begin{pmatrix} 1 & \frac{m_{g,t,1}(X)}{p_g(X)} & -\frac{m_{\infty,t,1}(X)}{p_\infty(X)} \\ \vdots & & \\ 1 & \frac{m_{g,t,g-1}(X)}{p_g(X)} & -\frac{m_{\infty,t,g-1}(X)}{p_\infty(X)} \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} (\bar{s})^{-1} & 0 & 0 \\ 0 & p_g(X)(1-p_g(X)) & -p_g(X)p_\infty(X) \\ 0 & -p_g(X)p_\infty(X) & p_\infty(X)(1-p_\infty(X)) \end{pmatrix} \\
&= \begin{pmatrix} (\bar{s})^{-1} & (1-p_g(X))m_{g,t,1}(X) + p_g(X)m_{\infty,t,1}(X) & -p_\infty(X)m_{g,t,1}(X) - (1-p_\infty(X))m_{\infty,t,1}(X) \\ \vdots & \vdots & \vdots \\ (\bar{s})^{-1} & (1-p_g(X))m_{g,t,g-1}(X) + p_g(X)m_{\infty,t,g-1}(X) & -p_\infty(X)m_{g,t,1}(X) - (1-p_\infty(X))m_{\infty,t,g-1}(X) \\ 0 & -p_g(X)(1-p_g(X)) & p_g(X)p_\infty(X) \\ 0 & p_g(X)p_\infty(X) & -p_\infty(X)(1-p_\infty(X)) \end{pmatrix} \\
&= \begin{pmatrix} ((\mathbf{1}_{g-1}(\bar{s})^{-1})', 0, 0)' & -\boldsymbol{\sigma}_g & -\boldsymbol{\sigma}_{g+1} \end{pmatrix}.
\end{aligned}$$

Multiplying the above matrix with $A_t(X)'$, we obtain

$$\begin{aligned}
L(X)B(X)^{-1}A_t(X)' &= \begin{pmatrix} ((\mathbf{1}_{g-1}(\bar{s})^{-1})', 0, 0)' & -\boldsymbol{\sigma}_g & -\boldsymbol{\sigma}_{g+1}. \end{pmatrix} \begin{pmatrix} -p_g(X) \\ ATT(g,t) - CATT(g,t,X) \\ 0 \end{pmatrix} \\
&= -(p_g(X)(\bar{s})^{-1}(\mathbf{1}'_{g-1}, 0, 0)' + (ATT(g,t) - CATT(g,t,X))\boldsymbol{\sigma}_g).
\end{aligned}$$

Multiplying the transpose of this matrix with the inverse of $\Sigma_2(X)$, we obtain

$$\begin{aligned}
&-(p_g(X)(\bar{s})^{-1}(\mathbf{1}'_{g-1}, 0, 0) + (ATT(g,t) - CATT(g,t,X))\boldsymbol{\sigma}'_g)(\boldsymbol{s}_1, \cdots, \boldsymbol{s}_{g+1}) \\
&= -p_g(X)(\bar{s})^{-1}((\mathbf{1}'_{g-1}, 0, 0)\boldsymbol{s}_1, \cdots, (\mathbf{1}'_{g-1}, 0, 0)\boldsymbol{s}_{g+1}) - (ATT(g,t) - CATT(g,t,X))(0, \cdots, 0, 1, 0) \\
&= -\frac{p_g(X)}{\bar{s}} \left(\sum_{j'=1}^{g-1} s_{1j'}, \cdots, \sum_{j'=1}^{g-1} s_{g+1,j'} \right) - (ATT(g,t) - CATT(g,t,X))(0, \cdots, 0, 1, 0)
\end{aligned}$$

because $\boldsymbol{\sigma}'_g \boldsymbol{s}_j = \mathbf{1}\{g = j\}$. Lastly, we multiply the above matrix with the first stage moments ρ_2 and obtain that

$$\begin{aligned}
&L(X)B(X)^{-1}A_t(X)'\Sigma_2(X)^{-1}\rho_2(W, h^*) \\
&= -\frac{p_g(X)}{\bar{s}} \sum_{j,j'=1}^{g-1} s_{j,j'} \left(CATT(g,t,X) - \frac{G_g(Y_t - Y_j)}{p_g(X)} + \frac{G_\infty(Y_t - Y_j)}{p_\infty(X)} \right)
\end{aligned}$$

$$
- \frac{p_g(X)}{\bar{s}} \sum_{j'=1}^{g-1} s_{g,j'}(G_g - p_g(X)) - \frac{p_g(X)}{\bar{s}} \sum_{j'=1}^{g-1} s_{g+1,j'}(G_\infty - p_\infty(X))
$$

$$
- (ATT(g,t) - CATT(g,t,X))(G_g - p_g(X)).
$$

This gives the expression for the efficient score. The efficient influence function is equal to the efficient score premultiplied by the efficiency bound:

$$
\begin{aligned}
\mathbb{EIF}_g(ATT(g,t)) = \frac{1}{\pi_g}\Bigg(& G_g(CATT(g,t,X) - ATT(g,t)) \\
& + \frac{p_g(X)}{\bar{s}} \sum_{j,j'=1}^{g-1} s_{j,j'} \left(\frac{G_g(Y_t - Y_j)}{p_g(X)} - \frac{G_\infty(Y_t - Y_j)}{p_\infty(X)} - CATT(g,t,X) \right) \\
& - \frac{p_g(X)}{\bar{s}} \sum_{j'=1}^{g-1} s_{g,j'}(G_g - p_g(X)) - \frac{p_g(X)}{\bar{s}} \sum_{j'=1}^{g-1} s_{g+1,j'}(G_\infty - p_\infty(X)) \Bigg).
\end{aligned}
$$

By direct calculation, we can show that the EIF can be written as

$$
\begin{aligned}
\mathbb{EIF}_g(ATT(g,t)) = \frac{1}{\pi_g}\Bigg(& G_g(CATT(g,t,X) - ATT(g,t)) + \sum_{j,j'=1}^{g-1} \frac{s_{j,j'}}{\bar{s}} G_g(Y_t - Y_j - m_{g,t,j}(X)) \\
& - \sum_{j,j'=1}^{g-1} \frac{s_{j,j'}}{\bar{s}} \frac{p_g(X)}{p_\infty(X)} G_\infty(Y_t - Y_j - m_{\infty,t,j}(X)) \Bigg) \\
& + \frac{p_g(X)}{\bar{s}}(G_g - p_g(X)) \sum_{j'=1}^{g-1} \left(\sum_{g',t'} s_{j,j'} \frac{m_{g,t,j}}{p_g} - s_{g,j'} \right), \qquad \text{(C.4)} \\
& + \frac{p_g(X)}{\bar{s}}(G_\infty - p_\infty(X)) \sum_{j'=1}^{g-1} \left(\sum_{g',t'} s_{j,j'} \frac{m_{\infty,t,j}}{p_\infty} - s_{g+1,j'} \right). \qquad \text{(C.5)}
\end{aligned}
$$

We want to show that the last two terms on the left-hand side of the above equation, (C.4) and (C.5), are zero. Notice that for each $1 \leqslant j' \leqslant g-1$, the term

$$
\sum_{g',t'} s_{j,j'} \frac{m_{g,t,j}}{p_g} - s_{g,j'}
$$

is equal to the j'th row of $\Sigma_2(X)^{-1}$ multiplied by the second column of $L(X)$. In other words, it is the (g,j')-th element of the matrix $L(X)'\Sigma_2(X)^{-1}$. By the previous analysis, this term is zero for any $j \leqslant g-1$, which implies that (C.4) evaluates to zero. Similarly, the term

$$
\sum_{g',t'} s_{j,j'} \frac{m_{\infty,t,j}}{p_\infty} - s_{g+1,j'}
$$

is the $(g+1, j')$-th element of the matrix $L(X)'\Sigma_2(X)^{-1}$, which is also equal to zero, implying that (C.5) is zero. The weights $s_{j,j'}/\bar{s}$ can be represented using $V^*_{gt}(X)$ as

$$\frac{s_{j,j'}}{\bar{s}} = \frac{(-1)^{j+j'}|\Sigma_{2,jj'}|/|\Sigma_2|}{\sum_{j,j'=1}^{g-1}(-1)^{j+j'}|\Sigma_{2,jj'}|/|\Sigma_2|} = \frac{(-1)^{j+j'}|V^*_{gt,jj'}(X)|/|V^*_{gt}(X)}{\sum_{j,j'=1}^{g-1}(-1)^{j+j'}|V^*_{gt,jj'}(X)|/|V^*_{gt}(X)|} = \frac{(j,j')\text{th entry of } V^*_{gt}(X)^{-1}}{\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}}.$$

Notice that the first part of the EIF, $G_g(CATT(g,t,X) - ATT(g,t))$, does not depend on j and can be put into the weighted average. The expression of the EIF becomes

$$\mathbb{EIF}_g(ATT(g,t)) = \frac{\mathbf{1}'V^*_{gt}(X)^{-1}}{\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}}\mathbb{IF}_g(ATT(g,t)).$$

