Effects of Family Disruption on Child Development: The Moderating Role of Residential Relocation*

Lucienne N.Y. Disch†
University of Pennsylvania

November 27, 2024

Abstract

This paper studies the consequences of family disruption and associated change of residence for human capital formation. I exploit variations in family stability that arise from changes in household composition due to the father's initial presence and subsequent absence. Using a dynamic within-child difference-in-differences approach, I compare longitudinal test scores of children who experience family disruption to those of children who have not yet been affected. Consistent with prior research – which often focuses only on married couples and uses the legal date of divorce as the point of separation – I find that, on average, family disruption leads to moderate but significant declines in test scores. However, I highlight that residential relocation emerges as a key factor in the context of family disruption. In the United States, 38% of children whose parents separate have to relocate, and 82% of those move more than a mile away. Using confidential geocoded NLSY data, I demonstrate that, on average, children who relocate to a new residence due to family disruption experience significant declines in school performance, particularly those who move more than a mile from their original home. In contrast, children who remain in their current residence or relocate within the same neighborhood exhibit less pronounced declines following family disruption. These findings indicate that the act of relocating, rather than family disruption itself, is the primary factor contributing to the observed test score gap. Consequently, targeted policies – such as assisting newly single mothers and their children in remaining within their familiar residential areas for at least three years following separation – could mitigate the negative consequences of long-distance moves on children's school performance.

*I am indebted to Francesco Agostinelli, Jere Behrman, Margaux Luflade, and Petra Todd for many fruitful discussions and suggestions. For helpful comments, I thank my discussant Susan Clampet-Lundquist, as well as seminar and conference participants at the University of Pennsylvania, the SEA 2023 conference at the University of Zurich, and the 2023-2024 Urban Studies Graduate Colloquium.

†University of Pennsylvania, Department of Economics. Email: brunnerl@sas.upenn.edu.

1 Introduction

Family is crucial for child development, but also the place growing up including its peers (Agostinelli, Doepke, Sorrenti, and Zilibotti, 2024). Family disruption can have an adverse impact on human capital formation. Striking, I find that 38% of children whose parents separate have to residentially relocate, and 82% of those move more than a mile away.[1] Father gone, friends gone? Little is known about the effects of family disruption on child development and what role residential relocation plays. What are the causal effects of parental separation on test scores and does it matter whether and how far away children have to move as a consequence?

To answer these questions, I propose a geospatial approach to identify the effects of family disruption. I exploit variation in family stability induced by changes in household composition due to initial presence and subsequent absence of the father. Equipped with a sequence of consecutive test score performances for both children experiencing family disruption (treatment group) and not-yet affected children staying in a two-parent household at the compared point in time (control group), I use an event study design, a staggered difference-in-differences model, to evaluate the causal effect of physical separation of the parents on the child's test scores.

In line with previous research, I find that family disruption has moderate but significant adverse effects on test scores. I highlight an important finding in the context of family disruption, which is *residential relocation* acting as a moderating variable. I suggest that the relationship between parental separation and test score performance depends on whether there is an associated move. On average, the phenomenon of moving contributes to the test score gap, rather than family disruption by itself. More specifically, the main transmission channel appears to work through the *distance* of the move. If they move not more than a mile away, test scores are less affected. In fact, staying mitigates – if not offsets – the adverse effects of family disruption. Consistent

[1] Calculations are based on a sample of school-aged children in the NLSY Geocode Data, the National Longitudinal Survey of Youth 1979 (NLSY79) confidential spatial data files from the U.S. Bureau of Labor Statistics (BLS).

with that, the further away from the familiar place of residence children have to move due to parental separation, the more adversely they are affected (although at a certain threshold, the distance does not matter anymore).

From a *policy perspective*, my results that the disadvantageous impact on test scores is concentrated among movers relocating to a new area, suggest that prevention can be done more easily to mitigate the adverse effects of family disruption (because intervention into the family would be harder and potentially counterproductive). Targeted policies, such as helping newly single mothers and their children remain in their familiar residential area for at least three years following separation, could mitigate the negative consequences of long-distance moves on children's school performance.

This paper is related to a growing literature studying the effects of social disruption, such as parental separation and residential relocation. While there is mounting evidence on the importance of the family background[2] and the power of place[3] for child development, less is known about the disruptive impact of a change in the social environment.[4] A key insight in this literature is that disruption occurrences are not random (e.g., Marinescu (2016); Greenwood et al. (2016)). For example, those families experiencing family disruption have different characteristics than more stable ones. Therefore, it is important to account for potential selection into "treatment" of family disruption. Because, when measuring the relationship between family disruption and outcome variables such as test scores, models can either overstate or understate the effect if there is an omitted variable that is both correlated with parental separation and is a determinant of test scores. In the last years, the literature has made substantial progress in measuring the causal effect of family disruption on the child's outcomes in the presence of sorting, identified with the help of changes in marital status and dynamic structural

[2]E.g., Cunha and Heckman (2007); Cunha and Heckman (2008); Agostinelli et al. (2024); Agostinelli and Wiswall (2024), among many others.

[3]E.g., Chetty et al. (2016); Chyn (2018), among many others.

[4]E.g., Chan and Liu (2022); Cordes et al. (2019), Chetty and Hendren (2018a), Selya et al. (2016), Tartari (2015).

modeling (e.g., Tartari (2015)), survey data and a control function approach (Chan and Liu, 2022), or administrative data and an event study analysis (Nielsen et al., 2024). A difficult problem in this context is the identification of the date of separation. Tartari (2015) addressed this issue by using the legal date of divorce. Chan and Liu (2022) and Nielsen et al. (2024) focus on changes in family structure based on survey and administrative data, respectively. The idea is that variations in family composition should be reflected in the living situations of both parents. The dissolution of a shared household typically includes a new residence for at least either the father or the mother-child pair.[5] These circumstances can lead to reduced financial and time investments in the child, which may negatively affect the child in ways that extend beyond changes within the family environment. A number of studies investigate changes in financial resources available to the child and mother as well as engagement of the father with the child after separating from the family (Del Boca et al. (2014); Chan and Liu (2022); Kalil et al. (2011)). Since family disruption may be accompanied with a change of the place of residence, this could also imply a change in the social environment (e.g., a new school and peers for the child) or being closer in proximity to supportive grandparents, for example. The challenge lies in the fact that transitions in family composition encompass numerous mechanisms that can influence the child, extending beyond mere changes in the household environment. This complexity underscores the difficulty in disentangling or causally identifying the underlying causes and consequences of family disruption on the child's outcomes.

This paper contributes to the literature by introducing a new source of information and a novel identification strategy that illuminate the role of geographic factors in family disruption and their impact on children. While I use survey information

[5]Since the NLSY handles the mother and the child as pairs, and treats the father separately, the underlying dataset does not allow me to account for (a significant number of) cases where the child lives with the father after parental separation. Further, family disruption typically incorporates for parents a transition from a shared household to two separate ones – implying that the father is still around, but just not living with the rest of the family anymore. In some cases, however, family disruption happens due to more involuntary incidents such as imprisonment or death of the father. I consider both scenarios separately as well as unified.

regarding the variation in household composition to identify family disruption events during school years when children take the standardized PIAT tests every other year, I rely on geocoded microdata to exploit geospatial variation in the place of residence to further identify the geography related to the family disruption. While I do not model the exact longitudinal and latitudinal geolocation explicitly, I show that proximity between the old and new residence contains valuable information for identification.[6] Living situations turn out to be strong mediators for the effect of family disruption on test scores. This is relevant as other proxies for family disruption, including Tartari (2015)'s approach of taking the legal date of divorce or Nielsen et al. (2024)'s survey usage, respectively, have been found to not identify the exact date of separation and exclude cohabitants, or ignore the potential changing living situation.

From a methodological perspective, my approach is closely related to event study methods aiming to evaluate the impact of interventions or events. In particular, I take advantage of a staggered difference-in-differences (DID) model with multiple time periods, to estimate the causal effect of family disruption on test scores. Roth et al. (2023) provide an overview of the new advances in staggered DID models. An important challenge in measuring the effect of family disruption is defining the appropriate counterfactual. The issue lies in the fact that stable two-parent households, which are unlikely to experience disruption, may not provide an accurate control group, as they may not face the same levels of tension or struggle as households that are on the verge of disruption. However, to accurately estimate the causal effect of the transition from a two-parent to a single-parent household, it is essential that the counterfactual is realistic.[7] Thus, I compare the sequence of test scores for children experiencing family disruption to those of children from similar families – who share the same environment (e.g., levels of tension) – but do not yet experience disruption at the time of

[6]The data allows me to identify their domiciles at the county level. While I also know whether they changed the ZIP codes when moving and the exact distance they moved, I cannot infer whether that includes a change in school zones or the like.

[7]To further ensure that the effect is causal, I am also reducing my sample to a *balanced* panel so that I am tracing the exact same individuals over time with no missing information for any year.

the comparison. Related to this, I pay attention to how trends of control and treatment groups are before the disruption event, to make sure they are parallel. This will also reveal potential anticipation effects affecting test scores in the immediate year(s) before family disruption. To account for potential pre-trends, anticipation, and the right counterfactual, Callaway and Sant'Anna (2021) propose a staggered DID estimator that essentially takes an average over multiple static 2x2 DIDs and offers to take not-yet affected individuals at the compared point in time as the control group. The key idea of this paper is to apply this approach to the child development literature, exploiting institutional features of PIAT and combining survey data with restricted geocoded spatial data. I also implement the staggered DID approach by Wooldridge (2021) as robustness check. Instead of most related research – that relies on the legal date of divorce, which excludes cohabiting parents splitting and often happens only several months or years after the actual physical and mental separation, or never at all – my spatial identification strategy allows me to shed light on those parents who transition from a shared household to two separate ones. This paper is not the first to apply such a difference-in-differences approach to investigate the effects of a transition of a two-parent household to a single-parent household on test scores (Nielsen et al. (2024)). To the best of my knowledge, however, this paper is the first to go beyond survey information and beyond just looking at the effects of parental separation per se, by looking at the geography of the living situation in association with family disruption.

