When should pre-trends be parallel?*

By Dalia Ghanem, Pedro H. C. Sant'Anna, and Kaspar Wüthrich†

In difference-in-differences (DiD) applications, it is common practice to test whether the trends of the treated and control units in the pre-treatment period (the "pre-trends") are parallel. Parallel pre-trends are then typically interpreted as supportive evidence for the plausibility of the parallel (post-)trends assumptions required for identifying causal effects using DiD methods.[1] Our goal is to scrutinize this common practice through the lens of how units select into treatment and to provide a better understanding of which selection mechanisms lead to pre-trends tests being informative about the validity of DiD.

To analyze pre-trends tests, it is helpful to distinguish four scenarios: (a) pre- and post-trends are parallel, (b) pre-trends are not parallel but post-trends are parallel, (c) pre-trends are parallel but post-trends are not, (d) pre- and post-trends are not parallel. In this paper, we provide necessary and sufficient conditions for scenario (a) under different restrictions on selection into treatment in the absence of structural breaks, building on the necessary and sufficient conditions for parallel post-trends in Ghanem, Sant'Anna and Wüthrich (2025). We focus on scenario (a) because this is when pre-trends tests are informative and interpreted as evidence in favor of DiD. Our necessary and sufficient conditions theoretically characterize when pre-trends tests can be informative, depending on how exactly units select into the treatment.

Our main findings are as follows. First, when selection is solely based on time-invariant unobservables ("fixed effects"), then parallel pre-trends and parallel post-trends holding jointly is equivalent to a time-homogeneity assumption on the mean of the time-varying unobservables conditional on the fixed effects, which is consistent with classical strict exogeneity assumptions. Thus, when selection is based on fixed effects, the pre-trends and post-trends are parallel under standard time series restrictions; implying that pre-trends tests can be informative.

Second, when selection is based on pre-treatment information, then pre-trends and post-trends being parallel implies that there are no idiosyncratic shocks in the pre-treatment period and a martingale restriction on the idiosyncratic shocks in the post-treatment period. In this case, the pre-trends may not be parallel in realistic scenarios, whereas the post-trends are parallel under classical time series restrictions. Consequently, pre-trends tests can lead to DiD designs being wrongly discarded.

Third, there are additional issues with pre-trends tests in settings with time-varying covariates. Our necessary and sufficient conditions imply that, when post-treatment potential outcomes depend on post-treatment covariates, controlling only for pre-treatment values of the covariates can lead to violations of parallel post-trends in settings where the pre-trends are parallel and the covariates are exogenous. Thus, pre-trends tests can be misleading in this case. Controlling for the entire time series of covariates can mitigate this issue when the covariates are exogenous.

Throughout this paper, we focus on the population identification problem. The issues we document are therefore distinct from the statistical issues arising with pre-trends tests (e.g., Roth, 2022).

I. Setup

We consider a DiD setting in which n units, indexed by $i = 1, \ldots, n$, are observed for three periods, indexed by $t \in \{-1, 0, 1\}$. Let Y_{it} denote the observed outcome of unit i in period t. In periods $t \in \{-1, 0\}$, no unit is treated. In period $t = 1$, some units select into the treatment, while others remain untreated. Let $G_i = 1$ if unit i is treated in period $t = 1$ and $G_i = 0$ if unit i is untreated in

* We are grateful to Andrew Goodman-Bacon for detailed comments. The usual disclaimer applies.

† Ghanem: University of California, Davis (dghanem@ucdavis.edu); Sant'Anna: Emory University (pedro.santanna@emory.edu); Wüthrich: University of Michigan (kasparwu@umich.edu).

[1] We will often refer to "parallel trends" as "parallel post-trends" to distinguish parallel post-trends from parallel pre-trends.

$t = 1$. We abstract from covariates for now and discuss additional issues with covariates in Section III. Throughout the paper, we focus on non-degenerate DiD designs (i.e., designs with $P(G_i = 1) \in (0,1)$). We focus on three-period, two-group DiD designs because this is the simplest setting for analyzing pre-trends tests. However, the basic issues we document are also relevant in the more general DiD designs that are common in empirical practice.