Lastly, we want to show that the weights can be represented by using $V_{gt}(X)$. Notice that $V_{gt}(X)$ is equal to

$$V_{gt}(X) = \frac{p_g(X)^2}{\pi_g^2}V^*_{gt}(X) + c(X)\mathbf{1}\mathbf{1}',$$

where $c(X)$ is

$$c(X) \equiv \frac{1}{\pi_g^2}p_g(X)(1-p_g(X))(CATT(g,t,X) - ATT(g,t))^2.$$

By the Sherman–Morrison formula, $V_{gt}(X)^{-1}$ is equal to

$$V_{gt}(X)^{-1} = \frac{\pi_g^2}{p_g(X)^2}V^*_{gt}(X)^{-1} - \frac{\frac{c(X)\pi_g^4}{p_g(X)^4}V^*_{gt}(X)^{-1}\mathbf{1}\mathbf{1}'V^*_{gt}(X)^{-1}}{1 + \frac{c(X)\pi_g^2}{p_g(X)^2}\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}}.$$

Therefore, we have

$$\begin{aligned}
\mathbf{1}'V_{gt}(X)^{-1} &= \frac{\pi_g^2}{p_g(X)^2}\mathbf{1}'V^*_{gt}(X)^{-1} - \frac{\frac{c(X)\pi_g^4}{p_g(X)^4}\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}\mathbf{1}'V^*_{gt}(X)^{-1}}{1 + \frac{c(X)\pi_g^2}{p_g(X)^2}\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}} \\
&= \mathbf{1}'V^*_{gt}(X)^{-1}\left(\frac{\pi_g^2}{p_g(X)^2} - \frac{\frac{c(X)\pi_g^4}{p_g(X)^4}\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}}{1 + \frac{c(X)\pi_g^2}{p_g(X)^2}\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}}\right),
\end{aligned}$$

and

$$\mathbf{1}'V_{gt}(X)^{-1}\mathbf{1} = \mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}\left(\frac{\pi_g^2}{p_g(X)^2} - \frac{\frac{c(X)\pi_g^4}{p_g(X)^4}\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}}{1 + \frac{c(X)\pi_g^2}{p_g(X)^2}\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}}\right).$$

Their ratio is equal to

$$\frac{\mathbf{1}'V_{gt}(X)^{-1}}{\mathbf{1}'V_{gt}(X)^{-1}\mathbf{1}} = \frac{\mathbf{1}'V^*_{gt}(X)^{-1}}{\mathbf{1}'V^*_{gt}(X)^{-1}\mathbf{1}}.$$

This completes the first part of the proof, which gives the EIF for $ATT(g,t)$ in the submodel constructed for a single time period t.

Part 2 For the second part of the proof, we examine the entire set of moments in Lemma 3.1 for all post-treatment periods:

$$\begin{aligned}\mathbb{E}[G_g(ATT(g,t)-CATT(g,t,X))] &= 0,\ \text{for all } t\in[g,T],\\ \mathbb{E}\left[CATT(g,t,X)-\frac{G_g(Y_t-Y_{g-1})}{p_g(X)}+\frac{G_\infty(Y_t-Y_{g-1})}{p_\infty(X)}\Big|X\right] &= 0,\ \text{for all } t\in[g,T],\\ \mathbb{E}\left[\frac{G_g(Y_{t'}-Y_1)}{p_g(X)}-\frac{G_\infty(Y_{t'}-Y_1)}{p_\infty(X)}\Big|X\right] &= 0, 2\leqslant t'\leqslant g-1,\\ \mathbb{E}[G_g-p_g(X)|X] &= 0.\end{aligned}$$

Intuitively, the additional moment conditions do not alter the efficiency bound of $ATT(g,t)$, and we aim to establish this formally. We first show that, to derive the EIF of a single $ATT(g,t)$, we may remove the unconditional moments (in the first line) corresponding time periods other than t. In other words, it suffices to examine the following model:

$$\begin{aligned}\mathbb{E}[G_g(ATT(g,t)-CATT(g,t,X))] &= 0,\\ \mathbb{E}\left[CATT(g,t,X)-\frac{G_g(Y_t-Y_{g-1})}{p_g(X)}+\frac{G_\infty(Y_t-Y_{g-1})}{p_\infty(X)}\Big|X\right] &= 0,\ \text{for all } t\in[g,T],\\ \mathbb{E}\left[\frac{G_g(Y_{t'}-Y_1)}{p_g(X)}-\frac{G_\infty(Y_{t'}-Y_1)}{p_\infty(X)}\Big|X\right] &= 0, 2\leqslant t'\leqslant g-1,\\ \mathbb{E}[G_g-p_g(X)|X] &= 0.\end{aligned} \tag{C.6}$$

To prove this point, we use the notations $\tilde{\rho}_1,\tilde{\rho}_2,\tilde{m}_1,\tilde{m}_2,\tilde{\Sigma}_1,\tilde{\Sigma}_2$ to denote the corresponding terms in the larger model. Denote $\theta=(ATT(g,t):t\in[g,T])$ as the vector of finite-dimensional parameter and $h=(CATT(g,g,\cdot),\cdots,CATT(g,T,\cdot),p_g)$ the nuisance parameters. The derivatives are

$$\begin{aligned}\frac{d\tilde{m}_1(\alpha^*)}{d\theta} &= \pi_g I, & \frac{d\tilde{m}_2(X,\alpha^*)}{d\theta} &= 0,\\ \frac{d\tilde{m}_1(\alpha^*)}{dh}[r] &= A(X)r(X), & \frac{d\tilde{m}_2(X,\alpha^*)}{dh}[r] &= \tilde{L}(X)r(X),\end{aligned}$$

where A essentially stacks all the A_t's, i.e.,

$$A(X)=(-p_g(X)I,(ATT(g,g)-CATT(g,T,X),\cdots,ATT(g,g)-CATT(g,T,X))).$$

We omit the specific expression of $\tilde{L}(X)$ for now. Define $\tilde{B}(X)\equiv\tilde{L}(X)'\tilde{\Sigma}_2(X)^{-1}\tilde{L}(X)$. We want to solve the following optimization:

$$\inf_r\quad(\pi_g I-\mathbb{E}[A(X)r(X)])\tilde{\Sigma}_1^{-1}(\pi_g-\mathbb{E}[A(X)r(X)])+\mathbb{E}[(\tilde{L}(X)r(X))'\tilde{\Sigma}_2(X)^{-1}\tilde{L}(X)r(X)].$$

By solving the first-order condition, we obtain the optimal r^* as

$$r^*(X)' = \pi_g(\tilde{\Sigma}_1 + \mathbb{E}[A(X)\tilde{B}(X)^{-1}A(X)'])^{-1}A(X)\tilde{B}(X)^{-1}.$$

Therefore, the efficient score is

$$\begin{aligned}&-\pi_g(\tilde{\Sigma}_1 + \mathbb{E}[A(X)\tilde{B}(X)^{-1}A(X)'])^{-1}\tilde{\rho}_1 \\ &+\pi_g(\tilde{\Sigma}_1 + \mathbb{E}[A(X)\tilde{B}(X)^{-1}A(X)'])^{-1}A(X)\tilde{B}(X)^{-1}\tilde{L}(X)'\tilde{\Sigma}_2(X)^{-1}\tilde{\rho}_2.\end{aligned}$$

The efficiency bound is the inverse of the expected outer product of the efficient score:

$$(\tilde{\Sigma}_1 + \mathbb{E}[A(X)\tilde{B}(X)^{-1}A(X)'])/\pi_g^2.$$

The efficient influence function is equal to the efficiency score pre-multiplied by the efficiency bound:

$$-\frac{1}{\pi_g}\tilde{\rho}_1 + \frac{1}{\pi_g}A(X)\tilde{B}(X)^{-1}\tilde{L}(X)'\tilde{\Sigma}_2(X)^{-1}\tilde{\rho}_2.$$

For each $ATT(g,t)$ as an entry of θ, its EIF is equal to the corresponding row in the above expression:

$$\begin{aligned}&-\frac{1}{\pi_g}p_g(X)(ATT(g,t) - CATT(g,t,X)) \\ &+\frac{1}{\pi_g}(0,\cdots,0,-p_g(X),0,\cdots,0,(ATT(g,t) - CATT(g,t,X)))\tilde{B}(X)^{-1}\tilde{L}(X)'\tilde{\Sigma}_2(X)^{-1}\tilde{\rho}_2.\end{aligned} \tag{C.7}$$

Observe that this is exactly the expression for the EIF we would obtain for the model given by (C.6). This proves that to derive the EIF of a single $ATT(g,t)$, we can remove the unconditional moments (in the first line) for time periods other than t.

