THREE ESSAYS IN APPLIED MICROECONOMICS

by

ROBERT JOHN MCDONOUGH II

A DISSERTATION

Presented to the Department of Economics
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2023

DISSERTATION APPROVAL PAGE

Student: Robert McDonough

Title: Three Essays in Applied Microeconomics

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Glen Waddell	Chairperson
Mark Colas	Core Member
Edward Rubin	Core Member
Chris Sinclair	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2023

DISSERTATION ABSTRACT

Robert John McDonough II

Doctor of Philosophy

Department of Economics

June 2023

Title: Three Essays in Applied Microeconomics

This dissertation examines three topics in applied microeconomics: econometric challenges created by student grade-point averaging, the causal effect of violent video games on crime, and spatial distortions created by the US social safety net.

Chapter 1 (with Glen Waddell) considers the underlying combinatorics of grade-point averaging, and the evolution of a GPA as students take classes. We illustrate the implications for inference that relies on the comparison of students with similar GPAs. In the context of a regression discontinuity, researchers are most exposed to this sensitivity with fewer classes contributing to GPA and at smaller bandwidths. While larger bandwidths shield such estimators from this challenge, this accommodation relies on the assumption of sufficient overlap of student types—to the extent there is not, identification is again threatened.

Chapter 2 (with Gretchen Gamrat) examines the causal relationship between violent video game releases and violent crime patterns. Using county-level variation in retail sales of "mature" video games, we leverage exogenous variation in exposure to identify corresponding changes in crime outcomes. Especially after high-profile violent crimes, policymakers and the news media frequently argue that increased exposure to violent games leads to increased violent crime. We find no such evidence. If anything, our analysis suggests that short-run decreases in violent crime, specifically violent sexual offenses, follow the release of mature video games.

Chapter 3 (with Mark Colas) studies the effect of the US social safety net on household location choice. US social transfer programs vary substantially across states, incentivizing

households to locate in states with more generous transfer programs. Further, transfer formulas often decrease in income, thereby rewarding low-income households for living in low-paying cities. We quantify these distortions by combining a spatial equilibrium model with a detailed model of transfer programs in the US.

Chapter 4 concludes this dissertation. This dissertation includes previously both previously published and unpublished and co-authored material.

ACKNOWLEDGEMENTS

I wish to express gratitude to my committee chair and collaborator, Glen Waddell, who provided honest guidance and support throughout my time as a graduate student. He has acted as a role model for academic life, and I look forward to our continued collaboration. Special thanks are also due to my coauthor and committee member Mark Colas, whose continued input and support improved every chapter of this dissertation. I would also like to thank the other members of my committee, Ed Rubin and Chris Sinclair, who provided valuable feedback on this dissertation.

I must express gratitude to John Buettler, my first mentor. John helped me take the first steps that led me here. I am forever in his debt. I am grateful for the unceasing love and support of my sisters, family, and friends. Finally, I am grateful beyond expression to Kelly, without whom this dissertation would not exist.

To Mother and Dad

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

COMBINATORICS, MEAN CONVERGENCE, AND GRADE-POINT AVERAGING

This dissertation includes unpublished co-authored material. This chapter (Chapter I) includes material that was coauthored with Glen Waddell: we jointly performed the statistical analysis and we both contributed to the writing of the manuscript. Chapter II includes material that was coauthored with Mark Colas and is forthcoming publication in the Journal of Labor Economics. Chapter III includes material that was coauthored with Gretchen Gamrat.

1.1 Introduction

In this paper we deconstruct grade-point averaging in a way that illustrates how variation in grade-point averages (GPAs) can challenge intuitions about local comparability. In so doing, we simultaneously demonstrate how rationing resources by GPA-based rules can be misguided, and how empirical tests that rely on the comparability of students with similar GPAs can be misleading.

Despite the prominent role this index of performance plays in resource allocation, there is very little analysis that formally considers the suitability of allocating resources to students differently based on marginal differences in GPA. Moreover, important decisions are often made around small differences in GPA. It is not uncommon for admissions decisions to implicate minimum-GPA requirements, for example, affording some students opportunity while denying others the same. It is also typical for academic probation and scholarship requirements to be GPA based. With resources being allocated this way, it is fundamentally important to understand the mechanisms that sort students to one side of a required GPA or the other.

While large differences in GPA can reliably signal differences in student ability, the underlying combinatorics of grade-point averaging (i.e., the systematic process by which discrete grades combine into an aggregate measure of student performance) implies that the sorting of students into GPA is itself not trivial. In particular, combinatorics regularly perturbs the ability-orderings of students across GPA—that is, there are many opportunities for the average student ability systematically *decreases* in GPA. This is most consequential when the number of contributing classes is small, which suggests that this would be most consequential early in students' academic careers. However, it is likewise important in any setting where the grades of a small number of select classes are aggregated to inform decision making. This also suggests that policies that rely on GPA-based metrics to allocate resources early in academic careers are particularly worthy of close scrutiny. For example, it implies that there are potential welfare benefits to reimagining the application of minimum-GPA cutoffs—such rules are clearly questionable if the average ability of students is systematically higher just *below* such minimums.

Given the prevalence of GPA-based rules, they also tend to be exploited to identify causal relationships running from treatment to outcomes. Regression-discontinuity designs, for example, construct empirical tests out of local environments where students have been treated differently on either side of a GPA (e.g., Bonilla et al., 2021; Bell, 2021; Bleemer and Mehta, 2020; Ost et al., 2018; Kane, 2003; van der Klaauw, 2002). Without coincident discontinuities in the type of student on either side of a GPA threshold, any discontinuities in average outcomes across that threshold are then reasonably interpretable as being induced by the different experiences students had on either side of it. However, as the combinatoric process naturally produces discontinuities in student type across GPA—even non-monotonicities are apparent—it is seemingly untenable to assume smoothness in student type through GPA-induced treatment rules without regard for the nature of the combinatoric problem at the relevant point in time.

In our decoding of the variation in grade-point averaging we draw out two characteristics of GPA. The first is the mean convergence process—like any stochastic process that informs better with a larger number of draws, GPA will tend to better represent a student's true ability as the number of classes contributing to GPA increases. The second is the combinatorics of grade-point averaging—though navigable, given the discretized letter grades and their associated grade points, the evolution of GPA is a combinatorics problem. More to the point, however, we suggest that as GPAs only converge to their means through the permissible combinatorics, understanding these two fundamental processes together becomes vital to how GPA-based rules for resource allocation should be evaluated and to the legitimacy of estimators that rely on local comparability in GPA.

In Section 1.2 we discuss these two mechanisms in turn. In so doing, we illustrate them in simple data-generating processes, limiting the scope of the problem by allowing for only a few levels of student ability. In Section 1.3 we then simulate course-taking behavior and consider an estimator that is sensitive to local variation in particular. This allows us to highlight with a known and transparent data-generating process the potential implications of assuming local comparability in GPA. In Section 1.4 we consider a version of the analysis that assumes continuous ability. While this possibly represents a more-extreme approach to capturing the relevant data-generating process, it will server to highlight that as long as the quantitative inputs (i.e., letter grades and their associated grade points) to constructing GPAs are discretized, the potential challenges to inference around local variation in GPA remain. That said, the continuous environment is informative insofar as it highlights that the challenges exist differentially across GPAs—smoothness in average ability is approached more quickly in middling GPAs while in the tails of the distribution of GPA, where the combinatorics are particularly cumbersome to sort through, discontinuities and non-monotonicities persist longer. This suggests that scrutiny be paid to the use of GPA where there are few contributing classes and in the tails in particular. We offer concluding remarks in Section 1.5.

1.2 Two characteristic components of GPA

In this section we separately consider the roles of mean convergence and combinatorics in contributing to the variation in grade-point averages. We then consider their co-existence in the context of treatment evaluation. We first approach this with a simpler notion of the data-generating process—discrete student types (e.g., A, B, and C students, or high-ability and low-ability students). In Section 1.4 we consider a setting where student ability is assumed to be continuous. In that environment, however, grading regimes still subject student performance to being discretize into letter grades—unsurprisingly, then, the problems associated with the combinatoric process by which GPAs converge to their mean values will persist.

1.2.1 Mean convergence

If students of different ability levels don't overlap in their distributions of performance then GPA is no more informative of a student's type than any individual grade would be. For example, if "B students" always draw Bs, and "C students" always draw Cs, it is enough to see one grade to immediately identify student types. For convergence to play a meaningful role, then, we consider environments in which students of different ability levels overlap in their distributions of performance and the associated grades they can experience.

For example, if B students can draw course grades between C and A, and C students can draw course grades between D and B, then a true C can outperform a true B in any given class. However, the inability for C students to *repeatedly* outperform B students will presumably be revealed over time—additional draws increase the strength of the signally component of GPA. That is to say, there is a systematic relationship between the informativeness of GPA and the number of contributing classes.

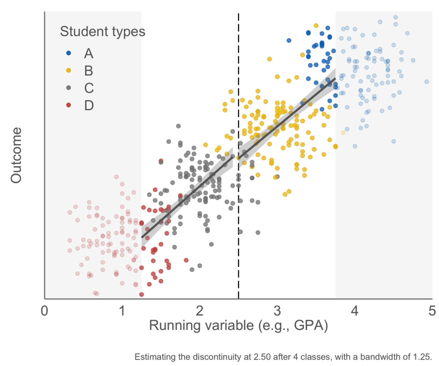
This is the first-order influence of the mean convergence process on the informativeness of GPA and what will ultimately challenge the establishment of all-else equal environments within local differences of GPA. In few classes, there is likely to be B students with GPAs below 2.50 and C students with GPAs above 2.50. As students engage in additional classes, however, B students and C students will eventually separate in GPA—common support will be lost, as students below a 2.50 are eventually all C students while those above a 2.50 are all B students.

In Figure 1 we illustrate the potential pitfalls of this convergence process in the context of regression-discontinuity (RD) designs, where variation in student outcomes local to a GPA threshold drive treatment estimates. In this exercise, we consider four student types and allow student ability to correlate with level differences in expected outcomes—higher-ability students experience both higher grades and better outcomes on average. We then simulate the course taking of 125 students of each type. Each type draws uniformly from the traditional grade points that are within one full grade point of their central tendencies. For example, B students draw 3.00 on average, from the set {2.00, 2.30, 2.70, 3.00, 3.30, 3.70, 4.00}. Likewise, A, C, and D students draw respectively from distributions centered on 4.00, 2.00, and 1.00. With respect to outcomes, we assume that A students have outcomes described by $N(40, 5)$, B students have outcomes described by $N(30, 5)$, C students have outcomes described by $N(20, 5)$, and D students have outcomes described by $N(10, 5)$. (While we visually distinguish student types in Figure 1, this information remains unobservable to the econometrician.)

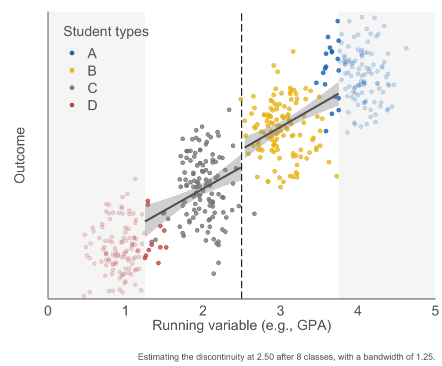
Despite there being no treatment occurring anywhere in the data-generating process, in each panel of Figure 1 we proceed to estimate a discontinuity as though a researcher was inquiring into evidence of treatment occurring at some GPA with a discontinuity estimator. In Panel A, for example, students have drawn four classes and we consider whether a researcher could establish evidence of treatment having fallen on students with GPAs at or above 2.50. Even on visual inspection, we see that there are C students on the right of 2.50 in Panel A and B students on the left—this is beneficial to the estimator, as an abrupt change in the makeup of students at 2.50 would be troubling. Despite level differences in outcomes

Figure 1: The effect of mean convergence on RD estimates in the absence of treatment

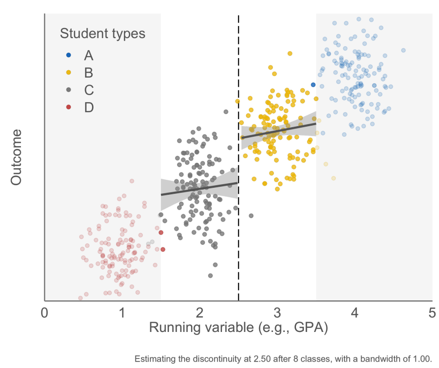
A: After one semester (4 classes)



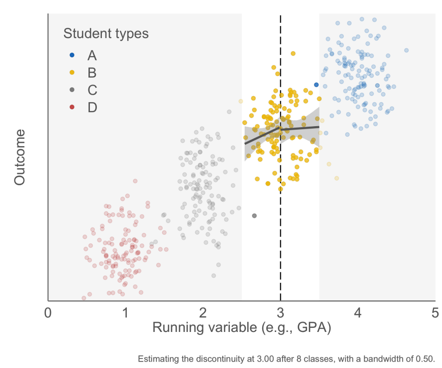
B: After one year (8 classes)



C: After one year with a smaller bandwidth



D: After one year with a smaller bandwidth but where the threshold "splits" a type of student



Notes: In all panels, students draw uniformly from grades (i.e., the traditional grade points) within ± 1 of their median grade, which is centered by their type (i.e., 1, 2, 3, or 4). In all panels there are 125 students of each type (i.e., $n = 125 \times 4 = 500$ students). See Section 1.2.1 for related discussion. (For a more flexible environment in which to explore the variation in RD estimates in simulated GPA data, see `https://glenwaddell.shinyapps.io/RD-in-GPA-data/`.)

across student type, fitting $y_i = f(\text{GPA}_i)$ on either side of a 2.50 GPA threshold yields a confidence interval within which the true $\beta = 0$ is contained. Across the relevant GPAs, there is sufficient overlap (of student types) that the 2.50 threshold does not yet separate C and B students.

In Panel B we reconsider the same students after they have each taken eight classes. As a general rule, we expect that students will begin to separate in GPA as they complete additional classes—again, visual inspection indicates that there are fewer C students above 2.50 in Panel B than in Panel A, and fewer B students below. More to the point, though, fitting $y_i = f(\text{GPA}_i)$ to the same students just one-semester later we see the beginning of mistaking what is unobserved heterogeneity in type (i.e., what we know are just level differences in outcomes in this example) for a discontinuity in y_i at 2.50. If we reduce the bandwidth around the 2.50 threshold, as in Panel C, estimates now suggest a significant discontinuity in outcomes despite the absence of treatment. Polynomial order is a researcher's choice, of course, and is best chosen in light of the specific data under investigation. Thus, our intent in providing these examples is to exemplify the systematic variation in GPA rather than to prescribe how researchers would ideally model regression discontinuity estimators. We consider functional form further in our discussion of treatment evaluation in Section 1.3.1.[1]

Valid RD designs require that potential outcomes are smooth at the discontinuity. However, this is increasingly unlikely to be true as GPA does a better job separating students by ability. What drives the estimated discontinuities in panels A through C is the location of the (placebo) threshold relative to the central tendencies of students in the vicinity of that threshold. With additional classes, higher-ability students are converging to the right of the 2.50 threshold while lower-ability students are converging to the left. Despite the density of students itself remaining smooth across GPA, there is no similar assurance that there is

[1] For a more flexible environment in which to explore the variation in RD estimates, see `https://glenwaddell.shinyapps.io/RD-in-GPA-data/`.

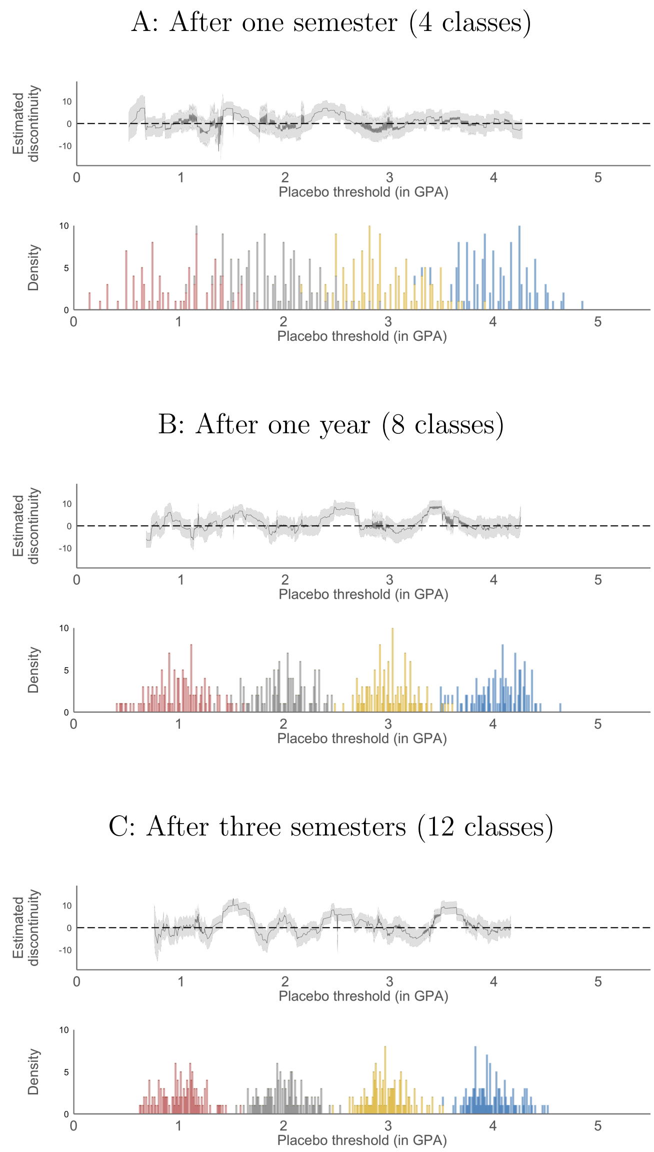
smoothness in student *type* at the threshold.[2] In the limit, for example, if only C types are on the left of the threshold and only B types are on the right, potential outcomes are likely to change discontinuously at the threshold separates them.

In Panel D of Figure 1 we estimate a discontinuity at a threshold that is safely in the middle of one type of student, and choose a bandwidth that only rarely puts weight on other student types when fitting $y_i = f(\text{GPA}_i)$. The implications of mean convergence should be less problematic here—indeed, the traditional RD analysis cannot reject that $\hat{\beta} = 0$, as type is not separated by GPA at the threshold of interest.

As there is an absence of treatment in this data-generating process, the sensitivity in $\hat{\beta}$ across the panels of Figure 1 is disconcerting. In this example, the degree to which different student types overlap with each other in the distribution of GPAs clearly drives treatment estimates—and common support in the relevant range of GPA itself depends on the number of classes students have taken. In Figure 2 we demonstrate the variability in treatment estimates at every two-digit GPA. With an absence of treatment, confidence intervals should generally include zero. Yet, as students draw additional classes and their GPAs converge to their central tendencies, the risks associated with mean convergence again become apparent—with additional classes, point estimates are increasingly likely to deviate from true β. (In each case, we adopt the optimally chosen bandwidth (Imbens and Kalyanaraman, 2012).) In this simulated environment, the central tendencies of students are observable. Thus, we see that the biases are largest at the placebo thresholds that tend to separate students by type—where the overlap of students is lost. Without overlap in student types, which occurs naturally as students engage in additional classes, the average

[2] In all panels of Figure 1 we fail to reject that the density is continuous (using the test provided in McCrary (2008)). As noted in Frandsen (2017), the McCrary test can over- or under-reject the assumption of smoothness when the running variable is discrete. However, the test proposed for use with a discrete running variable is also inappropriate in GPA data, as it relies on the assumption that the support of the running variable has equally spaced intervals. Grade points themselves are unequally spaced, and combinatorics yields unequal spacing. See Lee and Card (2008) and Kolesár and Rothe (2018) for discussions of standard-error estimation and the inference problems associated with research designs in which treatment is determined by a discrete covariate. This relates to our discussion as GPA should arguably be considered discrete data—especially in small numbers of classes.

Figure 2: RD estimates across GPA in the absence of any treatment (i.e., true $\beta = 0$)



Notes: In all panels, students draw uniformly from grades (i.e., the traditional grade points) within ± 1 of their median grade, which is centered on their type (i.e., 1, 2, 3, or 4). In all panels there are 125 students of each type (i.e., $n = 125 \times 4 = 500$ students). See Section 1.2.1 for related discussion.

ability of students on the left side of a cutoff is meaningfully different than that of students on the right side of the same cutoff, and there is a larger change in type coincident with passing over the threshold.

With few discrete types of student, the position of a treatment threshold relative to the expected GPAs that students are converging to influences the comparability of students around GPAs. In Section 1.4 we consider a data-generating process in which student ability is continuous across the GPA domain. There, the implications associated with the number of classes will again be evident. In particular, introducing continuous types highlights that the assumption that the composition of students across GPA changes smoothly is unlikely to be supported in the tails of the GPA domain (where the convergence process unfolds more slowly) or if student performance is itself more diffuse.

1.2.2 Combinatorics

While in the preceding data-generating process we highlight that GPA will tend to separate students by type over time, this convergence process is itself combinatorial. As a combinatorics problem, it is then prone to exhibit discontinuous sorting of students. Thus, in the absence of a smooth convergence process generally, the evolution of GPA over time will hide within it many opportunities for local variation in GPA to be confounded by unobserved heterogeneity.

To consider this process, we derive discrete probability mass functions (PMFs) across the number of classes that contribute to GPA and demonstrate two relevant contributing sources of heterogeneity that are likely to materialize as students of different ability sort into GPAs. First, we demonstrate the implications of students of different ability drawing the same number of classes but from different distributions of performance. Second, we demonstrate the implications of students of different ability drawing a slightly different number of classes, albeit from the same distribution of performance. (In practice, anticipating a mix of both may be most relevant.) In all cases we restrict our attention to GPAs measured at two-digit precision.[3]

Traditionally, there are 13 possible "grade points" that can contribute to a student's GPA.[4] At k classes, this implies 13^k potential grade permutations (i.e., 13^k possible transcripts), and fewer distinct grade combinations.[5] For example, if there is positive weight on all 13 grade points, after two classes there are 91 combinations of grades (in 169 different permutations) that can be realized, and 39 unique two-digit GPAs. With four classes, however, there are 1,820 combinations (28,561 permutations) and 133 two-digit GPAs. By the end of a typical undergraduate experience of 32 classes, there are 21,090,682,613 combinations (13^{32} permutations), at which time all but one of the two-digit GPA between 0 and 4.30 are feasible—a GPA of 0.01 remains unattainable at 32 classes.[6,7] It is within this process that governs how students of different ability sort into GPAs.

Consider the selection into a bandwidth around one such GPA—consider a minimum-GPA requirement of 3.00, for example. As students can often face such a requirement if they are attempting to enroll in a capacity constrained school or major, we first consider this margin at the end of relatively few (four) classes. The combinatorics are also tractable at this point. While there are 37 combinations that result in a GPA of 3.00 after four classes,

[3] While reporting two-digit GPAs is most common, some institutions do report GPA to three-digit precision. Adding precision in this way can exaggerate the combinatoric complexity, but does not change the concerning implications of combinatorics we illustrate.

[4] The traditional letter grades of {F, D-, D, D+, C-, C, C+, B-, B, B+, A-, A, A+}, for example, traditionally map into grade points as $\Gamma = \{0, 0.7, 1.0, 1.3, 1.7, 2.0, 2.3, 2.7, 3.0, 3.3, 3.7, 4.0, 4.3\}$.

[5] In short, permutations require an ordering of the elements, while combinations do not. For example, (A+, B) and (B, A+) are distinct *permutations,* while representing only one *combination.*

[6] For n potential grade points, there are n^k grade permutations that a student can receive over a sequence of k classes. As an unordered sample with replacement, the set of possible grade-point combinations is then

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}\ .$$

. See Brualdi (2009) for an introduction to unordered sampling with replacement.

[7] The set of combinations of grades determines the number and set of feasible GPAs after students have taken some number of courses. To our knowledge, there is no formulaic solution to determine the number of distinct values that are yielded when averaging a full set of combinations with replacement, to say nothing of determining what those distinct values will be. In our setting, this problem is made more complicated by the grade points themselves being unevenly spaced, as well as the fact that we are rounding to some level of precision after calculating an average.