We write the potential outcomes with and without the treatment as $Y_{it}(1)$ and $Y_{it}(0)$, respectively, assuming no anticipatory effects (e.g., Roth et al., 2023). For $i \in \{1,\dots,n\}$, the observed outcomes are $Y_{it} = Y_{it}(0)$ for $t \in \{-1,0\}$ and $Y_{i1} = Y_{i1}(1)G_i + Y_{i1}(0)(1-G_i)$.

The parameter of interest is the average treatment effect on the treated (ATT) in period $t = 1$, $E[Y_{i1}(1) - Y_{i1}(0)|G_i = 1]$. To identify the ATT, DiD methods rely on the following parallel post-trends assumption.

ASSUMPTION PT: $E[Y_{i1}(0) - Y_{i0}(0)|G_i = 1] = E[Y_{i1}(0) - Y_{i0}(0)|G_i = 0]$.

Assumption PT is fundamentally untestable. Therefore, to assess its validity, researchers routinely assess the validity of the following pre-treatment analog of Assumption PT.

ASSUMPTION PPT: $E[Y_{i0}(0) - Y_{i(-1)}(0)|G_i = 1] = E[Y_{i0}(0) - Y_{i(-1)}(0)|G_i = 0]$.

Unlike Assumption PT, Assumption PPT is directly testable since it implies that

$$E[Y_{i0} - Y_{i(-1)}|G_i = 1] = E[Y_{i0} - Y_{i(-1)}|G_i = 0].$$

We assume that $Y_{it}(0)$ is generated by a standard two-way fixed effects model,

$$Y_{it}(0) = \alpha_i + \lambda_t + \varepsilon_{it}, \quad E[\varepsilon_{it}] = 0, \quad i \in \{1,\dots,n\}, \quad t \in \{-1,0,1\}. \tag{1}$$

Model (1) encompasses the idea that the underlying structure remains the same in the pre- and the post-treatment period, a necessary (but not sufficient) condition for pre-trends tests to be informative. Therefore, our analysis should be viewed as a "best-case" analysis of pre-trends tests. If there are unrestricted structural breaks, then pre-trends tests are inherently uninformative.

Following Ghanem, Sant'Anna and Wüthrich (2025), we allow the units' selection decision to depend on the unobservable determinants of $\{Y_{it}(0)\}$ as well as additional time-invariant and time-varying unobservables, $(\mu_i, \eta_{i(-1)}, \eta_{i0}, \eta_{i1})$. We consider the following selection mechanism,

$$G_i = g(\omega_i, \nu_i), \quad i \in \{1,\dots,n\}, \tag{2}$$

where ω_i is a subvector of $(\alpha_i, \varepsilon_{i(-1)}, \varepsilon_{i0}, \varepsilon_{i1}, \mu_i, \eta_{i(-1)}, \eta_{i0}, \eta_{i1})$ and ν_i is a selection-specific error term that is independent of all other unobservables. Specifically, we impose the following assumption, which is Assumption SEL in Ghanem, Sant'Anna and Wüthrich (2025), adapted to our setup with an additional pre-treatment period. For a random variable A_{it}, let $A_i^t = (A_{i(-1)},\dots,A_{it})$.

ASSUMPTION SEL: *The error term ν_i has a non-degenerate distribution, $P(\nu_i > v) \in (0,1)$ for some $v \in \mathbb{R}$, and is independent of $(\alpha_i, \varepsilon_i^1, \mu_i, \eta_i^1)$.*

We denote by $\mathscr{G}_\omega$ the set of all selection mechanisms $g : \mathrm{supp}(\omega_i, \nu_i) \mapsto \{0,1\}$.

II. Necessary and sufficient conditions for parallel pre-trends and parallel post-trends

Here, we derive necessary and sufficient conditions for Assumptions PPT and PT to hold jointly under different assumptions on selection, building on Ghanem, Sant'Anna and Wüthrich (2025) who provide necessary and sufficient conditions for Assumption PT. We consider two leading cases that differ with respect to what the units select on (i.e., ω_i). To focus on nonparametric and robust instances of Assumptions PPT and PT, we provide necessary and sufficient conditions for Assumptions PPT and PT to hold jointly for all $g \in \mathscr{G}_\omega$. We refer to Ghanem, Sant'Anna and Wüthrich (2025) for a discussion of why such conditions are relevant in practice.