My results indicate that the geographical living situation of parents in association with a family disruption can have a significant impact on test scores. The unified consideration of family disruption and residential relocation reveals novel evidence of an important mechanism behind the adverse effects of parental separation on academic performance, which is (the distance of) moving, that acts as the moderator in this context. Staying in the same residential area after parental separation could mitigate – even outweigh – the adverse impact of family disruption on test scores. While children are adversely impacted by a transition to a new place of residence if it is associated with family disruption, children of stable parents, instead, do better after a move than before. Thus, effects of the move depend on the reason why families move, confirming that mobility is endogenous. In this sense, I also contribute to the literature on the role of residential relocation on children's outcomes (Chetty and Hendren (2018a); Chetty and Hendren (2018b); Chetty et al. (2016)). Overall, my spatial identification strategy reveals novel evidence of an important mechanism behind the adverse effects of parental separation on academic performance: my findings highlight the importance of accounting for residential relocation in association with family disruption, as solely focusing on the general effects of family disruption can overstate the actual impact, which is crucial for (targeted) policy interventions.

Overview. The paper proceeds as follows. In the next section, I talk about the data, provide institutional background on the standardized test scores, and discuss the geospatial identification strategy. Section 3 covers the econometric approach and Section 4 presents the results on the effects of family disruption. Section 5 looks into the mechanism by distinguishing movers from stayers. Section 6 provides a couple of robustness checks and Section 7 concludes.

2 Identification Strategy

Measuring parental separation is inherently challenging. While the legal date of divorce is a common metric, it is insufficient for at least two reasons. First, divorce may occur years after parents have emotionally and physically separated – or it may not happen at all. Second, focusing solely on divorce dates excludes cohabiting families who were never married and overlooks more "involuntary" family disruptions caused by imprisonment or the death of a parent. I propose a geospatial identification strategy. I follow school-aged children who take standardized tests every other year while simultaneously tracking their living situations – including family composition and residential stability. This approach allows me to identify family disruptions and residential relocation events, which can serve as indicators of shocks in a child's life that may impact their test score performance. In this chapter, I provide data sources, background information on the standardized test scores, the measurement of family disruption, the tracking of children's living situations, and the construction of the panel data frame.

2.1 Data Sources and Measurement of Standardized Test Scores

Data sources. The main data source for this project consists of panel data from the National Longitudinal Survey of Youth (NLSY) by the U.S. Bureau of Labor Statistics (BLS).[8] In particular, I base my analysis on the NLSY cohort of Americans born between $1957-64$ (NLSY79). The cohort originally included $12,686$ respondents ages $14-22$ when first interviewed in 1979 and data are now available from Round 1 (1979 survey year) to Round 29 (2020 survey year). I link the women in the NLSY79 with their biological children in the NLSY79 Children and Young Adults cohort (NLSY79 Child/YA). To date, a total of $11,545$ children have been identified as born to interviewed NLSY79 mothers. Data are now available from 1986 to 2018, representing 17 survey rounds for the child sample and 13 for young adults in that time span. I further match these linked files with the corresponding confidential NLSY79 and NLSY79 Child/YA Geocode data. The latter provide information collected about the respondent's geographic location.

These geocoded files reveal information such as from which county to which county a person moved. They also provide information regarding the geographical distance of the places of residence between two consecutive interview rounds at a more granular level. While the confidential Geocode data files do not reveal the identified ZIP codes where people live, they do provide information regarding whether they changed the ZIP or not between two survey years. On top of that, the distance between two places of residence between two interview rounds is reported more specifically (0 miles (non-mover), $0-999$ feet, 1000 feet-1 mile, $1-5$ miles, ..., 500+ miles). The data also allows

[8] For more information and details, consult the NLSY website of the U.S. Bureau of Labor Statistics: https://www.nlsinfo.org/content/cohorts, accessed November 2024.

me to infer the identification of those moving abroad. While latitudes and longitudes are generated from exact address matches in most cases, the migration distances were calculated based on the ZIP centroid if an address was invalid.

The combination of the survey and geospatial matched data from NLSY allows me to obtain a number of characteristics defining a child's living situation every other year – consisting of information where these children are located at the county level; whether they moved to a new place of residence between two survey waves; and conditional on moving, the exact geographical distance between the old and the new place of residence; whether the father lives in the same household or not; his engagement with the child; and the geographical distance between the father and the child in case of a split of a two-parent household into two one-parent households. In addition, the matched data file consists of school-aged children for whom I observe their biennial standardized test score performances for at least three consecutive periods. These tests are the main outcome variables to measure the effects of disruption, as I will explain next.

Standardized Test Scores. Key measurement variables in this paper are the raw Peabody Individual Achievement Test (PIAT) test scores. The PIAT test is one of the most popular used assessments of academic achievements in the United States, primarily thanks to its high test-retest reliability and concurrent validity. Children take these standardized tests every other year during K-12 (between five and 22 years of age). I focus on the three most widely used sub-assessments of the PIAT, which are available 1986–2020 in the NLSY data files: Mathematics, Reading Recognition, and Reading Comprehension. All children take the same PIAT tests in a particular year. Relatively older children are expected to score higher than younger ones, although there are some adjustments in the difficulty depending on which grade the child is enrolled in, which I will highlight in more detail when describing the PIAT Reading Recognition sub-test. However, given that the baseline test is the same irrespective of the age of the child, this also implies that children typically improve their performance in the tests as they get older throughout their school years. A detailed explanation of what these assessments contain in evaluating a child's skills in both reading and mathematics can be found in *Appendix A*, as well as on the NLSY website of the U.S. Bureau of Labor Statistics.[9].

The PIAT consists of many institutional features that are very advantageous and appealing when comparative studies of academic performance among children are desirable. Given that all children take the same baseline standardized test every other year during K-12, irrespective of their age, school attended, and neighborhood they live in, it is very comparable, for example, across different children of similar age. Not only from a cross-sectional viewpoint, but also from a longitudinal one, as it allows to compare different trends in raw test scores throughout life across children. Children take these PIAT tests up to five times in their life, so I observe at most five consecutive biennial test scores performances for the same child in my sample. Thus, the longitudinal sample period does not exceed ten years.

Transformation of time domain. Without loss of generality, I transform the years in which I observe test scores into $1, 2, 3, 4, 5$ for every child. This transformation does not affect the validity of the technique or results, but the reason I am doing this is to gain more power, and thus precision, when comparing children whose parents separate in a particular year to not-yet affected children staying in a two-parent household during the compared point in time. During the investigated time period, $1986 - 2014$, I would need to have a sufficiently large number of families disrupting in every single year, so that the number of treated units is a large enough sample size and can be compared to the larger control group. Transforming the time domain essentially guarantees a decrease in the number of parameters without any loss of generality. In addition, The panel data becomes more concise and well-structured when the time period is consistently defined as "Year 1" to "Year 5" for each child.

[9]https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/assessments/piat-reading-reading-recognitionreading, accessed November, 2024.

2.2 Measuring Family Disruption

To identify family disruption events, I employ a panel data approach by tracking school-aged children over ten years, for whom I have biennial PIAT test scores. I include only those children who are living in two-parent households at the time of their initial entry into the panel, which is also when they take their first of five PIAT tests. To track children who transition from a two-parent household to a single-parent household over the ten-year period, I examine changes in family composition. Based on this analysis, I create two additional indicator variables, defined below.

The first variable, $FamDisrup_i$, indicates whether and when child i experienced a family disruption event. For children who experienced disruption, $FamDisrup_i$ takes on the values 2, 3, 4, or 5, corresponding to the year of disruption. For those who did not experience disruption, $FamDisrup_i = 0$. "Year 1" serves as the reference or baseline period where all children are living in two-parent households, so no disruptions occur then (i.e., $Fam.Disrup_i \neq 1$). This setup provides four consecutive years (Years 2 to 5) during which a disruption event can occur. I group all children who experience disruption in the same year into the same group (or "cohort").

The second indicator variable, $Treat.Fam_i$, is a binary variable indicating whether child i experienced family disruption at any point during the ten-year period. Specifically, $Treat.Fam_i = 1$ if the child experienced a disruption, and $Treat.Fam_i = 0$ if the child remained in a two-parent household throughout the entire period.

