Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Parallel Trends Forest: Data-Driven Control Sample Selection in Difference-in-Differences

Yesol Huh and Matthew Vanderpool Kling

2025-091

Please cite this paper as:
Huh, Yesol, and Matthew Vanderpool Kling (2025). "Parallel Trends Forest: Data-Driven Control Sample Selection in Difference-in-Differences," Finance and Economics Discussion Series 2025-091. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2025.091.

Parallel Trends Forest: Data-Driven Control Sample Selection in Difference-in-Differences *

Yesol Huh and Matthew Vanderpool Kling

Federal Reserve Board

September 23, 2025

Abstract

This paper introduces parallel trends forest, a novel approach to constructing optimal control samples when using difference-in-differences (DiD) in a relatively long panel data with little randomization in treatment assignment. Our method uses machine learning techniques to construct an optimal control sample that best meet the parallel trends assumption. We demonstrate that our approach outperforms existing methods, particularly with noisy, granular data. Applying the parallel trends forest to analyze the impact of post-trade transparency in corporate bond markets, we find that it produces more robust estimates compared to traditional two-way fixed effects models. Our results suggest that the effect of transparency on bond turnover is small and not statistically significant when allowing for constrained deviations from parallel trends. This method offers researchers a powerful tool for conducting more reliable DiD analyses in complex, real-world settings.

*The views of this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. Federal Reserve Board, 20th St. and Constitution Avenue, NW, Washington, DC, 20551. Please send comments to *yesol.huh@frb.gov*

1 Introduction

Difference-in-differences (DiD) is one of the most frequently used methodologies in economics and social sciences for establishing causality and estimating the impact of policy changes. Its effectiveness relies critically on the parallel trends assumption—the notion that, in the absence of treatment, the difference between the treatment and control groups would have remained constant over time. Because this critical assumption involves unobserved values, that is, what the outcome values would be for the treated sample if there were no treatment, it is nearly impossible to prove that the parallel trends assumption holds within empirical work.

Researchers have used different approaches to convince their audience that the parallel trends assumption holds. The most robust is using randomness in treatment assignment. On the extreme, if treated units were decided completely randomly, the treated group and control group would be similar otherwise, and with enough units parallel trends would hold. As long as there is some randomness in assignment (for instance, treatment assignment is random conditional on certain covariates), researchers can select treatment and control samples for which the assignment would be random within those samples.

Unfortunately, because we rarely get to oversee an experiment in economics, many cases in which researchers would want to use DiD, such as a policy change for a subset of the population, have zero or very little randomness in treatment assignment. Because natural experiments are fairly rare to begin with, we may still want to exploit the setting as much as possible to gain empirical knowledge. In this case, researchers often pick treatment and control samples that look relatively similar to each other in terms of important characteristics, plot the time series in the pre-treatment period, and argue that the series roughly follow parallel trends with one another. Recent literature has used the pre-treatment data to formally test whether the treatment and control data have different trends and proposed allowing explicitly for deviations from parallel trends (Rambachan and Roth, 2023).

In this paper, we address the aforementioned challenge by introducing a novel approach to estimating treatment effects called parallel trends forest. The goal of the parallel trends forest is to find the optimal control sample for which the outcome variable moves in parallel with the treated sample in the absence of the treatment. We estimate the optimal control sample using the pre-treatment data and assume that the parallel movements would continue in absence of treatment. Specifically, for each treatment unit, we find a set of optimal weights on the control units in which the weighted average outcome of the control units would move in parallel with the outcome of the treatment unit. Our approach builds on the recent literature on synthetic control (Abadie and Gardeazabal, 2003; Abadie et al., 2010) and synthetic DiD (Arkhangelsky et al., 2021) in that the optimal control is assumed to be a linear combination of the control sample and the weights are estimated using the pre-treatment period. Methodologically, we use random forests, allowing us to work with a large number of covariates and a large number of treatment and control units. We also introduce a different measure of deviation from "perfect" parallel trends that works with noisy granular data. Using random forests also has the added advantage of automatically selecting covariates that turn out to be important in how the outcome variable moves, rather than selecting on covariates that the researchers assume are important.

We then demonstrate the performance of the parallel trends forest using the introduction of post-trade transparency in corporate bond markets. Between 2002 and 2005, the Financial Industry Regulatory Authority (FINRA) phased in post-trade transparency in which trade information such as price and quantity became available to market participants shortly after the trade. Because bonds were divided into several groups and each group was phased in at different times, this setup allows for a natural experiment through which we can study the impact of post-trade transparency. However, phase lists were determined by issue size, ratings, and turnover—which are variables that are generally considered to impact liquidity, the outcome variable of interest—and had very little randomness built in. Nevertheless, given the importance of the question and the lack of other natural experiments that can study the impact of transparency on liquidity, many academic papers have used this event (Edwards et al., 2007; Bessembinder et al., 2006; Asquith et al., 2019). The data are granular, noisy, and tend to have non-normal distributions, especially for outcome variables such as weekly trading volume or turnover, which have a high share of zeros.

We first show, using a placebo test, that parallel trends forest works better than existing methods such as synthetic control, synthetic DiD, and matrix completion (Athey et al., 2021). The existing methods fail to produce an optimal control that fits the treated sample closely in the pre-treatment period or fits perfectly in-sample but performs poorly out-of-sample. We further show via Monte Carlo simulations that parallel trends forest does almost as well as two-way fixed effects estimator that already knows the correct control sample.

We then use one of the phases in the post-transparency introduction to further illustrate the parallel trends forest and compare it with the two-way fixed effects (TWFE) estimator. The TWFE estimator, which is the estimate from a pooled regression with unit fixed effects and time fixed effects, is what most applied researchers use when employing a DiD method. We study the impact of post-trade transparency on bond turnover for Phase 2, the phase in which medium-sized bonds with high ratings became transparent. The TWFE estimator, when using various "reasonable" control samples, gives different estimates and sometimes even flips sign depending on the control sample used. Thus, an algorithm to select the optimal control in a data-driven way is important. Using the parallel trends forest, we show that the magnitude of the average treatment effect is smaller than what is estimated through TWFE. Furthermore, if we explicitly allow for constrained deviations from parallel trends (Rambachan and Roth, 2023), we show that the treatment effect is not statistically significant. We also illustrate a parallel trends forest with honesty, which builds a tree on one sample and estimates the weight using the constructed tree and data from another sample. This prevents the algorithm from overfitting in the pre-treatment period.

Overall, the parallel trends forest proposed in this paper is a data-driven method for selecting the optimal control that most closely moves in parallel with the treated sample. While a large-scale randomized trial would be the ideal setting, in cases with zero or little randomization in treatment assignment our method can help researchers pick the best control sample for DiD studies.

In the literature, synthetic control (Abadie and Gardeazabal, 2003; Abadie et al., 2010), synthetic DiD (Arkhangelsky et al., 2021), and matrix completion (Athey et al., 2021) are most closely related to this paper in that they use pre-treatment period to find the set of weights on the control sample that would most closely fit the treatment sample in the absence of a treatment. Synthetic control and synthetic DiD are generally meant for a small number of highly-aggregated data and optimize over the weights directly, which is difficult when the universe of control units are large. Also, with granular data, allowing the algorithm to use covariates to find which set of units follow the parallel trends more closely leads to less overfitting than optimizing the weights directly. Lastly, as we will outline in Section 2, we use a different objective function that is more effective than the usual sum of squared errors for granular, non-normally distributed data.

Methodologically, our paper is most closely related to the causal forest developed in Wager and Athey (2018) and Athey et al. (2019). However, the objective of the causal forest is completely different from our parallel trends forest. The goal in causal forests is to estimate heterogeneous treatment effects, so trees are split such that units with similar treatment effect would be in the same leaf. However, in our parallel trends forest, the goal is to group units by those that have parallel trends with one another. We adopt many of the techniques from causal forests—namely, "honesty" and calculation of weights from the forest.

Our paper is also related to an emerging literature in economics that studies the DiD methodology more carefully.[1] A growing number of papers carefully analyze whether parallel trends hold in pre-treatment data before running DiD, for instance, see He and Wang (2017). However, Roth (2022) shows that if the parallel trends assumption is violated, then using only the samples that pass the pretests may bias the treatment

[1]For a more comprehensive review on DiD, see Roth et al. (2023) and Baker et al. (2025).

effect estimates even more. Because we select the optimal control sample such that they move roughly in parallel with the treated sample, this critique on overfitting could apply to our method. There are two aspects of our methodology that help address this concern. First, overfitting is relatively mild in random forests due to averaging across many trees and using a subset of possible covariates at every split, especially when compared to simply choosing units that pass the pretests. Second, in Section 4.5, we develop a parallel trends forest with honesty, which builds the tree on one sample and estimates the weights on the other. This further addresses overfitting.

