# Causal Duration Analysis with Diff-in-Diff*

Ben Deaner† and Hyejin Ku‡
University College London

November 22, 2024

**Abstract**

In economic program evaluation, it is common to obtain panel data in which outcomes are indicators that an individual has reached an absorbing state. For example, they may indicate whether an individual has exited a period of unemployment, passed an exam, left a marriage, or had their parole revoked. The parallel trends assumption that underpins difference-in-differences generally fails in such settings. We suggest identifying conditions analogous to those of difference-in-differences, but which apply to hazard rates rather than mean outcomes. These alternative assumptions motivate estimators that retain the simplicity and transparency of standard diff-in-diff, and we suggest analogous specification tests. Our approach can be adapted to include general linear restrictions between the hazard rates of different groups, motivating duration analogues of the triple differences and synthetic control methods. We apply our procedures to examine the impact of a policy that increased the generosity of unemployment benefits, using a cross-cohort comparison.

*We thank seminar participants at Brown University and The University of Surrey, and those who attended the presentation of this paper at the European Winter Meeting of the Econometric Society 2023 for helpful comments and feedback. We would also like to thank Guido Imbens, Xavier D'Haultfoeuille, anonymous referees, and colleagues at UCL for their insights, particularly Christian Dustmann, Dennis Kristensen, Liyang Sun, and Andrei Zeleneev, and others who attended internal presentations of this work.

†b.deaner@ucl.ac.uk
‡h.ku@ucl.ac.uk

Many important topics in applied economics involve durations. To name a few, the impact of unemployment insurance on the length of unemployment spells (Katz & Meyer (1990), Hunt (1995), Lalive *et al.* (2006), Lalive (2007), Card *et al.* (2007), Chetty (2008), Schmieder *et al.* (2012), Schmieder *et al.* (2016), Lichter & Schiprowski (2021), and others), the effect of divorce laws on marriage duration (Friedberg (1998), Gruber (2004), Wolfers (2006)), the strength of residency rules on the rate at which refugees pass language tests (Arendt *et al.* (2024)), the impact of insurance schemes on fertility rates (Lambert & Rossi (2016)), mandatory retirement rules on rates of retirement (Ashenfelter & Card (2002)), and the consequences of criminal justice policies on the rate of recidivism or probation revocation (Schmidt & Witte (1989), Bhuller *et al.* (2020), Rose (2021)). In settings like these, available data may consist of panels or repeated cross-sections in which the outcome is a binary indicator that an individual has entered an absorbing state. For example, an indicator that an individual has exited unemployment by a particular date. Difference-in-differences is a popular tool for policy-evaluation, but when the data take the form just described, the parallel trends assumption generally fails.

To fix ideas, consider the case in which the outcome indicates exit from unemployment. Suppose that some individuals receive an increase in unemployment benefits at a particular point in time while others do not. The foundational assumption of diff-in-diff is that, absent the policy of interest, the difference in mean outcomes between the treated and untreated groups would remain fixed. If a sufficiently large share of individuals eventually exit unemployment, then mean outcomes will tend to converge over time, even absent any treatment effect.[1] This entails a failure of parallel trends and may result in severely biased and inconsistent estimates of treatment effects.

It is well-known that the parallel trends assumption is generally incompatible with settings in which outcomes are discrete or truncated (see Wooldridge (2023)). What makes this problem particularly acute in duration settings is that the resulting bias is systematic. Suppose a higher proportion of individuals in the treated group are unemployed just prior to treatment. Then under a counterfactual in which no one is treated, the gap in unemployment rates between the treated and untreated groups would tend to narrow over the post-treatment period. Standard diff-in-diff estimates attribute this narrowing to the effect of treatment, leading to over-estimation of the effect of treatment on job-finding.

[1] Apart from in the special case in which mean outcomes are identical between the two groups.

By contrast, in cases in which the binary outcomes reflect say, choices among discrete alternatives, it is less clear that the misspecification of the model would systematically bias the estimates in any particular direction.

In response, we consider alternative approaches based on insights from duration analysis. Rather than assume parallel trends between mean potential outcomes, we suppose there is a constant linear relationship between group-specific counterfactual hazard rates. This condition is consistent with the convergence of counterfactual mean outcomes over time, and thus the convergence does not imply inconsistency of the corresponding causal estimates. The alternative assumption motivates a simple fix. Instead of performing diff-in-diff or related methods using mean outcomes, we apply the same procedures using a particular function of the mean outcomes and the time period. We sidestep any estimation of the hazard rates themselves, and our analysis entirely avoids the need to specify an explicit model for the hazard function.

To be precise, in place of the mean outcome $E[Y_{i,t}|G_i = k]$, where $i$ indexes individuals, $t$ is the time period, and $G_i$ is the group to which individual $i$ belongs, we use the 'time-average hazard', denoted $\bar{H}_{k,t}$, which can be written in terms of mean outcomes. This object is the average of the hazard rate over the continuous period between times 1 and $t$. If we assume parallel trends in group-specific counterfactual hazard rates, then there is a fixed level difference between the time-average hazards of the different groups under the no-treatment counterfactual. Thus we can perform difference-in-differences using this alternative object. We may then invert the definition of the time-average hazard to recover counterfactual mean outcomes and average treatment effects. Because we use only group-specific mean outcomes, our methods are applicable to repeated cross-sections as well as panel data.

Note that the counterfactual mean outcome, understood as a function of time since the start of a spell, is precisely the cumulative distribution function of counterfactual durations. Thus our approach can produce estimates of the distribution of counterfactual durations, rather than say, the mean duration under a no-policy counterfactual.

Diagnostics like placebo tests and the standard test for pre-treatment parallel trends can be adapted straight-forwardly to our setting. Where standard tests use the mean outcome, we again use the time-average hazard rate. Analogous with vanilla diff-in-diff, one can perform informal visual inspections to assess whether parallel trends in time-average hazards holds in the pre-treatment period.

Adjusting for covariates can be important for credibly identifying causal effects in difference-in-differences. If there is covariate imbalance between treatment and control groups in diff-in-diff, then parallel trends may fail, even if it holds within each covariate stratum (say, within each demographic subgroup). We adapt our hazard approach to allow for such cases by employing a simple propensity score weighting when we calculate the time-average hazard. This avoids the difficulties of partial likelihood estimation of proportional hazard models with discretised time periods.

As stated above, our approach extends beyond the assumption of a fixed difference between hazard rates. Rather, we can accommodate any fixed linear relationship between the hazard rates of different groups. Thus our approach applies if there is a fixed difference in log hazard rates, which is equivalent to a particular proportional hazards specification. In the case of multiple untreated groups we can obtain a duration analogue of the triple differences estimator and of the synthetic control method. As with our duration diff-in-diff approach, estimation differs from standard approaches only in the use of transformed mean outcomes.

We apply our methods to the setting of Lalive *et al.* (2006). The authors in that study evaluate the impact of a policy that increased the generosity of unemployment insurance benefits for unemployed Austrian workers. The authors identify causal effects by exploiting the presence of individuals who were ineligible for the benefit changes. They estimate a flexible parametric duration model and from their estimates they recover causal effects.

In contrast, we identify causal effects using a cross-cohort comparison. We employ our methods in order to estimate the impact of an extension to the potential benefit duration (PBD). We adjust for the calendar date at which an unemployment spell begins using our covariate re-weighting strategy. Thus our estimates are robust to differential trends in job-seeking between individuals who become unemployed in different parts of the year. We obtain similar results to the authors of the original study, however we do so while avoiding estimation and specification of the hazard function, and without the need for numerical optimization of a likelihood. Moreover, the transparency of our approach allows us to both visually and formally assess whether there are parallel trends over the pre-treatment period. We find a statistically significant positive impact of PBD on unemployment duration with strongly positive estimates shortly following treatment which then taper off. We are unable to reject parallel trends in hazard rates even at the 50% level.

We evaluate the finite-sample performance of our methods in a simulation study. The results also demonstrate the potential for standard diff-in-diff to produce severely misleading estimates in duration settings. The simulation results are available in the supplementary appendices.

In sum, the present paper suggests a simple means of adapting existing difference-in-differences and related methods to settings with duration data. Whereas standard procedures extrapolate the relationship between mean outcomes for different groups forward in time, we instead extrapolate the relationship between these groups' time-average hazard rates, which are known functions of mean outcomes and therefore easily estimated from the data. What is crucial here is that we retain the intuitive appeal of diff-in-diff. One of the key benefits of diff-in-diff and related methods is that one can visually assess whether the assumption of a stable linear relationship holds in the pre-treatment period, and assess the magnitude of any deviation from this condition. Because we simply shift the objects to which these methods are applied, our approach allows researchers to perform similar visual inspections, as required for effective causal event studies.

## Related Literature

We are not the first to suggest an extension of differences-in-differences to duration settings. Proportional hazard models (Cox (1972)) with a diff-in-diff-type linear index are considered by Hunt (1995), Wu & Wen (2022), Lalive *et al.* (2006), and Marinescu (2009). Wu & Wen (2022) show that parallel trends cannot hold when the data are generated by the proportional hazards model. Hunt (1995) suggests an approximate partial likelihood estimation procedure for the coefficients of the linear index in a proportional hazard diff-in-diff specification, Lalive *et al.* (2006) and Marinescu (2009) employ maximum likelihood estimation. Wu & Wen (2022) consider estimation in the two-period case. In both Wu & Wen (2022) and Hunt (1995), interest is in the estimation of the coefficients on the binary indicators in the linear index, rather than average treatment effects.

The proportional hazard diff-in-diff model is a special case of the more general linear restrictions that we consider in the present work. Our approach avoids maximum likelihood estimation in favor of simple and transparent nonparametric imputation of time-average hazards that closely resembles classic diff-in-diff. The simplicity and transparency of our approach is important not

only in that it facilitates practical application of our methods, but because it allows us to both visually and formally test for pre-treatment parallel trends, much as in standard diff-in-diff. Our preferred specification assumes a fixed level difference in the hazard rates rather than a fixed ratio. This has the advantage of allowing for a simple means of flexibly incorporating covariates via propensity score weighting.

Also related to our approach is the literature on non-linear difference-in-differences. These papers, usually motivated by a limited dependent variable, specify a generalized linear model (GLM) in which the outcome is a non-linear transformation of a linear index with a diff-in-diff form. A number of empirical papers estimate the coefficients in a GLM diff-in-diff model. For example, Gruber & Poterba (1994) and Eissa (1996). Puhani (2011) considers the interpretation of the coefficients in these models.

Motivation for GLM diff-in-diff models is discussed in Blundell *et al.* (2004), Blundell & Costa Dias (2009), Lechner (2011), and most thoroughly in Wooldridge (2023). The general strategy in Blundell *et al.* (2004) and Wooldridge (2023) is to transform mean outcomes by inverting the link function in a GLM and to perform diff-in-diff on the transformed means. Re-applying the link function then recovers counterfactual mean outcomes and thus treatment effects. Their specifications are motivated by latent variable models for the discrete outcomes. Thus in those works, the link function is determined by an a priori assumption that an unobserved noise term follows a known parametric probability distribution. In our case, the form of link function follows from the duration structure of the data. For example, from the proportional hazards version of our identifying assumption we obtain the same restriction on mean outcomes as the exponential model in Wooldridge (2023), however we obtain this model from a distinct starting point. Our approach also covers the case in which there are more than two groups (treated and untreated) and allows for general linear restrictions.

In addition, our approach differs from Wooldridge (2023) in how we adjust for covariates. Rather than incorporating covariates in the linear index inside of a link function, we instead suggest a propensity score weighting approach and provide formal justifications for this method. As we discuss later on, this approach may be better suited to the duration setting and has some additional advantages.

Works in which the link function is estimated using sufficiently rich pre-treatment data include Ashenfelter & Greenstone (2004) and Athey & Imbens (2006). The latter suggest a nonparametric approach that is valid under weaker

conditions than the GLM diff-in-diff model.