A. Selection on time-invariant unobservables

We start by discussing the classical case where selection is only based on time-invariant unobservables ("fixed effects"), so that $\omega_i = (\alpha_i, \mu_i)$. In the context of studying the effect of job training programs on earnings, fixed effects encompass the permanent component of earnings (e.g., Ashenfelter and Card, 1985), whereas when evaluating state-level policies, such as the Medicaid expansion, fixed effects would capture state-specific time-invariant factors. The following proposition provides necessary and sufficient conditions for Assumptions PPT and PT to hold jointly in this case. We say that a selection mechanism g is *nondegenerate* if $P(g(\omega_i, \nu_i) = 1) \in (0,1)$.

PROPOSITION 1: *Suppose that $\omega_i = (\alpha_i, \mu_i)$, Assumption SEL and Assumption REG in Appendix A hold, and the model is* (1). *Then, Assumptions PPT and PT hold jointly for all nondegenerate $g \in \mathscr{G}_\omega$ if and only if $E[\varepsilon_{i1}|\alpha_i, \mu_i] = E[\varepsilon_{i0}|\alpha_i, \mu_i] = E[\varepsilon_{i(-1)}|\alpha_i, \mu_i]$.*[2]

The proof of Proposition 1 follows from an application of Lemma 1 in Appendix A, which gives necessary and sufficient conditions for any choice of ω_i. The "if" direction follows from the law of iterated expectations (LIE). To prove the "only if" direction, the key insight is to note that if Assumptions PPT and PT hold for all $g \in \mathscr{G}_\omega$, then we can choose two selection mechanisms $g_1, g_2 \in \mathscr{G}_\omega$ so that Assumptions PPT and PT holding for g_1 and g_2 imply that $E[\varepsilon_{i1} - \varepsilon_{i0}|\alpha_i, \mu_i] = E[\varepsilon_{i0} - \varepsilon_{i(-1)}|\alpha_i, \mu_i] = 0$. Assumption SEL and a weak regularity condition (Assumption REG in Appendix A) ensure that g_1 and g_2 are nondegenerate.

The necessary and sufficient condition in Proposition 1 is a time-homogeneity assumption on the conditional mean of ε_{it} given (α_i, μ_i). This condition is related to classical strict exogeneity assumptions in fixed effects models (Ghanem, Sant'Anna and Wüthrich, 2025). Thus, in settings where selection is on fixed effects (and absent structural breaks), Assumptions PPT and PT hold under standard conditions. Thus, pre-trends tests can be informative in this case.

B. Selection on pre-treatment information

Suppose now that the units select into treatment based on the information available in the pre-treatment period, so that $\omega_i = (\alpha_i, \varepsilon_i^0, \mu_i, \eta_i^0)$, which is often plausible in DiD applications. In the context of evaluating job training programs, selection into treatment may be on pre-program earnings (e.g., Ashenfelter and Card, 1985) or the discounted sum of future earnings conditional on the individuals' pre-treatment information sets. See Ghanem, Sant'Anna and Wüthrich (2025) for further discussions and examples of selection mechanisms. The following proposition provides necessary and sufficient conditions for this case.

PROPOSITION 2: *Suppose that $\omega_i = (\alpha_i, \varepsilon_i^0, \mu_i, \eta_i^0)$, Assumption SEL and Assumption REG in Appendix A hold, and the model is* (1). *Then, Assumptions PPT and PT hold jointly for all nondegenerate $g \in \mathscr{G}_\omega$ if and only if $\varepsilon_{i(-1)} = \varepsilon_{i0}$ and $E[\varepsilon_{i1}|\alpha_i, \varepsilon_i^0, \mu_i, \eta_i^0] = \varepsilon_{i0}$.*

The main implication of Proposition 2 is that Assumptions PPT and PT hold jointly for all $g \in \mathscr{G}_\omega$ only if there are no time-varying shocks in the pre-treatment period.[3] This condition likely fails in most applications. However, Assumption PT may hold when selection is on pre-treatment information. Indeed, the martingale condition $E[\varepsilon_{i1}|\alpha_i, \varepsilon_i^0, \mu_i, \eta_i^0] = \varepsilon_{i0}$ implies Assumption PT (Ghanem, Sant'Anna and Wüthrich, 2025). We emphasize that this martingale condition is not innocuous and needs to be carefully assessed in applications.

