Selection and parallel trends*

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

First draft on arXiv: March 17, 2022. This draft: August 8, 2022

Comments welcome!

Abstract

One of the perceived advantages of difference-in-differences (DiD) methods is that they do not explicitly restrict how units select into treatment. However, when justifying DiD, researchers often argue that the treatment is "quasi-randomly" assigned. We investigate what selection mechanisms are compatible with the parallel trends assumptions underlying DiD. We derive necessary and sufficient conditions for parallel trends that clarify whether and how selection can depend on time-invariant and time-varying unobservables. We also suggest a menu of interpretable primitive sufficient conditions for parallel trends, thereby providing the formal underpinnings for justifying DiD based on contextual information about selection into treatment. We provide results for both separable and nonseparable outcome models and show that this distinction has implications for the use of covariates in DiD analyses. 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 Dmitry Arkhangelsky, Brantly Callaway, Clément de Chaisemartin, Gordon Dahl, Stefan Hoderlein, Nikolay Kudrin, David McKenzie, Vitor Possebom, Niklas Potrafke, Jonathan Roth, Yuya Sasaki, seminar participants at UCSD, and conference participants at the California 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. dghanem@ucdavis.edu

‡Microsoft and Department of Economics, Vanderbilt University. pedro.h.santanna@vanderbilt.edu

§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 potential selection mechanisms, they are often not explicit about what constitutes a "quasi-random" assignment, arguably due to the lack of formal guidance. In this paper, we investigate the connection between selection and parallel trends assumptions and thereby establish formal underpinnings for justifying parallel trends in practice.

Consider the classical 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).[1] 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, formally

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

where $G_i = 1$ indicates the treatment group and $G_i = 0$ indicates the control group.

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

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

[1] Several recent papers show that more general DiD setups with multiple periods and multiple groups can be thought of as a sequence of DiD problems with two groups and two periods (e.g., de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). Thus, for expositional simplicity and clarity, we focus on the classical DiD setting. Other multiple-period, multiple-group DiD settings includes, e.g., Borusyak, Jaravel, and Spiess (2021); Gardner (2021); Marcus and Sant'Anna (2021); Wooldridge (2021).

where α_i and ε_{it} are the time-invariant and time-varying unobservables, respectively. Equation (1) imposes separability in the unobservable determinants of the *untreated* potential outcome and allows for a transparent discussion of our main theoretical results.[2] 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 unobservables $(\nu_i, \eta_{i1}, \eta_{i2})$,

$$G_i = g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}). \tag{2}$$

Here ν_i and η_{it} are vectors of time-invariant and time-varying unobservables, respectively, and $g(\cdot)$ is an arbitrary function. The general selection mechanism in Equation (2) accommodates selection based on untreated potential outcomes, selection based 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 key assumption. We consider two scenarios to characterize the trade-offs between restrictions on selection and time-varying unobservables. First, suppose that researchers are not willing to impose a specific model for the selection mechanism. 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, λ_t. We therefore consider restricted selection mechanisms: (i) if selection does not depend on ε_{i2}, parallel trends implies a martingale-type property on ε_{it}; (ii) if selection does not 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 conditions on the conditional expectation of G_i, the latter two necessary conditions are also sufficient for parallel trends.

Alternatively, suppose that researchers are not willing to impose any restrictions on the distribution of (time-varying) unobservables, then parallel trends holds if and only if selection is independent of the time-varying unobservable determinants of the untreated potential outcome. Together with the other necessary and sufficient conditions, this result implies that parallel trends cannot hold absent additional restrictions on the selection mechanism or the distribution of (time-varying) unobservables.

These necessary and sufficient 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 unobserv-

[2]Note that, even under the separable outcome model (1), parallel trends may not hold without further restrictions on the selection mechanism.

ables 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 selection mechanisms where selection is based on untreated potential outcomes and treatment effects (Roy-style selection).

We then examine the role of (time-varying) covariates in DiD analyses.[3] We start by incorporating them into the separable model,

$$Y_{it}(0) = \alpha_i + \lambda_t + \gamma_t(X_{it}) + \varepsilon_{it}, \tag{3}$$

where $\gamma_t(\cdot)$ is an arbitrary, potentially time-varying nonparametric function of the covariates. We provide interpretable conditions that imply conditional parallel trends. These conditions generalize the sufficient conditions for unconditional parallel trends by allowing covariates to enter the selection mechanism and by conditioning on them in all distributional restrictions.

Our analysis highlights 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. In particular, they 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.

Finally, we generalize our analysis to a nonseparable model, where covariates and unobservables can interact,

$$Y_{it}(0) = \mu(X_{it}^{\mu}, \alpha_i^{\mu}, \varepsilon_{it}^{\mu}) + \lambda_t(X_{it}^{\lambda}, \alpha_i^{\lambda}, \varepsilon_{it}^{\lambda}). \tag{4}$$

Here we allow the selection mechanism to depend on all covariates as well as the unobservables that enter the time-invariant component of the structural function, $\mu(\cdot)$. 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. It also sheds light on the connection between parallel trends and identifying assumptions in nonseparable panel models (e.g. Altonji and Matzkin, 2005; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013) as well as unconfoundedness

[3] We assume that covariates are not affected by the treatment. See Caetano, Callaway, Payne, and Rodrigues (2022) for some recent results relaxing this assumption.

(e.g., Imbens, 2004; Imbens and Wooldridge, 2009).

Our analysis has important implications for empirical practice. First, our necessary and sufficient conditions demonstrate that (implicit) assumptions on selection are unavoidable in DiD analyses. Second, we expand the set of formally grounded arguments for justifying DiD based on contextual knowledge about selection, thereby providing formal guidance on what constitutes "quasi-random" assignment. Finally, our results provide guidance on which covariates to include in DiD analyses to render the required identification conditions more plausible. We discuss these implications in detail in Section 6.

This paper contributes to several branches of the causal inference literature. Our first contribution is to the classical literature on canonical DiD setups without covariates. See, e.g., Ashenfelter (1978), Ashenfelter and Card (1985), Heckman and Robb (1985), Card (1990), Card and Krueger (1994), Meyer, Viscusi, and Durbin (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'Haultfœuille (2021) and Roth, Sant'Anna, Bilinski, and Poe (2022) for surveys. Within this strand of the literature, our paper is most closely related to Roth and Sant'Anna (2021), Arkhangelsky and Imbens (2022), and Arkhangelsky, Imbens, Lei, and Luo (2021), 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 and Imbens (2022) and Arkhangelsky, Imbens, Lei, and Luo (2021) propose doubly robust estimation methods that leverage restrictions on outcome models and/or selection models with unconfoundedness-type restrictions; see also Athey, Bayati, Doudchenko, Imbens, and Khosravi (2021). Our results complement theirs as we maintain the parallel trends assumption and discuss the types of restrictions on selection compatible with it.

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; Chabé-Ferret, 2015; Verdier, 2020; Marx, Tamer, and Tang, 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.

Finally, we connect the DiD assumptions to the literature on nonparametric identification in nonseparable panel models.[4] 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, Fernández-Val, Hahn, and Newey, 2013) or restricting individual heterogeneity via nonparametric correlated random effects assumptions (e.g. Altonji and Matzkin, 2005; Bester and Hansen, 2009). We establish an explicit connection between DiD and the literature on nonseparable panel models. 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 $i = 1, \ldots, N$ 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 $\overset{d}{=}$ to denote equality of distribution. For random variables, X_i, Z_i, and W_i, $Z_i|W_i, X_i \overset{d}{=} Z_i|X_i, W_i$ denotes that $F_{Z_i|W_i,X_i}(z|w, x) = F_{Z_i|X_i,W_i}(z|w, x)$ for $(z, w, x) \in \mathcal{Z} \times \mathcal{W} \times \mathcal{X}$.

[4] See, e.g., Altonji and Matzkin (2005); Bester and Hansen (2009); Hoderlein and White (2012); Chernozhukov, Fernández-Val, Hahn, and Newey (2013); 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 Setup and parallel trends assumptions

Consider the classical DiD setup with two groups and two periods. While several recent papers have considered more general setups with multiple periods and groups, they typically show that these more general setups can be thought of as a sequence of DiD problems with two groups and two periods.[5] Thus, for expositional simplicity and clarity, we focus on the classical two-period, two-group setup.

Let D_{it} and Y_{it} denote the treatment status and outcome for individual (or unit) i in period t. The treatment group ($G_i = 1$) selects the following 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.[6]

Identification in DiD settings relies on parallel trends assumptions.[7] In Section 3, we abstract from covariates and consider the following unconditional parallel trends assumption.

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

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

Under Assumption PT, the unconditional average treatment effect on the treated (ATT) is identified as

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

In many applications, parallel trends may only be plausible conditional on covariates (e.g., Heckman, Ichimura, and Todd, 1997; Abadie, 2005; Sant'Anna and Zhao, 2020). 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.

In Section 4, we examine the model in (3), which is separable in observables and unobservables. We consider the following parallel trends assumption, which is conditional on the time series of the covariates.

[5] See, e.g., de Chaisemartin and D'Haultfœuille (2020); Callaway and Sant'Anna (2021); Goodman-Bacon (2021); Sun and Abraham (2021).

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

[7] For identification strategies that rely on alternative assumptions, see, e.g., Athey and Imbens (2006); Bonhomme and Sauder (2011); Callaway, Li, and Oka (2018); de Chaisemartin and D'Haultfœuille (2017); Callaway and Li (2019); D'Haultfœuille, Hoderlein, and Sasaki (2021).

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

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

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

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

The unconditional ATT can then be obtained by integrating out with respect to the distribution of X_i conditional on $G_i = 1$,

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

In Section 5, we examine the model in (4), which is nonseparable in observables and unobservables. In this context, it is crucial to differentiate between the covariates that interact with the unobservable determinants of selection, X^μ_{it}, and those that do not, X^λ_{it}. Intuitively, this is because in general we cannot have parallel trends between treatment and control subpopulations that experience changes in X^μ_{it} over time. We therefore examine the following modified version of Assumption PT-X.

