When Is Parallel Trends Sensitive to Functional Form?*

Jonathan Roth†   Pedro H.C. Sant'Anna‡

October 21, 2021

Abstract

This paper assesses when the validity of difference-in-differences and related estimators depends on functional form. We provide a novel characterization: the parallel trends assumption holds under all strictly monotonic transformations of the outcome if and only if a stronger "parallel trends"-type condition holds for the cumulative distribution function of untreated potential outcomes. This condition for parallel trends to be insensitive to functional form is satisfied if and essentially only if the population can be partitioned into a subgroup for which treatment is effectively randomly assigned and a remaining subgroup for which the distribution of untreated potential outcomes is stable over time. We introduce falsification tests for the insensitivity of parallel trends to functional form. We also show that it is impossible to construct any estimator that is consistent for the average treatment effect on the treated (ATT) without either imposing functional form restrictions or imposing assumptions that identify the full distribution of untreated potential outcomes. Our results suggest that researchers who wish to point-identify the ATT should either (i) argue treatment is as-if randomly assigned, (ii) provide a method for inferring the full counterfactual distribution for the treated group, or (iii) justify the validity of the specific chosen functional form.

†Brown University. Email: jonathanroth@brown.edu
‡Vanderbilt University and Microsoft. Email: pedro.h.santanna@vanderbilt.edu

1 Introduction

Difference-in-differences (DiD) is one of the most popular strategies in the social sciences for estimating causal effects in non-experimental contexts. The DiD design allows for identification of the average treatment effect on the treated (ATT) under the parallel trends assumption, which states that the mean outcome for the treated and comparison populations would have evolved in parallel if the treatment had not occurred. It is well-known that the parallel trends assumption is weaker than (as-if) random assignment of treatment. This paper studies the content of the parallel trends assumption in settings where treatment may not be (as-if) randomly assigned.

We focus on the extent to which the assumptions underlying DiD and other estimators of the ATT depend on the functional form of the outcome. Following Athey and Imbens (2006), we say that an assumption is insensitive to functional form if it is invariant to strictly monotonic transformations of the outcome — i.e., if the assumption holds for potential outcomes $Y(\bullet)$, then it also holds if the potential outcomes are replaced with any strictly monotonic function of the original potential outcomes $g(Y(\bullet))$.[1] Intuitively, this invariance requires that the validity of the assumption does not depend on the units in which the outcome is measured.

Our goal in studying this property is to clarify the different ways that a researcher can justify the validity of a DiD design. One way to justify the validity of a DiD design is to verify conditions that ensure that parallel trends holds for all functional forms of the outcome. Alternatively, if the parallel trends assumption may be sensitive to functional form, then the researcher can provide context-specific knowledge or an economic model that justifies the parallel trends assumption specifically for the chosen functional form. Our results make precise the conditions needed to justify parallel trends for all functional forms, and suggest that the researcher should be careful to give a functional-form specific justification for the parallel trends assumption in settings where these conditions are implausible.

Understanding when parallel trends is sensitive to functional form is particularly important in settings where it is not clear from economic theory that parallel trends should hold particularly for the chosen functional form. The choice of functional form for a DiD analysis is often motivated by which ATT is most policy-relevant, but it may not be obvious that parallel trends will hold specifically for that transformation. For example, depending on the exact question they have in mind, a researcher studying a labor market intervention may be most interested in the policy's effect on earnings measured in levels, logs, or percentiles

---

[1] Athey and Imbens (2006, FN 11) use the phrase "invariance to scale" to describe this property; we use "invariance to transformations" instead to make clear that the transformations may be non-linear.

relative to the national distribution. However, it will not always be clear that the policy variation studied will generate parallel trends in the functional form for which the ATT is most relevant – the researcher may be most interested in the ATT in levels even though parallel trends only holds in logs. We thus may be skeptical of the parallel trends assumption if its validity depends on the exact choice of functional form.[2] Such concerns about sensitivity to functional form in fact have a long history in econometrics — see, e.g., Leamer's (1983; 1985) influential critiques of applied work in the 1980s.

This paper's first main contribution is to characterize when the parallel trends assumption is insensitive to functional form. We prove that the parallel trends assumption is insensitive to functional form if and only if a "parallel trends"-type condition holds for the entire cumulative distribution function (CDF) of untreated potential outcomes. We further show that this parallel trends of CDFs condition holds if and only if the distribution of untreated potential outcomes for each group and time period can be represented as a mixture of a common time-varying distribution that does not depend on group (with weight θ) and a group-specific distribution that does not depend on time (with weight $1-\theta$). There are thus three cases in which parallel trends is insensitive to functional form: first, if the distribution of untreated potential outcomes is the same for both groups, as occurs under random assignment of treatment. Second, if the untreated potential outcome distributions for each group are stable over time. And third, a hybrid of the first two cases in which the population is effectively a mixture of a sub-population that is (as-if) randomized between treatment and control and another sub-population that has stable untreated potential outcome distributions over time. In settings where the treatment is not (as-if) randomly assigned, the assumptions needed for the insensitivity of parallel trends to functional form will thus often be restrictive.

The second contribution of this paper is to introduce tests that can reject the null hypothesis that parallel trends is insensitive to functional form. Such tests can be useful in flagging situations where the researcher should be particularly careful about justifying the parallel trends assumption for the functional form chosen for the analysis. We illustrate how such tests can be used in a stylized analysis of the effects of the minimum wage.

The third contribution of this paper is to analyze the sensitivity to functional form of other estimators beyond DiD. We prove that the ATT is identified under all strictly monotonic transformations of the outcome if and only if the entire distribution of untreated potential outcomes for the treated group is identified. Thus, to obtain *any* consistent estimator of the ATT, one must impose assumptions that either are sensitive to functional form *or* that

---

[2]These concerns may be particularly severe in settings where the researcher wishes to estimate the ATT for multiple transformations. Sensitivity to functional form may also be of concern in settings where the outcome has no natural units, such as test scores (Cunha and Heckman, 2008).

identify the full counterfactual distribution of potential outcomes.