DiD methodology is intimately related to the potential outcomes framework. Both use a treatment and a control sample to estimate the causal effect from some treatment. The potential outcomes framework usually employs a much stricter approach of aiming to approximate a randomized experiment (Rubin, 2008). A critical assumption in this "design-based" approach is that the probability of being treated is bounded away from zero and one; that is, there is randomness in treatment assignment. If the probability of treatment differs between the treatment and control groups, propensity score matching can be used to achieve balance. On the other hand, DiD hinges on a much weaker parallel trends assumption. Our setting of allowing treatment assignment to be completely based on observed covariates clearly does not fit into the design-based approach. However, while our setting may not be the optimal scenario, it is still a useful and important one to study. Given how rare large-scale natural experiments are in observational studies, researchers would take every opportunity to exploit them, and employing our methodology would be preferable to blindly applying a DiD. Also, in settings where we observe the outcome variables for a large number of pre-treatment time periods, it is possible to find a reasonable set of optimal control sample that moves in parallel as long as it exists.

2 Parallel Trends Forest

Following the convention in the potential outcomes framework, $Y_{u,t}$ denotes the observed outcome variable for unit u at time t, and $Y_{u,t}^{(d)}$ is the (potential) outcome variable if the treatment status is d at time t, with $d = 0$ being untreated and $d = 1$ being treated. We observe $t \in \{1, 2, \ldots, T\}$ time periods and a total of $M + N$ units with $j \in \{1, \ldots, M\}$ not changing treatment status during this time period ("control sample") and $i \in \{M + 1, \ldots, M + N\}$ receiving treatment at $t = T_1 + 1$ where $T_1 < T$ ("treatment sample"). We allow for the possibility that treatment assignment can be deterministic, that is, treatment assignment can fully depend on covariates X. Observation units are granular—each time period is fairly short, and each unit is small—and the observed outcome variables can be noisy and have non-normal distributions.

The goal of the parallel trends forest is for each treatment unit $i \in [M+1, M+N]$, to find a set of weights $w_j(i), j \in [1, M]$ with $\sum_{j=1}^{M} w_j(i) = 1$ that would predict the unobserved counterfactual of the treated unit i as a weighted average of observed outcome values of the control sample:

$$Y_{i,t}^{(0)} = u_i + \sum_{j=1}^{M} w_j(i) Y_{j,t} \quad t \in [T_1 + 1, T] \tag{1}$$

In other words, the goal is to find for treatment unit i a synthetic weighted control unit that satisfy the parallel trends assumption, with u_i allowing for a constant difference between the two.

In its strictest form, a treated unit i and a control unit j satisfy the parallel trends assumption if (2) holds for all $t_1, t_2 \in [1, T]$.

$$E[Y_{i,t_1}^{(0)}] - E[Y_{i,t_2}^{(0)}] = E[Y_{j,t_1}] - E[Y_{j,t_2}] \tag{2}$$

Equation (2) implies that without treatment, the outcome variables for i and j move in parallel with one another. DiD uses this assumption to infer what the outcome variable would have been for the treated units without treatment at $t \geq T_1 + 1$.

While there is consensus on the definition of parallel trends, there is no unanimous measure for the degree of deviation from parallel trends that two time series exhibit from one-another. Suppose we want to measure how far from parallel two time series $Y_i = \{Y_{i,t}\}_{t=1}^{T_1}$ and $Y_j = \{Y_{j,t}\}_{t=1}^{T_1}$ are.

Consider the often-used deviation measure of

$$\frac{1}{T_1} \sum_{t=1}^{T_1} (Y_{i,t} - \bar{Y}_i - Y_{j,t} + \bar{Y}_j)^2 \tag{3}$$

where $\bar{Y}_u = \frac{1}{T_1} \sum_{t=1}^{T_1} Y_{u,t}$. To see why this would measure the deviation from parallel trends, note that if parallel trends hold exactly then $E[Y_{i,t}] - \bar{Y}_i = E[Y_{j,t}] - \bar{Y}_j$ for all t. Synthetic control, synthetic DiD, and matrix completion method all minimize some variation of this deviation measure to find the "optimal" synthetic control sample. This measure works well when data is highly aggregated and has very little noise. However, in cases where units are small, observations are noisy, and observed outcomes are drawn from a non-normal or unusual distribution, this measure does not work well, as we will demonstrate in Section 3.

We instead define our own deviation measure $||Y_i, Y_j||$ as:

$$c(Y_i, Y_j, \tau) = \left(\frac{1}{T_1 - \tau} \sum_{\tau+1}^{T_1} Y_{i,t} - \frac{1}{\tau} \sum_{1}^{\tau} Y_{i,t} \right) - \left(\frac{1}{T_1 - \tau} \sum_{\tau+1}^{T_1} Y_{j,t} - \frac{1}{\tau} \sum_{1}^{\tau} Y_{j,t} \right)$$

$$= \frac{1}{T_1 - \tau} \sum_{\tau+1}^{T_1} (Y_{i,t} - Y_{j,t}) - \frac{1}{\tau} \sum_{1}^{\tau} (Y_{i,t} - Y_{j,t}) \quad (4)$$

$$||Y_i, Y_j|| = \sqrt{\frac{1}{T_1 - 1} \sum_{\tau=1}^{T_1-1} c(Y_i, Y_j, \tau)^2} \quad (5)$$

To illustrate the intuition, consider a placebo test with observed data $Y_i = \{Y_{i,t}\}_{t=1}^{T_1}$ and $Y_j = \{Y_{j,t}\}_{t=1}^{T_1}$. In this test, we hypothetically assume that unit i receives the treatment at time $\tau + 1$, while unit j does not change treatment status throughout. The estimated treatment effect will be $c(Y_i, Y_j, \tau)$. If the two series satisfy the parallel trends assumption and neither are treated during the period, $E[c(Y_i, Y_j, \tau)] = 0$ for all $\tau \in [1, T_1 - 1]$. Thus, $||Y_i, Y_j||$ would measure how far the placebo treatment effect is from zero when the placebo treatment date is assigned at random. Measuring deviation from parallel trends in this way ties much more closely to the treatment effect, and we find our method behaves better with noisy data and non-normal distributions.

We now move on to outlining our method, which we call "parallel trends forest." The aim of the parallel trends forest is to find the optimal control sample that moves in parallel for each treatment unit.

We first build a tree that uses the deviation measure $||Y_i, Y_j||$ to group units that have parallel trends together based on only the pre-treatment period data. This is similar to the classification and regression tree (CART) but with two major differences. First, trees are trained on a panel data instead of the usual cross sectional data, so we only allow splits along units and not along the time dimension. Second, we use a different splitting function from what is normally used in a CART.

Consider a set of units $\mathcal{J}$ in the parent node that we aim to partition into two subsets, $\mathcal{S}_1$ and $\mathcal{S}_2$. We define the deviation from parallel trends across units within the leaf $\mathcal{J}$ as:

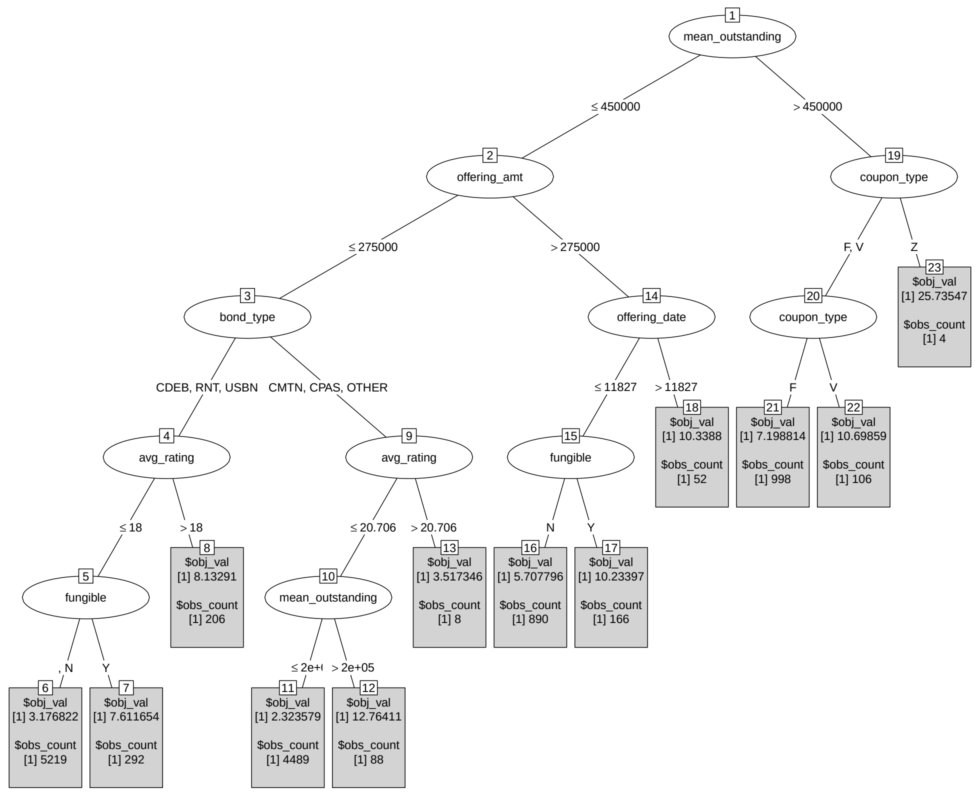
$$L(\mathcal{J}) = \sum_{i \in \mathcal{J}} ||Y_i, \, mean_{j \in \mathcal{J}}(Y_j)||^2. \quad (6)$$

We split the units into two groups, $\mathcal{S}_1$ and $\mathcal{S}_2$, in a way that minimizes $L(\mathcal{S}_1) + L(\mathcal{S}_2)$.