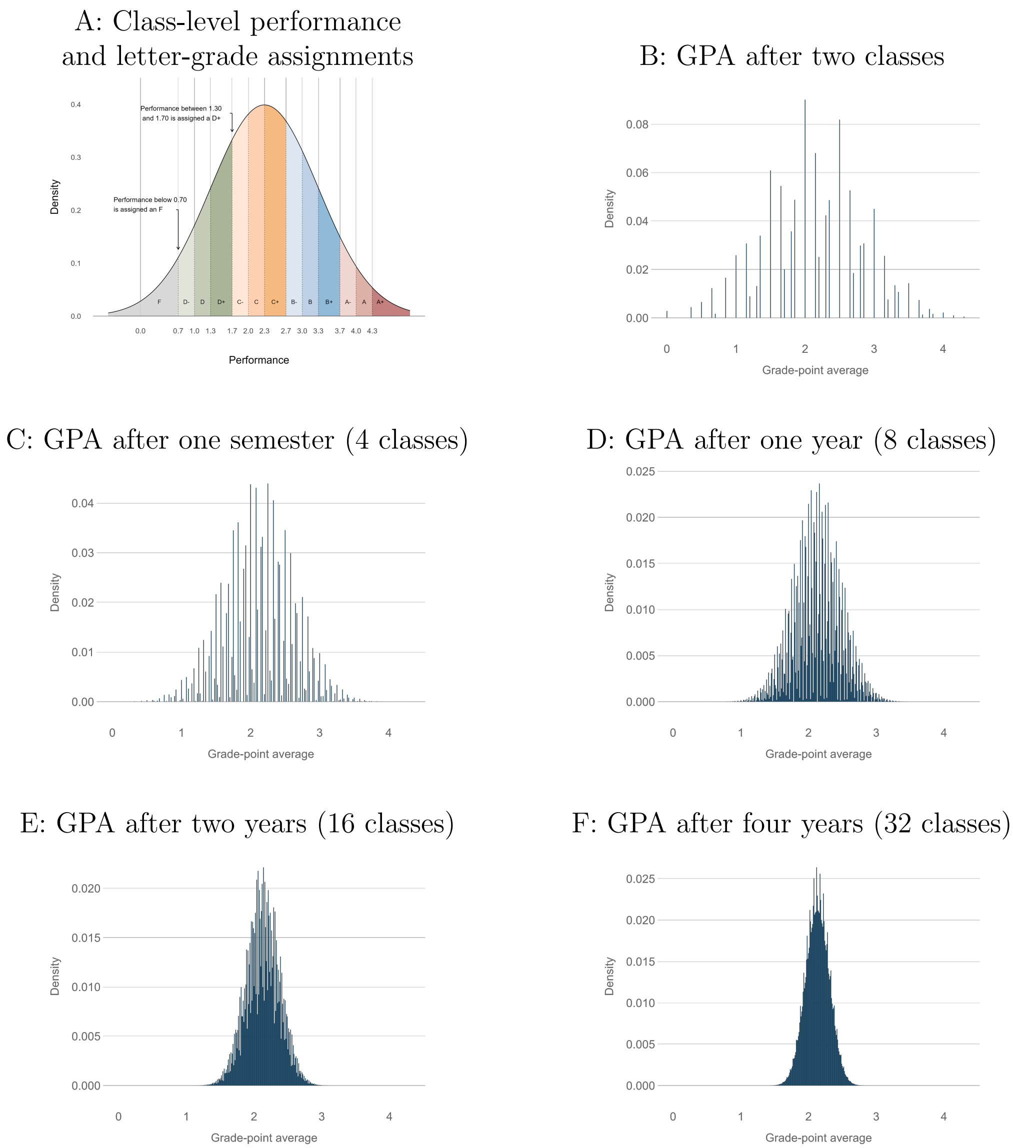
there are only 12 combinations that result in a GPA of 2.98 (the closest feasible GPA to the left after four classes) and 7 combinations that result in a GPA of 3.03 (the closest feasible GPA to the right). Across similar GPAs, then, there are large differences in the number of combinations that can lead to them. If selection into these paths is random we worry less that local comparisons that rely on the comparability of students at these GPAs is misleading. However, these paths are often quite dissimilar. For example, 75 percent of the combinations that lead to 2.98 (8 of 12) include at least one A+. Meanwhile, only 41 percent of the combinations that lead to 3.00 (15 of 37) include an A+, and no student with a 3.03 after four classes can have had an A+. In settings where identification rests on notions of local comparability, to disregard the selection into GPA is to potentially assume away meaningful unobserved heterogeneity.

In fact, given the many combinations that lead to a given GPA it is also important to consider the non-random selection across those who have the same GPA. Among the 37 unique grade combinations that lead to a 3.00 GPA at the end of one term of four classes, for example, some students will have received grades of {B, B, B, B}, while others will have received grades of {B-, B, B, B+}, while still others will have received grades of {C+, B, B, A-}. If course-level grades correlate with potential outcomes—if those paths to 3.00 are selected into differently by meaningfully different students—there is clearly scope for sources of unobserved heterogeneity to plague GPA-related inference. That is, even when GPA is equal, assuming an equivalency across students with that GPA is to assume that each of the combinations that lead to that GPA afford "all-else-equal" comparisons (e.g., 37 combinations at 3.00).

In Figure 3 we derive a series of PMFs that follow the evolution of GPA for a representative student. To generate a set of course-level probability weights we assume that a student draws course-level performance from a continuous probability distribution centered on their ability, and that performance maps into discrete letter-grade outcomes. We visualize this in Panel A for a student who draws performance from the normal distribution $N(2.30, 1)$ which

Figure 3: The combinatorics of GPA for an individual student: PMFs as students complete classes

In all panels, class-level performance is drawn from $N(2.3, 1)$ and letter grades are assigned as the highest grade point surpassed, as in the PMF in Panel A. In panels B through F we plot the probability mass functions over GPAs associated with repeated draws.



Notes: In each panel we plot the probability of earning each GPA between 0.00 and 4.30 in increments of 0.01. See Section 1.2.2 for related discussion.

is then mapped into a letter grade by rounding down to the closest grade point exceeded. For example, a letter grade of C+ is assigned to student i in class c when performance P_{ic} meets or exceeds 2.30 but does not meet the 2.70 that would be required for a B-. That is, $P_{ic} \in [2.30, 2.70)$ maps to a C+, while $P_{ic} \in [2.70, 3.00)$ maps to a B-. In subsequent panels of Figure 3 we plot the PMF that results from a student randomly drawing grades according to this process, at the end of two classes, at the end of the first and second semesters (i.e., 4 and 8 classes), at the end of two years (i.e., 16 classes), and at the end of four years (i.e., 32 classes).

1.2.2.1 When students differ in average performance

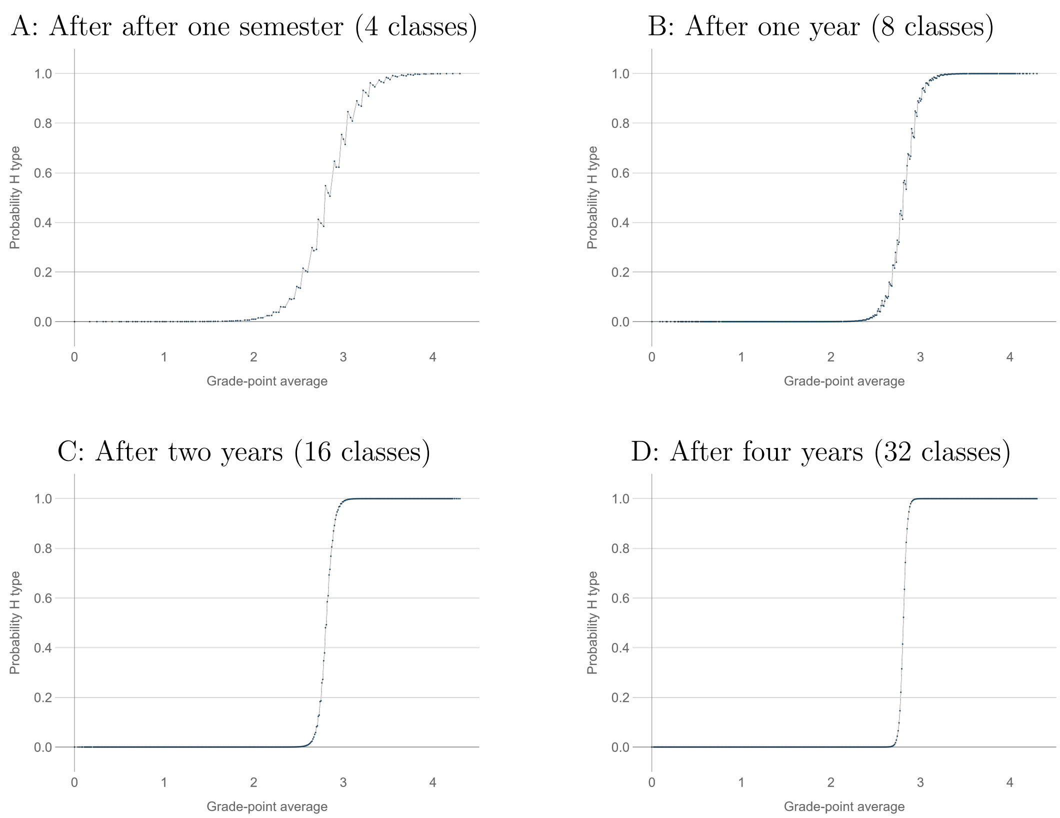
In Figure 4 we introduce a second type of student to the data-generating process and plot the probability that the student at a given two-digit GPA is this new type—an "H-type" who draws performance from $N(3.70, 1)$ but is subject to the same mapping into letter grades. In so doing, we begin to visualize the problems generated by the combinatoric process that sorts student types into GPAs.

The probability that a student is one type or the other should clearly change across GPA—this is how GPA informs type, generally. However, what is noteworthy in Panel A of Figure 4 is that this probability does not change smoothly across GPA, or even monotonically. In short, through the combinatorics of GPA, the aggregation of student performance can induce complex and irregular variation in the student types occupying neighboring GPA—to assume that two students with "similar" GPAs are actually similar in type is at odds with this process. Moreover, where treatment variation is coincident with any such discontinuity in expected type, estimators that rely on local comparability will confound treatment effects and discrete changes in type.

Across panels in Figure 4 it is also clear that this pattern of discontinuity (and non-monotonicity) changes with the number of contributing classes. While there are still non-

Figure 4: Non-random sorting into GPA: When H types outperform L types, on average, how likely is a randomly drawn student an H type?

We evaluate a DGP with two student types—L types and H types, drawing performance from $N(2.3, 1)$ and $N(3.7, 1)$, respectively. For each GPA between 0.00 and 4.30 (in increments of 0.01) we plot the probability that a student with that observed GPA is an H type (i.e., the type with mean performance of 3.7). Across panels, we therefore demonstrate how the sorting of students into GPAs systematically evolves.



Notes: In each case, the population of students is split equally between the two types. See Section 1.2.2 for related discussion.

monotonicities evident after two semesters (Panel B), by the end of three semesters (Panel C) student type is mostly monotonic in GPA.[8] Given these patterns, we cast the discontinuities (and non-monotonicities) that are systematically induced by the combinatorics of GPA as a "small number of classes" problem. In the relevant environments, however, important decisions are often made well before one or two years of classes, so this is not easily ignored in practice. Major choice, for example, will typically occur well before students have taken enough classes for researchers to not worry about this "small n" problem. Moreover, schools and departments often admit students who earn a minimum GPA across very few introductory courses. For example, UC Santa Cruz maintains a 2.80 GPA threshold over three introductory economics classes, which Bleemer and Mehta (2020) exploits in evaluating the return to an economics major.[9] The number of classes at which it is safe to assume that the distribution of expected ability is smooth will depend on the underlying grading distributions, of course. However, as a general rule, the less overlapping are the distributions of performance (e.g., the less overlap there is in the performance of H and L types) the sooner will the relationship between expected ability and GPA become smooth—we consider this in Section 1.4, where we introduce continuously varying ability and varying degrees of overlap in student performance.

1.2.2.2 When students differ in their rates of class taking

In Figure 5 we take an alternative approach to the introduction of student heterogeneity. Namely, we consider a setting in which L and H types perform similarly—we assume $N(3.0, 1)$ for the performance of both types—but that H types engage in additional course taking. For example, H types may have access to resources that L types do not, so differ in their ability

[8] Small and infrequent non-monotonicities do persist through 26 classes. However, after eight classes (two semesters), remaining non-monotonicities are concentrated at low GPAs (e.g., 0.08) and become vanishingly small (e.g., decreases in probability at 26 classes are on the order of 10^{-40}).

[9] Bleemer and Mehta (2020) recognizes the small number of feasible GPAs around their identifying variation, and the relative rarity of some of those that are feasible at just three classes. Adopting a relatively large bandwidth (only excluding students with GPAs below 1.00 in their main specification) increases confidence that combinatoric sorting is less consequential in that setting.

to

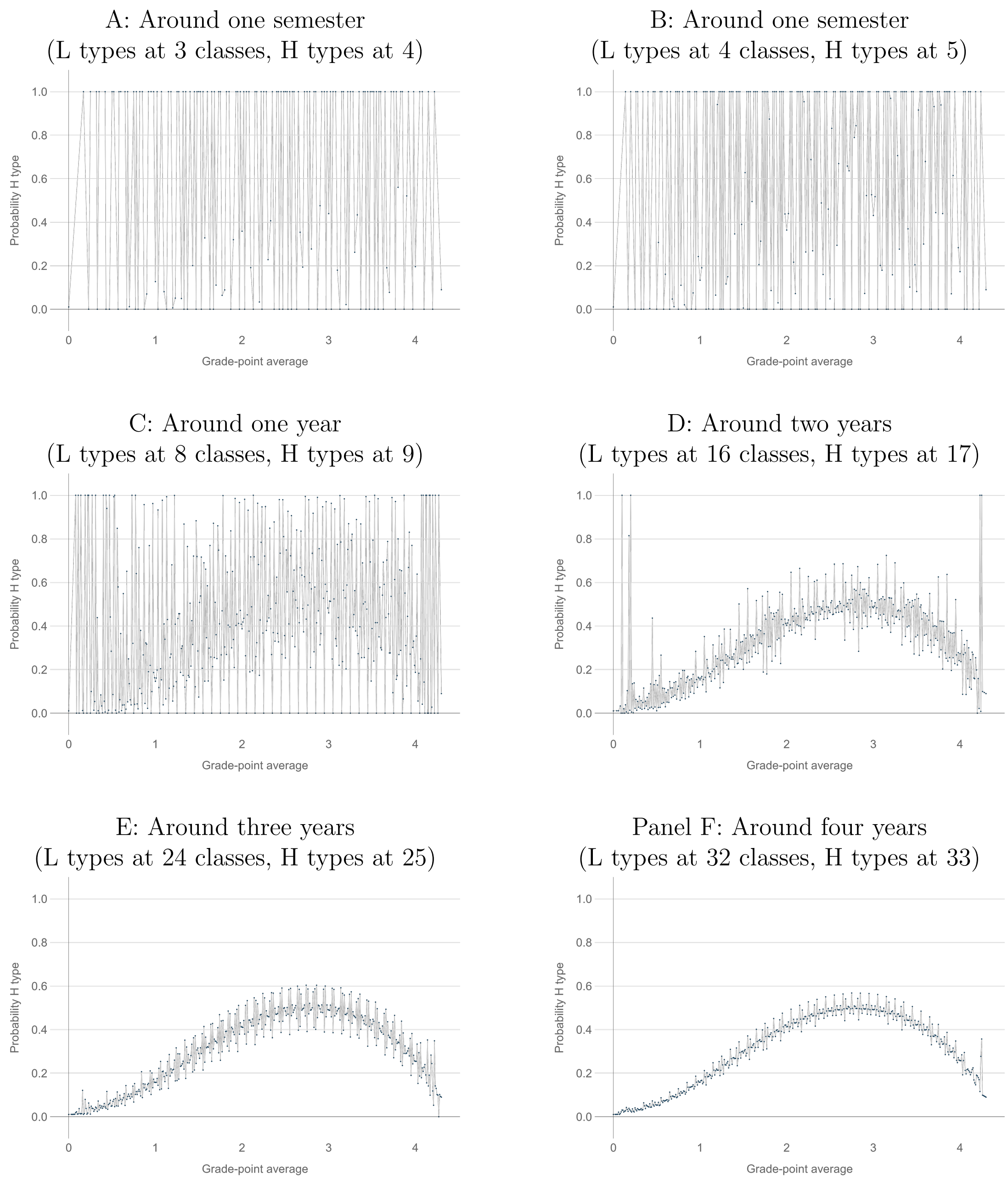
make progress—we presume that these resources are unobservable to the econometrician but correlate with outcomes, thus leading to the potential for confounding sources of variation. Notably, in terms of the combinatoric sorting it can induce, this dimension of heterogeneity will be particularly more stubborn and the problems more persistent.

Across panels of Figure 5 we consider what we anticipate researchers or practitioners possibly doing—assuming that two students are comparable at various points in time without regard for the number of classes they may have taken. For example, in Panel A we consider two students who are both at the end of one semester of coursework with one of them having taken three classes while the other having taken four. In Panel B, we do the same comparison at four and five classes. In each case, for two-digit GPAs between 0.00 and 4.30 we plot the probability that a student with that GPA is the H type—the type to have taken the extra class. If this probability was smooth through the range of GPA, or even locally smooth in places, we would be less concerned—we are used to accommodating smooth changes in the fraction of students who are of a particular type by simply controlling for GPA (e.g., as the running variable in an RD). However, we again see that the underlying combinatorics of GPA yields a significant amount of troubling variation in student type across the domain space of GPA. Across two otherwise-identical students, even taking one additional class can fundamentally change which GPAs are possible. Given the frequent discontinuities—many of them non-monotonicities here—if there is any degree of non-random selection into taking an extra class, then it is reasonable to anticipate that these students are selecting into distinct sets of potential GPAs. This requires attention beyond our typical approach to policy evaluation. Similarly, this pattern can be generated by two students having taken the same number of classes but a slightly different number of credit hours.[10]

[10] For a more flexible environment in which to explore the patterns of average student ability see `https://glenwaddell.shinyapps.io/average-ability-in-GPA/`.)

Figure 5: Non-random sorting into GPA: What if H types are distinguished by taking an extra class?

We evaluate a DGP with two student types—L types and H types both draw performance from $N(3.0, 1)$*, but L types draw* c *grades from that PMF and H types draw* $c+1$ *grades from that PMF. For each GPA between 0.00 and 4.30 (in increments of 0.01) we plot the probability that a student with that observed GPA is an H type (i.e., the type to have taken* $c+1$ *classes). Across panels, we therefore demonstrate how the sorting of students into GPAs systematically evolves.*



Notes: In each panel, the population of students is split equally between the two types. See Section 1.2.2 for related discussion.

1.3 Evaluating GPA-determined treatment

The combinatorics-induced discontinuities and frequent non-monotonicities in student type across GPA are fundamentally challenging to research designs that rely on the comparability of students around small changes in GPA. In this section we consider the two thought experiments of the preceding section in simulated environments. First, we consider the scenario where a student's type is only learned over time as she draws from either better and worse grade distributions. Second, we consider the scenario where all students draw grades similarly, but H types are distinguished by taking a larger number of classes—as the number of class is presumably observable, here we also demonstrate the implications of controlling for this distinguishing factor. This will both demonstrate the importance of considering the component parts of GPA and suggest ways in which identifying treatment in such an environment might be salvageable.

Given the construction of GPA, and the potential for non-random sorting into local GPAs in particular, identifying unbiased estimates of treatment in GPA data is non-trivial. This is especially true the more local is the identifying variation, as concerns over combinatorics are first-order in more-local comparisons. As demonstrated in Section 1.2, a clear violation exists in designs that rely on smoothness around a treatment threshold, for example, and the discontinuities and non-monotonicities in student type across GPAs should give us pause as we consider analyses where GPA is the treatment-assignment variable. (Moreover, with interest in collapsing on *smaller* bandwidths where power allows, we should be particularly concerned that the implications of combinatorics around treatment thresholds could lead to questionable inference from well-powered RD designs.)

We first revisit the data-generating process in which there are two student types (i.e, L types and H types, as in Section 1.2.2) and formally allow for them to experience different average outcomes. In particular, suppose that each student realizes weekly wages, w_i, according to the simple process,

$$w_i = \alpha + \delta \mathbb{1}(\mathrm{H}_i = 1) + e_i \ . \tag{1}$$

The parameterization of (1) will be immaterial, so we roughly mimic the 25^{th}, 50^{th}, and 75^{th} percentiles of weekly incomes among college graduates in the United States in 2020, assuming that H types experience δ-higher average wages.[11] Other than from the level differences in outcomes associated with being an H or L type, then, wages vary randomly. To be clear, there is no treatment-induced variation in outcomes, so we are in an environment in which well-identified models should fail to reject the null hypothesis that there is no systematic discontinuity in outcomes. Nonetheless, this environment exposes researchers to the risk of identifying a discontinuity in w_i, as the combinatorics process itself facilitates a source of non-random selection into GPA by type of student.

1.3.1 When students differ in average performance

In producing Figure 6, we run 1,000 regression discontinuity models at bandwidth of 0.01 through 0.50, in increments of 0.01. (As we produce estimates for models that model outcomes as linear, quadratic, or cubic, in producing Figure 6 we run 150,000 models, in total.) In the creation of 10,000-student panels we follow Section 1.2.2—L types draw performance from $N(2.3, 1)$ and H types draw performance from $N(3.7, 1)$. We then consider whether there is evidence of a systematic discontinuity in outcomes for those with GPAs of 3.00 or above—the midpoint between 2.30 and 3.70.

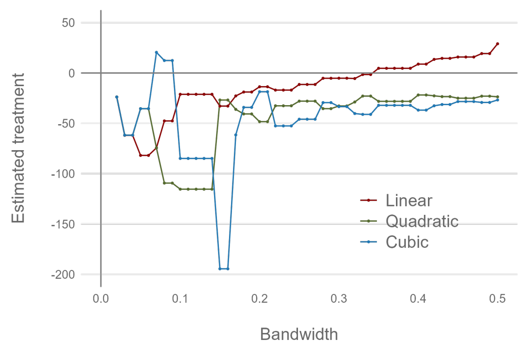
In Panel A we plot mean point estimates in the form of a bandwidth sensitivity plot. While the absence of any treatment supports the expectation that there should be no system-

[11] Assuming $\alpha = \$1,133$, $\delta = \$566$, and $e_i \sim N(0, 300)$ centers our DGP on the weekly incomes of college graduates, approximating the first quartile (\$977), median (\$1,416), and third quartile (\$2,110), according to the Usual Weekly Earnings of Wage and Salary Workers section of the Current Population Survey. (See Bureau of Labor Statistics (2020) for details.)

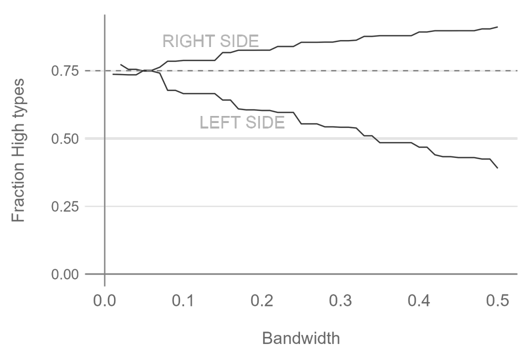
Figure 6: Bandwidth sensitivity in RD estimates reveals evidence of combinatorics

In the absence of any treatment, we retrieve estimates of the discontinuity in outcomes at GPAs at or above 3.0. We evaluate a DGP with two student types—L types and H types drawing performance from $N(2.3, 1)$ and $N(3.7, 1)$, respectively. Due to combinatorics, the sorting of L and H types into GPAs potentially leaves behind discontinuities in type, violating the smoothness assumption around the RD threshold. As H types are level different in outcomes, this violation is transmitted through to estimated treatment effects. At smaller bandwidths, estimates increasingly reflect combinatoric sorting. See Section 1.3 for related discussion.

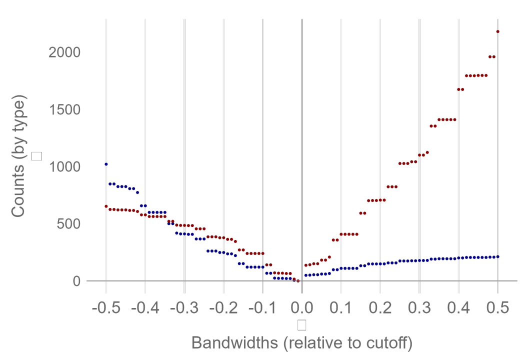
A: Bandwidth sensitivity after one semester (4 classes)



B: On either side of the cutoff, the fraction of "High" types



C: The combinatorics-induced sorting of types across bandwidths



Notes: In each panel we consider bandwidths in increments of 0.01 and report mean treatment estimates over 1,000 simulated panels of 10,000 students.

atically occurring discontinuities at the threshold, the estimated treatment effects are clearly deviating from zero. In this environment, the bias also tends to be large in magnitude—where average weekly wages are $1,416, estimates discontinuities in wages as large as $-81 (-0.19$\sigma$) when wages are fit to a linear function of GPA. The addition of more flexibility in the fitting of w_i does not solve this problem—a cubic function of GPA produces mean discontinuities in wages as large as $-194 (-0.46$\sigma$), for example. Note that negative discontinuity estimates at smaller bandwidths mirror the decreasing probability that an H type is observed at a GPA of 3.00, first seen in Panel A of Figure 4.

Of particular interest in Panel A of Figure 6, however, is the sensitivity to bandwidth choice revealed in this exercise. In panels B and C we use this simple environment to visualize the underlying mechanics behind the sensitivity of point estimates. In Panel B we plot the fraction of students who are H types across the same range of bandwidths (0.01 through 0.50) separately for those to the left- and right-hand sides of the 3.00 cutoff. The lumpy introduction of H and L types to the sample as bandwidth increases is clearly evident—other than noise, this imbalance is the only factor that drives estimates away from zero in Panel A. We see this from a different perspective in Panel C, where we plot counts of student type across bandwidths. While there are bandwidth adjustments that do not trigger changes in the number of observations at all—by its nature, combinatoric sorting will yield GPAs that are not occupied, for example, so changes in bandwidth need not always lead to changes in sample size. When increasing the bandwidth does allow "new" GPAs to enter the sample there are discrete changes in the number of H and L types. It is at these same GPAs that mass then shifts discretely toward one type of student or the other—this abrupt tipping of the balance explains fully the change in point estimates across bandwidths.