Definition 1 (Family Disruption Start):

$$Fam.Disrup_i = \begin{cases} 2 & \text{if} \quad (i_{t=1} \in \text{TP}) \cap (i_{t=2} \in \text{SP}) \\ 3 & \text{if} \quad (i_{t=1} \in \text{TP}) \cap (i_{t=2} \in \text{TP}) \cap (i_{t=3} \in \text{SP}) \\ 4 & \text{if} \quad (i_{t=1} \in \text{TP}) \cap (i_{t=2} \in \text{TP}) \cap (i_{t=3} \in \text{TP}) \cap (i_{t=4} \in \text{SP}) \\ 5 & \text{if} \quad (i_{t=1} \in \text{TP}) \cap (i_{t=2} \in \text{TP}) \cap (i_{t=3} \in \text{TP}) \cap (i_{t=4} \in \text{TP}) \cap (i_{t=5} \in \text{SP}) \\ 0 & \text{if} \quad (i_{t=1} \in \text{TP}) \cap (i_{t=2} \in \text{TP}) \cap (i_{t=3} \in \text{TP}) \cap (i_{t=4} \in \text{TP}) \cap (i_{t=5} \in \text{TP}), \end{cases}$$

where TP stands for "Two-Parent-Household" and SP stands for "Single-Parent-Household".

Definition 2 (Family Disruption Event):

$$Treat.Fam_i = \begin{cases} 1 & \text{if} \quad Fam.Disrup_i \in [2,3,4,5] \\ 0 & \text{otherwise.} \end{cases}$$

I can then construct a panel data frame as shown in Table 1, that captures all the relevant data for my analysis. In particular, I have one row per child (ID) per time period ($Year$), along with the child's standardized PIAT test scores ($Test.Scores$) for Year 1 to Year 5. Note that, even though I call them "Year 1", "Year 2", etc., there is a two-year gap between two such time periods, because PIAT tests are only taken every other year. I further have my two indicator variables that inform whether there was a potential family disruption event ($Treat.Fam$) and in what year ($Fam.Disrup$). Lastly, I include time-invariant control variables such as an age fixed effect (Age.FE), race fixed effect, a fixed effect for both the mother's AFQT score (AFQT.Mom) and the mother's years of education (Educ.Mom), and a siblings fixed effect, indicating whether the child has siblings or not.

To ensure consistency in data and facilitate meaningful comparisons, I restrict the analysis to children who consistently live with their biological mother. In this setup, any transition to a "single-parent" household is defined as either the biological father leaving the household or the mother and child moving out to form a new household. Additionally, the sample is further narrowed to children who live with their biological father during the first period, excluding cases where a stepfather or another individual is present in the household. This focus is partly due to data limitations, as the dataset does not provide sufficient observations of single-parent fathers, among other scenarios.

Table 1: Illustration of the cleaned input data for the analysis at hand

ID	Year	Test.Scores	Fam.Disrup	Treat.Fam	Age.FE	Race	AFQT.Mom	Educ.Mom	Siblings
244	1	66	2	1	8	0	1	16	1
244	2	45	2	1	8	0	1	16	1
244	3	41	2	1	8	0	1	16	1
244	4	38	2	1	8	0	1	16	1
244	5	62	2	1	8	0	1	16	1
817	1	88	0	0	6	1	4	12	0
817	2	93	0	0	6	1	4	12	0
817	3	79	0	0	6	1	4	12	0
817	4	98	0	0	6	1	4	12	0
817	5	99	0	0	6	1	4	12	0
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

2.3 Measuring Residential Relocation

In my baseline event study, I leverage the full available time period and exclude children with fewer than five consecutive years of test score observations. This approach provides a comprehensive overview of how test scores evolve in the years before and after a potential family disruption treatment. However, for analyses that account for both family disruption and residential relocation, I lower the threshold to three consecutive test score observations. This adjustment increases statistical power, as the interaction between parental separation and residential relocation significantly reduces the number of treated children in the sample when using the stricter five-year criterion.

The identification and definition of residential relocation follow a similar approach to that used for family disruption: I construct a variable $Move_i \in [2, 3, 4, 5, 0]$ and an additional variable $Treat.Move_i \in [1, 0]$. To capture the combined experience of family disruption and relocation, I create unified measures. These include $Treat_i \in [2, 3, 4, 5]$ and $First.Treat_i = 1$, identifying children who experience both events. Conversely, children who remain in two-parent households and the same residence throughout the sample period are characterized by $Treat_i = 0$ and $First.Treat_i = 0$.

2.4 Diagnostics of the Counterfactual (Control Group)

A critical aspect of analyzing the effects of disruption events on children's outcomes is selecting an appropriate counterfactual or comparison group, as this decision fundamentally shapes the interpretation of the causal effect. Comparing families undergoing disruption to very stable two-parent households can lead to biased estimates of the causal effect. Families experiencing disruption often face pre-existing tension in the years leading up to the event. However, comparable levels of tension may also exist in families that ultimately do not separate. Overlooking this nuance risks conflating the effects of family tension with the causal impact of the disruption itself. Another key challenge is sorting—families with certain underlying characteristics may be more prone to experiencing disruption, leading to selection into treatment. If not addressed, this selection process can distort the validity of comparisons. To mitigate these concerns, my baseline analysis uses a control group consisting of individuals who have not yet experienced family disruption. This approach ensures greater similarity in pre-treatment characteristics and family dynamics between treated and control groups, providing a more robust basis for causal inference. To address concerns regarding the selection of an appropriate comparison group, I will also present results from a sensitivity analysis in which the control group is restricted to individuals who never experience family disruption. This approach provides additional insight into whether children in families that will eventually undergo disruption may already be exposed to pre-existing tension in the years prior, which could negatively impact their test scores.

3 Descriptive Evidence

In this chapter, I begin by presenting summary statistics for the outcome variable, *test scores*. I then conduct an initial analysis of the relationship between social disruption events and test score performance using descriptive ordinary least squares (OLS) regression models.

3.1 Summary Statistics

Table 2 presents detailed summary statistics on the evolution of test score performance for children with at least three years of observations. The sample includes all children in the dataset, encompassing both those who experienced a disruption during the observed period and those who did not.

Table 2: Summary Statistics

Obs	Min	Max	Avg	Std
Year 1	0	67	21.8	9.6
Year 2	0	78	36.3	11.1
Year 3	0	80	46.1	10.7

Reading Comprehension

Obs	Min	Max	Avg	Std
Year 1	0	64	21.1	10.3
Year 2	6	84	37.3	11.9
Year 3	0	84	49.4	10.5

Mathematics

In my sample at hand, the probability that children experience family disruption in a given year is 12%. This number is higher for Black (21%) than for non-Black children (11%). Note that there are almost five times more non-Black than Black children in my sample. Among all children experiencing family disruption, less than a third are Black (29%). Furthermore, conditional on experiencing family disruption, 33% of the children have to move to a new residence with their mother. 37% of them are Black. Among stayers (conditional on family disruption), 25% of children are Black. Plus, while the child's age as well as mother's years of education are similar for both movers and stayers, children who move have a mother with a lower AFQT score (on average, 2.41) as opposed to stayers (on average, 2.94). Movers are also less likely to have siblings, although the majority of all movers and stayers experiencing a family disruption have at least one sibling.

3.2 OLS Rregression Model (FamDisrup)

As an initial analysis of the relation between parental disruption (FamDisrup) and test score performance (y_{it}), I estimate descriptive OLS regressions of PIAT test scores of the child on a family disruption dummy controlling for time-invariant child characteristics. I control for the following time-invariant child characteristics, which are summarized in **X**: race, age at first observation, presence of siblings, mother's years of education, and mother's AFQT score. Thus, I estimate OLS regressions of the following form:

$$y_{it} = \beta_0 + \beta_1 FamDisrup_{i0} + \mathbf{X}_i' \gamma + \varepsilon_{i0},$$

where $FamDisrup_{i0}$ is a binary variable equal to 1 if the child experiences family disruption in period 0 and equal to 0 if the child stays in a two-parent household with both biological parents throughout the ten years during which I observe five PIAT test score outcomes. Vector **X** incorporates all control variables and ε_{it} is the error term. The outcome variable y_{it} stands for raw PIAT test scores for child i at time

$t \in [-3, -1, +1, +3, +5]$, with t reflecting the distance to the disruption event in 0. For this descriptive analysis, I assume that, from a child's perspective, parental relationship status is exogenous conditional on controls. Thus, identification comes from exogenous changes in family stability: $\mathbb{E}\left[\varepsilon_{it}|FamDisrup_{i0}, \mathbf{X}_i\right] = 0$. I run these regressions five times in order to incorporate the effects on test scores for different years before and after a potential family disruption event in $t = 0$.

I restrict my analysis to children who stay in a two-parent household for the first two periods 1 and 2 (i.e., in $t-3$ and $t-1$) and then stay in a single-parent household with their mother in all subsequent periods $3-5$ (i.e., in $t+1$, $t+3$, and $t+5$). Recall that there is a two-year period between any two consecutive "years" observed. Thus, all these children experience family disruption between the second and the third PIAT test score observation (which I normalize to period 0). I focus on this particular group because it allows me to see how test scores evolved a few years before and after a potential change in the family composition – because, the effect of a family disruption in $t = 0$ on test scores could be of different magnitude 1 period after the event, in $t+1$, versus 5 periods after the event, in $t+5$. It also happens to be the largest of all groups, i.e. most children in the sample experience family disruption between periods 2 and 3 (as opposed to between periods $1-2$, $3-4$ or $4-5$), which should thus also be the relatively most precise group when it comes to statistical power. Table 3 presents estimates from the above regression model. It shows the impact of family disruption in $t = 0$ on PIAT test scores (reading comprehension) as a function of time span between family disruption in t and the timing of test scores measurement. There is a dummy for the targeted explanatory variable equal to 1 if family disruption, and 0 otherwise. The estimates are conditional on controls. I do not report estimates without controlling for characteristics of the child, but results are in line and effects are even stronger.