Our results have important implications for practitioners who wish to estimate the ATT using difference-in-differences or related research designs. In light of our results, researchers interested in point-identifying the ATT should make one of the following three justifications. First, they may argue that treatment is (as-if) randomly assigned. In this case, parallel trends will hold under all transformations of the outcome. Second, the researcher may justify an assumption that pins down the entire counterfactual distribution of potential outcomes for the treated group, and then choose an appropriate estimator that is valid without any additional functional form restrictions under this assumption. To justify the difference-in-differences estimator without further functional form assumptions, for instance, the researcher should justify parallel trends of CDFs. Third, the researcher may give up on insensitivity to functional form, and argue for the validity of the particular chosen functional form. If none of these justifications is appealing, the researcher may instead impose weaker assumptions on the data-generating process that do not point-identify the ATT, e.g. using partial identification tools that do not impose that the parallel trends assumption holds exactly (Manski and Pepper, 2018; Rambachan and Roth, 2020).

Several previous papers have noted that the parallel trends assumption may be sensitive to functional form, with particular attention paid to the logs versus levels specifications (e.g. Meyer, 1995; Athey and Imbens, 2006). To our knowledge, however, we are the first to provide a complete characterization of when parallel trends is sensitive to functional form. In particular, the previous literature does not appear to have realized that parallel trends can be insensitive to functional form even if baseline distributions differ between the treated and comparison groups and outcome distributions vary over time.[3] We also provide novel tests of the conditions for parallel trends to be insensitive to functional form.

Our work relates to several papers that consider identification of quantile treatment effects in difference-in-differences settings (Athey and Imbens, 2006; Bonhomme and Sauder, 2011; Callaway and Li, 2019). These papers introduce new sets of assumptions that differ from the usual parallel trends assumption and allow for identification of the full distribution of untreated potential outcomes. One feature of the set of assumptions introduced in Athey and Imbens (2006) is that it is invariant to transformations of the outcome, unlike the usual parallel trends assumption. By contrast, we derive conditions on the distributions of potential outcomes under which the usual parallel trends assumption is invariant to transformations. These conditions are different from, and non-nested with, the conditions provided for identification of the full distribution of potential outcomes in previous work.

---

[3]For example, Kahn-Lang and Lang (2020) write that parallel trends cannot hold in both levels and logs "unless the distribution of outcomes is initially the same for the experimental and control groups."

Our results also draw connections between approaches to causal identification that rely on an unconfoundedness assumption (e.g. Imbens and Rubin (2015)) and difference-in-differences designs, which rely on the parallel trends assumption. We show, roughly speaking, that parallel trends is either a functional form restriction or a combination of unconfoundedness and stationarity assumptions. Our paper clarifies when differencing in both the time and group dimensions allows for identification of the ATT without further functional form assumptions, and thus also relates to Rosenbaum (1987)'s discussion of the advantages and limitations of using a second control group in observational studies, albeit in a somewhat different context.

We are not aware of any previous papers showing the simple but important fact that identification of the ATT under all monotonic transformations is equivalent to identification of the distribution of counterfactual outcomes for the treated group. A powerful implication of this result is that the consistency of any estimator for the ATT must either depend on functional form assumptions or assumptions that identify the full distribution of untreated potential outcomes. This result does not depend on the multi-period structure of DiD, and may be relevant in other contexts where ATTs are of interest.

2   Model

We consider a canonical two-period difference-in-differences model. There are two periods $t = 0, 1$, and units indexed by i come from one of two populations denoted by $D_i \in \{0, 1\}$. Units in the $D_i = 1$ (treated) population receive treatment beginning in period $t = 1$, and units in the $D_i = 0$ (comparison) population never receive treatment. We denote by $Y_{it}(1), Y_{it}(0)$ the potential outcomes for unit i in period t under treatment and control, respectively, and we observe the outcome $Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0)$. We assume that there are no anticipatory effects of treatment, so that $Y_{i0}(1) = Y_{i0}(0)$ for all i. The average treatment effect on the treated is defined as

$$\tau_{ATT} = \mathbb{E}\left[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1\right].$$

Remark 1 (Multiple periods and staggered timing). We consider a two period model for expositional simplicity. Several recent papers have considered settings with multiple periods and staggered treatment timing under a generalized parallel trends assumption that imposes the two-period, two-group version of parallel trends for multiple pairs of groups and periods (e.g., Assumptions 4 and 5 in Callaway and Sant'Anna (2020), Assumption 1 in Sun and Abraham (2020), or Assumption 5 in de Chaisemartin and D'Haultfœuille (2020)). Our results on the parallel trends assumption in this simple model thus have immediate implica-

tions for the building blocks of the generalized parallel trends assumption in the staggered case.

Remark 2 (Conditional parallel trends)**.** Likewise, for simplicity we consider a model that does not condition on unit-specific covariates. However, the same results would go through if all probability statements were conditional on some value of unit-specific covariates X_i. Our results thus have implications for the conditional parallel trends assumptions considered in, for example, Abadie (2005) and Sant'Anna and Zhao (2020).

3   Invariance of Parallel Trends

The classical assumption that allows for point identification of the ATT in the DiD design is the parallel trends assumption, which imposes that

$$\mathbb{E}\left[Y_{i1}(0) \,|\, D_i = 1\right] - \mathbb{E}\left[Y_{i0}(0) \,|\, D_i = 1\right] = \mathbb{E}\left[Y_{i1}(0) \,|\, D_i = 0\right] - \mathbb{E}\left[Y_{i0}(0) \,|\, D_i = 0\right]. \quad (1)$$

Under the parallel trends assumption, the ATT is identified: $\tau_{ATT} = (\mu_{11} - \mu_{10}) - (\mu_{01} - \mu_{00})$, where $\mu_{dt} = \mathbb{E}\left[Y_{it} \,|\, D_i = d\right]$. We assume throughout that the four expectations in (1) exist and are finite. Following Athey and Imbens (2006), we say that the parallel trends assumption is invariant to transformations if the parallel trends assumption holds for all strictly monotonic transformations of the outcome.

Definition 1. We say that the parallel trends assumption is invariant to transformations (a.k.a. insensitive to functional form) if

$$\mathbb{E}\left[g(Y_{i1}(0)) \,|\, D_i = 1\right] - \mathbb{E}\left[g(Y_{i0}(0)) \,|\, D_i = 1\right] = \mathbb{E}\left[g(Y_{i1}(0)) \,|\, D_i = 0\right] - \mathbb{E}\left[g(Y_{i0}(0)) \,|\, D_i = 0\right]$$

for all strictly monotonic functions g such that the expectations above are finite.[4]

Our first result characterizes when parallel trends is invariant to transformations.

