

Selection and parallel trends*

Dalia Ghanem[†] Pedro H. C. Sant'Anna[‡] Kaspar Wüthrich[§]

First draft on arXiv: March 17, 2022. This draft: December 12, 2022

Abstract

We study the connection between selection into treatment and the parallel trends assumptions underlying difference-in-differences (DiD) designs. We start by deriving necessary and sufficient conditions for the parallel trends assumption. These conditions theoretically clarify the empirical content of this key assumption and demonstrate trade-offs between restrictions on selection and the distribution of time-varying unobservables. We then consider different restrictions on the selection mechanism and provide a menu of interpretable primitive sufficient conditions, which constitute a formal framework for justifying DiD in practice. We illustrate these sufficient conditions with two examples of selection into treatment: (i) selection on untreated potential outcomes and (ii) Roy-style selection. We derive results for both separable and nonseparable outcome models and show that this distinction has implications for the role of covariates in parallel trends assumptions. Building on our analysis of nonseparable models, we connect DiD to the literature on nonparametric identification in panel models.

Keywords: causal inference, conditional parallel trends, covariates, difference-in-differences, selection mechanism, time-invariant and time-varying unobservables, treatment effects

JEL Codes: C21, C23

*We are grateful to Isaiah Andrews, Manuel Arellano, Dmitry Arkhangelsky, Stéphane Bonhomme, Christoph Breunig, Federico Bugni, Brantly Callaway, Ivan Canay, Clément de Chaisemartin, Gordon Dahl, Joachim Freyberger, Bryan Graham, Lena Janys, Stefan Hoderlein, Christian Hansen, Nikolay Kudrin, David McKenzie, Eduardo Morales, Vitor Possebom, Niklas Potrafke, Jonathan Roth, Yuya Sasaki, seminar participants at the Bank of Spain, Monash University, Northwestern University, University of Bonn, UC Davis, UC San Diego, University of Zürich, and conference participants at the California Econometrics Conference 2022, the CEMFI Fall Econometrics Conference 2022, and the CESifo Area Conference on Labor Economics 2022 for comments. The usual disclaimer applies.

[†]Department of Agricultural & Resource Economics, University of California, Davis. One Shields Ave, Davis CA 95616; dghanem@ucdavis.edu

[‡]Microsoft and Department of Economics, Vanderbilt University. 14820 NE 36th St, Redmond 98052, WA; psantanna@microsoft.com

[§]Department of Economics, University of California San Diego, 9500 Gilman Dr. La Jolla, CA 92093; CESifo; Ifo Institute. kwuthrich@ucsd.edu

... while the new papers [in the DiD literature] clarify very well the statistical assumptions needed for estimation, effective use of these methods also requires being able to understand what the threats to these assumptions are in different contexts, and to make a plausible rhetorical argument as to why we should think the assumptions hold.

— David McKenzie, *World Bank Development Impact Blog* (McKenzie, 2022)

1 Introduction

Difference-in-differences (DiD) designs are widely used in practice to estimate causal effects. One of the perceived advantages of DiD is that it does not require explicit assumptions on how units select into treatment but instead relies on parallel trends assumptions. However, when justifying DiD in empirical applications, researchers often argue that the treatment is “quasi-randomly” assigned. Although these discussions allude to selection mechanisms, they are often not explicit about what constitutes “quasi-random” assignment, arguably due to the lack of formal guidance.

In this paper, we study parallel trends assumptions through the lens of selection mechanisms. Our goals are threefold: (i) explicitly examine the role of selection into treatment in the context of the parallel trends assumption, (ii) provide practitioners with a menu of theory-based templates for justifying parallel trends in applications, and (iii) provide foundations for comparing DiD to competing methods and developing new ones.

Consider the classical 2×2 DiD setup, where we observe N units over two time periods. In the first period, none of the units is treated; in the second period, some units select into treatment (treatment group), while others remain untreated (control group). Let $Y_{it}(0)$ denote the untreated potential outcome for unit $i = 1, \dots, N$ in time period $t = 1, 2$. The identifying assumption of DiD is the parallel trends assumption. This assumption requires that the expected change across time in the untreated potential outcome, $Y_{it}(0)$, is identical in the treatment and control group,

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1] = E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0],$$

where $G_i = 1$ indicates the treatment group and $G_i = 0$ indicates the control group. We focus on the 2×2 case for expositional simplicity. Our results directly extend to general DiD designs with multiple groups and multiple periods; see Appendix B.3.

We begin our analysis with a separable model for the untreated potential outcome

$$Y_{it}(0) = \alpha_i + \lambda_t + \varepsilon_{it}, \quad (1)$$

where α_i and ε_{it} are time-invariant and time-varying unit-specific unobservables, respectively, and λ_t is a common time effect. Equation (1) imposes separability in the unobservable determinants of the *untreated* potential outcome and allows for a transparent discussion of our main theoretical results. Our results directly generalize to fully nonseparable, time-varying models

$$Y_{it}(0) = \xi_t(\alpha_i, \varepsilon_{it}), \quad (2)$$

where $\xi_t(\cdot)$ is an arbitrary time-varying function, and α_i , ε_{i1} , and ε_{i2} are random vectors. See Appendix B.2 for details.

To study the role of selection into treatment, we consider a general selection mechanism that depends on the unobservable determinants of the untreated potential outcomes as well as additional time-invariant and time-varying unobservables $(\nu_i, \eta_{i1}, \eta_{i2})$,

$$G_i = g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}). \quad (3)$$

The selection mechanism in Equation (3) accommodates selection on time-invariant unobservables (“fixed effects”), selection on untreated potential outcomes, selection on treatment effects (Roy-style selection), and other economic models of selection.

Our first contribution is to provide necessary and sufficient conditions for parallel trends, which clarify the empirical content of this assumption. We first consider a scenario where researchers are not willing to impose a specific model for the selection mechanism.¹ We show that absent any restrictions on how selection depends on ε_{i1} and ε_{i2} , parallel trends holds if and only if ε_{it} is time-invariant. This condition is generally implausible since it implies that the untreated potential outcomes are constant across time up to location shifts, $\lambda_2 - \lambda_1$. We therefore consider restricted classes of selection mechanisms: (i) if selection does not depend on ε_{i2} , parallel trends implies a martingale-type property on this time-varying unobservable; (ii) if selection does not

¹In Appendix B.1, we provide necessary and sufficient conditions under an alternative scenario where researchers are not willing to impose any restrictions on the distribution of (time-varying) unobservables.

depend on $(\varepsilon_{i1}, \varepsilon_{i2})$, parallel trends implies that the conditional mean of ε_{it} given α_i does not vary across time. Under high-level assumptions on the conditional expectation of G_i , the latter two necessary conditions are also sufficient for parallel trends.

Taken together, the necessary and sufficient conditions imply that parallel trends cannot hold absent restrictions on the selection mechanism and/or the distribution of (time-varying) unobservables. These conditions raise two questions: (i) Can parallel trends hold if $Y_{it}(0)$ varies across time beyond location shifts *and* selection depends on both ε_{i1} and ε_{i2} , albeit in a restricted way? (ii) Are there interpretable primitive sufficient conditions for parallel trends when researchers are willing to restrict which (if any) time-varying unobservables enter the selection mechanism? We show that the answer to both questions is yes. We provide three sets of primitive sufficient conditions that differ in terms of whether and how ε_{i1} and ε_{i2} affect selection. These sufficient conditions constitute theory-based templates for justifying parallel trends based on contextual information about selection. We illustrate the sufficient conditions based on two examples of mechanisms where selection is either based on untreated potential outcomes or treatment effects (Roy-style selection). In addition, we discuss the implications of our results for justifying parallel trends in practice in Section 3.3.

We then examine the role of (time-varying) covariates in DiD analyses.² We start by providing primitive sufficient conditions for parallel trends for models that are separable in covariates and unobservables. Our results highlight the importance of time-varying covariates in weakening the sufficient conditions for parallel trends. In all our sufficient conditions, time-varying covariates can enter the selection mechanism in an unrestricted way and do not have to obey the restrictions imposed on the time-varying unobservables. Furthermore, conditioning on both time-invariant and time-varying covariates makes the restrictions on the distribution of unobservables more plausible.

We then generalize our analysis to a nonseparable model where covariates and unobservables can interact. We show that many of the insights from our analysis of the separable model remain valid. However, nonseparability between the covariates and the unobservables determining selection implies that parallel trends can only hold within subpopulations for which these covariates do not vary across time. This analysis highlights that the role of covariates in DiD analyses depends on how they enter the outcome model. See Section 4.5 for further discussion of the practical implications.

²We assume that covariates are not affected by the treatment. See Caetano et al. (2022) for some recent results relaxing this assumption.

This paper contributes to several branches of the literature on causal inference using panel data. Our first contribution is to the classical literature on canonical DiD setups. See, e.g., Ashenfelter (1978), Ashenfelter and Card (1985), Heckman and Robb (1985), Card (1990), Card and Krueger (1994), Meyer et al. (1995), and Angrist and Krueger (1999) for early developments, and Section 2 of Lechner (2010) for a historical perspective. Our contribution is to provide foundations for the parallel trends assumption to hold in non-experimental settings, where selection into treatment may depend on time-invariant and time-varying unobservables.

Our second contribution is to the more recent literature on DiD methods. See, e.g., de Chaisemartin and D’Haultfoeuille (2021) and Roth et al. (2022) for surveys. Within this strand of the literature, our paper is most closely related to Roth and Sant’Anna (2021), Arkhangelsky et al. (2021), and Arkhangelsky and Imbens (2022), though our focus greatly differs from theirs. Roth and Sant’Anna (2021) discuss necessary and sufficient conditions under which the parallel trends assumption is satisfied for all (monotonic) transformations of the untreated potential outcome. We, on the other hand, take the outcome model (and thus the specific transformation) as given and study the connection between parallel trends and selection into treatment. Arkhangelsky et al. (2021) and Arkhangelsky and Imbens (2022) propose doubly robust estimation methods that leverage restrictions on outcome models and/or selection models with unconfoundedness-type restrictions; see also Athey et al. (2021). Our results complement theirs as we maintain the parallel trends assumption and discuss the types of restrictions on selection compatible with it. Moreover, our analysis shows that parallel trends is compatible with various types of selection on unobservables, unlike standard unconfoundedness assumptions (e.g., Imbens, 2004; Imbens and Wooldridge, 2009).

Our third contribution is to the literature imposing explicit selection and/or outcome models to develop and compare different methods for estimating treatment effects, including DiD (e.g., Ashenfelter and Card, 1985; Heckman and Robb, 1985; Card and Hyslop, 2005; Chabé-Ferret, 2015; de Chaisemartin and D’Haultfoeuille, 2018; Verdier, 2020; Marx et al., 2022). These selection mechanisms were developed for economic models, some of which are tailored to applications such as job training and technology adoption. Our results complement this strand of the literature in several ways. First, our necessary and sufficient conditions are derived for general selection and outcome models that nest models considered in this literature. Our

conditions thus clarify trade-offs between assumptions on selection and time-varying unobservables that are relevant for those models. Second, our primitive sufficient conditions nest several of the existing application-specific restrictions. Third, we provide results for both separable and nonseparable models and clarify the role of covariates in the context of parallel trends assumptions. It is worth noting that while most papers in this literature examine sharp DiD designs, as we do, de Chaisemartin and D’Haultfoeuille (2018) and Marx et al. (2022) also consider fuzzy DiD designs.

Finally, we establish an explicit connection between DiD and the literature on nonseparable panel models.³ A strand of this literature has analyzed the identification of average effects either by allowing for fixed effects and imposing time homogeneity (e.g. Hoderlein and White, 2012; Chernozhukov et al., 2013) or restricting individual heterogeneity via nonparametric correlated random effects assumptions (e.g. Altonji and Matzkin, 2005; Bester and Hansen, 2009). We show that our sufficient conditions for parallel trends imply combinations of time homogeneity and (correlated) random effects restrictions. Our results demonstrate how restrictions on the selection mechanism can be used to justify identification assumptions in the nonseparable panel literature.

Notation. For a random vector W_{it} , where $i = 1, \dots, N$ and $t = 1, 2$, we denote its time series by $W_i \equiv (W_{i1}, W_{i2})$. We use F_W to denote the distribution of the random vector W . Let $f(z, w)$ be a function defined on $\mathcal{Z} \times \mathcal{W}$. We say that $f(z, w)$ is a trivial function of w if $f(z, w) = f(z, w') = h(z)$ for all $z \in \mathcal{Z}$, $w \neq w'$, and $(w, w') \in \mathcal{W}^2$. We say that $f(z, w)$ is a symmetric function in z and w if $f(z, w) = f(w, z)$ for all $(z, w) \in \mathcal{Z} \times \mathcal{W}$. For a vector W_i , W_i^j is the j^{th} element of W_i . We use the notation $\stackrel{d}{=}$ to denote equality of distribution. For random variables, X_i , Z_i , and W_i , $Z_i|W_i, X_i \stackrel{d}{=} Z_i|X_i, W_i$ denotes that $F_{Z_i|W_i, X_i}(z|w, x) = F_{Z_i|X_i, W_i}(z|w, x)$ for $(z, w, x) \in \mathcal{Z} \times \mathcal{W} \times \mathcal{X}$.

³See, e.g., Altonji and Matzkin (2005); Athey and Imbens (2006); Bester and Hansen (2009); Hoderlein and White (2012); Chernozhukov et al. (2013); Arellano and Bonhomme (2016); Ghanem (2017). This work extends notions of fixed effects and correlated random effects that originated in the linear model (Mundlak, 1961, 1978; Chamberlain, 1982, 1984). Recent surveys (Arellano and Honoré, 2001; Arellano and Bonhomme, 2011) and textbook treatments (Arellano, 2003; Wooldridge, 2010) further describe the role of restrictions on time and individual heterogeneity in linear and nonlinear models. Such restrictions have been imposed in the context of identification in limited dependent variable models (e.g. Manski, 1987; Honoré, 1993; Kyriazidou, 1997; Honoré and Kyriazidou, 2000a,b) and random coefficient models (e.g. Chamberlain, 1992; Graham and Powell, 2012; Arellano and Bonhomme, 2012). Nonparametric identification of panel models with additivity restrictions has been examined, e.g., in Evdokimov (2010) and Freyberger (2017).

2 Necessary and sufficient conditions for parallel trends

2.1 Setup

We consider the classical DiD setup with two groups and two periods and abstract from covariates. We discuss the role of covariates in Section 4 and generalize our results to DiD designs with multiple groups and multiple periods in Appendix B.3. Let D_{it} and Y_{it} denote the treatment status and outcome for unit i in period t . Here the index i refers to the unit making the decision to select into treatment. This could be an individual or a more aggregate administrative unit, such as county or state. The treatment group ($G_i = 1$) selects the treatment path $D_i = (0, 1)$; the control group ($G_i = 0$) selects $D_i = (0, 0)$. The potential outcomes with and without the treatment are $Y_{it}(1)$ and $Y_{it}(0)$, respectively.⁴

We consider the standard parallel trends assumption. Throughout the paper, we assume that all relevant moments exist.

Assumption PT. *The (unconditional) parallel trends assumption holds:*

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1] = E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0].$$

To simplify the exposition and convey the main ideas, we consider a model for the potential outcome without the treatment that is separable in the time-invariant and time-varying unobservables.

Assumption SP.

$$Y_{it}(0) = \alpha_i + \lambda_t + \varepsilon_{it}, \quad E[\varepsilon_{it}] = 0, \quad i = 1, \dots, N, \quad t = 1, 2.$$

In Assumption SP, α_i is the time-invariant unobservable, λ_t is the (non-stochastic) time fixed effect, and ε_{it} is the time-varying, unit-specific unobservable.⁵ The assumption that the time-varying unobservables have mean zero, $E[\varepsilon_{it}] = 0$, is a normalization. It is without loss of generality since we can always redefine λ_t such that this assumption holds. In what follows, we denote the supports of α_i and ε_{it} as \mathcal{A} and \mathcal{E} ,

⁴We assume that the units do not anticipate their treatment. As a result, at $t = 1$ we observe $Y_{i1}(0)$ for all units, while at $t = 2$ we observe $Y_{i2}(1)$ for treated and $Y_{i2}(0)$ for untreated units.

⁵The assumption that λ_t is non-stochastic is w.l.o.g. since we can always reparametrize the model as $Y_{it}(0) = \alpha_i + \tilde{\lambda}_t + \tilde{\varepsilon}_{it}$, where $\tilde{\lambda}_t$ is stochastic, $E[\tilde{\lambda}_t] = \lambda_t$, and $\tilde{\varepsilon}_{it} = \varepsilon_{it} - (\tilde{\lambda}_t - \lambda_t)$.

respectively.⁶

In Section 2.3.2 and Appendix B.2, we consider the fully nonseparable, time-varying potential outcome model, $Y_{it}(0) = \xi_t(\alpha_i, \varepsilon_{it})$. We show that all our results extend straightforwardly to this nonseparable model when we replace ε_{it} by the centered potential outcome $Y_{it}(0) - E[Y_{it}(0)]$.

We work with separable and nonseparable models that allow us to distinguish between time-invariant and time-varying unobservables. This is necessary to define selection mechanisms that can directly depend on these unobservables.⁷ We specifically consider a general class of selection mechanisms in which units select into treatment based on $(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ as well as additional time-invariant and time-varying vectors of random variables, $(\nu_i, \eta_{i1}, \eta_{i2})$,

$$G_i = g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}). \quad (4)$$

Note that since $G_i = D_{i2}$, $g(\cdot)$ can equivalently be viewed as the selection mechanism for D_{i2} . Let \mathcal{G}_{all} denote the set of all selection mechanisms $g(\cdot)$, mapping from the support of the unobservables to $\{0, 1\}$.

The class of selection mechanisms \mathcal{G}_{all} accommodates many different types of selection, including random assignment, selection on time-invariant unobservables, selection on untreated potential outcomes, Roy-style selection on treatment effects, and other selection mechanisms based on economic decision problems (e.g. Heckman and Robb, 1985; Chabé-Ferret, 2015; Marx et al., 2022).

For selection mechanisms $g \in \mathcal{G}_{\text{all}}$, Assumption SP does not imply the PT assumption without further restrictions. In Section 2.2, we examine the trade-offs between restrictions on the selection mechanism and the distribution of unobservables by deriving necessary and sufficient conditions for Assumption PT.

Remark 1 (Two-way fixed effects model). *Assumption SP does not impose the standard two-way fixed effects model for the realized outcome, $Y_{it} = \delta D_{it} + \alpha_i + \lambda_t + \varepsilon_{it}$, which imposes treatment effect homogeneity. Since Assumption SP does not restrict the potential outcome with the treatment, $Y_{it}(1)$, it is consistent with arbitrary treatment effect heterogeneity.* \square

⁶For simplicity, we assume that the supports do not depend on (i, t) .