Table 3: OLS Regression (Family Disruption)

	Test Scores measured around Family Disruption Event in t				
Time	**t − 3**	**t − 1**	**t + 1**	**t + 3**	**t + 5**
Family Disruption (Dummy)	**-0.70****	**-2.03*****	**-3.21*****	**-5.03*****	**-5.34*****
	(0.35)	(0.64)	(0.64)	(0.70)	(0.76)
Observations (balanced panel)	3,435	3,435	3,435	3,435	3,435
Experiencing Fam Disruption	6.8%	6.8%	6.8%	6.8%	6.8%
R^2	0.16	0.12	0.07	0.11	0.11

Note: *p<0.1; **p<0.05; ***p<0.01

Note: Impact of family disruption in $t = 0$ on test scores as a function of time span between family disruption in t and the timing of test scores measurement.

Table 3, presenting estimates of OLS regressions of family disruption on test scores, shows significant negative estimates. For example, the estimated effect is a decrease of 5.34 raw points in test scores 5 years after experiencing family disruption.

This descriptive evidence shows that, at any point in time since family disruption, there is a gap between the affected group and those not affected. However, it is not clear whether that gap is increasing or not since family disruption. While the values do increase over time, the trends do not necessarily need to do so. The trending gap before family disruption is $(-0.70) - (-2.03) = 1.33$. Then, the gap is 1.18, 1.82, and 0.31. So, it could be that part of the trend is just continuing. And if such a trend is continuing, then this implies extra losses around 1.5 every other year, which are not because of what happened in between. However, a difference-in-differences analysis in the next chapter will test that. Before that, I conduct another descriptive analysis where I investigate the interaction between family disruption and residential relocation and their effects on a child's test scores based on an OLS regression model.

3.3 OLS Rregression Model (FamDisrup×Move)

Since I want to learn how the effect on test scores of a change in family stability depends on whether children also have to move at the same time, I also execute an initial analysis of the effect of family disruption in interaction with a simultaneous residential relocation event (Move) on test scores. To analyze the effects of family disruption in interaction with a simultaneous residential relocation event on test scores, I estimate the following regression models of test scores of the child on a family disruption dummy, a moving dummy, and the interaction between these two dummies ($FamDisrup_{i0} \times Move_{i0}$), controlling for the same time-invariant child characteristics as before:

$$y_{it} = \beta_0 + \beta_1 FamDisrup_{i0} + \beta_2 Move_{i0} + \beta_3 (FamDisrup_{i0} \times Move_{i0}) + \mathbf{X}_i' \gamma + \varepsilon_{i0},$$

where now, on top of what the OLS regression model before included, I also have the additional binary variable $Move_{i0}$, which is equal to 1 if the child has to move in period 0 and equal to 0 if the child stays in the same place throughout the ten years during which I observe five PIAT test score outcomes. Coefficient estimate $\hat{\beta}_1$ measures the average difference in test scores between individuals with and without experiencing family disruption; $\hat{\beta}_2$ is the residential stability differential in test scores, ceteris paribus; and $\hat{\beta}_3$ measures the difference in the effect of having a family disruption for movers versus stayers. Adding up $\hat{\beta}_1$ and $\hat{\beta}_2$ gives us an idea how the effect on test scores of a change in family disruption depends on whether the child experienced a simultaneous move. Identification, again, comes from exogenous changes in family stability and/or place of residence for the child: $\mathbb{E}\left[\varepsilon_{it} | FamDisrup_{i0}, Move_{i0}, FamDisrup_{i0} \times Move_{i0}, \mathbf{X}_i\right] = 0$. Table 4 presents OLS estimates from the above regression model.

Table 4: OLS Regression (Family Disruption X Move)

	Test Scores around Disruption Event in t		
Time	**t − 1**	**t + 1**	**t + 3**
FamDisrup (Dummy)	1.71***	1.35***	1.01**
	(0.35)	(0.52)	(0.53)
Move (Dummy)	-0.34	0.47	1.28***
	(0.23)	(0.34)	(0.35)
FamDisrup x Move	-2.14***	-4.71***	-4.39***
	(0.59)	(0.87)	(0.89)
Observations	4,410	4,410	4,410
R^2	0.54	0.28	0.19
Note:	*p<0.1; **p<0.05; ***p<0.01		

Note: Impact of *Family Disruption* × *Move* in $t = 0$ on test scores as a function of time span between simultaneous disruption events in t and the timing of test scores measurement.

Table 4 shows the resulting impact of family disruption in interaction with residential relocation on test scores. For example, 3 years after family disruption, for stayers, the estimated effect is an *increase* of 1.01 raw points, while for children who had to move, the predicted *decrease* in test scores amounts to $1.01 - 4.39 = -3.38$ raw points. OLS estimates are significant (and would be even bigger in the not reported case without control variables).

3.4 OLS Rregression Model (FamDisrup×DistMove)

Now, I proceed as in the previous subsection, but instead of accounting for whether children moved or not, I focus on the distance of the move (DistMove). Again, I run simple regression models where family disruption is interacted with residential relocation, but split between two samples: those moving at most a mile and those moving more than a mile away. Table 5 presents the results.

Table 5: OLS Regression (Family Disruption X DistMove)

	Test Scores around Disruption Event in t		
MOVE ≤ 1 MILE	**$t-1$**	**$t+1$**	**$t+3$**
FamDisrup (Dummy)	1.70***	1.43***	1.02**
	(0.35)	(0.52)	(0.54)
Move (Dummy)	-0.95***	0.39	1.00
	(0.48)	(0.71)	(0.74)
FamDisrup x Move	1.43	-0.57	-0.55
	(1.14)	(1.69)	(1.75)
MOVE > 1 MILE	**$t-1$**	**$t+1$**	**$t+3$**
FamDisrup (Dummy)	1.71***	1.33***	1.03**
	(0.35)	(0.52)	(0.53)
Move (Dummy)	-0.21	0.51	1.34
	(0.22)	(0.37)	(0.37)
FamDisrup x Move	-2.94***	-5.69***	-5.26***
	(0.63)	(0.93)	(0.95)
Note:	*p<0.1; **p<0.05; ***p<0.01		

Note: Impact of *Family Disruption* × *Distance of Move* in $t = 0$ on test scores as a function of time span between simultaneous disruption events in t and the timing of test scores measurement.

Table 5 provides descriptive evidence demonstrating that children who move more than a mile away, in connection with family disruption, experience significantly greater adverse effects on their test scores compared to those who move within a mile.

Such descriptive regressions, as estimated in these OLS regressions in this section, have many desirable properties, but are limited to showing a correlational relationship between disruptions and test scores and potentially suffer from the following issues. First, they could be affected by a selection bias, since there could be sorting into "treatment" making parental separation not random. In fact, the summary statistics results revealed that those families experiencing a disruption have somehow different characteristics than other more stable families in the sample. One could miss potential different trends due to the fact that level differences between the affected and the not-affected groups were already there before the potential event of family disruption. Related, OLS regression models could be impacted by omitted variable bias and either overstate or understate the effect. If there is an omitted variable that is correlated with parental separation and is a determinant of test scores, then assignment is not random and I get an unbiased measure of the effect of family disruption on test scores. In addition, such static pairwise comparisons focusing on the group of children experiencing family disruption between periods 2 and 3 are not exploiting the full sample, as, instead, a staggered event study analysis would do, for example. Lastly, the control group consists of two-parent households who never separate during the considered time period. However, this might not be the right counterfactual and I may want to account for potential tension before treatment or tension in families who do not separate. Comparing it to a group including not-yet treated children seems to be more accurate in order to trace out the causal effect of the transition from a two-parent household to a single-parent one.

To overcome these challenges and identifying the parameters as a causal effects, I will thus focus on an empirical event-study strategy. In particular, I will use the method by Callaway and Sant'Anna (2021) and thus focus on a staggered difference-in-differences model for my baseline analysis to identify the causal effects of disruption on test scores. The dynamic roll-out setting will reveal potential pre-trends and is accounting for the fact that not all children experienced family disruption at the same time, such as in the same year and when of same age, and allows to take not-yet affected children to act as control group.

4 Econometric Approach

In this chapter, I use a dynamic within-child difference-in-differences approach à la Callaway and Sant'Anna (2021) to compare longitudinal test scores of children who experience family disruption, and potential associated change of residence, to those of children who have not yet been affected. The estimator allows for staggered treatment and heterogeneous treatment effects for children and will quantify the causal effects of family disruption, and potential associated change in the place of residence, on test scores.

4.1 Framework of Analysis

Balanced Panel. The sample I consider has no missing values for any child or year and can be described as $\{(Y_{i1}, Y_{i2}, ..., Y_{iT}, D_{i1}, D_{i2}, ..., D_{iT}, \mathbf{X}_i)\}_{i=1}^{N}$, where Y_{it} represents PIAT test scores for child i at time t. Since children take these biennial tests at most five times, I can follow a child's performance over a period for up to ten years. In my baseline analysis, I exploit this longest longitudinal period of time possible, $T = 5$, and eliminate all children with fewer observations. The sample consists of $i = 1, ..., N$ children and includes the following time-invariant covariates, $\mathbf{X}_i$, which must be measured before treatment takes place: race, age at first observation, presence of siblings, mother's years of education, and mother's AFQT score. Lastly, D_{it} is a dummy variable taking the value 1 if child i experienced family disruption, and 0 otherwise. I design it so that once a child is treated, it stays so throughout the rest of the time.