Proposition 3.1. *Parallel trends is invariant to transformations if and only if*

$$F_{Y_{i1}(0)|D_i=1}(y) - F_{Y_{i0}(0)|D_i=1}(y) = F_{Y_{i1}(0)|D_i=0}(y) - F_{Y_{i0}(0)|D_i=0}(y), \text{ for all } y \in \mathbb{R} \quad (2)$$

where $F_{Y_{it}(0)|D_i=d}(y)$ is the cumulative distribution function of $Y_{it}(0) \,|\, D_i = d$.

Proof. If (2) holds, then from integrating on both sides of the equation it is immediate that

$$\int g(y)dF_{Y_{i1}(0)|D_i=1} - \int g(y)dF_{Y_{i0}(0)|D_i=1} = \int g(y)dF_{Y_{i1}(0)|D_i=0} - \int g(y)dF_{Y_{i0}(0)|D_i=0} \quad (3)$$

---

[4]Note that all strictly monotonic functions are Lebesgue-measurable, since their contour sets are intervals.

for any strictly monotonic g such that the integrals exist and are finite, and hence parallel trends is invariant to transformations.

Conversely, if parallel trends is invariant to transformations, then (3) holds for every strictly monotonic g such that the expectations exist and are finite. In particular, it holds for the identity map $g_1(y) = y$ as well as the map $g_2(y) = y - 1[y \leqslant \tilde{y}]$ for any given $\tilde{y} \in \mathbb{R}$. Then, it follows that

$$\int y dF_{Y_{i1}(0)|D_i=1} - \int y dF_{Y_{i0}(0)|D_i=1} = \int y dF_{Y_{i1}(0)|D_i=0} - \int y dF_{Y_{i0}(0)|D_i=0}, \text{ and}$$

$$\int (y - 1[y \leqslant \tilde{y}]) dF_{Y_{i1}(0)|D_i=1} - \int (y - 1[y \leqslant \tilde{y}]) dF_{Y_{i0}(0)|D_i=1} =$$

$$\int (y - 1[y \leqslant \tilde{y}]) dF_{Y_{i1}(0)|D_i=0} - \int (y - 1[y \leqslant \tilde{y}]) dF_{Y_{i0}(0)|D_i=0}.$$

Subtracting the second equation from the first in the previous display, we obtain

$$\int 1[y \leqslant \tilde{y}] dF_{Y_{i1}(0)|D_i=1} - \int 1[y \leqslant \tilde{y}] dF_{Y_{i0}(0)|D_i=1} = \int 1[y \leqslant \tilde{y}] dF_{Y_{i1}(0)|D_i=0} - \int 1[y \leqslant \tilde{y}] dF_{Y_{i0}(0)|D_i=0},$$

which is equivalent to (2) by the definition of the CDF and the fact that $\tilde{y}$ is arbitrary. □

Proposition 3.1 shows that parallel trends is invariant to transformations if and only if a "parallel trends"-type assumption holds for the CDFs of the untreated potential outcomes. We note that if the outcome is continuous, then parallel trends of CDFs is equivalent to parallel trends of PDFs (almost everywhere). The following result provides a characterization of how distributions satisfying this assumption can be generated.

Proposition 3.2. *Suppose that the distributions $Y_{it}(0)|D_i = d$ for all $d, t \in \{0, 1\}$ have a Radon-Nikodym density with respect to a common dominating, positive σ-finite measure.*[5] *Then parallel trends is invariant to transformations if and only if there exists $\theta \in [0, 1]$ and CDFs $G_t(\cdot)$ and $H_d(\cdot)$ depending only on time and group, respectively, such that*

$$F_{Y_{it}(0)|D_i=d}(y) = \theta G_t(y) + (1 - \theta) H_d(y) \text{ for all } y \in \mathbb{R} \text{ and } d, t \in \{0, 1\}. \tag{4}$$

Proof. See Appendix A. □

Proposition 3.2 shows that parallel trends of CDFs is satisfied if and only if the untreated potential outcomes for each group and time can be represented as a mixture of a common time-varying distribution that does not depend on group (with weight θ) and a group-specific

---

[5]The Radon-Nikodym density exists if $Y_{it}(0)|D = d$ is continuously distributed (using the probability density function and Lebesgue measure), or discrete with finite support (using the probability mass function and counting measure). It will also exist for many non-pathological mixed distributions.

distribution that does not depend on time (with weight $1 - \theta$). This implies that there are three cases in which parallel trends will be insensitive to functional form, depending on the value of θ.

Case 1: Random assignment. ($\boldsymbol{\theta = 1}$). The case $\theta = 1$ corresponds with imposing that the distributions of $Y(0)$ for the treated and comparison groups are the same in each period, $F_{Y_{it}(0)|D_i=1}(y) = F_{Y_{it}(0)|D_i=0}(y)$, as occurs under (as-if) random assignment of treatment.

Case 2: Stationary $\boldsymbol{Y(0)}$. ($\boldsymbol{\theta = 0}$). The case $\theta = 0$ corresponds with imposing that the distribution of $Y(0)$ for both the treated and comparison populations does not depend on time, i.e. $F_{Y_{i1}(0)|D_i=d}(y) = F_{Y_{i0}(0)|D_i=d}(y)$.

Case 3: Non-random assignment and non-stationarity. ($\boldsymbol{\theta \in (0,1)}$). The case $\theta \in (0,1)$ corresponds with a hybrid of Cases 1 and 2. In each period, we can partition the treated and comparison groups so that θ fraction of each group have the same distribution (G_t), as if they were randomly assigned, and $1 - \theta$ fraction have a group-specific distribution that does not depend on time (H_d). This is perhaps the most interesting case, since in the other two cases a single difference (either across time or across groups) would suffice to identify the ATT.

Remark 3 (Use of phrase "essentially only if")**.** The simplest way for Case 3 to hold is to have a θ fraction of the population that is as-if randomized between treatment and control and a $1 - \theta$ fraction that has stationary potential outcomes. In principle, though, it is possible for the units in the θ and $1 - \theta$ partitions to change across periods in Case 3, although it is difficult to imagine scenarios where this would be the case in practice. We thus write in the abstract that parallel trends can be invariant to transformations "essentially only if" the population can be partitioned into groups such that one is effectively randomized between treatment and control, and the others have stable potential outcomes over time.