⁷If, instead, we were to work directly with potential outcomes, this would rule out important examples of selection mechanisms such as selection on time-invariant unobservables (e.g., Ashenfelter and Card, 1985).

Remark 2 (Parallel trends and functional form). *Throughout this paper, for both separable and nonseparable models, we take the functional form of the outcome as given. We thereby abstract from the issues arising from the sensitivity of DiD to functional form specification; see Roth and Sant’Anna (2021) for a discussion.* \square

2.2 Necessary and sufficient conditions

To better understand the implications of parallel trends, we derive necessary and sufficient conditions for this assumption. We consider a scenario where researchers are not willing to impose a specific model for the selection mechanism $g(\cdot)$ and want parallel trends to hold for any fixed $g \in \mathcal{G}$, where $\mathcal{G} \subseteq \mathcal{G}_{\text{all}}$ is a (potentially restricted) class of selection mechanisms. Our focus on classes of selection mechanisms is motivated by parallel trends not relying on an explicit model of selection.

We start by analyzing a scenario where researchers are not willing to make any assumptions on the selection mechanism so that parallel trends needs to hold for any fixed $g \in \mathcal{G}_{\text{all}}$ and then also consider two scenarios where parallel trends holds for restricted versions of \mathcal{G}_{all} .

To ensure non-degeneracy of the selection mechanisms we use to derive necessary and sufficient conditions for parallel trends, we impose the following weak regularity condition.

Assumption SEL. *There exists a component of ν_i , labeled ν_i^1 (w.l.o.g.), such that $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ and $P(\nu_i^1 > c) \in (0, 1)$ for some $c \in \mathbb{R}$.*

The following proposition provides a necessary and sufficient condition for parallel trends absent any restrictions on the selection mechanism.

Proposition 2.1 (Necessary and sufficient condition for $g \in \mathcal{G}_{\text{all}}$). *Suppose that Assumptions SP and SEL hold and $P(G_i = 1) \in (0, 1)$. Then Assumption PT holds for any fixed $g \in \mathcal{G}_{\text{all}}$ if and only if $\varepsilon_{i1} = \varepsilon_{i2}$ almost surely (a.s.).*

The “if” direction of the proof is straightforward. The “only if” direction follows by noting that if Assumption PT holds for any (fixed) $g \in \mathcal{G}_{\text{all}}$, then it holds for $g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{\varepsilon_{i2} \geq \varepsilon_{i1}\}$.⁸ By Lemmas A.1 and A.2, this specific choice of selection mechanism can be shown to imply the result.

⁸Note that under Assumption SEL, the condition $P(G_i = 1) \in (0, 1)$ implies $P(\varepsilon_{i2} \geq \varepsilon_{i1}) > 0$.

Proposition 2.1 shows that Assumption PT holds for any fixed selection mechanism $g(\cdot)$ if and only if the time-varying unobservables are in fact time-invariant. Put simply, if one were to allow for an *unrestricted* selection mechanism, one would need to rule out time-varying shocks. Given that this condition is implausible in many applications, we next provide necessary and sufficient conditions under restricted versions of the selection mechanism.

We consider two restricted classes of selection mechanisms. First, we examine a class of selection mechanisms in which selection does not depend on the time-varying unobservable determinant of $Y_{i2}(0)$,⁹

$$\mathcal{G}_1 = \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, e_2, v, t_1, t_2) \text{ is a trivial function of } e_2\}.$$

Second, we do not allow the time-varying unobservable determinants of $Y_{i1}(0)$ and $Y_{i2}(0)$ to enter the selection mechanism and consider the class

$$\mathcal{G}_2 = \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, e_2, v, t_1, t_2) \text{ is a trivial function of } (e_1, e_2)\}.$$

The next two propositions provide necessary and sufficient conditions for parallel trends when the selection mechanism belongs to \mathcal{G}_1 and \mathcal{G}_2 , respectively.

Proposition 2.2 (Necessary and sufficient condition for $g \in \mathcal{G}_1$). *Suppose that Assumptions SP and SEL hold and $P(G_i = 1) \in (0, 1)$. If Assumption PT holds for any fixed $g \in \mathcal{G}_1$, then $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$ a.s. If, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i, \varepsilon_{i1}]$ a.s., then $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$ a.s. is also sufficient for Assumption PT.*

Proposition 2.3 (Necessary and sufficient condition for $g \in \mathcal{G}_2$). *Suppose that Assumptions SP and SEL hold and $P(G_i = 1) \in (0, 1)$. If Assumption PT holds for any fixed $g \in \mathcal{G}_2$, then $E[\varepsilon_{i1}|\alpha_i] = E[\varepsilon_{i2}|\alpha_i]$ a.s. If, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i]$ a.s., then $E[\varepsilon_{i1}|\alpha_i] = E[\varepsilon_{i2}|\alpha_i]$ a.s. is also sufficient for Assumption PT.*

The two propositions demonstrate that while parallel trends is compatible with the presence of time-varying unobservables under the restricted classes of selection mechanisms, it implies time series restrictions on ε_{it} . Proposition 2.2 shows that for Assumption PT to hold for any fixed selection mechanism that is a trivial function

⁹The case where selection does not depend on the time-varying unobservable determinant of $Y_{i1}(0)$ is symmetric. However, the resulting necessary condition would be implausible from a time-series perspective.

of ε_{i2} ($g \in \mathcal{G}_1$), it is necessary that $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$, a martingale-type property.¹⁰ This property is satisfied, for example, if $\varepsilon_{i2} = \varepsilon_{i1} + \zeta_{i2}$, where ζ_{it} is white noise. Moreover, when the conditional expectation of G_i given $(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ does not depend on ε_{i2} , this martingale-type property is also sufficient for parallel trends.¹¹

When we further restrict selection to be a trivial function of both ε_{i1} and ε_{i2} , Proposition 2.3 shows that parallel trends implies that the conditional mean of ε_{it} given α_i is stable over time. If $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i]$, this condition is also sufficient for parallel trends. In general, the stability of the conditional mean, $E[\varepsilon_{i1}|\alpha_i] = E[\varepsilon_{i2}|\alpha_i]$, is implied by (and weaker than) the textbook strict exogeneity assumption, $E[\varepsilon_{it}|G_i, \alpha_i] = 0$, since in our framework selection may depend on additional unobservables $(\nu_i, \eta_{i1}, \eta_{i2})$.

Taken together, our necessary and sufficient conditions show that Assumption PT cannot hold absent additional restrictions on the selection mechanism and/or the distribution of unobservables. In particular, these results highlight the role of restrictions on time-varying unobservables, either in terms of how they vary over time or how they determine selection. As a result, researchers using DiD approaches cannot avoid making meaningful and nontrivial assumptions on selection and time-varying unobservables.

2.3 Extensions

In this section, we discuss three extensions. We summarize the main results here and refer to Appendix B for details.

2.3.1 Parallel trends for any distribution

In Appendix B.1, we provide necessary and sufficient conditions for an alternative scenario where researchers are not willing to restrict the distribution of unobservables. Specifically, suppose researchers want parallel trends to hold for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$, where \mathcal{F} is a class of distributions.

¹⁰Due to the presence of α_i , this martingale property does not imply that $Y_{it}(0)$ is a martingale.

¹¹The result in Proposition 2.2 relates to the consistency of the first-differences estimator under violations of strict exogeneity when the idiosyncratic shocks follow a unit root. In this case, $\Delta\varepsilon_{i2} = \zeta_{i2}$. Since $\zeta_{i2} \perp (G_i, \alpha_i)$ by the white-noise assumption, it follows that $E[\Delta\varepsilon_{i2}|G_i, \alpha_i] = 0$, even though strict exogeneity may be violated. We thank Stéphane Bonhomme for pointing out this connection.

We show that if \mathcal{F} is a complete class of distributions, then Assumption PT holds for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$ if and only if

$$P(G_i = 1 | \varepsilon_{i1}, \varepsilon_{i2}) = P(G_i = 1) \text{ a.s. for all } F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}.$$

That is, parallel trends (holding for any distribution of unobservables) is equivalent to selection being independent of the time-varying unobservable determinants of the untreated potential outcome. Intuitively, completeness of \mathcal{F} , which is formally defined in Definition B.1, requires that the class of possible distributions of unobservables is “rich enough.” This condition is trivially satisfied if \mathcal{F} is completely unrestricted.

2.3.2 Nonseparable models

In Appendix B.2, we derive necessary and sufficient conditions for a fully nonseparable potential outcome model

$$Y_{it}(0) = \xi_t(\alpha_i, \varepsilon_{it}), \quad i = 1, \dots, N, \quad t = 1, 2, \quad (5)$$

where $\xi_t(\cdot)$ is an arbitrary time-varying function, and α_i , ε_{i1} , and ε_{i2} are finite-dimensional vector-valued random variables. The nonseparable model (5) is fully general and does not impose any restrictions on the potential outcomes.

We show that the results for the separable model in Assumption SP extend in a straightforward manner to the nonseparable model. The necessary and sufficient condition for parallel trends for any fixed $g \in \mathcal{G}_{\text{all}}$ (Proposition B.2) is

$$Y_{i1}(0) - E[Y_{i1}(0)] = Y_{i2}(0) - E[Y_{i2}(0)] \text{ a.s.}$$

That is, the untreated potential outcome does not vary across time except for location shifts as in Proposition 2.1. It is worth noting that for outcomes with finite support, this condition would generally rule out location shifts.

Similar to Propositions 2.2 and 2.3, Propositions B.3 and B.4 show that parallel trends for any fixed $g \in \mathcal{G}_1$ and $g \in \mathcal{G}_2$ in the context of the fully nonseparable outcome model implies a martingale-type property and stability of the conditional mean of the centered untreated potential outcome across time, respectively.

2.3.3 Multiple periods and multiple groups

In Appendix B.3, we extend our results to DiD designs with multiple periods and multiple groups.¹² Specifically, we consider a staggered adoption setting with T periods, where no units are treated at $t = 1$ and some units remain untreated at $t = T$. The group indicator G_i denotes the first period in which units select into the treatment. We set $G_i = \infty$ for the never-treated units so that $G_i \in \{2, \dots, T, \infty\}$.

We consider the following standard parallel trends assumption on the never-treated potential outcome $Y_{it}(\infty)$,

$$E[Y_{it}(\infty) - Y_{it-1}(\infty) | G_i = g] = E[Y_{it}(\infty) - Y_{it-1}(\infty) | G_i = \infty] \quad \text{for all } (g, t). \quad (6)$$

Under this assumption and a no-anticipation condition, group-time ATTs are identified (e.g., Callaway and Sant’Anna, 2021).

For simplicity, we maintain a separable model for $Y_{it}(\infty)$,

$$Y_{it}(\infty) = \alpha_i + \lambda_t + \varepsilon_{it}, \quad E[\varepsilon_{it}] = 0.$$

Following the same arguments as in Section 2.3.2 and Appendix B.2, all results extend straightforwardly to fully nonseparable models. Selection into groups can depend on the unobservable determinants of the untreated potential outcomes as well as additional unobservables, $\zeta_i = (\nu_i, \eta_{i1}, \dots, \eta_{iT})$,

$$G_i = g(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i).$$

We generalize Propositions 2.1, 2.2, and 2.3 to the multiple-group, multiple-period case. As in the 2×2 case, we denote the set of all selection mechanisms by \mathcal{G}_{all} and consider the following restricted classes of selection mechanisms,

$$\begin{aligned} \mathcal{G}_1 &= \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, \dots, e_T, z) \text{ is a trivial function of } e_T\}, \\ \mathcal{G}'_1 &= \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, \dots, e_T, z) \text{ is a trivial function of } (e_2, \dots, e_T)\}, \\ \mathcal{G}_2 &= \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, \dots, e_T, z) \text{ is a trivial function of } (e_1, \dots, e_T)\}, \end{aligned}$$

where \mathcal{G}_1 and \mathcal{G}'_1 are different extensions of \mathcal{G}_1 in the 2×2 case.

¹²Our setup and notation build on Callaway and Sant’Anna (2021), Sun and Abraham (2021), and Roth et al. (2022).

We give four necessary and sufficient conditions for the parallel trends assumption in Equation (6). First, parallel trends holds for any fixed $g \in \mathcal{G}_{\text{all}}$ if and only if $\varepsilon_{i1} = \dots = \varepsilon_{iT}$. Second, parallel trends for any fixed $g \in \mathcal{G}_1$ implies that $E[\varepsilon_{it}|\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{it-1}] = \varepsilon_{it-1}$ for $t \geq 2$. Third, parallel trends for any fixed $g \in \mathcal{G}'_1$ implies that $E[\varepsilon_{it}|\alpha_i, \varepsilon_{i1}] = E[\varepsilon_{it-1}|\alpha_i, \varepsilon_{i1}]$ for $t \geq 2$. Finally, parallel trends for any fixed $g \in \mathcal{G}_2$ implies $E[\varepsilon_{i1}|\alpha_i] = \dots = E[\varepsilon_{iT}|\alpha_i]$. Under additional conditions on the distribution of $G_i|\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}$, the necessary conditions for $g \in \mathcal{G}_1$, $g \in \mathcal{G}'_1$, and $g \in \mathcal{G}_2$ are also sufficient. All these conditions can be viewed as natural generalizations of the results for the 2×2 case.

3 Primitive sufficient conditions for parallel trends

The results in Section 2 illustrate that restrictions on time-varying unobservables are necessary for parallel trends to hold. If researchers are not willing to rule out that the time-varying unobservable determinants of the untreated potential outcomes enter the selection mechanism in an unrestricted way, then parallel trends implies that the untreated potential outcomes are constant over time up to location shifts (Proposition 2.1). While, naturally, this condition is sufficient for parallel trends to hold, it is implausible in practice. Propositions 2.2 and 2.3 demonstrate necessary conditions that allow the untreated potential outcomes to vary over time beyond location shifts. However, these conditions are only sufficient under further high-level restrictions on the conditional expectation of G_i .

Our analysis in Section 2 thus raises two questions: (i) Can we allow for selection to depend on both ε_{i1} and ε_{i2} , albeit in a restricted way, while allowing $Y_{it}(0)$ to vary across time beyond location shifts? (ii) What are primitive sufficient conditions that imply parallel trends if $g \in \mathcal{G}_1$ and $g \in \mathcal{G}_2$, respectively? The goal of this section is to answer these two questions and illustrate the primitive conditions in the context of two classical examples of selection mechanisms: selection based on untreated potential outcomes and selection based on treatment effects (Roy-style selection).

3.1 Sufficient conditions for parallel trends

The first primitive sufficient condition demonstrates a case where selection depends on both ε_{i1} and ε_{i2} , *and* the untreated potential outcomes can vary across time beyond

location shifts. Define the class of symmetric selection mechanisms as

$$\mathcal{G}_{\text{sym}} = \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, e_2, v, t_1, t_2) \text{ is a symmetric function in } e_1 \text{ and } e_2\}.$$

Assumption SC1. *The following conditions hold:*

1. $G_i = g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2})$, where $g \in \mathcal{G}_{\text{sym}}$.
2. $(\nu_i, \eta_{i1}, \eta_{i2}) | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2} \stackrel{d}{=} (\nu_i, \eta_{i1}, \eta_{i2}) | \alpha_i, \varepsilon_{i2}, \varepsilon_{i1}$.
3. $\varepsilon_{i1}, \varepsilon_{i2} | \alpha_i \stackrel{d}{=} \varepsilon_{i2}, \varepsilon_{i1} | \alpha_i$.

While Assumption SC1.1 allows selection to depend on both ε_{i1} and ε_{i2} , it requires the selection mechanism to be symmetric in them. Assumptions SC1.2 and SC1.3 require two different types of exchangeability restrictions. Assumption SC1.2 requires that the conditional distribution of $(\nu_i, \eta_{i1}, \eta_{i2})$ is exchangeable in ε_{i1} and ε_{i2} after conditioning on α_i . This notion of exchangeability has been employed, for example, in Altonji and Matzkin (2005). By contrast, Assumption SC1.3 requires the distribution of $(\varepsilon_{i1}, \varepsilon_{i2})$ to be exchangeable conditional on α_i .

In sum, Assumption SC1 consists of symmetry restrictions on how ε_{i1} and ε_{i2} enter the selection mechanism and on the distribution of unobservables. These symmetry conditions imply that $E[G_i \varepsilon_{i1} | \alpha_i] = E[G_i \varepsilon_{i2} | \alpha_i]$, where G_i and ε_{it} may be correlated even after conditioning on α_i . Therefore, Assumption SC1 provides a case where parallel trends holds due to the time invariance of the bias from selecting on ε_{it} conditional on α_i .

The next two assumptions provide sufficient conditions for the restricted classes of selection mechanisms \mathcal{G}_1 and \mathcal{G}_2 .

Assumption SC2. *The following conditions hold:*

1. $G_i = g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2})$, where $g \in \mathcal{G}_1$.
2. $(\nu_i, \eta_{i1}, \eta_{i2}) | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2} \stackrel{d}{=} (\nu_i, \eta_{i1}, \eta_{i2}) | \alpha_i, \varepsilon_{i1}$.
3. $E[\varepsilon_{i2} | \alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$.

Assumption SC3. *The following conditions hold:*

1. $G_i = g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2})$, where $g \in \mathcal{G}_2$.
2. $(\nu_i, \eta_{i1}, \eta_{i2}) | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2} \stackrel{d}{=} (\nu_i, \eta_{i1}, \eta_{i2}) | \alpha_i$.

$$3. E[\varepsilon_{i1}|\alpha_i] = E[\varepsilon_{i2}|\alpha_i].$$

Assumptions SC2.3 and SC3.3 impose the necessary conditions for parallel trends to hold when $g \in \mathcal{G}_1$ and $g \in \mathcal{G}_2$, respectively. The remaining conditions in Assumptions SC2 and SC3 provide primitive conditions on $g(\cdot)$ and the conditional distribution of $(\nu_i, \eta_{i1}, \eta_{i2})$ that imply the additional conditional mean independence restriction on G_i in Propositions 2.2 and 2.3, respectively.

The combinations of restrictions in Assumptions SC2 and SC3 imply parallel trends through different channels. Assumption SC2 implies that G_i is conditionally uncorrelated with $(\varepsilon_{i2} - \varepsilon_{i1})$ given $(\alpha_i, \varepsilon_{i1})$, whereas Assumption SC3 implies that G_i is conditionally uncorrelated with $(\varepsilon_{i2} - \varepsilon_{i1})$ given α_i .¹³

The next proposition shows that Assumptions SC1, SC2, and SC3 are indeed sufficient for Assumption PT.