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

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i, X^\lambda_i, X^\mu_{i1} = X^\mu_{i2}] = E[Y_{i2}(0) - Y_{i1}(0)|X^\lambda_i, X^\mu_{i1} = X^\mu_{i2}] \text{ a.s.}$$

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, $E[Y_{i2}(1) - Y_{i2}(0)|G_i = 1, X^\lambda_i, X^\mu_i]$. Instead, we can identify the following conditional ATT,

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

After integrating out with respect to the distribution of covariates, we can identify the ATT for subpopulations that do not experience changes in X^μ_{it},

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

It is important to note that if X^μ_{it} is time-invariant, then $X^\mu_{i1} = X^\mu_{i2}$ holds by definition such that Assumptions PT-X and PT-NSP are equivalent.

Remark 1 (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.* □

3 Selection and parallel trends in separable models

In this section, we examine the trade-off between restrictions on the selection mechanism and the distribution of unobservables in the context of the parallel trends assumption. In order to keep the presentation transparent, we start with a separable outcome model without covariates here and extend the analysis to covariates in Section 4 and to nonseparable models in Section 5.

3.1 Model

We consider a model for the potential outcomes without the treatment that is separable in the time-invariant and time-varying unobservables.

Assumption SP.

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

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, individual-specific unobservable.[8] 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 $\mathcal{E}$, respectively.[9]

Remark 2 (Two-way fixed effects model)**.** *Assumption SP does not impose the standard two-way fixed effects model for the realized outcomes,*

$$Y_{it} = \delta D_{it} + \alpha_i + \lambda_t + \varepsilon_{it},$$

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. □

To analyze the role of selection, it is useful to express Assumption PT equivalently as an orthogonality condition.

[8] The 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 and $E[\tilde{\lambda}_t] = \lambda_t$ and $\tilde{\varepsilon}_{it} = \varepsilon_{it} - (\tilde{\lambda}_t - \lambda_t)$.

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

Lemma 3.1 (Equivalence)**.** *Suppose that Assumption SP holds and that $P(G_i = 1) \in (0, 1)$. Then Assumption PT is equivalent to $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$.*

Recall that under Assumption SP, the counterfactual trend for each unit i, $Y_{i2}(0) - Y_{i1}(0)$, consists of two components: a common component, $\lambda_2 - \lambda_1$, and an individual-specific component, $\varepsilon_{i2} - \varepsilon_{i1}$. Lemma 3.1 allows us to state Assumption PT as an orthogonality condition between the selection indicator and the individual-specific component.

The key implication of Lemma 3.1 is that for Assumption PT to hold, we need to impose additional restrictions on the selection mechanism and/or the distribution of the time-varying unobservables. To formalize these additional restrictions, we consider a general selection mechanism in which units may select into treatment based on the unobservable determinants of the untreated potential outcomes, $(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$, as well as additional time-invariant and time-varying vectors of random variables, $(\nu_i, \eta_{i1}, \eta_{i2})$,

$$G_i = g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}). \tag{5}$$

This general selection mechanism accommodates many different types of selection, including random assignment, selection based on past outcomes, Roy-style selection based on treatment effects, and other selection mechanisms based on economic decision problems (e.g. Heckman and Robb, 1985; Chabé-Ferret, 2015). Let $\mathcal{G}_{\text{all}}$ denote the set of all selection mechanisms $g(\cdot)$ that map from the support of the unobservables to $\{0, 1\}$.

Under the selection mechanism (5), Assumption SP does not imply the PT assumption without further restrictions. In Section 3.2, 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.

3.2 Necessary and sufficient conditions for parallel trends

To better understand the implications of the parallel trends assumption, we derive necessary and sufficient conditions for this assumption in two scenarios. The first scenario is where researchers are not willing to impose a model for the selection mechanism $g(\cdot)$ and want parallel trends to hold for any $g(\cdot)$ in a prespecified class (Scenario I). In the second scenario, researchers are not willing to restrict the distribution of unobservables $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}}$ and want parallel trends to hold for all $F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}}$ (Scenario II). These two scenarios clarify the trade-offs between restrictions on selection into treatment and time-varying unobservables.

Remark 3 (Nonseparable models)**.** *In Appendix A, we derive necessary and sufficient con-*

ditions for a fully nonseparable, potentially time-varying potential outcome model,

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

where α_i*,* ε_{i1} *and* ε_{i2} *are random vectors. As we discuss in Remarks 4 and 5, the results for the nonseparable model highlight similar trade-offs as those for separable models.* □

3.2.1 Scenario I: Parallel trends for any selection mechanism

Suppose that 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 $\mathcal{G}$ is a (potentially restricted) class of selection mechanisms. This requirement can be interpreted as a "robustness" requirement. 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 $\mathcal{G}_{\text{all}}$.

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

Proposition 3.1 (Necessary and sufficient condition for $g \in \mathcal{G}_{\text{all}}$)**.** *Suppose that Assumption SP holds. Suppose further that* $P(G_i = 1) \in (0, 1)$*,* $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$*,* $P(\nu_i^1 > c) \in (0, 1)$ *for some* $c \in \mathbb{R}$*, and* $P(\varepsilon_{i2} \geq \varepsilon_{i1}) > 0$*. Then Assumption PT holds for any fixed* $g \in \mathcal{G}_{all}$ *if and only if* $\varepsilon_{i1} = \varepsilon_{i2}$ *a.s.*

The "if" direction of the proof is straightforward. The "only if" direction follows by noting that if Assumption PT holds for all $g \in \mathcal{G}_{\text{all}}$, then it holds for $g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{\varepsilon_{i2} \geq \varepsilon_{i1}\}$. By Lemmas 3.1 and B.1, this specific choice of selection mechanism can be shown to imply the result.

Proposition 3.1 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 restrictions on the selection mechanism. First, we examine a class of selection mechanisms in which selection does not depend on the time-varying unobservable determinant of $Y_{i2}(0)$,[10]

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

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 following class of selection mechanisms,

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

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 3.2 (Necessary and sufficient condition for $g \in \mathcal{G}_1$)**.** *Suppose that Assumption SP holds. Suppose further that $P(G_i = 1) \in (0,1)$, $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$, $P(\nu_i^1 > c) \in (0,1)$ for some $c \in \mathbb{R}$, and $P(E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] \geq \varepsilon_{i1}) > 0$. 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., $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$ a.s. is also sufficient for Assumption PT.*

Proposition 3.3 (Necessary and sufficient condition for $g \in \mathcal{G}_2$)**.** *Suppose that Assumption SP holds. Suppose further that $P(G_i = 1) \in (0,1)$, $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$, $P(\nu_i^1 > c) \in (0,1)$ for some $c \in \mathbb{R}$, and $P(E[\varepsilon_{i2}|\alpha_i] \geq E[\varepsilon_{i1}|\alpha_i]) > 0$. 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., $E[\varepsilon_{i1}|\alpha_i] = E[\varepsilon_{i2}|\alpha_i]$ 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 3.2 shows that for Assumption PT to hold for any fixed selection mechanism that is a trivial function 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. 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.

When we further restrict selection to be a trivial function of both ε_{i1} and ε_{i2}, Proposition 3.3 shows that parallel trends implies that the conditional mean of ε_{it} given α_i is stable over time. This condition is implied by (and weaker than) the textbook strict exogeneity assumption, $E[\varepsilon_{it}|G_i, \alpha_i] = 0$, as well as the time homogeneity assumption in Chernozhukov,

[10] The 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.

Fernández-Val, Hahn, and Newey (2013). If $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i]$, this condition is also sufficient for parallel trends.

Remark 4 (Nonseparable models: Scenario I)**.** *The results for fully nonseparable outcome models, $Y_{it}(0) = \xi_t(\alpha_i, \varepsilon_{it})$, in Appendix A highlight similar trade-offs as in the separable model. The necessary and sufficient condition for parallel trends for any fixed $g \in \mathcal{G}_{all}$ is $Y_{i1}(0) - E[Y_{i1}(0)] = Y_{i2}(0) - E[Y_{i2}(0)]$ a.s. (Proposition A.1). That is, the untreated potential outcome does not vary across time except for location shifts as in Proposition 3.1. It is worth noting that for outcomes with finite support, this condition would generally rule out location shifts.*

Similar to Propositions 3.2 and 3.3, Propositions A.2 and A.3 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 demeaned untreated potential outcome across time, respectively. □

3.2.2 Scenario II: Parallel trends for any distribution of unobservables

Consider a 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 3.1 (Completeness of a class of distributions)**.** *Let W be a vector of random variables. A family of distributions $\mathcal{F}$ is complete if*

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

implies

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

Proposition 3.4 (Necessary and sufficient condition for any distribution of unobservables)**.** *Suppose that Assumptions 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 3.4, 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}.

Remark 5 (Nonseparable models: Scenario II)**.** *For the fully nonseparable model $Y_{it}(0) = \xi_t(\alpha_i, \varepsilon_{it})$, Proposition A.4 shows that parallel trends holds if and only if $P(G_i = 1|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) = P(G_i = 1)$ a.s. This condition 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 3.4.* $\square$

Taken together, our necessary and sufficient conditions show that Assumption PT cannot hold absent additional restrictions on the selection mechanism and 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.

3.3 Primitive sufficient conditions for parallel trends

The results in the previous section 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, then parallel trends implies that the untreated potential outcomes are constant over time up to location shifts (Proposition 3.1). While, naturally, this condition is sufficient for parallel trends to hold, it is implausible in practice. Propositions 3.2 and 3.3 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 the previous section thus raises two questions: (i) Is there a set of primitive sufficient conditions that allows for selection on both ε_{i1} and ε_{i2} *and* allows $Y_{it}(0)$ 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).

3.3.1 Sufficient conditions

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 beyond location shifts across time. To state the condition, we define the class of symmetric selection mechanisms:

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

Assumption SC1. *The following conditions hold:*

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

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. Overall, Assumption SC1 consists of symmetry restrictions on how ε_{i1} and ε_{i2} enter the selection mechanism and the distribution of unobservables.