By proposing a simple and transparent extension of diff-in-diff to duration settings, we extend the ever-growing literature on difference-in-differences. For recent surveys see for example Roth *et al.* (2023) and de Chaisemartin & D'Haultfœuille (2023). More general works that consider the failure of parallel trends under alternative functional forms include Roth & Sant'Anna (2023) and Rambachan & Roth (2023). Our work is more narrow in that we focus on a particular case in which standard parallel trends generally fails, but we show that a particular non-linear diff-in-diff method and some related approaches are valid in this setting.

In our empirical application we employ a cross-cohort comparison in which the time period for a given individual is relative to the start of their unemployment duration. This is the same approach taken in Van Den Berg (2020). However in that work, the authors employ an identification strategy based on regression discontinuity design for nonparametrically estimated hazard rates. In addition, Van Den Berg (2020) focus on identification and estimation of average differences in counterfactual hazard rates, whereas our objects of interest are counterfactual mean outcomes (i.e., the counterfactual distribution of durations).

There is a sizable literature on causal analysis using duration data (see e.g., Abbring & Heckman (2007) for a survey). In the seminal paper of Abbring & Van Den Berg (2003) and similar works, researchers achieve identification under assumptions on the treatment process, its relation to individual heterogeneity, and the separability of unobserved heterogeneity in individual-level hazard rates. Our approach differs from this in its foundation. Rather than begin with an individual-level duration model incorporating heterogeneity, our analysis is premised upon an assumption concerning group-level hazard rates in spirit of classic diff-in-diff.

Abbring & van den Berg (2005) identify and estimate causal effects from duration data using instrumental variables. Vikström (2017) suggests a dynamic inverse propensity score weighting to estimate treatment effects under an unconfoundedness assumption using discrete time duration data. van den Berg & Vikström (2022) use a type of inverse propensity score weighting to identify causal objects under a sequential unconfoundedness condition when treatment can only be applied to individuals in a particular state (e.g., individuals can only receive an unemployment benefit during when they are unemployed). Cui *et al.* (2023) develop a method based on random forests to estimate conditional

average treatment effects under the assumption that treatment assignment and censoring are ignorable conditional on covariates.

Unlike Abbring & van den Berg (2005), we do not require the presence of a valid instrument, instead we rely on a type of parallel trends assumption. To compare with the other works which rely on unconfoundedness/ignorability, note that our identification approach with covariate adjustment (outlined in Section 2.1) rests on an assumption that must hold under unconfoundedness (that is, when allocation to the treatment group is independent of counterfactual outcomes conditional on covariates). On the other hand, we do not consider settings with an eligibility state like van den Berg & Vikström (2022), nor do we identify conditional average causal effects like Cui *et al.* (2023). Moreover, we only consider adjustment for time-invariant covariates unlike Vikström (2017) and van den Berg & Vikström (2022).

In some cases it may be possible to apply difference-in-differences using the durations themselves as outcomes as in Lichter & Schiprowski (2021). Consider again the unemployment example and suppose some individuals are ineligible for the increase in benefits. One might perform diff-in-diff by taking the difference in the mean unemployment durations between those who became unemployed before and after the reform and seeing how this varies with eligibility for the benefits increase. This approach differs from ours both in the settings to which it is applicable and in the causal objects it identifies.

First, consider that this alternative strategy identifies causal effects using variation in the start date of unemployment spells, whereas our method is applicable even if all spells in the data begin on the same date. Moreover, such methods require some form of correction for censoring if either the length of spells is discretized as is common in panel data, or some spells are incomplete (that is, if an individual's unemployment lasts beyond the final period of the data). In addition, suppose that the policy change applies to those in ongoing unemployment spells, as is common in practice. Then the group of eligible individuals who became unemployed prior to the reform will contain some treated individuals, and dropping these from the sample would lead to selection bias. Our approach avoids this problem.

Using durations as outcomes identifies distinct causal objects compared with our approach. Suppose that the policy impacts not only time spent in unemployment, but also who becomes unemployed. This compositional change represents an additional channel through which the policy might impact unemployment durations. Because we follow the same fixed set of individuals over time, we isolate

the behavioral response of individuals in ongoing spells. Finally, our approach recovers a counterfactual cumulative distribution of durations evaluated at certain points in its support, whereas the alternative identifies the counterfactual mean of durations.

# 1 Motivation and Background

We sample binary outcomes $Y_{i,t}$ for individuals $i = 1, ..., n$ at periods $t = 1, .., T$. Each individual belongs to a group $G_i$ where group membership is constant over time. An outcome of 1 indicates that an individual has entered an 'absorbing state' and so all future outcomes for that individual are also equal to 1. Individuals in Group 1 receive an intervention at some point between periods $t^* - 1$ and $t^*$ and we wish to assess the impact of this intervention on the evolution of the outcomes of individuals in that group.

Although we focus on panel data, a nice feature of our analysis is that it applies equally to repeated cross-sections. Moreover, the data need not contain the exact lengths of spells. Rather, we only need to know whether or not a spell has ended by a particular length of time and we accommodate incomplete spells. For example, if the time increments are weeks, then we need only know whether or not a given duration has ended by the $t$-th week for $t = 1, ..., T$, not the exact moment at which it ended. While we focus on the case of discrete time increments, our approach can be adapted straightforwardly to accommodate time increments of varying lengths.

In order to define causal effects of the treatment, we consider a counterfactual in which there is no intervention on Group 1. We denote by $Y_{i,t}^{(0)}$ the outcome under this counterfactual at time $t$ for individual $i$. We sometimes refer to $Y_{i,t}^{(0)}$ as an 'untreated potential outcome', although this differs from the standard definition in that the counterfactual is defined in terms of an intervention on Group 1 rather than an individual-level treatment. We use the superscript '(0)' to indicate counterfactual values throughout this work.

Our primary object of interest is the time-$t$ average treatment effect for individuals in the treated group where $t^* \leq t$. This is the average difference between the outcome for a randomly sampled individual in the treated Group 1 at time $t$, and that individual's outcome in the counterfactual world in which there is no intervention. This is defined as follows:

$$\tau_t = E[Y_{i,t} - Y_{i,t}^{(0)} | G_i = 1]$$

Note that $E[Y_{i,t}^{(0)}|G_i = 1]$ is the counterfactual cumulative distribution function (CDF) of the durations of individuals in Group 1 evaluated at $t$. Therefore, if treatment effects are identified, then so too are the values of the counterfactual CDF of durations evaluated at each discrete time increment.

To illustrate, consider two schools 1 and 2. We sample $n$ students, each of whom attends one of the two schools. If $G_i = k$ then individual $i$ attends school $k$. Data is available from periods $t = 1, ..., T$. The students in each school have the opportunity to sit and pass an English proficiency exam. If a student $i$ has passed the exam by time $t$ then $Y_{i,t} = 1$ and otherwise $Y_{i,t} = 0$. We suppose that at a time between $t^* - 1$ and $t^*$, the students in school 1 receive some educational intervention, where $t^*$ is known.

The binary indicator $Y_{i,t}^{(0)}$ is equal to 1 if and only if student $i$ would have passed the test by time $t$ in a counterfactual world in which there is never any intervention on school 1. In this setting, the time-$t$ average treatment effect $\tau_t$ is the difference between the share of the student population in school 1 who have passed the test by time $t$ versus the proportion who would have passed under the counterfactual in which there is no intervention on school 1.

Assumptions 1 and 2 formally impose some elementary properties of the factual and counterfactual outcomes implicit in the discussion above.

**Assumption 1 (Absorbing State).** $Y_{i,t}$ is a binary random variable and $Y_{i,t} = 1$ implies $Y_{i,s} = 1$ for all $t \leq s$. The same holds for the potential outcomes $Y_{i,t}^{(0)}$.

**Assumption 2 (No Anticipation/Spill-Overs).** i. $1 < t^*$ and for all $t < t^*$, $Y_{i,t} = Y_{i,t}^{(0)}$, ii. For all $t \geq t^*$, if $G_i > 1$ then $Y_{i,t} = Y_{i,t}^{(0)}$.

Assumption 1 states that having an outcome of 1 is absorbing state. This means that if an individual has an outcome of 1 at time $t$, then that individual's outcome is equal to 1 in all future periods, and similarly for potential outcomes. In the schools setting this follows simply from the definition of the outcome: a student cannot un-pass the exam.

Assumption 2 imposes conditions on potential outcomes that are standard in difference-in-differences. We assume that individuals do not anticipate treatment, and so observed outcomes in periods strictly prior to $t^*$ are identical to those in the counterfactual world in which there is no intervention. A no-anticipation condition was introduced into the literature on causal duration analysis by Abbring & Van Den Berg (2003). In addition, we assume there are

no spill-overs between groups. That is, the treatment of students in school 1 has no impact on students in other schools. More formally, the potential outcomes of students in other schools under the no-treatment counterfactual are equal to their realized outcomes. However, this does not rule out spill-overs between individuals in the same group.

## 1.1 Standard Diff-in-Diff and Related Methods

Difference-in-differences identifies causal effects under an assumption that there is a fixed level difference between the mean outcomes of the treated and untreated groups. Formally, this parallel trends condition imposes that there is some constant $c$ so that the following equation holds in all periods.

$$E[Y_{i,t}^{(0)}|G_i = 1] - E[Y_{i,t}^{(0)}|G_i = 2] = c \tag{1.1}$$

In the schools example, the condition above states that the difference in mean potential outcomes for the two schools is constant over time. Under Assumptions 1 and 2, the condition identifies the average treatment effect $\tau_t$ and motivates a simple estimator. Let $\bar{Y}_{k,t}$ be the sample average outcome for individuals in Group $k$ at time $t$. That is, for each $t$ and $k$:

$$\bar{Y}_{k,t} = \frac{1}{n_k} \sum_{i=1}^{n} 1\{G_i = k\} Y_{i,t}$$

Under Assumptions 1, 2, and parallel trends, an unbiased estimate $\hat{c}$ of $c$ is given below, where $\{\alpha_s\}_{s=1}^{t^*-1}$ are some positive weights that sum to 1.

$$\hat{c} = \sum_{s=1}^{t^*-1} \alpha_s (\bar{Y}_{1,s} - \bar{Y}_{2,s})$$

An unbiased estimator $\hat{\tau}_t$ of the time-$t$ average treatment effect is as follows.

$$\hat{\tau}_t = (\bar{Y}_{1,t} - \bar{Y}_{2,t}) - \hat{c}$$

The estimator above can be expressed equivalently in terms of ordinary least squares regression and can be adapted to include covariates. A number of inference methods have been proposed in the literature. For example, the block bootstrap method of Bertrand *et al.* (2004). With two periods of data (so that $t^* = 2$), the estimator above reduces to the simple difference-in-differences

below:

$$\hat{\tau}_2 = (\bar{Y}_{1,2} - \bar{Y}_{2,2}) - (\bar{Y}_{1,1} - \bar{Y}_{2,1})$$

Diff-in-diff is one example of a general class of methods that identify causal effects by assuming that there exists a fixed linear relationship between the counterfactual mean outcomes in different groups. That is, there exist parameters $W_1$,$W_2$,...,$W_K$ so that the following holds in all periods:

$$E[Y_{i,t}^{(0)}|G_i = 1] = W_1 + \sum_{k=2}^{K} W_k E[Y_{i,t}^{(0)}|G_i = k] \tag{1.2}$$

Imposing additional conditions on the coefficients yields alternative identification approaches. With two groups, difference-in-differences specializes the above by fixing $W_2 = 1$, in which case (1.1) holds with $c = W_1$. Another case that fits into this framework is triple differences, which corresponds to $K = 4$, $W_2 = W_3 = 1$, and $W_4 = -1$. This approach is applicable when the four groups correspond to different combinations of two binary characteristics, and only individuals with one of the four combinations are treated. For example, suppose individuals are drawn from two regions $A$ and $B$ and only individuals that satisfy eligibility criteria and live in region $A$ are treated. Then Group 1 may consist of eligible individuals in region $A$, Group 2 of eligible individuals in region $B$, group 3 of ineligible individuals in region $A$, and group 4 of ineligible individuals in region $B$. In this case the condition reduces to the following for some $c$.