Parameter of Interest. The main parameter that I try to recover in this paper is the "average treatment effect on the treated", ATT, which is the average effect on test scores for those children who experienced family disruption. In particular, I am interested in the average treatment effect at time period t for children that experienced the family disruption event in period g, $ATT(g, t)$. Children in my baseline sample can either be affected in period 2, 3, 4, or 5. I combine all children who experienced family disruption in the same period, resulting in four groups (or "cohorts"). Each of these groups is going to have their own $ATT(g,t)$. The main building block of the paper can be summarized with the following equation (Callaway and Sant'Anna, 2021):

$$ATT(g,t) = \mathbb{E}\left[Y_t(g) - Y_t(0)|G_g = 1\right], \quad \text{for t} \geq g,$$

which is the *group-time average treatment effect*, i.e., the ATT for children who are members of a particular "treatment" group g at a particular time t, $ATT(g,t)$. So, the ATT is allowed to vary across groups g and calendar time t, $ATT(g,t)$, and estimates will show whether and by what magnitude. Since the maximum amount of biennial PIAT tests that one can take is five, all children in my baseline sample are roughly of same age when entering the panel in period 1 (roughly six years old) and taking their first PIAT tests as well as when taking their fifth PIAT tests and subsequently leaving the panel. Thus, the four cohorts (groups), which are created according to the time of treatment, represent children of different ages. For instance, children who experienced family disruption in period 2, and therefore belong to group 2, are six years younger than children who experienced family disruption in period 5, and therefore belong to group 5. On top of potential variation across groups who are of different average age, the $ATT(g,t)'s$ can also vary across time to treatment, i.e., depend on the distance to treatment. For example, the instantaneous effects of family disruption on test scores can be stronger than the one five years after the event.

However, in order to get the overall ATT, I then also have to aggregate the $ATT(g,t)'s$ across groups and time. In the $ATT(g,t)$ equation above, I fix a group g, $G_g = 1$, and I then have the treatment effect at time period t, $Y_t(g)$. While the expected value of Y_t among treated groups g, $\mathbb{E}[Y_t(g)|G_g = 1]$, is observed from the data, the challenge is to recover $\mathbb{E}[Y_t(0)|G_g = 1]$. To generate this missing counterfactual measure in a difference-in-differences setup, two key assumptions must hold (Callaway and Sant'Anna, 2021). Identification of the ATT is achieved via the following two main identification assumptions.

Assumption 1 (No Treatment Anticipation).
$\mathbb{E}[Y_t(g)|X, G_g = 1] = \mathbb{E}[Y_t(0)|X, G_g = 1]$ *almost surely* $\forall g \in \mathcal{G}, t \in T$, *such that* $t < g$.

This assumption states that before treatment takes place ($t < g$), there is no treatment effect. And if no treatment, then the treatment effect for this group G is also zero. For example, test score performance outcome in period 1 is not affected by treatment status in period 2 (whether the child experiences family disruption or not). This assumption guarantees that the difference-in-differences estimand can be interpreted as a causal effect in period 2.
While parental separation is assumed to be exogenous from the child's point of view, it is possible that these school-aged children were already informed and anticipated a parental breakup before the actual event, which could have impacted their school performance due to the announcement or anticipation. While there is a two-year gap between the immediate test scores taken before and after such a family disruption, and it is probably less likely that parents already announced their breakup two years before the event, I cannot be certain. Therefore, I will test this in the next section. First, I also have to impose a second assumption.

Assumption 2 (Parallel Trends).
For each $(s, t) \in \{2, ..., T\}$x$\{2, ..., T\}, g \in \mathcal{G}$ such that $t \geq g, s \geq t$,

$\mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = \mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, D_s = 0, G_g = 0]$ *almost surely.*

This is a parallel trends assumption conditional on observables and comparing the affected children to a control group consisting of children who are not-yet affected at the time of the comparison, meaning that they could either be affected later or never during the sample observed. To be more specific, *Assumption 2* states that, conditional on a vector of covariates X, the evolution of the outcome in the absence of treatment, $Y_t(0) - Y_{t-1}(0)$, among this group g, is the same as among the not-yet-treated children on average, where $D_s = 0$ means not-yet-treated by time s. The notation $(t \geq g, s \geq t)$ restricts parallel pre-trends, as required when using not-yet treated children as control group (Callaway and Sant'Anna, 2021).

Nonparametric Identification. Given both restrictions, *Assumption 1* and *Assumption 2*, hold, I can then nonparametrically point-identify the group-time average treatment effects, $ATT(g,t)'s$, via the following representation (Callaway and Sant'Anna, 2021):

$$ATT(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_{g,t}(X)(1-D_t))}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_{g,t}(X)(1-D_t))}{1-p_{g,t}(X)}\right]}\right)(y_t - y_{g-1} - \mathbb{E}[y_t - y_{g-1}|X, D_t = 0, G_g = 0])\right]$$

Even though I have multiple groups of children affected by family disruption at different times and multiple periods of test scores observations around such a potential disruption event in my sample, the subsetting of the data into $ATT(g,t)'s$ is letting me regard each $ATT(g,t)$ as a two-group and two-period case. So the comparison group changes over time. Thus, I have two groups (or "cohorts"): G_g representing children from affected group g and the group not-yet affected, D. Plus, I have two time periods (t and $g-1$), as well as stabilizing weights, $Y_t - Y_g - m(X)$. Also, $p_{g,t}(X)$ is the time-varying propensity score of being in group g as a function of the vector of additional control variables, X. I am going to model both the probability of belonging to the treatment group, denoted by the propensity score $p_g(X)$, and I am going to put more weights on children who are more similar to the group's g, G_g. So, $p_g(X)$ reweights the comparison groups (D) such that they look similar to the cohort g, G_g. Further, it includes an outcome regression model, $E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$, which models the outcome evolution of the not-yet affected children. Implicitly, $y_t - y_{g-1} - E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$ is a difference-in-difference with recentered covariates ("residuals"): $y_t - y_{g-1} - E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$, where $E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$ is the outcome regression. And then I am reweighting the residuals with the large bracket on the left-hand side, so that I am balanced. See Callaway and Sant'Anna (2021) for more details. The idea is that, as long as my model for the propensity score, $p_g(X)$, *or* my model for the outcome evolution, $E(y_t - y_{g-1}|X, D_t = 0, G_g = 0)$, are correctly specified, this is going to recover the ATT. Since this is giving me two "chances" to estimate the ATT, this estimator is also called the "doubly-robust" estimator. This estimator by Callaway and Sant'Anna (2021) extends Heckman et al. (1998); Abadie (2005); Sant'Anna and Zhao (2000). Without covariates, Callaway and Sant'Anna (2021)'s procedure would be similar to Sun and Abraham (2001) and de Chaisemartin and D'Haultfœuille (2020).

Aggregation. To enhance precision and regain statistical power, I can aggregate the subsetted data to construct a unified overall $ATT(g, t)$. The idea is to calculate a weighted average by summarizing the $ATT(g, t)$ components for every group g and every time period t. In particular, Callaway and Sant'Anna (2021) propose taking the weighted averages of the $ATT(g, t)'s$ of the following form:

$$\sum_{g=2}^{T} \sum_{t=2}^{T} \mathbb{1}\{g \leq t\} w_{gt} ATT(g, t),$$

where the indicator function guarantees that I have zero weights before treatment takes place and non-zero weights afterwards – because the goal is to average whatever happens post-treatment (and not pre-treatment). Aggregation then follows by computing weighted averages of the $ATT(g, t)'s$ with given weights w_{gt} that can vary across groups g and across time t.

While there are different types of weights possible, for my baseline analysis, I focus on event-study type weights (or dynamic treatment effect weights). This weighting method seems most feasible, because the effect of family disruption on test scores may depend on the length of exposure to a single-parent household, so I want to emphasize how treatment effects may vary with elapsed treatment. However, I also check alternative weighting methods as robustness. Average effect of participating in the treatment for the group of children that have been exposed to the treatment for exactly e time periods:

$$\theta_D(e) = \sum_{g=2}^{T} \mathbb{1}\{g + e \leq T\} ATT(g, g+e) P(G = g | G + e \leq T, C \neq 1)$$

The time period is now reparameterized to $t = g + e$, where e stands for time since treatment took place. So, $e = 0$ would be the instantaneous treatment effect, $e = 1$ would be the treatment effect 1 period after treatment, etc. Since there is a two-year gap between any two consecutive observations, $e = 0$ can be interpreted as one year after family disruption or two years after the previous observations (where they were still living in a two-parent household). I then average the $ATT(g,t)'s$ for a fixed e across different groups $G = g$. I will also report robustness checks based on two alternative weighting methods.

First, I undertake an aggregation that takes the time average of these $ATT(g,t)'s$ for each group. This gives me the average effect of "participating" in the treatment of family disruption that children in group g experienced, which allows me to highlight heterogeneity across groups. Second, I fix time period t and average across groups. This allows me to highlight heterogeneity across time periods (calendar-time heterogeneity). The result is an average effect of participating in the treatment in time period t for groups that have participated in the treatment by time period t.[10]. For estimation, Callaway and Sant'Anna (2021) allow to choose between the outcome regression, the inverse probability weighting, and the doubly-robust one.