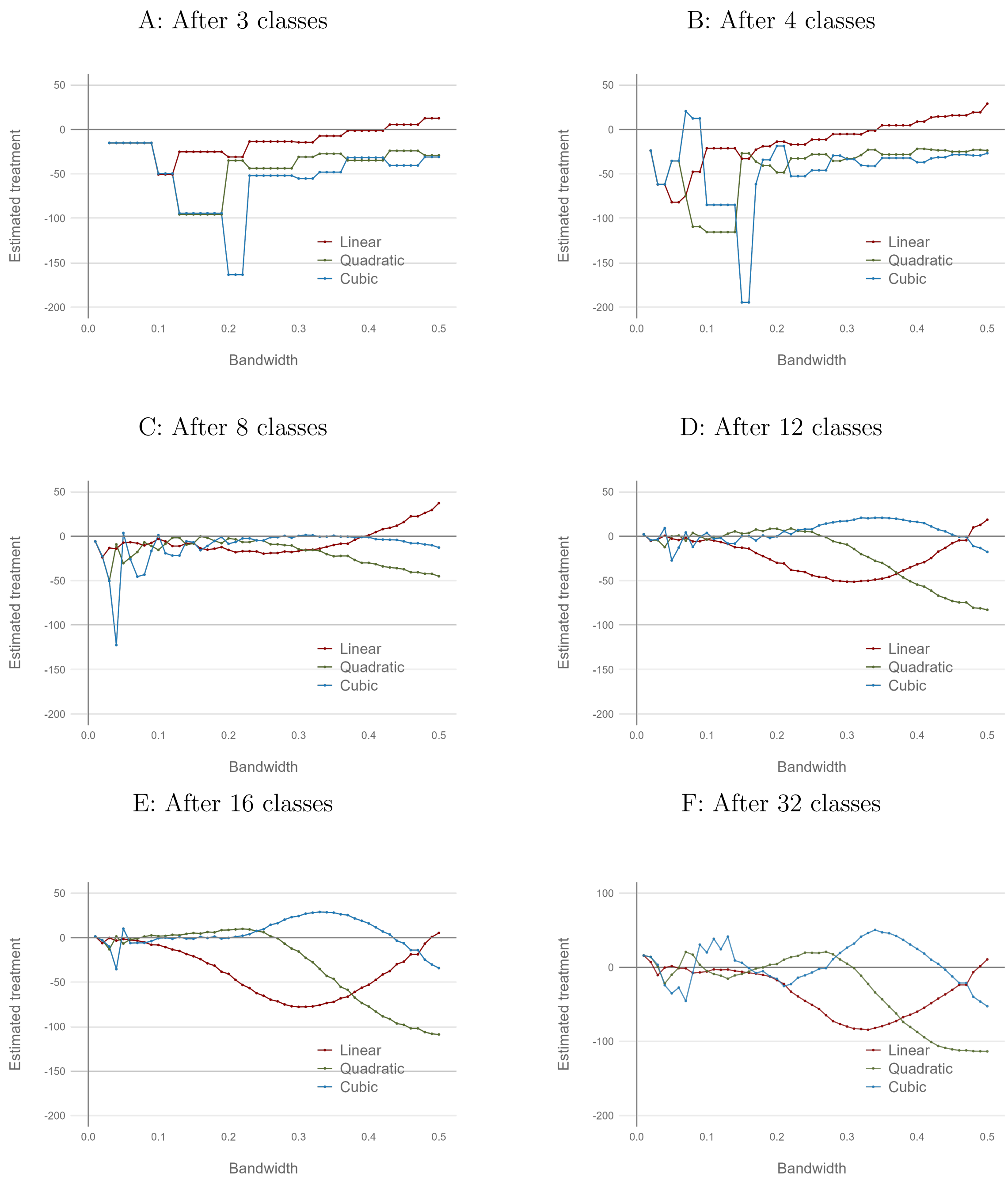
Notably, even with combinatorics playing an active role in facilitating the non-random sorting of student types into GPA, the density of students can still appear smooth. In fact, our DGPs routinely pass standard tests for changes in the density of students around the threshold (i.e., McCrary 2008; Frandsen 2017). It is simply the composition of ability types at those GPAs that drives the discontinuities. Despite their heterogeneity, the underlying combinatorics of GPA will sort students into close proximity.

In Figure 7 we demonstrate the systematic evolution of this sorting over time, with bandwidth sensitivity plots at various increments of progress through a number of classes. In each case, we see more sensitivity of discontinuity estimates to combinatoric sorting in fewer classes and at smaller bandwidths. After three classes (Panel A), point estimates evidence combinatorics even at larger bandwidths—there are notable jumps in discontinuity estimates occurring throughout the range we report. In the relevant literature, there is no commonly adopted bandwidth—Kane (2003) adopts 0.02, for example, while the preferred specification in Ost et al. (2018) adopts 0.50, Bell (2021) adopts 1.00, and (Bleemer and Mehta, 2020) implicitly adopts 1.80 on the left side of a 2.80 treatment threshold and 1.5 on the right. At four classes (Panel B), the role of combinatorics in driving discontinuity estimates is even more evident—the fourth class opens up additional GPAs at which L and H types can sort combinatorially, and the discrete change in composition is responsible for the abrupt changes in estimates. After eight classes, and again at twelve, the composition of students who contribute to the identifying variation across bandwidths changes less discretely—this is generally true as the number of classes increases. Thus, with more classes contributing to GPA, the imprint of combinatorics fades and estimates change more smoothly with bandwidth. Notably, the sensitivity of point estimates to combinatoric-induced bias remains evident in smaller bandwidths, even at 16 and 32 classes.

As students engage in additional classes, their GPAs converge to their expected values. As such, discontinuity estimates are generally less sensitive to combinatorics. Consistent with combinatorics and mean convergence trading off, however, the composition of students on either side of the treatment threshold begins to drive estimates. This sorting does not reverse with additional classes, and larger bandwidths make the difference between those on the left and right of the threshold larger. With student types disproportionately selecting onto the left- and right-hand sides of the threshold, the point estimates at higher numbers of classes are

Figure 7: Bandwidth sensitivity in RD estimates at different points in time

As in Panel A of Figure 6, we retrieve estimates of the discontinuity at GPAs at or above 3.00. (As there is no treatment at 3.00, these should be zero.) This demonstrates the important role of combinatorics at smaller numbers of classes and at smaller bandwidths, trading off with the challenges associated with mean convergence at larger numbers of classes and at larger bandwidths. See Section 1.3.1 for related discussion.



Notes: In each we panel consider bandwidths in increments of 0.01 and report mean treatment estimates across 1,000 simulated panels of 10,000 students.

also particularly sensitive to the shape of $f(\text{GPA}_i)$, and are often driven away from zero. This sensitivity partially reflects that there are only two ability types in this environment. If we assume continuous ability (Section 1.4), outcomes are still identified by differences in student composition on either side of the treatment threshold, but the differences are less stark since there is more overlap of student types throughout the distribution of GPA, and less potential for the complete loss of overlap that occurs here. However, this sensitivity also reflects the more general concern regarding choices of function form in regression discontinuity designs, and polynomial order in particular (Gelman and Imbens, 2019).

In the end, the roles of combinatorics and mean convergence are in tension, and differentially so as students progress through course taking. At smaller numbers of classes, estimates are more sensitive to combinatorics. Here, larger bandwidths can act as a mitigating device as they leave parameter estimates less sensitive to the *types* of student selecting into particular GPAs. However, at larger numbers of classes combinatorics induces less local variation in student type and smaller bandwidths can become appropriate. As evident in Panel D, however, larger bandwidths can expose researchers to lost comparability. In this case, after twelve classes, student GPAs have converged sufficiently to their expected values that there are only H types on the right of 3.00, while there is still a mix of both L and H types on the left of 3.00. The sign of the bias induced by this loss of overlap will depend on the interaction of bandwidth selection and the functional form assumptions (e.g., the particular polynomial allowed for in modeling $w_i = f(\text{GPA}_i)$.

1.3.2 When students differ in their rates of class taking

In this section we recast the problem as one in which both types draw repeatedly from the same grade distribution, but H types simply take one extra class (i.e., they take classes at a faster rate) than L types. (In this experiment we shut down entirely on any heterogeneity coming from grades themselves. All students draw all performance from $N(3.00, 1)$, which are converted to letter grades following the simple assignment rule we have used above.) This highlights the problem that can arise solely due to the mechanistic sorting of combinatorics, which allows some students to populate GPAs that other students simply cannot—or cannot without one additional classes. As the number of classes contributing to a student's GPA is often observable, we follow this by considering the implications of controlling for the number of classes in regression analyses. In all panels of Figure 8 we estimate discontinuities in outcomes at 3.00, for bandwidths between 0.01 and 0.50 in increments of 0.01.

In Panel A, L types have taken eight classes and H types have taken nine classes. By construction, there should again be no discontinuity in outcomes in this environment. Yet, at small bandwidths there is again the imprint of the combinatoric sorting of H and L types differentially—here capturing inclinations of H types to take one more class. Consistent with combinatoric sorting, the bias is unsignable in general and particularly problematic at smaller bandwidths where the density of student types can be different either side of a given GPA and can change abruptly as different bandwidths allow different GPAs to populate the estimator. As both types are converging to the same expected value in this setting, higher bandwidths do not expose the estimator to biases associated with a loss of overlap—estimated discontinuities approach zero. More to the point, controlling for the source of this confoundedness (i.e., the number of classes contributing to GPA) yields estimated discontinuities that are not different from zero throughout the range of bandwidths.

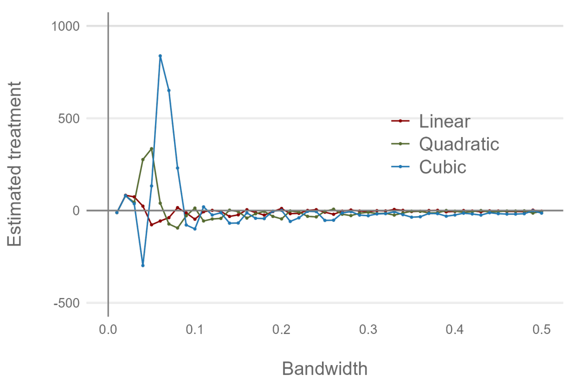
In Panel B we consider the same experiment at a larger number of classes. While the magnitude is smaller, we again see a systematic sensitivity to combinatorics. At the same time, it is notably larger than what was evident in Figurerfig:rd-classes, where heterogeneity was introduced through differences in mean performance. This is consistent with the combinatorics problem at c classes yielding a different set of paths to a given GPA than is the case at $c+1$ classes. Moreover, given the difference in course taking, there is less opportunity for L and H types to both occupy the same GPA. Thus, in this setting, changes in bandwidth are more likely to expose the estimator to discrete changes in the type of student contributing

Figure 8: Bandwidth sensitivity in RD estimates with variation in the number of classes contributing to GPA

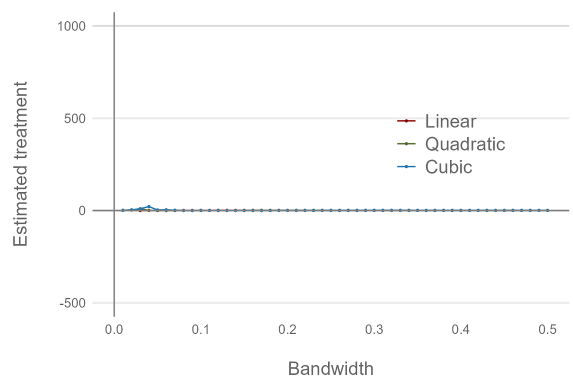
As in Figure 5, students draw performance from $N(3.0, 1)$ and have letter grades at the class level assigned accordingly. However, L types draw c grades from that PMF and an equal number of H types draw $c+1$ grades from that PMF. As H types are level different in outcomes, RD estimates again reflect combinatoric sorting of different students types into close proximity around any threshold. The bias induced by combinatorics is eliminated when we control for the number of classes taken—here, this is the only source of heterogeneity in course taking, as L and H types draw performance similarly. See Section 1.3.2 for related discussion.

A: L types completing 8 classes, H types 9

Not controlling for number of classes

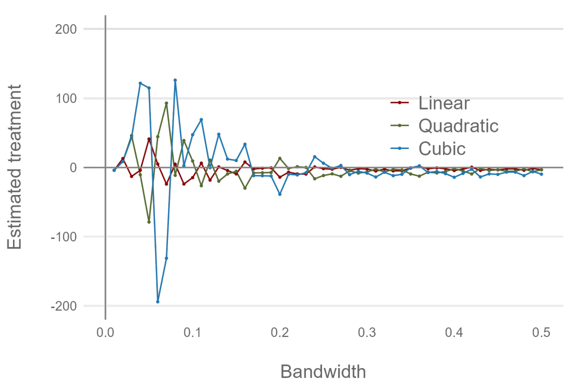


Controlling for number of classes

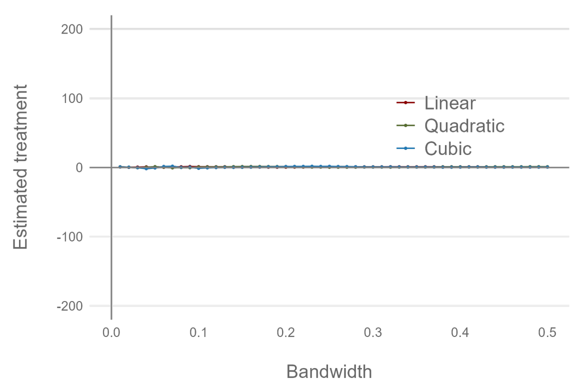


B: L types completing 12 classes, H types 13

Not controlling for number of classes



Controlling for number of classes



Notes: In each we panel consider bandwidths in increments of 0.01 and report mean treatment estimates across 1,000 simulated panels of 10,000 students.

to identification, and for longer (i.e., for more classes).

1.4 Continuous student ability

While it is possibly natural to consider discrete student types at the traditional letter-grade categories, considering a version of the data-generating process in which student ability is continuously distributed further demonstrates the challenges associated with local comparisons of GPA. Thus, we assume that student ability, A_i, is uniformly distributed across the domain of grade-point averages (i.e., $A_i \in [0, 4.30]$). In order to consider the underlying dispersion in student performance, we also vary the noise permitted in student performance across three difference scenarios. Specifically, we allow for low, medium, and high signal-to-noise ratios by reporting (i.e,. $\sigma = 2$, $\sigma = 1$, and $\sigma = 0.5$, respectively). Class-level performance, P_{ic}, is then drawn from $N(A_i, \sigma)$.[12]

In Panel A(i) of Figure 9 we plot representative distributions of potential class-level performance faced by students of different ability levels, with $P_{ic} \sim N(A_i, 2)$. In an environment where performance is noisier, there is considerable overlap in the potential grades students receive in this environment—the mean convergence is slower in this environment. In Panel A(ii) of Figure 9 we then plot the expected ability of students given their observed GPA. As these expectations change over time, we plot the relationship separately for various numbers of classes. Despite continuously distributed ability, the distribution of grades themselves is discrete and students are again only sorted into GPAs as they become feasible—the sparseness of GPA when the number of classes is small is immediately evident in the (ii) plots of Figure 9. Observations in Panel A(ii) fill in, for example, as additional classes contribute to GPA according to the combinatorics process we have discussed earlier. This is not different

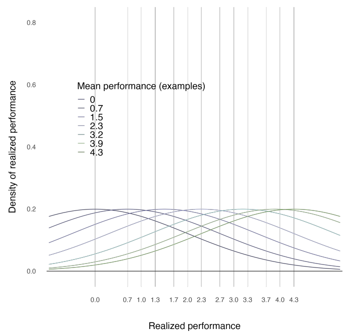
[12] This clearly implies that students are drawing performance outside of the bounds of [0, 4.30]. As an alternative way of modeling the underlying performance of students, assuming that performance is drawn from truncated normal distributions leads to qualitatively similar patterns and magnifies the concerns we document—this is expected, as low- and high-ability students are less able to separate in small numbers of draws. The concerns we document are evident in many ad hoc distributions of performance, mimicking common conceptions for how students might arrive at grades (e.g., department norms for the assignment of

Figure 9: Continuous ability: Expected student ability at realized GPAs

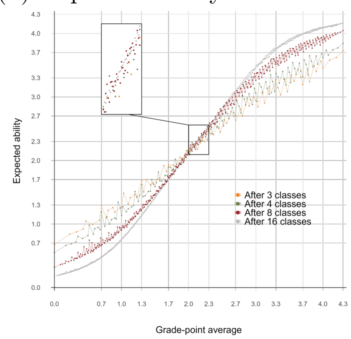
We evaluate a DGP with continuous student types—drawing ability A_i uniformly between 0 and 4.30, and performance from $N(A_i, \sigma)$. For each two-digit GPA between 0.00 and 4.30 we plot the average ability among students with that observed GPA. See Section 1.4 for related discussion.

A: Low signal-to-noise in class performance ($\sigma = 2.0$)

(i) Class-level performance across ability

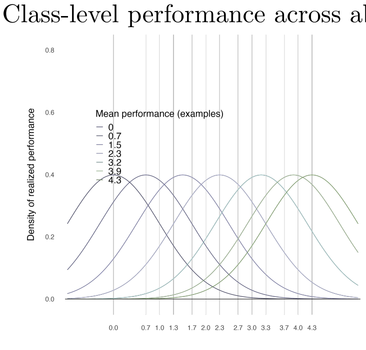


(ii) Expected ability across GPA

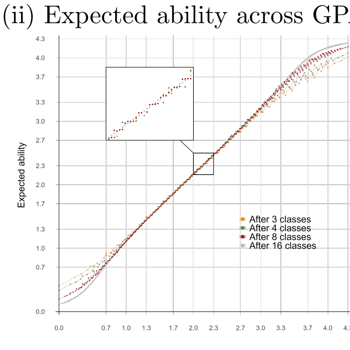


B: Medium signal-to-noise in class performance ($\sigma = 1.0$)

(i) Class-level performance across ability

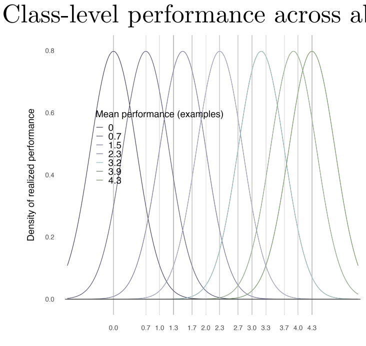


(ii) Expected ability across GPA

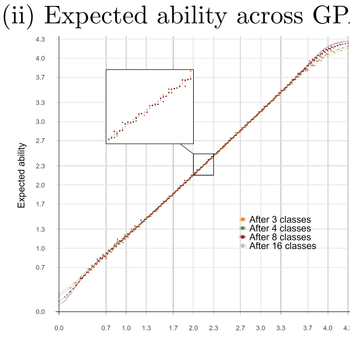


C: High signal-to-noise in class performance ($\sigma = 0.5$)

(i) Class-level performance across ability



(ii) Expected ability across GPA



in an environment with continuous ability.

Also consistent with the simpler DGPs we consider above, while large differences in GPA are again associated with the anticipated change in expected student ability, close inspection again reveals the tell-tale signs of the combinatoric sorting of students of different ability into close proximity. Indeed, the non-monotonicities still occur regularly. Though the relationship between ability and GPA is smoother with continuous types (compared to Figure 4, for example), we therefore continue to characterize the relationship as highly non-linear, with non-monotonicities particularly prevalent when there are fewer classes contributing to GPA, when there is more variance in student performance, and in the tails of the distributions of observed GPA generally. As it is often in small numbers of classes that decisions are made (e.g., major choice, re-enrollment) and in the tails of GPA that policies often bind (e.g., probationary status, admittance into over-subscribed programs), it is particularly problematic in practice that the effects of this sorting are persistent in these areas.

As non-monotonicities are easily quantifiable across the domain of GPA, we report the frequency of these combinatorics-driven non-monotonicities for various numbers of classes in Table 1. In so doing, we note two aspects to what regularly occurs. First, as students accumulate classes and the set of feasible GPAs increases, the frequency of non-monotonicities increase and then decreases. In our simulated environment, this peaks at eight classes. There are 385 feasible GPAs at the end of eight classes, implying that there are 384 opportunities to evaluate an increase in GPA. Of them, roughly 27-to-50 percent of the times that GPA increases, average ability actually declines. This frequency is falling off considerably by 16 classes (i.e., 6-to-29 percent). There are no non-monotonicities at the end of 32 classes. The second pattern notable in Table 1 is the role played by the informativeness of performance, and therefore grades. Across columns (2) through (4) we increase the signal-to-noise ratio in underlying student performance. This, generally, tends to speed up the convergence process and the imprint of combinatorics on the variation in GPAs fades more quickly. However,

letter grades).

Table 1: Frequency of non-monotonicities in expected student ability across realized GPAs

We evaluate a DGP with continuous student types—drawing ability A_i uniformly between 0 and 4.30, and performance from $N(A_i, \sigma)$. With GPA observable at two-digit precision, there are a maximum of 430 GPAs and thus 429 possible non-monotonicities across GPAs. Non-monotonicity is defined as any decrease in expected ability between adjacent viable GPAs, at 0.01 precision. See Section 1.4 for related discussion.

Number of contributing classes	Feasible two-digit GPAs	Number of GPAs at which there is a decline in average ability		
		$\sigma = 2.0$	$\sigma = 1.0$	$\sigma = 0.5$
	(1)	(2)	(3)	(4)
3	79	34 (43.6%)	21 (26.9%)	15 (19.2%)
4	133	70 (53.0%)	44 (33.3%)	30 (22.7%)
8	385	193 (50.3%)	124 (32.3%)	104 (27.1%)
16	420	124 (29.6%)	44 (10.5%)	25 (6.0%)
32	430	0 (0%)	0 (0%)	0 (0%)

even with less noise in student performance the frequency of them is still meaningful in the range of classes where GPA-based policies are likely to be relevant to decision making. At four classes, for example, even where performance distributions are the least diffuse (i.e., drawn from $N(A_i, 0.5)$), 23 percent of the times that GPA increases, average ability declines. At eight classes, this occurs 27 percent of the times that GPA increases.

That non-monotonicities are more prevalent when there are fewer classes contributing to GPA reflects that combinatorics initially dominates the sorting of students into GPA. However, as students converge to their expected GPAs, combinatoric sorting is less first-order and the heterogeneity in ability at a given GPA diminishes. At the simplest level, while even very low- or very high-performing students will draw middling grades on occasion—this is what propagates through the combinatorics creating non-monotonicities—continuing to draw middling grades is a very low probability path for them. Thus, low- and high-performing students are more likely to be present in middling GPAs initially, but they vacate middling GPAs quickly with repeated draws.

In the tails, however, non-monotonicities notably persist. This is due to the asymmetry with which students in the tails experience letter grades that reflect variation in performance. The lower is student ability, the more common it is for students to be shielded from downside risk (i.e., from draws of $P_{ic} < 0$). The higher is student ability, the more common it is for students to not be credited for upside risk (i.e., for draws of $P_{ic} > 4.30$). Middling students experience both positive and negative (i.e., offsetting) shocks.[13] Thus, in the tails of the distribution of ability, expected GPAs do not separate students as effectively, and the combinatorics process of sorting students into GPA unfolds more slowly. The noise in student performance also operates in this asymmetry—greater variance in the distribution of performance implies that the shocks to performance that go unmeasured by grades are more common, which implies that the stickiness in the tails of GPA is itself exaggerated. (See panels of Figure 9.)

As expected ability is smoother in observed GPA with continuous types, it is also the case that identification strategies that rely on local smoothness will be more robust to the challenges associated with the combinatoric sorting of students into GPA. However, at any

[13] Assigning continuous performance to letter grades (and grade points) according to a "meets or exceeds" rule implies that P_{ic} will always exceed the assigned grade—student i will always receive a grade point less than her actual level of performance. As such, the felt weight of downside risk is magnified in measured performance, which universally leads to the expected ability of a student at a given GPA being higher than that GPA—in that way, higher ability students are positively selected into observed GPAs. (Though we do not show the 45-degree lines in Figure 9, the relationships between expected ability and GPA is lying mostly above the 45-degree line.) There is also an asymmetry on either end of the GPA domain, however—at low and at high GPAs. This shape is due to an underlying asymmetry in the potential to offset this downward bias in average ability. While a 4.30-ability student is shielded from upside risk (i.e., never credited with performance realizations above 4.30), a 0.00-ability student is shielded from *downside* risk (i.e., draws below zero). Moreover, while being shielded from upside risk goes in the same direction as the universal implications of a "meets or exceeds" rule, being shielded from downside risk works to offset this bias. In general, the lower is student ability the more likely the student is being shielded from downside risk and the larger is the offset to the effect of the "meets or exceeds" rule. (If we instead adopt a simple rounding rule in the assignment of grade points, the asymmetry of course remains.)

of the GPAs where there is a systematically occurring discontinuity in expected ability, treatment will be inseparable from coincident differences in student ability. At smaller bandwidths, in particular, naturally occurring discontinuities in expected ability (some of them changing sign) clearly still confound treatment, well into the relevant range of course taking and decision making. In general, though, researchers do well to avoid attributing differences in outcomes to locally varying treatment at any of the anticipated discontinuities in expected ability across GPAs. (While we can identify the locations of such discontinuities in our simulated environment, knowing where they occur in practice will require knowledge of the underlying DGP—something we anticipate being unobservable to researchers.)