Using this splitting rule, we construct B trees to get a parallel trends forest. For the most part, we follow the literature on random forests (Breiman, 2001). For each tree, we take a random sample (without

Figure 1: **Example tree**

The following picture shows a truncated tree with first few levels.



replacement) of K units from $M + N$ total units, and at every split, we allow the algorithm to split on $\hat{k}$ randomly-selected covariates out of k covariates.[2] We use $K = 0.5(M + N)$ and $\hat{k} = \frac{1}{3}k$. We split until all leaves have at most 100 units.

Figure 1 shows an example tree generated by our algorithm. The outcome variable of interest is weekly turnover (i.e., weekly trading volume divided by amount outstanding) for corporate bonds. Trees sometimes split on covariates, like amount outstanding, that are intuitively highly correlated with turnover, but at other times split on covariates, such as coupon type or bond type, whose correlation with turnover is less intuitive. When trees split on seemingly less important covariates, it may be that they are uncovering potentially complex relationships that are missed otherwise or they may be overfitting. Regardless, we will show in Section 3 that our parallel trends forest, which is an ensemble of trees, performs quite well.

[2]Random forests, including the implementation in Breiman (2001), typically sample with replacement to create training data for building trees, but several papers have shown that subsampling without replacement behave better in certain situations (Bühlmann and Yu, 2002; Strobl, Boulesteix, Zeileis, and Hothorn, 2007). We follow Athey et al. (2019) and use subsampling without replacement, but results are similar when we sample with replacement.

Lastly, we find the optimal control sample weights for each treatment unit in a similar manner to the construction done in Athey et al. (2019). For each treated unit i, denote $\mathcal{G}(i)$ as the set of trees, out of B trees, in which i appears in and have at least one control unit in the same leaf. Then, for every pair of treated unit i and control unit j, the weight $w_j(i)$ is calculated as:

$$w_{b,j}(i) = \frac{1(j \text{ is in the same leaf as } i \text{ in tree } b)}{\text{Number of control bonds in same leaf as } i \text{ in tree } b}, \quad b \in \mathcal{G}(i) \tag{7}$$

$$w_j(i) = \frac{1}{\mathcal{N}(\mathcal{G}(i))} \sum_{b \in \mathcal{G}(i)} w_{b,j}(i) \tag{8}$$

where $\mathcal{N}(\mathcal{G}(i))$ is the number of trees in $\mathcal{G}(i)$. The weight $w_j(i)$ captures the frequency with which control unit j falls into the same leaf as i. $\sum_j w_j(i)Y_j$ is the optimal control sample for treated unit i. We can approximate $Y_{i,t}^{(0)}$ for treated bond i on time t as:

$$\hat{Y}_{i,t}^{(0)} = u_i + \sum_{j=1}^{M} w_j(i)Y_{j,t} \tag{9}$$

$$u_i = \frac{1}{T_1}\sum_{t=1}^{T_1} Y_{i,t} - \frac{1}{T_1}\sum_{t=1}^{T_1}\sum_{j=1}^{M} w_j(i)Y_{j,t} \tag{10}$$

Because we do not force individual leaves to contain both treated and control units, there could exist leaves that have only treated units or only control units. Thus, in principle, if a treated unit behaves very differently from all control units, it may not have any control units in the same leaf for all trees in the forest and thus may be not possible to calculate the weights for the particular treated unit. While this may seem problematic at first, excluding those treatment units that do not have good optimal controls is preferable since the ATT for them would be biased. One may also choose to omit treatment units with a small $\mathcal{N}(\mathcal{G}(i))$ for similar reasons. In our particular use case, all of our treated units have optimal controls.

As is standard in the literature, we are interested in the average treatment effect for the treated (ATT). ATT is estimated as:

$$ATT = \frac{1}{N(T-T_1)} \sum_{i=M+1}^{M+N} \sum_{t=T_1+1}^{T} (Y_{i,t} - \hat{Y}_{i,t}^{(0)}). \tag{11}$$

It is worth noting that we include units that are treated before the sample period starts ("earlier-treated" units) in the control sample as long as they do not change treatment status during the sample period. Goodman-Bacon (2021) argues that if the treatment effect is time-varying rather than a shift in levels, using the earlier-treated units as controls can be problematic. In our case, the control sample is just the set of candidates that can be included in the optimal control. If the effect of the treatment for the earlier-treated units has not been fully phased in by the start of the sample period, those units would likely have a different trend from the treated units and would end up having zero or very small weights in the optimal control.

3 Comparison with Existing Methods

Three methodologies from the literature—synthetic control, synthetic DiD, and matrix completion—share our goal of finding the weights to construct an optimal control. In this section, we demonstrate that in a setting with noisy, granular, and non-normally distributed data, these existing methods may fall short while our parallel trends forest approach performs well.

3.1 Description of existing methods

We first quickly describe synthetic control, synthetic DiD, and matrix completion at a high level. For a more complete explanation, readers are referred to the citations within. Synthetic control was proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) to estimate the treatment effect in situations where the number of treated units, N, is small, and each unit is at an aggregated level such as a state or a country. Using the setting from Doudchenko and Imbens (2016) but with N possibly larger than 1, synthetic control minimizes:[3]

$$L_{sc} = \frac{1}{NT_1} \sum_{i=M+1}^{M+N} \sum_{t=1}^{T_1} \left(Y_{i,t} - \sum_{j=1}^{m} w_j(i) Y_{j,t} \right)^2 + \text{regularization term.} \tag{12}$$

Synthetic DiD also minimizes a very similar function:

$$L_{sdid} = \frac{1}{T_1} \sum_{t=1}^{T_1} \left(\frac{1}{N} \sum_{i=M+1}^{M+N} Y_{i,t} - w_0 - \sum_{j=1}^{m} w_j Y_{j,t} \right)^2 + \text{regularization term.} \tag{13}$$

The regularization is done slightly differently for each approach and synthetic DiD includes an intercept term, but otherwise the objective function is very similar and uses the deviation measure (3). Estimation of the treatment effect is somewhat different because synthetic DiD allows for unit fixed effects and time fixed effects as well as including time weights.

Matrix completion (Athey et al., 2021) takes a somewhat different approach. This approach sets matrix $\hat{Y}$, a $(M + N) \times T$ matrix consisting of Y values if there is no change in treatment status for all $(M + N)$

[3] The original setting from Abadie et al. (2010) is slightly different in that they include covariates in the objective function.

units, as:

$$\hat{Y} = \begin{pmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,T_1} & Y_{1,T_1+1} & \cdots & Y_{1,T} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,T_1} & Y_{2,T_1+1} & \cdots & Y_{2,T} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Y_{M+1,1} & Y_{M+1,2} & \cdots & Y_{M+1,T_1} & ? & \cdots & ? \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Y_{M+N,1} & Y_{M+N,2} & \cdots & Y_{M+N,T_1} & ? & \cdots & ? \end{pmatrix}, \tag{14}$$

where the lower right $N \times (T - T_1)$ quadrant is missing because we do not observe the outcome of the treated units in the absence of treatment after time T_1. The missing values of $\hat{Y}$ are then imputed by using a low-rank approximation of what is observed in $\hat{Y}$.[4]

3.2 Comparison with synthetic control, synthetic DiD, and matrix completion

We compare the performance of the three existing methods and our parallel trends forest via a placebo test using the first 36 weeks of trading data from 2006.[5] This 36 week period is well removed from the last TRACE phase-in date of February 7, 2005, so most likely any phase-in effects would have been fully incorporated by the start of 2006. In this placebo test we designate a "treatment" group and an "control" group such that half of the bonds in the control group have similar characteristics to the bonds in the treatment group, the other half not. If a method is effective, it should assign higher weights to bonds that are more similar to the treatment group, and lower weights to those dissimilar. The treatment group contains 663 randomly selected investment-grade bonds of \$500M or larger ("large IG bonds"). The control group is comprised of two sets of bonds: 663 large IG bonds (the similar bonds to the treatment sample) and 663 randomly-selected high-yield bonds that are \$100M or smaller ("small HY bonds").[6]

We first confirm in Figure 2 that turnover for large IG bonds in the control sample indeed behaves similarly to the treatment sample by moving in parallel, whereas the small HY part of the control sample moves differently.