To illustrate, suppose that $\{\varepsilon_{it}\}$ follows an AR(1) process, $\varepsilon_{it} = \rho \varepsilon_{i(t-1)} + \zeta_{it}$, where $\zeta_{it} \sim \mathrm{WN}(0, \sigma^2)$ and $\mathrm{WN}(0, \sigma^2)$ denotes a mean-zero white noise process with variance σ^2. If $\rho = 1$, this model satisfies the martingale condition $E[\varepsilon_{i1}|\alpha_i, \varepsilon_i^0, \mu_i, \eta_i^0] = \varepsilon_{i0}$ and thus Assumption PT. By contrast, for

[2]Here and below, we interpret equalities involving random variables as holding almost surely.

[3]This result complements the discussion on pre- and post-trends in the presence of feedback in Bonhomme (2025, Section 5.2).

$\varepsilon_{i0} = \varepsilon_{i(-1)}$ to hold, we would further require a highly restrictive condition, specifically that the white noise process is degenerate ($\sigma^2 = 0$).

The discussion in this section shows that when selection is based on pre-treatment information, pre-trends tests can be uninformative and lead to DiD designs being wrongly discarded.

C. Selection on time-varying unobservables

Section II.B shows that the necessary and sufficient conditions are particularly restrictive if selection is on pre-treatment information. It turns out that this is a leading example of a broader phenomenon. The necessary and sufficient conditions tend to be restrictive whenever selection is on time-varying unobservables and may require the time series dependence between $\varepsilon_{i(-1)}$ and ε_{i0} to differ from the time series dependence between ε_{i0} and ε_{i1}.

To illustrate, consider the case where units select on the (known to them) treatment effect in period $t = 1$, $\tau_{i1} = Y_{i1}(1) - Y_{i1}(0)$. In this case, Lemma 1 in Appendix A implies the following necessary and sufficient conditions (i) $E[\varepsilon_{i0} - \varepsilon_{i(-1)}|\tau_{i1}] = 0$ and (ii) $E[\varepsilon_{i1} - \varepsilon_{i0}|\tau_{i1}] = 0$. Suppose that $\varepsilon_{it} \sim \mathrm{WN}(0, \sigma^2)$. Then, condition (ii) generally fails since τ_{i1} is typically a function of ε_{i1} (absent additional restrictions), whereas condition (i) holds assuming that τ_{i1} is not a function of ε_i^0. This demonstrates a case where pre-trends tests can be misleading.

III. Additional issues with time-varying covariates

Here we show that pre-trends tests can be misleading if researchers condition on pre-treatment values of time-varying covariates, even if these covariates are exogenous (not affected by the treatment). Specifically, suppose that researchers have access to a vector of exogenous time-varying covariates, $X_i = (X_{i(-1)}, X_{i0}, X_{i1})$.

We allow the outcome and selection model to depend on covariates. Ghanem, Sant'Anna and Wüthrich (2025) show that conditional parallel post-trends assumptions imply separability assumptions on the outcome model, restricting how the unobservable determinants of selection can interact with the covariates. We therefore consider the following separable two-way model for $Y_{it}(0)$,

$$Y_{it}(0) = \gamma_t(X_{it}) + \alpha_i + \lambda_t + \varepsilon_{it}, \quad E[\varepsilon_{it}|X_i] = 0, \quad i \in \{1, \dots, n\}, \quad t \in \{-1, 0, 1\}, \tag{3}$$

where $\gamma_t(\cdot)$ is a general time-varying function (e.g., $\gamma_t(X_{it}) = X_{it}\beta_t$). Suppose that selection is based on pre-treatment values of the covariates X_i^0, so that $G_i = g\left(X_i^0, \omega_i, \nu_i\right)$, and let $\mathscr{G}_{\omega,X^0}$ be the corresponding class of selection mechanisms.

To study the issues with pre-trends tests when researchers control for pre-treatment values of the time-varying covariates, X_i^0, we consider conditional versions of Assumptions PT and PPT.