Example 1 (Binary outcomes)**.** Suppose the outcome is binary, $Y_i \in \{0,1\}$. Then for any $y \in [0,1)$, $F_{Y_{i1}(0)|D_i=1}(y) = 1 - \mathbb{E}\left[Y_{i1}(0) \mid D_i = 1\right]$, and analogously for the other CDFs. Thus, (2) is equivalent to the parallel trends assumption (1). Proposition 3.1 thus implies that whenever the parallel trends assumption holds, it also holds for all monotonic transformations of the outcome (i.e. replacing $\{0,1\}$ with $\{a,b\}$). This is intuitive, as the expectation of a binary outcome fully characterizes its distribution. Note that this does not imply that the parallel trends assumption necessarily holds for binary outcomes, only that it does not depend on the transformation of the outcome.

Example 2 (Normally distributed outcomes). Equation (2) can be highly restrictive for non-binary outcomes. For example, if the distribution of $Y_{it}(0)|D_i = d$ is normally distributed with positive variance for all (d, t), then (2) can hold only if either both groups have the same distribution of $Y(0)$ in each period or the group-specific distributions of $Y(0)$ do not change over time (Cases 1 and 2).[6]

Example 3 (Mixtures of distributions). Case 3 will hold if treatment is as-if randomly assigned among a subset of the population for which there are time trends in $Y(0)$ (e.g., younger workers on an upward earnings trajectory), but there are also units in the population for whom treatment status is endogenous but whose untreated potential outcomes are stable over time (e.g. older workers with stable earnings trajectories).[7] As an illustration, suppose $\theta = \frac{1}{3}$ fraction of workers are young. Young workers are effectively randomly assigned to locations, and have untreated earnings in period t corresponding with the $lognormal(t, 1)$ distribution (whose CDF we will denote G_t) regardless of their location. The $1 - \theta$ fraction of older workers endogenously sort to locations, but have stable earnings over time. In each period, older workers in locations with treatment d have untreated earnings distributed $lognormal(2 + d, 2)$ (whose CDF we will denote H_d). By construction, the distribution of untreated earnings satisfies the decomposition (4), and thus parallel trends will hold for all functional forms, despite the fact that the treatment and comparison groups have different baseline means and their distributions change over time.

3.1 Relationship to Prior Work

Remark 4 (Empirical papers using parallel trends of CDFs). Several empirical papers have used a DiD design to estimate the effects of a treatment on the entire distribution of an outcome. For example, Stepner (2019) uses DiD to estimate the impact of layoffs on the density of income at each point in the (discretized) income distribution. This analysis is valid if parallel trends holds throughout the income distribution — precisely the condition that Proposition 3.1 shows is needed for the usual parallel trends assumption to be insensitive to functional form. Other recent papers (e.g. Almond, Hoynes and Schanzenbach (2011), Cengiz, Dube, Lindner and Zipperer (2019)) have conducted related distributional analyses.

Remark 5 (Relationship to distributional DiD Models). Condition (2) implies that the full counterfactual distribution for the treated group, $Y_{i1}(0)|D_i = 1$, is identified. In particular,

---

[6]Appendix A.2 in the working paper version of this paper provides a proof.

[7]Of course, if data on worker age were available, then the researcher could restrict attention to younger workers, in which case we would be back in Case 1.

by re-arranging terms in (2), we obtain that

$$F_{Y_{i1}(0)|D_i=1}(y) = F_{Y_{i0}(0)|D_i=1}(y) + F_{Y_{i1}(0)|D_i=0}(y) - F_{Y_{i0}(0)|D_i=0}(y), \tag{5}$$

where the terms on the right-hand side correspond with CDFs of identified distributions. Condition (2) may thus be reminiscent of distributional DiD models such as Athey and Imbens (2006)'s Changes-in-Changes (CiC) model, which infers the counterfactual distribution by assuming that the mapping between quantiles of $Y(0)$ for the treated and comparison populations remains stable over time, i.e.,

$$F_{Y_{i1}(0)|D_i=1}(y) = F_{Y_{i0}(0)|D_i=1}(F^{-1}_{Y_{i0}(0)|D_i=0}(F_{Y_{i1}(0)|D_i=0})(y))). \tag{6}$$

The two ways of inferring the counterfactual distribution agree in Cases 1 and 2 above, but are generally non-nested otherwise. For instance, the CiC model will not generally hold in Case 3 given above (except for special choices of the distributions), since the mapping between quantiles for the treated and untreated groups need not be preserved across periods.[8] Conversely, one can construct examples where the CiC model holds when $Y(0)$ is normally distributed conditional on group and time period (with distinct means by group/period), whereas parallel trends of CDFs necessarily fails in this case as discussed in Example 2.[9] It is also straightforward to show that condition (5) is non-nested with the distributional DiD models of Bonhomme and Sauder (2011) and Callaway and Li (2019).

3.2 Testable Implications of Invariance to Transformations

We now show that condition (2) has testable implications, and thus can be rejected by the data. Recall that we can re-arrange the terms in (2) to obtain that

$$F_{Y_{i1}(0)|D_i=1}(y) = F_{Y_{i0}(0)|D_i=1}(y) + F_{Y_{i1}(0)|D_i=0}(y) - F_{Y_{i0}(0)|D_i=0}(y). \tag{7}$$

This shows that under condition (2), the CDF of the counterfactual distribution for the treated group at $t = 1$ is a linear combination of identified CDFs. The left-hand side of (7) is a CDF and thus must be weakly increasing in y, but this is not guaranteed of the right-hand side. We can thus falsify condition (7) if we reject the null that the right-hand side of (7) is weakly increasing in y. This is in fact a *sharp* testable implication, since the right-hand side of (7) is guaranteed to be right-continuous and to have limits of 0 and 1 as $y \to \pm\infty$ from

---

[8]As a concrete counterexample, suppose $\theta = 0.5$, G_t is the CDF for the uniform distribution on $[0, 1]$ for $t = 0$ and the uniform distribution on $[1, 2]$ for $t = 1$, whereas H_d corresponds with the CDF of a point mass at 1 and 0 for $d = 0$ and $d = 1$, respectively. Then $F_{Y_{i1}(0)|D_i=1}(1.2) = 0.6$, whereas equation (6) implies it is 0.1.

[9]For example, the CiC model holds if $Y_{it}(0)|D_i = d \sim \mathcal{N}(2t + d, 1)$.

the properties of the CDFs in the linear combination.