Proposition 3.1 (Sufficient conditions for parallel trends). *Suppose that Assumption SP holds and $P(G_i = 1) \in (0, 1)$. Then (i) Assumption SC1 implies Assumption PT, (ii) Assumption SC2 implies Assumption PT, and (iii) Assumption SC3 implies Assumption PT.*

The key step in the proof of Proposition 3.1 is to show that the conditions imposed on $g(\cdot)$ and the conditional distribution of $(\nu_i, \eta_{i1}, \eta_{i2})$ imply specific properties of the projected selection mechanism, $\bar{g}(a, e_1, e_2) = E[G_i|\alpha_i = a, \varepsilon_{i1} = e_1, \varepsilon_{i2} = e_2]$. In (i), we show that the symmetry conditions on $g(\cdot)$ and the conditional distribution of $(\nu_i, \eta_{i1}, \eta_{i2})$ imply the symmetry of $\bar{g}(\cdot)$ in e_1 and e_2 . Similarly, in (ii) and (iii), we show that the conditions on $g(\cdot)$ and the conditional distribution of $(\nu_i, \eta_{i1}, \eta_{i2})$ imply that $\bar{g}(\cdot)$ is a trivial function of e_2 and (e_1, e_2) , respectively. Together with the restrictions on the distribution of $(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$, these conditions on $\bar{g}(\cdot)$ imply parallel trends by the law of iterated expectations (LIE) and Lemma A.1.

¹³To see this formally, first note that under Assumption SC2 and SC3, $E[\varepsilon_{i2} - \varepsilon_{i1}|\alpha_i, \varepsilon_{i1}] = 0$ and $E[\varepsilon_{i2} - \varepsilon_{i1}|\alpha_i] = 0$, respectively. Now note that under Assumption SC2, $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|\alpha_i, \varepsilon_{i1}] = E[G_i|\alpha_i, \varepsilon_{i1}]E[\varepsilon_{i2} - \varepsilon_{i1}|\alpha_i, \varepsilon_{i1}] = 0$, where the first equality follows from the conditional independence of G_i and ε_{i2} . Under Assumption SC3, $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|\alpha_i] = E[G_i|\alpha_i]E[\varepsilon_{i2} - \varepsilon_{i1}|\alpha_i] = 0$, where the first equality follows from the conditional independence of G_i and $(\varepsilon_{i1}, \varepsilon_{i2})$. Note that under Assumption SC3, unlike under Assumption SC1, G_i will also be uncorrelated with ε_{it} conditional on α_i , specifically $E[G_i\varepsilon_{it}|\alpha_i] = E[G_i|\alpha_i]E[\varepsilon_{it}|\alpha_i]$.

3.2 Illustration based on two examples of selection mechanisms

In this section, we illustrate the primitive sufficient conditions for parallel trends using examples of classical selection mechanisms. We start by considering selection mechanisms where selection into treatment is based on untreated potential outcomes.

Example 1 (Ashenfelter and Card (1985)-style selection mechanisms). *We consider a class of threshold-crossing selection mechanisms inspired by Ashenfelter and Card (1985), who study the effect of training programs on earnings. We generalize their analysis to accommodate different information sets \mathcal{I}_i available to the units when deciding whether to participate in the training program. Specifically, we consider the following mechanism*

$$G_i = 1 \{E[Y_{i1}(0) + \beta Y_{i2}(0)|\mathcal{I}_i] \leq c\} = 1 \{E[(1 + \beta)\alpha_i + \varepsilon_{i1} + \beta\varepsilon_{i2}|\mathcal{I}_i] \leq \tilde{c}\}, \quad (7)$$

where G_i indicates participation in a job training program, $Y_{it}(0)$ denotes the untreated potential earnings, $\beta \in [0, 1]$ is a discount factor, and $\tilde{c} = c - \lambda_1 - \beta\lambda_2$.¹⁴ Following Ashenfelter and Card (1985), we interpret α_i as the “permanent” component and ε_{it} as the “transitory” component of earnings.

Consider first the case where $\mathcal{I}_i = \mathcal{I}_i^f \equiv \{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}\}$ and, therefore, $E[Y_{it}(0)|\mathcal{I}_i^f] = Y_{it}(0)$ for $t = 1, 2$ as in Ashenfelter and Card (1985). Here, selection depends only on the unobservable determinants of the untreated potential outcomes $(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$. Under model (7) with \mathcal{I}_i^f , Assumption SC1.1 requires that there is no discounting, $\beta = 1$, so that selection depends on the sum of $Y_{i1}(0)$ and $Y_{i2}(0)$, $G_i = 1 \{2\alpha_i + \varepsilon_{i1} + \varepsilon_{i2} \leq \tilde{c}\}$. By contrast, Assumption SC2.1 requires that there is full discounting, $\beta = 0$, so that selection depends only on $Y_{i1}(0)$, $G_i = 1 \{\alpha_i + \varepsilon_{i1} \leq \tilde{c}\}$ as in p.651 in Ashenfelter and Card (1985). Finally, a simple example of a selection mechanism satisfying Assumption SC3.1 is $G_i = 1\{\alpha_i \leq c\}$, which corresponds to the selection mechanism on p.650 in Ashenfelter and Card (1985).

If instead $\mathcal{I}_i = \mathcal{I}_i^1 \equiv \{\alpha_i, \varepsilon_{i1}\}$, then the selection mechanism simplifies to $G_i = 1\{Y_{i1}(0) + \beta E[Y_{i2}(0)|\mathcal{I}_i^1] \leq c\} = 1\{(1 + \beta)\alpha_i + \varepsilon_{i1} + \beta E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] \leq \tilde{c}\}$. Regardless of the value of the discount factor β , this selection mechanism only depends on α_i and ε_{i1} and thus is an element of \mathcal{G}_1 , thereby fulfilling Assumption SC2.1.

¹⁴Our sufficient conditions accommodate additional unobservables determining selection. For example, if $G_i = 1 \{(1 + \beta)\alpha_i + \varepsilon_{i1} + \beta E[\varepsilon_{i2}|\mathcal{I}_i] + \nu_i^1 \leq \tilde{c}\}$, where the “random shock” ν_i^1 satisfies Assumption SEL, the distributional assumptions on the additional unobservables (Assumptions SC1.2, SC2.2, and SC3.2) are satisfied.

Finally, if $\mathcal{I}_i = \mathcal{I}_i^0 \equiv \{\alpha_i\}$, then the selection mechanism (7) depends on α_i only and therefore belongs to \mathcal{G}_2 , fulfilling Assumption SC3.1, regardless of the value of β . \square

Next, we consider Roy-style selection based on treatment effects. An important takeaway from Proposition 3.1 is that none of the primitive sufficient conditions impose any restrictions on how the additional unobservables $(\nu_i, \eta_{i1}, \eta_{i2})$ determine selection. This implies that parallel trends can be consistent with general selection mechanisms based on treatment effects, as we illustrate in the following example.

Example 2 (Roy-style selection). *Consider the following random coefficients model for the observed outcome, $Y_{it} = \alpha_i + \delta_{it}D_{it} + \lambda_t + \varepsilon_{it}$. Suppose that selection depends on the expected gains from the treatment $E[\delta_{i2}|\mathcal{I}_i]$ as well as the expected cost of treatment, $E[c_i|\mathcal{I}_i]$,*

$$G_i = 1\{E[\delta_{i2}|\mathcal{I}_i] > E[c_i|\mathcal{I}_i]\}.$$

Suppose that $\mathcal{I}_i = \{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \delta_{i1}, \delta_{i2}, c_i\}$, then this selection mechanism does not depend on $(\varepsilon_{i1}, \varepsilon_{i2})$ so that the conditions on $g(\cdot)$ in Assumptions SC1, SC2, and SC3 hold immediately. We therefore only have to impose the distributional restrictions in Assumptions SC1, SC2, and SC3. Specifically, for Assumption SC1 to hold, the conditional distribution of the treatment effects and costs has to be exchangeable in $(\varepsilon_{i1}, \varepsilon_{i2})$, i.e., $(\delta_{i2}, c_i)|(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) \stackrel{d}{=} (\delta_{i2}, c_i)|(\alpha_i, \varepsilon_{i2}, \varepsilon_{i1})$. For Assumption SC2 to hold, the treatment effects and costs have to be independent of ε_{i2} conditional on $(\alpha_i, \varepsilon_{i1})$, formally $(\delta_{i2}, c_i)|(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) \stackrel{d}{=} (\delta_{i2}, c_i)|(\alpha_i, \varepsilon_{i1})$. Finally, for Assumption SC3 to hold, it is sufficient that the treatment effects and costs are independent of $(\varepsilon_{i1}, \varepsilon_{i2})$ conditional on α_i , that is $(\delta_{i2}, c_i)|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2} \stackrel{d}{=} (\delta_{i2}, c_i)|\alpha_i$.

This example illustrates that the time-varying treatment effects can enter the selection mechanism in an unrestricted way. Indeed, we could replace δ_{i2} by any arbitrary function of δ_{i1} and δ_{i2} without affecting the main implications of the example. As a result, in the context of Roy-style selection where perfect foresight is plausible, researchers do not have to impose any restrictions on how selection depends on the treatment effects or costs. Instead, they have to justify the required distributional restrictions.

If instead the information available to the units is $\mathcal{I}_i = \{\alpha_i, \varepsilon_{i1}\}$ ($\mathcal{I}_i = \{\alpha_i\}$), then the selection mechanism is an element of \mathcal{G}_1 (\mathcal{G}_2), and the researchers have to justify the remaining conditions from Assumption SC2 (SC3). \square

On the one hand, the examples in this section show that parallel trends can be consistent with a wide range of selection mechanisms. On the other hand, because parallel trends is an assumption on the untreated potential outcomes, nontrivial restrictions may be required when selection is based on those outcomes. Furthermore, the examples elucidate the crucial role information sets play in how unobservables determine selection into treatment. We next discuss the practical implications of our theoretical results in light of these examples.

3.3 Implications for practice

Restrictions on selection are unavoidable in DiD designs. Our necessary and sufficient condition in Proposition 2.1 underscores that if researchers are not willing to impose any restrictions on selection, then parallel trends rules out time-varying unobservables. Therefore, in realistic settings, relying on parallel trends assumptions implicitly imposes restrictions on the time-varying unobservables and how selection depends on them.

Parallel trends can be compatible with selection on time-varying unobservables. It is well-understood that selection on time-invariant unobservables is compatible with parallel trends in the classical two-way fixed effects model under strict exogeneity. Our primitive sufficient conditions provide cases where parallel trends could hold despite selection depending on time-invariant *and* time-varying unobservables.

Parallel trends can be compatible with selection on outcomes and treatment effects. Building on the primitive sufficient conditions, we provide two classical examples of selection mechanisms: selection on untreated potential outcomes (Example 1) and selection on treatment effects (Example 2). In Example 1, the restrictions on the selection mechanism correspond to restrictions on the units' information set and on how the second period's outcome is discounted. In Example 2, our sufficient conditions constitute restrictions on the available information set and on how the expected treatment effects depend on the time-varying unobservable determinants of the untreated potential outcome.

Contextual knowledge on selection can be used to justify parallel trends assumptions. The menu of primitive sufficient conditions provides practitioners with explicit theory-based templates for justifying parallel trends assumptions. These conditions consist of different combinations of restrictions on (i) which/how unobservables determine

selection and (ii) how their distribution varies over time. We recommend that applied researchers relying on these conditions use contextual information to assess and explicitly discuss which determinants of the untreated potential outcome affect selection. In doing so, it is crucial to consider the timing of the decision as well as the information set available to the units. Once a suitable selection mechanism is identified, the next step is to discuss the plausibility of the corresponding assumption on the distribution of the unobservables. In this context, periodicity is crucial both to distinguish between time-invariant and time-varying factors and to justify the distributional assumptions.

4 Covariates

In many applications, parallel trends may only be plausible conditional on covariates (e.g., Heckman et al., 1997; Abadie, 2005; Sant’Anna and Zhao, 2020; Callaway and Sant’Anna, 2021). Therefore, we study the role of covariates in parallel trends assumptions. While many existing approaches focus on time-invariant covariates, we explicitly allow for a vector of both time-invariant and time-varying covariates, X_{it} , assuming that X_{it} is not affected by the treatment.

We focus on primitive sufficient conditions for conditional parallel trends assumptions. We examine both separable and nonseparable models and show that this distinction has important implications for the information set under which the conditional parallel trends assumption can be maintained. Finally, we establish connections between conditional parallel trends and identification assumptions in the nonseparable panel data literature.

4.1 Conditional parallel trends assumptions

In Section 4.2, we examine models that are separable in observables and unobservables. We consider the following parallel trends assumption, which is conditional on the time series of the covariates.¹⁵

Assumption PT-X. *The conditional parallel trends assumption holds:*

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1, X_i] = E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0, X_i] \text{ a.s.}$$

¹⁵See Appendix D for a discussion of when unconditional parallel trends (Assumption PT) holds despite covariates entering the outcome equation.

Under Assumption PT-X, the unconditional ATT is identified as

$$\begin{aligned} E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1] &= E[E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1, X_i]|G_i = 1] \\ &= E[E[Y_{i2} - Y_{i1}|G_i = 1, X_i] - E[Y_{i2} - Y_{i1}|G_i = 0, X_i]|G_i = 1] \end{aligned}$$

In Section 4.3, we examine nonseparable models. A simple example is the correlated random coefficients model (e.g., Chamberlain, 1992), $Y_{it}(0) = \alpha_i X_{it}^\mu + \lambda_t + \varepsilon_{it}$. For nonseparable models, it is crucial to differentiate between the covariates that interact with the unobservable determinants of selection, X_{it}^μ , and those that do not, X_{it}^λ . Intuitively, this is because in general we cannot have parallel trends between treatment and control subpopulations that experience changes in X_{it}^μ over time.

Motivated by this discussion, we examine the following modified version of Assumption PT-X.

Assumption PT-NSP. *The (modified) conditional parallel trends assumption holds:*

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1, X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] = E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0, X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \text{ a.s.}$$

Under Assumption PT-NSP, we can no longer identify the ATT, $E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1]$, because we cannot identify the conditional ATT, $E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1, X_i^\lambda, X_i^\mu]$. Instead, we can identify $E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1, X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]$. After integrating out with respect to the distribution of covariates, we can identify the ATT for subpopulations that do not experience changes in X_{it}^μ ,

$$E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1, X_{i1}^\mu - X_{i2}^\mu = 0].$$

Note that if X_{it}^μ is time-invariant, then $X_{i1}^\mu = X_{i2}^\mu$ holds by definition such that Assumptions PT-X and PT-NSP are equivalent.

4.2 Covariates in separable models

Consider the following separable model with covariates.

Assumption SP-X.

$$Y_{it}(0) = \alpha_i + \lambda_t + \gamma_t(X_{it}) + \varepsilon_{it}, \quad E[\varepsilon_{it}] = 0, \quad i = 1, \dots, N, \quad t = 1, 2. \quad (8)$$

Assumption SP-X allows for nonparametric covariate-specific trends, which is a key reason for incorporating covariates in DiD analyses. It nests commonly used parametric specifications such as $\gamma_t(X_{it}) = X'_{it}\beta_t$. Recall that we assume that the treatment does not affect X_{it} . In what follows, we denote the support of X_{it} as \mathcal{X} .

To focus the discussion on the different roles played by the time-varying observable and unobservable determinants of $Y_{it}(0)$, we state our sufficient conditions in terms of the projected selection mechanism,

$$\bar{g}(a, x_1, x_2, e_1, e_2) = E[G_i | \alpha_i = a, X_{i1} = x_1, X_{i2} = x_2, \varepsilon_{i1} = e_1, \varepsilon_{i2} = e_2].$$

Assumption SC1-X. *The following conditions hold:*

1. $\bar{g}(a, x_1, x_2, e_1, e_2)$ is a symmetric function in e_1 and e_2 .
2. $\varepsilon_{i1}, \varepsilon_{i2} | \alpha_i, X_i \stackrel{d}{=} \varepsilon_{i2}, \varepsilon_{i1} | \alpha_i, X_i$.

Assumption SC2-X. *The following conditions hold:*

1. $\bar{g}(a, x_1, x_2, e_1, e_2)$ is a trivial function of e_2 .
2. $E[\varepsilon_{i2} - \varepsilon_{i1} | \alpha_i, \varepsilon_{i1}, X_i] = E[\varepsilon_{i2} - \varepsilon_{i1} | X_i]$.

Assumption SC3-X. *The following conditions hold:*

1. $\bar{g}(a, x_1, x_2, e_1, e_2)$ is a trivial function of e_1 and e_2 .
2. $E[\varepsilon_{i1} | \alpha_i, X_i] = E[\varepsilon_{i2} | \alpha_i, X_i]$.

Assumptions SC1-X, SC2-X, and SC3-X demonstrate that incorporating time-varying covariates makes the restrictions on the selection mechanism more plausible. Specifically, none of the assumptions impose any restrictions on how the time-varying covariates determine selection. Assumptions SC1-X.2, SC2-X.2, and SC3-X.2 are conditional versions of Assumptions SC1.3, SC2.3, and SC3.3, respectively.¹⁶ Conditioning on covariates weakens those distributional restrictions, since they are more likely to be satisfied once we focus on subpopulations with the same evolution of time-varying covariates.

The following proposition shows that Assumptions SC1-X, SC2-X, and SC3-X are sufficient for Assumption PT-X.

¹⁶To see that Assumption SC2-X.2 is the conditional version of Assumption SC2.3, note that the latter can be equivalently stated as $E[\varepsilon_{i2} - \varepsilon_{i1} | \alpha_i, \varepsilon_{i1}] = E[\varepsilon_{i2} - \varepsilon_{i1}]$, given the normalization $E[\varepsilon_{it}] = 0$.

Proposition 4.1. *Suppose that Assumption SP-X holds and $P(G_i = 1|X_i) \in (0, 1)$ a.s. Then (i) Assumption SC1-X implies Assumption PT-X, (ii) Assumption SC2-X implies Assumption PT-X, and (iii) Assumption SC3-X implies Assumption PT-X.*

The results in this section have implications for the choice of covariates to be included in DiD analyses. Proposition 4.1 provides several avenues for justifying the inclusion of covariates in DiD analyses. A key takeaway from Proposition 4.1 is that time-invariant and time-varying covariates play different roles in ensuring that Assumption PT-X holds. Any (observable) time-varying factors that asymmetrically affect selection should be included as covariates. In addition, practitioners should include time-invariant and time-varying covariates that render the distributional restrictions plausible in their application.

All sufficient conditions in Proposition 4.1 allow for selection on unobservable determinants of the untreated potential outcome. This is in contrast with the unconfoundedness assumptions commonly used in cross-sectional studies (e.g., Imbens, 2004; Imbens and Wooldridge, 2009). Therefore, these results elucidate the differences between conditional parallel trends and unconfoundedness-type assumptions.

4.3 Covariates in nonseparable models

The following nonseparable model nests the models in Assumptions SP and SP-X.

Assumption NSP-X.

$$Y_{it}(0) = \mu(X_{it}^\mu, \alpha_i^\mu, \varepsilon_{it}^\mu) + \lambda_t(X_{it}^\lambda, \alpha_i^\lambda, \varepsilon_{it}^\lambda), \quad i = 1, \dots, N, \quad t = 1, 2,$$

where X_{it}^μ , X_{it}^λ , α_i^μ , α_i^λ , ε_{it}^μ , and ε_{it}^λ are finite-dimensional random vectors.