ASSUMPTION PTX: $E\left[Y_{i1}(0) - Y_{i0}(0)|G_i = 1, X_i^0\right] = E\left[Y_{i1}(0) - Y_{i0}(0)|G_i = 0, X_i^0\right].$

ASSUMPTION PPTX: $E\left[Y_{i0}(0) - Y_{i(-1)}(0)|G_i = 1, X_i^0\right] = E\left[Y_{i0}(0) - Y_{i(-1)}(0)|G_i = 0, X_i^0\right].$

Under model (3), the necessary and sufficient conditions for Assumptions PPTX and PTX to hold jointly for all $g \in \mathscr{G}_{\omega,X^0}$ are

$$0 = E\left[\varepsilon_{i0} - \varepsilon_{i(-1)}|X_i^0, \omega_i\right] \tag{4}$$

$$0 = E\left[\varepsilon_{i1} - \varepsilon_{i0}|X_i^0, \omega_i\right] + E\left[\gamma_1(X_{i1})|X_i^0, \omega_i\right] - E\left[\gamma_1(X_{i1})|X_i^0\right] \tag{5}$$

See Appendix C for a detailed derivation. If $\omega_i = (\alpha_i, \mu_i)$, (4) and (5) generalize the necessary and sufficient conditions in Proposition 1; if $\omega_i = (\alpha_i, \varepsilon_i^0, \mu_i, \eta_i^0)$, (4) and (5) generalize the necessary and sufficient condition in Proposition 2.

Condition (4) holds if the conditional mean of $\varepsilon_{it}|X_i^0, \omega_i$ is constant over time. The first term of condition (5) is the post-treatment analog of condition (4). The additional terms reflect the correlation between X_{1i} and ω_i after conditioning on X_i^0.

To illustrate, suppose that $\omega_i = \alpha_i$ and $E\left[\varepsilon_{it}|X_i^0, \alpha_i\right]$ is time-invariant for all $t \in \{-1, 0, 1\}$, inspired by Section II.A. Then, Assumption PPTX holds for all $g \in \mathscr{G}_{\alpha, X^0}$. By contrast, Assumption PTX can fail if $E\left[\gamma_1(X_{i1})|X_i^0, \alpha_i\right] \neq E\left[\gamma_1(X_{i1})|X_i^0\right]$. $E[\gamma_1(X_{i1})|X_i^0, \alpha_i]$ is not equal to $E[\gamma_1(X_{i1})|X_i^0]$ if X_{i1} is correlated with α_i, which is likely in empirical applications. Indeed, an important reason for using fixed effects methods in the first place is to allow for such correlation. Intuitively, Assumption PTX can fail even if $\omega_i = \alpha_i$ because differencing does not fully eliminate the confounding effect of the fixed effect α_i if α_i is correlated with X_{i1} and X_{i1} is not conditioned on. Thus, pre-trends tests can be misleading when researchers control for pre-treatment values of time-varying covariates.

If the covariates are indeed exogenous, the issue documented above can be mitigated by conditioning on the entire time series of covariates, X_i, in Assumptions PPTX and PTX. In this case, maintaining that $\omega_i = \alpha_i$, the necessary and sufficient conditions become $E[\varepsilon_{i1}|X_i, \alpha_i] = E[\varepsilon_{i0}|X_i, \alpha_i] = E[\varepsilon_{i(-1)}|X_i, \alpha_i]$. Exogeneity of the covariates is a potentially strong assumption that needs to be carefully assessed in applications. See, for example, Caetano et al. (2022) for a discussion.

IV. Implications for practice

Our analysis has several practical implications. Most importantly, we show that even in the absence of structural breaks, looking at pre-trends can be uninformative about the validity of parallel post-trends assumptions, except under specific restrictions on selection into treatment. This finding suggests that correctly interpreting pre-trends tests requires understanding how units select into treatment and that pre-trends tests cannot and should not replace economic arguments for justifying parallel post-trends assumptions.

There are additional issues with pre-trends tests when the untreated potential outcomes depend on contemporaneous time-varying covariates. Controlling only for the pre-treatment values of such covariates can lead to failures of parallel post-trends assumptions, even if the pre-trends are parallel and the covariates are exogenous. Thus, pre-trends tests can be misleading in this case. Controlling for the entire time series of covariates can mitigate this issue under the potentially restrictive assumption that the covariates are exogenous.

REFERENCES

Ashenfelter, Orley C., and David Card. 1985. “Using the longitudinal structure of earnings to estimate the effect of training programs.” *The Review of Economics and Statistics*, 67(4): 648–660.

Bonhomme, Stéphane. 2025. “Back to Feedback: Dynamics and Heterogeneity in Panel Data.” Based on a presentation given at the Sargan lecture of the Royal Economic Society in Birmingham, July 2025.

Caetano, Carolina, Brantly Callaway, Stroud Payne, and Hugo Sant’Anna Rodrigues. 2022. “Difference in Differences with Time-Varying Covariates.” *arXiv:2202.02903*.