Robustness. While my baseline analysis consists of aggregating the $ATT(g,t)$'s such that I can make statements how average treatment effects vary by length of exposure to the treatment, I also experiment with other aggregation methods proposed by Callaway

[10] An alternative way of combining $ATT(g,t)'s$ across g and t could just be to take a simple weighted average (give weight 1 to everybody, i.e., everybody has the same weight such that these weights sum up to 1). However, this is not very appealing as some groups are much larger than others (group g could be $10x$ bigger than group g', so cohort g deserves to be more ?¿'informative"). An alternative would be to attach weights on the size of group g. However, earlier groups get larger weights.

and Sant'Anna (2021). In particular, I also aggregate group-time average treatment effects into group-specific average treatments effects in order to focus on average effects specific to each group treated at the same time. In addition, I also do a calendar-time aggregation that gives me the average effect of experiencing family disruption in a particular time period for all groups that were treated in that time period. However, I do not report a simple weighted average of all group-time average treatment effects with weights proportional to the group size, since that tends to overweight the effect of early-treated groups, because more of them are observed during post-treatment periods.

5 Results

In this section, I present the baseline results. I begin by examining the average effects of family disruption on test scores, focusing on how these effects vary by the length of exposure to the parental separation event ("treatment"). This analysis sheds light on the evolution of average treatment effects over time for specific groups. I differentiate between two categories of disruption: (1) "All Disruptions", which encompass family disruptions due to parental breakup, the death of the father, his imprisonment, or any other reasons, and (2) "Breakup", which considers only parental relationship splits. Furthermore, I refine the analysis by distinguishing between movers and stayers among those experiencing family disruption and by considering the distance of relocation.

5.1 Effects of Family Disruption on Test Scores

In order to provide evidence of the aggregate effects of family disruption on test scores, I aggregate group-time average treatment effects, $ATT(g, t)$'s, in the spirit of Callaway and Sant'Anna (2021). Table 6 shows the dynamic effects on the raw PIAT reading comprehension score for all available distances to the family disruption event. The "Overall ATT" reported in the table averages the treatment effects across all lengths of exposure to the family disruption event.

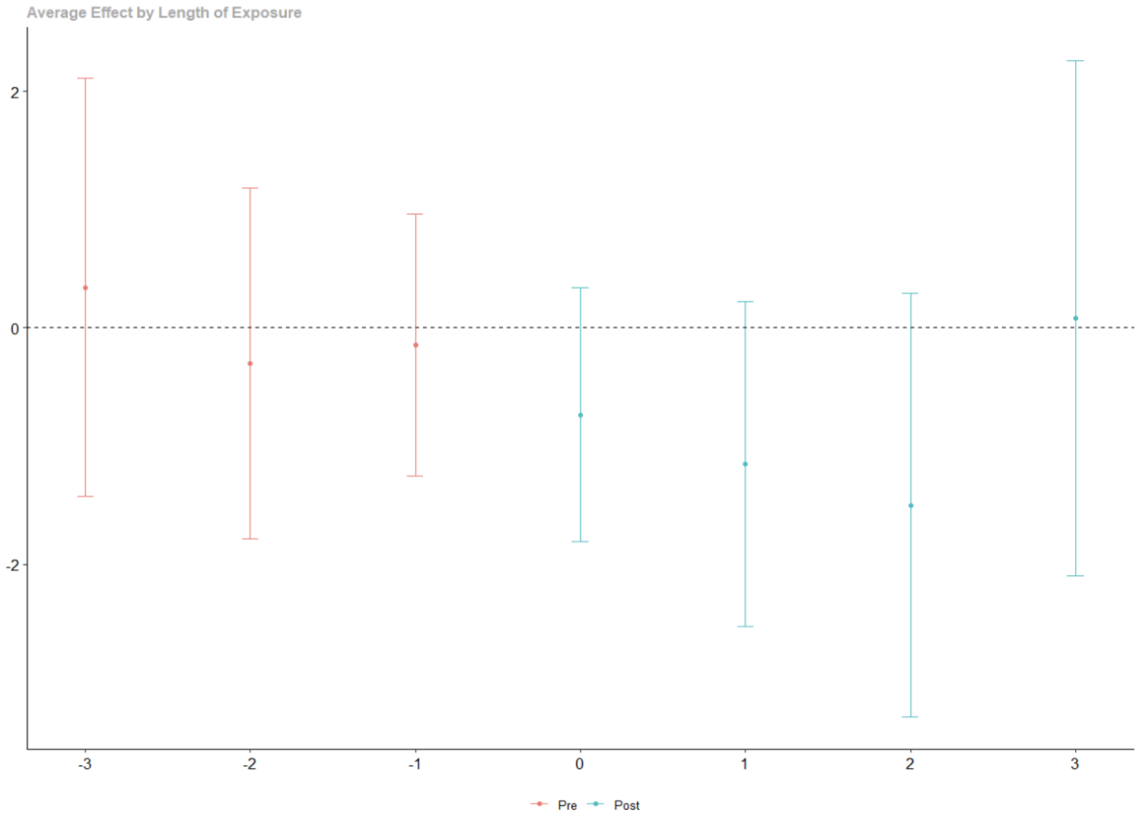
Table 6: Dynamic Aggregation of Group-Time Average Treatment Effects

Event Time	All Disrup	All Disrup	Breakup	Breakup
-3	0.29	0.34	-0.06	-0.08
	(0.64)	(0.68)	(0.84)	(0.83)
-2	-0.36	-0.30	-0.55	-0.29
	(0.58)	(0.57)	(0.75)	(0.78)
-1	-0.25	-1.14	-0.32	-0.10
	(0.43)	(0.42)	(0.59)	(0.59)
0	-0.96	-0.74	-1.46^{*}	-1.05
	(0.40)	(0.38)	(0.48)	(0.48)
$+1$	-1.70^{*}	-1.15	-1.99^{*}	-1.14
	(0.51)	(0.51)	(0.70)	(0.64)
$+2$	-2.79^{*}	-1.50	-3.02^{*}	-1.41
	(0.62)	(0.62)	(0.86)	(0.83)
$+3$	-1.97	0.08	-2.61	-0.29
	(0.93)	(0.81)	(1.12)	(1.07)
Overall ATT:	-1.85^{*}	-0.83	-2.27^{*}	-0.97
	(0.48)	(0.46)	(0.58)	(0.55)
Controls		✓		✓
# Obs	7,345/5 = 1,469 kids		6,240/5 = 1,248 kids	
Fam Disrup	39%		28%	

Results in Table 6 reveal that the average effects of a parental breakup on the child's test scores ("Breakup") are slightly higher compared to all types of disruptions ("All Disrup"). In addition, while estimates are significant when not controlling for observable characteristics, lowering test scores by roughly 2 raw points, they are not significant otherwise, but still lower test scores by roughly 1 raw point.

Figure 1 illustrates the average effect by length of exposure to the family disruption event for the sample including all types of family disruption and including control variables.

Figure 1: Dynamic Aggregation of Group-Time Average Treatment Effects



Note: Graph shows aggregated average treatment effects on treated by length of exposure. The dynamic aggregation of group-time average treatment effects includes all types of family disruption. Whisker plots represent 95% *confidence interval constructed from bootstrapped standard errors with 1000 iterations.*

The same figure based on a sample of parental separation only as a potential family disruption event (excluding a father's absence due to other reasons such as death or imprisonment) is shown in *Appendix B*. I will now refine the analysis by distinguishing between movers and stayers among those experiencing family disruption.

5.2 Effects of (Family Disruption × Move) on Test Scores

Table 7: Effects of (Family Disruption × Move) on Test Scores

Event Time	(1) disrup(all)	(2) disrupXmove −	(3) stableXmove	(4) = disrup(cond)	(5) disrupXstay
t−1	-0.16	-0.89	-0.17	-0.72	0.89
	(0.49)	(1.34)	(0.67)	(0.67)	(0.78)
t+1	-0.75^{*}	-0.58	0.95	-1.53	-0.96
	(0.34)	(0.84)	(0.45)	(0.39)	(0.61)
t+3	-1.22	-0.47	1.77^{*}	-2.24^{*}	-0.08
	(0.55)	(1.32)	(0.56)	(0.76)	(0.94)
Overall ATT	-0.99^{*}	-0.53	1.36^{*}	-1.89^{*}	-0.52
	(0.38)	(0.94)	(0.41)	(0.53)	(0.60)
AgeFE	✓	✓	✓	✓	✓
RaceFE	✓	✓	✓	✓	✓
AFQTMomFE	✓	✓	✓	✓	✓
EducMomFE	✓	✓	✓	✓	✓
SiblingsFE	✓	✓	✓	✓	✓
# obs	9,762/3=3,254	3,015/3=1,005	4,296/3=1,432		3,456/3=1,152
Treated	22%	11%	38%		22%

The unrestricted effect of family disruption on children's test scores for the considered sample with three observation periods is a decrease of 1 raw test score point, as reflected by the "Overall ATT" in column (1). When family disruption is associated with a move (column 4), then that implies a decrease of 1.9 raw test score points. However, for stayers experiencing family disruption (column 5), the decrease in raw test score points is 0.5 only and not significant.

5.3 Effects of (Family Disruption × Distance of Move) on Test Scores

Now, conditional on moving associated with family disruption, I will further account for the distance of the residential relocation.