The next two sufficient conditions provide restrictions on the selection mechanism and the distribution of unobservables 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. $(\nu_i, \eta_{i1}, \eta_{i2})|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2} \overset{d}{=} (\nu_i, \eta_{i1}, \eta_{i2})|\alpha_i, \varepsilon_{i1}$.
3. $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] = \varepsilon_{i1}$.

Assumption SC3. *Suppose that 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. $(\nu_i, \eta_{i1}, \eta_{i2})|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2} \overset{d}{=} (\nu_i, \eta_{i1}, \eta_{i2})|\alpha_i$.
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 3.2 and 3.3, respectively.

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

Proposition 3.5 (Sufficient conditions for parallel trends)**.** *Suppose that Assumption SP holds. Suppose further that $P(G_i = 1) \in (0,1)$. 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 3.5 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,

$$\bar{g}(a, e_1, e_2) = E[G_i | \alpha_i = a, \varepsilon_{i1} = e_1, \varepsilon_{i2} = e_2]. \tag{6}$$

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 and Lemma 3.1.

3.3.2 Examples

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)**.** *Consider the following class of threshold-crossing selection mechanisms inspired by the selection mechanisms considered by Ashenfelter and Card (1985),*

$$G_i = 1\left\{Y_{i1}(0) + \beta Y_{i2}(0) \leq c\right\} = 1\left\{(1+\beta)\alpha_i + \varepsilon_{i1} + \beta\varepsilon_{i2} \leq \tilde{c}\right\}, \tag{7}$$

where $\beta \in [0,1]$ is a discount factor and $\tilde{c} = c - \lambda_1 - \beta\lambda_2$. In Equation (7), selection depends *only on the unobservable determinants of the untreated potential outcomes* $(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ *so that the projected selection mechanism is equal to the selection mechanism* $G_i = \bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$*.*

Under the selection model in Equation (7), Assumption SC1.1 requires that there is no discounting, $\beta = 1$*, so that selection depends on the sum of* $Y_{i1}(0)$ *and* $Y_{i2}(0)$*,*

$$G_i = 1\left\{Y_{i1}(0) + Y_{i2}(0) \leq c\right\} = 1\left\{2\alpha_i + \varepsilon_{i1} + \varepsilon_{i2} \leq \tilde{c}\right\}. \tag{8}$$

By contrast, Assumption SC2.1 requires that there is full discounting, $\beta = 0$*, so that selection depends only on* $Y_{i1}(0)$*,*

$$G_i = 1\left\{Y_{i1}(0) \leq c\right\} = 1\left\{\alpha_i + \varepsilon_{i1} \leq \tilde{c}\right\}. \tag{9}$$

The selection mechanism in Equation (9) corresponds to that considered on 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).* □

Next, we present an example of Roy-style selection based on treatment effects. An important takeaway from Proposition 3.5 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 (arbitrary functions of) time-varying treatment effects, as we illustrate in the following example.

Example 2 (Roy-style selection)**.** *Consider the following random coefficients model for the observed outcome,*

$$Y_{it} = \alpha_i + \delta_{it} D_{it} + \lambda_t + \epsilon_{it}. \tag{10}$$

Suppose that selection depends on an arbitrary function $f(\cdot)$ *of the treatment effects* $(\delta_{i1}, \delta_{i2})$ *as well as an individual-specific cost of treatment,* c_i*, that may depend on* α_i*,*

$$G_i = 1\{f(\delta_{i1}, \delta_{i2}) > c_i\}. \tag{11}$$

Since the selection mechanism in (11) does not depend on $(\varepsilon_{i1}, \varepsilon_{i2})$*, 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.,* $(\delta_{i1}, \delta_{i2}, c_i)|(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) \overset{d}{=} (\delta_{i1}, \delta_{i2}, c_i)|(\alpha_i, \varepsilon_{i2}, \varepsilon_{i1})$*. For Assumption SC2* *to hold, the treatment effects and costs have to be independent of* ε_{i2} *conditional on* $(\alpha_i, \varepsilon_{i1})$, *formally* $(\delta_{i1}, \delta_{i2}, c_i)|(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) \stackrel{d}{=} (\delta_{i1}, \delta_{i2}, c_i)|(\alpha_i, \varepsilon_{i1})$. *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* $(\delta_{i1}, \delta_{i2}, c_i) \perp (\varepsilon_{i1}, \varepsilon_{i2})|\alpha_i$.

This example illustrates that the time-varying treatment effects can enter the selection mechanism in an unrestricted way. In fact, $f(\cdot)$ *can depend on* δ_{i1} *and* δ_{i2} *asymmetrically (e.g.,* $f(\cdot)$ *could be a trivial function of one or the other, or there could be discounting). As a result, in the context of Roy-style selection, researchers do not have to impose any restrictions on how the selection mechanism depends on the treatment effects or costs. Instead, they have to justify the required distributional restrictions.* □

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. For example, Ashenfelter and Card (1985)-style selection on outcomes requires assumptions on discounting, while Roy-style selection does not require any restrictions on the discounting of the treatment effects. We note, however, that for parallel trends to hold in the Roy-style selection example, additional restrictions on the conditional distribution of the treatment effects have to be satisfied.

4 Covariates in the separable model

In this section, we introduce covariates in the separable model. Our goals in this section are two-fold. First, we show how conditioning on covariates weakens the sufficient conditions. Second, we demonstrate how the sufficient conditions can allow for selection on both observable and unobservable determinants of the untreated potential outcomes.

The following model extends Assumption SP to include a vector of time-invariant and time-varying covariates, X_{it}.

Assumption SP-X.

$$Y_{it}(0) = \alpha_i + \lambda_t + \gamma_t(X_{it}) + \varepsilon_{it}, \quad E[\varepsilon_{it}] = 0, \quad i = 1, \ldots, N, \quad t = 1, 2. \tag{12}$$

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,

$$\bar{g}(a, x_1, x_2, e_1, e_2) = E[G_i|\alpha_i = a, X_{i1} = x_1, X_{i2} = x_2, \varepsilon_{i1} = e_1, \varepsilon_{i2} = e_2]. \tag{13}$$

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. $\varepsilon_{i1}, \varepsilon_{i2}|\alpha_i, X_i \overset{d}{=} \varepsilon_{i2}, \varepsilon_{i1}|\alpha_i, X_i$.

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. $E[\varepsilon_{i2} - \varepsilon_{i1}|X_i, \alpha_i, \varepsilon_{i1}] = E[\varepsilon_{i2} - \varepsilon_{i1}|X_i]$.

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.[11] 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.

Proposition 4.1. *Suppose that Assumptions SP-X holds. Suppose further that* $P(G_i = 1|X_i) \in (0, 1)$ *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.*

[11] To 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, $E[\varepsilon_{it}] = 0$.

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 the sufficient conditions in Proposition 4.1 allow for selection on unobservable determinants of the untreated potential outcome. This is in contrast with the unconfoundedness-type 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.

The conclusions in this section crucially depend on the separability between covariates and the unobservables that determine selection. Assumption PT-X will generally not hold in models where α_i and X_{it} interact. A simple example is a correlated random coefficients model (e.g., Chamberlain, 1992), $Y_{it}(0) = \alpha_i X_{it} + \lambda_t + \varepsilon_{it}$, where the scalar X_{it} and α_i enter the selection mechanism. In the next section, we relax the separability restriction and demonstrate the implications of relaxing separability for the type of conditional parallel trends assumptions that can hold in this setting.

Remark 6 (Unconditional parallel trends and covariate-specific trends)**.** *A natural question is whether and when unconditional parallel trends (Assumption PT) continue 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.* □

5 Selection in a nonseparable model with covariates

So far, we have studied separable models to keep the presentation transparent. However, since DiD is a model-agnostic reduced-form approach, it is crucial to generalize the results to nonseparable models for both practical and theoretical reasons. In doing so, we establish interesting connections between our sufficient conditions for parallel trends and identifying assumptions from the nonseparable panel literature. Our results have important implications for the choice of covariates in DiD analyses. See Section 6 for additional discussion of the practical implications.

The necessary and sufficient conditions we provide in Section 3 illustrate the trade-offs between restrictions on selection and the time-varying unobservables. Because the nonseparable model nests the separable model as a special case, these trade-offs remain relevant for this more general class of models as we show in Appendix A. We therefore focus on primitive sufficient conditions for Assumption PT-NSP in this section.

5.1 Model

We consider the following nonseparable model, which nests the models in Assumptions SP and SP-X.

Assumption NSP.

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

where X_{it}^{μ}, X_{it}^{λ}, α_i^{μ}, α_i^{λ}, ε_{it}^{μ}, and $\varepsilon_{it}^{\lambda}$ are finite-dimensional vector-valued random variables.

The above model consists of time-invariant and time-varying nonseparable components. Without further restrictions on the unobservables, the additive structure in Assumption NSP is without loss of generality and the superscripts μ and λ are merely labels. Indeed, it is possible that $X_{it}^{\mu} = X_{it}^{\lambda}$, $\alpha_i^{\mu} = \alpha_i^{\lambda}$, and $\varepsilon_{it}^{\mu} = \varepsilon_{it}^{\lambda}$, which implies that the model is fully nonseparable and time-varying in an arbitrary way. In the following, we use $\mathcal{X}_{\mu}$, $\mathcal{X}_{\lambda}$, $\mathcal{A}$, 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.[12] We therefore impose the following condition on the projected selection mechanism.

Assumption SEL-NSP.

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

Assumption SEL-NSP allows the projected selection mechanism to depend on all covariates, but only on the unobservables that enter the time-invariant component of the structural

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

function. In view of Assumption SEL-NSP, we define the following notation:

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

5.2 Sufficient conditions

Here 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. $\varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu|\alpha_i^\mu, X_i^\mu, X_i^\lambda \overset{d}{=} \varepsilon_{i2}^\mu, \varepsilon_{i1}^\mu|\alpha_i^\mu, X_i^\mu, X_i^\lambda$.
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 distribution of $(\varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)$ to be exchangeable conditional on $(\alpha_i^\mu, X_i^\mu, X_i^\lambda)$. 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 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. $\varepsilon_{i1}^\mu | \alpha_i^\mu, X_i^\mu, X_i^\lambda \overset{d}{=} \varepsilon_{i2}^\mu | \alpha_i^\mu, X_i^\mu, X_i^\lambda$.