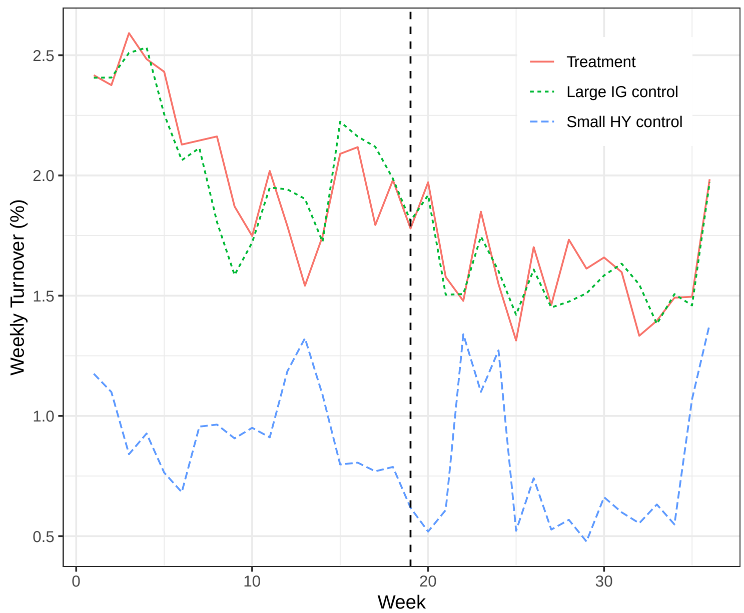
Figure 3 plots the outcome of the four methods. We pretend that treatment begins at week 19, so only the data from the first 18 weeks are used to fit the weights in synthetic control, synthetic DiD, and parallel trends

[4]Matrix completion model does not explicitly approximate the counterfactual outcomes for the treated sample as a weighted average of the control sample; rather it approximates the values directly. However, given that $\hat{Y}$ is approximated with a low-rank matrix, we can use singular value decomposition to get an approximate set of weights.

[5]We will describe the setting and data in more detail in Sections 4.1 and 4.2.

[6]There are 1,326 total large IG bonds that are outstanding throughout the sample period. We randomly select half of them to be in the treatment group, and the other half to be in the control group.

Figure 2: **Average weekly turnover for the placebo test data**
This figure plots the average weekly turnover for the treatment sample, large IG bonds in the control sample, and small HY bonds in the control sample.



forest.[7] In the case of matrix completion, the first 18 weeks of treatment data and all 36 weeks of control data are used for the estimation. All methods except for synthetic control allow for a constant difference between the treatment sample and the optimal control sample. Panel (a) plots the average turnover for treated and control units as well as the fitted average values using synthetic control and synthetic DiD. Panels (c) and (d) plot the values using matrix completion and parallel trends forest, respectively.

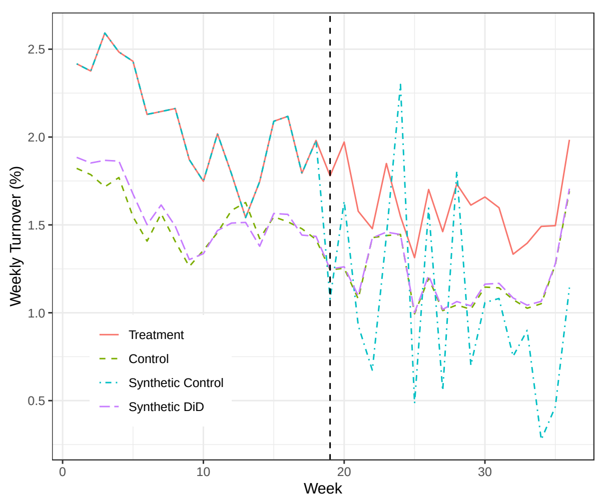
Panel (a) indicates that synthetic control fits the data very closely in-sample (in the pre-treatment period) but very poorly out-of-sample. Synthetic DiD produces a counterfactual that is very similar to the control sample. Panel (b) illustrates the distribution of weights produced by the synthetic DiD. Specifically, it displays a Lorenz curve of $\bar{w}_j = \frac{1}{N}\sum_{i=M+1}^{M+N} w_j(i)$, where $\bar{w}_j$ represents the average weight assigned to control bond j across all treated bonds. To construct the curve, we first sort $\bar{w}_j$ in ascending order and then plot the cumulative sum against the number of bonds. The resulting straight line implies that the synthetic DiD assigns almost equal weights to all control bonds, which is clearly far from optimal. Panel (c) shows the results for matrix completion both with and without time fixed effects. The counterfactual with the fixed effects is clearly similar to the average control turnover up to a constant shift, indicating that weights are similar across all control bonds, which is far from optimal as one half (large IG) of the control sample is

[7] Strictly speaking, synthetic DiD uses pre-treatment period data to fit the unit weights and all control data to fit the time weights. For our comparison, unit weights are more important.

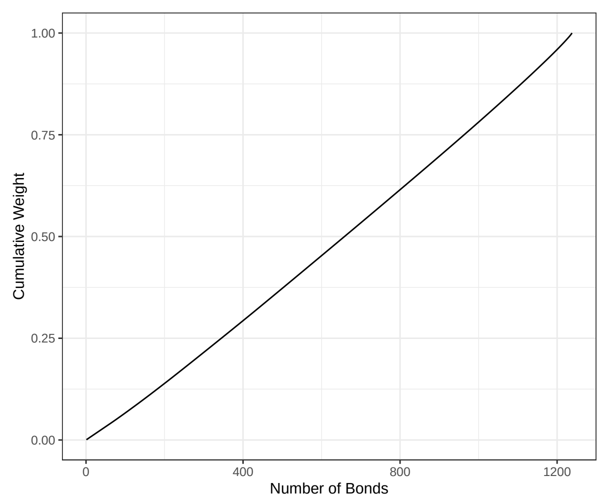
Figure 3: **Performance of the four methods**

Figures below plot the outcome of the placebo test using synthetic control, synthetic DiD, matrix completion, and parallel trends forest. Panels (a), (c), and (d) plot the average weekly turnover for the treatment sample, the control sample, and the fitted optimal control sample using synthetic control and synthetic DiD (panel a), matrix completion (panel c), and parallel trends forest (panel d). Panel (b) presents the Lorenz curve for the optimal weights derived by the synthetic DiD method.

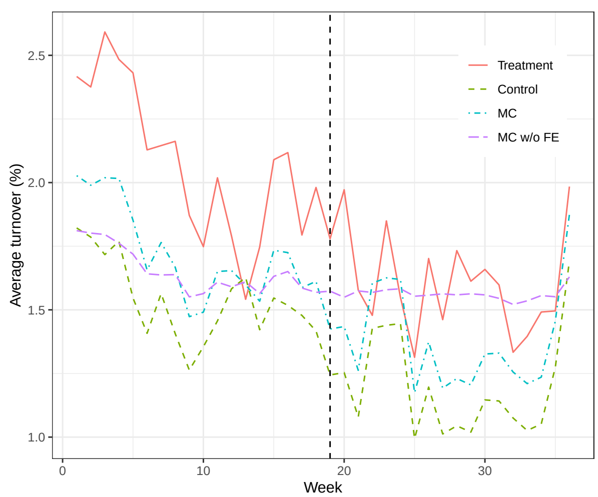
(a) Synthetic control and synthetic DiD



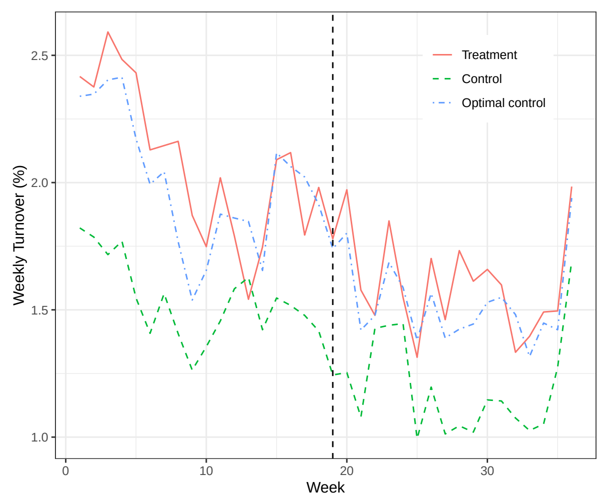
(b) Lorenz curve for synthetic DiD weights



(c) Matrix completion



(d) Parallel trends forest



clearly more similar to the treatment sample than the other half (small HY). The counterfactual without fixed effects is much flatter and far from what the average treatment sample turnover looks like.