The above model consists of time-invariant and time-varying nonseparable components. Without further restrictions on the unobservables, the additive structure in Assumption NSP-X is without loss of generality and the superscripts μ and λ are merely labels. Indeed, if $X_{it}^\mu = X_{it}^\lambda$, $\alpha_i^\mu = \alpha_i^\lambda$, and $\varepsilon_{it}^\mu = \varepsilon_{it}^\lambda$, the model is fully nonseparable and time-varying in an arbitrary way. In the following, we use \mathcal{X}_μ , \mathcal{X}_λ , \mathcal{A} , and \mathcal{E} to denote the supports of X_{it}^μ , X_{it}^λ , α_i^μ , and ε_{it}^μ , respectively.

In view of our analysis of the separable models, it is natural to consider selection based on the unobservables entering $\mu(\cdot)$, since they can be viewed as the counterparts

of the unobservables in the separable model.¹⁷ We therefore impose the following condition on the projected selection mechanism.

Assumption SEL-CI-X.

$$E[G_i|\alpha_i^\mu, \alpha_i^\lambda, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda] = E[G_i|\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu].$$

Assumption SEL-CI-X allows the projected selection mechanism to depend on all covariates, but only on the unobservables that enter the time-invariant component of the structural function. In view of Assumption SEL-CI-X, we define

$$\begin{aligned} & \bar{g}(a^\mu, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_1^\mu, e_2^\mu) \\ &= E[G_i|\alpha_i^\mu = a^\mu, X_{i1}^\mu = x_1^\mu, X_{i2}^\mu = x_2^\mu, X_{i1}^\lambda = x_1^\lambda, X_{i2}^\lambda = x_2^\lambda, \varepsilon_{i1}^\mu = e_1^\mu, \varepsilon_{i2}^\mu = e_2^\mu]. \end{aligned}$$

We present three sets of sufficient conditions for Assumption PT-NSP. Each set of conditions consists of assumptions on the projected selection mechanism as well as distributional restrictions on the unobservables. Our first sufficient condition allows selection to depend on all covariates as well as the unobservables that enter the time-invariant component of the structural function, while imposing a symmetry restriction on the projected selection mechanism similar to Assumptions SC1 and SC1-X.

Assumption SC1-NSP. *The following conditions hold:*

1. $\bar{g}(a^\mu, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_1^\mu, e_2^\mu)$ is a symmetric function in e_1^μ and e_2^μ .
2. $(\varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)|\alpha_i^\mu, X_i^\mu, X_i^\lambda \stackrel{d}{=} (\varepsilon_{i2}^\mu, \varepsilon_{i1}^\mu)|\alpha_i^\mu, X_i^\mu, X_i^\lambda$.
3. $(\alpha_i^\mu, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu) \perp (\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda)|X_i^\mu, X_i^\lambda$.

Here we require the conditional distribution of $(\varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)|\alpha_i^\mu, X_i^\mu, X_i^\lambda$ to be exchangeable. Since the projected selection mechanism depends on $(\alpha_i^\mu, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)$, we require them to be independent of the unobservables entering $\lambda_t(\cdot)$ conditional on (X_i^μ, X_i^λ) .

As noted earlier, the exchangeability restriction in Assumption SC1-NSP is different from the exchangeability assumption in Altonji and Matzkin (2005). The exchangeability assumption in Altonji and Matzkin (2005) requires the conditional distribution of all unobservables that enter $\mu(\cdot)$ and $\lambda_t(\cdot)$ to be invariant to permutations of covariates in the conditioning set, which is a nonparametric correlated

¹⁷To see this, note that the separable model in Assumption SP-X is nested in Assumption NSP-X by setting $\mu(X_{it}^\mu, \alpha_i^\mu, \varepsilon_{it}^\mu) = \alpha_i^\mu + \varepsilon_{it}^\mu$ and $\lambda_t(X_{it}^\lambda, \alpha_i^\lambda, \varepsilon_{it}^\lambda) = \lambda_t + \gamma_t(X_{it}^\lambda)$.

random effects restriction (Ghanem, 2017). By contrast, we assume that the time-varying unobservables are exchangeable conditional on $(\alpha_i^\mu, X_i^\mu, X_i^\lambda)$ without imposing any restrictions on the distribution of $\alpha_i^\mu | G_i, X_i^\mu, X_i^\lambda$.

Next, in the spirit of Assumptions SC2 and SC2-X, we consider a projected selection mechanism that is a trivial function of ε_{i2}^μ in the following sufficient condition.

Assumption SC2-NSP. *The following conditions hold:*

1. $\bar{g}(a^\mu, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_1^\mu, e_2^\mu)$ is a trivial function of e_2^μ .
2. $(\alpha_i^\mu, \varepsilon_{i1}^\mu) \perp \Delta_{\mu,i} | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu$, where $\Delta_{\mu,i} \equiv \mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu)$.
3. $(\alpha_i^\mu, \varepsilon_{i1}^\mu) \perp (\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda) | X_i^\mu, X_i^\lambda$.

Assumption SC2-NSP.2 implicitly imposes separability conditions on $\mu(\cdot)$ and restrictions on time series dependence. The independence condition in Assumption SC2-NSP.3 requires that the unobservable determinants of selection are independent of the unobservables that enter $\lambda_t(\cdot)$, conditional on the times series of the covariates.

The last sufficient condition restricts the projected selection mechanism to only depend on covariates and the time-invariant unobservables.

Assumption SC3-NSP. *The following conditions hold:*

1. $\bar{g}(a^\mu, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_1^\mu, e_2^\mu)$ is a trivial function of e_1^μ and e_2^μ .
2. $\varepsilon_{i1}^\mu | \alpha_i^\mu, X_i^\mu, X_i^\lambda \stackrel{d}{=} \varepsilon_{i2}^\mu | \alpha_i^\mu, X_i^\mu, X_i^\lambda$.
3. $\alpha_i^\mu \perp (\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda) | X_i^\mu, X_i^\lambda$.

Assumption SC3-NSP requires the distribution of ε_{it}^μ , which enters $\mu(\cdot)$, to be time-invariant conditional on $(\alpha_i^\mu, X_i^\mu, X_i^\lambda)$. The unobservables entering $\lambda_t(\cdot)$, $(\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda)$, are required to be independent of the unobservables that determine selection, α_i^μ , conditional on (X_i^μ, X_i^λ) .

Each of the sufficient conditions consists of three components: (i) a restriction on how/which unobservables determine the projected selection mechanism, (ii) a restriction on the unobservables entering the time-invariant component of the structural function, and (iii) an independence assumption that ensures that the time-varying component of the structural function is independent of G_i conditional on the time series of the covariates.

The following proposition formally establishes sufficiency of each set of conditions.

Proposition 4.2. *Suppose that Assumptions NSP-X and SEL-CI-X hold and $P(G_i = 1 | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu) \in (0, 1)$ a.s. Then (i) Assumption SC1-NSP implies Assumption PT-NSP, (ii) Assumption SC2-NSP implies Assumption PT-NSP, and (iii) Assumption SC3-NSP implies Assumption PT-NSP.*

4.4 Selection, fixed effects, and correlated random effects

DiD methods have traditionally been motivated using two-way fixed effects models. Fixed effects assumptions allow for unrestricted dependence between time-invariant unobservables and the regressors in separable and nonseparable models, thereby implicitly allowing for selection on time-invariant unobservables.¹⁸ In this paper, we explicitly analyze the connection between selection mechanisms and the parallel trends assumptions underlying DiD. Therefore, a natural question is how our sufficient conditions relate to the identification assumptions in the nonseparable panel literature.

The literature on nonseparable panel models has considered two broad categories of identification assumptions. First, time homogeneity conditions (e.g., Hoderlein and White, 2012; Chernozhukov et al., 2013) require the distribution of time-varying unobservables to be stationary across time while allowing for unrestricted individual heterogeneity (fixed effects). Second, nonparametric correlated random effects restrictions (e.g., Altonji and Matzkin, 2005; Bester and Hansen, 2009) allow for unrestricted time heterogeneity by imposing restrictions on individual heterogeneity, generalizing the classical notion of correlated random effects (e.g., Mundlak, 1978; Chamberlain, 1984). However, neither category of assumptions is explicit about the selection mechanism and, in particular, about how unobservables determine selection.

The existing identification results based on time homogeneity or correlated random effects assumptions suggest a trade-off between restrictions on time and individual heterogeneity. Here we show that our sufficient conditions for parallel trends constitute interpretable primitive conditions on the selection mechanism that imply *combinations of* time homogeneity and correlated random effects restrictions from the nonseparable panel literature.

The following assumption is the time homogeneity assumption from Chernozhukov et al. (2013) imposed on ε_{it}^μ in Assumption NSP-X, conditional on the time series of all covariates that enter the outcome equation.

¹⁸See, e.g., Chamberlain (1984); Arellano (2003); Evdokimov (2010); Wooldridge (2010); Hoderlein and White (2012); Chernozhukov et al. (2013).

Assumption TH. $\varepsilon_{i1}^\mu | G_i, X_i^\mu, X_i^\lambda, \alpha_i^\mu \stackrel{d}{=} \varepsilon_{i2}^\mu | G_i, X_i^\mu, X_i^\lambda, \alpha_i^\mu$

Assumption TH requires the distribution of ε_{it}^μ to be homogeneous across time conditional on $G_i, X_i^\mu, X_i^\lambda$, and α_i^μ . However, it does not impose any restrictions on the conditional distribution of ε_{it}^μ . Furthermore, there are no restrictions imposed on the distribution of $\alpha_i^\mu | G_i, X_i^\mu, X_i^\lambda$, consistent with the notion of fixed effects.

The next assumption is a nonparametric correlated random effects assumption (e.g., Altonji and Matzkin, 2005; Ghanem, 2017).

Assumption CRE. $(\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda) | G_i, X_i^\mu, X_i^\lambda \stackrel{d}{=} (\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda) | X_i^\mu, X_i^\lambda$.

Assumption CRE is a conditional independence condition between G_i and the unobservables that enter the time-varying component of the structural function, $\lambda_t(\cdot)$. This assumption does not imply conditional random assignment, $(Y_{i1}(0), Y_{i2}(0)) \perp G_i | X_i^\mu, X_i^\lambda$, since selection into treatment can depend on the unobservables entering the time-invariant component $\mu(\cdot)$.

It is straightforward to show that Assumptions TH and CRE imply Assumption PT-NSP (see Appendix C for the formal result). In the following proposition, we show that Assumptions SC1-NSP and SC3-NSP are primitive sufficient conditions on the selection mechanism for the nonseparable model satisfying Assumptions TH and CRE.¹⁹ This result demonstrates how restrictions on selection can be used to justify combinations of Assumptions TH and CRE.

Proposition 4.3. *Suppose that Assumption NSP-X holds. Suppose further that $G_i = g(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)$. Then (i) Assumption SC1-NSP with $g(\cdot)$ in lieu of $\bar{g}(\cdot)$ implies Assumptions TH and CRE if $P(G_i = 1 | \alpha_i^\mu, X_i^\mu, X_i^\lambda) \in (0, 1)$ a.s., (ii) Assumption SC3-NSP with $g(\cdot)$ in lieu of $\bar{g}(\cdot)$ implies Assumptions TH and CRE.*

Proposition 4.3 sheds light on the connection between selection, fixed effects, and correlated random effects in the nonseparable DiD framework. On the one hand, Assumptions SC1-NSP and SC3-NSP allow the distribution of $\alpha_i^\mu | G_i, X_i^\mu, X_i^\lambda$ to be unrestricted, consistent with the notion of fixed effects. On the other hand, both conditions require $(\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda)$ to be conditionally independent of the unobservable determinants of selection and therefore conditionally independent of G_i , consistent with the notion of correlated random effects.

¹⁹In the context of correlated random coefficient models, Graham and Powell (2012) impose a similar structure on their model.

4.5 Implications for practice

Time-varying covariates weaken the assumptions required for parallel trends. Our theoretical results clarify how covariates can weaken restrictions on selection. Specifically, time-varying covariates do not have to obey the strong symmetry and exclusion restrictions required for time-varying unobservables. Thus, researchers should include time-varying factors that asymmetrically determine selection into treatment as covariates, assuming they are unaffected by the treatment. Moreover, conditioning on covariates makes the distributional restrictions on the unobservables more plausible.

How to condition on covariates depends on how they enter the outcome model. By analyzing parallel trends through the lens of nonseparable panel models, we demonstrate the implications of separability restrictions on the outcome model for how researchers should condition on time-varying covariates in their DiD analyses. If covariates and unobservable determinants of selection enter the outcome model separably, researchers should condition on the entire time series of covariates. If, in addition, there are time-varying covariates that interact with the unobservable determinants of selection in the outcome model, researchers have to condition on these covariates not changing over time.

Restrictions on nonseparable outcome models can also be used to justify parallel trends. An implication of Section 4.4 is that parallel trends is consistent with a nonseparable outcome model satisfying a combination of time homogeneity and correlated random effects (see also Appendix C). This provides researchers with an alternative avenue for justifying parallel trends based on restrictions on the untreated potential outcome and its unobservable determinants.

References

- Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1):1–19.
- Altonji, J. G. and Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73(4):1053–1102.
- Angrist, J. D. and Krueger, A. B. (1999). Chapter 23 - empirical strategies in labor economics. volume 3 of *Handbook of Labor Economics*, pages 1277–1366. Elsevier.
- Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.

- Arellano, M. and Bonhomme, S. (2011). Nonlinear panel data analysis. *Annual Review of Economics*, 3:395–424.
- Arellano, M. and Bonhomme, S. (2012). Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies*, 79(3):987–1020.
- Arellano, M. and Bonhomme, S. (2016). Nonlinear panel data estimation via quantile regressions. *The Econometrics Journal*, 19(3):C61–C94.
- Arellano, M. and Honoré, B. (2001). Panel data models: Some recent developments. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5. Elsevier Science.
- Arkhangelsky, D. and Imbens, G. W. (2022). Doubly robust identification for causal panel data models. *The Econometrics Journal*, 25(3):649–674.
- Arkhangelsky, D., Imbens, G. W., Lei, L., and Luo, X. (2021). Double-robust two-way-fixed-effects regression for panel data. *arXiv:2107.13737 [econ]*.
- Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, 60(1):47–57.
- Ashenfelter, O. C. and Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, 67(4):648–660.
- Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix Completion Methods for Causal Panel Data Models. *Journal of the American Statistical Association*, 116(536):1716–1730.
- Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497.
- Bester, C. A. and Hansen, C. (2009). Identification of marginal effects in a non-parametric correlated random effects model. *Journal of Business and Economic Statistics*, 27(2):235–250.
- Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting Event Study Designs: Robust and Efficient Estimation. *arXiv:2108.12419 [econ]*.
- Caetano, C., Callaway, B., Payne, S., and Rodrigues, H. S. (2022). Difference in Differences with Time-Varying Covariates. *arXiv:2202.02903*.
- Callaway, B. and Sant’Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.
- Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *ILR*

- Review*, 43(2):245–257.
- Card, D. and Hyslop, D. R. (2005). Estimating the effects of a time-limited earnings subsidy for welfare-leavers. *Econometrica*, 73(6):1723–1770.
- Card, D. and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4):772–793.
- Chabé-Ferret, S. (2015). Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes. *Journal of Econometrics*, 185(1):110–123.
- Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5–46.
- Chamberlain, G. (1984). Chapter 22 panel data. volume 2 of *Handbook of Econometrics*, pages 1247–1318. Elsevier.
- Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica*, 60(3):567–596.
- Chernozhukov, V., Fernández-Val, I., Hahn, J., and Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2):535–580.
- de Chaisemartin, C. and D’Haultfœuille, X. (2018). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, 85(2):999–1028.
- de Chaisemartin, C. and D’Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996.
- de Chaisemartin, C. and D’Haultfœuille, X. (2021). Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey. SSRN Scholarly Paper ID 3980758, Social Science Research Network, Rochester, NY.
- Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. *Department of Economics, Princeton University*, 1.
- Freyberger, J. (2017). Non-parametric Panel Data Models with Interactive Fixed Effects. *The Review of Economic Studies*, 85(3):1824–1851.
- Gardner, J. (2021). Two-stage differences in differences. *Working Paper*.
- Ghanem, D. (2017). Testing identifying assumptions in nonseparable panel data models. *Journal of Econometrics*, 197(2):202–217.
- Graham, B. S. and Powell, J. L. (2012). Identification and estimation of average par-

- tial effects in “irregular” correlated random coefficient panel data models. *Econometrica*, 80(5):2105–2152.
- Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 64(4):605–654.
- Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1):239–267.
- Hoderlein, S. and White, H. (2012). Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics*, 168(2):300–314.
- Honoré, B. and Kyriazidou, E. (2000a). Estimation of Tobit-type models with individual specific effects. *Econometric Reviews*, 19:341–366.
- Honoré, B. E. (1993). Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics*, 59(1–2):35–61.
- Honoré, B. E. and Kyriazidou, E. (2000b). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874.
- Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *The Review of Economics and Statistics*, 86(1):4–29.
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.
- Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica*, 65(6):1335–1364.
- Lechner, M. (2010). The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends in Econometrics*, 4(3):165–224.
- Lehmann, E. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics.
- Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55(2):357–362.
- Marcus, M. and Sant’Anna, P. H. C. (2021). The role of parallel trends in event study settings: An application to environmental economics. *Journal of the Association of Environmental and Resource Economists*, 8(2):235–275.
- Marx, P., Tamer, E., and Tang, X. (2022). Parallel trends and dynamic choices. *arXiv:2207.06564*.

- McKenzie, D. (2022). A new synthesis and key lessons from the recent difference-in-differences literature. World Bank Blogs (Link). Accessed: 2022-02-22.
- Meyer, B. D., Viscusi, W. K., and Durbin, D. L. (1995). Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment. *The American Economic Review*, 85(3):322–340.
- Mundlak, Y. (1961). Empirical production function free of management bias. *Journal of Farm Economics*, 43(1):44–56.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.
- Roth, J. and Sant’Anna, P. H. C. (2021). When is parallel trends sensitive to functional form? *arXiv:2010.04814 [econ, stat]*.
- Roth, J., Sant’Anna, P. H. C., Bilinski, A., and Poe, J. (2022). What’s trending in difference-in-differences? A synthesis of the recent econometrics literature. *arXiv:2201.01194 [econ]*.
- Sant’Anna, P. H. C. and Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122.
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.
- Verdier, V. (2020). Average treatment effects for stayers with correlated random coefficient models of panel data. *Journal of Applied Econometrics*, 35(7):917–939.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. The MIT Press, Cambridge, MA and London, England.
- Wooldridge, J. M. (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. *Working Paper*.