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 the time-invariant component, to be time-invariant conditional on $(\alpha_i^\mu, X_i^\mu, X_i^\lambda)$. On the other hand, the unobservables that enter the time-varying component, $(\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda)$, are required to be independent of the unobservables that determine selection, α_i^μ, conditional on (X_i^μ, X_i^λ).

Looking across the three sufficient conditions, we can see that each of them 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 shows that each of these conditions is sufficient for Assumption PT-NSP.

Proposition 5.1. *Suppose that Assumptions NSP and SEL-NSP hold. Suppose further that* $P(G_i = 1 | X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu) \in (0, 1)$ *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.*

5.3 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.[13] 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 panel literature, and those pertaining to nonseparable models in particular.

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, Fernández-Val, Hahn, and Newey, 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. To simplify the exposition, we abstract from covariates and assume that the selection mechanism only depends on the unobservables that enter $\mu(\cdot)$, $G_i = g(\alpha_i^\mu, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)$.

The following assumption is the time homogeneity assumption from Chernozhukov, Fernández-Val, Hahn, and Newey (2013) imposed on ε_{it}^μ in Assumption NSP.

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

Assumption TH requires the distribution of ε_{it}^μ to be homogeneous across time conditional on G_i and α_i^μ. However, it does not impose any restrictions on the conditional distribution of ε_{it}^μ given G_i and α_i^μ. Furthermore, there are no restrictions imposed on the distribution of $\alpha_i^\mu | G_i$, consistent with the notion of fixed effects in the nonseparable panel literature (Evdokimov, 2010; Hoderlein and White, 2012; Chernozhukov, Fernández-Val, Hahn, and Newey, 2013).

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

Assumption RE. $(\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda) | G_i \overset{d}{=} (\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda)$.

[13] See, e.g., Chamberlain (1984); Arellano (2003); Evdokimov (2010); Wooldridge (2010); Hoderlein and White (2012); Chernozhukov, Fernández-Val, Hahn, and Newey (2013).

Assumption RE is an 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 random assignment, $(Y_{i1}(0), Y_{i2}(0)) \perp G_i$, since selection into treatment can depend on the unobservables entering the time-invariant component $\mu(\cdot)$. With covariates, Assumption RE is a correlated random effects restriction that takes the form of a conditional independence assumption.

It is straightforward to show that Assumptions TH and RE imply Assumption PT. 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 RE.[14] This result demonstrates how restrictions on the selection mechanism can be used to justify combinations of Assumptions TH and RE.

Proposition 5.2. *Suppose that Assumption NSP holds with $X_{it}^{\mu} = X_{it}^{\lambda} = \emptyset$. Suppose further that $G_i = g(\alpha_i^{\mu}, \varepsilon_{i1}^{\mu}, \varepsilon_{i2}^{\mu})$. Then (i) Assumption SC1-NSP with $g(\cdot)$ in lieu of $\bar{g}(\cdot)$ implies Assumptions TH and RE if $P(G_i = 1|\alpha_i^{\mu} = a) \in (0, 1)$ for all $a \in \mathcal{A}$ and (ii) Assumption SC3-NSP with $g(\cdot)$ in lieu of $\bar{g}(\cdot)$ implies Assumptions TH and RE.*

Given the selection mechanisms in Assumptions SC1-NSP and SC3-NSP, the results in Proposition 5.2 follow from two observations. First, the restrictions on the unobservables that enter $\mu(\cdot)$ constitute primitive conditions for Assumption TH. Second, the independence conditions imposed on the unobservables determining selection and those entering $\lambda_t(\cdot)$ imply Assumption RE.

Proposition 5.2 sheds light on the connection between selection, fixed effects, and correlated random effects in our nonseparable DiD framework. On the one hand, Assumptions SC1-NSP and SC3-NSP allow the distribution of $\alpha_i^{\mu}|G_i$ 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 independent of the determinants of selection and therefore independent of G_i, consistent with the notion of random effects.

6 Implications for practice

This paper studies the connection between selection and parallel trends in DiD analyses. We first provide necessary and sufficient conditions that demonstrate that researchers relying on parallel trends assumptions implicitly impose restrictions on how selection depends on unobservables. We then derive primitive sufficient conditions on selection for parallel trends

[14]In the context of correlated random coefficient models, Graham and Powell (2012) impose a similar structure on their random coefficient. They assume that the random coefficient consists of a time-invariant and time-varying component.

assumptions with and without covariates. These conditions provide practitioners with new and explicit theory-based templates for justifying parallel trends assumptions based contextual information on the selection mechanism.

The menu of primitive sufficient 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 our conditions use contextual information to assess and explicitly discuss which determinants of the untreated potential outcome affect selection. 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. These restrictions are typically more plausible the closer the pre- and post-treatment period are.

Our theoretical results have important implications for the role of covariates in DiD analyses. First, they clarify how covariates can weaken restrictions on selection. We show that 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. Second, conditioning on covariates makes the distributional restrictions on unobservables weaker/more plausible. However, even after conditioning on covariates, researchers still have to take a stance on how selection depends on the remaining unobservables. We emphasize that our recommendations only apply to covariates that are unaffected by the treatment.

Finally, 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 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 covariates that interact with the unobservable determinants of selection in the outcome model, researchers have to condition on these covariates not changing over time. Indeed, even if treatment and control groups experience the same change in these covariates, the two groups may not exhibit the same counterfactual trends.

References

Abadie, A. (2005): "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 72(1), 1–19. 6

Altonji, J. G., and R. L. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73(4), 1053–1102. 3, 5, 14, 21, 23

Angrist, J. D., and A. B. Krueger (1999): "Chapter 23 - Empirical Strategies in Labor Economics," vol. 3 of *Handbook of Labor Economics*, pp. 1277–1366. Elsevier. 4

Arellano, M. (2003): *Panel Data Econometrics*. Oxford University Press. 5, 23

Arellano, M., and S. Bonhomme (2011): "Nonlinear Panel Data Analysis," *Annual Review of Economics*, 3, 395–424. 5

Arellano, M., and S. Bonhomme (2012): "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *The Review of Economic Studies*, 79(3), 987–1020. 5

Arellano, M., and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5. Elsevier Science. 5

Arkhangelsky, D., and G. W. Imbens (2022): "Double-Robust Identification for Causal Panel Data Models," *arXiv:1909.09412 [econ]*. 4

Arkhangelsky, D., G. W. Imbens, L. Lei, and X. Luo (2021): "Double-Robust Two-Way-Fixed-Effects Regression For Panel Data," *arXiv:2107.13737 [econ]*. 4

Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, 60(1), 47–57. 4

Ashenfelter, O. C., and D. Card (1985): "Using the longitudinal structure of earnings to estimate the effect of training programs," *The Review of Economics and Statistics*, 67(4), 648–660. 4, 15, 16, 17

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2021): "Matrix Completion Methods for Causal Panel Data Models*," *Journal of the American Statistical Association*, 116(536), 1716–1730. 4

Athey, S., and G. W. Imbens (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74(2), 431–497. 6

Bester, C. A., and C. Hansen (2009): "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," *Journal of Business and Economic Statistics*, 27(2), 235–250. 5, 23

Bonhomme, S., and U. Sauder (2011): "Recovering Distributions in Difference-in-Differences models: A Comparison of Selective and Comprehensive Schooling," *Review of Economics and Statistics*, 93(May), 479–494. 6

Borusyak, K., X. Jaravel, and J. Spiess (2021): "Revisiting Event Study Designs: Robust and Efficient Estimation," *arXiv:2108.12419 [econ]*. 1

Caetano, C., B. Callaway, S. Payne, and H. S. Rodrigues (2022): "Difference in Differences with Time-Varying Covariates," *arXiv:2202.02903*. 3

Callaway, B., and T. Li (2019): "Quantile treatment effects in difference in differences models with panel data," *Quantitative Economics*, 10(4), 1579–1618. 6

Callaway, B., T. Li, and T. Oka (2018): "Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods," *Journal of Econometrics*, 206(2), 395–413, Special issue on Advances in Econometric Theory: Essays in honor of Takeshi Amemiya. 6

Callaway, B., and P. H. C. Sant'Anna (2021): "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 225(2), 200–230. 1, 6

Card, D. (1990): "The Impact of the Mariel Boatlift on the Miami Labor Market," *ILR Review*, 43(2), 245–257. 4

Card, D., and A. B. Krueger (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. 4

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. 4, 9

Chamberlain, G. (1982): "Multivariate regression models for panel data," *Journal of Econometrics*, 18(1), 5–46. 5

——— (1984): "Chapter 22 Panel data," vol. 2 of *Handbook of Econometrics*, pp. 1247–1318. Elsevier. 5, 23

——— (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60(3), 567–596. 5, 19

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81(2), 535–580. 3, 5, 11, 23

de Chaisemartin, C., and X. D'Haultfœuille (2017): "Fuzzy Differences-in-Differences," *The Review of Economic Studies*, 85(2), 999–1028. 6

——— (2020): "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110(9), 2964–2996. 1, 6

——— (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. 4

D'Haultfœuille, X., S. Hoderlein, and Y. Sasaki (2021): "Nonparametric difference-in-differences in repeated cross-sections with continuous treatments," arXiv preprint arXiv:2104.14458. 6

Evdokimov, K. (2010): "Identification and estimation of a nonparametric panel data model with unobserved heterogeneity," *Department of Economics, Princeton University*, 1. 5, 23

Freyberger, J. (2017): "Non-parametric Panel Data Models with Interactive Fixed Effects," *The Review of Economic Studies*, 85(3), 1824–1851. 5

Gardner, J. (2021): "Two-stage differences in differences," *Working Paper*. 1

Ghanem, D. (2017): "Testing identifying assumptions in nonseparable panel data models," *Journal of Econometrics*, 197(2), 202–217. 5, 21, 23

Goodman-Bacon, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225(2), 254–277. 1, 6

Graham, B. S., and J. L. Powell (2012): "Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models," *Econometrica*, 80(5), 2105–2152. 5, 24