Results for parallel trends forest in Figure 3(d) indicate that parallel trends forest performs significantly better. The optimal control tracks the treatment sample quite well, and 93.1% of the weight is in large IG bonds.

3.3 Monte Carlo simulations

Having established that parallel trends forest works significantly better than synthetic control, synthetic DiD, and matrix completion in our setting, we now further study the performance of parallel trends forest and compare it to the usual DiD approach (TWFE estimator) using Monte Carlo simulations. For each Monte Carlo simulation, we generate a placebo test sample as outlined in Section 3.2 and extract the ATT estimate from the parallel trends forest as well as the TWFE estimate from the usual DiD setup for comparison. The TWFE estimate is the estimated β from the following pooled regression:

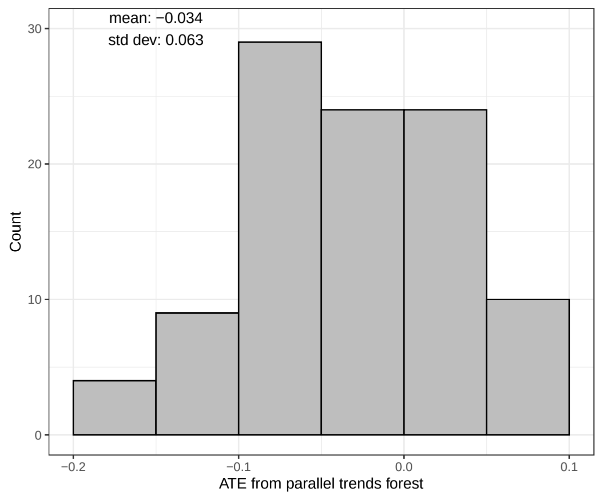
$$Y_{u,t} = A_u + B_t + \beta \mathbf{1}(u \in trmt)\mathbf{1}(t > T_1) + \epsilon_{u,t} \tag{15}$$

where A_u is the bond fixed effect, B_t is the time fixed effect, $\mathbf{1}(u \in trmt)$ indicates whether bond u is in the treatment sample, and $\mathbf{1}(t > T_1)$ indicates whether time t is in the post-treatment period.

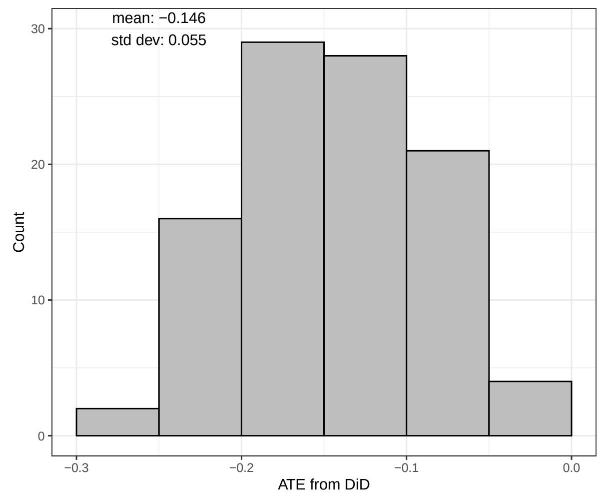
Figure 4 presents the results of 100 Monte Carlo simulations. Since this is a placebo test, a good estimator would give estimates close to zero. The TWFE estimates that use all control bonds (Panel (b)) clearly perform poorly, as the mean estimate is -0.146 with a standard deviation of 0.055. In comparison, parallel trends forest, presented in Panel (a), does quite well with a mean estimate of -0.034 and standard deviation of 0.063. It does slightly worse than the case where we know the optimal control sample ex ante (large IG bonds), as presented in Panel (c). Overall, the parallel trends forest performs quite well in selecting the correct weights.

Figure 4: **Monte Carlo simulation**
The figures below plot the distribution of the average treatment effect estimates from 100 Monte Carlo simulations. Panel (a) uses ATT estimates from parallel trends forest. Panels (b) and (c) use TWFE estimates, where Panel (b) uses all control bonds as controls and Panel (c) uses only large IG bonds as controls. Mean and standard deviation of average treatment effect estimates are also presented.

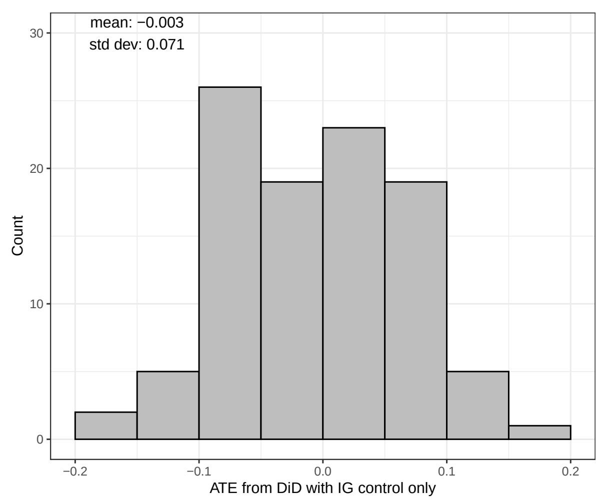
(a) Parallel trends forest



(b) DiD using all control bonds



(c) DiD using large IG bonds only



4 Effect of Post-Trade Transparency: Analysis using Parallel Trends Forest

4.1 TRACE Introduction

Post-trade transparency—that is, the release of transaction data after the execution of a trade—was introduced in the corporate bond market over multiple phases between 2002 and 2005 through a system called Trade Reporting and Compliance Engine (TRACE). Multiple academic papers have exploited the phase-in design of TRACE introduction to study the effect of transparency on trading costs (Bessembinder et al., 2006; Edwards et al., 2007; Goldstein et al., 2007). The majority of academic papers agree with the finding that transparency decreased trading costs for customers in all phases of the introduction.

Although the effect transparency has on trading cost is well-studied and consistent, the opposite is true for the effect on trading activity. Dealers opposed increasing transparency because they believe it would make it more difficult to offload positions acquired from market making, and according to Bessembinder and Maxwell (2008), many market participants have indicated greater difficulty trading post-TRACE. However, additional studies reach different conclusions. Goldstein et al. (2007) find that trading volume did not change around the FINRA 120 event. Asquith et al. (2019) study the effect of transparency on various measures of trading activity separately for all phases and find that trading activity remains unchanged for all phases, except the last in which the less-frequently traded high-yield bonds are phased in. For those illiquid bonds, they find that number of trades decreases with TRACE introduction.

Because TRACE was phased in over multiple disjoint events, it is a prime candidate for a natural experiment to establish causality. However, the choice of which bonds phasing in when was not random except for one small phase. Bonds with similar ratings, size, and trade frequencies were phased in together, and these characteristics are generally considered important for bond liquidity. Therefore, one cannot argue that the control bonds are similar to treated bonds, potentially violating the parallel trends assumption that is crucial for a DiD study. Furthermore, any effects of TRACE introduction seems to be gradual, making it necessary to look over a relatively long sample period and rely heavily on the parallel trends assumption. Researchers have tried to circumvent this problem by carving out a subset of control bonds that look relatively similar to the treated bonds in terms of ratings and bond size, but because there is little common support the authors still implicitly make a strong parallel trends assumption.

FINRA began collecting trade information such as traded volume and price from dealers through TRACE starting July 1, 2002. Dealers were required to submit trade information to TRACE for all TRACE-eligible corporate bonds, but FINRA did not disseminate all of this information. They initially disseminated information to the market for a subset of bonds (Phase 1). Over the next few years, FINRA expanded the dissemination to cover all TRACE-eligible bonds that are not 144a bonds by February 2005. The phases were as follows.

- Phase 1, July 1, 2002: Investment grade bonds with issue size $1 billion or greater
- FIPS 50, July, 1, 2002: 50 high-yield bonds disseminated under Fixed Income Pricing System (FIPS)
- Phase 2, Mar 3, 2003: Investment grade bonds of issue size $100 million or greater, or bonds rated A- or higher
- FINRA 120, Apr 14, 2023: 120 chosen BBB-rated bonds
- Phase 3A, Oct 1, 2003: Bonds rated BBB, and more-frequently traded bonds within those rated BB+ or lower
- Phase 3B, Feb 7, 2005: Less-frequently traded bonds within those rated BB+ or lower

In this paper, we illustrate our parallel trends forest by using the methodology to estimate the impact of Phase 2 implementation on bond turnover. We also compare our results with TWFE estimates.