1.5 Conclusion

We first model the process through which students arrive at GPAs, highlighting that students converge to their expected GPAs through a combinatoric process. For a given number of number of classes and grades, the set of feasible GPAs varies systematically with the underlying combinatorics. This problem determines both the paths by which students arrive at given GPAs and, ultimately, the paths they will have experienced in converging to their expected GPAs over time. In the context of treatment evaluation, we show how these two mechanisms (i.e., mean convergence and combinatorics) tradeoff as students engage in additional classes and thereby interfere with the interpretation of variation in GPA in rather complex ways.

While collapsing on students with more-similar GPA sounds like an approach to satisfying an "all-else-equal" condition, combinatorics is most active in local comparisons. As such, it tends to confound differences in GPA *more* when making more-local comparisons. This is especially true where the number of classes is also small, and where student-level performance is itself more diffuse—with additional noise in performance, combinatorics can bring students with very different ability levels into surprisingly close proximity. In the context of treatment evaluation, we demonstrate that this induces a form of combinatoric selection through which students on either side of a given GPA can end up being different, which calls into question the validity of identification strategies that rely on local comparisons across GPA. In short, there is little reason to anticipate that there is smoothness in student *type* across GPA. Moreover, where treatment variation is coincident with any such discontinuity in expected type, estimators that rely on local comparability will confound treatment effects and discrete changes in outcomes that are associated with changes in the composition of students.

CHAPTER II

THE EFFECT OF VIOLENT VIDEO GAMES ON VIOLENT CRIME

The analysis in this chapter was coauthored with Gretchen Gamrat. Gretchen performed much of the cleaning and wrangling of the Nielsen product sales data necessary for this project. I was responsible for the cleaning of the NIBRS crime data used herein, as well as the development of our empirical strategy, as well the implementation of that strategy. The writing contained in this chapter is entirely mine.

2.1 Introduction

Around 40 percent of Americans believe that there is a relationship between violent video games and violent behaviors, and 32 percent of those who play video games believe the same.[14] With the media's support (Markey et al., 2015), this belief is understandable. Moreover, given that roughly three out of four Americans play video games—and that the average gamer plays 14 hours per week[15]—any positive causal effect of violent games on violence would be a societal concern. Yet, there is little evidence to support the hypothesis that violent video gaming leads to an increase in violent criminality.

We contribute to this area by examining whether increased exposure to violent video games has an effect on violent crime rates. We measure exposure to violent games at the county level through a dataset that contains video game sales records from a set of retail chains starting in 2007. This dataset allows us to exploit both the timing of violent video game releases and variation in county-level purchase of those releases to estimate the effect that violent video games have on violent crime.

Our identification amounts to an instrumental variables strategy that leverages a technological limitation on the ability to access video games. Namely, video game disks work only when paired with their matching platform. For instance, an Xbox 360 game disk works

[14] See `https://www.pewresearch.org/internet/2015/12/15/gaming-and-gamers/`.
[15] See `https://www.npd.com/news/press-releases/2020/more-people-are-gaming-in-the-us/`.

only on an Xbox 360 console. When *Halo 3*, a violent science-fiction game, was released in 2007 exclusively for the Xbox 360, the only way to play the game was to buy the disk and insert it into an Xbox 360. Since all game disks are coded for a specific platform, a person's ability to access a new game that releases on one or several platforms varies based on the platforms she already owns. By exploiting this hardware-software link, we isolate variation in exposure to newly released games that is exogenous from any unobserved determinants of crime. To implement this identification strategy, we use variation in lagged platform-specific game sales to predict the sale of newly released violent video games. We use these predictions to then estimate the effect that increased exposure to a violent game release has on weekly agency-level violent crime rates obtained from the National Incident-Based Reporting System (NIBRS). We find no evidence that violent game releases lead to increased short-term violent crime rates. Further, the point estimates that we recover are small enough in magnitude to rule out the large effects that have been suggested by some others. We also find evidence suggesting that increased exposure to violent games causes a decrease in the rate of violent sex offenses.

The public impression that violent video games *should* lead to violent criminality is informed by a body of social science research on the connection between violent media consumption and aggressive behavior. In laboratory experiments, for example, the effect of violent video game consumption has been studied extensively. In this setting, researchers have found that playing a violent game leads to increased aggression compared to playing a nonviolent video game. The laboratory has allowed researchers to establish, for instance, that violent video games cause an immediate increase in bodily production of stress hormones linked to fight-or-flight responses (Gentile, Bender and Anderson, 2017). To take these findings outside of the laboratory, researchers have also conducted surveys to explore the relationship between violent gameplay and "real-life" behavior (e.g., Möller and Krahé (2009)). These studies find that those who play violent video games more frequently are also more likely to report anti-social and violent behavior, and the reported effect sizes are similar in magnitude to laboratory studies. Collectively, this body of work suggests that the correlation between violent gameplay and aggression is large enough in magnitude to constitute a public health concern, given the number of people playing violent video games. In a meta-analysis of studies of violent gameplay and aggression, Anderson (2003) finds that the correlation between violent video games and aggressive outcomes is similar in strength to the correlation between passive secondhand smoke exposure and lung cancer.

Yet, surveys cannot account for the fact that those who play violent video games in real life have selected into the activity, and are likely different from those who do not select into violent gameplay. For example, Ward (2010) finds that controlling for gender, race, and geography eliminates a positive association between increased video game play and adolescent fighting for all but the most heavy video game users. Beyond demographic variables, other unobserved third factors could easily cause selection into both violent gameplay and violent behavior. The simple fact that video game players *believe* in the cathartic effect of violent gameplay (Olson, Kutner and Warner, 2008) suggests one such confounder. Namely, those with heightened latent aggression turn to video games in an attempt to relieve that aggression, and are also more likely to resort to real violence when no simulated violence is available (e.g., starting a fight in the schoolyard). In the presence of selection like this, correlational studies will not identify the causal effect of playing violent video games on violent crime. Further, the causal effects of violent gameplay on aggression that have been well-identified in the laboratory lack the external validity to inform us about effects on criminality for a variety of reasons. Normally, laboratory researchers are constrained to studying only the immediate effects of exposure to a violent game, as outcomes are normally measured in the minutes or hours following gameplay. Also, laboratory studies cannot evaluate the impact of violent gameplay in the research subjects' normal lives—the setting in which we are concerned about the effect of violent media.[16]

[16] Researchers also debate how to interpret the results of laboratory experiments *vis-à-vis* "aggression." For instance, some argue that outcome measurements used in laboratory studies (e.g., whether a participant chooses to conclude an open ended story prompt with a violent or non-violent resolution) have an unclear link to an aggressive state of mind. Ferguson (2007b) provides a good overview of this and other common

Policymakers' interest in violent video games is predicated on reducing high-impact violent actions.[17] Thus, while research on the correlation between video games and aggressive behavior is a useful starting point, policymakers need evidence on the presence or absence of a causal link between violent video games and violent crimes. We help to provide this evidence by finding and leveraging exogenous variation in exposure to violent video games, following a strand of economic literature that had sought to provide causal evidence on the behavioral effects of popular media. Dahl and DellaVigna (2009), one of the first studies in this area, estimates the short-term effects of violent movies on violent crime. By exploiting variation in the violence of movies shown in a given movie theater across different days, Dahl and DellaVigna (2009) finds that violent movies cause a contemporaneous decline in violent crime rates. In the first causal analysis of the effect of video games on crime in the United States, Cunningham, Engelstätter and Ward (2016) uses game quality and recency of game release as instruments that generate exogenous time series variation in the sale of video games in the United States. This research finds that violent video game exposure caused by new game releases actually leads to small declines in national violent crime rates, but the authors caution that their IV estimates should be viewed more as robustness checks. Indeed, one conclusion the paper draws is that an analysis using cross-sectional variation in game sales—the sort of variation we use—would be a fruitful area for further research.

The identification strategy that we use is also reminiscent of that used in Kearney and Levine (2015) and more recently in Lindo, Swenson and Waddell (2022), both of which examine television media. Kearney and Levine (2015) examines the effect that *16 and Pregnant* had on teen birth rates, while Lindo, Swenson and Waddell (2022) examines the impact that *The Ultimate Fighter*—a reality TV show about mixed martial arts fighting—had on violent crime rates. Both of these studies rely on the notion that television viewers exhibit habit persistence with regard to the TV stations they watch. When a new show

criticisms.

[17] See, for instance, the opening statements made during the original 1993 U.S. Senate hearing on the potential regulation of violent video games: https://www.govinfo.gov/content/pkg/CHRG-109shrg28337/html/CHRG-109shrg28337.htm.

premiers, there is variation in exposure to that show driven by how many people were already in the habit of watching a certain channel.[18]

While some of the same habit persistence likely does exist for the platforms that gamers use to access their games, our identification strategy does not rest on this behavioral assumption. Rather, we are exploiting the technical fact that gamers can access a new video game only if they have the correct hardware. Thus, when a new video game releases for one or more platforms, an individual's actual exposure to that game is determined partially by whether they already own the appropriate hardware. The video game sales that we observe are platform-specific (i.e., we can distinguish between an Xbox disk and a PlayStation disk), meaning that we can use lagged platform-specific game sales as a proxy for the stock of gaming platforms in a county. When new violent games then release for a certain set of platforms, counties are differentially exposed to those games based on their prior stock of hardware. Also, since we can measure the national sale of game releases in our data, we can leverage differences in the popularity of specific games—and differences in a game's popularity between platforms—to increase the precision of our estimates.

Our results do not show that increased exposure to new violent video games after their release is associated with any increase in violent crime outcomes, as measured by weekly crime rates reported by police agencies. In most cases, we estimate treatment effects that are precisely centered on zero. With regard to violent sexual offenses, we actually find that increased exposure to violent video games leads to statistically significant *decreases* in these crimes. More striking than the statistical significance of any individual result is the fact that our point estimates are consistently small in magnitude. Our estimates of the effect of additional exposure to violent game releases on assaults, for instance, are generally on the order of 0.1 percent of a standard deviation in assault rates.[19]

[18] Lindo, Swenson and Waddell (2022) actually goes further than this by leveraging the fact TV viewers turn to specific channels during specific time slots. We summarize both of these studies in more detail in Section 2.2.

[19] The standard deviations in crime rates that we refer to are calculated cross-sectionally across agencies for a given time period. We describe our standard deviation calculation in detail in Section 2.3.

Using an event-study specification, we then ask whether the effect of violent video game releases varies across weeks. Estimating the week-by-week effect of game exposure in this way helps us to assess whether the null results from our pooled model are obfuscating any important time-varying effects. For instance, other researchers have pointed out that *incapacitation*, the mechanical fact that video game play displaces time which could be used for alternative activities, could exist alongside other mechanisms. Effectively, this implies that any point estimate for the effect of video game exposure on crime could reflect an agglomeration of both incapacitation effects (negative) and psychological effects (uncertain). We find no evidence for any differences in treatment effects across weeks using our event study models, suggesting that the precise null results from our pooled estimator reflect the absence of any causal relationship between violent video games and violent crime.

Since we exploit changes in the ability to play new violent video games over the short run among populations that have already expressed an interest in video gaming, our analysis is most well-equipped to isolate effects along the intensive margin of violent game playing– we find no effect on violent crime of an increase in playing violent games by communities that are already playing violent games to some extent. Our results are largely silent with regard to the effect of changes in violent gaming along the extensive margin—the effect of communities (and individuals) playing violent games over long periods of time, compared to the counterfactual in which those communities are not playing violent games. This implicates our ability to detect psychological mechanisms in particular, as these could manifest slowly and over longer time horizons than we consider.

2.2 Background

2.2.1 Violent media, behavioral outcomes, and crime

In 1993, the Senate held the first congressional hearings on violent video games. Of particular concern to Senators was the marketing of violent games to children, the increasing realism of violence available through games, and the lack of cohesive regulation of violent game sales. The consequence of the 1993 Senate hearings on video game violence was the voluntary adoption and enforcement of a content rating system administered by the Entertainment Software Rating Board (ESRB). Akin to the Motion Picture Association of America (MPAA) responsible for the American movie rating system, the ESRB rates games on an escalating scale based on content. Games that have pervasive and realistic violence are rated *M: for mature*, which the ESRB defines as being "generally suitable for ages 17 and up. May contain intense violence, blood, and gore."[20] Video game retailers aere supposed to sell M-rated games only to those above the age of 17, and participating retailers are supposed to verify customer age with photo ID.

The 1993 Senate's concern over violent video games was predicated on the longstanding theory that engagement with violent media leads to an increase in violent behavior, including violent crime. Theories on the relationship between media and aggression can be traced to antiquity, but the general aggression model (GAM) (Anderson and Bushman, 2002a) is the predominant contemporary theory used to explain how something like violent gameplay would lead to aggressive behaviors such as violent crime. The GAM can be traced back to Bandura's social learning theory (Bandura, 1977), but also incorporates refinements such as script theory (Heusmann, 1988), cultivation theory (Gerbner et al., 1994), and others. In this framework, violent video games provide users with repeated opportunities to participate in simulated violent behavior that is rewarding and satisfying, leading gamers to develop internal scripts in which aggression is a viable course of action (Anderson and Dill, 2000), and also teaching players to expect hostile and aggressive actions from others (Anderson and Bushman, 2002b).

Empirically, the GAM has largely been tested through laboratory work. Laboratory experiments are generally conducted by randomizing participants to play either a violent or

[20] The two other common rating categories are *E: for everyone*, for games that are entirely suitable for all ages; and *T: for teen*, for games with some violence but minimal blood and no extreme violence. See www.esrb.com for a complete history and description of the content rating system.

non-violent video game for a period of time, and shortly thereafter measuring some outcome variable linked to aggression.[21] Proponents of the GAM point to these lab studies as cohesively showing that video games lead to relatively high levels of increased aggression. In a meta-analysis, Anderson (2003) finds that across multiple aggression measurements, playing violent video games leads to a 0.26α increase in aggression, which the author notes is a larger effect size than that found for secondhand cigarette smoke and lung cancer. However, Ferguson (2007b) and other critics have raised concerns about whether the outcomes used in these studies are well suited to measure aggression, as well as how to interpret the magnitude of these treatment effects. Overall, though, the claim that violent video games have some short-term impact on aggression in the laboratory is well supported.

In order to take the GAM's predictions about video games outside the laboratory, researchers have largely relied on correlation analyses in cross-sectional and panel surveys of adolescents (e.g., Möller and Krahé (2009), Anderson et al. (2008) and Gentile et al. (2004)). While these studies find that higher exposure to violent games is associated with increased aggression and aggressive behavior, critics have pointed out that these research designs fail to account for potential problems such as omitted variable bias, or selection into violent gameplay. Indeed, many such surveys do not report how the correlation between violent gameplay and aggression changes after controlling for covariates known to be associated with aggression and gameplay, such as gender. Using the CDC's Youth Risk Behavior Survey, Ward (2010) finds that controlling for demographic characteristics of youth in the sample leads the observed correlation between video game play and increased fighting propensity to shrink and become insignificant for groups who play up to four hours of video games daily. Even for those adolescents who report playing over 5 hours of video games *daily*—the group for whom the observed correlation between gameplay and fighting is strongest—Ward (2010) finds that increased fighting propensity relative to those who do not play games is reduced

[21] Commonly used outcomes meant to measure aggression include how a subject chooses to complete an open-ended scenario prompt (Anderson and Bushman, 2002b), the choice to act aggressively or punitively against opponents in simple games (Anderson and Benjamin, 2004), and even physiological markers linked to flight-or-fight responses such as cortisol levels and cardiovascular arousal (Gentile et al., 2017).

from 13.4 percent down to only 6 percent with the addition of these covariates. Ward argues that even if this remaining positive association is interpreted as causal, the magnitude of the effect is likely not large enough to merit policy intervention. Ferguson (2007a) also notes the descriptive fact that youth-involved crime rates have fallen steadily across the United States during the same period of time in which violent video games have become massively popular.

In contrast to the GAM, other theories of aggression and behavior predict that violent video games would reduce acts of aggression and violent crime. Some researchers have proposed that violent video games reduce violent behaviors by providing gamers with a simulated environment in which to sate aggressive urges, in what is commonly referred to as a theory of catharsis. Konečni and Doob (1972) provides a treatment of modern theories of catharsis that is useful when considering video games by distinguishing a mechanism of catharsis through displaced aggression that is applicable to video games. By providing an outlet through which aggression can be expressed virtually, violent video games relieve the gamer's urge to engage in violence against some real-life cause of aggressive feeling, such as a bully. Notably, many violent video game players believe in the cathartic effect of these games. In a survey of eighth-grade boys who play violent video games, Olson et al. (2008) finds that one common reason why adolescents choose to play violent games is to relieve feelings of aggression, or even to displace a specific urge to act violently in real life. Empirical research, however, finds limited evidence for catharsis effects. Kersten and Greitemeyer (2021) argues that video game players who report catharsis are conflating a general improvement in mood following violent gameplay with actual reductions in aggression.

Finally, violent video games could also serve to reduce crime through a mechanism of incapacitation, which draws on the basic notion of time use popularized by Becker (1965). Entertainment media is generally time-consuming, and violent video games are no exception. When an individual chooses to spend time consuming violent media, they are substituting time away from other activities. In the most straightforward story, some individuals drawn to violent behavior may be directly substituting time away from violent real-life activity in order to spend more time with the virtual violence of a game. More generally, video games would create an incapacitation effect whenever a game player chooses to spend time gaming that would otherwise be spent in an activity with a higher chance of leading to aggression. Indeed, even non-violent games could create incapacitation effects, so long as the time spent playing the game is not being substituted for time spent on an equally non-violent activity. Empirical work focusing on the incapacitating effect of video games is limited, though Ward (2018) finds that popular video games cause gamers to reduce their school attendance during the period when they are completing the game.

While these theories provide potential mechanisms through which violent video games could affect violent crime, well-identified causal analysis of this question is rare. Cunningham, Engelstätter and Ward (2016) is one of the first studies to focus on this causal relationship, using variation in violent game exposure created by new game releases to identify the impact of violent video games on national crime rates, finding evidence for either a null effect or actually a negative relationship between exposure to violent media and violent crime outcomes. Suziedelyte (2021), on the other hand, finds cross-sectional variation in violent video game exposure at the individual level through the timing of interview dates for families in the Panel Study of Income Dynamics (PSID). PSID interview dates are random throughout a given year, meaning that some families are interviewed soon after the release of major violent video games, while others are not. Suziedelyte (2021) finds that adolescents more recently exposed to violent video games before their PSID interview were less likely to engage in violent behavior, though their parents were more likely to report destructive behavior not against people (e.g., damaging school property).

2.2.2 Modern video game hardware and software

By the 21st century, the video game industry had become a major component in U.S. entertainment spending. In 2008, 53 percent of American adults reported playing video games. By 2020, that share had risen to 75 percent. As the video game market grew, so too did the number of new video games released each year. The online entertainment database IMDb.com contains a listing of 824 new games released during 2010, a 25 percent increase from the number of games released in 2000. Violent video games make up a sizeable share of these new releases. Between 2007 and 2011, roughly one quarter of the best-selling games each year were given the mature content rating, as measured by VGchartz.com, a video game research firm that publishes yearly lists of the top-100 video game releases. Based on VGchartz.com's list, a top-selling mature game was released every 3 weeks on average. As the industry grew and video game releases became more numerous, video game publishers also became increasingly sophisticated with regard to the marketing, product differentiation, and strategic timing of video game releases. Engelstätter and Ward (2018) finds that major video game publishers strategically choose the release date for their new games in order to avoid competition with other games of a similar genre or ESRB rating.

Modern video games need to be understood as a combination of software and hardware. Video games themselves are software developed by video game publishers. A video game can be played only when that software is run on a compatible video game platform, where a platform is defined as any piece of computing hardware and operating system that can play a video game. Given the need for game platforms, the video game market is used as a modern example of a two-sided market (Davidovici-Nora and Bourreau (2012), Rysman (2009)), in which platform manufacturers serve as the intermediary between consumers and software producers. In the early 2000s, common video game platforms were the personal computer, the video game console (e.g., Xbox 360, PlayStation 3), handheld gaming devices (e.g., the Gameboy), and even mobile phones.

As with any other software, video games must be coded to function on specific platforms, meaning that game publishers must choose in advance a set of platforms on which their games will work. When consumers choose to buy a video game, they must then choose a platform-specific version of that game; a game disk designed for one platform cannot be used to play on another, though gameplay is largely identical across platforms. For example, *Call of Duty 4: Modern Warfare* was one of the top game releases in 2007, and was released on PC and on most major game consoles, but was not originally released on any handheld device or on the Nintendo Wii. A consumer who bought *Call of Duty 4* for the Xbox 360 would be able to use the disk to play on any Xbox 360 console, but not on PS3 or a PC.

Most games are released on multiple different gaming platforms. Releases are synchronized in time across both platform and geography, and video game price is also synchronized across software versions. In fact, there is little variation in price even between games sold by competing publishers. Occasionally, a video game is released as an "exclusive" for one platform or family of platforms. For instance, new games in the *Halo* franchise, a series of violent science-fiction games, are published by Xbox Game Studios and released exclusively for the latest Xbox consoles. When *Halo 3* was released in 2007, anyone with an Xbox 360 could buy and play the game, while anyone without access to an Xbox 360 could not.

Video game consoles—computers designed specifically to run video game software in the home—were second only to personal computers as the platform of choice during the early 2000s. Video game consoles are expensive, long-term purchases that remain functional to play the latest video games for roughly a decade. For the past several decades, three major firms have dominated the video game console market: Sony, Microsoft, and Nintendo. In November 2005, Microsoft launched the Xbox 360, while in November 2006 Sony released the Playstation 3 (PS3) and Nintendo released the Wii. Together these three are referred to as the 7$^{\text{th}}$ generation of consoles, and these consoles remained dominant in the video game hardware market until the next generation of consoles was released starting in 2012. 7$^{\text{th}}$ generation consoles were wildly successful products, and led to technological advances including the rapid proliferation of online console gaming. In 2008, 53 percent of adult gamers and 89 percent of teenage gamers used a console at least some of the time. At release, the Xbox 360 sold for \$400, the PS3 sold for \$500, and the Wii sold for \$250.[22] Console

[22] These were the prices for the baseline or standard console version sold at launch. Both the PS3 and the Xbox 360 were available in several models at release, with different models largely distinguished by hard

manufacturers generally price their hardware at a significant loss that is then recouped via software licensing fees. As a consequence of this pricing strategy, console manufacturers compete to grow a large and loyal user base that will purchase enough games over the lifetime of the console to ensure profitability (Williams, 2002).

Those who play video games usually possess more than one platform with which to access video game software overall (Davidovici-Nora and Bourreau, 2012); however Derdenger (2014) finds little evidence that gamers purchase more than one console of the same generation. While some gamers are drawn to one console or another for idiosyncratically reasons (e.g., the desire to play a specific exclusive available on only one console), consoles are effective substitutes, since each provides access to hundreds of new video games each year. Moreover, while some exclusives are popular, many non-exclusive games are equally or more popular. And specifically with regard to violent games, many violent exclusive and non-exclusive games were available for both the Xbox 360 and the PlayStation 3.

2.3 Data

We measure video game sales at the county level using the Nielsen Retail Scanner Data (scanner data) spanning from 2007 through 2011. The scanner dataset contains weekly pricing and volume data for products sold in over 35,000 participating stores, comprising roughly 90 retail chains and capturing roughly 30 percent of mass merchandise sales and 50 percent of food and drug sales. In participating stores, Nielsen measures total weekly sales volume and price at the product level using the universal product code (UPC) system. The scanner data contains reported sales for over 2 million distinct UPCs, which are grouped into roughly 1,100 product categories. In the "video and computer games" product category, we have retail sales records for 22,328 UPCs. With regard to video games, it is useful to think of UPCs as each representing a distinct instantiation of a product. For example, *Call of Duty 4: Modern Warfare* was one of the most popular games released in 2007. Not only

drive size and the appearance of the console exterior.

was it initially released on several consoles, but it was also re-released several times due to its popularity. Nielsen's scanner data contains about a dozen distinct UPCs corresponding to a game disk for *Call of Duty 4*. Of those, one UPC corresponds to the original game disk used to play *Call of Duty 4* on the Xbox 360, while another corresponds to the equivalent game disk used for the PlayStation 3.

We use giantbomb.com, an online wiki and database for information about video game software and hardware, to identify the set of video games released from 2007 to 2011, as well as to capture product information such as release dates and content ratings. Using this information, we manually identify the set of UPCs associated with each video game from this time period that was released for either the Xbox 360, the PlayStation 3, or the Nintendo Wii. We successfully identify at least one UPC code for 84.5 percent of the 2,329 games that released on at least one of the three consoles noted above. Out of the roughly 500 games for which we found no sales records, the majority were either international games that saw no major U.S. release, or games that were released only via digital storefronts (e.g., the Xbox Live Arcade Game Pack, a collection of arcade games released directly on Microsoft's digital storefront). While each UPC corresponds to a console-specific version of a game, the product descriptions included in Nielsen's scanner data do not usually indicate this detail. To identify the console associated with each UPC, we rely on upcitemdb.com, an online UPC database. Out of the 22,000 video game UPCS for which Nielsen has sales records, we match over 16,000 of those UPCS to a specific console.