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64(4), 605–654. 6

Heckman, J. J., and R. Robb (1985): "Alternative methods for evaluating the impact of interventions: An overview," *Journal of Econometrics*, 30(1), 239–267. 4, 9

Hoderlein, S., and H. White (2012): "Nonparametric identification in nonseparable panel data models with generalized fixed effects," *Journal of Econometrics*, 168(2), 300–314. 5, 23

Honoré, B., and E. Kyriazidou (2000a): "Estimation of Tobit-type models with individual specific effects," *Econometric Reviews*, 19, 341–366. 5

Honoré, B. E. (1993): "Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables," *Journal of Econometrics*, 59(1–2), 35–61. 5

Honoré, B. E., and E. Kyriazidou (2000b): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68(4), 839–874. 5

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics*, 86(1), 4–29. 4, 19

Imbens, G. W., and J. M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86. 4, 19

Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65(6), 1335–1364. 5

Lechner, M. (2010): "The Estimation of Causal Effects by Difference-in-Difference Methods," *Foundations and Trends in Econometrics*, 4(3), 165–224. 4

Lehmann, E., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Springer Texts in Statistics. 12, 5, 9

Manski, C. F. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55(2), 357–362. 5

Marcus, M., and P. H. C. Sant'Anna (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. 1

Marx, P., E. Tamer, and X. Tang (2022): "Parallel Trends and Dynamic Choices," arXiv:2207.06564. 4

McKenzie, D. (2022): "A new synthesis and key lessons from the recent difference-in-differences literature," World Bank Blogs (Link), Accessed: 2022-02-22. 1

Meyer, B. D., W. K. Viscusi, and D. L. Durbin (1995): "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment," *The American Economic Review*, 85(3), 322–340. 4

Mundlak, Y. (1961): "Empirical Production Function Free of Management Bias," *Journal of Farm Economics*, 43(1), 44–56. 5

——— (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1), 69–85. 5, 23

Roth, J., and P. H. C. Sant'Anna (2021): "When Is Parallel Trends Sensitive to Functional Form?," *arXiv:2010.04814 [econ, stat]*. 4, 8

Roth, J., P. H. C. Sant'Anna, A. Bilinski, and J. Poe (2022): "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *arXiv:2201.01194 [econ]*. 4

Sant'Anna, P. H. C., and J. Zhao (2020): "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, 219(1), 101–122. 6

Sun, L., and S. Abraham (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225(2), 175–199. 1, 6

Verdier, V. (2020): "Average treatment effects for stayers with correlated random coefficient models of panel data," *Journal of Applied Econometrics*, 35(7), 917–939. 4

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*. The MIT Press, Cambridge, MA and London, England. 5, 23

Wooldridge, J. M. (2021): "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators," *Working Paper*. 1

Appendix to *Selection and Parallel Trends*

Dalia Ghanem Pedro H. C. Sant'Anna Kaspar Wüthrich

A Necessary and sufficient conditions for 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 and a general selection mechanism that can depend on all unobservable determinants of the outcome as well as additional unobservables $(\nu_i, \eta_{i1}, \eta_{i2})$, where ν_i, η_{i1} and η_{i2} are vector-valued random variables.

Assumption A.1 (Nonseparable model)**.**

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

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

Consider the following general selection mechanism

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

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

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

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

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 A.1 (Necessary and sufficient condition for $g \in \mathcal{G}_{\text{all}}$)**.** *Suppose that Assumption A.1 holds. Suppose further that $P(G_i = 1) \in (0, 1)$, $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$, $P(\nu_i^1 > c) \in (0, 1)$ for some $c \in \mathbb{R}$, and $P(\tilde{Y}_{i2}(0) \geq \tilde{Y}_{i1}(0)) > 0$. Assumption PT holds for any fixed $g \in \mathcal{G}_{all}$ if and only if $\tilde{Y}_{i1}(0) = \tilde{Y}_{i2}(0)$ a.s.*

Proof. **"$\Longrightarrow$":** 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

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

By Lemma B.3, Assumption PT holding for $\breve{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0) \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0, \tag{14}$$

which, by $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ and $P(\nu_i^1 > c) > 0$, is equivalent to

$$E[1\{\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0) \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0. \tag{15}$$

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 B.1.

"$\Longleftarrow$": 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.3. This completes the proof. $\square$

Proposition A.2 (Necessary and sufficient condition for $g \in \mathcal{G}_1$)**.** *Suppose that Assumption A.1 holds. Suppose further that $P(G_i = 1) \in (0, 1)$, $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$, $P(\nu_i^1 > c) \in (0, 1)$ for some $c \in \mathbb{R}$, and $P(E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] \geq \varepsilon_{i1}) > 0$. 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, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i, \varepsilon_{i1}]$, $E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] = \tilde{Y}_{i1}(0)$ a.s. is also sufficient for Assumption PT.

Proof. **"$\Longrightarrow$":** 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 $\breve{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0) \geq 0\}$. By Lemma B.3, Assumption PT holding for $\breve{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0) \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0, \tag{16}$$

which, by $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ and $P(\nu_i^1 > c) > 0$, is equivalent to

$$E[1\{E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0) \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0. \tag{17}$$

By the law of iterated expectations, this is further equivalent to

$$E[1\{E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0) \geq 0\}(E[\tilde{Y}_{i2}(0)|\alpha_i, \varepsilon_{i1}] - \tilde{Y}_{i1}(0))] = 0. \tag{18}$$

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 B.1.

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

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

where the first equality follows from the law of iterated expectations. 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.3, this completes the proof. ∎

Proposition A.3 (Necessary and sufficient condition $g \in \mathcal{G}_2$)**.** *Suppose that Assumption A.1 holds. Suppose further that $P(G_i = 1) \in (0, 1)$, $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$, $P(\nu_i^1 > c) \in (0, 1)$ for some $c \in \mathbb{R}$, and $P(E[\tilde{Y}_{i2}(0)|\alpha_i] \geq E[\tilde{Y}_{i1}(0)|\alpha_i]) > 0$. 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, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i]$, then $E[\tilde{Y}_{i1}(0)|\alpha_i] = E[\tilde{Y}_{i2}(0)|\alpha_i]$ a.s. is also sufficient for Assumption PT.*

Proof. **"$\Longrightarrow$":** 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 $\breve{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] \geq 0\}$. By Lemma B.3, Assumption PT holding for $\breve{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{E[E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0, \tag{20}$$

which, by $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ and $P(\nu_i^1 > c) > 0$, is equivalent to

$$E[1\{E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] \geq 0\}(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] = 0. \tag{21}$$

By the law of iterated expectations, this is further equivalent to

$$E[1\{E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i] \geq 0\}E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i]] = 0. \tag{22}$$

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 B.1.

"$\Longleftarrow$": We show that if, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i]$ holds, 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

$$\begin{aligned} E[G_i(\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))] &= E[E[E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}](\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))|\alpha_i]] \\ &= E[E[E[G_i|\alpha_i](\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0))|\alpha_i]] \\ &= E[E[G_i|\alpha_i]E[\tilde{Y}_{i2}(0) - \tilde{Y}_{i1}(0)|\alpha_i]] = 0, \end{aligned} \tag{23}$$

where the first equality follows by the law of iterated expectations. 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.3, this completes the proof. □

Proposition A.4 (Necessary and sufficient condition for parallel trends for any distribution of unobservables)**.** *Suppose that Assumptions A.1 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 $P(\tilde{Y}_{i1}(0) = \tilde{Y}_{i2}(0)) = 0$ 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 $F_{\alpha_i,\varepsilon_{i1},\varepsilon_{i2},\nu_i,\eta_{i1},\eta_{i2}} \in \mathcal{F}$.*

Proof. **"$\Longrightarrow$":** By Lemma B.3, 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

$$E[\bar{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})(\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})-\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2}))]=0, \tag{24}$$

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 $\tilde{\xi}_t(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})=\xi_t(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})-E[Y_{it}(0)]$ for $t=1,2$. The equivalence between $E[G_i(\tilde{Y}_{i2}(0)-\tilde{Y}_{i1}(0))]=0$ and (24) follows by the law of iterated expectations 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 holds for all $F_{\alpha_i,\varepsilon_{i1},\varepsilon_{i2},\eta_{i1},\eta_{i2}}\in\mathcal{F}$ is equivalent to

$$E[\bar{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})(\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})-\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2}))]=0, \tag{25}$$

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 (Lehmann and Romano, 2005, p.115) that

$$P(\bar{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})(\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})-\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2}))=0)=1 \quad \text{for all } F_{\alpha_i,\varepsilon_{i1},\varepsilon_{i2},\eta_{i1},\eta_{i2}}\in\mathcal{F}. \tag{26}$$

Now note that the left-hand side of (26) can be simplified as follows,

$$\begin{aligned}
&P(\bar{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})(\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})-\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})=0)\\
=&P(\bar{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})(\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})-\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})),\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})=\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2}))\\
&+P(\bar{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})(\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})-\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})),\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})\neq\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2}))\\
=&P(\bar{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})(\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})-\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2}))|\tilde{\xi}_2(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})\neq\tilde{\xi}_1(\alpha_i,\varepsilon_{i1},\varepsilon_{i2}))\\
=&P(\bar{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})=0)=1,
\end{aligned} \tag{27}$$

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

$$P(E[g(\alpha_i,\varepsilon_{i1},\varepsilon_{i2},\nu_i,\eta_{i1},\eta_{i2})|\alpha_i,\varepsilon_{i1},\varepsilon_{i2}]=E[G_i])=1 \quad \text{for all } F_{\alpha_i,\varepsilon_{i1},\varepsilon_{i2},\nu_i,\eta_{i1},\eta_{i2}}\in\mathcal{F}. \tag{28}$$

"$\Longleftarrow$": The if statement follows by the law of iterated expectations. 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 law of iterated expectations implies the following equality

$$\begin{aligned}
E[G_i(\tilde{Y}_{i2}(0)-\tilde{Y}_{i1}(0))]&=E[E[G_i|\alpha_i,\varepsilon_{i1},\varepsilon_{i2}](\tilde{Y}_{i2}(0)-\tilde{Y}_{i1}(0))]\\
&=E[E[G_i](\tilde{Y}_{i2}(0)-\tilde{Y}_{i1}(0))]=0.
\end{aligned} \tag{29}$$