Statistical Testing. We now describe how one can conduct such tests in practice. For simplicity, we will focus on testing in the case where y has finite support, in which case the null is equivalent to testing that the implied distribution has non-negative mass at all support points. That is, we are interested in testing the null that

$$f_{Y_{i1}(0)|D_i=1}(y) = f_{Y_{i0}(0)|D_i=1}(y) + f_{Y_{i1}(0)|D_i=0}(y) - f_{Y_{i0}(0)|D_i=0}(y) \geqslant 0 \text{ for all } y \in \mathbb{R},$$

where $f_{Y_{it}(0)|D_i=d}(y) = \mathbb{E}\left[1[Y_{it}(0) = y] \mid D_i = d\right]$ is the mass at y of $Y_{it}(0)|D_i = 1$. To test this null, we can form estimates

$$\hat{f}_{Y_{i1}(0)|D_i=1}(y) = \hat{f}_{Y_{i0}|D_i=1}(y) + \hat{f}_{Y_{i1}|D_i=0}(y) - \hat{f}_{Y_{i0}|D_i=0}(y),$$

where the $\hat{f}$ terms on the right-hand side are sample frequencies, e.g., $\hat{f}_{Y_{i0}|D_i=1}(y) = \frac{1}{\sum_i D_i}\sum_i D_i 1[Y_{i0} = y]$. Under standard regularity conditions, the $\hat{f}_{Y_{i1}(0)|D_i=1}(y)$ will be jointly asymptotically normally distributed, and thus one can test the null hypothesis that $f_{Y_{i1}(0)|D_i=1}(y) \geqslant 0$ for all y in the support of Y using methods from the moment inequality literature (Canay and Shaikh (2017) provide a review). When Y has continuous support, one can apply a similar approach by first discretizing the outcome variable, or by using methods for testing a continuum of moment inequalities (e.g. Andrews and Shi (2013)).

In addition to these formal falsification tests, a researcher may further assess the plausibility of (2) by (i) estimating the counterfactual distribution implied by (7) and assessing whether it is economically reasonable, and (ii) assessing whether "parallel trends of CDFs" holds in the pre-treatment period, e.g. by constructing an event-study plot for the transformed outcome $1[y \leqslant \bar{y}]$ for different values of $\bar{y}$.

Empirical Illustration. We illustrate how such tests can be used with a stylized application to the effects of the minimum wage. Our pre-treatment period is 2007 or 2010 (depending on the specification), our post-treatment period is 2015, and the treatment is whether a state raised its minimum wage at any point between the pre-treatment and post-treatment periods. The outcome of interest is individual wages W_i (where $W_i = 0$ if i is not working). We use data from Cengiz et al. (2019), who compile panel data on state-level minimum wages and employment-to-population ratios in 25-cent wage-bins.[10] Note that the employment-to-population ratio in wage bin w corresponds with the mass function of W_i at w. For each

---

[10]Following Cengiz et al. (2019), wages are adjusted for inflation using the CPI-UR-S and expressed in constant 2016 dollars, and we exclude from the treated group states that only had small minimum wage changes of less than 25c (in 2016 dollars) or which affected less than 2 percent of the workforce.

wage bin w, we infer the treated population's counterfactual employment-to-population ratio in wage bin w as $\hat{f}_{Y_{i,post}(0)|D_i=1}(w) = \hat{f}_{Y_{i,pre}|D_i=1}(w) + (\hat{f}_{Y_{i,post}|D_i=0}(w) - \hat{f}_{Y_{i,pre}|D_i=0}(w))$, where $\hat{f}_{Y_{i,t}|D_i=d}$ is the sample employment-to-population ratio in period t in states in treatment group d. In all calculations, we weight states by their population.

Figure 1: Implied Employment-to-Population Ratios for Treated States in 2015
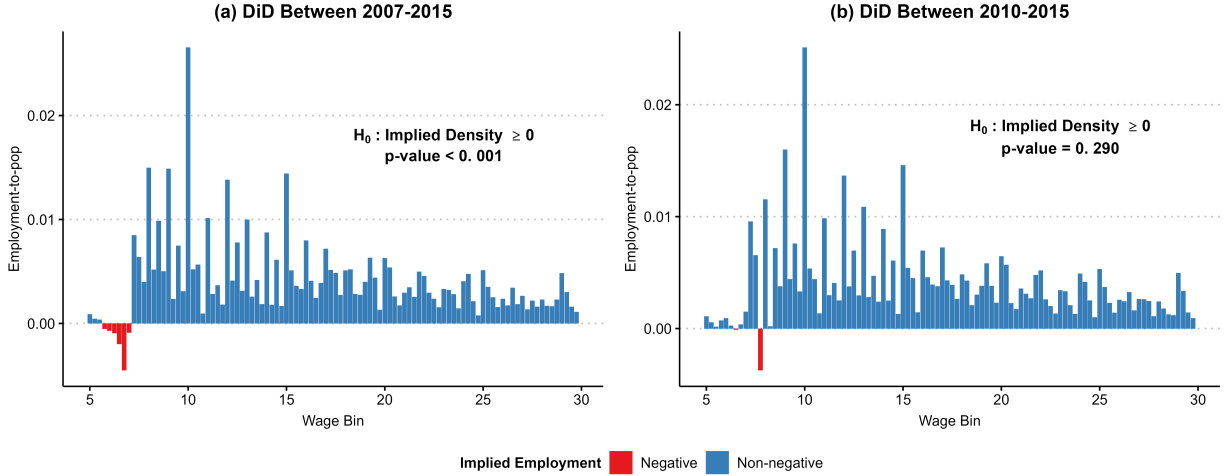