The remaining task for the second part of the proof is to show that the conditional moments defining the irrelevant/redundant $CATT(g,t'',\cdot), t'' \neq t$ can also be removed. If this is true, then we can claim that the EIFs derived in the first part (based on the model given in (C.1)) coincide with the EIFs for the entire set of moment restrictions in Lemma 3.1. This can be shown by induction. Assume first that there is only one additional conditional moment for an irrelevant $CATT(g,t'',X)$. For convenience, this moment is placed at the end of the model:

$$\begin{aligned}\mathbb{E}[G_g(ATT(g,t) - CATT(g,t,X))] &= 0,\\ \mathbb{E}\left[CATT(g,t,X) - \frac{G_g(Y_t - Y_{g-1})}{p_g(X)} + \frac{G_\infty(Y_t - Y_{g-1})}{p_\infty(X)}\Big|X\right] &= 0,\\ \mathbb{E}\left[\frac{G_g(Y_{t'} - Y_1)}{p_g(X)} - \frac{G_\infty(Y_{t'} - Y_1)}{p_\infty(X)}\Big|X\right] &= 0, 2 \leqslant t' \leqslant g-1,\\ \mathbb{E}[G_g - p_g(X)|X] &= 0,\\ \mathbb{E}\left[CATT(g,t'',X) - \frac{G_g(Y_{t''} - Y_{g-1})}{p_g(X)} + \frac{G_\infty(Y_{t''} - Y_{g-1})}{p_\infty(X)}\Big|X\right] &= 0.\end{aligned}$$

The parameters of this model are $\theta = ATT(g,t)$ and $h = (CATT(g,t,\cdot), p_g, CATT(g,t'',\cdot))$. The second term of the EIF given in (C.7) now becomes

$$\frac{1}{\pi_g}(A_t(X), 0)\tilde{B}(X)^{-1}\tilde{L}(X)'\tilde{\Sigma}_2(X)^{-1}\tilde{\rho}_2. \tag{C.8}$$

The goal is to show that this is the same as the one for the model derived in the first part of the proof

$$\frac{1}{\pi_g}A_t(X)B(X)^{-1}L(X)'\Sigma_2(X)^{-1}\rho_2.$$

Since the redundant $CATT(g, t'', \cdot)$ does not appear in any other moments except the last one, we can decompose $\tilde{L}$ as

$$\tilde{L} = \begin{pmatrix} L & \mathbf{0} \\ \ell' & 1 \end{pmatrix},$$

where the specific expression for ℓ is not important here. We write the inverse matrix $\tilde{\Sigma}_2(X)^{-1} \equiv \tilde{S}$ as the following 2×2 block matrix

$$\tilde{S} = \begin{pmatrix} \tilde{S}_{UL} & \tilde{S}_{LL}' \\ \tilde{S}_{LL} & \tilde{S}_{LR} \end{pmatrix}.$$

The matrix $\tilde{B}$ is equal to

$$\begin{aligned}\tilde{B} = \tilde{L}'\tilde{S}\tilde{L} &= \begin{pmatrix} L' & \ell \\ \mathbf{0}' & 1 \end{pmatrix}\begin{pmatrix} \tilde{S}_{UL} & \tilde{S}_{LL}' \\ \tilde{S}_{LL} & \tilde{S}_{LR} \end{pmatrix}\begin{pmatrix} L & \mathbf{0} \\ \ell' & 1 \end{pmatrix} \\ &= \begin{pmatrix} L'\tilde{S}_{UL}L + \ell\tilde{S}_{LL} + L'\tilde{S}_{LL}'\ell' + \ell\tilde{S}_{LR}\ell' & L'\tilde{S}_{LL}' + \ell\tilde{S}_{LR} \\ \tilde{S}_{LL}L + \tilde{S}_{LR}\ell' & \tilde{S}_{LR} \end{pmatrix}.\end{aligned}$$

Using Schur complement for block matrix inversion, we know that the upper-left block of $\tilde{B}^{-1}$ is equal to the inverse of the Schur complement of $\tilde{S}_{LR}$:

$$\begin{aligned}&\left(L'\tilde{S}_{UL}L + \ell\tilde{S}_{LL} + L'\tilde{S}_{LL}'\ell' + \ell\tilde{S}_{LR}\ell' - (L'\tilde{S}_{LL}' + \ell\tilde{S}_{LR})S_{LR}^{-1}(\tilde{S}_{LL}L + \tilde{S}_{LR}\ell')\right)^{-1} \\ =&(L'(\tilde{S}_{UL} - \tilde{S}_{LL}'\tilde{S}_{LR}^{-1}\tilde{S}_{LL})L)^{-1} = (L'\Sigma_2^{-1}L)^{-1},\end{aligned}$$

where the second inequality follows from the fact that Σ_2 is the upper-left block of the inverse of $\tilde{S}$. Similarly, we can show that the upper-right block of $\tilde{B}^{-1}$ is equal to $-(L'\Sigma_2^{-1}L)^{-1}(L'\tilde{S}_{LL}' + \ell\tilde{S}_{LR})\tilde{S}_{LR}^{-1}$. Now the expression in (C.8) becomes

$$\begin{aligned}&\frac{1}{\pi_g}A_t(X)\left((L'\Sigma_2^{-1}L)^{-1} \quad -(L'\Sigma_2^{-1}L)^{-1}(L'\tilde{S}_{LL}' + \ell\tilde{S}_{LR})\tilde{S}_{LR}^{-1}\right)\begin{pmatrix} L'\tilde{S}_{UL}' + \ell\tilde{S}_{LL} & L'\tilde{S}_{LL}' + \ell\tilde{S}_{LR} \\ \tilde{S}_{LL} & S_{LR} \end{pmatrix}\tilde{\rho}_2 \\ =&\frac{1}{\pi_g}A_t(X)(L'\Sigma_2^{-1}L)^{-1}(L'\Sigma_2^{-1}L, 0)\tilde{\rho}_2 \\ =&\frac{1}{\pi_g}A_t(X)B(X)^{-1}L(X)'\Sigma_2(X)^{-1}\rho_2,\end{aligned}$$

where the last equality follows from the fact that $\tilde{\rho}_2$ augments ρ_2 by including an additional moment for $CATT(g, t'', \cdot)$ at the end. This proves that the efficiency bound remains unchanged when we include an additional moment based on an irrelevant $CATT(g, t'', \cdot), t'' \neq t$. We can then sequentially include further moments for irrelevant CATTs, and, by induction, show that these extra moments leave the efficient influence function unchanged. This completes the second part of the proof. □

Proof of Corollary 3.1. In the exactly identified model under Assumption PT-Post, there is only one influence function. Therefore, the weight $\frac{\mathbf{1}'V_{gt}(X)^{-1}}{\mathbf{1}'V_{gt}(X)^{-1}\mathbf{1}}$ is equal to one. The efficient influence function coincides with the only influence function. □

Proof of Theorem 3.2. We first orthogonalize the unconditional moment conditions by replacing G_g with $p_g(X)$ as in the proof of Theorem 3.1. The orthogonalized unconditional moments become the following: for each $g \in \mathcal{G}_{\text{trt}}$,

$$\mathbb{E}[\pi_g - p_g(X)] = 0,$$
$$\mathbb{E}[p_g(X)(ATT(g,t) - CATT(g,t,X))] = 0, g \leqslant t \leqslant T.$$

We reuse the notations $\rho_1, \rho_2, m_1, m_2, \Sigma_1, \Sigma_2$ in the proof of Theorem 3.1 to denote the corresponding terms in the set of moment restrictions in Lemma 3.2. Notice that the parameters π_g and $ATT(g,t)$ are just identified by the unconditional moments in ρ_1. In particular, each parameter only appears in a single moment equation. Therefore, we can derive the efficient influence function for each parameter separately.[3] The following proof is divided into two parts: (i) the efficient influence function for the group probability π_g, and (ii) the efficient influence function for the treatment effect ATT(g,t).

EIF for π_g For a single π_g, its efficient influence function can be derived equivalently using the following model

$$\mathbb{E}[\pi_g - p_g(X)] = 0,$$
$$\mathbb{E}\left[CATT(g,t,X) - \frac{G_g(Y_t - Y_1)}{p_g(X)} + \frac{G_\infty(Y_t - Y_1)}{p_\infty(X)}\Big|X\right] = 0, g \leqslant t \leqslant T, g \in \mathcal{G}_{\text{trt}},$$
$$\mathbb{E}\left[\frac{G_{g'}(Y_{t'} - Y_1)}{p_{g'}(X)} - \frac{G_\infty(Y_{t'} - Y_1)}{p_\infty(X)}\Big|X\right] = 0, 2 \leqslant t' \leqslant g' - 1, g' \in \mathcal{G}_{\text{trt}},$$
$$\mathbb{E}[G_{g'} - p_{g'}(X)|X] = 0, g' \in \mathcal{G}_{\text{trt}},$$
$$\mathbb{E}[G_\infty - p_\infty(X)|X] = 0.$$

Notice that the second line contains only definitions of CATT instead of restrictions. Since the CATTs are not involved in the definition of π_g, we can remove them from the model without affecting the calculation of the efficient influence function. Then the model becomes

$$\mathbb{E}[\pi_g - p_g(X)] = 0,$$

[3]Such claims can be verified more formally using the induction approach in the second part of the proof for Theorem 3.1. These proofs are omitted to avoid repetition.