A Proofs

A.1 Proof of Proposition 2.1

“ \implies ”: We show that if Assumption PT holds for any fixed $g \in \mathcal{G}_{\text{all}}$, then $\varepsilon_{i1} = \varepsilon_{i2}$ a.s. Since Assumption PT holds for any fixed $g \in \mathcal{G}_{\text{all}}$, then it holds for

$$\check{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{\varepsilon_{i2} - \varepsilon_{i1} \geq 0\}.$$

By Lemma A.1,²⁰ Assumption PT holding for $\check{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{\varepsilon_{i2} - \varepsilon_{i1} \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0, \quad (9)$$

which, by Assumption SEL, implies

$$E[1\{\varepsilon_{i2} - \varepsilon_{i1} \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0. \quad (10)$$

Since $E[\varepsilon_{i2} - \varepsilon_{i1}] = 0$ by assumption, the above equality implies $\varepsilon_{i2} - \varepsilon_{i1} = 0$ a.s. by Lemma A.2.

“ \Leftarrow ”:
We show that if $\varepsilon_{i1} = \varepsilon_{i2}$ a.s., then Assumption PT holds. This is immediate, since if $\varepsilon_{i1} = \varepsilon_{i2}$ a.s., then $G_i(\varepsilon_{i2} - \varepsilon_{i1}) = 0$ a.s. As a result, $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$, which is equivalent to Assumption PT by Lemma A.1. This completes the proof. \square

A.2 Proof of Proposition 2.2

“ \Rightarrow ”:
We show that if Assumption PT holds for any fixed $g \in \mathcal{G}_1$, then $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$ a.s. Since Assumption PT holds for any fixed $g \in \mathcal{G}_1$, then it holds for

$$\check{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} \geq 0\}.$$

By Lemma A.1, Assumption PT holding for $\check{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0, \quad (11)$$

which, by Assumption SEL, is equivalent to

$$E[1\{E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0. \quad (12)$$

By the LIE, this is further equivalent to

$$E[1\{E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} \geq 0\}(E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1})] = 0. \quad (13)$$

²⁰Note that under Assumption SEL, requiring $P(G_i = 1) \in (0, 1)$ for selection mechanism $\check{g}(\cdot)$ implies that $P(\varepsilon_{i2} \geq \varepsilon_{i1}) > 0$.

Since $E[E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1}] = 0$ by assumption, the above equality implies $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} = 0$ a.s. by Lemma A.2.

“ \Leftarrow ”:
We show that if, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i, \varepsilon_{i1}]$ a.s., then $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$ a.s. is also sufficient for Assumption PT. Note that

$$\begin{aligned} E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] &= E[E[E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}](\varepsilon_{i2} - \varepsilon_{i1})|\alpha_i, \varepsilon_{i1}]] \\ &= E[E[E[G_i|\alpha_i, \varepsilon_{i1}](\varepsilon_{i2} - \varepsilon_{i1})|\alpha_i, \varepsilon_{i1}]] \\ &= E[E[G_i|\alpha_i, \varepsilon_{i1}](E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1})] = 0. \end{aligned}$$

The first equality follows by the LIE. The second equality follows by the conditional mean independence restriction imposed on G_i . The third equality follows since $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$ a.s. implies that $E[G_i|\alpha_i, \varepsilon_{i1}](E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1}) = 0$ a.s. As a result, its expectation is zero. Since $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$ is equivalent to Assumption PT by Lemma A.1, this implies the result. The proof is complete. \square

A.3 Proof of Proposition 2.3

“ \implies ”:
We show that if Assumption PT holds for any $g \in \mathcal{G}_2$, then $E[\varepsilon_{i1}|\alpha_i] = E[\varepsilon_{i2}|\alpha_i]$ a.s. Since Assumption PT holds for any fixed $g \in \mathcal{G}_2$, then it holds for

$$\check{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{E[\varepsilon_{i2}|\alpha_i] - E[\varepsilon_{i1}|\alpha_i] \geq 0\}.$$

By Lemma A.1, Assumption PT holding for $\check{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{E[\varepsilon_{i2}|\alpha_i] - E[\varepsilon_{i1}|\alpha_i] \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0, \quad (14)$$

which, by Assumption SEL, is equivalent to

$$E[1\{E[\varepsilon_{i2}|\alpha_i] - E[\varepsilon_{i1}|\alpha_i] \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0. \quad (15)$$

By the LIE, this is further equivalent to

$$E[1\{E[\varepsilon_{i2}|\alpha_i] - E[\varepsilon_{i1}|\alpha_i] \geq 0\}(E[\varepsilon_{i2}|\alpha_i] - E[\varepsilon_{i1}|\alpha_i])] = 0. \quad (16)$$

Since $E[E[\varepsilon_{i2} - \varepsilon_{i1}|\alpha_i]] = 0$ by assumption, the above equality implies that $E[\varepsilon_{i2}|\alpha_i] - E[\varepsilon_{i1}|\alpha_i] = 0$ a.s. by Lemma A.2.

“ \Leftarrow ”: We show that if, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i]$ a.s., then $E[\varepsilon_{i1}|\alpha_i] = E[\varepsilon_{i2}|\alpha_i]$ a.s. is sufficient for Assumption PT. Note that

$$\begin{aligned} E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] &= E[E[E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}](\varepsilon_{i2} - \varepsilon_{i1})|\alpha_i]] \\ &= E[E[E[G_i|\alpha_i](\varepsilon_{i2} - \varepsilon_{i1})|\alpha_i]] \\ &= E[E[G_i|\alpha_i](E[\varepsilon_{i2}|\alpha_i] - E[\varepsilon_{i1}|\alpha_i])] = 0. \end{aligned}$$

The first equality follows by the LIE. The second equality follows by the conditional mean independence restriction imposed on G_i . The third equality follows, since $E[\varepsilon_{i1}|\alpha_i] = E[\varepsilon_{i2}|\alpha_i]$ a.s. implies that $E[G_i|\alpha_i](E[\varepsilon_{i2}|\alpha_i] - E[\varepsilon_{i1}|\alpha_i]) = 0$ a.s. It therefore has zero expectation, which implies the equivalent condition of Assumption PT in Lemma A.1. This completes the proof. \square

A.4 Proof of Proposition 3.1

We prove the three statements separately. By Lemma A.1, it suffices to show that $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$.

(i) We first show that Assumptions SC1.1-SC1.2 imply the symmetry of $\bar{g}(a, e_1, e_2) = E[G_i|\alpha_i = a, \varepsilon_{i1} = e_1, \varepsilon_{i2} = e_2]$ in e_1 and e_2 . To do so, we note that Assumptions SC1.1-SC1.2 imply the following for $(a, e_1, e_2) \in \mathcal{A} \times \mathcal{E}^2$

$$\begin{aligned} \bar{g}(a, e_1, e_2) &= \int g(a, e_1, e_2, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2}|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}}(v, t_1, t_2|a, e_1, e_2) \\ &= \int g(a, e_2, e_1, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2}|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}}(v, t_1, t_2|a, e_2, e_1) = \bar{g}(a, e_2, e_1), \end{aligned} \quad (17)$$

where the penultimate equality follows by the symmetry of $g(\cdot)$ and $F_{\nu_i, \eta_{i1}, \eta_{i2}|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}}$ in ε_{i1} and ε_{i2} imposed in Assumptions SC1.1 and SC1.2, respectively.

Next, by the LIE, we can decompose $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})]$ and then invoke the symmetry restrictions on $\bar{g}(\cdot)$ and $F_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}$ implied by Assumptions SC1.1–SC1.2 and Assumption SC1.3, respectively:

$$\begin{aligned}
E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] &= E[E[\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i2}|\alpha_i] - E[\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i1}|\alpha_i]] \\
&= \int \left(\int \bar{g}(a, e_1, e_2)e_2 dF_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}(e_1, e_2|a) - \int \bar{g}(a, e_1, e_2)e_1 dF_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}(e_1, e_2|a) \right) dF_{\alpha_i}(a) \\
&= \int \left(\int \bar{g}(a, e_2, e_1)e_2 dF_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}(e_2, e_1|a) - \int \bar{g}(a, e_1, e_2)e_1 dF_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}(e_1, e_2|a) \right) dF_{\alpha_i}(a) = 0.
\end{aligned}$$

The second equality follows from the symmetry restrictions on $\bar{g}(\cdot)$ and $F_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}$. Together, they imply that both conditional expectations in the parentheses equal $E[\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i1}|\alpha_i]$, and therefore the difference between them is zero. As a result, Assumption SC1 implies Assumption PT.

(ii) We first show that Assumptions SC2.1 and SC2.2 imply that $\bar{g}(a, e_1, e_2)$ is a trivial function of e_2 . To do so, we note that Assumptions SC2.1 and SC2.2 imply the following for $(a, e_1, e_2, e'_2) \in \mathcal{A} \times \mathcal{E}^3$, $e_2 \neq e'_2$,

$$\begin{aligned}
\bar{g}(a, e_1, e_2) &= \int g(a, e_1, e_2, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2}|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}}(v, t_1, t_2|a, e_1, e_2) \\
&= \int g(a, e_1, e'_2, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2}|\alpha_i, \varepsilon_{i1}}(v, t_1, t_2|a, e_1) = \bar{g}(a, e_1, e'_2), \tag{18}
\end{aligned}$$

where the penultimate equality follows from Assumption SC2.1, the definition of a trivial function, and the conditional independence assumption in Assumption SC2.2. As a result, we can define $\check{g}(a, e_1) = \bar{g}(a, e_1, e_2)$.

Next, by the LIE, we can decompose $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})]$ as follows,

$$\begin{aligned}
E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] &= E[E[\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1})|\alpha_i, \varepsilon_{i1}]] \\
&= E[\check{g}(\alpha_i, \varepsilon_{i1})(E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1})] = 0, \tag{19}
\end{aligned}$$

where the second equality follows since $\bar{g}(\cdot)$ is a trivial function of ε_{i2} implied by Assumptions SC2.1-SC2.2. The last equality follows from Assumption SC2.3. As a result, Assumption SC2 implies Assumption PT.

(iii) We first show that Assumptions SC3.1-SC3.2 imply that $\bar{g}(a, e_1, e_2)$ is a trivial function of e_1 and e_2 . To do so, we show that Assumptions SC3.1-SC3.2 imply the

following for $(a, e_1, e_2, e'_1, e'_2) \in \mathcal{A} \times \mathcal{E}^4$, where $e_1 \neq e'_1$ and $e_2 \neq e'_2$,

$$\begin{aligned} \bar{g}(a, e_1, e_2) &= \int g(a, e_1, e_2, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2} | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}}(v, t_1, t_2 | a, e_1, e_2) \\ &= \int g(a, e'_1, e'_2, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2} | \alpha_i}(v, t_1, t_2 | a) = \bar{g}(a, e'_1, e'_2), \end{aligned} \quad (20)$$

where the penultimate equality follows by Assumption SC3.1, the definition of a trivial function, and the conditional independence assumption imposed in Assumption SC3.2. As a result, we can define $\check{g}(a) = \bar{g}(a, e_1, e_2)$.

Next, by the LIE, we can decompose $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})]$ as follows,

$$E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = E[E[\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) | \alpha_i]] = E[\check{g}(\alpha_i)E[\varepsilon_{i2} - \varepsilon_{i1} | \alpha_i]] = 0. \quad (21)$$

The second equality follows since $\bar{g}(\cdot)$ is a trivial function of ε_{i1} and ε_{i2} , and the last follows by Assumption SC3.3. Thus, Assumption SC3 implies Assumption PT. This completes the proof. \square

A.5 Proof of Proposition 4.1

In this proof, all equalities involving random variables are understood to hold a.s. By Lemma A.3, it suffices to show that each assumption implies $E[G_i(\varepsilon_{i2} - \varepsilon_{i1}) | X_i] = E[G_i | X_i]E[\varepsilon_{i2} - \varepsilon_{i1} | X_i]$.

(i) The exchangeability restrictions in Assumption SC1-X imply the following:

$$\begin{aligned} &E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i1} | \alpha_i = a, X_i = (x_1, x_2)] \\ &= \int \bar{g}(a, x_1, x_2, e_1, e_2) e_1 dF_{\varepsilon_{i1}, \varepsilon_{i2} | \alpha_i, X_i}(e_1, e_2 | a, x_1, x_2) \\ &= \int \bar{g}(a, x_1, x_2, e_2, e_1) e_1 dF_{\varepsilon_{i1}, \varepsilon_{i2} | \alpha_i, X_i}(e_2, e_1 | a, x_1, x_2) \\ &= E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i2} | \alpha_i = a, X_i = (x_1, x_2)], \end{aligned} \quad (22)$$

a.e. $(a, x_1, x_2) \in \mathcal{A} \times \mathcal{X}^2$.

Integrating out $\alpha_i | X_i$ in the above yields the following a.e. equality:

$$\begin{aligned}
& \int E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i1} | \alpha_i = a, X_i = (x_1, x_2)] dF_{\alpha_i | X_i}(a | (x_1, x_2)) \\
&= \int E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i2} | \alpha_i = a, X_i = (x_1, x_2)] dF_{\alpha_i | X_i}(a | (x_1, x_2)). \quad (23)
\end{aligned}$$

As a result, by the LIE, we have that $E[G_i(\varepsilon_{i2} - \varepsilon_{i1}) | X_i] = 0$. This completes the proof, since by Assumption SC1-X.2 $\varepsilon_{i1} | X_i \stackrel{d}{=} \varepsilon_{i2} | X_i$ by Lemma A.4 and therefore $E[\varepsilon_{i2} - \varepsilon_{i1} | X_i] = 0$.

(ii) Since under Assumption SC2-X, $\bar{g}(\cdot)$ is a trivial function of ε_{i2} , we can define $\check{\bar{g}}(a, x_1, x_2, e_1) = \bar{g}(a, x_1, x_2, e_1, e_2)$. Note that

$$\begin{aligned}
E[G_i(\varepsilon_{i2} - \varepsilon_{i1}) | X_i] &= E[E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) | X_i, \alpha_i, \varepsilon_{i1}] | X_i] \\
&= E[\check{\bar{g}}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}) E[\varepsilon_{i2} - \varepsilon_{i1} | X_i, \alpha_i, \varepsilon_{i1}] | X_i] \\
&= E[\check{\bar{g}}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}) E[\varepsilon_{i2} - \varepsilon_{i1} | X_i] | X_i] \\
&= E[G_i | X_i] E[\varepsilon_{i2} - \varepsilon_{i1} | X_i], \quad (24)
\end{aligned}$$

where the first equality follows by the LIE. The second equality follows from Assumption SC2-X.1. The third equality follows by Assumption SC2-X.2, which implies the result in the last equality.

(iii) Since $\bar{g}(\cdot)$ is a trivial function of ε_{i1} and ε_{i2} under Assumption SC3-X, we can define $\check{\bar{g}}(a, x_1, x_2) = \bar{g}(a, x_1, x_2, e_1, e_2)$.

$$\begin{aligned}
E[G_i(\varepsilon_{i2} - \varepsilon_{i1}) | X_i] &= E[E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) | X_i, \alpha_i] | X_i] \\
&= E[\check{\bar{g}}(\alpha_i, X_{i1}, X_{i2}) E[\varepsilon_{i2} - \varepsilon_{i1} | X_i, \alpha_i] | X_i] = 0. \quad (25)
\end{aligned}$$

The first equality follows by the LIE. The second equality follows by Assumption SC3-X.1. The last equality follows from $E[\varepsilon_{i1} | X_i, \alpha_i] = E[\varepsilon_{i2} | X_i, \alpha_i]$ under Assumption SC3-X.2. The result then follows from noting that $E[\varepsilon_{i2} - \varepsilon_{i1} | X_i] = 0$ under this assumption, which completes the proof. \square

A.6 Proof of Proposition 4.2

In this proof, all equalities involving random variables are understood to hold a.s.

First, note that Lemma A.5 applies here by simply changing the conditioning set. As a result, Assumption PT-NSP under Assumption NSP-X holds if and only if

$$\begin{aligned} & E[G_i(Y_{i2}(0) - Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[G_i|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0) - Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]. \end{aligned} \quad (26)$$

Next, we state some preliminary observations and then proceed to show each statement separately.

Note that, by the LIE, Assumption SEL-CI-X and the definition of $\bar{g}(\cdot)$, the LHS of (26) equals the following

$$\begin{aligned} & E[G_i(Y_{i2}(0) - Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[E[G_i|\alpha_i^\mu, \alpha_i^\lambda, X_i^\mu, X_i^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda](Y_{i2}(0) - Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(Y_{i2}(0) - Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]. \end{aligned} \quad (27)$$

Similarly, by the LIE, the RHS of (26) equals the following,

$$\begin{aligned} & E[G_i|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0) - Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0) - Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \end{aligned} \quad (28)$$

As a result, in the following, to show that Assumptions SC1-NSP, SC2-NSP, and SC3-NSP are sufficient for Assumption PT-NSP, it suffices to show that each assumption implies the following equality,

$$\begin{aligned} & E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(Y_{i2}(0) - Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0) - Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \end{aligned}$$

(i) By Assumption NSP-X, it follows that

$$\begin{aligned} & E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(Y_{i2}(0) - Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ & \quad + E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu], \end{aligned} \quad (29)$$

We first examine the first term on the RHS of the above equality. Note that by the symmetry restrictions in Assumptions SC1-NSP.1 and SC1-NSP.2, it follows that a.e.

$$(a, x^\mu, x_1^\lambda, x_2^\lambda) \in \mathcal{A} \times \mathcal{X}_\mu \times \mathcal{X}_\lambda^2$$

$$\begin{aligned} & E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu) \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu) | X_i^\lambda = (x_1^\lambda, x_2^\lambda), X_{i1}^\mu = X_{i2}^\mu = x^\mu, \alpha_i^\mu = a] \\ &= \int \bar{g}(a, x^\mu, x^\mu, x_1^\lambda, x_2^\lambda, e_1, e_2) \mu(x^\mu, a, e_1) dF_{\varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu, \alpha_i^\mu}(e_1, e_2 | (x_1^\lambda, x_2^\lambda), x^\mu, a) \\ &= \int \bar{g}(a, x^\mu, x^\mu, x_1^\lambda, x_2^\lambda, e_2, e_1) \mu(x^\mu, a, e_1) dF_{\varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu, \alpha_i^\mu}(e_2, e_1 | (x_1^\lambda, x_2^\lambda), x^\mu, a) \\ &= E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu) \mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) | X_i^\lambda = (x_1^\lambda, x_2^\lambda), X_{i1}^\mu = X_{i2}^\mu = x^\mu, \alpha_i^\mu = a]. \quad (30) \end{aligned}$$

As a result, the first summand in (29) equals zero by (30) and the LIE.

Next, we consider the second summand in (29),

$$\begin{aligned} & E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu) (\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda)) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] E[Y_{i2}(0) - Y_{i1}(0) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]. \quad (31) \end{aligned}$$

The first equality follows from the conditional independence assumption in Assumption SC1-NSP.3. The last equality follows from the time homogeneity of $F_{\varepsilon_{it}^\mu | X_i^\mu, X_i^\lambda, \alpha_i^\mu}$, which follows from the exchangeability restriction in Assumption SC1-NSP.2 by Lemma A.4, and implies that $E[\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu, \alpha_i^\mu] = 0$ and

$$E[Y_{i2}(0) - Y_{i1}(0) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] = E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]$$

by the LIE. As a result, the above implies that Assumption PT-NSP holds.