Figure 1 shows the implied counterfactual densities $\hat{f}_{Y_{i,post}(0)|D_i=1}(w)$ under parallel trends of distributions. In the left panel, where the pre-treatment period is 2007, the figure shows that the implied density is *negative* for wages between approximately \$5-7/hour. Intuitively, this occurs because the decrease in employment in such wage bins in comparison states between 2007 and 2015 is larger than the initial employment in treated states (who had lower baseline levels in these bins). One possible explanation for this pattern is that the increase in the federal minimum wage over this time period had a disproportionate impact on low-wage employment in states that did not have state-level minimum wages. To formally test the null that the implied density $f_{Y_{i,post}(0)|D_i=1}(w)$ is positive for all w, we estimate the variance-covariance matrix of the $\hat{f}_{Y_{i,post}(0)|D_i=1}(w)$ using a bootstrap at the state level, and then compare the minimum studentized value to a "least-favorable" critical value for moment inequalities that assumes all of the moments have mean 0 (see, e.g., Section 4.1.1 of Canay and Shaikh (2017)). Using such tests, we are able to reject the null hypothesis that all of the implied densities are positive ($p < 0.001$). We thus reject the null of parallel trends of CDFs in this context, which in turn implies that parallel trends cannot hold for all monotonic transformations of the outcome. This suggests that a researcher using such a DiD analysis to estimate the ATT for an outcome of the form $g(W_i)$ should be careful to justify the validity of the parallel trends assumption for the particular choice of functional form. Likewise, these results suggest that equation (7) will not be valid for inferring the entire counterfactual

distribution of wages in this setting. By contrast, in the right panel of Figure 1, which shows results using the period 2010-2015, we see that the estimated counterfactual distribution has positive density nearly everywhere, and we cannot formally reject the hypothesis that it is positive everywhere ($p = 0.29$). This does not necessarily imply that parallel trends holds for all transformations of the outcome, but insensitivity to functional form is not rejected by the data in this example.

Caveats. The tests described above may be useful in detecting situations where the data suggests that parallel trends will clearly be sensitive to functional form. We note, however, that failure to reject such tests should not be interpreted as strong evidence that the choice of functional form does not matter. First, such tests are for the null that there is *some* possible counterfactual distribution for the treated group in period 1 such that parallel trends is insensitive to functional form. Yet even if the null is satisfied there is no guarantee that the distribution that leads to such insensitivity would have been realized under the counterfactual. In other words, parallel trends of CDFs is *falsifiable* but *not verifiable*. Second, as with tests of pre-existing trends (Roth, 2021), there are concerns that such tests may have low power against violations of the null in finite samples, and conditioning the analysis on the result of such a pre-test may distort estimation and inference. It is also worth highlighting that even if parallel trends is sensitive to functional form it may still be valid for a particular functional form of interest – thus, if the tests described above reject, this does not imply that the DiD design is necessarily invalid, but rather that the researcher should provide justification for the validity of parallel trends for the chosen functional form.

3.3 Extensions

Remark 6 (Different classes of transformations)**.** Following Athey and Imbens (2006), we define parallel trends to be invariant to transformations if it holds for all strictly monotonic (measurable) functions. It is straightforward to show, however, that if (2) holds, then parallel trends holds for all (measurable) g. Hence, parallel trends for all strictly monotonic functions is equivalent to parallel trends for *all* measurable functions. We note that this equivalence does not hold for other assumptions – e.g., the CiC model of Athey and Imbens (2006) is invariant to *strictly* monotonic transformations but not to all measurable transformations. We note further that once one requires parallel trends to hold for a "sufficiently rich" set of transformations, this will imply that it holds for all transformations. For instance, if $Y_{it}(0)|D = d$ has a moment-generating function (MGF) for all (d, t), then it is sufficient to

consider the set of smooth exponential transformations $g_\lambda(y) = exp(\lambda y)$ for $\lambda \in \mathbb{R}$.[11] In the working paper version of this paper, we showed that that the distribution of untreated potential outcomes may only be partially identified if g is restricted to smaller classes of functions (see Appendix B).

Remark 7 (Use of pre-treatment periods). In settings where multiple pre-treatment periods are available, researchers may be inclined to use pre-treatment data to justify the parallel trends assumption for a particular functional form. However, the fact that parallel trends holds in the pre-treatment period (for a particular functional form) need not imply that it holds in the post-treatment period (for that functional form) – see Kahn-Lang and Lang (2020) for intuitive counterexamples. Indeed, we showed in the working paper version of this paper that when the support of $Y(0)$ is sufficiently rich, there will be multiple transformations g for which parallel trends holds in the pre-treatment period and a counterfactual post-treatment distribution such that parallel trends fails for at least one of these transformations. Moreover, in practice tests of pre-trends may have low power, and relying on such tests introduces statistical issues related to pre-testing (Roth, 2021). We thus encourage researchers to provide an *ex ante* justification for the necessary parallel trends assumption.

4 Invariance of other estimators

The results in the previous section show that for parallel trends to be invariant to transformations, we require an assumption that pins down the entire distribution of counterfactual potential outcomes. A natural question is whether we might be able to construct a different estimator that allows for consistent estimation of the ATT for all monotonic transformations under weaker assumptions that do not pin down the full counterfactual distribution. The following result shows that the answer is no.

Proposition 4.1. *Define*

$$\tau_{ATT}(g) = \mathbb{E}\left[g(Y_{i1}(1)) - g(Y_{i1}(0)) \mid D_i = 1\right].$$

Let $\mathcal{G}$ be the set of strictly monotonic functions g for which $\tau_{ATT}(g)$ is finite, and assume the identity map is in $\mathcal{G}$. Then $\tau_{ATT}(g)$ is identified for all $g \in \mathcal{G}$ if and only if $F_{Y_{i1}(0)|D_i=1}(\cdot)$ is identified.

---

[11]Specifically, parallel trends for this class of functions implies a "parallel trends"-type assumption for MGFs. Inverting both sides of the equation via inverse Laplace transforms then yields parallel trends of CDFs.

Proof. Suppose first $F_{Y_{i1}(0)|D_i=1}(\cdot)$ is identified. Then $\mathbb{E}\left[g(Y_{i1}(0))\,|\,D_i=1\right]$ is identified, since it is equal to $\int g(y)dF_{Y_{i1}(0)|D_i=1}$. Further, $\mathbb{E}\left[g(Y_{i1}(1))\,|\,D_i=1\right]=\mathbb{E}\left[g(Y_{i1})\,|\,D_i=1\right]$, and thus is also identified. Hence $\tau_{ATT}(g)=\mathbb{E}\left[g(Y_{i1}(1))\,|\,D_i=1\right]-\mathbb{E}\left[g(Y_{i1}(0))\,|\,D_i=1\right]$ is identified.