$$\mathbb{E}\left[\frac{G_{g'}(Y_{t'}-Y_1)}{p_{g'}(X)}-\frac{G_\infty(Y_{t'}-Y_1)}{p_\infty(X)}\Big|X\right]=0, 2\leqslant t'\leqslant g'-1, g'\in\mathcal{G}_{\text{trt}},$$
$$\mathbb{E}[G_{g'}-p_{g'}(X)|X]=0, g'\in\mathcal{G}_{\text{trt}},$$
$$\mathbb{E}[G_\infty-p_\infty(X)|X]=0.$$

In this model, the parameters are $\theta=\pi_g$ and $h=(p_g,g\in\mathcal{G})$. The derivatives of m_1 and m_2 with respect to π_g are

$$\frac{dm_1(\alpha^*)}{d\pi_g}=1,\ \frac{dm_2(X,\alpha^*)}{d\pi_g}=0.$$

The derivatives of m_1 and m_2 with respect to h are

$$\frac{dm_1(\alpha^*)}{dh}[r]=\mathbb{E}[-\boldsymbol{e}_g'r(X)],$$
$$\frac{dm_2(X,\alpha^*)}{dh}[r]=L(X)r(X),$$

where $\boldsymbol{e}_g$ is the one-hot vector with the gth entry being 1 and the remaining entries being zero, and $L(X)$ is defined by

$$L(X)\equiv\begin{pmatrix}\left(\begin{pmatrix}-\frac{m_{g',2,1}(X)}{p_{g'}(X)}\\ \vdots\\ -\frac{m_{g',g'-1,1}(X)}{p_{g'}(X)}\end{pmatrix}\boldsymbol{e}_{g'}',\begin{pmatrix}\frac{m_{\infty,2,1}(X)}{p_\infty(X)}\\ \vdots\\ \frac{m_{\infty,g'-1,1}(X)}{p_\infty(X)}\end{pmatrix}\right)_{g'\in\mathcal{G}_{\text{trt}}}\\ -I\end{pmatrix},$$

where I denotes the identity matrix (of dimension $|\mathcal{G}|$). Similar to the proof of Theorem 3.1, we follow Theorem 2.1 of Ai and Chen (2012) and solve the following minimization problem:

$$\inf_r\quad(1+\mathbb{E}[\boldsymbol{e}_g'r(X)])\Sigma_1^{-1}(1+\mathbb{E}[\boldsymbol{e}_g'r(X)])+\mathbb{E}[(L(X)r(X))'\Sigma_2(X)^{-1}L(X)r(X)].$$

The corresponding solution is given by

$$r^*(X)=\frac{-B(X)^{-1}\boldsymbol{e}_g}{\mathbb{E}[\boldsymbol{e}_g'B(X)^{-1}\boldsymbol{e}_g]+\Sigma_1},$$

with $B(X)=L(X)'\Sigma_2(X)^{-1}L(X)$. The efficient score of π_g is given by

$$\begin{aligned}&-\left(\frac{dm_1(\alpha^*)}{d\pi_g}-\frac{dm_1(\alpha^*)}{dh}[r^*]\right)'\Sigma_1^{-1}\rho_1-\left(\frac{dm_2(X,\alpha^*)}{d\pi_g}-\frac{dm_2(X,\alpha^*)}{dh}[r^*]\right)'\Sigma_2(X)^{-1}\rho_2\\ =&-(\mathbb{E}[\boldsymbol{e}_g'B(X)^{-1}\boldsymbol{e}_g]+\Sigma_1)^{-1}(\rho_1+\boldsymbol{e}_g'B(X)^{-1}L(X)'\Sigma_2(X)^{-1}\rho_2).\end{aligned}$$

The semiparametric efficiency bound is hence equal to $(\mathbb{E}[\boldsymbol{e}_g'B(X)^{-1}\boldsymbol{e}_g]+\Sigma_1)^{-1}$. Therefore, the efficient influence function for π_g is equal to

$$-(\rho_1+\boldsymbol{e}_g'B(X)^{-1}L(X)'\Sigma_2(X)^{-1}\rho_2).$$

The analysis of $L(X)'\Sigma_2(X)^{-1}$ and $B(X)$ is similar to the proof of Theorem 3.1. Define the bottom right $|\mathcal{G}| \times |\mathcal{G}|$ submatrix of $\Sigma_2(X)$ as

$$\Pi \equiv \begin{pmatrix} p_2(1-p_2) & -p_3p_2 & -p_4p_2 & \cdots & -p_\infty p_2 \\ -p_3p_2 & p_3(1-p_3) & -p_4p_3 & \cdots & -p_\infty p_3 \\ \vdots & \vdots & \vdots & & \vdots \\ -p_\infty p_2 & -p_\infty p_3 & -p_\infty p_4 & \cdots & p_\infty(1-p_\infty) \end{pmatrix}.$$

We can decompose the covariance matrix $\Sigma_2(X)$ into the following block matrix:

$$\Sigma_2 = \begin{pmatrix} \Sigma_{2,UL} & \Sigma_{2,LL}' \\ \Sigma_{2,LL} & \Pi \end{pmatrix}.$$

Here $\Sigma_{2,LL}$ is a matrix with $|\mathcal{G}|$ rows. Each row of $\Sigma_{2,LL}$ is $(\sigma_{g,g',t'}(X), 2 \leqslant t' \leqslant g'-1, g' \in \mathcal{G}_{\text{trt}})$, where $\sigma_{g,g',t'}$ is defined as with

$$\begin{aligned} \sigma_{g,g',t'}(X) &\equiv \mathbb{E}\left[(G_g - p_g(X))\left(\frac{G_{g'}(Y_{t'}-Y_1)}{p_{g'}(X)} - \frac{G_\infty(Y_{t'}-Y_1)}{p_\infty(X)}\right)|X\right] \\ &= \begin{cases} -p_g(X)(m_{g',t',1}(X) - m_{\infty,t',1}(X)), & \text{if } g \notin \{g',\infty\}, \\ (1-p_g(X))m_{g,t',1}(X) + p_g(X)m_{\infty,t',1}(X), & \text{if } g = g', \\ -p_\infty(X)m_{g,t',1}(X) - (1-p_\infty(X))m_{\infty,t',1}(X), & \text{if } g = \infty. \end{cases} \end{aligned}$$

Notice that $L(X)$ is related to $\Sigma_2(X)$ as $L(X)\Pi = -(\Sigma_{2,LL}, \Pi)'$, which implies that $L(X)' = -(\Pi^{-1}\Sigma_{2,LL}, I)$. Therefore, we have

$$\begin{aligned} L(X)'\Sigma_2(X)^{-1} &= -(\mathbf{0}, \Pi^{-1}), \\ L(X)'\Sigma_2(X)^{-1}L(X) &= \Pi^{-1}, \\ B(X)^{-1} &= \Pi, \\ B(X)^{-1}L(X)'\Sigma_2(X)^{-1} &= -(\mathbf{0}, I). \end{aligned}$$

The efficient influence function of π_g is hence equal to

$$-(\rho_1 + \boldsymbol{e}_g' B(X)^{-1}L(X)'\Sigma_2(X)^{-1}\rho_2) = -(\rho_1 - \boldsymbol{e}_g'(\mathbf{0}, I)\rho_2) = G_g - \pi_g,$$

which is the same influence function in a model without any restrictions.

EIF for $ATT(g,t)$ The efficient influence function of $ATT(g,t)$ can be derived similarly. The corresponding model can be reduced to the following:

$$\begin{aligned} \mathbb{E}[p_g(X)(ATT(g,t) - CATT(g,t,X))] &= 0, \\ \mathbb{E}\left[CATT(g,t,X) - \frac{G_g(Y_t-Y_1)}{p_g(X)} + \frac{G_\infty(Y_t-Y_1)}{p_\infty(X)}\Big|X\right] &= 0, \end{aligned}$$

$$\mathbb{E}\left[\frac{G_{g'}(Y_{t'}-Y_1)}{p_{g'}(X)}-\frac{G_\infty(Y_{t'}-Y_1)}{p_\infty(X)}\Big|X\right]=0, 2\leqslant t'\leqslant g'-1, g'\in\mathcal{G}_{\text{trt}},$$
$$\mathbb{E}[G_{g'}-p_{g'}(X)|X]=0, g'\in\mathcal{G}_{\text{trt}},$$
$$\mathbb{E}[G_\infty-p_\infty(X)|X]=0.$$

After rotation, the model is equivalently represented as

$$\mathbb{E}[p_g(X)(ATT(g,t)-CATT(g,t,X))]=0,$$
$$\mathbb{E}\left[CATT(g,t,X)-\frac{G_g(Y_t-Y_1)}{p_g(X)}+\frac{G_\infty(Y_t-Y_1)}{p_\infty(X)}\Big|X\right]=0,$$
$$\mathbb{E}\left[CATT(g,t,X)-\frac{G_g(Y_t-Y_1)}{p_g(X)}+\frac{G_\infty(Y_t-Y_{t'})}{p_\infty(X)}+\frac{G_{g'}(Y_{t'}-Y_1)}{p_{g'}(X)}\Big|X\right]=0, 2\leqslant t'\leqslant g'-1, g'\in\mathcal{G}_{\text{trt}},$$
$$\mathbb{E}[G_{g'}-p_{g'}(X)|X]=0, g'\in\mathcal{G}_{\text{trt}},$$
$$\mathbb{E}[G_\infty-p_\infty(X)|X]=0.$$