(ii) By Assumption SC2-NSP.1, we can define $\check{g}(a^\mu, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_1^\lambda) = \bar{g}(a^\mu, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_1^\lambda, e_2^\lambda)$. By Assumption NSP-X, it follows that

$$\begin{aligned} & E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu) (Y_{i2}(0) - Y_{i1}(0)) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu) \Delta_{\mu,i} | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &\quad + E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu) (\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda)) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] E[\Delta_{\mu,i} | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &\quad + E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] E[Y_{i2}(0) - Y_{i1}(0) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \quad (32) \end{aligned}$$

The second equality follows from the conditional independence conditions in Assumptions SC2-NSP.2 and SC2-NSP.3. The last equality follows from Assumption NSP-X. Equation (32) then implies Assumption PT-NSP.

(iii): By Assumption SC3-NSP.1, we can define $\check{g}(a^\mu, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda) = \bar{g}(a^\mu, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_1^\lambda, e_2^\lambda)$. Now by the Assumption NSP-X and SC3-NSP.1, it follows that

$$\begin{aligned}
& E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda)(Y_{i2}(0) - Y_{i1}(0)) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&= E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda)(\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu)) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&\quad + E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda)(\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda)) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&= E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda) E[\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu, \alpha_i^\mu] | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&\quad + E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&= E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&= E[\check{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] E[Y_{i2}(0) - Y_{i1}(0) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu],
\end{aligned}$$

where the first equality follows from Assumption NSP-X. The second equality follows by applying the LIE to the first term and the conditional independence imposed in Assumption SC3-NSP.3 to the second term. The first term on the RHS of the second equality equals zero by the conditioning on $X_{i1}^\mu = X_{i2}^\mu$ and the time homogeneity condition in Assumption SC3-NSP.2. The last equality follows from noting, similar as in the proof of (i), that since $E[\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu, \alpha_i^\mu] = 0$,

$$E[Y_{i2}(0) - Y_{i1}(0) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] = E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda) | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]$$

by the LIE. This completes the proof. \square

A.7 Proof of Proposition 4.3

Throughout this proof, equalities involving conditioning statements are understood to hold *a.e.* We proceed to show each result separately.

(i) It suffices to show (i.a) Assumptions SC1-NSP.1 and SC1-NSP.2 imply Assumption TH and (i.b) Assumptions SC1-NSP.1 and SC1-NSP.3 imply Assumption CRE.

(i.a) Consider

$$F_{\varepsilon_{i1}^\mu, G_i | X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_1, g | x^\mu, x^\lambda, a) = F_{G_i | \varepsilon_{i1}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(g | e_1, x^\mu, x^\lambda, a) F_{\varepsilon_{i1}^\mu | X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_1 | x^\mu, x^\lambda, a),$$

where $x^\mu = (x_1^\mu, x_2^\mu)$ and $x^\lambda = (x_1^\lambda, x_2^\lambda)$. Assumption SC1-NSP.2 implies $F_{\varepsilon_{i1}^\mu | X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e | x^\mu, x^\lambda, a) = F_{\varepsilon_{i2}^\mu | X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e | x^\mu, x^\lambda, a)$ as well as $F_{\varepsilon_{i1}^\mu | \varepsilon_{i2}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_1 | e_2, x^\mu, x^\lambda, a) = F_{\varepsilon_{i2}^\mu | \varepsilon_{i1}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_1 | e_2, x^\mu, x^\lambda, a)$

by Lemma A.4, which implies

$$\begin{aligned}
& F_{G_i|\varepsilon_{i1}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(g|e_1, x^\mu, x^\lambda, a) \\
&= \int 1\{g(a, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_1, e_2) \leq g\} dF_{\varepsilon_{i2}^\mu|\varepsilon_{i1}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_2|e_1, x^\mu, x^\lambda, a) \\
&= \int 1\{g(a, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda, e_2, e_1) \leq g\} dF_{\varepsilon_{i1}^\mu|\varepsilon_{i2}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_2|e_1, x^\mu, x^\lambda, a) \\
&= F_{G_i|\varepsilon_{i2}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(g|e_1, x^\mu, x^\lambda, a). \tag{33}
\end{aligned}$$

As a result,

$$\begin{aligned}
& F_{\varepsilon_{i1}^\mu, G_i|X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_1, g|x^\mu, x^\lambda, a) = F_{G_i|\varepsilon_{i1}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(g|e_1, x^\mu, x^\lambda, a) F_{\varepsilon_{i1}^\mu|X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_1|x^\mu, x^\lambda, a) \\
&= F_{G_i|\varepsilon_{i2}^\mu, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(g|e_1, x^\mu, x^\lambda, a) F_{\varepsilon_{i2}^\mu|X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_1|x^\mu, x^\lambda, a) \\
&= F_{\varepsilon_{i2}^\mu, G_i|X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e_1, g|x^\mu, x^\lambda, a). \tag{34}
\end{aligned}$$

This implies Assumption TH by the definition of a conditional distribution,

$$F_{\varepsilon_{it}^\mu|G_i, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e|g, x^\mu, x^\lambda, a) = \frac{F_{\varepsilon_{it}^\mu, G_i|X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e, g|x^\mu, x^\lambda, a)}{F_{G_i|X_i^\mu, X_i^\lambda, \alpha_i^\mu}(g|x^\mu, x^\lambda, a)},$$

where $F_{G_i|X_i^\mu, X_i^\lambda, \alpha_i^\mu}(g|x^\mu, x^\lambda, a) > 0$ a.s. for $g = 0, 1$ by assumption.

(i.b) This statement follows in a straightforward manner from the definition of G_i in Assumption SC1-NSP.1 and the conditional independence condition in Assumption SC1-NSP.3 which together imply Assumption CRE. This completes the proof of (i).

(ii) To show the result, it suffices to show that (ii.a) Assumptions SC3-NSP.1 and SC3-NSP.2 imply Assumption TH and (ii.b) Assumptions SC3-NSP.1 and SC3-NSP.3 imply Assumption CRE.

(ii.a) Under Assumptions SC3-NSP.1 and SC3-NSP.2, $G_i = g(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda)$ is a degenerate random variable equaling either zero or one with probability one conditional on X_i^μ, X_i^λ and α_i^μ . As a result,

$$\begin{aligned}
& F_{\varepsilon_{it}^\mu | G_i, X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e | g, x^\mu, x^\lambda, a) \\
&= \sum_{g=0,1} P(\varepsilon_{it}^\mu \leq e | G_i = g(a, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda), X_i^\mu = x^\mu, X_i^\lambda = x^\lambda, \alpha_i^\mu = a) 1\{g(a, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda) = g\} \\
&= \sum_{g=0,1} P(\varepsilon_{it}^\mu \leq e | X_i^\mu = x^\mu, X_i^\lambda = x^\lambda, \alpha_i^\mu = a) 1\{g(a, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda) = g\} \\
&= \sum_{g=0,1} F_{\varepsilon_{it}^\mu | X_i^\mu, X_i^\lambda, \alpha_i^\mu}(e | a, x^\mu, x^\lambda) 1\{g(a, x_1^\mu, x_2^\mu, x_1^\lambda, x_2^\lambda) = g\}. \tag{35}
\end{aligned}$$

As a result, Assumption SC3-NSP.1 together with the time homogeneity of $F_{\varepsilon_{it}^\mu | X_i^\mu, X_i^\lambda, \alpha_i^\mu}$ in Assumption SC3-NSP.2 is sufficient for the time homogeneity of $F_{\varepsilon_{it}^\mu | G_i, X_i^\mu, X_i^\lambda, \alpha_i^\mu}$, which yields Assumption TH.

(ii.b) The statement (ii.b) is immediate from noting that Assumption SC3-NSP.3 together with $G_i = g(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda)$ imply that $g(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda) \perp (\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda) | X_i^\mu, X_i^\lambda$, which is equivalent to Assumption CRE. This completes the proof of (ii). \square

A.8 Supplementary lemmas

Lemma A.1. *Suppose that Assumption SP holds and $P(G_i = 1) \in (0, 1)$. Then Assumption PT is equivalent to $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$.*

Proof. Under Assumption SP, Assumption PT can be rewritten as

$$E[\varepsilon_{i2} - \varepsilon_{i1} | G_i = 1] = E[\varepsilon_{i2} - \varepsilon_{i1} | G_i = 0].$$

This is equivalent to $E[\varepsilon_{i2} - \varepsilon_{i1} | G_i = g] = 0$ for $g \in \{0, 1\}$ since $E[\varepsilon_{i2} - \varepsilon_{i1}] = 0$ by assumption. Since $G_i \in \{0, 1\}$ and $P(G_i = 1) \in (0, 1)$ by assumption, this is in turn equivalent to $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$. \square

Lemma A.2. *For a scalar random variable W_i , let $\tilde{W}_i = W_i - E[W_i]$. If $E[\tilde{W}_i 1\{\tilde{W}_i \geq 0\}] = 0$, then $W_i = E[W_i]$ a.s.*

Proof. First, note that by definition $E[\tilde{W}_i] = 0$, which is equivalent to

$$E[\tilde{W}_i^+] = E[\tilde{W}_i^-], \tag{36}$$

where $\tilde{W}_i^+ = |\tilde{W}_i| 1\{\tilde{W}_i > 0\}$ and $\tilde{W}_i^- = |\tilde{W}_i| 1\{\tilde{W}_i < 0\}$.

Now suppose that $E[\tilde{W}_i 1\{\tilde{W}_i \geq 0\}] = 0$ holds, which is equivalent to

$$E[\tilde{W}_i^+ 1\{\tilde{W}_i \geq 0\}] = E[\tilde{W}_i^- 1\{\tilde{W}_i \geq 0\}], \quad (37)$$

since, by the definition, $\tilde{W}_i = \tilde{W}_i^+ - \tilde{W}_i^-$. Note that the right-hand side equals zero by the definition of \tilde{W}_i^- . As a result, $E[\tilde{W}_i^+ 1\{\tilde{W}_i \geq 0\}] = E[\tilde{W}_i^+] = 0$. Since $\tilde{W}_i^+ \geq 0$, this implies that $P(\tilde{W}_i^+ = 0) = 1$. Now note that $P(\tilde{W}_i^+ = 0) = P(|\tilde{W}_i| 1\{\tilde{W}_i > 0\} = 0) = P(1\{\tilde{W}_i > 0\} = 0) = 1$, which implies $P(\tilde{W}_i > 0) = 0$.

Since $E[\tilde{W}_i] = 0$, (36) further implies that $E[\tilde{W}_i^-] = E[\tilde{W}_i^+] = 0$. Since $\tilde{W}_i^- \geq 0$, it follows that $P(\tilde{W}_i^- = 0) = 1$. Now note that $P(\tilde{W}_i^- = 0) = P(|\tilde{W}_i| 1\{\tilde{W}_i < 0\} = 0) = P(1\{\tilde{W}_i < 0\} = 0) = 1$, which implies $P(\tilde{W}_i < 0) = 0$.

Together, $P(\tilde{W}_i > 0) = 0$ and $P(\tilde{W}_i < 0) = 0$ imply that $P(\tilde{W}_i = 0) = 1 - (P(\tilde{W}_i < 0) + P(\tilde{W}_i > 0)) = 1$, which completes the proof. \square

Lemma A.3. *Suppose that Assumption SP-X holds and $P(G_i = 1|X_i) \in (0, 1)$ a.s. Then Assumption PT-X holds if and only if $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|X_i] = E[G_i|X_i]E[\varepsilon_{i2} - \varepsilon_{i1}|X_i]$ a.s.*

Proof. Recall that Assumption PT-X is given by

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1, X_i] = E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0, X_i] \text{ a.s.} \quad (38)$$

Subtracting $E[Y_{i2}(0) - Y_{i1}(0)|X_i]$ from both sides yields the equivalent equality

$$E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|G_i = 1, X_i] = E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|G_i = 0, X_i] \text{ a.s.}, \quad (39)$$

where $\tilde{Y}_{it}(0) = Y_{it}(0) - E[Y_{it}(0)|X_i]$. Since, by definition, $E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|X_i] = 0$ a.s., Equation (39) is equivalent to $E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|G_i = g, X_i] = 0$ a.s. for $g \in \{0, 1\}$. Since $P(G_i = 1|X_i) \in (0, 1)$ a.s., by arguments similar to Lemma A.1, but conditioning on X_i , Assumption PT-X holds iff $E[G_i(Y_{i2}(0) - Y_{i1}(0))|X_i] = E[G_i|X_i]E[Y_{i2}(0) - Y_{i1}(0)|X_i]$ a.s. By Assumption SP-X, the left-hand side of the previous equality simplifies to

$$\begin{aligned} & E[G_i(Y_{i2}(0) - Y_{i1}(0))|X_i] \\ &= E[G_i|X_i](\lambda_2 - \lambda_1) + E[G_i|X_i](\gamma_2(X_i) - \gamma_1(X_i)) + E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|X_i] \text{ a.s.} \end{aligned}$$

Thus, Assumption PT-X holds if and only if $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|X_i] = E[G_i|X_i]E[\varepsilon_{i2} - \varepsilon_{i1}|X_i]$

$\varepsilon_{i1}|X_i]$ a.s. □

Lemma A.4. *Let $(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ denote a vector of random variables. Suppose that $\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i \stackrel{d}{=} \varepsilon_{i2}, \varepsilon_{i1}|\alpha_i$ holds. Then*

$$(i) F_{\varepsilon_{i1}|\alpha_i}(e|a) = F_{\varepsilon_{i2}|\alpha_i}(e|a) \text{ a.e. } (a, e) \in \mathcal{A} \times \mathcal{E}$$

$$(ii) F_{\varepsilon_{i1}|\varepsilon_{i2}, \alpha_i}(e_1|e_2, a) = F_{\varepsilon_{i2}|\varepsilon_{i1}, \alpha_i}(e_1|e_2, a) \text{ a.e. } (a, e_1, e_2) \in \mathcal{A} \times \mathcal{E}^2.$$

Proof. (i) By the definition of the marginal distribution, the conditional exchangeability restriction implies (i) by the following, a.e.

$$F_{\varepsilon_{i1}|\alpha_i}(e_1|a) = \lim_{e_2 \rightarrow \infty} F_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}(e_1, e_2|a) = \lim_{e_2 \rightarrow \infty} F_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}(e_2, e_1|a) = F_{\varepsilon_{i2}|\alpha_i}(e_1|a). \quad (40)$$

(ii) By the definition of the conditional distribution and (i) of this lemma, the conditional exchangeability restriction implies (ii) by the following

$$F_{\varepsilon_{i1}|\varepsilon_{i2}, \alpha_i}(e_1|e_2, a) = \frac{F_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}(e_1, e_2|a)}{F_{\varepsilon_{i2}|\alpha_i}(e_2|a)} = \frac{F_{\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i}(e_2, e_1|a)}{F_{\varepsilon_{i1}|\alpha_i}(e_2|a)} = F_{\varepsilon_{i2}|\varepsilon_{i1}, \alpha_i}(e_1|e_2, a), \quad (41)$$

a.e. $(a, e_1, e_2) \in \mathcal{A} \times \mathcal{E}^2$. □

Lemma A.5. *Suppose Assumption NSP-X holds and $P(G_i = 1|X_i) \in (0, 1)$ a.s. Then Assumption PT-X holds if and only if $E[G_i(Y_{i2}(0) - Y_{i1}(0))|X_i] = E[G_i|X_i]E[Y_{i2}(0) - Y_{i1}(0)|X_i]$ a.s.*

Proof. Recall that Assumption PT-X is given by

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1, X_i] = E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0, X_i] \text{ a.s.} \quad (42)$$

Subtracting $E[Y_{i2}(0) - Y_{i1}(0)|X_i]$ from both sides of the equality yields the following equivalent condition

$$E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|G_i = 1, X_i] = E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|G_i = 0, X_i] \text{ a.s.}, \quad (43)$$

where $\tilde{Y}_{it}(0) = Y_{it}(0) - E[Y_{it}(0)|X_i]$ for $t = 0, 1$. Since, by definition, $E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|X_i] = 0$ a.s. and $P(G_i = 1|X_i) \in (0, 1)$ a.s., (43) is equivalent to $E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))|X_i] = 0$ a.s. by arguments similar to Lemma A.1. The result now follows by the definition of $\tilde{Y}_{it}(0)$. □

Online Appendix

B Necessary and sufficient conditions: Extensions	1
B.1 Parallel trends for any distribution	1
B.2 Nonseparable models	2
B.3 Multiple periods and multiple groups	4
C Time homogeneity and correlated random effects imply parallel trends	7
D Unconditional parallel trends and covariate-specific trends	7
E Proofs of results in Online Appendix	7
E.1 Proof of Proposition B.1	7
E.2 Proof of Lemma B.1	9
E.3 Proof of Proposition B.2	9
E.4 Proof of Proposition B.3	9
E.5 Proof of Proposition B.4	10
E.6 Proof of Proposition B.5	11
E.7 Proof of Proposition B.6	13
E.8 Proof of Proposition B.7	13
E.9 Proof of Proposition B.8	14
E.10 Proof of Proposition B.9	14
E.11 Proof of Proposition C.1	15
E.12 Supplementary lemmas Online Appendix	16

B Necessary and sufficient conditions: Extensions

B.1 Parallel trends for any distribution

In the main text, we derive necessary and sufficient conditions for a scenario where researchers are not willing to choose a specific selection mechanism. Here we consider an alternative scenario where researchers are not willing to impose any restrictions on the distribution of unobservables, $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}}$, and require parallel trends to hold for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}}$.

The following proposition shows that Assumption PT holds for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}}$ in a complete class if and only if selection is independent of the time-varying unobservables $(\varepsilon_{i1}, \varepsilon_{i2})$. Before we state the proposition, we recall the definition of a complete class of distributions (Equations (4.8)–(4.9) on p.115 in Lehmann and Romano, 2005).

Definition B.1 (Completeness of a class of distributions). *Let W be a vector of random variables. A family of distributions \mathcal{F} is complete if*

$$E[f(W)] = 0 \quad \text{for all } F_W \in \mathcal{F}$$

implies

$$f(w) = 0 \quad \text{almost everywhere (a.e.) } \mathcal{F}.$$

Proposition B.1 (Necessary and sufficient condition for any distribution of unobservables). *Suppose that Assumption SP holds. Suppose further that $g \in \mathcal{G}_{all}$ and $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$, where \mathcal{F} is a complete family of distributions satisfying $P(\varepsilon_{i1} \neq \varepsilon_{i2}) = 1$, $E[\varepsilon_{i1}] = E[\varepsilon_{i2}] = 0$, and $P(G_i = 1) \in (0, 1)$. Assumption PT holds for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$ if and only if $P(G_i = 1 | \varepsilon_{i1}, \varepsilon_{i2}) = P(G_i = 1)$ a.s. for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$.*

In Proposition B.1, we require $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}}$ to belong to a complete family of distributions, \mathcal{F} . Completeness requires that the class of possible distributions of unobservables is rich enough. This condition is key for showing that parallel trends implies that selection is independent of ε_{i1} and ε_{i2} .