4.2 Data

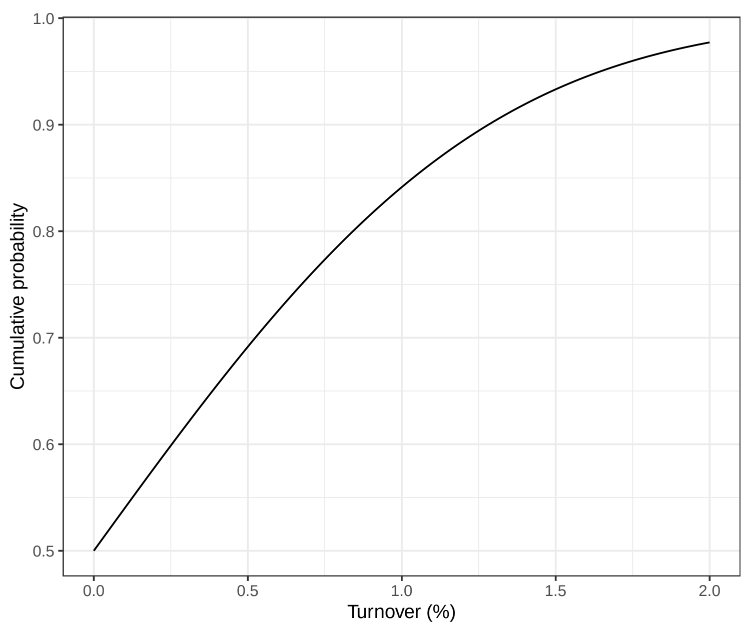
We use the enhanced TRACE data from WRDS to gather information on dissemination status and trading activity. TRACE data includes a list of bonds that are eligible to be reported to TRACE ("TRACE-eligible bonds," which is a superset of bonds eligible for dissemination) and an indicator for whether each bond is eligible for post-trade dissemination for every trading day.[8] We create the treatment and control samples for Phase 2 by using this list. The control sample is all bonds that exist throughout the 36-week period surrounding the phase-in date and do not change dissemination status, and thus include bonds that are disseminated throughout (Phase 1 and FIPS 50 bonds) as well as bonds that are not disseminated throughout (bonds with lower ratings and smaller bonds). The treatment sample are the bonds that exist throughout the 36-week period and change their dissemination status to start dissemination exactly on the

[8]Some bonds may be outstanding and eligible to report to TRACE but may not trade at all over the sample period, and for our goal of studying the impact of transparency on turnover, it is important to include these bonds in the sample. Thus, cleaning the list of TRACE-eligible bonds correctly is crucial. The TRACE master file in WRDS often contains bonds that have already matured or have zero outstanding; we delete those bond-days. We also delete convertible bonds, exchangeable bonds, and 144a bonds.

phase-in date and do not change dissemination status on any other dates during the 36-week period. There are 2,204 bonds in the treatment sample and 10,314 bonds in the control sample.

The outcome variable of interest is weekly turnover, calculated as the week's trading volume divided by the bond's outstanding amount. We allow the algorithm to split on more than 25 covariates, including variables that are usually considered to impact turnover such as average rating, outstanding amount, age, and past turnover, as well as potentially less-related covariates such as seniority and industry. Most of these bond characteristics variables are from Mergent FISD. Figure 5 plots the cumulative probability distribution of the weekly turnover data. About half of the bond-week observations are zero, and the rest roughly follows an exponential distribution.

Figure 5: **Distribution of weekly turnover**
This figure plots the cumulative probability distribution for weekly turnover. Each observation is at the bond-week level.



4.3 Two-way fixed effects estimator

We first present the results from the TWFE estimator, the most commonly used estimator in the DiD literature. This specification assumes that the control sample that is used in the regression satisfies the parallel trends assumption. Because the treatment assignment is not random and treatment criteria of ratings and issue size are highly correlated with the variable of interest (turnover), we follow the existing literature and use a sample of control bonds that have somewhat similar characteristics as the treated bonds.

In particular, we follow Edwards et al. (2007) and construct three different control samples: bonds that are transparent throughout the sample and are rated A or higher ("Transparent & $\geq$ A"), bonds that are not transparent throughout and are rated A or higher ("Not Transparent & $\geq$ A"), and bonds that are not transparent throughout that are BBB-rated and have size between 100M and 1B ("BBB & 100M–1B").[9] The first two control samples are similar in ratings but differ in issue size from the treated sample, and the third control sample is similar in issue size but differ in ratings. We also test using all control bonds as an additional control sample.

Figure 6 shows the average weekly turnover for the treatment sample, full control sample, and the three control samples. It is clear that even in the pre-treatment period the different control samples behave quite differently. For instance, the Transparent & $\geq$A sample have a much stronger trend downwards during the pre-treatment period compared to the treatment sample.

Figure 6: **Average weekly turnover for Phase 2 data**
This figure plots the average weekly turnover for treated bonds, all control bonds, and various control samples.

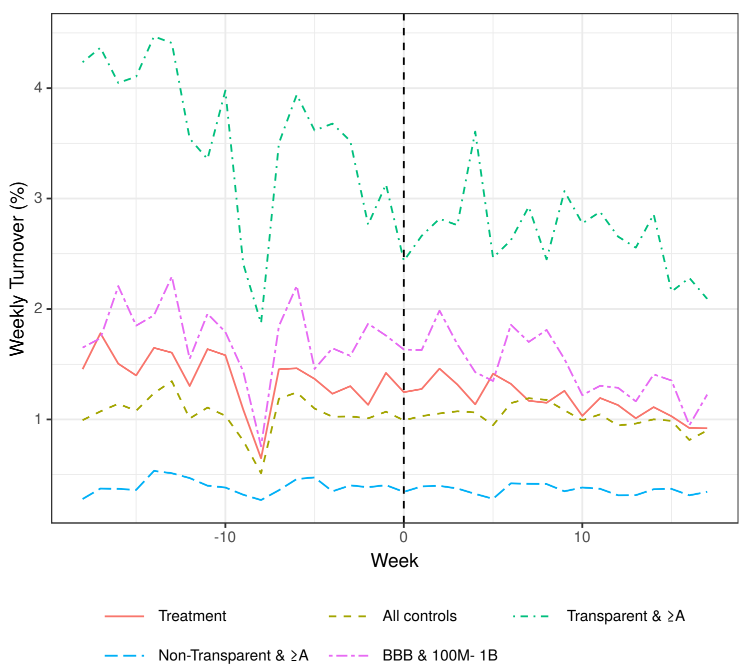


Table 1 presents the TWFE estimator using the various control samples. The estimates for the treatment effect changes with the control sample used, which underscores the need for a method that chooses which

[9] Edwards et al. (2007) study the effect of post-trade transparency on bid-ask spreads, not turnover, so our results is not a critique of the paper.

control sample works best. Moreover, the most optimal control sample may not be one of the four used here.

Table 1: **Two-way fixed effects estimator**
The following table presents the treatment effects estimated by regressing weekly turnover on bond fixed effects, time fixed effects, and the interaction of post-treatment indicator and treated indicator. For their respective control samples, column (1) uses all control bonds, column (2) uses Transparent & $\geq$ A bonds, column (3) uses Not Transparent & $\geq$ A bonds, and column (4) uses BBB & 100M–1B bonds.

	All	Transparent & $\geq$ A	Not Transparent & $\geq$ A	BBB, 100M-1B
	(1)	(2)	(3)	(4)
Treated × Post	-0.185^{***}	0.719^{***}	-0.184^{**}	0.060
	(0.052)	(0.179)	(0.068)	(0.077)
date f.e.	Yes	Yes	Yes	Yes
cusip f.e.	Yes	Yes	Yes	Yes
Observations	450,648	87,768	200,520	134,928
R^2	0.232	0.175	0.138	0.165
Adjusted R^2	0.210	0.151	0.113	0.141