Then this model is essentially the same as the one studied in the proof of Theorem 3.1. Following the same steps, we can show that the efficient influence function is obtained by optimally weighting the influence functions in $\mathbb{IF}(ATT(g,t))$. □

We introduce some definitions. Recall that $r_{g,g'}:=p_g/p_{g'}$ for any $g,g'\in\mathcal{G}$, and denote $r\equiv(r_{g,g'},g,g'\in\mathcal{G})$. Denote $\mathcal{H}_w$, $\mathcal{H}_m$, and $\mathcal{H}_p$ as the nuisance parameter spaces containing respectively the true values of w, m, and r and their estimates. For a generic $\mathcal{H}$ and norm $\|\cdot\|$, the covering number $N(\epsilon,\mathcal{H},\|\cdot\|)$ is the minimal number of N for which there exist ϵ-balls $\{\{f:\|f-h_j\|\leqslant\epsilon\},\|h_j\|<\infty,j=1,\cdots,N\}$ to cover $\mathcal{H}$.

Assumption C.1. *The following regularity assumptions are imposed for Theorem 4.1.*

(1) Second moment: Each outcome Y_t has finite second moment.

(2) Proper weighting: The estimated weights $\hat{w}^{att(g,t)}$ sum to one and is bounded in probability, i.e., $\|\hat{w}^{att(g,t)}\|_\infty=O_p(1)$.

(3) Donsker property: For each $j=w,m,r$, we have

$$\int_0^\infty\sup_Q\sqrt{\log N(\epsilon\|H\|_{L_2(Q)},\mathcal{H}_j,\|\cdot\|_{L_2(Q)})}d\epsilon<\infty,$$

where the supremum is taken over all finitely discrete probability measures Q on the support of X, H denotes an envelope of $\mathcal{H}$, and $L_2(Q)$ denotes the L_2 measure under Q.

(4) Overlap: For each $g,g'\in\mathcal{G}_{trt}\times\mathcal{G}$, $\mathbb{E}[r_{g,g'}(X)^2]<\infty$.

(5) Uniform consistency: $\|\hat{w}^{att(g,t)}-w^{att(g,t)}\|_\infty=o_p(1)$, $\|\hat{m}-m\|_\infty=o_p(1)$, and $\|\hat{r}-r\|_\infty=o_p(1)$, where $\|\cdot\|_\infty$ denotes the sup norm.

(6) Rate requirement:

$$\left\|\hat{m}_{\infty,t,t''}-m_{\infty,t,t''}\right\|_{L_2(X)}\|\hat{r}_{g,\infty}-r_{g,\infty}\|_{L_2(X)}=o_p(n^{-1/2}), t''<t,$$

$$\left\|\hat{m}_{g',t'',1} - m_{g',t'',1}\right\|_{L_2(X)} \left\|\hat{r}_{g,g'} - r_{g,g'}\right\|_{L_2(X)} = o_p(n^{-1/2}), t'' < g', g' \in \mathcal{G}_{trt},$$

where $\|\cdot\|_{L_2(X)}$ *represents the* L_2 *norm under the marginal distribution of* X*.*

A popular nonparametric class that may satisfy the Donsker condition is the smoothness class defined in Theorem 2.7.1 in van der Vaart and Wellner (1996).

Proof of Theorem 4.1. To simplify the notation in the proof, we focus on a single ATT's estimation and drop the superscript att(g,t) and subscript stg. We make explicit the dependence of θ on the nuisance parameters by writing it as $\theta(W; p, m; \pi)$. The ATT estimator is now written as $\widehat{ATT} = \mathbb{E}_n[\hat{w}(X)'\theta(W; \hat{p}, \hat{m}; \hat{\pi})]$. Define the infeasible estimator constructed using the true nuisance parameters (instead of their estimators) as

$$\widetilde{ATT} \equiv \mathbb{E}_n[w(X)'\theta(W; \hat{p}, \hat{m}; \hat{\pi})].$$

Our goal is to show that $\widehat{ATT}$ and $\widetilde{ATT}$ are first-order equivalent, i.e., $\sqrt{n}(\widehat{ATT} - \widetilde{ATT}) = o_p(1)$. Once this is established, the influence function of $\widehat{ATT}$ will be the same as that of $\widetilde{ATT}$. Since $\widetilde{ATT}$ is a ratio between two sample averages, its influence function is straightforwardly obtained by using the delta method, which is equal to the efficient influence function specified in Theorem 3.2. Then the asymptotic distribution is obtained by using the central limit theorem under the assumption that the second moment of the efficient influence functions exists. The efficiency of $\widehat{ES}(e)$ follows from another use of the delta method. In the remaining part of the proof, we focus on establishing the first-order equivalence between $\widehat{ATT}$ and $\widetilde{ATT}$.

The term $\hat{\pi}_g$ in the denominator has no impact on the asymptotic convergence of $\sqrt{n}(\widehat{ATT} - \widetilde{ATT})$ given that $\pi_g > 0$. Therefore, we treat $\hat{\pi}_g$ as one and simply write $\theta(W_i; \hat{p}, \hat{m})$ instead of $\theta(W_i; \hat{p}, \hat{m}; 1)$. The difference $\widehat{ATT} - \widetilde{ATT}$ can be decomposed as

$$\begin{aligned}
\widehat{ATT} - \widetilde{ATT} &= \frac{1}{n}\sum_{i=1}^{n} \hat{w}(X_i)'\theta(W_i; \hat{p}, \hat{m}) - w(X_i)'\theta(W_i; p, m) \\
&= \underbrace{\frac{1}{n}\sum_{i=1}^{n}(\hat{w}(X_i) - w(X_i))'\theta(W_i; p, m)}_{\equiv E_1} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} \hat{w}(X_i)'(\theta(W_i; \hat{p}, \hat{m}) - \theta(W_i; p, m))}_{\equiv E_2}.
\end{aligned}$$

Denote the generic entries of w, $\hat{w}$, and θ by using the subscript g', t''. Recall that θ is written as

$$\begin{aligned}
\theta_{g',t''}(W; \hat{p}, \hat{m}) = G_g(Y_t - Y_1 - m_{g,t,1}(X)) - \frac{p_g(X)}{p_\infty(X)} G_g(Y_t - Y_{t'} - m_{\infty,t,t''}(X)) \\
- \frac{p_g(X)}{p_{g'}(X)} G_{g'}(Y_{t''} - Y_1 - m_{g',t'',1}(X)) + G_g \underbrace{(m_{g,t,1}(X) - m_{\infty,t,t''}(X) - m_{g',t'',1}(X))}_{=CATT(g,t,X)}.
\end{aligned}$$

Notice that the first and last terms on the right-hand side, $G_g(Y_t - Y_1 - m_{g,t,1}(X))$ and $G_g CATT(g, t, X)$, do not depend on g' or t''. These two terms will not contribute to E_1 since both $\hat{w}$ and w sum to one, and hence

$$(\hat{w}(X_i) - w(X_i))'\mathbf{1}(G_{g,i} CATT(g, t, X_i)) = 0,$$
$$(\hat{w}(X_i) - w(X_i))'\mathbf{1}(G_{g,i}(Y_{i,t} - Y_{i,1} - m_{g,t,1}(X_i))) = 0.$$

Therefore, the term E_1 is equal to the sum of the following two terms

$$\frac{1}{n}\sum_{g',t''}\sum_{i=1}^{n}(\hat{w}_{g',t''}(X_i)-w_{g',t''}(X_i))'\frac{p_g(X_i)}{p_\infty(X_i)}G_{g,i}(Y_{i,t}-Y_{i,t''}-m_{\infty,t,t''}(X_i)),$$
$$\frac{1}{n}\sum_{g',t''}\sum_{i=1}^{n}(\hat{w}_{g',t''}(X_i)-w_{g',t''}(X_i))'\frac{p_g(X_i)}{p_{g'}(X_i)}G_{g',i}(Y_{i,t''}-Y_{i,1}-m_{g',t'',1}(X_i)).$$

Given that the convergence rate of the two terms can be derived using the same method, we choose to illustrate the convergence of the second term. Additionally, since the summation over g' and t'' is finite, it suffices to analyze the convergence of a single term within the summation:

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{w}_{g',t''}(X_i)-w_{g',t''}(X_i))'\frac{p_g(X_i)}{p_{g'}(X_i)}G_{g',i}(Y_{i,t''}-Y_{i,1}-m_{g',t'',1}(X_i)).$$