B.2 Nonseparable models

Here we provide necessary and sufficient conditions for parallel trends in the nonseparable model. To simplify the exposition, we abstract from covariates. We derive the necessary and sufficient conditions in the context of a fully nonseparable, time-varying outcome model.

Assumption NSP.

$$Y_{it}(0) = \xi_t(\alpha_i, \varepsilon_{it}), \quad i = 1, \dots, N, \quad t = 1, 2,$$

where α_i , ε_{i1} and ε_{i2} are finite-dimensional vector-valued random variables.

Consider a general selection mechanism that can depend on all unobservable determinants of $Y_{it}(0)$ as well as additional unobservables $(\nu_i, \eta_{i1}, \eta_{i2})$, where ν_i , η_{i1} and η_{i2} are vector-valued random variables,

$$G_i = g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}).$$

Let \mathcal{G}_{all} denote the set of all selection mechanisms $g(\cdot)$ that map from the support of the unobservables to $\{0, 1\}$. We also consider the following restricted versions of \mathcal{G}_{all} ,

$$\begin{aligned} \mathcal{G}_1 &= \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, e_2, v, t_1, t_2) \text{ is a trivial function of } e_2\}, \\ \mathcal{G}_2 &= \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, e_2, v, t_1, t_2) \text{ is a trivial function of } (e_1, e_2)\}. \end{aligned}$$

As for the separable model, we impose the following weak regularity condition that ensures non-degeneracy of the selection mechanisms used in the proofs of the necessary and sufficient conditions.

Assumption SEL-NSP. *There exists a component of ν_i , labeled ν_i^1 (w.l.o.g.), such that $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ and $P(\nu_i^1 > c) \in (0, 1)$ for some $c \in \mathbb{R}$.*

Before we proceed to the necessary and sufficient conditions, we provide an equivalent condition for Assumption PT in the context of the nonseparable model.

Lemma B.1. *Suppose that $P(G_i = 1) \in (0, 1)$. Then Assumption PT holds if and only if $E[G_i(Y_{i2}(0) - Y_{i1}(0))] = E[G_i]E[Y_{i2}(0) - Y_{i1}(0)]$.*

The following propositions extend the necessary and sufficient conditions for the separable model to the fully nonseparable model in Assumption NSP. To simplify exposition, we use $\tilde{Y}_{it}(0)$ to denote the centered potential outcome without the treatment, $\tilde{Y}_{it}(0) = Y_{it}(0) - E[Y_{it}(0)]$ for $t = 1, 2$.

Proposition B.2 (Necessary and sufficient condition for $g \in \mathcal{G}_{\text{all}}$). *Suppose that Assumptions NSP and SEL-NSP hold and $P(G_i = 1) \in (0, 1)$. Assumption PT holds for any fixed $g \in \mathcal{G}_{\text{all}}$ if and only if $\tilde{Y}_{i1}(0) = \tilde{Y}_{i2}(0)$ a.s.*

Proposition B.3 (Necessary and sufficient condition for $g \in \mathcal{G}_1$). *Suppose that Assumptions NSP and SEL-NSP hold and $P(G_i = 1) \in (0, 1)$. If Assumption PT holds for any fixed $g \in \mathcal{G}_1$, then $E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] = \tilde{Y}_{i1}(0)$ a.s. If, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i, \varepsilon_{i1}]$ a.s., then $E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] = \tilde{Y}_{i1}(0)$ a.s. is also sufficient for Assumption PT.*

Proposition B.4 (Necessary and sufficient condition $g \in \mathcal{G}_2$). *Suppose that Assumptions NSP and SEL-NSP hold and $P(G_i = 1) \in (0, 1)$. If Assumption PT holds for any fixed $g \in \mathcal{G}_2$, then $E[\tilde{Y}_{i1}(0)|\alpha_i] = E[\tilde{Y}_{i2}(0)|\alpha_i]$ a.s. If, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i]$ a.s., then $E[\tilde{Y}_{i1}(0)|\alpha_i] = E[\tilde{Y}_{i2}(0)|\alpha_i]$ a.s. is also sufficient for Assumption PT.*

Proposition B.5 (Necessary and sufficient condition for parallel trends for any distribution of unobservables). *Suppose that Assumption NSP holds. Suppose further that $g \in \mathcal{G}_{all}$ and $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$, where \mathcal{F} is a complete family of probability distributions satisfying $P(\tilde{Y}_{i1}(0) \neq \tilde{Y}_{i2}(0)) = 1$ and $P(g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1) \in (0, 1)$. Assumption PT holds for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$ if and only if $P(G_i = 1|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) = P(G_i = 1)$ a.s. for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$.*

The necessary and sufficient conditions in Propositions B.2, B.3, and B.4 are similar to their counterparts for the separable model. The necessary and sufficient condition in Proposition B.5 requires selection to be independent of all unobservable determinants of the untreated potential outcomes. It is stronger than the corresponding result for the separable model in Proposition B.1.

B.3 Multiple periods and multiple groups

Here we generalize our results to DiD designs with multiple periods and multiple groups. The setup and notation are based on Callaway and Sant’Anna (2021), Sun and Abraham (2021), and Roth et al. (2022). We focus on necessary and sufficient conditions; primitive sufficient conditions can be obtained in a straightforward manner as in Section 3.

Let $t \in \{1, 2, \dots, T\}$ index the periods. Suppose that at time $t = 1$, no units are treated, at $t = 2$, some units become treated, while others remain untreated, and so on. Previously treated units remain treated for all periods. Units can be categorized based on their treatment adoption pattern $D_i = (D_{i1}, \dots, D_{iT})$. We define the group indicator G_i as the first period in which units are treated, $G_i = \min\{t \in \{1, \dots, T\} : D_{it} = 1\}$, and set $G_i = \infty$ for the never-treated units so that $G_i \in \{2, \dots, T, \infty\}$.²¹

Potential outcomes are indexed by the entire treatment sequence $(d_1, \dots, d_T) \in \{0, 1\}^T$, $Y_{it}(d_1, \dots, d_T)$. Since treatment is an absorbing state, the potential outcomes can be indexed by the first treatment period only. Define $Y_{it}(g) = Y_{it}(\mathbf{0}_{g-1}, \mathbf{1}_{T-g+1})$

²¹Since G_i is a random variable with finite support, we emphasize that $\{\infty\}$ is merely a label.

for $g \in \{2, \dots, T\}$ and $Y_{it}(\infty) = Y_{it}(\mathbf{0}_T)$, where $\mathbf{0}_s \equiv (0, \dots, 0) \in \mathbb{R}^s$ and $\mathbf{1}_s \equiv (1, \dots, 1) \in \mathbb{R}^s$. Observed outcomes are given by $Y_{it} = \sum_{g \in \{2, \dots, T, \infty\}} 1\{G_i = g\} Y_{it}(g)$. We maintain a standard no-anticipation assumption (e.g., Roth et al., 2022).

Assumption NA. For $g \in \{2, \dots, T, \infty\}$ and $t < g$, $Y_{it}(g) = Y_{it}(\infty)$.

Our objects of interest are the group-time ATTs,

$$\text{ATT}(g, t) = E[Y_{it}(g) - Y_{it}(\infty) | G = g]. \quad (44)$$

We impose the following parallel trends assumption to identify the $\text{ATT}(g, t)$.²²

Assumption PT-MP. For $(g, t) \in \{2, \dots, T\}^2$

$$E[Y_{it}(\infty) - Y_{it-1}(\infty) | G_i = g] = E[Y_{it}(\infty) - Y_{it-1}(\infty) | G_i = \infty] \quad (45)$$

As in Section 2, we focus on a separable model to simplify the exposition.

Assumption SP-MP.

$$Y_{it}(\infty) = \alpha_i + \lambda_t + \varepsilon_{it}, \quad E[\varepsilon_{it}] = 0, \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

Selection into treatment can depend on the unobservable determinants of $Y_{it}(\infty)$ as well as additional unobservables, $\zeta_i = (\nu_i, \eta_{i1}, \dots, \eta_{iT})$,

$$G_i = g(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i).$$

As before, let \mathcal{G}_{all} denote the set of all selection mechanisms $g(\cdot)$ and define the following classes of restricted selection mechanisms, which are natural analogs of those considered in Section 2.

$$\mathcal{G}_1 = \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, \dots, e_T, z) \text{ is a trivial function of } e_T\}$$

$$\mathcal{G}'_1 = \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, \dots, e_T, z) \text{ is a trivial function of } (e_2, \dots, e_T)\}$$

$$\mathcal{G}_2 = \{g \in \mathcal{G}_{\text{all}} : g(a, e_1, \dots, e_T, z) \text{ is a trivial function of } (e_1, \dots, e_T)\}$$

²²In our setting, this parallel trends assumption corresponds to the ones made by Borusyak et al. (2021), Callaway and Sant'Anna (2021), Gardner (2021), Sun and Abraham (2021), and Wooldridge (2021); see also de Chaisemartin and D'Haultfœuille (2020) and Marcus and Sant'Anna (2021) for related assumptions.

The following assumption generalizes Assumption SEL to the multiple-period multiple-group setting. It ensures that the selection mechanisms used to establish the necessary and sufficient conditions for parallel trends are non-degenerate.

Assumption SEL-MP. *There exists a component of ν_i , labeled ν_i^1 (w.l.o.g.), such that $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT})$. In addition, there exists a non-overlapping partition of $\text{supp}(\nu_i^1)$, $\{B_g\}_{g=2}^T$, such that $P(\nu_i^1 \in B_g) \in (0, 1)$ for $g \in \{2, \dots, T\}$.*

The following four propositions extend the necessary and sufficient conditions in Propositions 2.1, 2.2, and 2.3 to the more general DiD setting in this section. All these conditions are natural generalizations of their counterparts in the 2×2 case.

Proposition B.6 (Necessary and sufficient condition for $g \in \mathcal{G}_{\text{all}}$). *Suppose that Assumptions NA, SP-MP, and SEL-MP hold and $P(G_i = g) \in (0, 1)$ for $g \in \{2, \dots, T, \infty\}$. Then Assumption PT-MP holds for any fixed $g \in \mathcal{G}_{\text{all}}$ if and only if $\varepsilon_{i1} = \dots = \varepsilon_{iT}$ a.s.*

Proposition B.7 (Necessary and sufficient condition for $g \in \mathcal{G}_1$). *Suppose that Assumptions NA, SP-MP, and SEL-MP hold and $P(G_i = g) \in (0, 1)$ for $g \in \{2, \dots, T, \infty\}$. If Assumption PT-MP holds for any fixed $g \in \mathcal{G}_1$, then $E[\varepsilon_{it} | \alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{it-1}] = \varepsilon_{it-1}$ a.s. for $t \in \{2, \dots, T\}$. If, in addition, $G_i | \alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT} \stackrel{d}{=} G_i | \alpha_i, \varepsilon_{i1}$,²³ then $E[\varepsilon_{it} | \alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{it-1}] = \varepsilon_{it-1}$ a.s. for $t \in \{2, \dots, T\}$ is also sufficient for Assumption PT-MP.*

Proposition B.8 (Necessary and sufficient condition for $g \in \mathcal{G}'_1$). *Suppose that Assumptions NA, SP-MP, and SEL-MP hold and $P(G_i = g) \in (0, 1)$ for $g \in \{2, \dots, T, \infty\}$. If Assumption PT-MP holds for any fixed $g \in \mathcal{G}'_1$, then $E[\varepsilon_{it} | \alpha_i, \varepsilon_{i1}] = E[\varepsilon_{it-1} | \alpha_i, \varepsilon_{i1}]$ a.s. for $t \in \{2, \dots, T\}$. If, in addition, $G_i | \alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT} \stackrel{d}{=} G_i | \alpha_i, \varepsilon_{i1}$, then $E[\varepsilon_{it} | \alpha_i, \varepsilon_{i1}] = E[\varepsilon_{it-1} | \alpha_i, \varepsilon_{i1}]$ a.s. for $t \in \{2, \dots, T\}$ is also sufficient for Assumption PT-MP.*

Proposition B.9 (Necessary and sufficient condition for $g \in \mathcal{G}_2$). *Suppose that Assumptions NA, SP-MP, and SEL-MP hold and $P(G_i = g) \in (0, 1)$ for $g \in \{2, \dots, T, \infty\}$. If Assumption PT-MP holds for any fixed $g \in \mathcal{G}_2$, then $E[\varepsilon_{i1} | \alpha_i] = \dots = E[\varepsilon_{iT} | \alpha_i]$ a.s. If, in addition, $G_i | \alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT} \stackrel{d}{=} G_i | \alpha_i$, then $E[\varepsilon_{i1} | \alpha_i] = \dots = E[\varepsilon_{iT} | \alpha_i]$ a.s. is also sufficient for Assumption PT-MP.*

²³It is important to note that this high-level condition rules out some selection mechanisms in \mathcal{G}_1 . We therefore consider the alternative extension to the multiple-period, multiple-group case with \mathcal{G}'_1 in Proposition B.8.

C Time homogeneity and correlated random effects imply parallel trends

Here we prove the claim that Assumptions TH and CRE imposed on the potential outcome model in Assumption NSP-X imply Assumption PT-NSP in Section 4.4.

Proposition C.1 (Time homogeneity and correlated random effects imply Assumption PT-NSP). *Suppose that Assumption NSP-X holds and $P(G_i = 1 | X_{i1}^\mu = X_{i2}^\mu, X_i^\lambda) \in (0, 1)$ a.s. Then Assumptions TH and CRE imply Assumption PT-NSP.*

D Unconditional parallel trends and covariate-specific trends

A natural question is whether and when unconditional parallel trends (Assumption PT) continues to hold despite covariates entering the outcome equation as in Assumption SP-X. One can show that this is possible if three conditions hold. First, the selection mechanism is conditionally mean independent of the covariates, $E[G_i | \alpha_i, X_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}]$. Second, the covariates are independent of the unobservable determinants of the untreated potential outcomes, $(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) \perp X_i$. Finally, selection is orthogonal to the change in ε_{it} , $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$.

While this demonstrates a case where covariates may not be required for identification, even if there are covariate-specific trends, the conditions are restrictive. Not only do they rule out covariates determining selection, but covariates would also have to be independent of the unobservable determinants of the untreated potential outcomes.

E Proofs of results in Online Appendix

E.1 Proof of Proposition B.1

Recall that \mathcal{F} is a complete family of distributions satisfying $P(\varepsilon_{i1} \neq \varepsilon_{i2}) = 1$, $E[\varepsilon_{i1}] = E[\varepsilon_{i2}] = 0$, and $P(g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1) \in (0, 1)$.

“ \implies ”: We show that if Assumption PT holds for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$, then $P(G_i = 1 | \varepsilon_{i1}, \varepsilon_{i2}) = P(G_i = 1)$ a.s. for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$. By Lemma A.1,

Assumption PT is equivalent to $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$, which in turn is equivalent to

$$E[\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1})] = 0, \quad (46)$$

where $\bar{g}(\varepsilon_{i1}, \varepsilon_{i2}) = E[g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) - E[g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2})] | \varepsilon_{i1}, \varepsilon_{i2}]$. This follows by the LIE and subtracting $E[G_i]E[\varepsilon_{i2} - \varepsilon_{i1}]$, noting that it equals zero by assumption.

It follows that Assumption PT holding for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$ is equivalent to

$$E[\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1})] = 0 \quad \text{for all } F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}. \quad (47)$$

By completeness of \mathcal{F} , the last equality implies (Lehmann and Romano, 2005, p.115) that

$$P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0) = 1 \quad \text{for all } F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}. \quad (48)$$

Now note that the left-hand side of (48) can be simplified as follows,

$$\begin{aligned} & P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0) \\ &= P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0, \varepsilon_{i1} = \varepsilon_{i2}) + P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0, \varepsilon_{i1} \neq \varepsilon_{i2}) \\ &= P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0 | \varepsilon_{i1} \neq \varepsilon_{i2}) \\ &= P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2}) = 0) = 1, \end{aligned} \quad (49)$$

where the penultimate equality follows since $P(\varepsilon_{i1} \neq \varepsilon_{i2}) = 1$ by assumption. As a result, by the definition of $\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})$, it follows that

$$P(E[g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) | \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i]) = 1 \quad \text{for all } F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}. \quad (50)$$

“ \Leftarrow ”:

 We now proceed to show the other direction. In the following, all statements are understood to hold for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$. If $P(G_i = 1 | \varepsilon_{i1}, \varepsilon_{i2}) = P(G_i = 1)$ a.s., which is equivalent to $E[G_i | \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i]$ a.s., then

$$E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = E[E[G_i | \varepsilon_{i1}, \varepsilon_{i2}](\varepsilon_{i2} - \varepsilon_{i1})] = E[E[G_i](\varepsilon_{i2} - \varepsilon_{i1})] = 0,$$

where the first equality follows by the LIE and the second equality follows by the conditional mean independence condition on G_i . The last equality follows from $E[\varepsilon_{it}] = 0$ for $t = 1, 2$. This completes the proof. \square

E.2 Proof of Lemma B.1

This is a special case of Lemma A.5, and we omit the proof for brevity. \square

E.3 Proof of Proposition B.2

“ \implies ”: We show that if Assumption PT holds for any fixed $g \in \mathcal{G}_{\text{all}}$, then $\tilde{Y}_{i1}(0) = \tilde{Y}_{i2}(0)$ a.s. Since Assumption PT holds for any fixed $g \in \mathcal{G}_{\text{all}}$, then it holds for

$$\check{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0) \geq 0\}.$$

By Lemma B.1, Assumption PT holding for $\check{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0) \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0, \quad (51)$$

which, by Assumption SEL-NSP, is equivalent to

$$E[1\{\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0) \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0. \quad (52)$$

Since $E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)] = 0$ by construction, the above equality implies that $\tilde{Y}_{i1}(0) = \tilde{Y}_{i2}(0)$ a.s. by Lemma A.2.