$$\big(E[Y_{i,t}^{(0)}|G_i = 1] - E[Y_{i,t}^{(0)}|G_i = 2]\big) - \big(E[Y_{i,t}^{(0)}|G_i = 3] - E[Y_{i,t}^{(0)}|G_i = 4]\big) = c$$

Under Assumptions 1, 2, and (1.2), treatment effects are identified so long as there are unique coefficients $W_1$, $W_2$, ..., $W_K$ that satisfy (1.2) in all pre-treatment periods, subject to any additional a priori constraints on these parameters. The coefficients can be estimated by weighted least squares:

$$\{\hat{W}_k\}_{k=1}^{K} = \underset{\{W_k\}_{k=1}^{K} \in \mathcal{W}}{\arg\min} \sum_{t=1}^{t^*-1} \alpha_t (\bar{Y}_{1,t} - W_1 - \sum_{k=2}^{K} W_k \bar{Y}_{k,t})^2$$

Having estimated the parameters $\{W_k\}_{k=1}^{K}$, the average treatment effect for $t \geq t^*$ may be estimated by plugging-in the coefficient estimates and observed

average outcomes into the linear model:

$$\hat{\tau}_t = \bar{Y}_{1,t} - \hat{W}_1 - \sum_{k=2}^{K} \hat{W}_k \bar{Y}_{k,t}$$

The parameter space $\mathcal{W}$ may incorporate constraints. For example, if $K = 2$ then constraining $W_2 = 1$ yields precisely the diff-in-diff estimator specified earlier in this section. A number of variations on the above are available in the literature. Abadie & Gardeazabal (2003) suggest a more general method for estimating the parameters $\{W_k\}_{k=1}^{K}$. In place of the mean outcomes in the objective they use vectors of group-specific covariates which may include average outcomes. Abadie & Gardeazabal (2003) constrain the optimization problem so that $W_1 = 0$ and the remaining coefficients are positive and sum to 1. Doudchenko & Imbens (2016) suggest adding an elastic net penalty to the synthetic control objective.

A number of methods exist for selecting the weights $\{\alpha_t\}_{t=1}^{t^*-1}$. For example, Abadie & Gardeazabal (2003) propose that the weights be chosen to minimize the objective above subject to the constraint that they are weakly positive and sum to 1 and Abadie *et al.* (2015) suggest choosing $\{\alpha_t\}_{t=1}^{t^*-1}$ by cross-validation.

## 1.2 Consequences of the Diff-in-Diff Assumption

In the settings that we consider in this paper, the parallel trends assumption (1.1) can be highly problematic. We consider cases in which the outcome is a binary indicator that an individual has reached an absorbing state by a given period. Therefore, the group mean outcome is the share of individuals in the group who have reached the absorbing state. For reasons that we describe below, these shares tend to converge over time, regardless of treatment. That is, the magnitude of the difference in group-mean outcomes tends to decrease, even if treatment has no effect. Standard diff-in-diff would erroneously interpret such a decrease as evidence of a treatment effect, leading to biased and inconsistent causal estimates. Note that this same problem persists if one performs diff-in-diff using changes in outcomes $\Delta Y_{i,t} := Y_{i,t} - Y_{i,t-1}$ in place of the level outcome $Y_{i,t}$. If mean outcomes converge then so do changes in mean outcomes.

Perhaps the most striking failure of parallel trends occurs when the shares in the absorbing state grow too close to 1. Recall that the shares must be bounded above by 1 and are weakly increasing over time. Consider the schools example and suppose that at some time prior to $t^*$, 60% of students in school 1 have

passed and 40% in school 2, and by some time after $t^*$, strictly more than 80% of students in school 2 have passed. Under the assumption of counterfactual trends in mean outcomes, we must immediately conclude that there is a negative treatment effect, even before observing the post-treatment outcomes in the treated school. This is because no more than 100% of students in school 1 can pass the test, and so the gap in mean outcomes must be strictly lower in the post-treatment period than the 20% gap observed in the pre-treatment period.

The example above is the result of a ceiling parallel trends places on the shares of individuals who can reach the absorbing state. Given that the shares are bounded above by 1, it follows from (1.1) that the counterfactual share who reach the absorbing state can never exceed $1 + c$ in school 1, and $1 - c$ in school 2. Similar restrictions hold under other linearity assumptions of the form in (1.2).

Convergence of mean outcomes is likely to arise even in cases where the shares do not reach the ceiling $1 \pm c$. Consider that as the share of the population who have reached the absorbing state increases, there are fewer individuals left to enter the absorbing state. This means that if a large share of individuals have reached the absorbing state, any further increases in mean outcomes will be small compared to the rate at which the remaining individuals reach that state.

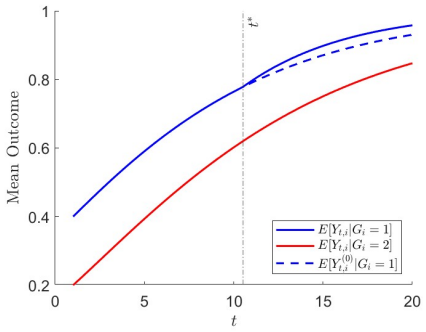
To demonstrate, consider the schools example and suppose that in school 1, 80% have passed the test by day $t$ and 60% in school 2. Thus the share who have not passed by day $t$ is twice as high in school 2 as in school 1. As such, for the difference in shares to remain consistent into the next period, students in school 1 who have not passed by day $t$ would need to be twice as likely to pass on that day as those in school 2. This effect strengthens as the shares who have reached the absorbing state grow. Suppose differences remain fixed and at some later period $t'$, 90% have passed in school 1 and 70% in school 2. For differences to remain fixed, the probability that a student in school 1 who has not yet passed by day $t'$ passes that day must now be three times higher than for school 2. We formalize this point at the end of this section by showing that the parallel assumption implies divergence of the counterfactual 'hazard rates', that is, the rates at which individuals in the two groups who have not yet reached the absorbing state reach that state.

For a graphical illustration, observe Figure 1.1(a) which shows population mean factual and counterfactual outcomes under a duration model specified in the supplementary appendices. We use this same model for our simulation ex-

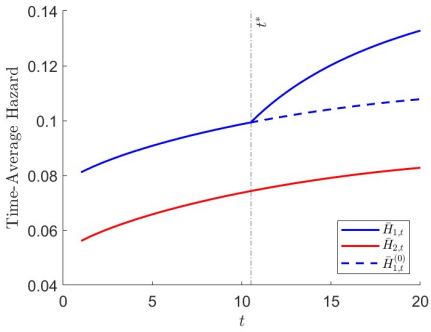
ercises. Factual mean outcomes, i.e., shares of the population that have reached the absorbing state, are plotted over time by solid lines, blue for Group 1 and red for Group 2. The mean outcome for Group 1 under a counterfactual of no treatment, is plotted by the dashed blue line.

Figure 1.1: Deviations from Parallel Trends

(a) Mean Outcomes




(b) Time-Average Hazards




In the figure, we see that mean outcomes for Group 1 are greater than for Group 2 in the initial period. Due to the duration nature of the setting, the difference decreases over time under the counterfactual of no intervention. In this example, the convergence is of a sufficient magnitude that the average difference in the observable factual outcomes is smaller over the post-treatment period than in the pre-treatment period. Therefore, in expectation standard diff-in-diff will estimate a negative treatment effect. However, the true treatment effect is positive, as seen by the the positive gap between the solid and dashed blue lines. In fact, in this model the bias of standard diff-in-diff is roughly four times the value of the true treatment effect. For simulation evidence of the poor performance of standard diff-in-diff under this model (and good performance of our proposed alternatives) see the supplementary appendices.

The discussion above suggests that whatever assumption we use to identify and estimate causal effects should be compatible with the type of convergence shown in Figure 1.1(a). In addition, we should avoid assumptions that imply divergence between the hazard rates. As an alternative, we suggest analogues of the diff-in-diff assumption that apply directly to the hazard rates. Instead of assuming parallel trends in the shares who have reached the absorbing state, we instead assume parallel trends (or some other fixed linear relationship) in

the hazard rates. Parallel trends in the hazard rates is consistent with the convergence of counterfactual mean outcomes. Indeed, the assumption holds in Figure 1.1(a) and thus the convergence in the figure does not result in misleading inference under this assumption.

Finally, we provide a brief formal argument that shows the parallel trends assumption implies divergence of counterfactual hazard rates. First we must define these objects, for concreteness we do so in the context of the schools example. Let $\Delta > 0$ and consider the probability that, under the counterfactual of no intervention, an individual in school $k$ passes the exam between times $t$ and $t + \Delta$, conditional on having not passed prior to $t$. If we scale this probability by $1/\Delta$, then the limit as the increment $\Delta$ shrinks to zero is the counterfactual hazard rate for school $k$. We denote the counterfactual hazard rate by $h_k^{(0)}(t)$ where the superscript indicates that this is the hazard rate for the outcomes under the no-intervention counterfactual. It is defined formally as follows:

$$h_k^{(0)}(t) = \lim_{\Delta \downarrow 0} \frac{P(Y_{i,t+\Delta}^{(0)} = 1 | Y_{i,t}^{(0)} = 0, G_i = k)}{\Delta}$$

We can define the factual hazard rate $h_k(t)$ analogously by replacing counterfactual outcomes with factual outcomes in the definition. In the schools example, the counterfactual hazard rate measures the rate at which students in school $k$ who have not passed the exam, pass the exam.

Suppose that $c$ is strictly positive and that for all $t$, we have $h_2^{(0)}(t) > 0$ and $E[Y_{i,t}^{(0)} | G_i = 2] < 1 - c$. Then under the diff-in-diff assumption (1.1) we have that for any $t$:

$$\frac{h_1^{(0)}(t)}{h_2^{(0)}(t)} = \frac{1 - E[Y_{i,t}^{(0)} | G_i = 2]}{1 - E[Y_{i,t}^{(0)} | G_i = 2] - c}$$

The right hand side grows with $t$ and it goes to infinity as the mean counterfactual outcome for Group 2 approaches the upper bound $1 - c$. So as the share in Group 2 who have reached the absorbing state approaches this upper bound, the ratio of the hazard rates grows to infinity.

In the context of the schools example, as $t$ increases, the probability that a student who has not passed in school 1 passes within the next $\Delta$ increment of time, must grow infinitely large compared to the same quantity for a student in school 2. If $c$ is negative and $E[Y_{i,t}^{(0)} | G_i = 1] < 1 + c$, then the hazard ratio shrinks to zero as $E[Y_{i,t}^{(0)} | G_i = 1]$ approaches $1 + c$.

A related point is made by Wu & Wen (2022) who show formally that the pa-

rameter estimated by diff-in-diff must evolve over time when data are generated by a hazard model (apart from in certain degenerate cases).

The analysis above is concerning because it suggests that the hazard rates exhibit aberrant behavior as the proportion of individuals who have entered into an absorbing state grows large. The hazard rate is a primitive building block of duration models, and it has a clear structural interpretation. If the difference-in-differences assumption implies counter-intuitive and surprising behavior of the hazard rates, then we should reject the assumption a priori.

## 2 Model and Identification

In light of the discussion above, we replace the standard parallel trends assumption, or a more general restriction of a fixed linear relationship between mean outcomes, with an analogous condition that applies directly to the hazard rates. By restricting hazards rather than mean outcomes, we ensure our modeling restrictions are consistent with the convergence of mean outcomes between groups. As such, the convergence of mean outcomes need not bias causal estimates based on our assumptions, in contrast to estimates from standard diff-in-diff.