Not all of the games released during our sample period saw widespread sale in the United States. In order to identify the video game releases that would be popular and accessible across the U.S., we also scrape yearly video game popularity data from VGchartz.com. We use these top-selling game lists to identify mature video game releases that were major enough to be used as treatment events. In Table 2, we provide a complete list of the top-selling mature game releases between 2007 and 2011. Within a given year, there is a sizeable gradient in sales between the games on VGchartz's bestsellers list. In 2007, for instance,

Table 2: Top-selling mature game releases 2007-2011

Name	Yearly Sales Rank	Release Date	Exclusive Release?
Crackdown	29	2007-02-20	Yes
Resident Evil 4: Wii Edition	66	2007-04-07	Yes
BioShock	32	2007-08-21	Yes
Halo 3	2	2007-09-25	Yes
The Orange Box	78	2007-10-09	No
Call of Duty 4: Modern Warfare	5	2007-11-05	No
Assassin's Creed	11	2007-11-13	No
Mass Effect	30	2007-11-20	No
Devil May Cry 4	70	2008-01-31	No
Army of Two	51	2008-03-04	No
God of War: Chains of Olympus	48	2008-03-04	Yes
Tom Clancy's Rainbow Six: Vegas 2	37	2008-03-18	No
Grand Theft Auto IV	6	2008-04-29	No
Ninja Gaiden II	91	2008-06-03	Yes
Metal Gear Solid 4: Guns of the Patriots	23	2008-06-12	Yes
SOCOM: U.S. Navy SEALs - Confrontation	95	2008-10-14	No
Saints Row 2	75	2008-10-14	No
Fable II	28	2008-10-21	Yes
Far Cry 2	96	2008-10-21	No
Fallout 3	38	2008-10-28	No
Resistance 2	65	2008-11-04	Yes
Gears of War 2	8	2008-11-07	Yes
Call of Duty: World at War	7	2008-11-11	No
Left 4 Dead	40	2008-11-18	No
Killzone 2	31	2009-02-27	Yes
Resident Evil 5	18	2009-03-05	No
Prototype	54	2009-06-09	No
Halo 3: ODST	8	2009-09-22	Yes
Borderlands	29	2009-10-20	No
Dragon Age: Origins	32	2009-11-03	No
Call of Duty: Modern Warfare 2	2	2009-11-10	No
Call of Duty: Modern Warfare Reflex Edition	85	2009-11-10	No
Assassin's Creed II	15	2009-11-17	No
God of War Collection	44	2009-11-17	Yes
Left 4 Dead 2	16	2009-11-17	No
Darksiders	88	2010-01-05	No
Heavy Rain	65	2010-01-25	No
Mass Effect 2	29	2010-01-26	Yes
BioShock 2	42	2010-02-09	No
Dante's Inferno	100	2010-02-09	No
Aliens vs. Predator	93	2010-02-16	No
Battlefield: Bad Company 2	22	2010-03-02	No
God of War III	15	2010-03-16	Yes
Tom Clancy's Splinter Cell: Conviction	49	2010-04-13	No
Red Dead Redemption	13	2010-05-18	No
Halo: Reach	5	2010-09-14	Yes
Medal of Honor	39	2010-10-12	No
Fallout: New Vegas	33	2010-10-19	No
Fable III	20	2010-10-26	Yes
Call of Duty: Black Ops	2	2010-11-09	No
Assassin's Creed: Brotherhood	20	2010-11-16	No
Dead Space 2	61	2011-01-25	No
Bulletstorm	94	2011-02-22	No
Killzone 3	37	2011-02-22	Yes
Dragon Age II	77	2011-03-08	No
Homefront	58	2011-03-15	No
Crysis 2	80	2011-03-22	No
Mortal Kombat	40	2011-04-19	No
SOCOM 4: U.S. Navy SEALs	99	2011-04-19	No
L.A. Noire	32	2011-05-17	No
Deus Ex: Human Revolution	87	2011-08-23	No
Dead Island	56	2011-09-06	No
Gears of War 3	6	2011-09-20	Yes
Rage	68	2011-10-04	No
Battlefield 3	7	2011-10-25	No
Call of Duty: Modern Warfare 3	1	2011-11-08	No
The Elder Scrolls V: Skyrim	8	2011-11-11	No
Assassin's Creed: Revelations	19	2011-11-15	No
Saints Row: The Third	53	2011-11-15	No
Halo: Combat Evolved Anniversary	44	2011-11-15	Yes

Notes: Games included and their yearly sales rank are based on top-selling games lists published by VGchartz.com. Status as an exclusive release based on data scraped from gematsu.com.

VGchartz lists Wii Sports as the best-selling game, with 6 million game units sold in that year in the US. For comparison, VGchartz reports that the 100th best-selling game in 2007 sold only 250 thousand copies. In Table 3, we break out the proportion of top-selling games that received each ESRB content rating, as well as the fraction of those top-selling games for which we see sales records in the Nielsen panel. Of the games on VGchartz.com's lists of top releases, roughly one quarter are rated M. We find sales records for 80 percent of the games on the top-100 lists, and 97 percent of M-rated games.[23] A large majority of the top-selling games for which we do not see game sales are E-rated games released exclusively for handheld video game platforms (e.g., Pokemon Diamond/Pearl Version, released only on the Nintendo DS).

In Figure 10, we show the time series of weekly mature video game sales revenue across our panel, plotted against the release dates for top mature releases. Visualizing the data this way, we can see that sales of mature games at retail stores in our dataset spike following the release of top-selling mature games.

To get a sense of how much of the total U.S. video game market is present in Nielsen's sample of retailers, we calculate the total monthly and yearly video game sales revenue present in the scanner data. Using national sales figures published by the NPD Group as the denominator, we estimate that Nielsen's scanner dataset captures between one and two percent of new physical video game sales. The most important factor driving this low figure is that physical sale of new video games in this time frame was heavily concentrated among a handful of mass-market retailers. Nielsen censors the name and exact location of all physical stores, and the data-sharing agreement with Nielsen prohibits attempts to uncover retailer names, as well as any mention of specific retailers that are present or absent from the dataset. Speaking generally, though, it is logical to conclude that the handful of retailers specializing in video game sales are not present in Nielsen's scanner data, since the presence of any one

[23] VGchartz ranks games separately by console, so for instance *Call of Duty 4* is in the top-100 game list for 2007 twice, once for the PS3 and once for the Xbox 360. This means that the number of unique games in each top-100 list is less than 100. From 2007 to 2011, there are 285 distinct games in VGchartz rankings, of which we find sales records for 226.

Table 3: Top game releases by content rating between 2006 and 2011

ESRB Content rating	Content rating description	Share of top games with this content rating	Share present in sales data
E: for Everyone	*"Content is generally suitable for all ages. May contain minimal cartoon, fantasy or mild violence and/or infrequent use of mild language."*	54.6%	67.7%
T: for Teen	*"Content is generally suitable for ages 13 and up. May contain violence, suggestive themes, crude humor, minimal blood, simulated gambling and/or infrequent use of strong language."*	19.3%	89.1%
M: for Mature	*"Content is generally suitable for ages 17 and up. May contain intense violence, blood and gore, sexual content and/or strong language."*	26.1%	97.3%

Notes: Set of top games determined using yearly top-100 lists from VGchartz.com. Content rating descriptions are determined and published by the Electronic Software Rating Board (ESRB). Games with the E-10 ESRB rating are grouped with E games. See Section 2.2.1 for more details on the ESRB system.

Figure 10: Weekly revenue from mature game sales and mature game release dates (2007-2011)

We plot weekly revenue (in thousands of dollars) from the sale of "mature" rated video games against the release dates for mature games in VGchartz.com's yearly top-selling list. Each red line represents one such mature game release. We observe spikes in mature game sales following the release dates of best-selling mature games.

of these retailers would lead to a higher observed fraction of total video game sales. This supposition also explains the lack of console sales in the scanner data. The normal profit margin on video game platforms like the Xbox 360 is far lower than that of video game disks. Thus while many stores sell video game disks, fewer general merchandise chains find it profitable to stock video game hardware.

Using scanner data, we produce county-by-week sales figures for each video game UPC observed. For each video released for either the Xbox 360, the PS3, or the Wii, we then calculate weekly sales figures by console for the first eight weeks after each video game release. We also aggregate across different video games and calculate the weekly video game units sold for each console in each county. Since video games must be purchased for a specific platform, these console-by-county software sales allow us to proxy for the hardware being used in each county, a decision we discuss in more detail when we outline our empirical strategy. Finally, we calculate the total weekly revenue spent on all products in the video and computer games category for each county.

We use the National Incident-Based Reporting System (NIBRS), accessed via the Inter-university Consortium for Political and Social Research at the University of Michigan, as our primary source for crime data. Intended as the successor for the Uniform Crime Reports (UCR), NIBRS is administered by the FBI and provides information on criminal activity at the incident-level. For a criminal incident, NIBRS records information about the date and time of the incident, the offenses committed, the victims, and (where known) the offenders. Crimes are reported into NIBRS by the law enforcement agency with jurisdiction where the incident occurred, meaning that weekly crime patterns can be constructed at the agency or county level. NIBRS offense codes are quite granular, including 46 specific crimes against either person (e.g., homicide), property (e.g., motor vehicle theft), or society (e.g., illegal gambling). However, participation in NIBRS by law enforcement agencies is not universal. As of 2011, only 36 states had law enforcement agencies that reported into NIBRS, either indirectly through a state reporting program or directly to the FBI. Only 32 percent of those agencies that report into the UCR reported into NIBRS in 2011. Despite these limitations, NIBRS provides the only data source with which weekly crime patterns can be analyzed at a granular geographic level, and has been increasingly used by empirical researchers. We aggregate NIBRS offense-level data to create agency-by-week counts for the offense categories included in NIBRS. NIBRS also includes an agency-by-year population estimate for the area served by each agency, allowing us to express our crime outcomes as rates *per capita*. This agency population variable also lets us reliably exclude law enforcement agencies with an unclear geographic jurisdiction, such as state patrols and university police departments.

We link the county-by-week sales figures for video game releases from Nielsen scanner data with the agency-by-week crime counts from NIBRS to construct a panel dataset with observations at the week-by-agency level. The law enforcement agencies reporting into NIBRS are nested within counties, meaning that county-level sales figures from Nielsen can potentially be matched to multiple agencies. However, we follow Lindo et al. (2022) and do not aggregate our dataset to the county level for our primary analysis. Some law enforcement agencies report data into NIBRS only sporadically, or stop reporting into NIBRS in certain years, or cease reporting into NIBRS altogether during our sample period. Thus, conducting our analysis at the agency level ensures that changes we observe in crime outcomes are not driven by changes in the agencies reporting into NIBRS. However, given that the exogenous variation we exploit is at the county level, we cluster our standard errors at the county level for our primary analysis.[24] Given that we observe only one to two percent of yearly video game sales, there are also many small counties for which we observe negligible sales of any product in the video game category. Likewise, since police agencies report data into NIBRS by incident, a lack of reporting by a given agency for a period of time could reflect sporadic reporting or simply an absence of crime, particularly when an agency is in a sparsely populated county. To mitigate the issue of sparse reporting of either video game sales or crime, we restrict our primary analysis to agencies that report a positive population within their

[24] See Maltz and Targonski (2002) for a discussion on the challenges of using county-level crime data.

jurisdiction that are nested within a county with a population of at least 100,000 overall. Our final dataset includes roughly 1500 police agencies spread across 200 counties.

2.4 Empirical Strategy

Our identification strategy takes advantage of the fact that playing a video game requires a specific combination of hardware and software. Once consumers have purchased one video game platform, they are more exposed to new video games that are available for that platform. For instance, when a violent video game comes out for the Xbox 360, individuals who own an Xbox 360 are more likely to purchase and play the Xbox 360 version of that game. Due to the requirement that software must match hardware, the consumers who are more exposed to one of these game releases are those who already own one of the video game platforms on which the new release can be played.

Since our available data covers the period of time in which the 7$^{\text{th}}$ generation consoles were released and became dominant in the U.S. video game market, we choose to focus on the Xbox 360, the PS3, and the Wii[25]. While we focus on these game releases for at least one of these three consoles, many of these games were also released on some other platforms, and most commonly on personal computers (e.g., *Call of Duty 4* was also released for PC and Mac). To account for this in a feasible way, we combine all other platforms into a single "other" category when measuring video game purchases. We leverage variation in take-up of these four platforms across U.S. counties as a source of exogenous variation in subsequent exposure to violent video games, asking whether counties that should be more exposed to new violent games due to pre-existing hardware popularity experience a relative change in crime rates after those violent games release. This identification strategy is similar to that used by Kearney and Levine (2015), in which the authors instrument for exposure to the television program *16 and Pregnant* using MTV viewership across the U.S. from before the

[25] While we include mature game sales for the Wii in our analysis, we see far fewer sales of such games on the Wii. This is due in large part to the fact that fewer mature games were released for the Wii compared to other consoles, as Nintendo focused on marketing the Wii to families and "non-gamers."

show's premiere date, asking whether increased exposure to *16 and Pregnant* led to a change in teen pregnancy outcomes.

For this strategy to identify the causal effect of violent video game releases on crime, however, variation in console popularity can influence crime outcomes only by influencing the extent to which different counties are exposed to new video games. If selection into purchasing the Xbox 360, the PS3, or the Wii is in any way correlated with crime outcomes, then the exogeneity assumption of our instrumental variables strategy would be violated. This is akin to the issues raised in Jaeger et al. (2016) and Jaeger et al. (2020) regarding the identification strategy used to estimate the effects of *16 and Pregnant* in Kearney and Levine (2015)—that counties with higher MTV viewership before the show premiered were different than counties with lower viewership with respect to trends in teen pregnancy. Here, we can likewise imagine that counties with a large stock of Xbox 360 hardware prior to the release of a violent game for the Xbox 360 are systematically different with respect to crime outcomes than counties with fewer Xbox 360s. For this reason, we identify off of variation in historic console-specific sales *controlling for overall video game spending per capita* in each county.[26] The thought experiment we then envision is that two locations have an equal propensity to consume video games, but vary in terms of the hardware medium with which residents consume those games. When a game like *Halo 3* comes out for only the Xbox 360, counties with an equal demand for violent games in general will still vary in their actual demand for *Halo 3*, solely because of variation in the popularity of the Xbox 360. This is similar to the identification strategy used in Lindo et al. (2022) to estimate the effect on violent crime of the TV show *The Ultimate Fighter*, a popular mixed martial arts fighting show that premiered on Spike TV. Since Spike TV marketed itself as "the first network for men" and tried to target a young male audience, one could easily imagine that general viewership of Spike TV correlates with violent crime rates. To overcome this possibility, this research

[26] Given the relative rarity of most violent crimes, we express crime outcomes and local video game sales as rates per 10,000 residents in most of our analysis. For simplicity in our main text, we refer to all such measurements as *per capita* rates.

actually controls for overall Spike TV viewership and exploits only variation in Spike TV viewership during the specific timeslot when *The Ultimate Fighter* was broadcast. Rather than needing to assume that there is no selection into Spike TV overall, this refinement means that Lindo et al. (2022) needs only assume that there is no selection into specific timeslots of Spike TV programming between counties that view Spike TV at similar rates.

By using only the variation in exposure to violent games coming from platform variation, we likewise loosen the exogeneity assumption under which our estimated treatment effects are well identified. Specifically, we must assume that the choice to use one console versus another is uncorrelated with crime outcomes in counties that have a similar level of video game spending overall. Our identifying assumption would be violated, for instance, if counties in which the Xbox 360 became more popular had different crime patterns than counties that played video games to a similar extent, but on the PS3. Since we have data on crime rates both before and after each video game releases, we can relax this assumption further by comparing changes in crime rates before and after violent games release. The validity of our analysis then relies on the assumption that crime trends are parallel between counties with varying console stock.

Most of the video games that we study were released on several platforms, meaning that there are likewise several channels through which a county could be more or less exposed to a given violent video game. Consequently, we can also recast our identification strategy as employing a shift-share or Bartik instrument (Bartik, 1991). The canonical example of a Bartik instrument, as in Blanchard and Katz (1992), uses local employment shares by industry alongside national employment growth rates by industry. Interacting these two variables creates a prediction for local employment growth rates by industry that can instrument for actual employment growth rates. In our setting, we use local console popularity alongside national platform-specific sales of violent video game releases to similar effect. When a video game releases for a given set of platforms, a county's exposure to that game is determined by the share of the county using those platforms to play video games, and the national popularity of that game on that platform (i.e., the shift). Framed as a shift-share instrument, the identification assumption required by our research design is unchanged: the pre-existing hardware popularity in counties must be conditionally exogenous from trends in crime rates (Goldsmith-Pinkham et al., 2020).

While the intuition for our identification strategy relies on the relationship between hardware stock and software sales, as noted in Section 2.3 we do not observe a usable quantity of hardware sales. Rather, we are limited to observing a consistent fraction of video game software sales. But, game disks function only on one platform, meaning that consumers always purchase a hardware-specific version of a game. Since Nielsen's scanner data lets us distinguish between game disks sold for different platforms, we can use platform-specific software sale in a county prior to the release of a video game as a proxy for stock of that platform at the time of game release. In plain language, we assume that if we see a county largely purchasing games for the Xbox 360, for instance, this is reflecting the fact that gamers in that county are largely playing games on the Xbox 360. In doing so, we must assume the same conditional exogeneity of software sales for a platform as we did for hardware stock of that platform.

This assumption further highlights that we must define the period over which we measure console popularity. Since we use a dataset with multiple game releases in our first stage, one option is to construct a game-specific instrument using lagged sales from one or several months prior to the release of each game. However, this lagged-sales instrument is arguably more vulnerable to violations of the exclusion restriction, especially given our reliance on software sales to infer hardware stock. For example, between 2007 and 2011 three new games were released within the *Halo* franchise. The *Halo* games released for Xbox 360 were popular and violent, and one could imagine that over time a subset of the population also predisposed to violent crime began to select into the Xbox 360 console to take advantage of violent opportunities like *Halo 3*. That is, prior software sales over time could reflect selection into some consoles correlated with violent crime. Such selection would mean that using a lagged prior sales instrument admits the very type of endogeneity that would have biased an estimator using raw game sales.

Alternatively, one could consider using as a prior sales instrument only those sales that took place directly after the release of the new generation of consoles. As argued above, consumers who buy a new console are likely to continue buying games for that console long after the purchase. When the 7^{th} generation of consoles was first released, the hit video games that might induce selection into one console or another were in the future for the most part, and unknown to consumers. Thus, a fixed prior sales instrument is more likely to satisfy the exogeneity requirement. However, as the gap in time between prior video game sales and new violent game releases grows, the predictive power of a first stage using a fixed instrument would naturally wane. This highlights a crucial tension in this identification strategy: by increasing the distance between our prior sales period and our "treatment events" we feel more confident in our exogeneity assumption being satisfied, but must accept a weaker first stage.

Consequently, we run our primary analysis using two versions of our instrument. First, we use a fixed prior sales instrument, using all game sales from the year 2007. The Xbox 360 released in late 2005 and the PS3 and Wii in late 2006, making 2007 the first year in which the entire 7^{th} generation of consoles was available for purchase. Thus 2007 was the period of time during which many consumers were making the decision regarding which new console to purchase. Second, we use three months (12 weeks) of lagged sales, leaving out the one month (4 weeks) of sales directly prior to each game release. For example, since *Halo: Reach* released in September, 2010, the lagged sales instrument would be constructed using county sales data from May, June, and July of 2010. We adopt this "leave-one-out" strategy with our lagged sales instrument specifically because we are using software sales to proxy for hardware sales. Prior to the release of a major video game, individuals who already possess a certain console may change their purchasing behavior (e.g., by saving their money for a month instead of buying a game). Thus, by leaving out the month directly prior to game releases, we assuage concerns that any such anticipatory behavior influences our first-stage predictions. Since we use sales from 2007 as one of our instruments, we run our first stage using release sales for all video games released between 2008 and 2011, including 60 M-rated video games. In Section 2.5 we show evidence for the validity of our first-stage design, as well as the tradeoff we face in terms of first-stage power.

One low-hanging argument for the violation of our parallel trends assumption comes from price difference between new consoles. While each of the 7$^{\text{th}}$ generation consoles were all relatively expensive, their prices at launch were not identical. Such price differences could have led counties to select into different consoles based on socioeconomic factors, which could easily lead to the type of endogeneity we describe above.[27] Consequently, we also control for county socioeconomic factors (i.e., poverty rates, unemployment rates, and per capita personal income) in our analysis.

To implement the first stage of this instrumental variables strategy, we use the county-level sales data available through Nielsen to predict how well newly released violent video games will sell in different counties as a function of those counties' prior platform-specific software sales. Prior software sales and release sales of new games are measured in units (i.e., game disks) *per capita*. In our second stage, we consider the effect of violent video game releases over time periods of varying length, asking how violent game releases impact violent crime in the week following release, in the two weeks following release, and so on. Thus, for each of these time periods we estimate our first stage using sales of the newly released game over the same period. For example, when estimating the effect that a violent game release has on violent crime rates in the two months after that game releases, our first-stage model predicts the total sale of that video game in the two months following its release. We also highlight that two violent video games occasionally release during the same week. To account for this, our dependent variable measuring release sales also aggregates across all of the mature games that release during the same week. That is, we predict total sales in

[27] Indeed, the higher price of the PS3 at launch ($500) compared to the Xbox 360 ($400) has been described as one of the reasons that the Xbox 360 was much more successful at launch than the PS3.

county c over a number of weeks w of the violent video games that release during week t:

$$\begin{aligned} ReleaseSales_{c,w,t} = &\alpha_1 \sum_{g \in G_t} B_{c,g,w,t} + \alpha_2 VideoGameRevenue_{c,t} + \\ &\alpha_3 X_{c,t} + \delta_{w,t} + \mu_{c,t}, \end{aligned} \tag{2}$$

regressing total sales on video game revenue *per capita*, socioeconomic controls, period fixed effects $\delta_{w,t}$, and on our shift-share instrument for a given violent game $B_{c,g,w,t}$. Since multiple games can release for different consoles during the same week, we sum our shift-share instrument across G_t, the set of game releases in week t. For a given game g, our shift-share instrument takes the form

$$B_{c,g,t} = \sum_p PriorSales_{c,p,t} NationalSales_{p,g,w,t,-c}$$

where $NationalSales_{g,p,t,-c}$, our measurement of the national popularity of a game g on a platform p, is the national total sales figure for that game, on that platform, over the same number of weeks. We exclude a county's own game sales when we construct these national sales estimates, hence the $-c$ subscript.

in the second stage, we then use the variation in predicted total sales of new violent video games released in week t to identify the effect of violent video game releases on crime. We first fit two-period difference-in-differences models to estimate the impact of violent video game sales on violent crime rates. As noted above, one advantage of this two-period model is that we can vary the window of time around violent video game releases—asking what is the overall effect of an increase in exposure to violent video games in the week following a game release, the two weeks following that release, and so on. The proposed mechanisms through which violent video games impact crime would very likely operate at different points in time after those games release. Any incapacitation effect of violent game releases are liekly to be concentrated in the initial weeks after game release, when most individuals are playing the new game. The increased aggression such as that predicted by the GAM could occur at any point after individuals begin playing the game, but would be more difficult to detect when incapacitation through gameplay was still occurring. An estimated treatment effect close to 0 in the single week following violent game releases could thus reflect a true null relationship between violent gameplay and violent crime, or an offsetting combination of incapacitation and increased aggression—individuals playing the new game are more inclined to violent crime, but have less time to pursue it. Later, we turn to event-study designs in order to estimate week-by-week treatment effects. By first pooling the weeks spanning the release period using a parsimonious difference-in-differences design, however, we can chart any evolution in the aggregate effect of violent game releases on crime while retaining more precision. Our second-stage difference-in-differences equation takes the form

$$\Delta y_{ac,w,t} = \beta_t \Delta Total\widehat{ReleaseSales}_{c,w,t} + \delta_{w,t} + \varepsilon_{ac,w,t}, \tag{3}$$

where $\Delta y_{ac,t}$ is the change in the rate of crime *per capita* in a given agency nested within a county, in the w weeks after one or more mature games release in week t. By expressing the model in differences, agency fixed effects and all of our covariates are differenced out.[28] The $\hat{\beta}_t$ estimate obtained from this model would then be interpreted as the change in crime *per capita* in a police agency's jurisdiction associated with one more violent game sale *per capita* in the corresponding county.