Table 8: Distance Large

Event Time	(1) disrup(all)	(2) disrupXmove −	(3) stableXmove	(4) = disrup(cond)	(5) disrupXstay
t−1	-0.16	-1.25	-0.18	-1.07	0.89
	(0.50)	(1.44)	(0.76)	(0.68)	(0.76)
t+1	-0.75^{*}	-0.67	0.93	-1.60	-0.96
	(0.32)	(0.83)	(0.47)	(0.36)	(0.57)
t+3	-1.22	-0.72	1.77^{*}	-2.49^{*}	-0.08
	(0.56)	(1.57)	(0.55)	(1.02)	(0.88)
Overall ATT	-0.99^{*}	-0.69	1.35^{*}	-2.04^{*}	-0.52
	(0.36)	(1.07)	(0.42)	(0.65)	(0.62)
AgeFE	✓	✓	✓	✓	✓
RaceFE	✓	✓	✓	✓	✓
AFQTMomFE	✓	✓	✓	✓	✓
EducMomFE	✓	✓	✓	✓	✓
SiblingsFE	✓	✓	✓	✓	✓
# obs	9,762/3=3,254	2,958/3=986	3,978/3=1,326		3,456/3=1,152
Treated	22%	9%	33%		22%

Table 9: Distance Small

Event Time	(1) disrup(all)	(2) disrupXmove −	(3) stableXmove	(4) = disrup(cond)	(5) disrupXstay
t−1	-0.16	1.39	-0.04	1.43	0.89
	(0.49)	(2.48)	(1.04)	(1.44)	(0.78)
t+1	-0.75^{*}	-0.11	1.06	-1.17	-0.96
	(0.32)	(2.21)	(1.03)	(1.18)	(0.55)
t+3	-1.22	0.82	1.76	-0.94	-0.08
	(0.52)	(2.74)	(1.22)	(1.52)	(0.92)
Overall ATT	-0.99^{*}	0.35	1.41	-1.06	-0.52
	(0.39)	(2.04)	(0.93)	(1.11)	(0.63)
AgeFE	✓	✓	✓	✓	✓
RaceFE	✓	✓	✓	✓	✓
AFQTMomFE	✓	✓	✓	✓	✓
EducMomFE	✓	✓	✓	✓	✓
SiblingsFE	✓	✓	✓	✓	✓
# obs	9,762/3=3,254	2,742/3=914	3,003/3=1,001		3,456/3=1,152
Treated	22%	2%	11%		22%

The above two tables show results for differences in the distance of moving. The top table shows results for large differences between the old and the new place of residence (> 1 mile) and the bottom table shows the corresponding results for a small distance of residential relocation (≤ 1 mile). Results reveal that for those children moving more than a mile away associated with family disruption experience a much larger and significant decrease in test scores (-2.04 raw points) compared to those moving within a mile (-1.06 raw points and not significant).

6 Sensitivity Analysis

In this section, I perform a series of robustness checks. In particular, I perform some additional aggregations of the $ATT(g, t)$'s. Figure 2 and Figure 3 show the average effect by group and by time period, respectively.

Figure 2: Average Effect By Group

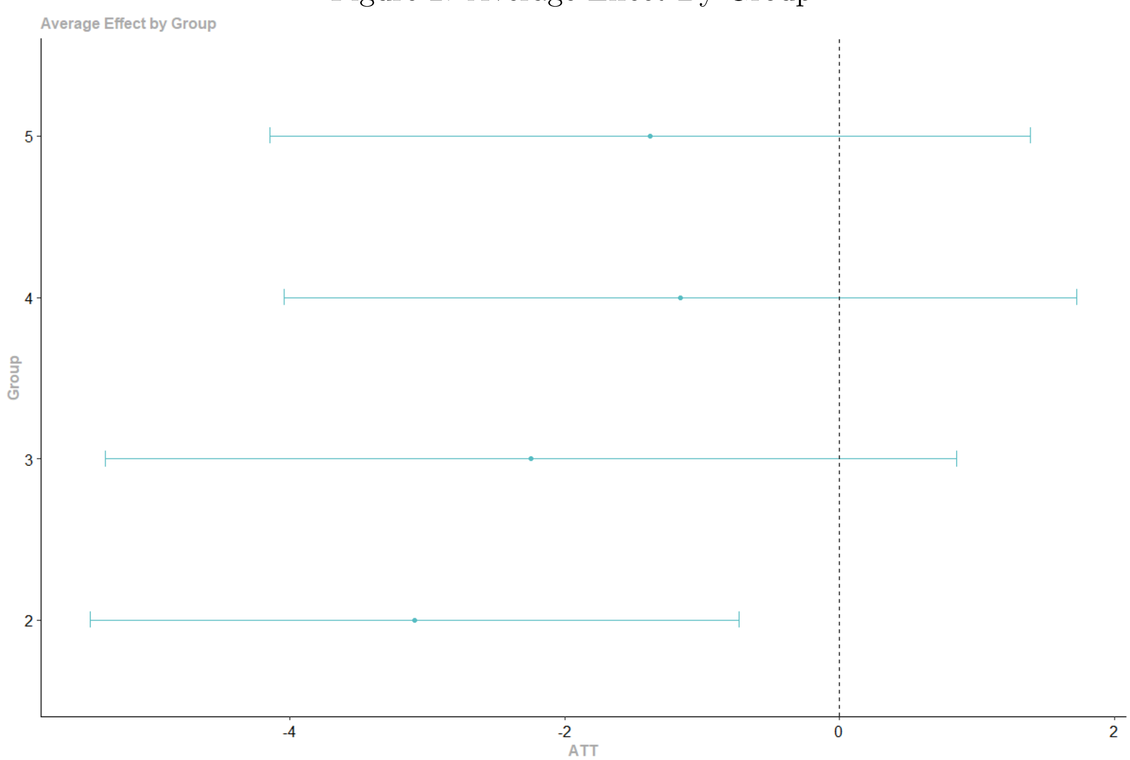
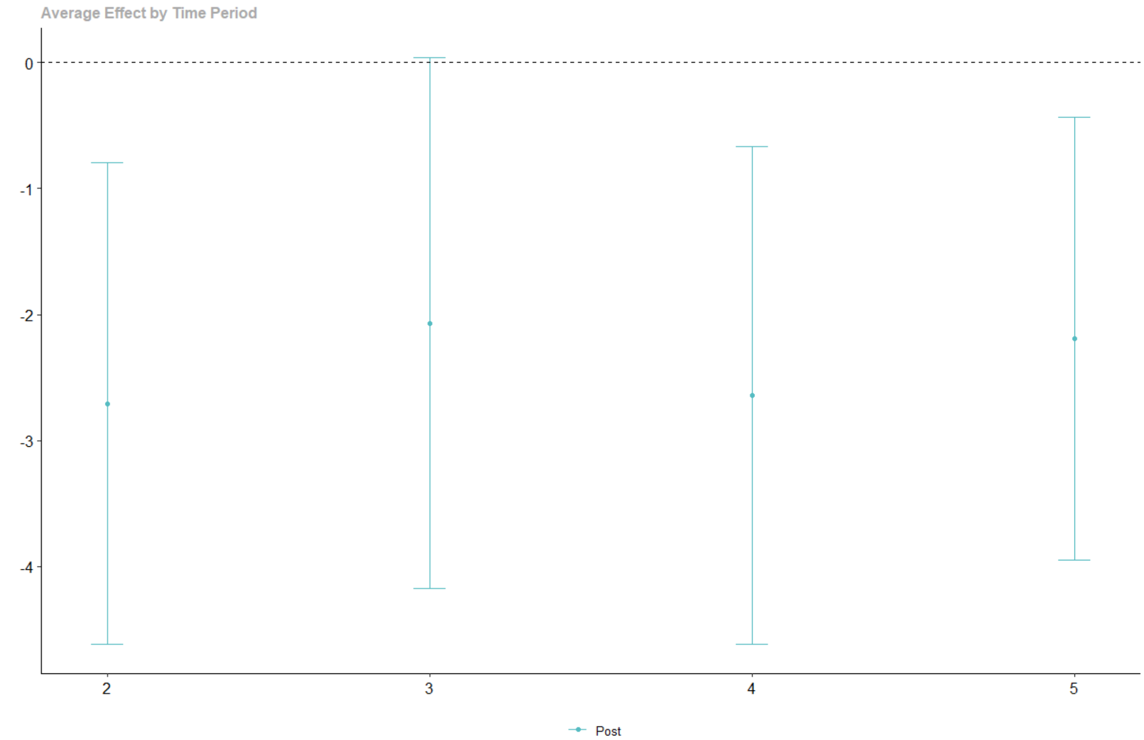


Figure 3: Average Effect by Time Period



7 Conclusion

It takes a village to raise a child. This paper quantifies the causal effects of family disruption on test scores of school-aged children in the United States and presents evidence that a simultaneous change of residence associated with parental separation is responsible for most of the adverse effects on test scores and not family disruption per se. In particular, the distance of the move is key, as a move within a mile is bearable.

My analysis opens avenues for several meaningful extensions. First, my findings can hopefully inspire future research to more fully elucidate the role of moving and potential heterogeneity in the effects due to the reason for parental separation. Future research may want to investigate the mechanism behind the moving issue: one potential mechanism behind the gap in test scores between the studied two-parent households and those experiencing family disruption could be peer disruption. I also find that new single mothers who move are somewhat different in terms of observable characteristics than those who stay after separating from the partner. Second, in my case, the time period cannot exceed 10 years, because children take the biennial PIAT tests at most five times. However, with a longer panel, equilibrium effects are likely to appear and neighbors being treated could affect the treatment. Testing general equilibrium effects with more frequent and/or long-term data offers avenues for future research.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies 72*(1), 1–19.

Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2024). It takes a village: The economics of parenting with neighborhood and peer effects. *Journal of Political Economy, forthcoming*.