Throughout, we continue to take $k = 1$ to be the unique treated group. Our key identifying assumption is that there exists a fixed linear relationship between the counterfactual hazard rates of the treated and untreated groups. Formally, we consider assumptions of the following form:

$$h_1^{(0)}(t) = W_1 + \sum_{k=2}^{K} W_k h_k^{(0)}(t) \tag{2.1}$$

Imposing a priori constraints on the coefficients $W_1, .., W_K$ yields a range of alternative modeling assumptions. For example, in the two-group case, setting $W_2 = 1$ we get a duration analogue of standard difference-in-differences.

$$h_1^{(0)}(t) - h_2^{(0)}(t) = c \tag{2.2}$$

The condition (2.2) states that the level difference in hazard functions between groups 1 and 2 is constant over time. The condition allows for common shocks that impact the counterfactual hazards of individuals in both the treated and untreated groups, so long as these shocks induce parallel movements in the hazard rates. Moreover, the condition allows for the possibility that the haz-

ard rates evolve over time, increasing or decreasing with the time spent in the $Y_{i,t} = 0$ state.

If we instead assume $W_1 = 0$ and leave $W_2$ unrestricted then we recover the proportional hazard restriction as in the models of Hunt (1995) and Wu & Wen (2022):

$$h_1^{(0)}(t)/h_2^{(0)}(t) = c \tag{2.3}$$

The above restriction admits shocks to the counterfactual hazard rates and allows for the possibility that the hazards evolve over time. In this case the restriction is that the shocks and time trends leave the difference in the log counterfactual hazard rates unchanged.

The general linear restriction also nests a hazard analogue of the triple differences specification in Section 1.1.

$$[h_1^{(0)}(t) - h_2^{(0)}(t)] - [h_3^{(0)}(t) - h_4^{(0)}(t)] = c$$

We show that linear restrictions on the hazard rates allow for simple and transparent identification and estimation methods that are close analogues of the corresponding diff-in-diff and related methods.

Note that the linear relationships described above are on group-level hazard rates. We do not begin with an individual-level model. This has the advantage that we do not need to take a stance on spill-overs between different individuals in the same group, and more generally it allows us to be agnostic about distributions of individual heterogeneity. It also allows us to use repeated cross-sections and to accommodate incomplete spells. On the other hand, one should not conclude that individual-level modeling requires strictly stronger conditions: our assumptions do not follow from individual-level hazard models with unobserved heterogeneity apart from in special cases. Instead, our assumptions provide a practical non-parametric alternative to individual-level modeling.

## 2.1 Hazard Diff-in-Diff

We establish identification of causal effects under linear restrictions on group-specific hazard rates of the form in (2.1). This includes the special cases of fixed level differences (2.2) and the fixed ratio restriction (2.3). What follows is a sketch of the identification argument.

The first step is to rewrite the condition (2.1) in terms of the time-average counterfactual hazard functions. Time-average hazards can be written in terms

of mean outcomes and are directly identified from the data. We denote the time-average counterfactual hazard function for Group $k$ at time $t$ by $\bar{H}^{(0)}_{k,t}$, and the time-average factual hazard by $\bar{H}_{k,t}$. The time-average hazard is, as the name suggests, the average over time of the hazard rate. Formally, these objects are defined as follows:

$$\bar{H}^{(0)}_{k,t} = \frac{1}{t-1}\int_1^t h^{(0)}_k(s)ds, \quad \bar{H}_{k,t} = \frac{1}{t-1}\int_1^t h_k(s)ds$$

From the condition (2.2) we obtain the equality below:

$$\bar{H}^{(0)}_{1,t} - \bar{H}^{(0)}_{2,t} = c \tag{2.4}$$

Alternatively, under the assumption of a fixed ratio between the group-specific hazards (2.3) we get:

$$\bar{H}^{(0)}_{1,t} / \bar{H}^{(0)}_{2,t} = c \tag{2.5}$$

The two examples above are special cases of the following relationship which follows from the general linear restrictions (2.1):

$$\bar{H}^{(0)}_{1,t} = W_1 + \sum_{k=2}^{K} W_k \bar{H}^{(0)}_{k,t} \tag{2.6}$$

Key to our analysis is the observation that the time-average of the counterfactual hazard rate can be written in terms of mean outcomes. This follows from a standard result in duration analysis that the 'cumulative hazard function' is equal to minus the log of the 'survivor function'. For $t > 1$ we have:

$$\bar{H}^{(0)}_{k,t} = \frac{1}{t-1} ln\left(\frac{1 - E[Y^{(0)}_{i,1}|G_i = k]}{1 - E[Y^{(0)}_{i,t}|G_i = k]}\right) \tag{2.7}$$

$$\bar{H}_{k,t} = \frac{1}{t-1} ln\left(\frac{1 - E[Y_{i,1}|G_i = k]}{1 - E[Y_{i,t}|G_i = k]}\right) \tag{2.8}$$

Note that the time-average factual hazard rate depends only on factual mean outcomes, and therefore it can be directly and straight-forwardly estimated from the observed outcomes. Moreover, under Assumptions 1 and 2, it is equal to the time-average counterfactual hazard when $k \neq 1$, or when $k = 1$ and $t < t^*$.

Putting these facts together, we can attain identification of causal effects. To illustrate, consider the case in which we assume a fixed level difference in

the hazard rates as in (2.2). This assumption implies that counterfactual time-average hazards exhibit parallel trends (2.4). The counterfactual time-average hazards can then be written in terms of counterfactual mean outcomes. This is illustrated graphically in Figure 1.1. In Figure 1.1(a), we plot group-specific mean factual and counterfactual mean outcomes over time from an underlying duration model that satisfies parallel trends in the hazard rates (2.2). As we discuss in Section 1.2, the solid blue curve represents the factual Group 1 mean outcome, the solid red line is the Group 2 mean outcome, and the dashed blue line, the counterfactual Group 1 mean outcome. Figure 1.1(b) shows the corresponding time-average hazards in the same colors and styles.

The curves in 1.1(b) are related to those in 1.1(a) by the formulas (2.7) and (2.8). By transforming the solid curves in 1.1(a) we obtain the factual time-average hazard in 1.1(b). Note that while parallel trends clearly fails in 1.1(a), the counterfactual time-average hazards are parallel. In order to identify average treatment effects, we perform difference-in-differences on the time-average hazards in Figure 1.1(b). Thus we impute the time-average counterfactual hazards for the treated group in the post-treatment period (the dashed blue curve). We can then recover the counterfactual mean outcomes in the post-treatment period by inverting (2.7), and thus we identify average treatment effects. That is, we apply the inverse to the dashed blue curve in 1.1(b) and obtain the dashed blue curve in 1.1(a), which plots the counterfactual mean outcomes for the treated group.

The identification argument above generalizes to other linear restrictions of the form in (2.1), including the fixed ratio restriction. Linear restrictions on group-specific counterfactual hazard rates imply linear restrictions on counterfactual time-average hazards. These are equivalent to linear restrictions on logarithmically transformed counterfactual mean outcomes. We identify the linear relationship using pre-treatment data, extrapolate to post treatment, and finally invert the logarithmic transformation. Thus we obtain counterfactual post-treatment mean outcomes for the treated group. This identification result is formalized in Theorem 1 below.

**Theorem 1.** *Suppose Assumptions 1 and 2 and (2.1) hold and define* $\bar{H}_{k,t}$ *as in (2.8). Then for every* $1 < t < t^*$*:*

$$\bar{H}_{1,t} = W_1 + \sum_{k=2}^{K} W_k \bar{H}_{k,t} \tag{2.9}$$

*And for any $t^* \leq t$ the counterfactual mean outcome is given by:*

$$E[Y_{i,t}^{(0)}|G_i = 1] = 1 - \big(1 - E[Y_{i,1}|G_i = 1]\big) exp\big(-(t-1)(W_1 + \sum_{k=2}^{K} W_k \bar{H}_{k,t})\big) \quad (2.10)$$

*Thus if (2.9) has a unique solution, then both $E[Y_{i,t}^{(0)}|G_i = 1]$ and $\tau_t$ are identified for all $t$.*

Theorem 1 identifies treatment effects when the equations (2.9) have a unique solution subject to the a priori constraints on the coefficients (that is, a unique solution that satisfies $\{W_1, W_2, ...\} \in \mathcal{W}$). Note that we must have at least three periods of data. Under the fixed level difference restriction, uniqueness always holds. For the fixed ratio constraint it suffices that the mean outcome for Group 2 is not constant over the whole pre-treatment period.

Suppose $T = t^* = 3$ and consider the fixed level differences assumption (2.2) in which $K = 2$, $W_1 = c$, and $W_2 = 1$. In this case, Theorem 1 identifies $c$ from

$$c = ln\left(\frac{1 - E[Y_{i,1}|G_i = 1]}{1 - E[Y_{i,2}|G_i = 1]}\right) - ln\left(\frac{1 - E[Y_{i,1}|G_i = 2]}{1 - E[Y_{i,2}|G_i = 2]}\right),$$

and counterfactual outcomes are then given by

$$E[Y_{i,3}^{(0)}|G_i = 1] = 1 - \big(1 - E[Y_{i,1}|G_i = 1]\big) exp\big(-2(c + E[Y_{i,3}^{(0)}|G_i = 2])\big).$$

If we do not constrain the coefficients $\{W_k\}_{k=1}^{K}$, then a necessary condition for uniqueness is that $t^* > K$ and uniqueness is generic whenever this holds. Given a sufficient number of pre-treatment periods, the parameters in (2.9) are over-identified. This suggests we can test the identifying restrictions using say placebo tests or, in the case of fixed level differences a test for parallel trends.

## 2.2 Incorporating Covariates

In difference-in-differences analysis it is common to adjust for covariates. One motivation is that parallel trends may hold within each stratum of the covariates but not in the aggregate (see e.g., Abadie (2005)). Of course, one could simply apply our our methods separately within each covariate stratum and aggregate to obtain average causal effects, but this may be impractical, particularly if covariates are continuous or discrete with many categories. For the case of $k = 2$, we suggest a simple means of adjustment for time-invariant covariates

based on a form of propensity-score weighting and provide a formal justification.

Before we provide a formal justification, let us first describe the reweighting approach. We define a weight function $\omega$ on the support of the covariates. In the case of continuous covariates one can replace the conditional probabilities with conditional probability densities and obtain a weighting of the kind introduced in DiNardo *et al.* (1996).

$$\omega(x) = \frac{P(X_i = x|Y_{i,1} = 0, G_i = 1)}{P(X_i = x|Y_{i,1} = 0, G_i = 2)} \tag{2.11}$$

By weighting individuals in the untreated group by $\omega(X_i)$ we down-weight those with values of the covariates that are more prevalent among survivors in Group 2 than Group 1 in the initial period, and up-weight those whose values were less prevalent. If the covariates have the same distribution across groups in the initial period, then the weight function reduces to $\omega(x) = 1$.

The weighting function can be written equivalently in terms of a type of propensity score (Rosenbaum & Rubin (1983)). Let $p(x)$ be the probability that an individual is in the treated group given they have covariate values $x$ and they are not in the absorbing state in the initial period. Formally, $p(x)$ is defined as follows:

$$p(x) = P(G_i = 1|Y_{i,1} = 0, X_i = x)$$

Then applying Bayes' rule we obtain the following which is well-defined if $p(x) < 1$:

$$\omega(x) = \frac{p(x)P(G_i = 2|Y_{i,t} = 0)}{\big(1 - p(x)\big)P(G_i = 1|Y_{i,t} = 0)} \tag{2.12}$$

With the weighting function in hand we can define the weighted time-average hazard $\bar{H}^*_{2,t}$ for the untreated group.

$$\bar{H}^*_{2,t} = \frac{1}{t-1} ln\left(\frac{1 - E[Y_{i,1}|G_i = 2]}{E[\omega(X_i)(1 - Y_{i,t})|G_i = 2]}\right) \tag{2.13}$$

For the above to be well-defined we need $P(p(X_i) < 1|Y_{i,1} = 0) = 1$.[2] In estimation, we use the estimates of the weighted time-average hazard in place of the unweighted time-average hazard.