Next, we use an event study to estimate the week-by-week effect of violent game releases on violent crime. In turning to an event-study specification, our goal is to consider whether the release of one or more violent video games has a changing effect on crime rates over time. Given the frequency of mature games releases, however, there are few opportunities from 2007 to 2011 to observe "long" periods over which there are no releases, then the release of a mature game, then a "long" period of time before another mature game is released. Rather, the release of most violent video games will be compounded in the following weeks

[28] Given that this pooled model has only a pre-period and post-period, a differences model is econometrically equivalent to a two-way fixed effects model. We make the choice to use a differences specification to avoid the computational burden of calculating fixed effects for thousands of police agencies.

by the release of another violent game. We can avoid the challenges associated with making inference when treatment is compounding in this way by focusing on the rare mature game releases that are isolated in time from other mature releases. For example, the longest period of weeks in the time series during which only one week contains a mature game release is plus/minus five weeks—this occurs twice, with the release of *Grand Theft Auto IV* (on the Xbox 360 and the PS3) and *Devil May Cry 4* (on the Xbox 360, the PS3, and Windows computers). But there is also a larger set of games that we can use to estimate the release-week effect of violent games more precisely—there are 36 mature games that were released with no other mature releases in the one week plus/minus. Thus, by re-estimating the effect of violent game releases on crime using stratified samples for which there are progressively longer periods of weeks over which only one mature game releases, we can retain more precision in our event study specification. In so doing, we identify off of more game releases when we consider smaller intervals of time (around release dates) than when we consider longer periods of time. Making sample restrictions of this kind, we estimate the following event-study specification:

$$\begin{aligned} y_{ac,t} = & \sum_{j=-J}^{J} \beta_j (WeeksFromRelease = j) Total\widehat{Sales}_{ac,4,t} + \\ & \sum_{j=-J}^{J} \theta_j (WeeksFromRelease = j) VideoGameRevenue_{c,t} + \\ & X_{ac,t} + \delta_t + \lambda_{ac} + \varepsilon_{ac,w,t} \end{aligned} \tag{4}$$

where J is the number of weeks over which there is no release of another violent video game and $j = -1$ (i.e., the week before a game releases) is used as the reference period. In our primary analysis, we use the estimated sale of new games during the month of their release as our measure of county exposure to new violent games. Notice also that the games with longer periods of isolation are subsets of those games with shorter periods of isolation. That is, since no mature games released in the plus/minus five weeks surrounding *Grand Theft Auto IV* and *Devil May Cry 4*, we can use these games when we estimate the release week effects of violent games. With all this in mind, we estimate this event-study specification for $J \in \{1, 2, 3, 4, 5\}$.

2.5 First-stage Relevance and Evidence for Validity

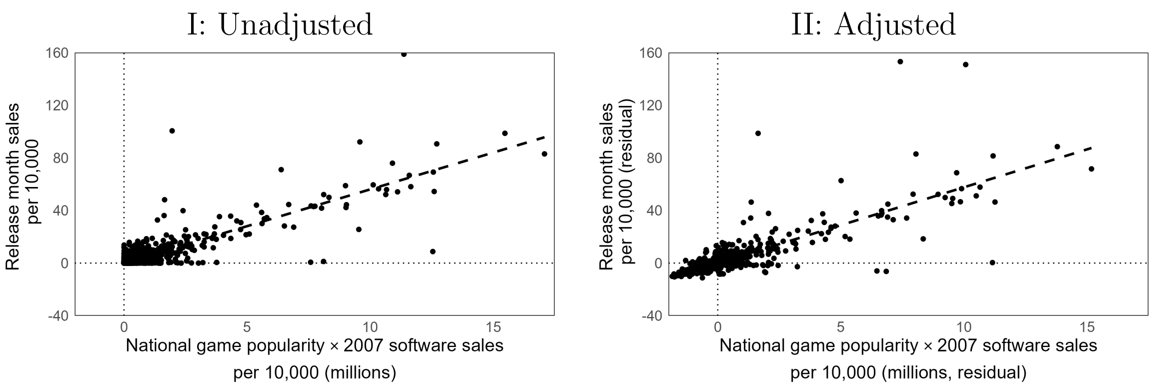
In Figure 11 we visualize the strength of our first stage in predicting the sale of just-released violent video games. As described in Section 2.4, we run this first stage separately to predict the release sales of violent games over periods of increasing length, ranging from one week to two months. Here we present the results from the models predicting release month sales, but our first-stage results are similar across all of the release window lengths that we consider. In Figure 11 we plot our instrument on the x-axis and the actual figures for the sale of newly-released violent video games during their release month on the y-axis. Recall that our instrument is the interaction of national game popularity (measured in total units) and county-level platform exposure (measured in units of platform-specific software *per capita*). In Figure 11 Panel A (I), we plot the relationship between our instrument and our release window sales. As we discuss above, we seek to exploit only the variation in exposure to a new violent video game arising from differences in platform popularity between counties that have similar demand for video games overall. In Panel A (II), we plot the same relationship after adjusting for the controls used in our first-stage specification. In both of these panels, there is a strong, visible relationship between our instrumental variable and our first-stage outcome.

We also present the results of our first-stage regression in Table 4. In Panel A, we present the first-stage results obtained using 2007 software sales as our instrumental variable. As shown in Panel A, the first-stage relationship between our shift-share instrument and sales of just-released violent games is statistically significant and positive, even with the addition of both our county-level video game spending control and socioeconomic controls. The

Figure 11: Graphical depiction of the first stage to predict new violent game sales

We instrument for the release month sale of new violent video games using 2007 platform-specific software sales interacted with national game popularity, as measured in units sold. Our instrument is highly predictive of release month sales, even after adjusting for county-level video game spending and socioeconomic characteristics. In Panel B, we consider how the performance of our instrument changes if we use platform-specific software sales alone. This exercise reveals that platform-specific software sales alone are a predictive instrument for release month sales, but highlights that the interaction with national game popularity is useful for generating accurate predictions. See Section 2.5 for details.

Panel A: Local platform popularity interacted with national game popularity



Panel B: Local platform popularity alone

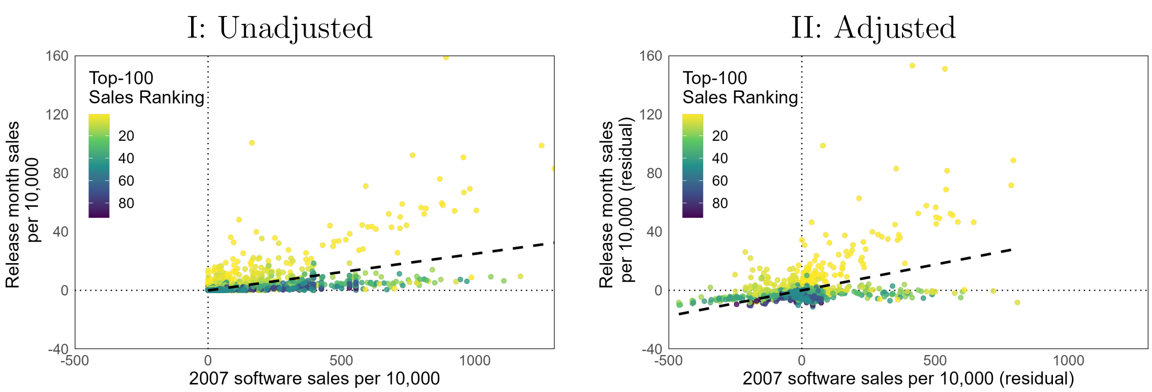


Table 4: First-stage estimates for video game sales during release month

	(1)	(2)	(3)
Panel A: 2007 instrument			
National game popularity · 2007 software sales	0.0000057***	0.0000056***	0.0000056***
	(0.0000004)	(0.0000004)	(0.0000004)
Per-10,000 game spending		0.0000116	0.0000130
		(0.0000083)	(0.0000082)
SES Controls	No	No	Yes
F-stat	22188	16290	16314
N	9693	9693	9693
Panel B: Lagged instrument			
National game popularity· lagged software sales	0.0000209***	0.0000216***	0.0000216***
	(0.0000009)	(0.0000011)	(0.0000011)
Per-10,000 game spending		-0.0001082***	-0.0001078***
		(0.0000242)	(0.0000243)
SES Controls	No	No	Yes
F-stat	29143	27049	27004
N	9838	9838	9838

Notes: Observations here are at the county level to be consistent with our source of identifying variation (video game sales across counties). Socioeconomic (SES) controls are per capita personal income, the poverty rate, and the unemployment rate. Standard-error estimates allow for clusters at the county level. The reported F-statistic is for the exclusion of the shift-share instrument interacting 2007 software sales with national game popularity. *, **, and ***, indicate statistical significance at the ten-, five-, and one-percent levels, respectively.

corresponding F-statistics are well beyond traditional thresholds used to asses first-stage strength. In Panel B of Table 4, we likewise present the first-stage results using the three-month leave-one-out version of our instrument described in Section 2.4. Using this alternate version of our instrument produces similar results, though the magnitude of the relationship between our instrument and new violent game sales is larger. This result makes sense, given that this alternate first stage is using sales data that occurs closer in time to the violent game releases we study.

The F-statistics in Table 4 demonstrate that our shift-share instrument is highly predictive of release-month sales for new violent video games. Beyond first-stage relevance, however, we are also interested in evidence for the validity of our research design, which relies on the argument that exposure to violent video games on a specific console should be modulated by the extent to which county residents have previously purchased that console. In order to get a better sense of how this variation in platform popularity contributes to the strength of our instrument, we also run an alternative specification of our first stage, where we replace the estimates of national (platform-specific) game popularity in Equation 2 with platform indicator functions, which turn on if a game is released for a given platform. Our instrument $B_{c,g,t}$ then takes the form

$$B_{c,g,t} = \sum_{p} PriorSales_{c,p,t} * \mathbb{1}_p\{NationalSales_{p,g,t} > 0\}$$

This specification of our first stage allows us to ask whether the story we seek to tell about our instrument is believable: that individuals are more exposed to new video games if they already possess the hardware to play those games.

Figure 11 Panel B and Table 4 Panel B show the results from this alternate first-stage specification. The most striking aspect of Panel B is that, while our first-stage relationship remains strong overall, there are two distinct patterns between 2007 software sales and new violent game sales. Specifically, there is a strong correlation between 2007 sales and new violent sales for one subset of the data (i.e., above the dotted best fit line), while there is almost no such correlation for another subset of the data. We argue that this clearly illustrates the useful role that national game popularity plays when used in our first stage. As noted in Section 2.3, the release-period sales for the best-selling game in a year are orders of magnitude higher than the corresponding sales for the hundredth best-selling game in a year. But without the estimates of national popularity, the model shown in Panel B is forced to estimate the relationship between prior hardware popularity and new violent game sales as if all of those games are equally important. The consequence of this is that the model underestimates the relationship between hardware popularity and new sale of the most popular games, and overestimates that relationship for the least popular games. Said differently, the choice to interact platform-specific game sales in 2007 with the national popularity of new games allows our first-stage model to weight how important each game release is, yielding a better estimate of the relationship between pre-existing hardware popularity and new software sales.

Panel B also demonstrates that, while including national game popularity improves the strength of our first stage, the platform-specific software sales that we employ are independently useful to predict new game sales. We interpret this as evidence that our instrumental variable is working due to the channel that we hypothesize: counties in which a given platform is already popular are more exposed to new game releases that happen for that console. Moreover, in Panel B (II), we again show that controlling for overall game spending in a county (along with socioeconomic variabes) does not eliminate the predictive power of our instrument.

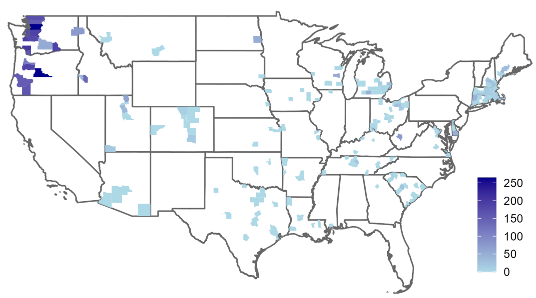
We have also argued that controlling for overall interest in video games is important to our exclusion restriction. Counties that have more interest in video games generally likely have more interest in new video game consoles, and a general interest in video games could be correlated with the unobserved determinants of crime. We want to only take advantage of variation in console popularity between counties that have similar interest in video games overall. In Figure 12, we show some evidence for this argument. In Panel A, we show that

Figure 12: Variation in exposure to game consoles in 2007

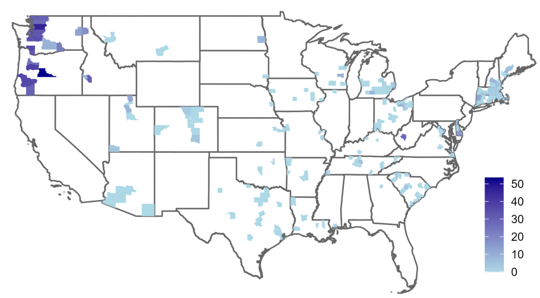
There is systematic variation in exposure to new video games on the Xbox 360 and the PS3 as measured in Nielsen scanner data. As seen in Panel A, both new consoles have heavily concentrated sales in the Pacific Northwest. After controlling for the overall level of video game spending as well as county socioeconomic characteristics, there is no systematic pattern in exposure to game releases for either console. See Section 2.5 for details.

Panel A: Unadjusted console game sales per 10,000 residents

I: Xbox 360

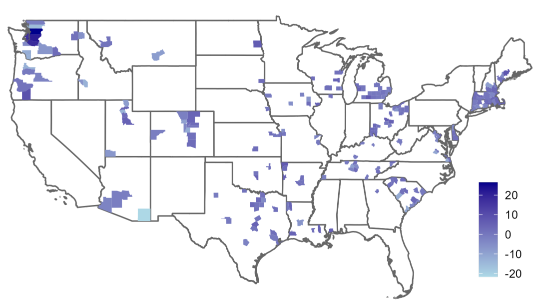


II: Playstation 3

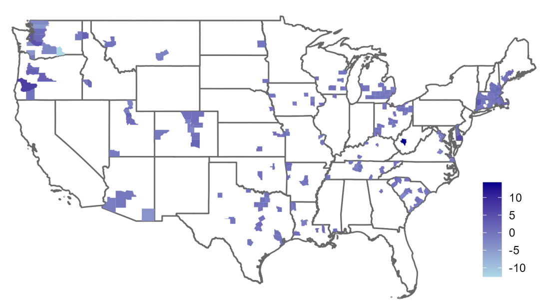


Panel B: Adjusted console game sales per 10,000 residents

I: Xbox 360



II: Playstation 3



the software sales *per capita* we observe in our dataset for the Xbox 360 and the Playstation 3 are indeed systematic across the United States. Sales *per capita* for both of these consoles are heavily concentrated in the Pacific Northwest. In Panel B, we then show the residual variation in console-specific game sales that is left after adjusting for video game spending and socioeconomic controls. Panel B demonstrates that controlling for *per capita* spending on video games in 2007 eliminates this systematic variation in console popularity. We argue that this provides suggestive evidence for our exclusion restriction—after controlling for general interest in video games, the variation in exposure to specific consoles that we exploit is not systematic across the United States.

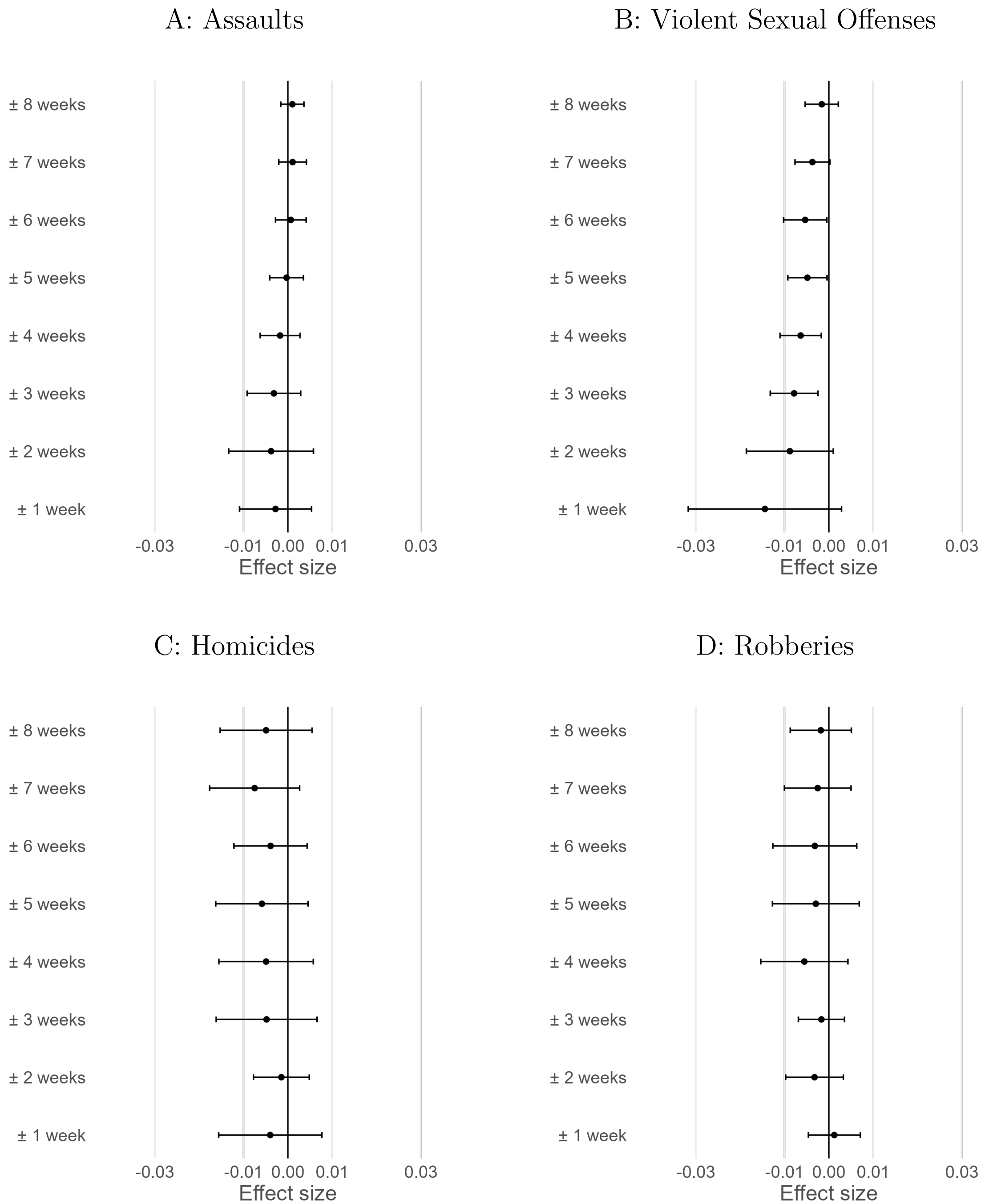
2.6 Results

Having shown evidence for the relevance of our first stage, and also some suggestive evidence for the validity of our exclusion restriction, we now turn to the results from our difference-in-differences and event-study specifications, which we consider to be our primary results. We focus on the four crime categories included in NIBRS that involve the use of force or violence: homicides, assaults, violent sexual offenses, and robberies.

In Figure 13 we plot the estimated effect of increased violent game sales on violent crime from the two-period difference-in-differences version of our identification strategy. For each crime outcome, we begin by estimating the aggregate effect of increased violent game sales on crime during the two months following a game release week, and tighten the number of weeks included in the post-period (and pre-period) window in one-week intervals. Plotting our results across tightening post-period windows in this way allows us to start thinking about the relative strength of the mechanisms through which violent video games could affect violence. To make our results more interpretable, we present all of our primary results in terms of effect size or impact. Given that our identification takes advantage of variation in game exposure across space, we use the average standard deviation in 2007 crime rates

Figure 13: Effect size estimates of the effect of violent game releases on violent crime, using 2007 platform shares

We estimate two-period difference-in-differences models to evaluate the effect of violent video game sales on violent crime rates. The proposed mechanisms through which violent games affect violent crime could vary in strength based on how much time has passed since game release. Thus we vary the length of the pre- and post-period in the model, estimating the effect for periods of ± 8 weeks down to ± 1 week. Here, we have normalized the point estimates obtained from these models by the standard deviation in crime rates across agencies over corresponding period of time in 2007. See Section 2.6 for details.



across agencies to calculate effect sizes.[29]

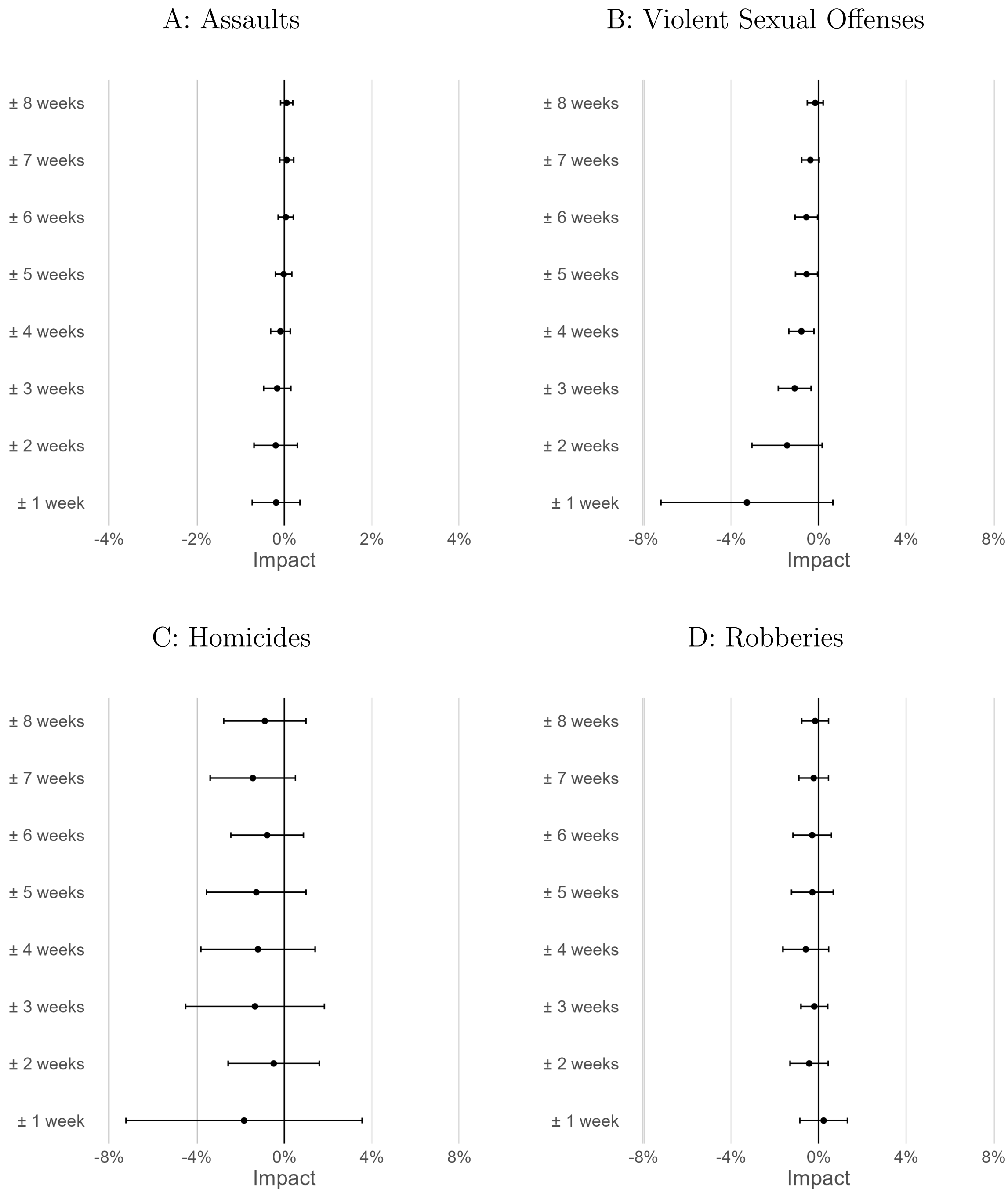
Our difference-in-differences results provide several main takeaways. First, across violent crime categories and post-period windows we find no evidence of any statistically significant positive relationship between increased violent game sales and violent crime. In fact, when we consider a post-period window of between three and six weeks, increased violent video game sales lead to a statistically significant decrease in the rate of violent sexual offenses committed (Panel B). The size of this decrease is roughly 0.5 percent of a standard deviation in violent sex offenses, on average. Though not statistically significant, the estimated effect of violent game sales on homicides (Panel C) and robberies (Panel D) are also negative. The estimated effect of violent game sales on assaults (Panel A) is negative when we consider any post-period between one and five weeks, and positive thereafter, though all of the point estimates for the relationship between mature game exposure and assault are statistically insignificant.

More important than the statistical significance of any individual point estimates, our difference-in-differences estimation allows us to rule out any systematic patterns or empirical regularities running from violent video game release to violent crime. The largest positive treatment effect we estimate is for violent game sales and assaults in the seven weeks after game releases, where we find that one additional violent game sale *per capita* residents increases assaults by .1 percent of a standard deviation over the two months following a game release. Figure 13 also shows that the corresponding 95 percent confidence intervals are tight enough to rule out positive effects larger than 0.6 percent of a standard deviation. Indeed, we rule out positive effect sizes greater than 1 percent across all post-period windows and violent crime categories. In Figure 14, we express the same results as impacts, again showing that violent video game releases are not associated with any large increase in any type of violent crime. Taken together, these results suggest a small incapacitation effect in

[29] Specifically, for the difference-in-differences estimator with a post-period of n weeks, we first calculate the standard deviation in crime rates across agencies for each n-week period in 2007. We then average across the standard deviations for each period. In the model with a post-period of just one week, for example, this amounts to calculating the average weekly standard deviation for a given crime type across agencies.

Figure 14: Impact estimates of violent game releases on violent crime, using 2007 platform shares

We estimate two-period difference-in-differences models to evaluate the impact of violent video game sales on violent crime rates. The proposed mechanisms through which violent games affect violent crime could vary in strength based on how much time has passed since game release. Thus we vary the length of the pre- and post-period in the model, estimating the effect for periods of ± 8 weeks down to ± 1 week. To present our results in terms of impact, we have normalized the point estimates obtained from these models by the average crime rate across agencies over the corresponding period of time in 2007. See Section 2.6 for details.



the weeks immediately following violent game releases.