Conversely, suppose $\tau_{ATT}(g)$ is identified for all $g\in\mathcal{G}$. By assumption, the identity map $g_1(y)=y$ is contained in $\mathcal{G}$. It follows that for any $\tilde{y}\in\mathbb{R}$, $g_2(y)=y-1[y\leqslant\tilde{y}]$ is also contained in $\mathcal{G}$, since it is the sum of g_1 and a bounded function. Now,

$$\tau_{ATT}(g_1)-\tau_{ATT}(g_2)=\mathbb{E}\left[1[Y_{i1}(1)\leqslant\tilde{y}]\,|\,D_i=1\right]-\underbrace{\mathbb{E}\left[1[Y_{i1}(0)\leqslant\tilde{y}\,|\,D_i=1\right]}_{=F_{Y_{i1}(0)|D_i=1}(\tilde{y})}$$

and hence

$$F_{Y_{i1}(0)|D_i=1}(\tilde{y})=\tau_{ATT}(g_2)-\tau_{ATT}(g_1)+\mathbb{E}\left[1[Y_{i1}(1)\leqslant\tilde{y}]\,|\,D_i=1\right].$$

However, the first two terms on the right-hand side of the previous display are identified by assumption, and the final term is equal to $\mathbb{E}\left[1[Y_{i1}\leqslant\tilde{y}]\,|\,D_i=1\right]$ and thus is identified. The result follows.

$\square$

Proposition 4.1 states that identification of $\tau_{ATT}(g)$ for all monotonic transformations g is equivalent to identification of the full distribution of untreated potential outcomes for the treated group. An immediate implication of this result is that there can exist a consistent or unbiased estimator of $\tau_{ATT}(g)$ for all g only if one imposes assumptions that identify the full distribution of untreated potential outcomes. Thus, the consistency or unbiasedness of an estimator for the ATT will necessarily be sensitive to functional form *unless* one imposes assumptions that identify the full distribution of counterfactual potential outcomes.

Remark 8 (Other classes of transformations). Analogous to Remark 6, identification of $\tau_{ATT}(g)$ for a sufficiently rich set of g functions is equivalent to identification of $\tau_{ATT}(g)$ for all measurable functions. For instance, if $Y_{i1}(d)|D_i=1$ has a moment generating function for $d=0,1$, then it suffices to consider the smaller set of transformations of the form $g_\lambda(y)=exp(\lambda y)$ for $\lambda\in\mathbb{R}$.

5   Implications for Applied Work

Our results imply that there are three paths that a researcher can take to justify point-identification of the ATT when considering using difference-in-differences or a related research design.

Path 1: Justify Randomization. If treatment is (as-if) randomly assigned, then the parallel trends assumption will hold regardless of the chosen functional form. Thus, sensitivity to functional form will not be an issue for DiD designs in settings with (as-if) random assignment. We note, however, that other estimators will also be valid under randomization of treatment and may be preferred over DiD on the basis of efficiency (McKenzie, 2012).

Path 2: Impose a distributional assumption. If treatment is not plausibly randomly assigned, then the researcher can impose an alternative assumption that pins down the full counterfactual distribution of $Y_{i1}(0)$ for the treated group, and then choose an estimator that is valid regardless of functional form under this assumption. If the researcher is confident in parallel trends of CDFs (e.g. they are in Case 3 above), then the DiD estimator will be valid regardless of functional form. Alternatively, the researcher may be more confident in the CiC model, in which case the estimator proposed by Athey and Imbens (2006) will be valid regardless of functional form.

Path 3: Justify a particular functional form. If the researcher is not willing to impose any assumptions that pin down the full counterfactual distribution, then Proposition 4.1 implies that the validity of any estimator for the ATT will depend on functional form. However, the researcher might use context-specific knowledge to argue for the validity of an assumption for a particular functional form. For example, in some contexts the researcher might be willing to impose a Cobb-Douglas production function for the untreated potential outcome, i.e. $Y_{it}(0) = A_{it} L_{it}^{\alpha} K_{it}^{1-\alpha}$, where L_{it} and K_{it} are capital and labor, and the productivity A_{it} can be written as $A_{it} = \nu_i \theta_t \eta_{it}$ for η_{it} a random shock that is independent of the other variables in the model (including treatment status D_i). Then, taking logs yields

$$\ln(Y_{it}(0)) = \underbrace{\lambda_i}_{=\ln(\nu_i)} + \underbrace{\phi_t}_{=\ln(\theta_t)} + \alpha L_{it} + (1-\alpha)K_{it} + \underbrace{\epsilon_{it}}_{=\ln(\eta_{it})}, \text{ where } \mathbb{E}\left[\epsilon_{it}|D_i, \lambda_i, \phi_t, L_{it}, K_{it}\right] = 0,$$

and thus parallel trends holds in logs (once we control for L_{it} and K_{it}).

Alternative approaches. In some settings, the researcher may not be confident in justifying the assumptions for any of the three possible paths for identification outlined above. In such cases, an appealing alternative may be to impose weaker assumptions that yield partial identification of the ATT. For example, Manski and Pepper (2018) and Rambachan and Roth (2020) provide tools for partial identification and inference on the ATT under assumptions that impose that parallel trends holds only approximately for a given functional form. These tools will be appropriate if the researcher thinks that parallel trends holds ap-

proximately for the chosen functional form but may not be confident that it holds exactly. Likewise, we showed in the working paper version of this article that these tools can be adapted to conduct inference on the ATT in levels under the assumption that parallel trends holds for some smooth transformation g. This type of approach may be attractive when the researcher is not sure what functional form is exactly correct but has a sense of the types of transformations that may be reasonable. We view these partial identification approaches as a potentially attractive middle ground between imposing that parallel trends holds exactly for a particular functional form and requiring it to hold for all functional forms.

Closing remarks. We suspect that there are contexts in which each of these approaches to identification may be most appropriate. Our hope is that the results in this paper will be useful in providing researchers with a menu of options for more clearly delineating the justification for their research design.

References

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

Almond, Douglas, Hilary W. Hoynes, and Diane Whitmore Schanzenbach, "Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes," *Review of Economics and Statistics*, May 2011, *93* (2), 387–403.

Andrews, Donald W. K. and Xiaoxia Shi, "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, *81* (2), 609–666.

Athey, Susan and Guido Imbens, "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 2006, *74* (2), 431–497.

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

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2020, (Forthcoming).

_ **and Tong Li**, "Quantile treatment effects in difference in differences models with panel data," *Quantitative Economics*, 2019, *10* (4), 1579–1618.

Canay, Ivan and Azeem Shaikh, "Practical and Theoretical Advances in Inference for Partially Identified Models," in Bo Honoré, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, eds., *Advances in Economics and Econometrics*, 2017, pp. 271–306.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The Effect of Minimum Wages on Low-Wage Jobs," *The Quarterly Journal of Economics*, August 2019, *134* (3), 1405–1454. Publisher: Oxford Academic.