Our covariate adjustment approach differs from that of Wooldridge (2023),

[2]Here we understand $\omega(X_i)(1 - Y_{i,t}) = 0$ if $Y_{i,t} = 1$, even if $\omega(X_i)$ is undefined due to division by zero.

who presents a means of incorporating covariates in non-linear diff-in-diff models. The approach in Wooldridge (2023) requires a functional form assumption for conditional mean outcomes. This is, in our view, less well-suited to duration settings. The reason for this is that with duration data it is more natural to begin with a model for hazard rates than a model for conditional mean outcomes. Indeed, a particular functional form for conditional mean outcomes may not be compatible with any underlying hazard model. Modeling covariate-specific hazard rates and deriving the resulting functional-form for conditional mean outcomes may be difficult or intractable, apart from in special cases. Moreover, this approach is contrary to the general character of our methods which are non-parametric and allow for transparent visual inspection of the underlying parallel trends assumption (or assumption of a fixed linear relationship). In contrast, our approach does not require us to specify a parametric model and it retains the simplicity and transparency of our methods without covariate adjustment: we reweight individuals when calculating group mean outcomes, and the rest of our analysis goes through unchanged.

The use of the weighted time-average hazards can be justified in two ways. Firstly, the a priori assumption of a fixed linear relationship (2.2) or (2.3) may be more plausible if the Group 1 and Group 2 populations have the same distribution of observed covariates. The weighting rebalances the Group 2 population so that this is the case. We can define the hazard rate for this rebalanced Group 2 as:

$$h_2^*(t) = \lim_{\Delta \downarrow 0} \frac{E\big[\omega(X_i) Y_{i,t+\Delta} | Y_{i,t} = 0, G_i = 2\big]}{\Delta}$$

To interpret the above, consider the case in which $X_i$ is discrete. Suppose we add or remove individuals at random from the Group 2 population with some probability that depends on their covariate values. Thus we obtain a new population Group $2^*$. Suppose we choose these probabilities so that the distribution of covariates is the same in the Group $2^*$ population as in Group 1. Then $h_2^*(t)$ is the hazard rate for Group $2^*$. The time-average hazard for Group $2^*$ is then $\bar{H}_{2,t}^*$. Thus, if we assume a fixed linear relationship between the hazard rates for the covariate-balanced groups, $h_1(t)$ and $h_2^*(t)$, rather than between $h_1(t)$ and $h_2(t)$, then our earlier identification arguments go through with $\bar{H}_{2,t}^*$ in place of $\bar{H}_{2,t}$.

A second and more formal justification is as follows. Again, suppose that there are two groups $k = 1, 2$ and for each individual $i$ we observe individual-specific and time-invariant covariates $X_i$. We can define group-specific coun-

terfactual hazard rates within each stratum of the covariates as follows. Let $h_k^{(0)}(t;x)$ denote the Group $k$ counterfactual hazard rate at time $t$ for the subpopulation for whom $X_i$ is equal to $x$.

$$h_k^{(0)}(t;x) = \lim_{\Delta \downarrow 0} \frac{P(Y_{i,t+\Delta}^{(0)} = 1 | Y_{i,t}^{(0)} = 0, G_i = k, X_i = x)}{\Delta}$$

Suppose the hazard diff-in-diff assumption holds within each stratum of the covariates and moreover, that the level difference does not depend on the stratum. That is, the following equation holds for each $x$ in the support of the covariates:

$$h_1^{(0)}(t;x) - h_2^{(0)}(t;x) = c \tag{2.14}$$

To motivate the above, consider that if treatment is independent of potential outcomes conditional on covariates (a condition sometimes known as 'unconfoundedness' or 'ignorability'), then the above holds with $c = 0$. Note that in this special case, the ratio of covariate-specific hazards is also fixed over time. However, the condition also allows for selection into treatment so long as the resulting level difference in hazards is constant.

If there is imbalance in the distribution of the covariates between the groups, then the condition (2.14) does not guarantee parallel trends for the marginal hazards (2.2), even when $c = 0$. However, by applying a weighting scheme we can recover covariate balance and identify treatment effects. Theorem 2 shows that an analogous result to Theorem 1 specialized to fixed level differences holds under the covariate-specific parallel trends assumption (2.14).

**Theorem 2.** *Suppose Assumptions 1, 2 and (2.14) hold, and additionally that* $P(p(X_i) < 1 | Y_{i,1} = 0) = 1$. *Define* $\bar{H}_{k,t}$ *as in (2.8) and* $\bar{H}^*_{2,t}$ *as in (2.13). Then for every* $1 < t < t^*$*:*

$$c = \bar{H}_{1,t} - \bar{H}^*_{2,t} \tag{2.15}$$

*And for any* $t^* \leq t$ *the counterfactual mean outcome is given by:*

$$E[Y_{i,t}^{(0)} | G_i = 1] = 1 - \big(1 - E[Y_{i,1} | G_i = 1]\big) exp\big( - (t-1)(c + \bar{H}^*_{2,t})\big) \tag{2.16}$$

The condition (2.14) allows for differential trends based on $x$, but it restricts the level difference $h_1^{(0)}(t;x) - h_2^{(0)}(t;x)$ to be independent of $x$. In standard diff-in-diff the converse is sometimes assumed. That is, one assumes that for all

periods $s$, $t$, and groups $k$, the following is the same for all $x$:

$$E[Y_{i,t}^{(0)}|G_i = k, X_i = x] - E[Y_{i,s}^{(0)}|G_i = k, X_i = x]$$

Thus trends do not differ between covariate strata. However, this allows for the possibility that level differences depend on $x$, that is

$$E[Y_{i,t}^{(0)}|G_i = 1, X_i = x] = E[Y_{i,t}^{(0)}|G_i = 2, X_i = x] = c(x).$$

One problem with this assumption is that it is not guaranteed to hold even in the case of unconfoundedness. That is, if potential outcomes are independent of $G_i$ conditional on $X_i$, the assumption of parallel trends may not hold. By contrast, our condition must hold under unconfoundedness.

## 2.3 Practical Considerations for Specification Choice

With sufficiently many time periods, one can identify causal effects under restrictions of the form (2.1) without any constraints on the parameters. In the two-group case, this means that researchers need not choose between the fixed level differences restriction (2.2) and fixed ratio constraint (2.3), but can rather identify effects under a more general restriction that nests both of these. However, in-line with standard empirical practice in diff-in-diff, researchers may wish to choose one of these two specifications. In the special case in which group-specific hazard rates are exactly equal, i.e., $c = 0$ in (2.2), then both the assumption of fixed-level differences and fixed ratios hold.[3]

In our empirical application we apply fixed level differences. Our motivation for this is two-fold. Firstly, this increases the familiarity of our approach to researchers who are used to standard diff-in-diff. Secondly, the covariate adjustment in Section 2.2 has a clearer formal justification in terms of fixed-level differences. In our application, the difference in time-average hazards is relatively small, $\approx 10\%$ of the level of the smallest time-average hazard, and as such, fixed level-differences and fixed ratios should yield similar results. It should be noted however that the fixed ratio assumption ensures that imputed time-average hazard rates are positive by construction, and thus it may be the more natural choice in settings in which time-average hazard rates differ substantially and trend towards zero.

[3] We thank an anonymous referee for pointing this out.

An important practical consideration in applying our method (and indeed related methods) is the definition of the time period $t$. In particular, whether $t$ reflects the actual calendar date, or $t$ represents the time that has elapsed since an individual began a particular spell, say an unemployment spell.

For example, in our empirical application, we take $t$ to represent the latter. In this setting, $Y_{i,t}$ is an indicator for whether individual $i$ has exited their unemployment spell within $t$ weeks of the start of that spell. We define $t$ in this way because in this particular setting, those who are treated receive treatment once they have been unemployed for a particular length of time, regardless of when that spell began.

While many empirical settings may share this feature, in other cases letting $t$ represent the calendar date may be more appropriate. For example, suppose we wish to assess the impact of divorce-law reform on marriage duration (e.g., Friedberg (1998), Gruber (2004), and Wolfers (2006)) by comparing rates of divorce in two states, one in which a reform is enacted, and one in which it is not. Suppose the data begin on October $25^{th}$ 1991 and the sample consists of individuals whose marriages began prior to that date. In this case, the reform is enacted on a particular date, and thus $t$ should be defined as the length of time that has elapsed since October $25^{th}$ 1991. This ensures that individuals in the treated state all receive treatment at the same period $t$.

If $t$ represents the time since the beginning of a spell, then it may be important to control for the start date of the spell. In the unemployment example, there may be seasonal differences in the job-finding rate, and thus trends in the hazard rate may differ by the date at which an individual becomes unemployed. Conversely, in the marriage example, in which $t$ is the calendar date, trends may differ by the time since the start of a marriage. Thus it may be prudent to control for either the date at the start of a spell or the length of the spell up to the initial period, depending on the definition of $t$. This emphasizes the importance of the covariate balancing in Section 2.2.

Finally, inference in duration settings may be complicated by random right-censoring. This refers to the case in which some individuals leave a study or are otherwise absent from the data after a certain period, prior to reaching the absorbing state. More formally, let $C_{i,t}$ be an indicator that individual $i$ has left the study prior to time $t$. Then instead of $Y_i$, we observe $\tilde{Y}_{i,t} = (1 - C_{i,t})Y_{i,t}$, and an indicator $c_i$ that equals 1 if and only if an individual is censored (i.e., if $C_{i,s} = 1$ and $Y_{i,s} = 0$ for some $s \leq t$). Censoring complicates identification and

estimation of the survivor function (given by $\frac{1-E[Y_{i,t}|G_i=k]}{1-E[Y_{i,1}|G_i=k]}$) and its covariate weighted counterpart. Fortunately, numerous techniques exist to adjust for random right-censoring in estimation of the survival function. In the case in which censoring is random ($C_{i,t}$ is independent of $Y_{i,t}$ given $G_i$) the Kaplan-Meier (Kaplan & Meier (1958)) estimator of the survival function is standard. If censoring is not independent, other techniques are available, see for example the review in Klein & Moeschberger (2003).

# 3 Estimation and Inference

The identification results in the previous section motivate plug-in estimates of counterfactual mean outcomes and treatment effects. We first define an estimate $\hat{H}_{k,t}$ of the time-average hazard $\bar{H}_{k,t}$ as follows:

$$\hat{H}_{k,t} = \frac{1}{t-1} ln\left(\frac{1-\bar{Y}_{k,1}}{1-\bar{Y}_{k,t}}\right) \tag{3.1}$$

With the time-average hazard estimates in hand, the first result in Theorem 1 motivates regression estimates of the parameters $\{W_k\}_{k=1}^K$. Using data from the pre-treatment periods we regress the estimate $\hat{H}_{1,t}$ on the estimated time-average hazards of the other groups. A weighted, and possibly constrained, least-squares estimator is given below where the weights are $\alpha_t$ for $t = 2, .., t^*-1$ and sum to unity.

$$\{\hat{W}_k\}_{k=1}^K = \underset{\{W_k\}_{k=1}^K \in \mathcal{W}}{\arg\min} \sum_{t=1}^{t^*-1} \alpha_t (\hat{H}_{1,t} - W_1 - \sum_{k=2}^{K} W_k \hat{H}_{k,t})^2 \tag{3.2}$$

The constraint set $\mathcal{W}$ can incorporate restrictions like positivity of the weights, or that the intercept $W_1$ is equal to zero. Having obtained parameter estimates, the second result in Theorem 1 motivates the following estimate of the time-$t$ average treatment effect:

$$\hat{\tau}_t = \bar{Y}_{1,t} - 1 + \big(1 - \bar{Y}_{1,1}\big) exp\big( - (t-1)(\hat{W}_1 + \sum_{k=2}^{K} \hat{W}_k \hat{H}_{k,t})\big)$$

In our practical application, we use equal weights (i.e., $\alpha_2, \alpha_3, ...$ all take the same value). More generally, they may be chosen either to a) place greater emphasis on those periods that are closer to the intervention, or b) minimize

the asymptotic variance of the estimates.