By the uniform consistency of $\hat{w}$, the above term multiplied by $\sqrt{n}$ is bounded by

$$\begin{aligned}&\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{w}_{g',t''}(X_i)-w_{g',t''}(X_i))'\frac{p_g(X_i)}{p_{g'}(X_i)}G_{g',i}(Y_{i,t''}-Y_{i,1}-m_{g',t'',1}(X_i))\right|\\ \leqslant&\sup_{\substack{\tilde{w}_{g',t''}\in\mathcal{C}^{\alpha}(\mathcal{X})_M:\\ \|\tilde{w}_{g',t''}-w_{g',t''}\|_\infty<\delta_n}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{w}_{g',t''}(X_i)-w_{g',t''}(X_i))'\frac{p_g(X_i)}{p_{g'}(X_i)}G_{g',i}(Y_{i,t''}-Y_{i,1}-m_{g',t'',1}(X_i))\right|+o_p(1),\end{aligned}$$

where $\delta_n\downarrow 0$ denotes a sequence that converges to zero slower than the uniform convergence rate of $\hat{w}$. The first term on the right-hand side is the standard stochastic equicontinuity term, which is of order $o_p(1)$ because of the Donsker condition, Theorem 2.5.2 in van der Vaart and Wellner (1996), and that $\frac{p_g(X)}{p_{g'}(X)}G_{g'}(Y_{t''}-Y_1-m_{g',t'',1}(X))$ is a fixed function with finite second moment by assumption. Following the same procedure, we can show that

$$\frac{1}{n}\sum_{g',t'}\sum_{i=1}^{n}(\hat{w}_{g',t''}(X_i)-w_{g',t''}(X_i))'\frac{p_g(X_i)}{p_\infty(X_i)}G_{g,i}(Y_{i,t}-Y_{i,t''}-m_{\infty,t,t''}(X_i))=o_p(n^{-1/2}),$$

Therefore, we have shown that $E_1=o_p(n^{-1/2})$. For the term E_2, define θ^1 and θ^2 as vectors respectively collecting the following term:

$$\theta^1_{g',t''}(W;p,m)\equiv\frac{G_gp_\infty(X)-G_\infty p_g(X)}{\pi_gp_\infty(X)}(Y_t-Y_{t'}-m_{\infty,t,t''}(X)),$$
$$\theta^2_{g',t''}(W;p,m)\equiv\frac{G_gp_{g'}(X)-G_{g'}p_g(X)}{\pi_gp_{g'}(X)}(Y_{t'}-Y_1-m_{g',t'',1}(X)).$$

We can decompose E_2 as

$$E_2=\underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat{w}(X_i)'(\theta^1(W_i;\hat{p},\hat{m})-\theta^1(W_i;p,m))}_{\equiv E_{2.1}}+\underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat{w}(X_i)'(\theta^2(W_i;\hat{p},\hat{m})-\theta^2(W_i;p,m))}_{\equiv E_{2.2}}.$$

The two terms can be analyzed analogously. We examine $E_{2.1}$ first. This term can be decomposed into three terms

$$E_{2.1.1} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{g',t''} \hat{w}_{g',t''}(X_i)\left(r_{g,\infty}(X_i) - \hat{r}_{g,\infty}(X_i)\right) G_{\infty,i}(Y_{i,t} - Y_{i,t''} - m_{\infty,t,t''}(X_i)),$$

$$E_{2.1.2} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{g',t''} \hat{w}_{g',t''}(X_i)\left(m_{\infty,t,t''}(X_i) - \hat{m}_{\infty,t,t''}(X_i)\right)\left(G_{g,i} - \frac{p_g(X_i)}{p_\infty(X_i)}G_{\infty,i}\right),$$

$$E_{2.1.3} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{g',t'} \hat{w}_{g',t'}(X_i) G_{\infty,i}\left(m_{\infty,t,t'}(X_i) - \hat{m}_{\infty,t,t'}(X_i)\right)\left(r_{g,\infty}(X_i) - \hat{r}_{g,\infty}(X_i)\right).$$

The convergence rate of $E_{2.1.1}$ and $E_{2.1.2}$ can be derived in the same way as that of E_1 under the smoothness and uniform consistency conditions. For the last term $E_{2.1.3}$, examine a single term in the summation over j:

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n} \hat{w}_{g',t''}(X_i) G_{\infty,i}\left(m_{\infty,t,t''}(X_i) - \hat{m}_{\infty,t,t''}(X_i)\right)\left(r_{g,\infty}(X_i) - \hat{r}_{g,\infty}(X_i)\right)\right| \\ \leqslant & O_p(1) \times \frac{1}{n}\sum_{i=1}^{n} \left|m_{\infty,t,t''}(X_i) - \hat{m}_{\infty,t,t''}(X_i)\right|\left|r_{g,\infty}(X_i) - \hat{r}_{g,\infty}(X_i)\right|\end{aligned}$$

because the estimated weights are bounded in probability. The second factor on the right-hand side is bounded as

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} \left|m_{\infty,t,t''}(X_i) - \hat{m}_{\infty,t,t''}(X_i)\right|\left|r_{g,\infty}(X_i) - \hat{r}_{g,\infty}(X_i)\right| \\ \leqslant & \sup_{(\tilde{m}_{\infty,t,t''},\tilde{r}_{g,\infty})\in\mathcal{F}_n} \frac{1}{n}\sum_{i=1}^{n}\Bigg(\left|m_{\infty,t,t''}(X_i) - \tilde{m}_{\infty,t,t''}(X_i)\right|\left|r_{g,\infty}(X_i) - \tilde{r}_{g,\infty}(X_i)\right| \\ &\quad - \mathbb{E}\left|m_{\infty,t,t''}(X) - \tilde{m}_{\infty,t,t''}(X)\right|\left|r_{g,\infty}(X) - \tilde{r}_{g,\infty}(X)\right|\Bigg) \\ &\quad + \sup_{(\tilde{m}_{\infty,t,t''},\tilde{r}_{g,\infty})\in\mathcal{F}_n} \mathbb{E}\left[\left|m_{\infty,t,t''}(X) - \tilde{m}_{\infty,t,t''}(X)\right|\left|r_{g,\infty}(X) - \tilde{r}_{g,\infty}(X)\right|\right],\end{aligned}$$

where $\mathcal{F}_n := \{\tilde{m}_{\infty,t,t''}, \tilde{r}_{g,\infty} \in \mathcal{C}^\alpha(\mathcal{X})_M : \|m_{\infty,t,t''} - \tilde{m}_{\infty,t,t''}\|_\infty \leqslant \delta_n, \|m_{\infty,t,t''} - \tilde{m}_{\infty,t,t''}\|_{L_2(X)} \leqslant \|m_{\infty,t,t''} - \hat{m}_{\infty,t,t''}\|_{L_2(X)}, \|r_{g,\infty} - \tilde{r}_{g,\infty}\|_\infty \leqslant \delta_n, \|r_{g,\infty} - \tilde{r}_{g,\infty}\|_{L_2(X)} \leqslant \|r_{g,\infty} - \hat{r}_{g,\infty}\|_{L_2(X)}\}$, with $\delta_n \downarrow 0$ being a sequence that converges to zero slower than the uniform convergence rates of $\hat{m}_{\infty,t,t''}$ and $\hat{r}_{g,\infty}$. By the fact that the finiteness of entropy integral is preserved under element-wise multiplication of function classes, we can use the previous stochastic equicontinuity argument to show that the first term on the right-hand side is of order $o_p(n^{-1/2})$. The second term on the right-hand side is also of order $o_p(n^{-1/2})$ by using Cauchy-Schwarz inequality together with the rate requirement on the nuisance estimators:

$$\begin{aligned}&\sup_{(\tilde{m}_{\infty,t,t''},\tilde{r}_{g,\infty})\in\mathcal{F}_n} \mathbb{E}\left[\left|m_{\infty,t,t''}(X) - \tilde{m}_{\infty,t,t''}(X)\right|\left|r_{g,\infty(X)} - \tilde{r}_{g,\infty}(X)\right|\right] \\ \leqslant & \sup_{(\tilde{m}_{\infty,t,t''},\tilde{r}_{g,\infty})\in\mathcal{F}_n} \|m_{\infty,t,t''} - \tilde{m}_{\infty,t,t''}\|_{L_2(X)}\|r_{g,\infty} - \tilde{r}_{g,\infty}\|_{L_2(X)} = o_p(n^{-1/2}).\end{aligned}$$

This proves that $E_{2.1} = o_p(n^{-1/2})$. The term $E_{2.2}$ can also be decomposed in a similar way as $E_{2.1}$:

$$E_{2.2.1} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{g',t''} \hat{w}_{g',t''}(X_i)\left(r_{g,g'}(X_i) - \hat{r}_{g,g'}(X_i)\right)G_{g',i}(Y_{i,t''} - Y_{i,1} - m_{g',t'',1}(X_i)),$$

$$E_{2.2.2} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{g',t''} \hat{w}_{g',t''}(X_i)\left(m_{g',t'',1}(X_i) - \hat{m}_{g',t'',1}(X_i)\right)\left(G_{g,i} - \frac{p_g(X_i)}{p_{g'}(X_i)}G_{g',i}\right),$$

$$E_{2.2.3} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{g',t''} \hat{w}_{g',t''}(X_i)G_{g',i}\left(m_{g',t'',1}(X_i) - \hat{m}_{g',t'',1}(X_i)\right)\left(r_{g,g'}(X_i) - \hat{r}_{g,g'}(X_i)\right).$$

The convergence rate of $E_{2.2.1}$ and $E_{2.2.2}$ can be derived in the same way as for $E_{2.1.1}$, $E_{2.1.2}$, and E_1. The term $E_{2.2.3}$ is similar to $E_{2.1.3}$. This proves that $E_{2.1} = o_p(n^{-1/2})$. In the end, we have shown that $\widehat{ATT} - \widetilde{ATT} = o_p(n^{-1/2})$, and therefore, $\widehat{ATT}(g,t)$ and $\widehat{ES}(e)$ have the desired asymptotic distributions.