We also estimate the same difference-in-differences model using lagged platform-specific software sales, as opposed to the equivalent sales from 2007. We show the results from this alternate specification in terms of effect size in Figure 15 and in terms of impact in Figure 16. Our results are largely unchanged using this alternate instrument. Again, we estimate treatment effects that are largely negative, and only statistically significant for violent sexual offenses.

Next, we consider the results from our event studies. In moving to an event-study specification, we are no longer estimating the aggregate effect of video games on violent crime over some post-period window, but are instead estimating week-by-week treatment effects. While this gives us more of an ability to tease out any dynamics or patterns in the impact of violent games over time, this ability comes at the cost of precision. For reasons we describe in Section 2.4, we estimate our event-study specification using only some mature game releases, specifically those that are isolated in time from other mature releases. This means that we lose precision not only because of the increased numbers of parameters that we estimate, but also due to the smaller available sample. In spite of this, the value of our event-study specification is that we can use it to check for any patterns or systematic trends in the week-by-week effect of mature game releases. In our difference-in-differences specification, we can estimate only the aggregate effect of a game release over some post-period window. A point estimate from this specification centered on zero—as in Panel A of Figure 13 showing the effect on assaults—could thus reflect either a true null relationship between violent video games and violent crime, or some offsetting combination of incapacitation and increased aggression. While they do not amount to a formal test of this possibility, the week-by-week effects that we can estimate in an event-study allow us to examine the evolution of crime rates after a game release for patterns that are suggestive of some mechanisms. If the point estimates for the release-week effect of mature games on assaults were negative but then trended towards zero in subsequent weeks, for instance, this would be consistent with

Figure 15: Effect size estimates of the effect of violent game releases on violent crime, using 4-month lagged leave-one-out platform shares

We again estimate two-period difference-in-differences models to evaluate the effect of violent video game sales on violent crime rates, this time instrumenting for release sales using lagged platform-specific sales from the 16 weeks directly prior to each game release. Given that these lagged sales could be influenced by anticipatory behaviors, we leave out the 4 weeks directly prior to each game release. Here, we have normalized the point estimates obtained from these models by the standard deviation in crime rates across agencies over corresponding period of time in 2007. See Section 2.6 for details.

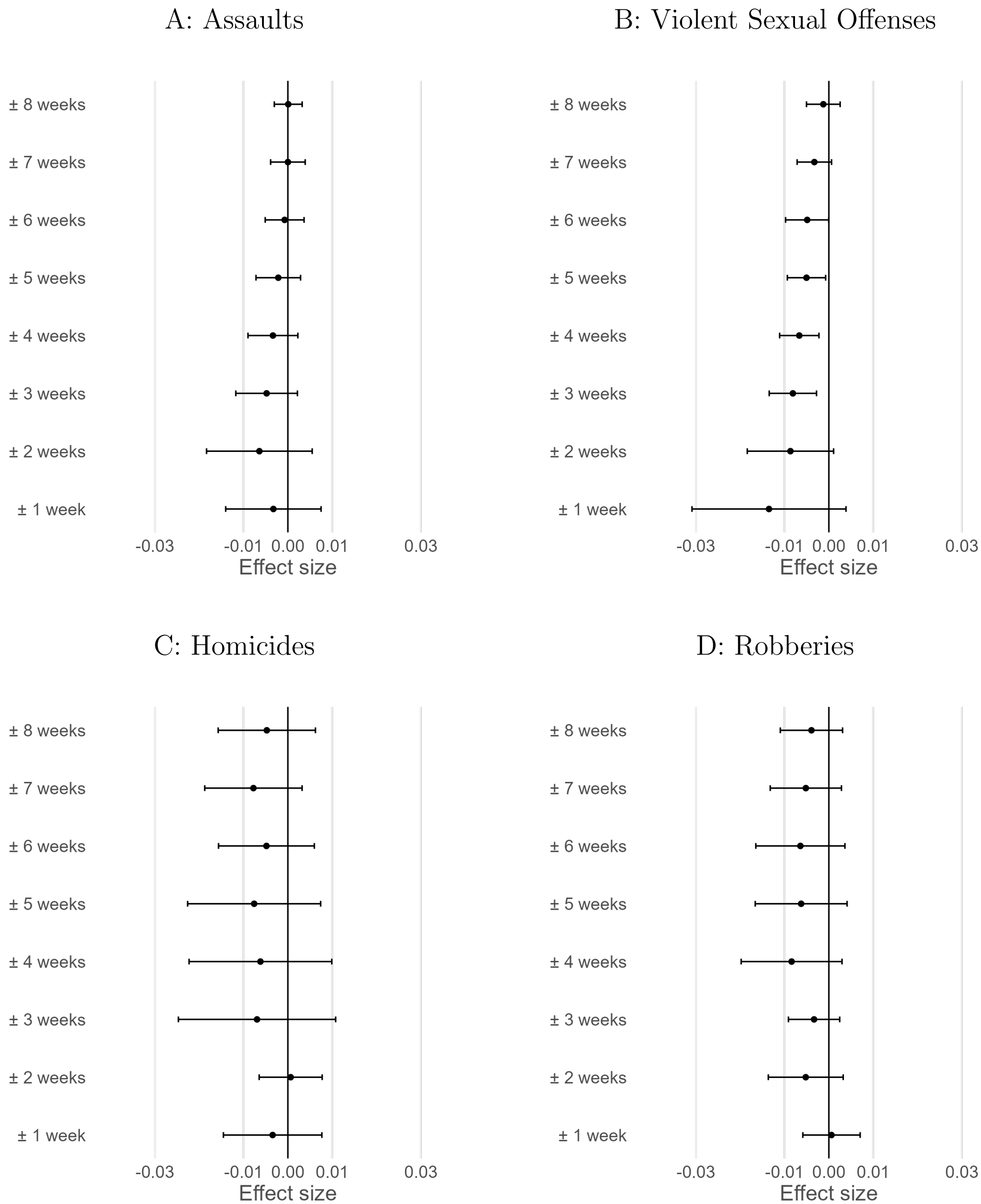
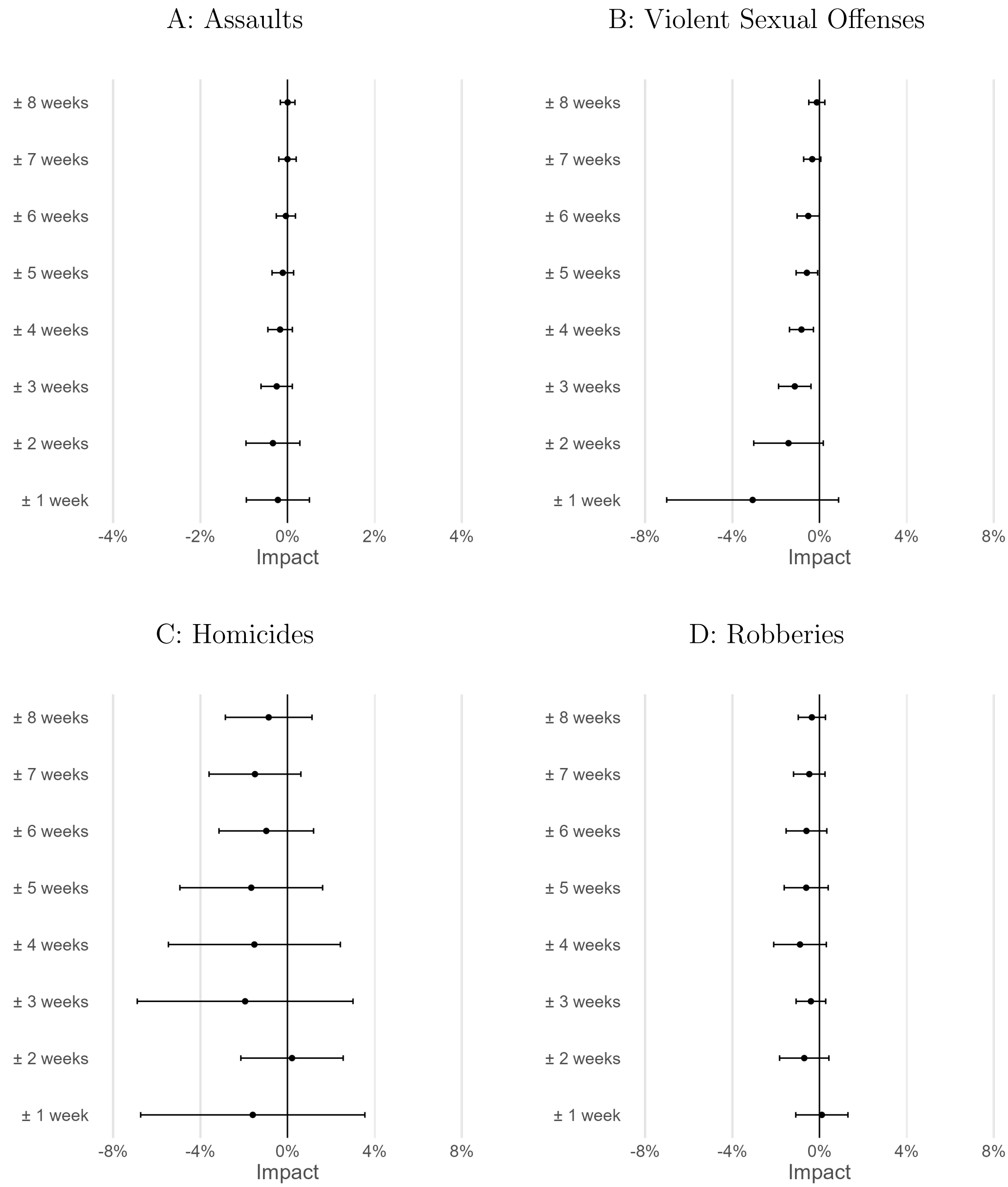


Figure 16: Impact estimates of violent game releases on violent crime, using 4-month lagged leave-one-out platform

We again estimate two-period difference-in-differences models to evaluate the effect of violent video game sales on violent crime rates, this time instrumenting for release sales using lagged platform-specific sales from the 16 weeks directly prior to each game release. Given that these lagged sales could be influenced by anticipatory behaviors, we leave out the 4 weeks directly prior to each game release. To present our results in terms of impact, we have normalized the point estimates obtained from these models by the average crime rate across agencies over the corresponding period of time in 2007. See Section 2.6 for details.



a story of incapacitation.

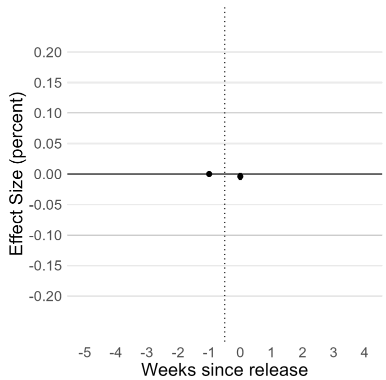
However, the results from our event-study specifications do not reveal any such patterns. In Figure 17 we show the results from this event-study analysis of violent game releases on assault rates. We again present our results in terms of effect sizes. In Panel A, we show the results for the set of games that are isolated in time from other mature releases by at least one week plus/minus. In Panel B, we likewise show the results from estimating our event-study model for the set of games that are isolated for plus/minus two weeks, and so on for Panels C, D, and E. While we see more noise and wider confidence intervals around these point estimates, no visible patterns or trends emerge. Rather, these panels again show point estimates that hew close to 0 and are statistically insignificant.

In Figure 18, Figure 19, and Figure 20, we visualize our event-study results for our other violent crime categories, sex offenses, homicides, and murders, respectively. Across these violent crime categories, we again find no evidence of a positive relationship between violent video games and violent crime. Moreover, while the confidence intervals around our estimates are larger than those from the difference-in-differences estimator, the point estimates themselves remain small for the most part. Further, while some of our point estimates do
move away from zero, this is largely happening as we enforce a longer period of isolation from other mature game releases in order to estimate treatment effects further out from release weeks. In Figure 19 Panel E, for example, the estimated effect on homicide in week 1 after the release week of a violent game is almost 0.1α. While this corresponds to a large effect, note that the confidence interval for this point estimate contains the confidence interval for the equivalent week 1 treatment effect in Panel B—recall also that the set of games informing the point estimates in Panel E is a subset of that in Panel B. This suggests that the large point estimate for week 1 in Panel E is simply a consequence of the heavy restrictions we have made on the dataset. When we use all of the available data to estimate the week 1 treatment effect, we see that the more precise estimate turns out to be close to zero.

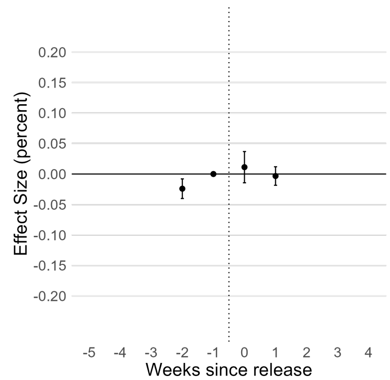
Figure 17: Event-study estimates of the effect of violent game releases on assault, using 2007 platform shares

We use event study specifications to estimate the week-by-week effect of exposure to mature game releases. Most mature video games release in close proximity to other mature video games, complicating our ability to interpret point estimates. To avoid this challenge while retaining precision, we estimate our event study models using stratified samples for which there are different periods of time over which only one mature game releases. For instance, Panel A shows the results from the model using the 36 mature games that were released such that the next-closest mature game release was at least one week away. See Section 2.6 for details.

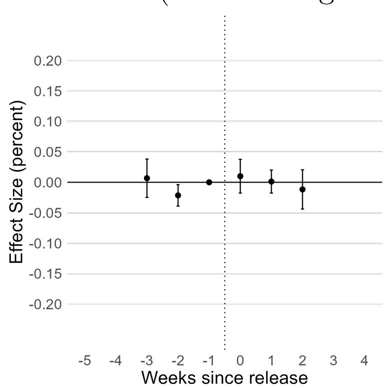
A: ±1 week (36 mature games)



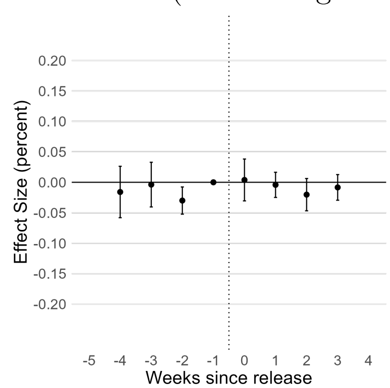
B: ±2 weeks (18 mature games)



C: ±3 weeks (11 mature games)



D: ±4 weeks (6 mature games)



E: ±5 weeks (2 mature games)

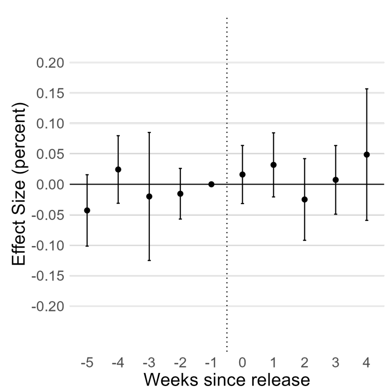


Figure 18: Event-study estimates of the effect of violent game releases on violent sex offenses, using 2007 platform shares

We use event study specifications to estimate the week-by-week effect of exposure to mature game releases. Most mature video games release in close proximity to other mature video games, complicating our ability to interpret point estimates. To avoid this challenge while retaining precision, we estimate our event study models using stratified samples for which there are different periods of time over which only one mature game releases. For instance, Panel A shows the results from the model using the 36 mature games that were released such that the next-closest mature game release was at least one week away. See Section 2.6 for details.

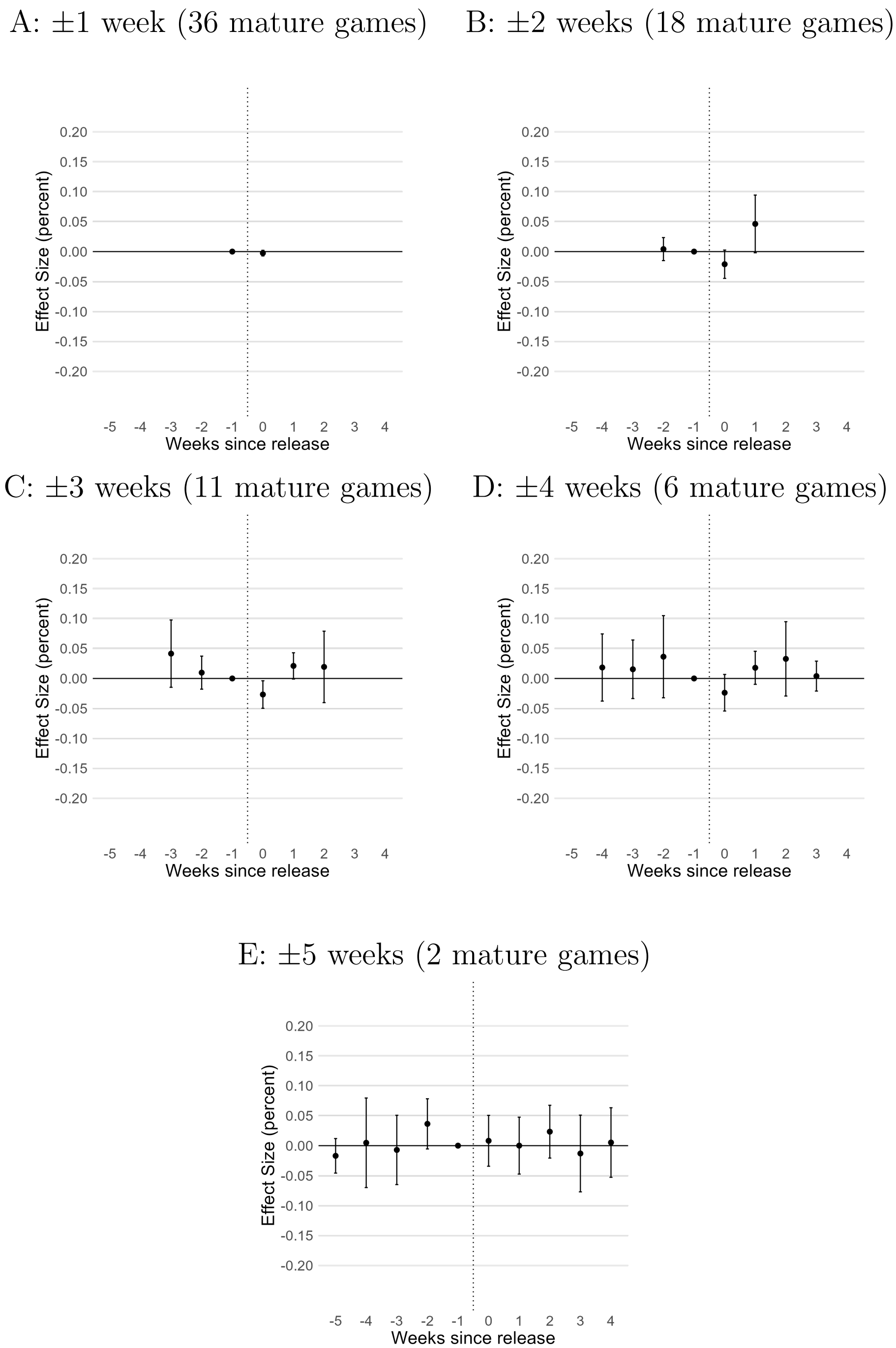


Figure 19: Event-study estimates of the effect of violent game releases on homicide, using 2007 platform shares

We use event study specifications to estimate the week-by-week effect of exposure to mature game releases. Most mature video games release in close proximity to other mature video games, complicating our ability to interpret point estimates. To avoid this challenge while retaining precision, we estimate our event study models using stratified samples for which there are different periods of time over which only one mature game releases. For instance, Panel A shows the results from the model using the 36 mature games that were released such that the next-closest mature game release was at least one week away. See Section 2.6 for details.

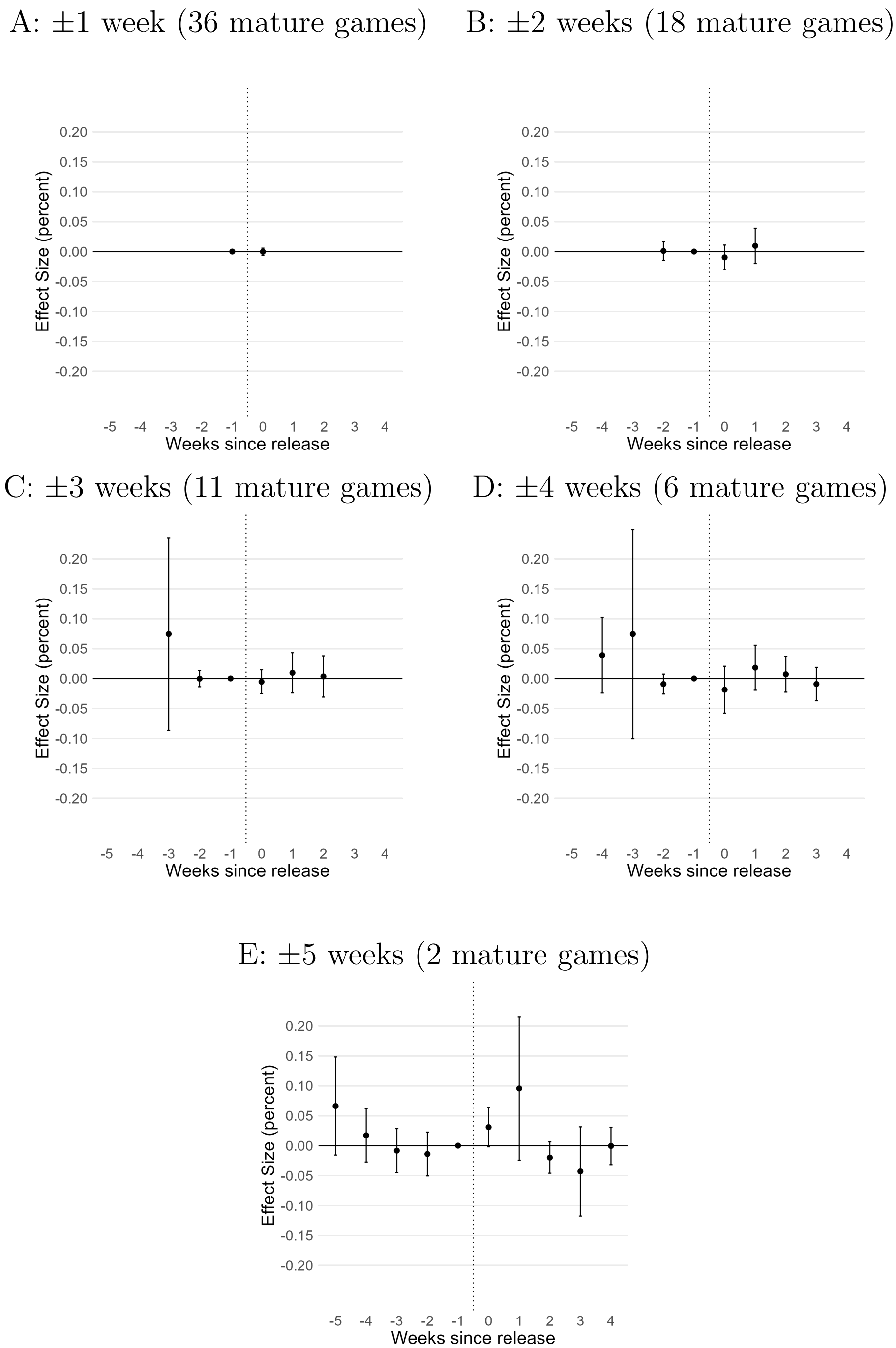
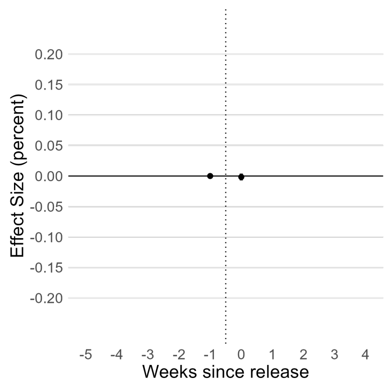


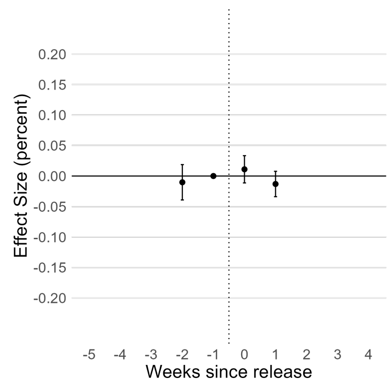
Figure 20: Event-study estimates of the effect of violent game releases on robberies, using 2007 platform shares

We use event study specifications to estimate the week-by-week effect of exposure to mature game releases. Most mature video games release in close proximity to other mature video games, complicating our ability to interpret point estimates. To avoid this challenge while retaining precision, we estimate our event study models using stratified samples for which there are different periods of time over which only one mature game releases. For instance, Panel A shows the results from the model using the 36 mature games that were released such that the next-closest mature game release was at least one week away. See Section 2.6 for details.