In hazard diff-in-diff, the estimator reduces to the formula below, where $\hat{c}$ is an estimate of $c$. This corresponds to the imputation approach in Wooldridge (2023).

$$\hat{c} = \sum_{t=2}^{t^*-1} \alpha_t (\hat{H}_{1,t} - \hat{H}_{2,t}) \tag{3.3}$$

$$\hat{\tau}_t = \bar{Y}_{1,t} - 1 + \big(1 - \bar{Y}_{1,1}\big) exp\big( - (t-1)(\hat{c} + \hat{H}_{2,t})\big) \tag{3.4}$$

In the proportional hazard model, we constrain $W_1 = 0$ and obtain the estimate of $W_2$, which we again denote by $\hat{c}$, and a corresponding treatment effect estimate as below.

$$\hat{c} = \sum_{t=2}^{t^*-1} \alpha_t \hat{H}_{1,t} \hat{H}_{2,t} / \sum_{t=2}^{t^*-1} \alpha_t \hat{H}_{1,t}^2$$

$$\hat{\tau}_t = \bar{Y}_{1,t} - 1 + \big(1 - \bar{Y}_{1,1}\big) exp\big( - (t-1)\hat{c}\hat{H}_{2,t}\big)$$

## 3.1 Estimation with Covariate Adjustment

In order to estimate treatment effects using the weighting scheme introduced in Section 2.3, one must first estimate the weighting function $\omega$. We suggest two alternatives.

The first estimator is based on (2.11) and is appropriate for discrete covariates. In this case we simply replace the conditional probabilities in the formula (2.11) with empirical frequencies. The estimate $\hat{\omega}(x)$ is defined for all $x$ for which there exists some individual $i$ in Group 2 for whom $X_i = x$ and $Y_{i,1} = 0$, and is given below:

$$\hat{\omega}(x) = \frac{n_2(1 - \bar{Y}_{2,1}) \sum_{i=1}^n 1\{G_i = 1\} 1\{X_i = x\}(1 - Y_{i,1})}{n_1(1 - \bar{Y}_{1,1}) \sum_{i=1}^n 1\{G_i = 2\} 1\{X_i = x\}(1 - Y_{i,1})} \tag{3.5}$$

If some covariates are continuous or discrete with many support points, then we suggest an approach based on the propensity score formulation (2.12). Let $\hat{p}(x)$ be an estimate of $p(x)$. One could obtain such an estimate using say, logistic regression of $1\{G_i = 1\}$ on $X_i$ using the sub-sample of individuals for whom $Y_{i,1} = 0$. We can then estimate the weights as follows:

$$\hat{\omega}(x) = \frac{\hat{p}(x)(1 - \bar{Y}_{2,1}) n_2}{\big(1 - \hat{p}(x)\big)(1 - \bar{Y}_{1,1}) n_1}$$

Having obtained an estimate of the weight function, we use the following estimate of the weighted time-average hazard function in (3.3) and (3.4).

$$\hat{H}_{2,t} = \frac{1}{t-1} ln\left( \frac{1 - \bar{Y}_{2,1}}{\frac{1}{n_2}\sum_{i=1}^{n} 1\{G_i = 2\}\hat{\omega}(X_i)(1 - Y_{i,t})} \right)$$

## 3.2 Bootstrap Inference

Bertrand *et al.* (2004) propose the block bootstrap (Efron & Tibshirani (1994)) for conducting inference in vanilla diff-in-diff. The asymptotic validity of the procedure rests on an assumption that the outcome histories of different individuals are independent. However, the outcomes of a given individual may exhibit arbitrary dependence over time. We suggest the use of the block bootstrap for inference in the duration settings in this paper. Implementation follows the same steps as for standard diff-in-diff with the distinction that the bootstrap samples are used to construct our estimator of the treatment effect rather than the usual diff-in-diff estimator.

To carry out block bootstrap inference, one independently resamples individuals uniformly with replacement and forms a new sample using the complete series of outcomes and covariates for each individual resampled. For example, if individual $i$ is sampled in the $b$-th bootstrap iteration, then that bootstrap sample will contain an individual whose outcome history is $Y_{i,1}, Y_{i,2}, ..., Y_{i,T}$.

Having obtained block bootstrap samples, one may then evaluate bootstrap standard errors as well as pointwise and uniform confidence bands in the usual way. In particular, for each bootstrap sample $b = 1, ..., B$, one computes the estimate $\hat{\tau}_t$ using the bootstrap sample in place of the original data, and thus obtains a bootstrap estimate $\hat{\tau}^*_{b,t}$.

The standard deviation $\hat{\sigma}_t$ of $\hat{\tau}^*_{b,t}$ over the bootstrap samples $b = 1, ..., B$ is taken as the standard error for $\hat{\tau}_t$. To form pointwise confidence intervals for $\tau_t$, let $\hat{q}_{1-\alpha,t}$ be the $1-\alpha$-quantile of $|\hat{\tau}^*_{b,t} - \hat{\tau}_t|/\hat{\sigma}_t$. Then a $1-\alpha$-level confidence interval has the form below:

$$CI_{1-\alpha,t} = [\hat{\tau}_t - \hat{q}_{1-\alpha}\hat{\sigma}_t, \hat{\tau}_t + \hat{q}_{1-\alpha}\hat{\sigma}_t]$$

The intervals described above are only designed to achieve correct pointwise coverage. Suppose we form confidence intervals for $\tau_t$ for each period from $t = t^*$ to $T$. Each of these intervals is specified so that it covers the corresponding period's average treatment effect with probability approximately $1-\alpha$. However,

the probability that **every** one of these intervals contains its corresponding period's treatment effect may be much lower.

In order to obtain a desired joint coverage probability, in place of the critical value $\hat{q}_{1-\alpha,t}$ defined above, we instead take $\hat{q}_{1-\alpha,t}$ to be the $1-\alpha$ quantile over $b = 1, ..., B$ of $\max_{t^* \leq s \leq T} |\hat{\tau}^*_{b,s} - \hat{\tau}_s|/\hat{\sigma}_s$ and otherwise form the confidence bands as above.

A more formal description of the pointwise and uniform inference procedures is given in Algorithm 1 in the supplementary appendices.

### 3.3 Specification Testing

The conditions (2.2) and (2.3) each have testable implications. Consider (2.2), if the assumption holds in all periods, then the difference in pre-treatment time-average hazards must be constant. This motivates a test analogous to the test for parallel trends in standard difference in differences. For simplicity we focus on the fixed level differences assumption, however the approach generalizes straight-forwardly to the assumption of a fixed ratio.

To motivate the test, note that from (2.4) and Assumptions 1 and 2, in all periods $t = 2, 3, ..., t^* - 2$ we have:

$$\delta_t = (\bar{H}_{1,t} - \bar{H}_{2,t}) - (\bar{H}_{1,t^*-1} - \bar{H}_{2,t^*-1}) = 0$$

We use this restriction to motivate a Wald test of the null that parallel trends holds in the pre-treatment period. In order to estimate $\delta_t$ we replace each population time-average hazard $\bar{H}_{k,t}$ with the corresponding estimate $\hat{H}_{k,t}$ specified in Section 2.1. We thus obtain an estimate $\hat{\delta}_t$:

$$\hat{\delta}_t = (\hat{H}_{1,t} - \hat{H}_{2,t}) - (\hat{H}_{1,t^*-1} - \hat{H}_{2,t^*-1}) = 0$$

We can then carry out a Wald test. Let $\hat{D} = (\hat{\delta}_1, \hat{\delta}_1, ..., \hat{\delta}_{t^*-2})'$. We can carry out the same calculation on the $b$-th bootstrap sample to obtain a bootstrap analogue $\hat{D}_b$. Let $\bar{D} = \frac{1}{B}\sum_{b=1}^{B} \hat{D}_b$ and $\hat{V}$ be the bootstrap variance-covariance estimator: $\hat{V} = \frac{1}{B}\sum_{b=1}^{B}(\hat{D}_b - \bar{D})(\hat{D}_b - \bar{D})'$.

A Wald statistic for a test of the null hypothesis of parallel trends can then be formed as $\hat{D}'\hat{V}^{-1}\hat{D}$. One can then compare this to the bootstrap distribution of this statistic (treating $\hat{V}$ as fixed) to obtain a test. In particular the p-value

is given by:

$$p = \frac{1}{B}\sum_{b=1}^{B} 1\{\hat{D}'\hat{V}^{-1}\hat{D} \geq \hat{D}_b'\hat{V}^{-1}\hat{D}_b\}$$

A level $1-\alpha$ test of parallel trends then rejects if and only if $p \leq \alpha$.

An alternative approach that allows for some a more visual inspection is to construct uniform (over $t = 2, .., t^* - 1$) bootstrap confidence bands for $\delta_t$, we describe this as Algorithm 2 in the supplementary appendices.

### 3.4 Asymptotic Validity

All of the procedures we have described can be written as generalized method of moments estimators or sequential generalized method of moments estimators. Thus, as the sample size grows (with $K$ and $T$ fixed) standard regularity conditions ensure the consistency of our estimates and asymptotically correct coverage of the bootstrap confidence intervals. These results are well-known and so, following the example of Wooldridge (2023), we omit a formal statement here.

Nonetheless, it is important to note three caveats. First, standard inferential results require identification of the nuisance parameters, which means that there must be unique coefficients $\{W_k\}_{k=1}^K$ that satisfy (2.9) subject to any priori constraints that $\{W_k\}_{k=1}^K \in \mathcal{W}$. Second, if this constraint is an inequality constraint and it binds, then treatment effect estimates may not be differentiable functions of mean outcomes. In this case, asymptotic normality generally fails and the bootstrap does not have correct coverage. Such settings may call for alternative inference procedures of the form described in Fang (2019). Finally, for the estimation with covariate adjustment, regular estimation generally requires that the weights $\omega(X_i)$ be bounded above, in which case the propensity scores must be bounded below away from zero (see Khan & Tamer (2010)).

## 4 Application: The Impact of Unemployment Insurance

We apply our methods to examine the impact of a policy change in Austria on the 1st of August 1989 that increased the generosity of unemployment benefits of eligible individuals. This setting was previously examined by Lalive *et al.* (2006) and we use the data accompanying their paper. Lalive *et al.* (2006) use

the data to estimate a piece-wise constant proportional hazards model with a linear index that interacts eligibility for various aspects of the policy with an indicator that the policy change has occurred. We instead employ a cross-cohort study using our methods.

In our view, the primary benefit of our analysis over the original study is its simplicity and transparency. Our approach allows us to avoid specifying a particular parametric form for the hazard rates and numerical maximization of the corresponding likelihood. It allows us to visually assess the presence of deviations over the pre-treatment period from our foundational assumption of parallel trends in the hazard rates, and to test the assumption formally. In addition, we control for the calendar date at which unemployment spells begin using our weighting approach. This ensures that our estimates are robust to differential trends between sub-populations who became unemployed on different dates.

Following the policy change, individuals aged 40-49 who had been employed for at least 312 weeks out of the previous ten years became eligible for 39 weeks of benefits rather than the previous 30 weeks. Some individuals also qualify for a (modest) increase in the replacement rate, which is the proportion of expected earnings given to individuals receiving benefits. To simplify our analysis we consider only the increase in the potential benefit duration (PBD) rather than the change in the level of benefits. As such, we exclude all individuals who qualify for the change in replacement rate from our sample.

To construct our treated group, we collect all individuals in the data who qualify for the extension of PBD from 30 to 39 weeks, excluding those who qualify for the increased replacement rate. Of these individuals we retain only those who became unemployed at or prior to the reform date, and less than 30 weeks after the reform date. The untreated group contains those individuals who satisfy the same demographic criteria as those in the treated group but who became unemployed at or prior to one year before the reform date, and less than 335 weeks before the reform date (thus the dates are shifted back exactly one year compared with the treated group). The treated group consists of $4,058$ individuals and the untreated group $4,207$.