Cunha, Flavio and James J. Heckman, "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, October 2008, *43* (4), 738–782. Publisher: University of Wisconsin Press.

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

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, 1 edition ed., New York: Cambridge University Press, April 2015.

Kahn-Lang, Ariella and Kevin Lang, "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications," *Journal of Business & Economic Statistics*, July 2020, *38* (3), 613–620. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07350015.2018.1546591.

Leamer, Edward E., "Let's Take the Con Out of Econometrics," *The American Economic Review*, 1983, *73* (1), 31–43. Publisher: American Economic Association.

_ , "Sensitivity Analyses Would Help," *The American Economic Review*, 1985, *75* (3), 308–313. Publisher: American Economic Association.

Manski, Charles F. and John V. Pepper, "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," *Review of Economics and Statistics*, 2018, *100* (2), 232–244.

McKenzie, David, "Beyond baseline and follow-up: The case for more T in experiments," *Journal of Development Economics*, 2012, *99* (2), 210–221. Publisher: Elsevier.

Meyer, Bruce D., "Natural and Quasi-Experiments in Economics," *Journal of Business & Economic Statistics*, 1995, *13* (2), 151–161. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Rambachan, Ashesh and Jonathan Roth, "An Honest Approach to Parallel Trends," *Working Paper*, 2020.

Rosenbaum, Paul R., "The Role of a Second Control Group in an Observational Study," *Statistical Science*, August 1987, *2* (3), 292–306. Publisher: Institute of Mathematical Statistics.

Roth, Jonathan, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *Working paper*, 2021.

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

Stepner, Michael, "The Insurance Value of Redistributive Taxes and Transfers," 2019.

Sun, Liyan and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2020, (Forthcoming).

A   Proof of Proposition 3.2

Proof. By Proposition 3.1, it suffices to show equivalence with (2). Observe that if (4) holds, then both sides of (2) reduce to $\theta(G_1(y) - G_0(y))$, and so (4) implies (2). To prove the converse, let $\mathcal{Y}$ denote the parameter space for $Y(0)$, and $\mathcal{Y}_y = \{\tilde{y} \in \mathcal{Y} \mid \tilde{y} \leqslant y\}$. By assumption, we can write

$$F_{Y_{it}(0)|D_i=d}(y) = \int_{\mathcal{Y}_y} f_{Y_{it}(0)|D_i=d}\, d\lambda,$$

where λ is the dominating measure and $f_{Y_{it}(0)|D_i=d}$ is the density (the Radon-Nikodym derivative). It is immediate from the previous display that if (2) holds, then $f_{Y_{i1}(0)|D_i=1} - f_{Y_{i0}(0)|D_i=1} = f_{Y_{i1}(0)|D_i=0} - f_{Y_{i0}(0)|D_i=0}$, λ a.e. The desired result then follows by applying Lemma A.1 to the CDFs on both sides of (2).

$\square$

Lemma A.1. *Suppose the CDFs F_1 and F_2 are such that $F_j(y) = \int_{\mathcal{Y}_y} f_j d\lambda$. Then we can decompose $F_j(y)$ as*

$$F_j(y) = (1 - \theta)F_{min}(y) + \theta\tilde{F}_j(y), \tag{8}$$

where F_{min} and $\tilde{F}_1, \tilde{F}_2$ are CDFs, $\theta \in [0, 1]$, and θ and $\tilde{F}_j$ depend on f_1 and f_2 only through $f_1 - f_2$.

Proof. To prove the claim, set $\theta = 1 - \int_{\mathcal{Y}} \min\{f_1, f_2\} d\lambda$. It is immediate that $\theta \in [0,1]$. Suppose first that $\theta \in (0,1)$. Define

$$f_{min} = \frac{\min\{f_1, f_2\}}{\int_{\mathcal{Y}} \min\{f_1, f_2\} d\lambda} = \frac{\min\{f_1, f_2\}}{1 - \theta}$$

and

$$\tilde{f}_j(x) = \frac{f_j - \min\{f_1, f_2\}}{\int_{\mathcal{Y}} (f_j - \min\{f_1, f_2\}) d\lambda} = \frac{f_j - \min\{f_1, f_2\}}{\theta} \text{ for } j = 1, 2.$$

By construction, f_{min} and the $\tilde{f}_j$ integrate to 1 and are non-negative, so that $F_{min}(y) = \int_{\mathcal{Y}_y} f_{min} d\lambda$ and $\tilde{F}_j(y) = \int_{\mathcal{Y}_y} \tilde{f}_j d\lambda$ are valid CDFs. Moreover, $f_j = (1-\theta)f_{min} + \theta\tilde{f}_j$ by construction, so that (8) holds. Finally, observe that $\min\{f_1, f_2\} = f_1 - (f_1 - f_2)_+$, where $(a)_+$ denotes the positive part of a. It follows that $\theta = 1 - \int_{\mathcal{Y}} (f_1 - (f_1 - f_2)_+) d\lambda = \int_{\mathcal{Y}} (f_1 - f_2)_+ d\lambda$, which depends only on $f_1 - f_2$. (In fact, note that $\int_{\mathcal{Y}} (f_1 - f_2)_+ d\lambda = \frac{1}{2} \int_{\mathcal{Y}} |(f_1 - f_2)| d\lambda$, and thus θ is the total variation distance between f_1 and f_2.) Likewise, $\tilde{f}_1 = (f_1 - f_2)_+/\theta$ and $\tilde{f}_2 = (f_2 - f_1)_+/\theta$, and so depend only on $f_1 - f_2$. This completes the proof for the case where $\theta \in (0,1)$. If $\theta = 0$, then $F_1(y) = F_2(y)$ and so the claim holds trivially with $F_{min}(y) = F_1(y) = F_2(y)$ and $\tilde{F}_j(y)$ arbitrary. If $\theta = 1$, then $\min\{f_1, f_2\} = 0$ λ a.e, and so $f_1 = (f_1 - f_2)_+$ λ a.e., and $f_2 = (f_2 - f_1)_+$ λ a.e.. Thus, the claim holds trivially with $\tilde{f}_j = f_j$, $\tilde{F}_j(y) = \int_{\mathcal{Y}_y} \tilde{f}_j d\lambda$, and $F_{min}(y)$ arbitrary. $\square$