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. $\square$

B Proofs

B.1 Proof of Lemma 3.1

Under Assumption SP, Assumption PT can be rewritten as

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

It follows that Assumption PT holds if and only if $E[\varepsilon_{i2} - \varepsilon_{i1}|G_i] = E[\varepsilon_{i2} - \varepsilon_{i1}] = 0$, where the last equality follows since $E[\varepsilon_{it}] = 0$ for $t = 1, 2$. It remains to show that (a) $E[\varepsilon_{i2} - \varepsilon_{i1}|G_i] = 0$ if and only if (b) $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$. Since $P(G_i = 1) \in (0, 1)$, the conclusion that (a) $\Rightarrow$ (b) is immediate. Therefore, we are left to show that (b) $\Rightarrow$ (a). Note that $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$ implies $E[\varepsilon_{i2} - \varepsilon_{i1}|G_i = 1] = 0$ since

$$E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = E[\varepsilon_{i2} - \varepsilon_{i1}|G_i = 1]P(G_i = 1)$$

and $P(G_i = 1) \in (0, 1)$ by assumption. To show that $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = 0$ implies $E[\varepsilon_{i2} - \varepsilon_{i1}|G_i = 0] = 0$, we subtract $E[\varepsilon_{i2} - \varepsilon_{i1}]$ from both sides of the former equality, multiply by -1, which yields

$$E[(1 - G_i)(\varepsilon_{i2} - \varepsilon_{i1})] = 0$$

since $E[\varepsilon_{i2} - \varepsilon_{i1}] = 0$ (Assumption SP). This implies $E[\varepsilon_{i2} - \varepsilon_{i1}|G_i = 0] = 0$ since $P(G_i = 0) \in (0, 1)$ and, by definition,

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

$\square$

B.2 Proof of Proposition 3.1

"$\Longrightarrow$": 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 $\breve{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{\varepsilon_{i2} - \varepsilon_{i1} \geq 0\}$. By Lemma 3.1, Assumption PT holding for $\breve{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{\varepsilon_{i2} - \varepsilon_{i1} \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0, \tag{30}$$

which, by $\nu_i^1 \perp (\varepsilon_{i1}, \varepsilon_{i2})$ and $P(\nu_i^1 > c) > 0$, is equivalent to

$$E[1\{\varepsilon_{i2} - \varepsilon_{i1} \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0. \tag{31}$$

Since $E[\varepsilon_{i2} - \varepsilon_{i1}] = 0$ by assumption, the above equality implies $\varepsilon_{i2} - \varepsilon_{i1} = 0$ a.s. by Lemma B.1.

"$\Longleftarrow$": 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 3.1. This completes the proof. □

B.3 Proof of Proposition 3.2

"$\Longrightarrow$": 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 $\breve{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) = 1\{\nu_i^1 > c\}1\{E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} \geq 0\}$. By Lemma 3.1, Assumption PT holding for $\breve{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1 > c\}1\{E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0, \tag{32}$$

which, by $\nu_i^1 \perp (\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})$ and $P(\nu_i^1 > c) > 0$, is equivalent to

$$E[1\{E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} \geq 0\}(\varepsilon_{i2} - \varepsilon_{i1})] = 0. \tag{33}$$

By the law of iterated expectations, this is further equivalent to

$$E[1\{E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} \geq 0\}(E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1})] = 0. \tag{34}$$

Since $E[\varepsilon_{i2} - \varepsilon_{i1}] = 0$ by assumption, the above equality implies $E[\varepsilon_{i2}|\alpha_i, \varepsilon_{i1}] - \varepsilon_{i1} = 0$ a.s. by Lemma B.1.

"$\Longleftarrow$": We show that if, in addition, $E[G_i|\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i|\alpha_i, \varepsilon_{i1}]$, 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 law of iterated expectations. 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 3.1, this implies the result. The proof is complete. $\square$

B.4 Proof of Proposition 3.3

"$\Longrightarrow$": We show that if Assumption PT holds for any $g\in\mathcal{G}_2$, then $E[\varepsilon_{i1}|\alpha_i]=E[\varepsilon_{i2}|\alpha_i]$. Since Assumption PT holds for any fixed $g\in\mathcal{G}_2$, then it holds for $\breve{g}(\alpha_i,\varepsilon_{i1},\varepsilon_{i2},\nu_i,\eta_{i1},\eta_{i2})=1\{\nu_i^1>c\}1\{E[\varepsilon_{i2}|\alpha_i]-E[\varepsilon_{i1}|\alpha_i]\geq 0\}$. By Lemma 3.1, Assumption PT holding for $\breve{g}(\cdot)$ is equivalent to

$$E[1\{\nu_i^1>c\}1\{E[\varepsilon_{i2}|\alpha_i]-E[\varepsilon_{i1}|\alpha_i]\geq 0\}(\varepsilon_{i2}-\varepsilon_{i1})]=0, \tag{35}$$

which, by $\nu_i^1\perp(\alpha_i,\varepsilon_{i1},\varepsilon_{i2})$ and $P(\nu_i^1>c)>0$, is equivalent to

$$E[1\{E[\varepsilon_{i2}|\alpha_i]-E[\varepsilon_{i1}|\alpha_i]\geq 0\}(\varepsilon_{i2}-\varepsilon_{i1})]=0. \tag{36}$$

By the law of iterated expectations, this is further equivalent to

$$E[1\{E[\varepsilon_{i2}|\alpha_i]-E[\varepsilon_{i1}|\alpha_i]\geq 0\}(E[\varepsilon_{i2}|\alpha_i]-E[\varepsilon_{i1}|\alpha_i])]=0. \tag{37}$$

Since $E[\varepsilon_{i2}-\varepsilon_{i1}]=0$ by assumption, the above equality implies that $E[\varepsilon_{i2}|\alpha_i]-E[\varepsilon_{i1}|\alpha_i]=0$ a.s. by Lemma B.1.

"$\Longleftarrow$": We show that if, in addition, $E[G_i|\alpha_i,\varepsilon_{i1},\varepsilon_{i2}]=E[G_i|\alpha_i]$, 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 law of iterated expectations. 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 by Lemma 3.1. This completes the proof. □

B.5 Proof of Proposition 3.4

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

"$\Longrightarrow$": 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 3.1, Assumption PT is equivalent to $E[G_i(\varepsilon_{i2}-\varepsilon_{i1})]=0$, which in turn is equivalent to

$$E[\bar{g}(\varepsilon_{i1},\varepsilon_{i2})(\varepsilon_{i2}-\varepsilon_{i1})]=0, \tag{38}$$

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 law of iterated expectations 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

$$E[\bar{g}(\varepsilon_{i1},\varepsilon_{i2})(\varepsilon_{i2}-\varepsilon_{i1})]=0 \quad \text{for all } F_{\alpha_i,\varepsilon_{i1},\varepsilon_{i2},\nu_i,\eta_{i1},\eta_{i2}}\in\mathcal{F}. \tag{39}$$

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

$$P(\bar{g}(\varepsilon_{i1},\varepsilon_{i2})(\varepsilon_{i2}-\varepsilon_{i1})=0)=1 \quad F_{\alpha_i,\varepsilon_{i1},\varepsilon_{i2},\nu_i,\eta_{i1},\eta_{i2}}\in\mathcal{F}. \tag{40}$$

Now note that the left-hand side of (40) can be simplified as follows,

$$
\begin{aligned}
&P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0) \\
=&P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0, \varepsilon_{i1} = \varepsilon_{i2}) + P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0, \varepsilon_{i1} \neq \varepsilon_{i2}) \\
=&P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) = 0 | \varepsilon_{i1} \neq \varepsilon_{i2}) \\
=&P(\bar{g}(\varepsilon_{i1}, \varepsilon_{i2}) = 0) = 1, \qquad (41)
\end{aligned}
$$

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

$$
P(E[g(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}) | \varepsilon_{i1}, \varepsilon_{i2}] = E[G_i]) = 1 \quad \text{for all } F_{\alpha_i, \varepsilon_{i1}, \varepsilon_{i2}, \nu_i, \eta_{i1}, \eta_{i2}} \in \mathcal{F}. \qquad (42)
$$

"$\Longleftarrow$": We now 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

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

where the first equality follows by the law of iterated expectations 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. $\square$

B.6 Proof of Proposition 3.5

We prove the three statements separately. By Lemma 3.1, 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 $(a, e_1, e_2) \in \mathcal{A} \times \mathcal{E}^2$

$$
\begin{aligned}
\bar{g}(a, e_1, e_2) &= \int g(a, e_1, e_2, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2} | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}}(v, t_1, t_2 | a, e_1, e_2) \\
&= \int g(a, e_2, e_1, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2} | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}}(v, t_1, t_2 | a, e_2, e_1) = \bar{g}(a, e_2, e_1), \qquad (43)
\end{aligned}
$$

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

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

The second equality follows from the symmetry restrictions on $\bar{g}(\cdot)$ and $F_{\varepsilon_{i1},\varepsilon_{i2}|\alpha_i}$. Together, they imply that the difference in the conditional expectations in parenthesis equals 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 $(a,e_1,e_2,e_2') \in \mathcal{A}\times\mathcal{E}^3$, $e_2 \neq e_2'$,

$$\begin{aligned}
\bar{g}(a,e_1,e_2) &= \int g(a,e_1,e_2,v,t_1,t_2) dF_{\nu_i,\eta_{i1},\eta_{i2}|\alpha_i,\varepsilon_{i1},\varepsilon_{i2}}(v,t_1,t_2|a,e_1,e_2) \\
&= \int g(a,e_1,e_2',v,t_1,t_2) dF_{\nu_i,\eta_{i1},\eta_{i2}|\alpha_i,\varepsilon_{i1}}(v,t_1,t_2|a,e_1) = \bar{g}(a,e_1,e_2'), \qquad (44)
\end{aligned}$$

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 $\check{\bar{g}}(a,e_1) = \bar{g}(a,e_1,e_2)$.