Note: *p<0.1; **p<0.05; ***p<0.01
double clustered standard errors

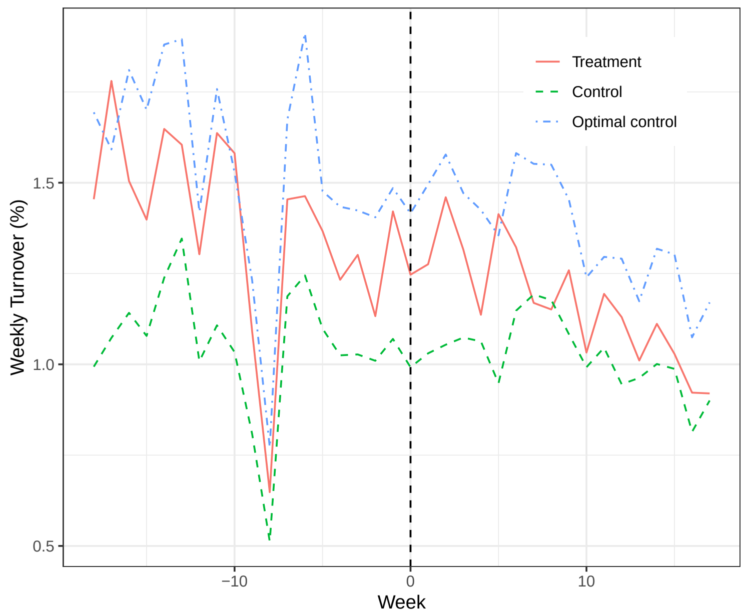
4.4 Parallel trends forest results

We construct a parallel trends forest with 1,000 trees based on pre-treatment data to estimate the weights. We then use those weights to construct the optimal control ($\hat{Y}_{i,t}^{(0)}$) for each treated bond i. When building the trees, we split until each leaf contains less than 100 bonds, but we do not restrict the leaves to have both treated and control bonds. We also allow the algorithm to split on ratings and bond issue size, which are the variables used to determine treatment assignment, so it is theoretically possible to have treated bonds that do not have any corresponding optimal control sample if none of the control bonds behave similarly; but in this sample that does not seem to be the case.

Figure 7 presents the average weekly turnover for the treated bonds, control bonds, and the optimal control sample. The optimal control sample follows the treated sample quite closely with a constant shift albeit not perfectly, especially compared to the average control sample, which behaves very differently from the treatment sample. The fact that the optimal control and treatment samples track each other quite well during the pre-treatment period indicates that the algorithm does quite well at selecting the weights in-sample. The two series track each other quite closely post-treatment, and there are no discontinuities around the treatment date, which implies that treatment effect is likely small.

Table 2 shows the composition of the optimal control by presenting the sum of $\bar{w}_j$ for the various control

Figure 7: **Parallel trends forest**
This figure plots the average weekly turnover for treated bonds, control bonds, and the optimal control sample derived using parallel trends forest. Parallel trends forest uses 1,000 trees.



samples used in the TWFE estimates. The optimal control sample, compared to the outstanding numbers, overweighs the Transparent & $\geq$ A bonds and the BBB 100M–1B bonds. Panel (b) shows that the optimal control sample matches the characteristics of the treatment sample more closely than the average control sample does.

Table 2: **Composition of Optimal Control Sample**

(a) Weights allocated to each subsample

	Transparent & $\geq$A	Not Transparent $\geq$A	BBB, 100M-1B	Other control
Sum(weight)	8.1%	26%	42.5%	23.5%
Share by count	2.3%	32.6%	15%	50.1%

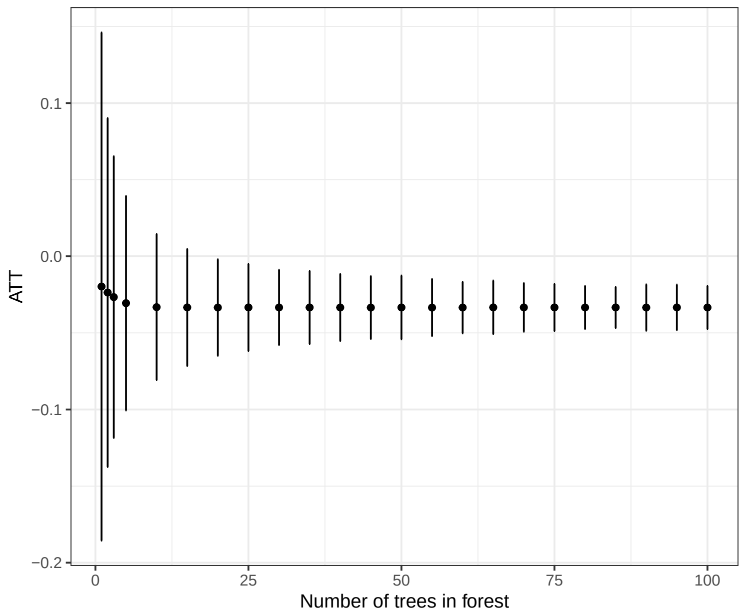
(b) Average characteristics of treated, optimal control, and control samples

Sample	Rating	Outstanding amt	Pre-sample turnover	Age	Time-to-maturity
Treatment	5.413	259	0.345	65.144	91.273
Optimal control	8.333	334	0.344	61.373	95.706
Control	9.222	150	0.242	57.584	80.050

We then look at how many trees we would need for our ATT estimates to be reasonably accurate. Using

4,000 trees, we plot the distribution of ATT estimates against the number of trees in each forest in Figure 8. Figure 8 indicates that the variance of ATT estimates are reasonably small with more than 50 trees. For instance, a forest with 100 trees would give an ATT estimate between -0.019 and -0.047 95% of the time. While estimating the standard error for parallel trends forest estimate in an analytical way is outside of the scope of this paper, our estimate constructed from 1,000 trees would have a small standard error.

Figure 8: **Convergence of parallel trends forest**
The following figure plots the average and 95% confidence interval of ATT estimate against the number of trees used in parallel trends forest construction. The dots indicate the average ATT estimate, and the lines indicate the 95% confidence interval calculated using the average and standard deviation of ATT estimates for the given number of trees.



The source of uncertainty calculated in the standard error of the parallel trends forest estimate mentioned above is about how accurately we can estimate the weights of the "most" optimal control that has the closest parallel trends to the treated sample. The results in Figure 8 indicate that this uncertainty is fairly low with a reasonable number of trees. The low number of trees needed for a reasonable estimate means that large computational resources are not necessary to employ parallel trends forest.

However, there are two other potential sources of uncertainty: how closely parallel trends hold and the uncertainties due to random sampling from a larger population.[10] In Figure 7, the treated and optimal control sample does not track each other exactly in the pre-treatment period. Some of the deviation could be

[10]Most empirical work in economics literature assume that researchers observe some random sample of a larger population. See Abadie et al. (2020) for discussion on sampling-based versus design-based uncertainties.

from inaccurate weights, but at least some of the deviation is because the parallel trends do not hold exactly. We deal with this issue by allowing for violation of parallel trends using the approach from Rambachan and Roth (2023).

We assume that the average deviation from parallel trends between the post and pre-treatment period is less than L times the average deviation from parallel trends between the first and the second half of the pre-treatment period. More specifically, denote β_t as the difference in Y on time t between the average treated bond and the average counterfactual derived from the optimal control sample, which can be estimated from the regression

$$Y_{i,t} - \hat{Y}_{i,t}^{(0)} = \beta_t + \epsilon_{i,t}. \tag{16}$$

Then we assume that

$$\left| \frac{1}{T - T_1} \sum_{t=T_1+1}^{T} \beta_t - \frac{1}{T_1} \sum_{t=1}^{T_1} \beta_t \right| \leq L \left| \frac{2}{T_1} \sum_{t=T_1/2+1}^{T_1} \beta_t - \frac{2}{T_1} \sum_{t=1}^{T_1/2} \beta_t \right|. \tag{17}$$

For brevity, we write (17) assuming T_1 is an even number. We do not use our definition of the deviation from parallel trends (equation (5)) here because it unfortunately does not satisfy the assumptions we need to use Rambachan and Roth (2023), but it has some similarities in that we use $|c(Y_i, Y_j, T_1/2)|$ from (4) as measure of deviation.

Table 3 presents the results for various values of L. The confidence intervals calculated here incorporate uncertainties arising from both deviations from parallel trends and random sampling, but assumes that the optimal control sample estimation is exact. The estimate for ATT is -0.032, smaller in magnitude compared to all the TWFE estimates in Table 1. This indicates that much of the large ATT estimate is capturing the deviation of parallel trends rather than true effect of treatment. In the case of $L = 0$, that is if parallel trends hold perfectly, the ATT estimate is almost statistically significant at the 10% level but otherwise the ATT estimate is not significant for other values of L. Overall, we conclude that the treatment effect is fairly small and not statistically significantly different from zero for reasonable values of L.

One may ask—if we still have to allow for violation from parallel trends, why do we bother with trying to find the optimal control? We can still get much more accurate estimates from having an optimal control in which the deviation from parallel trends is much smaller.

Table 3: **Allowing for violation of parallel trends**
The following table presents the 90% confidence interval for various values of L. The ATT estimate remains constant.