Agostinelli, F. and M. Wiswall (2024). Estimating the technology of children's skill formation. *Journal of Political Economy forthcoming.*

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

Chan, M. K. and K. Liu (2022). Changing families: Family relationships, parental decisions and child development. *Journal of Labor Economics, Forthcoming*.

Chetty, R. and N. Hendren (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R. and N. Hendren (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics 133*(3), 1163–1228.

Chetty, R., N. Hendren, and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*(4), 855–902.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review 108*(10), 3028–56.

Cordes, S. A., A. E. Schwartz, and L. Stiefel (2019). The effect of residential mobility on student performance: Evidence from new york city. *American Educational Research Journal 56*(4), 1380–1411.

Cunha, F. and J. J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources 43*(4), 738–782.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964—-96.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *The Review of Economic Studies 81*(1), 137–185.

Dunn, L. and F. Markwardt (1970). Manual for the peabody individual achievement test. *Circle Pines, MN: American Guidance Service*.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2016). Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics 8*(1), 1–41.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies 65*(2), 261–294.

Kalil, A., M. Mogstad, M. Rege, and M. Votruba (2011). Divorced fathers' proximity and children's long-run outcomes: Evidence from norwegian registry data. *Demography 48*(3), 1005–1027.

Marinescu, I. (2016). Divorce: What does learning have to do with it? *Labour Economics 38*(2), 90–105.

Nielsen, M. M., P. Fallesen, and M. Gähler (2024). Parental union dissolution and children's emotional and behavioral problems: Addressing selection and considering the role of post-dissolution living arrangements. *Working Paper*.

Roth, J., P. 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*(2), 2218–2244.

Sant'Anna, P. and J. Zhao (2000). Doubly robust difference-in-differences estimators. *Journal of Econometrics 219*(1), 101–122.

Selya, A., E. Engel-Rebitzer, L. Dierker, E. Stephen, J. Rose, D. Coffman, and M. Otis (2016). The causal effect of student mobility on standardized test performance: A case study with possible implications for accountability mandates within the elementary and secondary education act. *Frontiers in Psychology 7*(1096), 1–10.

Sun, L. and S. Abraham (2001). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

Tartari, M. (2015). Divorce and the cognitive achievement of children. *International Economic Review 56*(2), 597–645.

Wooldridge, J. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Working Paper*.

Appendices

Appendix A. Details on PIAT Tests

Key outcome and measurement variables in this paper are PIAT test scores. To assess a child's reading skills, the PIAT encompasses two sub-tests: Reading Comprehension and Reading Recognition. The PIAT Reading Comprehension test measures the child's ability to understand and interpret a sentence. This sub-test consists of 66 items of increasing difficulty, where the child has to select one out of four pictures that best describes what is written. The PIAT Reading Recognition test, on the other hand, measures word recognition and pronunciation ability, which are essential components of reading achievements. It contains 84 items, each with four options, which increase in difficulty from preschool to high school levels. Skills assessed include matching letters, naming names, and reading single words aloud. Even though understanding the context appears to be more important than being able to read it, I incorporate the reading recognition test on top of the reading comprehension test, because adverse events such as family disruption could have an impact on the (compounding) development of the child more generally. Furthermore, Dunn and Markwardt (1970) emphasize that reading aloud is a useful skill throughout life in a wide range of situations also outside of school. The authors also stress that, as the child gets older, performance becomes more and more confounded with the acculturation factors. The third assessment from the PIAT that I consider is Mathematics. It consists of 84 multiple-choice items of increasing difficulty. It begins with recognizing numerals and progresses to measuring advanced concepts in geometry and trigonometry. The child looks at each problem on a flipboard and then chooses an answer by pointing to or naming one of four answer options. More information regarding the PIAT test can be found on the NLSY website of the U.S. Bureau of Labor Statistics.[11]

[11] https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/assessments/piat-reading-reading-recognitionreading, accessed November, 2024.

Appendix B. Additional Tables and Figures

Figure 4: Dynamic Aggregation of Group-Time Average Breakup Effects

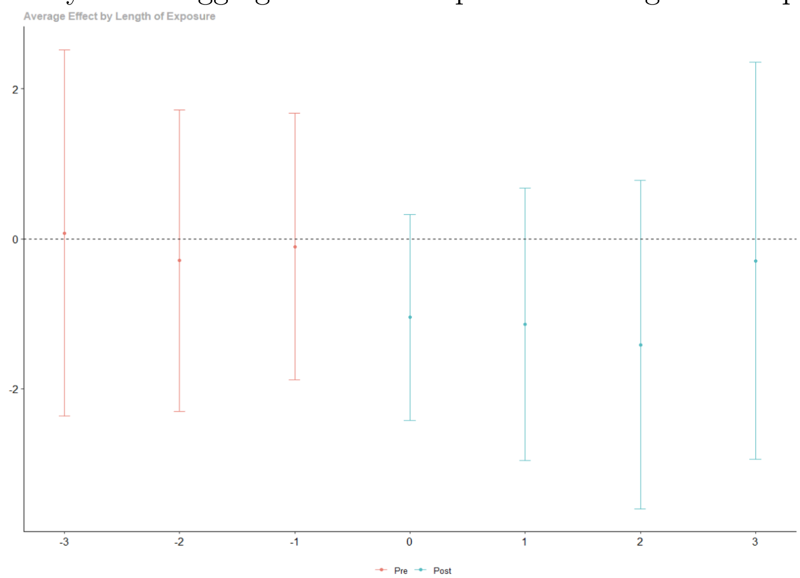


Table 10: Treatment Effects for Breakups only

Event Time	(1) disrup(all)	(2) disrupXmove −	(3) stableXmove	(4) = disrup(cond)	(5) disrupXstay
t−1	-0.38	0.73	-0.04	0.77	0.34
	(0.61)	(1.56)	(0.64)	(0.92)	(1.13)
t+1	-0.24	-0.44	0.94	-1.38	0.16
	(0.39)	(0.99)	(0.42)	(0.57)	(0.74)
t+3	-0.63	0.86	1.73*	−0.87*	0.47
	(0.59)	(1.48)	(0.55)	(0.93)	(1.20)
Overall ATT	−0.43	0.21	1.34*	−1.13*	0.31
	(0.41)	(1.07)	(0.42)	(0.65)	(0.83)
AgeFE	✓	✓	✓	✓	✓
RaceFE	✓	✓	✓	✓	✓
AFQTMomFE	✓	✓	✓	✓	✓
EducMomFE	✓	✓	✓	✓	✓
SiblingsFE	✓	✓	✓	✓	✓
# obs	9,093/3=3,031	2,904/3=968	4,266/3=1,422		3,060/3=1,020
Treated	16%	8%	37%		13%
R^2	0.67	0.68	0.68		0.68

Appendix C. Results based on a Static TWFE Model

To further investigate the popular phenomenon of residential relocation in the context of family disruption, I focus on a triple difference-in-differences framework. For this, I focus on a specific simple difference-in-differences case of the staggered version, considering the group of individuals who get treated between period 2 and 3. Since in this case, I don't have a roll-out design anymore, a two-way fixed effects model is valid. Homogeneity in treatment effect irrespective of the year being treated is also feasible. Treatment effects can still be heterogeneous through time, so that depending on the distance to treatment, the strength of treatment can vary.

To be more specific, I estimate the following panel data model with child and time fixed effects twice, once for movers and once for stayers, and the difference of these two groups then yields the triple difference-in-differences.

$$y_{it} = \alpha_i + \delta_2 f_t^2 + \delta_3 f_t^3 + \delta_4 f_t^4 + +\delta_5 f_t^5 + \omega_2 d_s f_t^2 + \tau_3 d_s f_t^3 + \tau_4 d_s f_t^4 + \tau_5 d_s f_t^5 + \varepsilon_{it},$$

for t=1,...,5.

- y_{it} = raw test scores for child i at time t
- α_i = child fixed effects
- δ_t = time fixed effects
- d_s = ever-treated dummy variable (=1 for treated i)
- $f_t^2, f_t^3, f_t^4, f_t^5$ = year-dummies for the 2nd, 3rd, 4th, and 5th period, respectively
- ε_{it} = error term
- Since the effect of family disruption can be heterogeneous through time, I estimate four event effects: ω_2, τ_3, τ_4, τ_5

From the child's point of view, parental relationship status is exogenous conditional on fixed effects. Thus, identification comes from exogenous changes in family stability. The simple specification gives us a triple difference-in-differences value of -2.51 in the year after family disruption and -2.20 three years after the disruption:

	Coefficient compared to t-1	
MOVERS*	**t + 1**	**t + 3**
DD Estimate (FamDisrup-Stable)	-3.19**	-3.24*
	(1.30)	(1.75)
Observations: 1,290		
Experiencing Fam Disruption: 21.13%		
R^2: 0.87; Adj. R^2: 0.81		
STAYERS	**t + 1**	**t + 3**
DD Estimate (FamDisrup-Stable)	-0.68	-1.04
	(1.03)	(1.21)
Observations: 3,120		
Experiencing Fam Disruption: 13.54%		
R^2: 0.87; Adj. R^2: 0.80		
Triple Difference-In-Difference:	**-2.51**	**-2.20**
Note:	*p<0.1; **p<0.05; ***p<0.01	

Since this "relative relative-to-relative" value is difficult to interpret, it is important to highlight that the effect of family disruption on test scores is more than three times higher for movers (and significant) than for stayers. Furthermore, the difference-in-differences (DD) estimates are higher for moves > 1 Mile and lower for moves ≤ 1 Mile.