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

$$E[G_i(\varepsilon_{i2}-\varepsilon_{i1})] = E[\bar{g}(a,e_1,e_2)(\varepsilon_{i2}-\varepsilon_{i1})|\alpha_i,\varepsilon_{i1}]] = E[\check{\bar{g}}(\alpha_i,\varepsilon_{i1})(E[\varepsilon_{i2}|\alpha_i,\varepsilon_{i1}]-\varepsilon_{i1})] = 0, \qquad (45)$$

where the second equality follows since $\bar{g}(\cdot)$ is a trivial function of ε_{i2} implied by Assumptions SC2.1 and 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 $e_2 \neq e_2'$,

$$\bar{g}(a, e_1, e_2) = \int g(a, e_1, e_2, v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2} | \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}}(v, t_1, t_2 | a, e_1, e_2)$$
$$= \int g(a, e_1', e_2', v, t_1, t_2) dF_{\nu_i, \eta_{i1}, \eta_{i2} | \alpha_i}(v, t_1, t_2 | a) = g(a, e_1', e_2'), \tag{46}$$

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

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

$$E[G_i(\varepsilon_{i2} - \varepsilon_{i1})] = E[E[\bar{g}(\alpha_i, \varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1}) | \alpha_i]] = E[\check{\bar{g}}(\alpha_i) E[\varepsilon_{i2} - \varepsilon_{i1} | \alpha_i]] = 0. \tag{47}$$

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

□

B.7 Proof of Proposition 4.1

In this proof, all equalities involving random variables are understood to hold a.s. By Lemma B.2, 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:

$$\begin{aligned} & E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2}) \varepsilon_{i1} | \alpha_i = a, X_i = (x_1, x_2)] \\ &= \int \bar{g}(a, x_1, x_2, e_1, e_2) e_1 dF_{\varepsilon_{i1}, \varepsilon_{i2} | \alpha_i, X_i}(e_1, e_2 | a, x_1, x_2) \\ &= \int \bar{g}(a, x_1, x_2, e_2, e_1) e_1 dF_{\varepsilon_{i1}, \varepsilon_{i2} | \alpha_i, X_i}(e_2, e_1 | a, x_1, x_2) \\ &= E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2}) \varepsilon_{i2} | \alpha_i = a, X_i = (x_1, x_2)], \end{aligned} \tag{48}$$

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:

$$\begin{aligned}
&\int E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i1}|\alpha_i = a, X_{i1} = x_1, X_{i2} = x_2]dF_{\alpha_i|X_{i1},X_{i2}}(a|x_1, x_2) \\
=&\int E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})\varepsilon_{i2}|\alpha_i = a, X_{i1} = x_1, X_{i2} = x_2]dF_{\alpha_i|X_{i1},X_{i2}}(a|x_1, x_2). \qquad (49)
\end{aligned}$$

As a result, by the law of iterated expectations we have that $E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|X_i] = 0$. This completes the proof, since Assumptions SP-X and Assumption SC1-X.2 directly imply that $E[\varepsilon_{i2} - \varepsilon_{i1}|X_i] = 0$.

(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

$$\begin{aligned}
&E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|X_i] \\
=&E[E[\bar{g}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1}, \varepsilon_{i2})(\varepsilon_{i2} - \varepsilon_{i1})|X_i, \alpha_i, \varepsilon_{i1}]|X_i] \\
=&E[\check{\bar{g}}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1})E[\varepsilon_{i2} - \varepsilon_{i1}|X_i, \alpha_i, \varepsilon_{i1}]|X_i] \\
=&E[\check{\bar{g}}(\alpha_i, X_{i1}, X_{i2}, \varepsilon_{i1})E[\varepsilon_{i2} - \varepsilon_{i1}|X_i]|X_i] \\
=&E[G_i|X_i]E[\varepsilon_{i2} - \varepsilon_{i1}|X_i], \qquad (50)
\end{aligned}$$

where the first equality follows by the law of iterated expectations. 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) Now since $\bar{g}(\cdot)$ is a trivial function of ε_{i1} and ε_{i2} under Assumption SC3-X, we can define $\check{\bar{g}}(a, x_1, x_2) = \bar{g}(a, x_1, x_2, e_1, e_2)$.

$$\begin{aligned}
E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|X_i] &= E[E[G_i|X_i, \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}](\varepsilon_{i2} - \varepsilon_{i1})|X_i, \alpha_i]|X_i] \\
&= E[\check{\bar{g}}(\alpha_i, X_{i1}, X_{i2})E[\varepsilon_{i2} - \varepsilon_{i1}|X_i, \alpha_i]|X_i] = 0. \qquad (51)
\end{aligned}$$

The first equality follows by the law of iterated expectations. 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. □

B.8 Proof of Proposition 5.1

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

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

$$E[G_i(Y_{i2}(0)-Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] = E[G_i|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0)-Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \tag{52}$$

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

Note that by the law of iterated expectations, Assumption SEL-NSP and the definition of $\bar{g}(\cdot)$, the LHS of (52) equals the following

$$\begin{aligned}
&E[G_i(Y_{i2}(0)-Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]\\
=&E[E[G_i|\alpha_i^\mu, \alpha_i^\lambda, X_i^\mu, X_i^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda](Y_{i2}(0)-Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]\\
=&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(Y_{i2}(0)-Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu].
\end{aligned} \tag{53}$$

Similarly, by the law of iterated expectations, the RHS of (52) equals the following,

$$\begin{aligned}
&E[G_i|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0)-Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]\\
=&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0)-Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]
\end{aligned} \tag{54}$$

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, it follows that

$$
\begin{aligned}
&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(Y_{i2}(0) - Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
=&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&+ E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu],
\end{aligned} \tag{55}
$$

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

$$
\begin{aligned}
&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)\mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu)|X_i^\lambda = (x_1^\lambda, x_2^\lambda), X_{i1}^\mu = X_{i2}^\mu = x^\mu, \alpha_i^\mu = a] \\
=&\int \bar{g}(a, x^\mu, x^\mu, x_1^\lambda, x_2^\lambda, e_1, e_2)\mu(x^\mu, a, e_1) dF_{\varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu, \alpha_i^\mu}(e_1, e_2|(x_1^\lambda, x_2^\lambda), x^\mu, a) \\
=&\int \bar{g}(a, x^\mu, x^\mu, x_1^\lambda, x_2^\lambda, e_2, e_1)\mu(x^\mu, a, e_1) dF_{\varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu, \alpha_i^\mu}(e_2, e_1|(x_1^\lambda, x_2^\lambda), x^\mu, a) \\
=&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu)|X_i^\lambda = (x_1^\lambda, x_2^\lambda), X_{i1}^\mu = X_{i2}^\mu = x^\mu, \alpha_i^\mu = a].
\end{aligned} \tag{56}
$$

As a result, the first term in (55) equals zero by (56) and the law of iterated expectations.

Next, we consider the second summand in (55),

$$
\begin{aligned}
&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)(\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
=&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
=&E[\bar{g}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu, \varepsilon_{i2}^\mu)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0) - Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu].
\end{aligned} \tag{57}
$$

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 exchangeabilty restriction in Assumption SC1-NSP.2 by Lemma B.4, and implies that $E[\mu(X_{i2}^\mu, \alpha_i^\mu, \varepsilon_{i2}^\mu) - \mu(X_{i1}^\mu, \alpha_i^\mu, \varepsilon_{i1}^\mu)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu, \alpha_i^\mu] = 0$ as well as

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

by the law of iterated expectations. As a result, the above implies that Assumption PT-NSP holds.

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

By Assumption NSP, it follows that

$$
\begin{aligned}
&E[\check{\bar{g}}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu)(Y_{i2}(0) - Y_{i1}(0))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
=&E[\check{\bar{g}}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu)\Delta_{\mu,i}|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&+ E[\check{\bar{g}}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu)(\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda))|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
=&E[\check{\bar{g}}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[\Delta_{\mu,i}|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
&+ E[\check{\bar{g}}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[\lambda_2(X_{i2}^\lambda, \alpha_i^\lambda, \varepsilon_{i2}^\lambda) - \lambda_1(X_{i1}^\lambda, \alpha_i^\lambda, \varepsilon_{i1}^\lambda)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \\
=&E[\check{\bar{g}}(\alpha_i^\mu, X_{i1}^\mu, X_{i2}^\mu, X_{i1}^\lambda, X_{i2}^\lambda, \varepsilon_{i1}^\mu)X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu]E[Y_{i2}(0) - Y_{i1}(0)|X_i^\lambda, X_{i1}^\mu = X_{i2}^\mu] \qquad (58)
\end{aligned}
$$

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, which implies Assumption PT-NSP.

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

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

where the first equality follows from Assumption NSP. The second equality follows by applying the law of iterated expectations 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 law of iterated expectations. This completes the proof. $\square$

B.9 Proof of Proposition 5.2

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

(i) Here 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 RE.