For each individual, the time period $t$ is taken to be the number of days since the beginning of their unemployment spell. Thus for an individual $i$, the period-$t$ outcome $Y_{i,t}$ is an indicator of whether that individual had exited unemployment at or prior to $t$ days after becoming unemployed.

We consider the treatment date to be the 30 week mark. This is the point

at which benefits end for those in the untreated group. Benefits for treated individuals last an additional 9 weeks. Note that this may be problematic for the no-anticipation assumption. Individuals in the treated group were likely aware that their benefits would not end at 30 weeks. Nonetheless, we note that any reduction in the pressure to search for employment prior to this date among the treated group would likely reduce the magnitude of our estimated treatment effects.

Given the importance of seasonal variation on job search, there may be differential trends in the rates of job-finding between individuals who became unemployed on different calendar dates. As we discuss in Section 2.2, such differential trends can be problematic if the distribution of the start dates of unemployment spells differs between the two cohorts. For this reason we apply the weighting scheme specified in (3.5) to re-balance these distributions between the groups. This requires us to drop from the sample individuals who became unemployed on a date on which no untreated individual was made unemployed (there are 17 such dates out of 240).

Figure 4.1: Full Panel

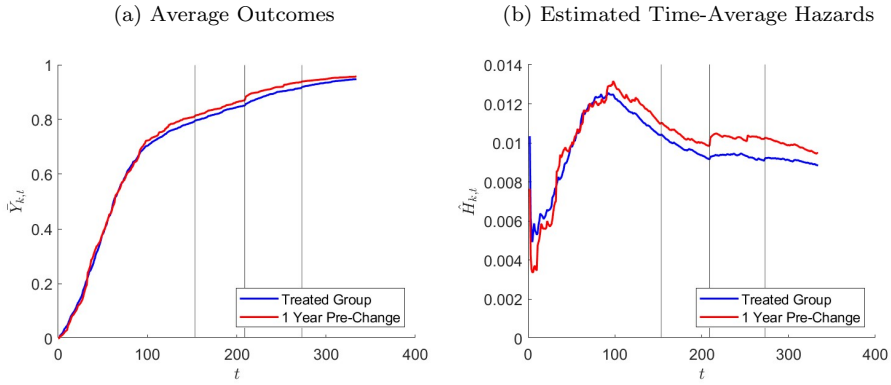



Figure 4.1(a) plots the weighted mean outcomes for the two groups. Recall that the mean outcome is the proportion of individuals who have left unemployment by a given day into their spell. The red curve is the mean outcome for the untreated cohort and blue for the treated cohort. The series for the two cohorts are relatively close, with some apparent divergence immediately following day 210, which is marked with a vertical dashed line and is the point at which benefits expire for individuals in the untreated cohort. Over the period

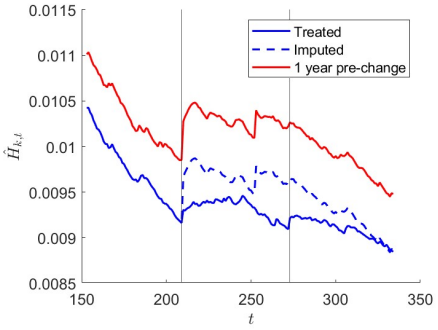
following the benefit extension the two curve eventually begin to converge. The rightmost vertical solid line marks day 273, at which time benefits expire for treated individuals.

Figure 4.1(b) contains the weighted time-average hazards for the two groups with the same color coding as in Figure 4.1(a). Note that these are logarithmic transformations of the values in Figure 4.1(a). We see that in the initial weeks following unemployment the time-average hazard rates for both groups increases steadily before stabilizing and declining. It then remains relatively steady for both groups following 30 weeks with a modest decline for the untreated group.

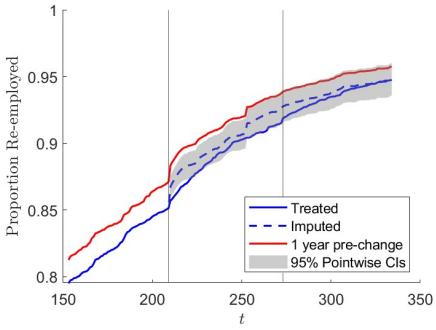
There is very marked jump in the time-average hazard rate for untreated individuals following the end of their benefit period. This may suggest a sudden increase in job search intensity after unemployment benefits run out, or it may reflect a deliberate delay in the start date of new employment until exactly the date of benefit expiration. Notably, there also appears to be a slight increase for the treated group at this same period despite the benefits of these individuals persisting for an additional 9 weeks. This may be explained by imperfect knowledge of the reform.

Figure 4.2: Imputations

(a) Imputed Time-Average Hazards




(b) Imputed Average Outcomes

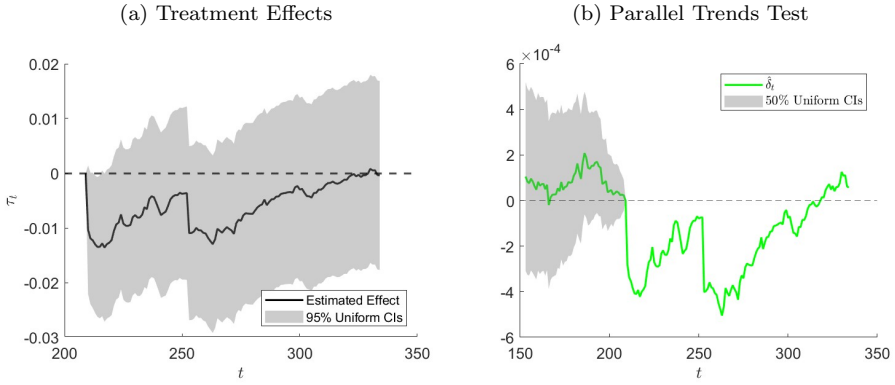



Following standard practice we apply our diff-in-diff approach using only periods within a relatively narrow window around the treatment date. We estimate the level difference between covariate weighted time-average hazards using the observations no more than eight weeks prior to the treatment date and extrapolate forward no more than 125 days after treatment. The choice of initial cut-off reflects that trends in time-average hazards before this point are

unrepresentative of those later on, as discussed above. The point eight weeks prior to the treatment date is marked by the leftmost solid black vertical line in Figures 4.1(a) and 4.1(b).

Figure 4.2(a) shows the weighted time-average hazards over this period. The dashed line indicates the imputed values for the treated cohort under the no-treatment counterfactual. Within this window, pre-treatment parallel trends appear plausible. Figure 4.2(b) plots the mean outcomes over this same window. Again, the dashed line plots the imputed mean outcomes for the treated group. The gray region gives the 95% pointwise confidence bands calculated using the bootstrap method described in Section 3.2. These are confidence bands for the average treatment effect centered on the dashed line.

Figure 4.3: Treatment Effects and Tests




In Figure 4.3(a) we plot the period-specific treatment effects. The treatment effect is strongly negative immediately following treatment but tends to decrease in magnitude over time. Notably, most of the steady reduction in magnitude follows the expiry of benefits for untreated individuals, which occurs at the time marked by the solid vertical line. The gray regions are 95% pointwise and uniform confidence bands. A close inspection reveals that the uniform bands do not contain zero at all periods, rather the upper band lies just below zero for a brief period shortly after treatment. That is, we find a statistically significant negative treatment effect.

Our Wald test for parallel trends produces a p-value of 0.31 so we cannot reject the null of parallel trends even at the 70% level. Figure 4.3(b) visualizes our test for pre-treatment parallel trends. The green curve plots the difference in

time-average hazard estimates relative to the value just prior to treatment. The vertical line indicates the treatment date. We see that this curve is strongly negative following treatment (in line with our negative treatment effect estimates), but is weakly positive prior to treatment. The 50% uniform confidence bands contain zero over the entire pre-treatment period. This indicates a failure to reject pre-treatment parallel trends in the time-average hazards even at the highly conservative 50% level.

# References


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

Abadie, Alberto, & Gardeazabal, Javier. 2003. The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, **93**, 113–132.

Abadie, Alberto, Diamond, Alexis, & Hainmueller, Jens. 2015. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, **59**, 495–510.

Abbring, Jaap H., & Heckman, James J. 2007. *Handbook of Econometrics*. North-Holland, Amsterdam. Chap. Chapter 72 Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation, pages 5145–5303.

Abbring, Jaap H., & Van Den Berg, Gerard J. 2003. The Nonparametric Identification of Treatment Effects in Duration Models. *Econometrica*, **71**, 1491–1517.

Abbring, Jaap H., & van den Berg, Gerard J. 2005. Social Experiments and Instrumental Variables with Duration Outcomes. *Tinbergen Institute Discussion Paper No. TI 05-047/3*.

Arendt, Jacob, Dustmann, Christian, & Ku, Hyejin. 2024. Permanent Residency and Refugee Immigrants' Skill Investment. *Journal of Labor Economics (Forthcoming)*.

Ashenfelter, Orley, & Card, David. 2002. Did the Elimination of Mandatory Retirement Affect Faculty Retirement? *American Economic Review*, **92**, 957–980.

Ashenfelter, Orley, & Greenstone, Michael. 2004. Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias. *Journal of Political Economy*, **94**, 454–460.

Athey, Susan, & Imbens, Guido W. 2006. Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, **74**, 431–497.

Bertrand, M., Duflo, E., & Mullainathan, S. 2004. How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, **119**, 249–275.

Bhuller, Manudeep, Dahl, Gordon B., Løken, Katrine V., & Mogstad, Magne. 2020. Incarceration, Recidivism, and Employment. *Journal of Political Economy*, **128**, 1269–1324.

Blundell, Richard, & Costa Dias, Monica. 2009. Alternative Approaches to Evaluation in Empirical Microeconomics. *The Journal of Human Resources*, **44**, 565–640.

Blundell, Richard W., Dias, Monica Costa, Meghir, Costas, & Van Reenen, John. 2004. Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*, **4**, 569–606.

Card, David E., Chetty, Raj, & Weber, Andrea. 2007. Cash-on-hand and competing models of intertemporal behavior. *Quarterly Journal of Economics*, **4**, 1511–1560.

Chetty, Raj. 2008. Moral Hazard versus Liquidity and Optimal Unemployment Insurance. *Journal of political Economy*, **116**, 173–234.

Cox, D. R. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B*, **34**, 187–202.

Cui, Yifan, Kosorok, Michael R., Sverdrup, Erik, Wager, Stefan, & Zhu, Ruoqing. 2023. Estimating heterogeneous treatment effects with right-censored data via causal survival forests. *Journal of the Royal Statistical Society Series B: Statistical Methodology,*, **85**, 179–211.

de Chaisemartin, Clément, & D'Haultfœuille, Xavier. 2023. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*, **26**, C1–C30.

DiNardo, John, Fortin, Nicole M., & Lemieux, Thomas. 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, **64**, 1001.

Doudchenko, Nikolay, & Imbens, Guido. 2016. Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. *NBER Working Paper*.

Efron, Bradley, & Tibshirani, R. J. 1994. An Introduction to the Bootstrap. *Monograph in Applied Statistics and Probability, No. 57*.

Eissa, Nada. 1996. *Empirical Foundations of Household Taxation*. University of Chicago Press. Chap. Labor Supply and the Economic Recovery Tax Act of 1981, pages 5–38.

Fang, Zheng. 2019. Inference on directionally differentiable functions. *The Review of Economic Studies*, **1**, 377–412.

Friedberg, Leora. 1998. Did unilateral divorce raise divorce rates? *American Economic Review*, **3**, 608–627.

Gruber, J., & Poterba, J. 1994. Tax Incentives and the Decision to Purchase Health Insurance: Evidence from the Self-Employed. *The Quarterly Journal of Economics*, **109**, 701–733.