□

The following lemma shows the consistency of the estimator based on the $\hat{K}$ in (4.2), i.e., asymptotically all basis functions are selected. Let $\ell(W,r) = r(X)^2 G_{g'} - 2r(X)G_g$ and $\bar{\ell}(r) = \mathbb{E}[\ell(W,r)]$. Denote the sieve space of r as $\mathcal{R}_K$. For example, in (4.2), $\mathcal{R}_K$ is the space spanned by $\Psi^K(X)$.

Lemma C.1. *Let the following conditions hold.*

(1) There is a sequence $c(K)$ such that $\inf_{\tilde{r}\in\mathcal{R}_K}[\bar{\ell}(\tilde{r}) - \bar{\ell}(r)] > c(K) > 0$ for all K.

(2) $\sup_{\tilde{r}\in\mathcal{R}_K} |\frac{1}{n}\sum_{i=1}^{n}(\ell(W_i,\tilde{r}) - \bar{\ell}(\tilde{r}))| = o_p(1)$.

(3) There is a sequence $K_n^* \to \infty$ such that $\|\hat{r}_{K_n^*} - r\|_\infty = o_p(1)$.

(4) For any sequence $\eta_n = o(1)$, $\sup_{\|\tilde{r}-r\|_\infty \leqslant \eta_n} |\frac{1}{n}\sum_{i=1}^{n}(\ell(W_i,\tilde{r}) - \bar{\ell}(r))| = o_p(1)$.

(5) $C_n K_n^*/n = o(1)$.

Then for any fixed K^*, we have $\hat{K} > K^*$ with probability approaching one.

Proof of Lemma C.1. For simplicity, we suppress the subscripts g and g' and denote the estimand and its estimator as r and $\hat{r}_K$, where the index K signifies the sieve dimension. Define the objective function in (4.2) as $I_n(K)$. For any fixed K^*, using conditions (2) and (5), we have

$$\begin{aligned} I_n(K^*) &= 2\bar{\ell}(\hat{r}_{K^*}) + \frac{2}{n}\sum_{i=1}^{n}(\ell(W_i,\hat{r}_{K^*}) - \bar{\ell}(\hat{r}_{K^*})) + C_n K^*/n \\ &= 2\bar{\ell}(\hat{r}_{K^*}) + o_p(1). \end{aligned}$$

On the other hand, using conditions (2)-(4) for the sequence K_n^*, we have

$$\begin{aligned} I_n(K_n^*) &= 2\bar{\ell}(r) + \frac{2}{n}\sum_{i=1}^{n}(\ell(W_i,\hat{r}_{K_n^*}) - \bar{\ell}(r)) + C_n K^*/n \\ &= 2\bar{\ell}(r) + o_p(1). \end{aligned}$$

Combining the above two results, we have

$$I_n(K^*) - I_n(K_n^*) = 2(\bar{\ell}(\hat{r}_{K^*}) - \bar{\ell}(r)) + o_p(1),$$

which, together with condition (1), implies that $I_n(K^*) - I_n(K_n^*) > 0$ with probability approaching one. This completes the proof. $\square$

Proof of Lemma B.1. Define the conditional LATT parameter as

$$CLATT(g, t, X) \equiv \mathbb{E}[Y_t(1) - Y_t(0) | G^{\text{IV}} = g, D_t(g) > D_t(\infty), X].$$

We first want to establish that

$$CLATT(g, t, X) = \frac{\mathbb{E}[Y_t - Y_{g-1} | G^{\text{IV}} = g, X] - \mathbb{E}[Y_t - Y_{g-1} | G^{\text{IV}} = \infty, X]}{\mathbb{E}[D_t - D_{g-1} | G^{\text{IV}} = g, X] - \mathbb{E}[D_t - D_{g-1} | G^{\text{IV}} = \infty, X]}.$$

For the numerator, notice that

$$\begin{aligned}
&\mathbb{E}[Y_t - Y_{g-1} | G^{\text{IV}} = g, X] - \mathbb{E}[Y_t - Y_{g-1} | G^{\text{IV}} = \infty, X] \\
=&\mathbb{E}[Y_t(D_t(g)) - Y_{g-1}(D_t(g)) | G^{\text{IV}} = g, X] - \mathbb{E}[Y_t(D_t(\infty)) - Y_{g-1}(D_t(\infty)) | G^{\text{IV}} = \infty, X] \\
=&\mathbb{E}[Y_t(D_t(g)) - Y_t(D_t(\infty)) | G^{\text{IV}} = g, X] \\
&+ \mathbb{E}[Y_t(D_t(\infty)) - Y_{g-1}(D_t(g))] - \mathbb{E}[Y_t(D_t(\infty)) - Y_{g-1}(D_t(\infty)) | G^{\text{IV}} = \infty, X] \\
=&\mathbb{E}[Y_t(D_t(g)) - Y_t(D_t(\infty)) | G^{\text{IV}} = g, X],
\end{aligned}$$

where the last equality follows from no anticipation and parallel trends in the outcome. For the right-hand side, notice that

$$\begin{aligned}
\mathbb{E}[Y_t(D_t(g)) | G^{\text{IV}} = g, X] &= \mathbb{E}[D_t(g)(Y_t(1) - Y_t(0)) | G^{\text{IV}} = g, X] + \mathbb{E}[Y_t(0) | G^{\text{IV}} = g, X], \\
\mathbb{E}[Y_t(D_t(\infty)) | G^{\text{IV}} = g, X] &= \mathbb{E}[D_t(\infty)(Y_t(1) - Y_t(0)) | G^{\text{IV}} = g, X] + \mathbb{E}[Y_t(0) | G^{\text{IV}} = g, X],
\end{aligned}$$

and therefore,

$$\begin{aligned}
\mathbb{E}[Y_t(D_t(g)) - Y_t(D_t(\infty)) | G^{\text{IV}} = g, X] &= \mathbb{E}[(D_t(g) - D_t(\infty))(Y_t(1) - Y_t(0)) | G^{\text{IV}} = g, X] \\
&= \mathbb{E}[Y_t(1) - Y_t(0) | G^{\text{IV}} = g, D_t(g) - D_t(\infty) = 1, X] \\
&\times \mathbb{P}(D_t(g) - D_t(\infty) = 1 | G^{\text{IV}} = g, X) \\
&= CLATT(g, t, X) \mathbb{P}(D_t(g) - D_t(\infty) = 1 | G^{\text{IV}} = g, X),
\end{aligned}$$

where the second line follows from monotonicity. Similarly, we can show that

$$\begin{aligned}
\mathbb{E}[D_t - D_{g-1} | G^{\text{IV}} = g, X] - \mathbb{E}[D_t - D_{g-1} | G^{\text{IV}} = \infty, X] &= \mathbb{E}[D_t(g) - D_t(\infty) | G^{\text{IV}} = g, X] \\
&= \mathbb{P}(D_t(g) - D_t(\infty) = 1 | G^{\text{IV}} = g, X).
\end{aligned}$$

This gives the identification of CLATT. To identify the unconditional LATT, we need to invoke the Bayes rule.

$$
\begin{aligned}
&LATT(g,t)\\
=&\mathbb{E}[CLATT(g,t,X)|G^{\text{IV}}=g,D_t(g)>D_t(\infty)]\\
=&\int\frac{\mathbb{E}[Y_t-Y_{g-1}|G^{\text{IV}}=g,x]-\mathbb{E}[Y_t-Y_{g-1}|G^{\text{IV}}=\infty,x]}{\mathbb{E}[D_t-D_{g-1}|G^{\text{IV}}=g,x]-\mathbb{E}[D_t-D_{g-1}|G^{\text{IV}}=\infty,x]}dF_{X|G^{\text{IV}}=g,D_t(g)>D_t(\infty)}(x)\\
=&\int\frac{\mathbb{E}[Y_t-Y_{g-1}|G^{\text{IV}}=g,x]-\mathbb{E}[Y_t-Y_{g-1}|G^{\text{IV}}=\infty,x]}{\mathbb{E}[D_t-D_{g-1}|G^{\text{IV}}=g,x]-\mathbb{E}[D_t-D_{g-1}|G^{\text{IV}}=\infty,x]}\\
&\times\frac{\mathbb{P}(D_t(g)>D_t(\infty)|G^{\text{IV}}=g,X=x)\mathbb{P}(G^{\text{IV}}=g|X=x)dF_X(x)}{\mathbb{P}(G^{\text{IV}}=g,D_t(g)>D_t(\infty))}\\
=&\frac{\mathbb{E}\big[(\mathbb{E}[Y_t-Y_{g-1}|G^{\text{IV}}=g,X]-\mathbb{E}[Y_t-Y_{g-1}|G^{\text{IV}}=\infty,X])G_g^{\text{IV}}\big]}{\mathbb{P}(G^{\text{IV}}=g,D_t(g)>D_t(\infty))}.
\end{aligned}
$$