“ \impliedby ”: We show that if $\tilde{Y}_{i1}(0) = \tilde{Y}_{i2}(0)$ a.s., then Assumption PT holds for any fixed $g \in \mathcal{G}_{\text{all}}$. Note that if $\tilde{Y}_{i1}(0) = \tilde{Y}_{i2}(0)$ a.s., then $G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)) = 0$ a.s. As a result, $E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0$, which is the equivalent condition of Assumption PT by Lemma B.1. This completes the proof. \square

E.4 Proof of Proposition B.3

“ \implies ”: We show that if Assumption PT holds for any fixed $g \in \mathcal{G}_1$, then $E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] = \tilde{Y}_{i1}(0)$ a.s. To show this, note that if Assumption PT holds for any fixed $g \in \mathcal{G}_1$, then it holds for $\check{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0) \geq 0\}$. By

Lemma B.1, Assumption PT holding for $\check{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0) \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0, \quad (53)$$

which, by Assumption SEL-NSP, is equivalent to

$$E[1\{E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0) \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0. \quad (54)$$

By the LIE, this is further equivalent to

$$E[1\{E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0) \geq 0\}(E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0))] = 0. \quad (55)$$

Since $E[E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0)] = 0$ by construction, the above equality implies that $E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] = \tilde{Y}_{i1}(0)$ a.s. by Lemma A.2.

“ \Leftarrow ”:

 We show that if, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i, \varepsilon_{i1}]$ a.s., then $E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] = \tilde{Y}_{i1}(0)$ a.s. is sufficient for Assumption PT. Note that

$$\begin{aligned} E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] &= E[E[E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}](\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i, \varepsilon_{i1})]] \\ &= E[E[E[G_i|\alpha_i, \varepsilon_{i1}](\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i, \varepsilon_{i1})]] \\ &= E[E[G_i|\alpha_i, \varepsilon_{i1}](E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0))] = 0, \end{aligned} \quad (56)$$

where the first equality follows from the LIE. The second equality follows from the conditional mean independence restriction imposed on G_i . The last equality follows, since $E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] = \tilde{Y}_{i1}(0)$ a.s. implies $E[G_i|\alpha_i, \varepsilon_{i1}](E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0)) = 0$ a.s. As a result, the latter term has zero expectation. Since $E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0$ is the equivalent condition for Assumption PT by Lemma B.1, this completes the proof. \square

E.5 Proof of Proposition B.4

“ \Rightarrow ”:

 We show that if Assumption PT holds for any fixed $g \in \mathcal{G}_2$, then $E[\tilde{Y}_{i1}(0)|\alpha_i] = E[\tilde{Y}_{i2}(0)|\alpha_i]$ a.s. To do so, note that if Assumption PT holds for any fixed $g \in \mathcal{G}_2$, then it holds for $\check{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] \geq 0\}$. By

Lemma B.1, Assumption PT holding for $\check{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{E[E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] \geq 0]\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0, \quad (57)$$

which, by Assumption SEL-NSP, is equivalent to

$$E[1\{E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0. \quad (58)$$

By the LIE, this is further equivalent to

$$E[1\{E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] \geq 0\}E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i]] = 0. \quad (59)$$

Since $E[E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i]] = 0$ by construction, the above equality implies $E[\tilde{Y}_{i1}(0)|\alpha_i] = E[\tilde{Y}_{i2}(0)|\alpha_i]$ a.s. by Lemma A.2.

“ \Leftarrow ”:
We show that if, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i]$ a.s., then $E[\tilde{Y}_{i1}(0)|\alpha_i] = E[\tilde{Y}_{i2}(0)|\alpha_i]$ a.s. is also sufficient for Assumption PT. Note that

$$\begin{aligned} E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] &= E[E[E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}](\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))|\alpha_i]] \\ &= E[E[E[G_i|\alpha_i](\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))|\alpha_i]] \\ &= E[E[G_i|\alpha_i]E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i]] = 0, \end{aligned} \quad (60)$$

where the first equality follows by the LIE. The second follows from the conditional mean independence restriction imposed on G_i . The last equality follows by noting that since $E[\tilde{Y}_{i2}(0)|\alpha_i] = E[\tilde{Y}_{i1}(0)|\alpha_i]$ a.s., then $E[G_i|\alpha_i]E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] = 0$ a.s., which therefore has zero expectation. Since $E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0$ is the equivalent condition for Assumption PT by Lemma B.1, this completes the proof. \square

E.6 Proof of Proposition B.5

“ \Rightarrow ”:
By Lemma B.1, Assumption PT is equivalent to $E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0$, which in turn is equivalent to the following

$$E[\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) - \tilde{\xi}_1(\alpha_i, \varepsilon_{i1}))] = 0, \quad (61)$$

where $\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) = E[g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) - E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}]]$ and $\tilde{\xi}_t(\alpha_i, \varepsilon_{it}) = \xi_t(\alpha_i, \varepsilon_{it}) - E[Y_{it}(0)]$ for $t = 1, 2$. The equivalence between $E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0$

and (61) follows by the LIE and subtracting $E[G_i]E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)]$, noting that it equals zero by construction.

It follows that Assumption PT holding for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$ is equivalent to

$$E[\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) - \tilde{\xi}_1(\alpha_i, \varepsilon_{i1}))] = 0, \quad (62)$$

for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$. By completeness of \mathcal{F} , the last equality implies the following (Lehmann and Romano, 2005, p.115)

$$P(\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) - \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})) = 0) = 1 \quad \text{for all } F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \eta_{i1}, \eta_{i2}} \in \mathcal{F}. \quad (63)$$

Now note that the left-hand side of (63) can be simplified as follows,

$$\begin{aligned} & P(\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) - \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})) = 0) \\ &= P(\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) - \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})), \tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) = \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})) \\ &\quad + P(\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) - \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})), \tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) \neq \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})) \\ &= P(\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) - \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})) | \tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) \neq \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})) \\ &= P(\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) = 0) = 1, \end{aligned} \quad (64)$$

where the penultimate equality follows since $P(\tilde{\xi}_2(\alpha_i, \varepsilon_{i2}) \neq \tilde{\xi}_1(\alpha_i, \varepsilon_{i1})) = P(\tilde{Y}_{i2}(0) \neq \tilde{Y}_{i1}(0)) = 1$ by assumption. As a result, by the definition of $\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$,

$$P(E[g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i]) = 1 \quad \text{for all } F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}. \quad (65)$$

“ \Leftarrow ”: The if statement follows by the LIE. All following statements are understood to hold for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}$. Note that $P(G_i = 1 | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) = P(G_i = 1)$ a.s. is equivalent to $E[G_i | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i]$ a.s. Next, the LIE implies the following equality

$$\begin{aligned} E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] &= E[E[G_i | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}](\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] \\ &= E[E[G_i](\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0. \end{aligned} \quad (66)$$

The second equality follows from $E[G_i | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i]$ a.s. The last equality follows from $E[\tilde{Y}_{it}(0)] = 0$ for $t = 1, 2$ by definition. \square

E.7 Proof of Proposition B.6

“ \implies ”: Since Assumption PT-MP holds for any fixed $g \in \mathcal{G}_{\text{all}}$, it holds for the following selection mechanism, where $\mathcal{G}_S = \{2, \dots, T\}$ denotes the set of switcher groups,

$$\check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) = \begin{cases} g & \text{if } 1\{\varepsilon_{ig} \geq \varepsilon_{ig-1}\}1\{\nu_i^1 \in B_g\} = 1, g \in \mathcal{G}_S \\ \infty & \text{if } \check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) \notin \mathcal{G}_S \end{cases}$$

By Lemma E.1, Assumption PT-MP implies that for any $g \in \mathcal{G}_S$,

$$E[1\{\check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) = g\}(\varepsilon_{ig} - \varepsilon_{ig-1})] = E[1\{\varepsilon_{ig} \geq \varepsilon_{ig-1}\}1\{\nu_i^1 \in B_g\}(\varepsilon_{ig} - \varepsilon_{ig-1})] = 0.$$

By Assumption SEL-MP, the last equality implies

$$E[1\{\varepsilon_{ig} \geq \varepsilon_{ig-1}\}(\varepsilon_{ig} - \varepsilon_{ig-1})] = 0.$$

Since $E[\varepsilon_{ig} - \varepsilon_{ig-1}] = 0$ by assumption, the above equality implies that $\varepsilon_{ig} - \varepsilon_{ig-1} = 0$ a.s. by Lemma A.2. This shows that $\varepsilon_{i1} = \dots = \varepsilon_{iT}$ a.s.

“ \impliedby ”: This direction of the proof is immediate. □

E.8 Proof of Proposition B.7

“ \implies ”: Since Assumption PT-MP holds for any fixed $g \in \mathcal{G}_1$, it holds for the following selection mechanism

$$\check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) = \begin{cases} g & \text{if } 1\{E[\varepsilon_{ig} | \alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{ig-1}] \geq \varepsilon_{ig-1}\}1\{\nu_i \in B_g\} = 1, g \in \mathcal{G}_S, \\ \infty & \text{if } \check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) \notin \mathcal{G}_S, \end{cases}$$

where \mathcal{G}_S is defined in the proof of Proposition B.6.

The proof now follows from similar arguments as in Propositions 2.2 and B.6.

“ \impliedby ”: For $g \in \{2, \dots, T, \infty\}$, consider

$$\begin{aligned} E[1\{G_i = g\}(\varepsilon_{i2} - \varepsilon_{i1})] &= E[E[E[1\{G_i = g\} | \alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}](\varepsilon_{i2} - \varepsilon_{i1}) | \alpha_i, \varepsilon_{i1}]] \\ &= E[E[1\{G_i = g\} | \alpha_i, \varepsilon_{i1}](E[\varepsilon_{i2} | \alpha_i, \varepsilon_{i1}] - \varepsilon_{i1})] = 0. \end{aligned}$$

The first equality follows by the LIE, the second follows from $G_i|\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT} \stackrel{d}{=} G_i|\alpha_i, \varepsilon_{i1}$, and the last by $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$ a.s. Next, for $g \in \{2, \dots, T, \infty\}$,

$$\begin{aligned} E[1\{G_i = g\}(\varepsilon_{i3} - \varepsilon_{i2})] &= E[E[E[1\{G_i = g\}|\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}](\varepsilon_{i3} - \varepsilon_{i2})|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}]] \\ &= E[E[E[1\{G_i = g\}|\alpha_i, \varepsilon_{i1}](\varepsilon_{i3} - \varepsilon_{i2})|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}]] \\ &= E[E[1\{G_i = g\}|\alpha_i, \varepsilon_{i1}](E[\varepsilon_{i3}|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] - \varepsilon_{i2})] = 0. \end{aligned}$$

Similar arguments show that $E[1\{G_i = g\}(\varepsilon_{it} - \varepsilon_{it-1})] = 0$ for $g \in \{2, \dots, T, \infty\}$ and $t \in \{4, \dots, T\}$. The result now follows by Lemma E.1. \square

E.9 Proof of Proposition B.8

“ \implies ”:

 Since Assumption PT-MP holds for any fixed $g \in \mathcal{G}'_1$, then it holds for the following selection mechanism

$$\check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) = \begin{cases} g & \text{if } 1\{E[\varepsilon_{ig}|\alpha_i, \varepsilon_{i1}] \geq E[\varepsilon_{ig-1}|\alpha_i, \varepsilon_{i1}]\}1\{\nu_i \in B_g\} = 1, g \in \mathcal{G}_S, \\ \infty & \text{if } \check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) \notin \mathcal{G}_S, \end{cases}$$

where \mathcal{G}_S is defined in the proof of Proposition B.6.

The proof now follows from similar arguments as in Propositions 2.2 and B.6.

“ \impliedby ”:

 For $(g, t) \in \{2, \dots, T, \infty\} \times \{2, \dots, T\}$, we have

$$\begin{aligned} E[1\{G_i = g\}(\varepsilon_{it} - \varepsilon_{it-1})] &= E[E[E[1\{G_i = g\}|\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}](\varepsilon_{it} - \varepsilon_{it-1})|\alpha_i, \varepsilon_{i1}]] \\ &= E[E[1\{G_i = g\}|\alpha_i, \varepsilon_{i1}](E[\varepsilon_{it}|\alpha_i, \varepsilon_{i1}] - E[\varepsilon_{it-1}|\alpha_i, \varepsilon_{i1}])] = 0. \end{aligned}$$

The first equality follows from the LIE, the second equality follows from $G_i|\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT} \stackrel{d}{=} G_i|\alpha_i, \varepsilon_{i1}$, and the last equality follows from $E[\varepsilon_{it}|\alpha_i, \varepsilon_{i1}] = E[\varepsilon_{it-1}|\alpha_i, \varepsilon_{i1}]$ a.s. for $t \in \{2, \dots, T\}$. The result now follows from Lemma E.1. \square

E.10 Proof of Proposition B.9

“ \implies ”:

 Since Assumption PT-MP holds for any fixed $g \in \mathcal{G}_2$, then it holds for the following selection mechanism

$$\check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) = \begin{cases} g & \text{if } 1\{E[\varepsilon_{ig}|\alpha_i] \geq E[\varepsilon_{ig-1}|\alpha_i]\}1\{\nu_i \in B_g\} = 1, g \in \mathcal{G}_S, \\ \infty & \text{if } \check{g}(\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}, \zeta_i) \notin \mathcal{G}_S, \end{cases}$$

where \mathcal{G}_S is defined in the proof of Proposition B.6.

The proof now follows from similar arguments as in Propositions 2.3 and B.6.

“ \Leftarrow ”: For $(g, t) \in \{2, \dots, T, \infty\} \times \{2, \dots, T\}$, we have

$$\begin{aligned} E[1\{G_i = g\}(\varepsilon_{it} - \varepsilon_{it-1})] &= E[E[E[1\{G_i = g\}|\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT}](\varepsilon_{it} - \varepsilon_{it-1})|\alpha_i]] \\ &= E[E[1\{G_i = g\}|\alpha_i](E[\varepsilon_{it}|\alpha_i] - E[\varepsilon_{it-1}|\alpha_i])] = 0. \end{aligned}$$

The first equality follows from the LIE, the second equality follows from $G_i|\alpha_i, \varepsilon_{i1}, \dots, \varepsilon_{iT} \stackrel{d}{=} G_i|\alpha_i$, and the last equality follows from $E[\varepsilon_{it}|\alpha_i] = E[\varepsilon_{it-1}|\alpha_i]$ a.s. for $t \in \{2, \dots, T\}$.

The result now follows from Lemma E.1. \square

E.11 Proof of Proposition C.1

Under Assumption NSP-X,

$$\begin{aligned} &E[Y_{i2}(0) - Y_{i1}(0)|G_i, X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\ &= E[\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu)|G_i, X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \end{aligned} \quad (67)$$

$$+ E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda)|G_i, X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]. \quad (68)$$

The remainder of the proof follows in two steps. First, we show that the term in (67) equals zero under our assumptions. Second, we show that the second term is conditionally mean independent of G_i , which implies Assumption PT-NSP.

We proceed to show that under Assumption TH the term in (67) equals zero by the following,

$$\begin{aligned} &E[\mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu)|G_i = g, X_i^\lambda = (x_1^\lambda, x_2^\lambda), X_{i1}^\mu = X_{i2}^\mu = x^\mu] \\ &= \int \mu(x^\mu, a^\mu, e^\mu) dF_{\alpha_i^\mu, \varepsilon_{i1}^\mu|G_i, X_i^\lambda}(a^\mu, e^\mu|g, (x^\mu, x^\mu), (x_1^\lambda, x_2^\lambda)) \\ &= \int \mu(x^\mu, a^\mu, e^\mu) dF_{\alpha_i^\mu, \varepsilon_{i2}^\mu|G_i, X_i^\lambda}(a^\mu, e^\mu|g, (x^\mu, x^\mu), (x_1^\lambda, x_2^\lambda)) \\ &= E[\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu)|G_i = g, X_i^\lambda = (x_1^\lambda, x_2^\lambda), X_{i1}^\mu = X_{i2}^\mu = x^\mu], \end{aligned} \quad (69)$$

where the first and last equalities follow by definition, whereas the penultimate equality follows from Assumption TH noting that it implies $\alpha_i^\mu, \varepsilon_{i1}^\mu|G_i, X_i^\mu, X_i^\lambda \stackrel{d}{=} \alpha_i^\mu, \varepsilon_{i2}^\mu|G_i, X_i^\mu, X_i^\lambda$.

Finally, we show that Assumption CRE implies the following for (68)

$$\begin{aligned}
& E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda) | G_i = g, X_i^\lambda = (x_1^\lambda, x_2^\lambda), X_{i1}^\mu = X_{i2}^\mu = x^\mu] \\
&= \int (\lambda_2(x_2^\lambda, a^\lambda, e_2^\lambda) - \lambda_1(x_1^\lambda, a^\lambda, e_1^\lambda)) dF_{\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda | G_i, X_i^\mu, X_i^\lambda}(a^\lambda, (e_1^\lambda, e_2^\lambda) | g, (x^\mu, x^\mu), (x_1^\lambda, x_2^\lambda)) \\
&= \int (\lambda_2(x_2^\lambda, a^\lambda, e_2^\lambda) - \lambda_1(x_1^\lambda, a^\lambda, e_1^\lambda)) dF_{\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda | X_i^\mu, X_i^\lambda}(a^\lambda, (e_1^\lambda, e_2^\lambda) | (x^\mu, x^\mu), (x_1^\lambda, x_2^\lambda)) \\
&= E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda) | X_i^\lambda = (x_1^\lambda, x_2^\lambda), X_{i1}^\mu = X_{i2}^\mu = x^\mu], \tag{70}
\end{aligned}$$

where the penultimate equality follows by Assumption CRE. This completes the proof. \square

E.12 Supplementary lemmas Online Appendix

Lemma E.1 (Equivalence with multiple periods). *Suppose that Assumptions NA and SP-MP hold and $P(G_i = g) \in (0, 1)$ for $g \in \{2, \dots, T, \infty\}$. Then Assumption PT-MP is equivalent to $E[1\{G_i = g\}(\varepsilon_{it} - \varepsilon_{it-1})] = 0$ for $g \in \{2, \dots, T, \infty\}$ and $t \in \{2, \dots, T\}$.*

Proof. Under Assumption SP-MP, Assumption PT-MP is equivalent to

$$E[\varepsilon_{it} - \varepsilon_{it-1} | G_i = g] = E[\varepsilon_{it} - \varepsilon_{it-1} | G_i = \infty] \quad \text{for } (g, t) \in \{2, \dots, T\}^2,$$

which, since $E[\varepsilon_{it}] = 0$, is also equivalent to

$$E[\varepsilon_{it} - \varepsilon_{it-1} | G_i = g] = 0 \quad \text{for } (g, t) \in \{2, \dots, T, \infty\} \times \{2, \dots, T\}. \tag{71}$$

Thus, we need to show that (71) is equivalent to $E[1\{G_i = g\}(\varepsilon_{it} - \varepsilon_{it-1})] = 0$ for $g \in \{2, \dots, T, \infty\}$ and $t \in \{2, \dots, T\}$. This follows because

$$E[\varepsilon_{it} - \varepsilon_{it-1} | G_i = g] = \frac{E[1\{G_i = g\}(\varepsilon_{it} - \varepsilon_{it-1})]}{P(G_i = g)} \quad \text{for } (g, t) \in \{2, \dots, T, \infty\} \times \{2, \dots, T\}$$

since $P(G_i = g) \in (0, 1)$ for $g \in \{2, \dots, T, \infty\}$ by assumption. \square