Ghanem, Dalia, Pedro H. C. Sant’Anna, and Kaspar Wüthrich. 2025. “Selection and parallel trends.” *arXiv 2203.09001v13*.

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

Roth, Jonathan, Pedro H. C. Sant’Anna, Alyssa Bilinski, and John Poe. 2023. “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature.” *Journal of Econometrics*, 235(2): 2218–2244.

Supplemental appendix to 'When should pre-trends be parallel?' by Ghanem, Sant'Anna, and Wüthrich

AUXILIARY LEMMA

Here we provide a general necessary and sufficient condition that is helpful for proving Propositions 1–2. It can be interpreted as a generalization of Theorem 3.1 in Ghanem, Sant'Anna and Wüthrich (2025) to the case where Assumptions PPT and PT hold jointly (specialized to model (1)). To prove the result, we impose the following additional weak regularity condition.

ASSUMPTION REG: $P(E[\varepsilon_{i1}-\varepsilon_{i0}|\omega_i]>0)<1$ *and* $P(E[\varepsilon_{i0}-\varepsilon_{i(-1)}|\omega_i]>0)<1$.

Assumption REG (together with Assumption SEL) ensures that the specific selection mechanisms we use to prove the "only if" direction in the proof of the lemma are nondegenerate.

LEMMA 1: *Suppose that Assumptions SEL and REG hold and the model is* (1). *Then, Assumptions PPT and PT hold jointly for all nondegenerate* $g \in \mathscr{G}_\omega$ *if and only if* $E[\varepsilon_{i1}-\varepsilon_{i0}|\omega_i]=E[\varepsilon_{i0}-\varepsilon_{i(-1)}|\omega_i]=0$.

PROOF:

We adopt the proof strategy in Ghanem, Sant'Anna and Wüthrich (2025, Theorem 3.1). The "if" direction follows by the LIE and Assumption SEL. We proceed to show the "only if" direction. Assumption PPT and PT holding for all $g \in \mathscr{G}_\omega$ implies that they hold for the following two mechanisms,

$$\begin{aligned} G_i &= g_1(\omega_i,\nu_i)=1\{\nu_i>\nu\}1\{E[\varepsilon_{i0}-\varepsilon_{i(-1)}|\omega_i]\le 0\}, \\ G_i &= g_2(\omega_i,\nu_i)=1\{\nu_i>\nu\}1\{E[\varepsilon_{i1}-\varepsilon_{i0}|\omega_i]\le 0\}. \end{aligned}$$

Both of these selection mechanisms are nondegenerate under the maintained assumptions.

We first consider the implications of Assumptions PPT and PT holding jointly for g_1. Under Assumption SEL, by Lemmas G.1 and G.2 in Ghanem, Sant'Anna and Wüthrich (2025), Assumption PPT holding for g_1 implies that

$$E[\varepsilon_{i0}-\varepsilon_{i(-1)}|\omega_i]=0. \tag{A1}$$

Similarly, Assumption PT holding for g_1 implies that

$$E[1\{E[\varepsilon_{i0}-\varepsilon_{i(-1)}|\omega_i]\le 0\}E[\varepsilon_{i1}-\varepsilon_{i0}|\omega_i]]=0, \tag{A2}$$

where we have used Assumption SEL. Since $1\{E[\varepsilon_{i0}-\varepsilon_{i(-1)}|\omega_i]\le 0\}=1$ by (A1), it follows that the last equality implies $E[\varepsilon_{i1}-\varepsilon_{i0}]=0$, which holds by assumption.

Next, we consider the implications of Assumptions PPT and PT holding jointly for g_2. Under Assumption SEL, by Lemmas G.1 and G.2 in Ghanem, Sant'Anna and Wüthrich (2025), Assumption PT holding for g_2 implies

$$E[\varepsilon_{i1}-\varepsilon_{i0}|\omega_i]=0. \tag{A3}$$

Similarly, Assumption PPT holding for g_2 implies that

$$E[1\{E[\varepsilon_{i1}-\varepsilon_{i0}|\omega_i]\le 0\}E[\varepsilon_{i0}-\varepsilon_{i(-1)}|\omega_i]]=0, \tag{A4}$$

where we have used Assumption SEL. Since $1\{E[\varepsilon_{i1}-\varepsilon_{i0}|\omega_i]\le 0\}=1$ by (A3), (A4) implies $E[\varepsilon_{i0}-\varepsilon_{i(-1)}]=0$, which holds by assumption.