(i.a) Consider

$$F_{\varepsilon_{i1}^\mu,G_i|\alpha_i^\mu}(e_1,g|a)=F_{G_i|\varepsilon_{i1}^\mu,\alpha_i^\mu}(g|e_1,a)F_{\varepsilon_{i1}^\mu|\alpha_i^\mu}(e_1|a) \tag{59}$$

Assumption SC1-NSP.2 implies $F_{\varepsilon_{i1}^\mu|\alpha_i^\mu}(e|a)=F_{\varepsilon_{i2}^\mu|\alpha_i^\mu}(e|a)$ as well as $F_{\varepsilon_{i1}^\mu|\varepsilon_{i2}^\mu,\alpha_i^\mu}(e_1|e_2,a)=F_{\varepsilon_{i2}^\mu|\varepsilon_{i1}^\mu,\alpha_i^\mu}(e_1|e_2,a)$ by Lemma B.4, which implies

$$\begin{aligned}F_{G_i|\varepsilon_{i1}^\mu,\alpha_i^\mu}(g|e_1,a)&=\int 1\{g(a,e_1,e_2)\le g\}dF_{\varepsilon_{i2}^\mu|\varepsilon_{i1}^\mu,\alpha_i^\mu}(e_2|e_1,a)\\ =\int 1\{g(a,e_2,e_1)\le g\}dF_{\varepsilon_{i1}^\mu|\varepsilon_{i2}^\mu,\alpha_i^\mu}(e_2|e_1,a)&=F_{G_i|\varepsilon_{i2}^\mu,\alpha_i^\mu}(g|e_1,a).\end{aligned} \tag{60}$$

As a result,

$$\begin{aligned}F_{\varepsilon_{i1}^\mu,G_i|\alpha_i^\mu}(e_1,g|a)&=F_{G_i|\varepsilon_{i1}^\mu,\alpha_i^\mu}(g|e_1,a)F_{\varepsilon_{i1}^\mu|\alpha_i^\mu}(e_1|a)=F_{G_i|\varepsilon_{i2}^\mu,\alpha_i^\mu}(g|e_1,a)F_{\varepsilon_{i2}^\mu|\alpha_i^\mu}(e_1|a)\\ =&F_{\varepsilon_{i2}^\mu,G_i|\alpha_i^\mu}(e_1,g|a).\end{aligned} \tag{61}$$

This implies Assumption TH by the definition of a conditional distribution $F_{\varepsilon_{it}^\mu|G_i,\alpha_i^\mu}(e|g,a)=\frac{F_{\varepsilon_{it}^\mu,G_i|\alpha_i^\mu}(e,g|a)}{F_{G_i|\alpha_i^\mu}}$, where $F_{G_i|\alpha_i^\mu}(g|a)>0$ by assumption.

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

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

$$F_{\varepsilon_{it}^\mu|G_i,\alpha_i^\mu}(e|g,a) = \sum_{g=0,1} P(\varepsilon_{it}^\mu \leq e|G_i = g(a), \alpha_i^\mu = a)1\{g(a) = g\}$$

$$= \sum_{g=0,1} P(\varepsilon_{it}^\mu \leq e|\alpha_i^\mu = a)1\{g(a) = g\} = \sum_{g=0,1} F_{\varepsilon_{it}^\mu|\alpha_i^\mu}(e|a)1\{g(a) = g\}. \tag{62}$$

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

(ii.b) The statement (ii.b) is immediate from noting that $G_i = g(\alpha_i^\mu)$ (Assumption SC3-NSP.1) and the independence condition in Assumption SC3-NSP.3 imply that $g(\alpha_i^\mu) \perp (\alpha_i^\lambda, \varepsilon_{i1}^\lambda, \varepsilon_{i2}^\lambda)$, which is equivalent to Assumption RE. This completes the proof of (i).

□

B.10 Supplementary lemmas

Lemma B.1. *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

$$E[\tilde{W}_i^+] = E[\tilde{W}_i^-], \tag{63}$$

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

$$E[\tilde{W}_i^+ 1\{\tilde{W}_i \geq 0\}] = E[\tilde{W}_i^- 1\{\tilde{W}_i \geq 0\}]. \tag{64}$$

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 $E[\tilde{W}_i] = 0$, (63) 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. □

Lemma B.2. *Suppose SP-X and $P(G_i = 1|X_i) \in (0, 1)$ a.s. hold. Assumption PT-X holds if and only if*

$$E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|X_i] = E[G_i|X_i]E[\varepsilon_{i2} - \varepsilon_{i1}|X_i] \ a.s. \tag{65}$$

Proof. Since $P(G_i = 1|X_i) \in (0, 1)$ a.s., Assumption PT-X holds iff

$$E[G_i(Y_{i2}(0) - Y_{i1}(0))|X_i] = E[G_i|X_i]E[Y_{i2}(0) - Y_{i1}(0)|X_i] \ a.s. \tag{66}$$

by arguments similar to Lemma 3.1 while conditioning on X_i. By Assumption SP-X, the left-hand side of the above simplifies to

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

As a result, Assumption PT-X holds iff

$$E[G_i(\varepsilon_{i2} - \varepsilon_{i1})|X_i] = E[G_i|X_i]E[\varepsilon_{i2} - \varepsilon_{i1}|X_i] \ a.s. \tag{67}$$

□

Lemma B.3. *Suppose $P(G_i = 1) \in (0, 1)$ holds. Then, Assumption PT holds iff $E[G_i(Y_{i2}(0) - Y_{i1}(0))] = E[G_i]E[Y_{i2}(0) - Y_{i1}(0)]$.*

Proof. Assumption PT can be written as

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1] = E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0]. \tag{68}$$

As a result, Assumption PT holds iff $E[Y_{i2}(0)-Y_{i1}(0)|G_i] = E[Y_{i2}(0)-Y_{i1}(0)]$. It remains to show that (a) $E[Y_{i2}(0)-Y_{i1}(0)|G_i] = E[Y_{i2}(0)-Y_{i1}(0)]$ holds if and only if (b) $E[G_i(Y_{i2}(0)-Y_{i1}(0))] = E[G_i]E[Y_{i2}(0) - Y_{i1}(0)]$. The result that (a) $\Rightarrow$ (b) is immediate. As for (b) $\Rightarrow$ (a), $E[G_i(Y_{i2}(0) - Y_{i1}(0))] = E[G_i]E[Y_{i2}(0) - Y_{i1}(0)]$ implies $E[Y_{i2}(0) - Y_{i1}(0)|G_i = 1] = E[Y_{i2}(0)-Y_{i1}(0)]$ by dividing the former equality by $E[G_i] = P(G_i = 1) \in (0, 1)$. It remains to show that $E[G_i(Y_{i2}(0) - Y_{i1}(0))] = E[G_i]E[Y_{i2}(0) - Y_{i1}(0)]$ implies $E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0] = E[Y_{i2}(0) - Y_{i1}(0)]$. To do so, we subtract $E[Y_{i2}(0) - Y_{i1}(0)]$ from both sides of the previous equality and multiply by -1.

$$E[(1 - G_i)(Y_{i2}(0) - Y_{i1}(0))] = E[1 - G_i]E[Y_{i2}(0) - Y_{i1}(0)] \tag{69}$$

Since $E[1-G_i] = P(G_i = 0) \in (0,1)$, the above implies that $E[Y_{i2}(0) - Y_{i1}(0)|G_i = 0] = E[Y_{i2}(0) - Y_{i1}(0)]$. This completes the proof. $\square$

Lemma B.4. *Suppose that Assumption SC1-NSP.2 holds, then*

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

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

Proof. **(i)** By the definition of the marginal distribution, Assumption SC1-NSP implies (i) by the following, *a.e.*

$$F_{\varepsilon_{i1}^{\mu}|\alpha_i^{\mu}}(e_1|a) = \lim_{e_2\to\infty} F_{\varepsilon_{i1}^{\mu},\varepsilon_{i2}^{\mu}|\alpha_i^{\mu}}(e_1,e_2|a) = \lim_{e_2\to\infty} F_{\varepsilon_{i1}^{\mu},\varepsilon_{i2}^{\mu}|\alpha_i^{\mu}}(e_2,e_1|a) = F_{\varepsilon_{i2}^{\mu}|\alpha_i^{\mu}}(e_1|a). \tag{70}$$

(ii) By the definition of the conditional distribution and (i) of this lemma, Assumption SC1-NSP implies (ii) by the following

$$F_{\varepsilon_{i1}^{\mu}|\varepsilon_{i2}^{\mu},\alpha_i^{\mu}}(e_1|e_2,a) = \frac{F_{\varepsilon_{i1}^{\mu},\varepsilon_{i2}^{\mu}|\alpha_i^{\mu}}(e_1,e_2|a)}{F_{\varepsilon_{i2}^{\mu}|\alpha_i^{\mu}}(e_2|a)} = \frac{F_{\varepsilon_{i1}^{\mu},\varepsilon_{i2}^{\mu}|\alpha_i^{\mu}}(e_2,e_1|a)}{F_{\varepsilon_{i1}^{\mu}|\alpha_i^{\mu}}(e_2|a)} = F_{\varepsilon_{i2}^{\mu}|\varepsilon_{i1}^{\mu},\alpha_i^{\mu}}(e_1|e_2,a), \tag{71}$$

a.e. $(a,e_1,e_2) \in \mathcal{A} \times \mathcal{E} \times \mathcal{E}_a$, where $\mathcal{E}_a$ is the support of $\varepsilon_{it}^{\mu}|\alpha_i^{\mu}$ for $t = 1,2$. $\square$

Lemma B.5. *Suppose Assumptions NSP and* $P(G_i = 1|X_i) \in (0,1)$ *a.s. hold. Then, Assumption PT-X holds iff*

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

Proof. We first note that Assumption PT-X holds iff

$$E[Y_{i2}(0) - Y_{i1}(0)|G_i, X_i] = E[Y_{i2}(0) - Y_{i1}(0)|X_i] \ a.s. \tag{72}$$

It remains to show that (a) $E[Y_{i2}(0) - Y_{i1}(0)|G_i, X_i] = E[Y_{i2}(0) - Y_{i1}(0)|X_i]$ a.s. iff (b) $E[G_i(Y_{i2}(0) - Y_{i1}(0))|X_i] = E[G_i|X_i]E[Y_{i2}(0) - Y_{i1}(0)|X_i]$ a.s. The result that (a) $\Rightarrow$ (b) follows immediately by the law of iterated expectations. As for (b) $\Rightarrow$ (a),

$$E[G_i(Y_{i2}(0) - Y_{i1}(0))|X_i] = E[G_i|X_i]E[Y_{i2}(0) - Y_{i1}(0)|X_i] \ a.s. \tag{73}$$

implies that

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

since $P(G_i = 1|X_i) \in (0,1)$ a.s. Now subtracting $E[Y_{i2}(0) - Y_{i1}(0)|X_i]$ from both sides of (73) and multiplying by -1 yields

$$E[(1 - G_i)(Y_{i2}(0) - Y_{i1}(0))|X_i] = E[(1 - G_i)|X_i]E[Y_{i2}(0) - Y_{i1}(0)|X_i] \ a.s. \tag{74}$$

Since $P(G_i = 0|X_i) \in (0,1)$ a.s., the above implies

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

This completes the proof. $\square$