L	Estimate	Lower Bound	Upper Bound
0	-0.032	-0.0643	0.0002
0.5	-0.032	-0.0794	0.0147
1	-0.032	-0.1069	0.0422
1.5	-0.032	-0.1363	0.0716
2	-0.032	-0.1677	0.103

4.5 Parallel trends forest with honesty

So far in our parallel trends forest, we select the optimal control sample such that parallel trends hold as closely as possible in the pre-treatment period. However, this does not necessarily guarantee that parallel trends will hold in the post-treatment period when there is no treatment; in general, it is impossible to construct an optimal control sample that we know for sure that moves in parallel with the treated sample in the counterfactual no-treatment world. Here we adopt a technique from Wager and Athey (2018) and Athey et al. (2019) and construct a parallel trends forest that is "honest" in order to reduce bias from overfitting in the pre-treatment period.

Construction of the forest is similar to that outlined in Section 2 but with some important differences. When constructing the tree b, we first divide the randomly-sampled K units further into two groups arbitrarily. We grow the tree on the first subsample, then put the second subsample into the tree to generate the weights $w_{b,j}(i)$. Because weights are calculated using out-of-sample data, if parallel trends do not hold well out-of-sample, the optimal control constructed from parallel trends forest with honesty would behave very differently from the treatment sample even in the pre-treatment period.

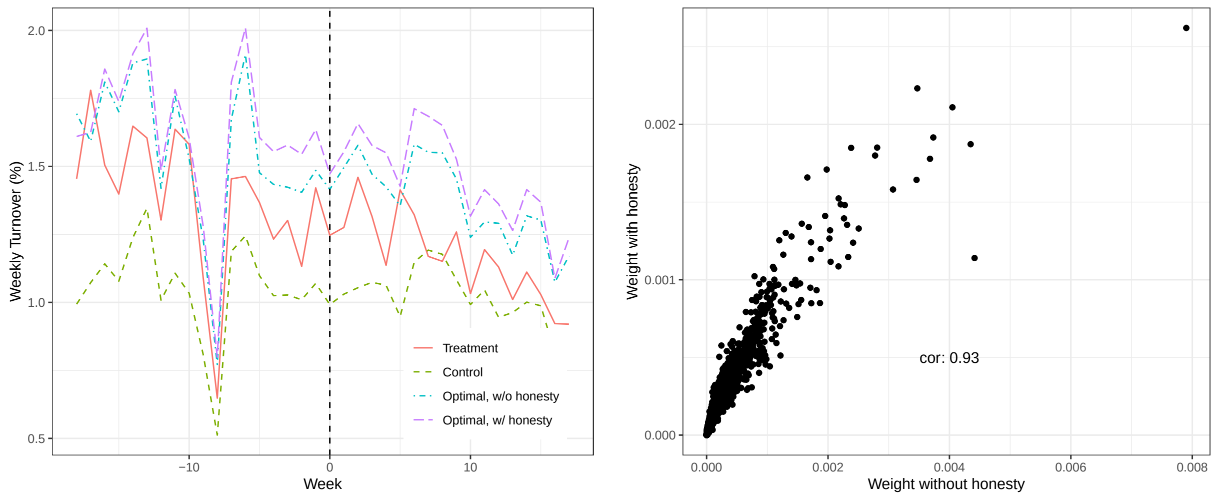
Figure 9 presents the results for the parallel trends forest with honesty. Panel (a) presents the average turnover values for the treatment sample, control sample, optimal control sample calculated from parallel trends forest without honesty (from Section 4.4), and optimal control sample calculated from parallel trends forest with honesty. The average turnover for the optimal control with honesty tracks the version without honesty quite closely, and both move with the treatment sample fairly closely. The optimal control with honesty fits the treatment sample slightly less closely compared to the version without honesty, especially in the pre-treatment period, which is to be expected since version without honesty uses out-of-sample data to calculate the weights.

Panel (b) of Figure 9 compares $\bar{w}_j = \frac{1}{N}\sum_{i=M+1}^{N} w_j(i)$, where $\bar{w}_j$ represents the average weight assigned

Figure 9: **Parallel trends forest with honesty**
The following figures show the results using parallel trends forest with honesty. Panel (a) shows the average turnover for the treatment sample, control sample, optimal control sample derived from parallel trends forest without honesty, and optimal control sample derived from parallel trends forest with honesty. Panel (b) plots $\bar{w}_j$, which represents the average weight assigned to control bond j across all treated bonds, obtained from parallel trends forest without honesty against that obtained from parallel trends forest with honesty.

(a) Time series

(b) Weights for parallel forest with and without honesty



to control bond j across all treated bonds, obtained from parallel trends forest with and without honesty. The two weights are highly correlated with a correlation of 0.93. Weights obtained from parallel trends forest without honesty tends to have a fatter right tail, which comes from fitting the data closely in-sample. Overall, given the high correlation in the optimal control between the two methods, we can conclude that in this use case, overfitting by the parallel trends forest is fairly mild.

We present ATT estimates that use the optimal control from parallel trends with honesty in Table 4. Similar to Table 3, we follow Rambachan and Roth (2023) and allow for deviations from parallel trends in the form of (17). The ATT estimate is slightly larger in magnitude but statistically significant only when $L = 0$.

Table 4: **ATT estimates using parallel trends with honesty**
The following table presents the 90% confidence interval ATT estimated using parallel trends with honesty for various values of L, where L denotes the maximum degree of deviation from parallel trends.

L	Estimate	Lower Bound	Upper Bound
0	-0.042	-0.075	-0.01
0.5	-0.042	-0.127	0.04
1	-0.042	-0.194	0.107
1.5	-0.042	-0.263	0.174
2	-0.042	-0.332	0.245

5 Conclusion

In this paper, we propose a novel method to select the optimal control sample that satisfies the parallel trends assumption as closely as possible. This parallel trends forest method is useful for natural experiments in which treatment assignment contains very little randomness but has a very large candidate of control samples and an observable long pre-treatment period that can be used for selection. We show using the introduction of post-trade transparency in the corporate bond market that in some settings our method works better than other existing data-driven methods as well as the usual TWFE estimator. It seems to be the case that the granular, non-normal distribution of the outcome variables plays a key role in why parallel trends forest outperforms other data-driven methods by a large margin in our use case. It would be interesting to study further in future work.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020): "Sampling-based versus design-based uncertainty in regression analysis," *Econometrica*, 88, 265–296.

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *The Journal of the American Statistical Association*, 105, 493–505.

Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," *The American Economic Review*, 93, 113–132.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021): "Synthetic Difference-in-Differences," *The American Economic Review*, 111, 4088–4118.

Asquith, P., T. Covert, and P. Pathak (2019): "The Effects of Mandatory Transparency in Financial Market Design: Evidence from the Corporate Bond Market," *Working Paper*.

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

Athey, S., J. Tibshirani, and S. Wager (2019): "Generalized Random Forests," *The Annals of Statistics*, 47, 1148–1178.

Baker, A., B. Callaway, S. Cunningham, A. Goodman-Bacon, and P. H. Sant'Anna (2025): "Difference-in-Differences Designs: A Practitioner's Guide," *forthcoming, Journal of Economic Literature*.

Bessembinder, H. and W. Maxwell (2008): "Markets: Transparency and the Corporate Bond Market," *Journal of Economic Perspectives*, 22, 217–234.

Bessembinder, H., W. Maxwell, and K. Venkataraman (2006): "Market Transparency, Liquidity Externalities, and Institutional Trading Costs in Corporate Bonds," *Journal of Financial Economics*, 82, 251–288.

Breiman, L. (2001): "Random forests," *Machine learning*, 45, 5–32.

Bühlmann, P. and B. Yu (2002): "Analyzing Bagging," *The Annals of Statistics*, 30, 927–961.

Doudchenko, N. and G. W. Imbens (2016): "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis," *National Bureau of Economic Research Working Paper*.

Edwards, A. K., L. E. Harris, and M. S. Piwowar (2007): "Corporate Bond Market Transaction Costs and Transparency," *The Journal of Finance*, 62, 1421–1451.

Goldstein, M. A., E. S. Hotchkiss, and E. R. Sirri (2007): "Transparency and Liquidity: A Controlled Experiment on Corporate Bonds," *Review of Financial Studies*, 20, 235–273.

Goodman-Bacon, A. (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225, 254–277.

He, G. and S. Wang (2017): "Do College Graduates Serving as Village Officials Help Rural China?" *American Economic Journal: Applied Economics*, 9, 186–215.

Rambachan, A. and J. Roth (2023): "A More Credible Approach to Parallel Trends," *The Review of Economic Studies*, 90, 2555–2591.

Roth, J. (2022): "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, 4, 305–322.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023): "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *Journal of Econometrics*, 235, 2218–2244.

Rubin, D. B. (2008): "For Objective Causal Inference, Design Trumps Analysis," *The Annals of Applied Statistics*, 808–840.

Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn (2007): "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution," *BMC Bioinformatics*, 8, 25.

Wager, S. and S. Athey (2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113, 1228–1242.