It follows that Assumption PPT and PT holding jointly for all $g \in \mathscr{G}_\omega$ imply $E[\varepsilon_{i0}-\varepsilon_{i(-1)}|\omega_i]=0$ and $E[\varepsilon_{i1}-\varepsilon_{i0}|\omega_i]=0$. Q.E.D.

PROOFS OF PROPOSITIONS

PROOF OF PROPOSITION 1:

The result follows from an application of Lemma 1 with $\omega_i = (\alpha_i, \mu_i)$. Q.E.D.

PROOF OF PROPOSITION 2:

The result follows from an application of Lemma 1 with $\omega_i = (\alpha_i, \varepsilon_i^0, \mu_i, \eta_i^0)$. Q.E.D.

DERIVATION OF NECESSARY AND SUFFICIENT CONDITIONS WITH COVARIATES

Here we provide a detailed derivation of the necessary and sufficient conditions with covariates. By the arguments in Appendix D of Ghanem, Sant'Anna and Wüthrich (2025) with X_i replaced by X_i^0 combined with the arguments in the proof of Lemma 1, the necessary and sufficient conditions for Assumptions PPTX and PTX to hold jointly for all $g \in \mathscr{G}_{\omega,X^0}$ are

$$\text{(C1)} \qquad E[Y_{i0}(0) - Y_{i(-1)}(0)|X_i^0, \omega_i] = E[Y_{i0}(0) - Y_{i(-1)}(0)|X_i^0],$$

$$\text{(C2)} \qquad E[Y_{i1}(0) - Y_{i0}(0)|X_i^0, \omega_i] = E[Y_{i1}(0) - Y_{i0}(0)|X_i^0].$$

Plugging model (3) into (C1) yields

$$\begin{aligned} & E[\gamma_0(X_{i0}) + \alpha_i + \lambda_0 + \varepsilon_{i0} - (\gamma_{-1}(X_{i(-1)}) + \alpha_i + \lambda_{-1} + \varepsilon_{i(-1)})|X_i^0, \omega_i] \\ & = E[\gamma_0(X_{i0}) + \alpha_i + \lambda_0 + \varepsilon_{i0} - (\gamma_{-1}(X_{i(-1)}) + \alpha_i + \lambda_{-1} + \varepsilon_{i(-1)})|X_i^0]. \end{aligned}$$

Simplifying this expression and noting that X_i^0 contains X_{i0} and $X_{i(-1)}$ yields

$$E[\varepsilon_{i0} - \varepsilon_{i(-1)}|X_i^0, \omega_i] = E[\varepsilon_{i0} - \varepsilon_{i(-1)}|X_i^0].$$

Noting that $E[\varepsilon_{i0} - \varepsilon_{i(-1)}|X_i^0] = E[E[\varepsilon_{i0} - \varepsilon_{i(-1)}|X_i]|X_i^0] = 0$ because $E[\varepsilon_{it}|X_i] = 0$ by assumption yields condition (4).

Plugging model (3) into (C2) yields

$$\begin{aligned} & E[\gamma_1(X_{i1}) + \alpha_i + \lambda_1 + \varepsilon_{i1} - (\gamma_0(X_{i0}) + \alpha_i + \lambda_0 + \varepsilon_{i0})|X_i^0, \omega_i] \\ & = E[\gamma_1(X_{i1}) + \alpha_i + \lambda_1 + \varepsilon_{i1} - (\gamma_0(X_{i0}) + \alpha_i + \lambda_0 + \varepsilon_{i0})|X_i^0]. \end{aligned}$$

Simplifying this expression and noting that X_i^0 contains X_{i0} but not X_{i1} yields

$$E[\gamma_1(X_{i1}) + \varepsilon_{i1} - \varepsilon_{i0}|X_i^0, \omega_i] = E[\gamma_1(X_{i1}) + \varepsilon_{i1} - \varepsilon_{i0}|X_i^0].$$

Rearranging this expression and noting that $E[\varepsilon_{i1} - \varepsilon_{i0}|X_i^0] = 0$ by $E[\varepsilon_{it}|X_i] = 0$ and the LIE as before yields condition (5).