Gruber, Jonathan. 2004. Is Making Divorce Easier Bad for Children? The Long-Run Implications of Unilateral Divorce. *Journal of Labor Economics*, **22**, 799–833.

Hunt, Jennifer. 1995. The effect of unemployment compensation on unemployment duration in Germany. *Journal of Labor Economics*, **1**, 88–120.

Kaplan, E. L., & Meier, Paul. 1958. Nonparametric Estimation from Incomplete Observations. *JASA*, **53**, 457.

Katz, Lawrence F., & Meyer, Bruce D. 1990. The impact of the potential duration of unemployment benefits on the duration of unemployment. *Journal of Public Economics*, **41**, 45–72.

Khan, Shakeeb, & Tamer, Elie. 2010. Irregular identification, support conditions, and inverse weight estimation. *Econometrica*, **78**(6), 2021–2042.

Klein, John P., & Moeschberger, Melvin L. 2003. *Survival analysis: Techniques for Censored and Truncated Data*. 2. ed. edn. @Statistics for biology and health. New York: Springer.

Lalive, Rafael. 2007. Unemployment Benefits, Unemployment Duration, and Post-Unemployment Jobs: A Regression Discontinuity Approach. *American Economic Review*, **97**, 108–112.

Lalive, Rafael, Ours, Jan Van, & Zwimuller, Josef. 2006. How Changes in Financial Incentives Affect the Duration of Unemployment. *The Review of Economic Studies*, **73**, 1009–1038.

Lambert, Sylvie, & Rossi, Pauline. 2016. Sons as widowhood insurance: Evidence from Senegal. *Journal of Development Economics*, **120**, 113–127.

Lechner, Michael. 2011. *The estimation of causal effects by difference-in-difference methods*. Foundations and trends in econometrics, nos. 2010,3. Boston [u.a.]: Now.

Lichter, Andreas, & Schiprowski, Amelie. 2021. Benefit duration, job search behavior and re-employment. *Journal of Public Economics*, **193**, 104326.

Marinescu, Ioana. 2009. Job Security Legislation and Job Duration: Evidence from the United Kingdom. *Journal of Labor Economics*, **27**, 465–486.

Puhani, Patrick A. 2011. The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. *Economics Letters*, **115**, 85–87.

Rambachan, Ashesh, & Roth, Jonathan. 2023. A More Credible Approach to Parallel Trends. *Review of Economic Studies*, **90**, 2555–2591.

Rose, Evan K. 2021. Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders. *Quarterly Journal of Economics*, **136**, 1199–1253.

Rosenbaum, Paul R., & Rubin, Donald B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70**, 41.

Roth, Jonathan, & Sant'Anna, Pedro H. C. 2023. When Is Parallel Trends Sensitive to Functional Form? *Econometrica*, **91**, 737–747.

Roth, Jonathan, Sant'Anna, Pedro H. C., Bilinski, Alyssa, & Poe, John. 2023. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, **235**, 2218–2244.

Schmidt, Peter, & Witte, Ann Dryden. 1989. Predicting criminal recidivism using 'split population' survival time models. *Journal of Econometrics*, **40**, 141–159.

Schmieder, J. F., von Wachter, T., & Bender, S. 2012. The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates Over 20 Years. *Quarterly Journal of Economics*, **127**, 701–752.

Schmieder, Johannes F., von Wachter, Till, & Bender, Stefan. 2016. The Effect of Unemployment Benefits and Nonemployment Durations on Wages. *American Economic Review*, **106**, 739–777.

Van Den Berg, Gerard J. 2020. Policy discontinuity and duration outcomes. *Quantitative Economics*, **3**, 871–916.

van den Berg, Gerard J., & Vikström, Johan. 2022. Long-Run Effects of Dynamically Assigned Treatments: A New Methodology and an Evaluation of Training Effects on Earnings. *Econometrica*, **90**(3), 337–1354.

Vikström, Johan. 2017. Dynamic treatment assignment and evaluation of active labor market policies. *Labour Economics*, **49**, 42–54.

Wolfers, Justin. 2006. Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results. *American Economic Review*, **96**, 1802–1820.

Wooldridge, Jeffrey M. 2023. Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, **3**, C31–C66.

Wu, Lawrence L., & Wen, Fangqi. 2022. Hazard Versus Linear Probability Difference-in-Differences Estimators for Demographic Processes. *Demography*, **59**, 1911–1928.

# A Proofs

*Proof of Theorem 1.* It is a standard result in duration analysis that the integrated hazard is equal to minus the logarithm of the survivor function. Applying this in our setting under the no-treatment counterfactual yields the following:

$$-ln\left(\frac{1-E[Y_{i,t}^{(0)}|G_i=k]}{1-E[Y_{i,1}^{(0)}|G_i=k]}\right)=\int_1^t h_k^{(0)}(r)dr \tag{A.1}$$

While this result is well-known, for completeness we derive it below using our notation. Recall the definition of $h_k^{(0)}(t)$:

$$\begin{aligned} h_k^{(0)}(t) &= \lim_{\Delta\downarrow 0}\frac{P(Y_{i,t+\Delta}^{(0)}=1|Y_{i,t}^{(0)}=0,G_i=k)}{\Delta} \\ &= \lim_{\Delta\downarrow 0}\frac{P(Y_{i,t+\Delta}^{(0)}=1,Y_{i,t}^{(0)}=0|G_i=k)}{\Delta(1-E[Y_{i,t}^{(0)}|,G_i=k])} \end{aligned}$$

Given $Y_{i,t}^{(0)}=1$ is an absorbing state by Assumption 1, we have:

$$P(Y_{i,t+\Delta}^{(0)}=1,Y_{i,t}^{(0)}=0|G_i=k)=E[Y_{i,t+\Delta}^{(0)}|G_i=k]-E[Y_{i,t}^{(0)}|G_i=k]$$

And so:

$$h_k^{(0)}(t)=\frac{\frac{d}{dt}E[Y_{i,t}^{(0)}|G_i=k]}{1-E[Y_{i,t}^{(0)}|G_i=k]}=-\frac{d}{dt}ln\big(1-E[Y_{i,t}^{(0)}|G_i=k]\big)$$

(A.1) follows from the above by the fundamental theorem of calculus. From (A.1) and the definition of $\bar{H}_{k,t}^{(0)}$, we see that:

$$\bar{H}_{k,t}^{(0)}=\frac{1}{t-1}ln\left(\frac{1-E[Y_{i,1}^{(0)}|G_i=k]}{1-E[Y_{i,t}^{(0)}|G_i=k]}\right)=\frac{1}{t-1}\int_1^t h_k^{(0)}(r)dr \tag{A.2}$$

Since this holds for all $k$ we get the following:

$$\bar{H}_{1,t}^{(0)}-\sum_{k=2}^K W_k\bar{H}_{k,t}^{(0)}=\frac{1}{t-1}\int_1^t\big(h_1^{(0)}(r)-\sum_{k=2}^K W_k h_k^{(0)}(r)\big)dr$$

Substituting $W_1 = h_1^{(0)}(r) - \sum_{k=2}^{K} W_k h_k^{(0)}(r)$ we obtain the equality below.

$$\bar{H}_{1,t}^{(0)} = W_1 + \sum_{k=2}^{K} W_k \bar{H}_{k,t}^{(0)} \tag{A.3}$$

Under Assumption 2 we have that for $t < t^*$, and $k = 1, ..., K$, $E[Y_{i,t}^{(0)}|G_i = k] = E[Y_{i,t}|G_i = k]$, in which case $\bar{H}_{k,t}^{(0)} = \bar{H}_{k,t}$. Substituting gives the first result.

Using (A.2) to substitute for $\bar{H}_{1,t}^{(0)}$ in (A.3) and solving for $E[Y_{i,t}^{(0)}|G_i = 1]$, we get:

$$E[Y_{i,t}^{(0)}|G_i = 1] = 1 - \big(1 - E[Y_{i,1}^{(0)}|G_i = 1]\big) exp\big( - (t-1)(W_1 + \sum_{k=2}^{K} W_k \bar{H}_{k,t}^{(0)})\big)$$

Using Assumption 2 we have that for all $1 \geq t$, and $k = 2, ..., K$, $E[Y_{i,t}^{(0)}|G_i = k] = E[Y_{i,t}|G_i = k]$, in which case $\bar{H}_{k,t}^{(0)} = \bar{H}_{k,t}$. And moreover, $E[Y_{i,1}^{(0)}|G_i = 2] = E[Y_{i,1}|G_i = 2]$. Using this and the above we then have for $t^* \leq t$:

$$E[Y_{i,t}^{(0)}|G_i = 1] = 1 - \big(1 - E[Y_{i,1}|G_i = 1]\big) exp\big( - (t-1)(W_1 + \sum_{k=2}^{K} W_k \bar{H}_{k,t})\big)$$

□

*Proof of Theorem 2.* Recall that $h_1^{(0)}(t;x) = c + h_2^{(0)}(t;x)$. Applying Proposition 1 within a single stratum of the covariates we obtain the following:

$$ln\left(\frac{1 - E[Y_{i,t}^{(0)}|G_i = 1, X_i = x]}{1 - E[Y_{i,1}^{(0)}|G_i = 1, X_i = x]}\right) = (1-t)c + ln\left(\frac{1 - E[Y_{i,t}^{(0)}|G_i = 2, X_i = x]}{1 - E[Y_{i,1}^{(0)}|G_i = 2, X_i = x]}\right)$$

Applying the exponential function to both sides we get:

$$\frac{1 - E[Y_{i,t}^{(0)}|G_i = 1, X_i = x]}{1 - E[Y_{i,1}^{(0)}|G_i = 1, X_i = x]} = exp\big((1-t)c\big) \frac{1 - E[Y_{i,t}^{(0)}|G_i = 2, X_i = x]}{1 - E[Y_{i,1}^{(0)}|G_i = 2, X_i = x]} \tag{A.4}$$

Define $\omega(x)$ as follows:

$$\omega(x) = \frac{P(X_i = x|G_i = 1, Y_{i,1}^{(0)} = 0)}{P(X_i = x|G_i = 2, Y_{i,1}^{(0)} = 0)}$$

With some applications of Bayes rule, we can re-write the above as follows:

$$\omega(x) = \frac{1 - E[Y_{i,1}^{(0)}|G_i = 1, X_i = x]}{1 - E[Y_{i,1}^{(0)}|G_i = 2, X_i = x]} \frac{E[1 - Y_{i,1}^{(0)}|G_i = 2]P(X_i = x|G_i = 1)}{E[1 - Y_{i,1}^{(0)}|G_i = 1]P(X_i = x|G_i = 2)}$$

Then using (A.4) to substitute out $\frac{1-E[Y_{i,1}^{(0)}|G_i=1,X_i=x]}{1-E[Y_{i,1}^{(0)}|G_i=2,X_i=x]}$ from the above, and applying a little algebra, we arrive at the following:

$$\begin{aligned}
&E[1 - Y_{i,1}^{(0)}|G_i = 1]E[\omega(x)(1 - Y_{i,1}^{(0)})|G_i = 2, X_i = x]P(X_i = x|G_i = 2) \\
=&exp\big((1-t)c\big)E[1 - Y_{i,1}^{(0)}|G_i = 1, X_i = x]P(X_i = x|G_i = 1)E[1 - Y_{i,1}^{(0)}|G_i = 2]
\end{aligned}$$

Summing over (or integrating over) the values of $x$, then taking logs and simplifying, we get the following:

$$\frac{1}{t-1} ln\left(\frac{1 - E[Y_{i,1}^{(0)}|G_i = 1]}{1 - E[Y_{i,t}^{(0)}|G_i = 1]}\right) = c + \frac{1}{t-1} ln\left(\frac{1 - E[Y_{i,1}^{(0)}|G_i = 2]}{E[\omega(X_i)(1 - Y_{i,t}^{(0)})|G_i = 2]}\right)$$

Applying definitions gives the result.

□