Lastly, the denominator on the right-hand side can be written as

$$
\begin{aligned}
\mathbb{P}(G^{\text{IV}}=g,D_t(g)>D_t(\infty))&=\mathbb{E}[\mathbb{E}[D_t(g)-D_t(\infty)|G^{\text{IV}}=g,X]\mathbb{P}(G^{\text{IV}}=g|X)]\\
&=\mathbb{E}\big[(\mathbb{E}[D_t-D_{g-1}|G^{\text{IV}}=g,X]-\mathbb{E}[D_t-D_{g-1}|G^{\text{IV}}=\infty,X])G_g^{\text{IV}}\big].
\end{aligned}
$$

This completes the proof. □

Proof of Lemma B.2. Similar to the proof of Lemmas 3.1 and 3.2, we want to construct a joint distribution of the potential outcomes and treatments that is consistent with the observed outcomes and treatments and satisfies the identification assumptions. Recall the notations Y and ΔY for the observed outcomes as defined in the proof of Lemmas 3.1 and 3.2. Similarly, define D and ΔD for the treatment. In the same way, define the vectors of potential variables $Y(g)$, $\Delta Y(g)$, $D(g)$, and $\Delta D(g)$, where the potential outcome $Y_t(g)$ is now shorthand for $Y_t(D_t(g))$. Let (D,Y,G^{IV}) denote the observed variables that already satisfy random sampling, overlap, and monotonicity. For each given g, we construct the potential outcomes and treatments as follows:

$$
\begin{aligned}
&\text{the vectors } (D(g'),Y(g')), g'\in\mathcal{G}^{\text{IV}} \text{ are jointly independent conditional on } G^{\text{IV}}=g,\\
&\text{for } g'=g\text{: } (D(g),Y(g))|\{G=g\}\overset{d}{=}(D,Y)|\{G=g\}, \text{ and}\\
&\text{for } g'\neq g\text{: } (D_1(g'),Y_1(g'))|\{G=g\}\overset{d}{=}(D_1(g),Y_1(g))|\{G=g\},\\
&\quad(\Delta D(g'),\Delta Y(g'))|\{G=g\}\overset{d}{=}(\Delta D,\Delta Y)|\{G=\infty\},(D_1(g'),Y_1(g'))\perp(\Delta D(g'),\Delta Y(g'))|\{G=g\}.
\end{aligned}
$$

This induces the observed outcomes and treatments through constructions. The no-anticipation and parallel trends conditions can be verified similarly to the proof of Lemma 3.2.

□

Proof of Corollary B.1. It is well-known that shape restrictions such as monotonicity do not affect the calculation of the semiparametric efficiency bound. Thus, we can focus on the moment equalities to derive the bound. Similar to the proof of Theorems 3.2 and 3.1, we can separately derive the efficient influence function for $LATT(g,t)_{num}$ and $LATT(g,t)_{den}$ and then combine them to obtain the efficient influence function for $LATT(g,t)_{num}/LATT(g,t)_{den}$. For a single $LATT(g,t)_{num}$, the model becomes the following

$$
\begin{aligned}
\mathbb{E}[G_g^{\mathrm{IV}}(LATT(g,t)_{num} - h_1(g,t,X))] &= 0, \\
\mathbb{E}\left[h_1(g,t,X) - \frac{G_g^{\mathrm{IV}}(Y_t - Y_{g-1})}{p_g^{\mathrm{IV}}(X)} + \frac{G_\infty^{\mathrm{IV}}(Y_t - Y_{g-1})}{p_\infty^{\mathrm{IV}}(X)}\Big| X\right] &= 0, \\
\mathbb{E}\left[\frac{G_g^{\mathrm{IV}}(Y_{t'} - Y_1)}{p_g^{\mathrm{IV}}(X)} - \frac{G_\infty^{\mathrm{IV}}(Y_{t'} - Y_1)}{p_\infty^{\mathrm{IV}}(X)}\Big| X\right] &= 0, \text{for all } 2 \leqslant t' \leqslant g-1, \\
\mathbb{E}\left[\frac{G_g^{\mathrm{IV}}(D_{t'} - D_1)}{p_g^{\mathrm{IV}}(X)} - \frac{G_\infty^{\mathrm{IV}}(D_{t'} - D_1)}{p_\infty^{\mathrm{IV}}(X)}\Big| X\right] &= 0, \text{for all } 2 \leqslant t' \leqslant g-1, \\
\mathbb{E}[G_g^{\mathrm{IV}} - p_g^{\mathrm{IV}}(X)|X] &= 0.
\end{aligned}
$$

We can rotate the conditional moment restrictions as in the proof of Theorem 3.1 and obtain that

$$
\begin{aligned}
\mathbb{E}\left[h_1(g,t,X) - \frac{G_g^{\mathrm{IV}}(Y_t - Y_1)}{p_g^{\mathrm{IV}}(X)} + \frac{G_\infty^{\mathrm{IV}}(Y_t - Y_1)}{p_\infty^{\mathrm{IV}}(X)}\Big| X\right] &= 0, \\
&\cdots \\
\mathbb{E}\left[h_1(g,t,X) - \frac{G_g^{\mathrm{IV}}(Y_t - Y_{g-1})}{p_g^{\mathrm{IV}}(X)} + \frac{G_\infty^{\mathrm{IV}}(Y_t - Y_{g-1})}{p_\infty^{\mathrm{IV}}(X)}\Big| X\right] &= 0, \\
\mathbb{E}\left[h_1(g,t,X) - \frac{G_g^{\mathrm{IV}}(Y_t - Y_1 + D_2 - D_1)}{p_g^{\mathrm{IV}}(X)} + \frac{G_\infty^{\mathrm{IV}}(Y_t - Y_1 + D_2 - D_1)}{p_\infty^{\mathrm{IV}}(X)}\Big| X\right] &= 0, \\
&\cdots \\
\mathbb{E}\left[h_1(g,t,X) - \frac{G_g^{\mathrm{IV}}(Y_t - Y_1 + D_{g-1} - D_1)}{p_g^{\mathrm{IV}}(X)} + \frac{G_\infty^{\mathrm{IV}}(Y_t - Y_1 + D_{g-1} - D_1)}{p_\infty^{\mathrm{IV}}(X)}\Big| X\right] &= 0, \\
\mathbb{E}[G_g^{\mathrm{IV}} - p_g^{\mathrm{IV}}(X)|X] &= 0.
\end{aligned}
$$

The structure of this set of moment restrictions aligns with that in Theorem 3.1. The term $Y_1 - (D_{t'} - D_1)$ can be treated as an outcome in a baseline period, enabling the derivation of the efficient influence function for $LATT(g,t)_{num}$ as $\mathbb{EIF}^{latt(g,t),num}/\pi_g^{\mathrm{IV}}$, where $\pi_g^{\mathrm{IV}} \equiv \mathbb{P}(G^{\mathrm{IV}} = g)$. Similarly, the efficient influence function for $LATT(g,t)_{den}$ is obtained as $\mathbb{EIF}^{latt(g,t),den}/\pi_g^{\mathrm{IV}}$. The efficient influence function for the LATT parameter then follows from the application of the Delta method. □

Proof of Theorem A.1. This is the well-known Hausman test, in which $\mathrm{aCov}(\widehat{ES} - \widetilde{ES})$ can be consistently estimated by $\widehat{\mathrm{aCov}}(\widetilde{ES}) - \widehat{\mathrm{aCov}}(\widehat{ES})$ because $\widehat{ES}$ achieves the semiparametric efficiency bound. The fact that this Hausman test has nontrivial power against all local alternatives is based on Theorem 3.3 and Remark 3.4 in Chen and Santos (2018). □

References

Ai, Chunrong and Xiaohong Chen, "The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions," *Journal of Econometrics*, 2012, *170* (2), 442–457.

Armstrong, Timothy B., Patrick Kline, and Liyang Sun, "Adapting to Misspecification," *arXiv:2305.14265v4*, 2024.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Chen, Xiaohong and Andres Santos, "Overidentification in Regular Models," *Econometrica*, 2018, *86* (5).

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Holm, Sture, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 1979, *6* (2), 65–70.

Miyaji, Sho, "Instrumented Difference-in-Differences with Heterogeneous Treatment Effects," *arXiv:2405.12083*, 2024.

Roth, Jonathan, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *4* (3), 305–322.

van der Vaart, Aad W. and Jon A. Wellner, *Weak Convergence and Empirical Processes*, New York: Springer, 1996.