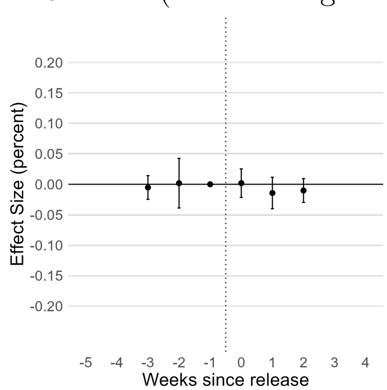
A: ±1 week (36 mature games)



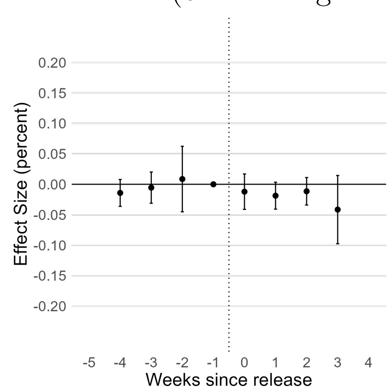
B: ±2weeks (18 mature games)



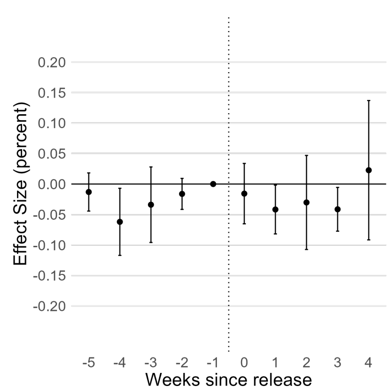
C: ±3 weeks (11 mature games)



D: ±4 weeks (6 mature games)



E: ±5 weeks (2 mature games)



Overall, the results from our event-study analysis do not reveal any systemic patterns in the impact of violent game releases in the five weeks following the release of such games. Combined with the results from our pooled difference-in-differences estimator, we find no evidence to suggest that violent video games lead to increases in any type of violent crime in the weeks after those games release.

2.7 Conclusion

We find no evidence that violent video game releases lead to increases in any type of violent crime. Our estimation strategy takes advantage of the fact that the video game hardware that gamers own influences their ability to access new violent video games. Effectively, we compare places that have similar interest in video games, but that vary in their ability to access new video games due to variation in the popularity of different game platforms in those places. With access to cross-sectional variation in video game sales, we are able to leverage this intuition to identify the short-run effects of violent video game releases on violent crime.

For all types of violent crime, we fail to recover any significant positive impact of mature game releases on violent crime. Indeed, our only statistically significant results suggest that mild *decreases* in violent sexual offenses occur after the release of violent video games. We also use an event-study specification in an attempt to tease out whether these null results could be the result of offsetting positive and negative causal channels between mature game releases and violent crime, but again fail to find any evidence for a relationship between violent crime and exposure to mature game releases.

We believe that our results are most useful in conjuction with other studies that have examined the impact of violent media exposure on violent crime using well-identified empirical strategies, including studies on the effect of violent movies (Dahl and DellaVigna, 2009), violent television (Lindo, Swenson and Waddell, 2022), and violent video games (Cunningham, Engelstätter and Ward, 2016). Taken together, these studies and our own provide a solid body of evidence that exposure to violent media does not lead to an increase in violent behavior, and in fact could lead to short-lived decreases in crime. More generally, this growing literature should invite skepticism of the common claim that video games contribute to violent crimes. Short-run effects of violent gameplay on aggression have been documented in a plethora of laboratory studies, but the evidence we present here calls into question how those results translate to outcomes outside of the laboratory. As video games continue to grow in popularity, calls for the regulation of violent games are unlikely to subside. We suggest that policymakers exercise caution before devoting resources to interventions rooted in a fear that violent video game play will lead to real-life violence.

2.8 Acknowledgements

Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

CHAPTER III

SOCIAL TRANSFERS AND SPATIAL DISTORTIONS

The analysis in this chapter was coauthored with Mark Colas and is forthcoming in the Journal of Labor Economics. Mark developed the structural model we use to explore location choice. I developed the quantification of the TANF welfare program embedded in the model. Mark performed the statistical analysis, and we both contributed to the writing of the manuscript.This work benefited from access to the University of Oregon high performance computer, Talapas.

3.1 Introduction

There is substantial variation in generosity of social transfers programs across states; a married household with two children and no income in 2017 could receive $1230 in monthly Temporary Assistance for Needy Families (TANF) benefits in New Hampshire, while the same married household in Louisiana would be ineligible for TANF.[30] Economists and policymakers have long debated whether these differences in transfer generosity lead poor households to migrate to locations with more generous transfer programs (so called "welfare magnets"), thereby distorting the distribution of households across space.[31]

Further, social transfers schedules are often decreasing in household income; household with lower income receive larger benefit payouts, all else equal.

The means-tested nature of these programs helps to reduce inequality and target resources at households with the greatest need. However, this same feature also reduces the returns of living in highly productive locations, as moving to a city where a household will receive higher pay may also lead to a reduction in transfers received. Therefore, means-tested social transfers may distort the location decisions of poor households by rewarding locating in less

[30]This applies to able-bodied households who meet the general eligibility criteria for TANF. We discuss the details of TANF eligibility in Appendix A.2.

[31]See e.g. Blank (1988), Borjas (1999), or Gelbach (2004).

productive cities.

To quantify these distortions, we build a quantitative spatial equilibrium model and embed within it a model of social transfer programs in the US. Locations vary in productivity levels, amenities, housing supply, and social transfer programs. Households choose the location which maximizes utility as a discrete choice. Wages and rents are determined in equilibrium. Transfer schedules vary across states, creating incentives for households to locate in states with more generous welfare programs. As in the models of Rosen (1979) and Roback (1982), households earn higher income in more productive cities, which creates an incentive to locate in these cities. However, these incentivizes are muted by the fact the social transfers schedules are decreasing in income; moving to a more productive city implies higher income but lower social transfers. Therefore, social transfers can lead to both an *earnings distortion* — an incentive to locate in low-wage cities — and a *generosity distortion* — an incentive to locate in states with more generous social transfer programs.

The model incorporates two social transfers programs, the Supplemental Nutrition Assistance Program (SNAP) and Temporary Assistance for Needy Families (TANF) programs, two of the largest social transfer programs in the US. Our model incorporates differences in TANF and SNAP programs across states, in addition to the non-linearities, kinks, and discontinuities present in the programs and the differences in eligibility and benefits allotment by household size and marital status.[32] This allows us to capture the complex system of spatial incentives created by these programs and to understand how these incentives differ across households. Further, our model incorporates both state and federal income taxes, allowing us to capture how distortions caused by transfer programs interact with the incentives created by income taxes.

Households are heterogeneous and vary in their race, marital status, number of children, experience group and education level. These household demographic characteristics play an important role in determining the amount of transfers a household receives. First,

[32]The SNAP benefits schedule is fixed across states, with the exceptions of Hawaii and Alaska. However, eligibility criteria and the ease at which households can apply for and receive benefits do vary across states.

demographic groups differ in their productivity and therefore their income levels. These income levels determine whether and where a household will be eligible for SNAP and TANF. Second, household demographics directly determine benefits through differences in demographic allotments in the social transfer functions. Finally, different demographics vary in their preferences over locations and thus their distribution across locations.

We quantify the model by utilizing data from the American Community Survey (ACS), the Survey of Income and Program Participation (SIPP), the tax simulator TAXSIM (Feenberg and Coutts, 1993), location-specific policy parameters of SNAP and TANF programs, and data on SNAP implementation across states. To quantify the parameters of household utility and therefore the household location choice, we combine data on household demographics, income, rent and location choice from the ACS with estimates of location choice elasticities from Colas and Hutchinson (2021). We use TAXSIM to quantify the state and federal income tax schedules. For our quantification of social transfer programs, we directly utilize location-specific formulas of TANF and SNAP. We use publications from the United States Department of Agriculture to quantify the SNAP benefit schedule. In cataloging the state variation in TANF programs, we rely heavily on the parameters and documentation collected by the Welfare Rules Database (The Urban Institute, 2019), in addition to state TANF manuals. We supplement this quantification of transfer programs with demographic specific take-up rates of TANF and SNAP that we estimate by combining SIPP data on program participation with data on SNAP application procedures and implementation across states from the SNAP Policy Database (Economic Research Service, 2019).

We then use the estimated model to quantify the spatial distortions caused by the current SNAP and TANF programs by comparing the equilibrium with the current programs to an equilibrium where social transfers are paid lump-sum.[33] We find that these programs lead to a substantial increase in the number of high school dropouts living in low-income cities

[33] There are two sources of inefficiency in the model: social transfers and taxes. Therefore all equilibria where both taxes and transfers are replaced by lump-sum transfers are efficient. Our main counterfactuals quantify the additional deadweight loss caused by social transfers, on top of the deadweight loss already caused by taxes.

and in locations with more generous transfer programs. Overall, the distortions caused by the current transfer programs lead to an increase in deadweight loss equal to 4.88%

Next, we consider three alternative transfer programs aimed at reducing the inefficiencies of the current programs. W first attempt to eliminate the earnings distortion by indexing the earnings used to calculate transfer benefits to local average earnings levels, such that households do not receive larger benefit amounts for locating in low-productivity cities. This leads to a roughly 50% decrease in deadweight loss of social transfers to 2.35% of total transfer spending. Second, we simulate the effects of harmonizing transfer schedules across states. This reduces locational inefficiency by considerably less than the earning index: the deadweight loss of social transfers decreases by only 14% to 4.19% of total transfer spending. Finally, we consider a combined program which both harmonizes transfer programs across locations and introduces an earnings index. We find that this combined policy intervention decreases deadweight loss of social transfers by 64%. Our results suggest that targeting both the earnings and generosity distortion caused by the current transfer programs can substantially reduce locational inefficiency while still preserving the fundamental means-tested nature of these programs.

A key limitation of our analysis is that we abstract away from externalitites arising from agglomeration effects, congestion effects, and endogenous amenities, all which have been shown to be empirically important in determining the distribution of populations across cities (see e.g. Glaeser and Gottlieb (2009), Diamond (2016), or Duranton and Puga (2020)). Further, we take state-level transfer policies as given, and do allow for the possibility that transfer functions are chosen by policy makers and may be endogenous to local population levels or prices.

This paper is related to a literature on "welfare migration" which analyzes the extent to which households move towards locations with more generous welfare programs (Blank, 1988; Walker, 1994; Enchautegui, 1997; Levine and Zimmerman, 1999; Meyer, 1998; Gelbach, 2004; Kennan and Walker, 2010). This paper incorporates differences in social transfer generosity across locations into a fully specified spatial equilibrium model and also highlights that the means-tested nature of welfare programs can disincentive households from moving to higher-paying locations. We show that in today's welfare environment, household location decisions are distorted predominately towards locations with low productivity, not towards so-called "welfare-magnet" states with generous transfer programs. To the best of our knowledge, ours is the first paper to quantify the locational inefficiency resulting from the progressivity of social transfers.

Notowidigdo (2020) studies the extent to which low out-migration rates of low-skilled workers in response to local labor market shocks can be explained by increases in transfers paid when local economic conditions deteriorate. While Notowidigdo (2020) focuses on the effect of local welfare programs on out-migration of workers from a given location, this paper focuses on the effects of transfer programs on the equilibrium distribution of heterogeneous households across cities.

A recent literature has quantified the distortionary effect of federal and state income taxes in spatial equilibrium (Albouy, 2009; Fajgelbaum et al., 2019; Coen-Pirani, 2021; Colas and Hutchinson, 2021).[34] This paper instead uses a spatial equilibrium model to study the distortion caused by social transfer programs, which 1) can vary spatially and 2) are generally decreasing in income.[35] Both these factors imply transfers can lead to spatial distortions. Further, while income taxes generally lead to larger distortions for high-skilled households (Colas and Hutchinson, 2021), the social transfers we highlight here almost exclusively affect low-income and low-skilled households.

Finally, this paper is related to a number of model-based papers quantifying the distortionary effects of social transfer programs on labor supply, household formation, and human

[34] Relatedly, Fajgelbaum and Gaubert (2020) characterize the optimal system of location- and group-specific transfers in a model with heterogeneous workers and spillovers. Rossi-Hansberg et al. (2019) study the optimal taxes and transfers in a spatial equilibrium model with multiple industries and occupation-specific externalities. Eeckhout and Guner (2017) study the optimal federal income tax schedule in a spatial equilibrium model.

[35] Albouy (2009) also analyzes differences in federal spending across locations affects his main conclusions about the efficiency costs of federal taxation. He concludes that differences in federal spending exacerbate the efficiency costs caused by federal taxation alone.

capital accumulation; and quantifying the resulting welfare consequences (see e.g., Greenwood et al. (2000), Keane and Wolpin (2010), Chan (2013), Blundell et al. (2016), Low et al. (2018), Guner et al. (2020), or Ortigueira and Siassi (2021)). In order to focus on the effects of social transfers on location choice, we abstract away from these margins in our paper.[36] We contribute to this literature by showing that the effect of social transfers on household location choice, previously absent from this literature, is responsible for a substantial efficiency cost.

3.2 Social Transfers Across Space

The federal government has provided food assistance and direct cash assistance to needy families for nearly a century under a variety of programs. SNAP and TANF, which offer food and cash benefits respectively, are two of the largest transfer programs for vulnerable households in the United States. In 2017, SNAP provided 64 billion dollars in food benefits to roughly 42 million households. TANF provided basic cash assistance totaling seven billion dollars during the same year.

Though SNAP and TANF are grounded in federal legislation, the amount of TANF or SNAP benefit a household receives is highly dependent on location choice. Indeed, whether or not a family is even eligible for SNAP or TANF is intimately tied to their place of residence. The dependence of social transfers on location is the consequence of two factors: (1) means testing and (2) policy variation between states. To see how these two factors influence transfer payments, first note that the formulas for SNAP benefits nationally and TANF benefits in most states follow the same basic structure.[37] To start, family size determines the maximum potential benefit a household can receive. To determine the actual benefit payment, a weakly increasing function of the household's unearned and earned income is subtracted from this maximum.

[36] We include an extension where we allow for endogenous labor supply in Section 3.6.2.

[37] We explain the SNAP and TANF formulas, including how the formulas vary across states, in Appendices A.1 and A.2.

Means Testing and the Earnings Distortion The amount of benefits a household receives based on this type of formula will vary with location due to means testing. Since household earnings enter into benefit calculation, differences in wage levels across US states and cities translate into differences in transfer payments. More concretely, Figure 21 displays the amount of monthly SNAP benefits as a function of monthly earnings for families with different numbers of children in 2017.[38] The graph on the left shows the schedules for single households and that on the right shows the schedule for married households. The benefits formulas are highly progressive: in the phase-out region of the benefits formulas each additional dollar of earnings leads to a 24 cent decrease in SNAP benefits.

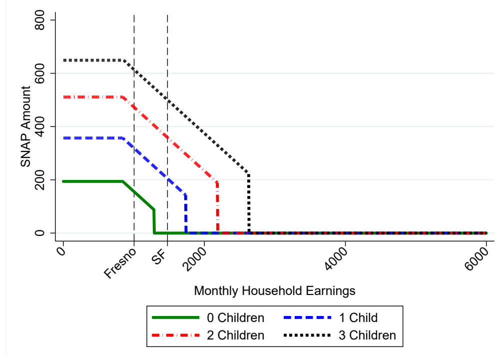
On this same figure we also plot the average household earnings for high school dropout households of the corresponding marital status who live in either Fresno, California; or the San Francisco Bay Area.[39] For both single and married households, the average household earning is considerably higher in San Francisco than in Fresno. These differences in earnings can lead to large differences in benefits. As an extreme example, consider two married households with three children, one who lives lives in San Francisco, one who lives in Fresno, and both who have earnings equal to the average earnings in their respective city. As a result of the differences in earnings, the household in Fresno would receive nearly $400 in monthly SNAP benefits while the household in San Francisco would not receive any benefits. More generally, we can see that households with San Francisco's average earnings receive less in transfers than households with Fresno's average earnings; however, the magnitude of the disparity depends on marital status and number of children. Furthermore, we can also imagine that higher-income households, such as households with higher education levels, may be ineligible for SNAP regardless of where they live.

[38]For this graph, we assume that 1) households are not made ineligible by asset tests or term limits, 2) their only source of income is earned income, 3) the household takes the maximum allowed excess-shelter deduction, and 4) the household only takes the standard deduction and the excess-shelter deduction.

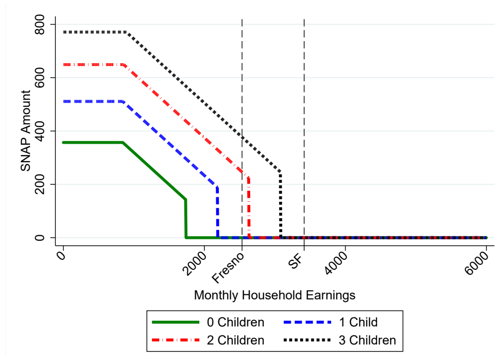
[39]This is average wage income of household head and spouse for households who's head is a high school dropout living in either the Fresno CBSA or the San Francisco CBSA. Calculations are from the 2017 ACS.

Figure 21: Monthly SNAP benefits as a function of earnings in 2017 for (a) single households and (b) married households.

For this graph, we assume that 1) households are not made ineligible by asset tests or term limits, 2) their only source of income is earned income, 3) the household takes the maximum allowed excess-shelter deduction, and 4) the household only takes the standard deduction and the excess-shelter deduction. The vertical lines give the average wage income of household head and spouse for high school dropout households living in either the Fresno CBSA or the San Francisco CBSA. Calculations are from the 2017 ACS.



(a) Single



(b) Married

Policy Variation and the Generosity Distortion Due to differences in state policy, though, holding earnings constant across location does not lead to equal benefit payments across space. Since the reform efforts of the 1990s, states have had substantial freedom— of which most states have taken advantage— to change their implementation of TANF, and to a lesser extent SNAP. First, states have wide latitude to alter the eligibility and accessibility parameters of both programs. Imagining that income is constant across location, a given family might be eligible for SNAP or TANF in one state but ineligible in another. Moreover, states also have considerable latitude to implement policies which do not alter a family's *de jure* eligibility for SNAP or TANF, but which nevertheless make it less likely that a family claims TANF or SNAP consistently (Currie and Grogger, 2001; Kabbani and Wilde, 2003; Bitler and Hoynes, 2010; Ganong and Liebman, 2018). For instance, Kabbani and Wilde (2003) find that frequency of re-certification requirements are associated with lower SNAP take-up among eligible households.

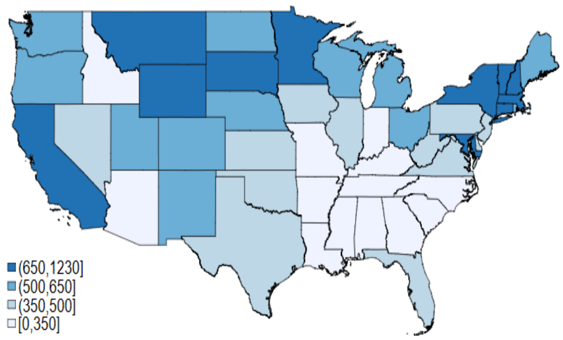
Specifically with regard to TANF, states also have broad authority to experiment with maximum benefits and levels of progressivity. In short, holding both income and also eligibility constant, TANF benefits still vary with location. As mentioned above, TANF in most states is calculated as a maximum benefit level minus some function of household income. However, both maximum benefit levels and benefit schedule progressivity differ massively across states.[40] Beyond simply altering the numbers in this traditional "welfare" formula, many states have experimented more drastically. Some have simplified their TANF payments, such as Wisconsin's implementation of a single, flat TANF payment for all eligible households. Other states have created more complex TANF systems.

To get a sense for how these differences in TANF policies translate into differences in benefits, Panel (a) of Figure 22 presents the maximum possible transfer for a married, able-bodied household with two children in 2017. In Louisiana, for example, two-parent households with

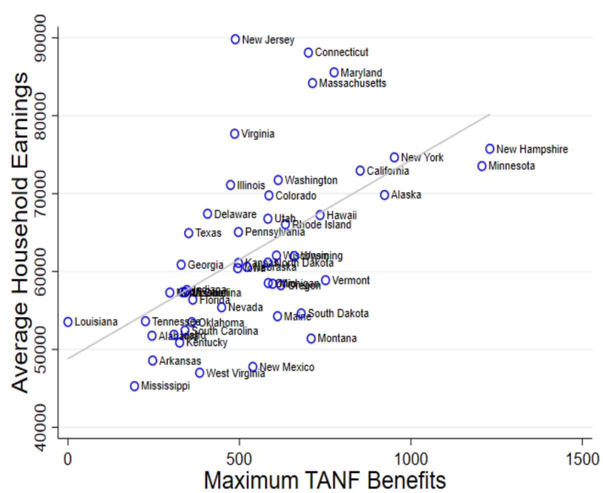
[40] In New Hampshire, for example, 100% of earned income can be deducted from a household's total income. Therefore, so long as a household remains eligible for TANF, increases in earnings do not lead to decreases in benefits. In Tennessee, on the other hand, a household can deduct a maximum of $250 in earned income each month, after which increases in earnings lead to dollar-for-dollar decreases in TANF benefits.

Figure 22: Maximum monthly TANF benefits across states and the relationship between average household earnings and TANF benefit generosity

Panel (a) shows the maximum possible TANF benefits (in dollars) for married households with able-bodied parents and two children in each state in 2017. Panel (b) is a scatterplot between state-level average household earnings and maximum possible TANF benefits for married households with able-bodied parents and two children in each state. Earnings are calculated as average wage income for the household head and spouse in the 2017 ACS.



(a)



(b)

able-bodied adults are categorically ineligible from receive TANF, and therefore the maximum benefit a family could receive is 0 dollars. On the other end of the spectrum, a married household with two children in New Hampshire with zero income would receive \$1230 each month. These differences create strong incentives to locate in states with generous TANF programs.

Interaction Between Generosity and Earnings Thus far, we have suggested that the social transfers system creates incentives to live in states with more generous transfer programs, and in locations where a given household will receive lower earnings. How do these incentives interact? First, note that since TANF schedules are generally set at the state level, the generosity distortion will mostly affect interstate location choice, while the earnings distortion can also affect intrastate location choice. Second, to get a sense of how these distortions jointly affect interstate location choices, in Panel (b) of Figure 22 we present a scatterplot of these maximum possible TANF benefits for married households with two children (X Axis) and average household earnings in each state (Y Axis). We can see there is a strong positive relationship: higher state-level earnings are generally associated with more generous TANF benefits. Therefore, these two incentives will generally work in opposite directions; the means-tested nature of these programs will encourage households to locate in states in which they receive lower earnings and therefore generally higher transfers, while differences in transfer generosity across states will incentive households to live in states with more generous benefits, which tend to have higher earnings.

Taken together, the evidence presented here suggests that social transfers differ substantially across space. The amount of transfers a household receives can therefore vary based on where a household chooses to live, potentially distorting the distribution of households across space. However, the magnitude of these distortions and what they imply for economic efficiency are open questions. To answer these questions, we now turn to our quantitative spatial equilibrium model.

3.3 Model

We build and estimate a spatial equilibrium model, in the tradition of Rosen (1979) and Roback (1982) and related to the recent models by Diamond (2016), Piyapromdee (2019), and Colas and Hutchinson (2021). Cities vary by wages, rents, amenities and social transfer programs. Households choose the city that maximizes utility as a discrete choice. Differences in wages and social transfer generosity across cities imply the amount of social transfers a household receives directly depend on a household's location choice. Wages and rents are determined in equilibrium.

Households differ in productivity, preferences, and household composition. These differences affect the menu of transfers households face. Therefore, the location decisions of some households, such as low-productivity households or households with children, will be more substantially distorted than those of high-productivity households without children. Households have idiosyncratic preferences over locations. The parameters which dictate the dispersion of these idiosyncratic preferences over locations play an important role in our analysis as they dictate the first-order extent to which differences in social transfers across locations affect the spatial distribution of households.

3.3.1 Household Location Choice

Individual households are indexed by i. Each household is endowed with a demographic group $d \in D$, which includes a household's education, experience group, marital status, number of children, and race.

Households choose a location j, and conditional on location, choose consumption of a tradeable good c and a housing level h_j. The price of the consumption good c is normalized to one. For now, we assume household labor supply is inelastic, conditional on location. Let Y_{dj} denote a household of demographic d's total post-tax, post-transfer income conditional on living in location j. We will refer to Y_{dj} as "net income" throughout. This is given by

$$Y_{dj} = I_{dj} + \Upsilon_d + b_{dj}\left(I_{dj}, \Upsilon_d\right) - \tau_{dj}\left(I_{dj}, \Upsilon_d\right),$$

where I_{dj} is the earned income of households of demographic d who live in city j, and Υ_d is unearned income for demographic group d. We assume unearned income, Υ_d, does not depend on the household's location. The function $b_{dj}(\cdot)$ represents SNAP and TANF transfers received by a household with demographic d in location j, and is written as a function of earnings I_{dj} and unearned income Υ_d.[41] We allow the transfer function to vary with j to allow for state-level differences in social transfer functions and by d to allow for differences in social transfer allotment by demographic groups, for example by number of children or marital status. Finally, the function τ_{dj} represents federal and state income taxes paid by the household as a function of earned income, unearned income, household demographics, and location.

Importantly, the transfer function depends on a household's location both through earnings, I_{dj}, and through location j directly. The dependence on earnings allows for an earnings distortion: households can choose locations where their earnings lead to larger transfer receipts. As transfer programs are generally decreasing in earnings, this implies that households are rewarded for locating in areas where they earn lower income. Second, the dependence on j implies location choices may be subject to a generosity distortion: households are rewarded for choosing locations with more generous transfer programs overall.

We allow for non-homothetic preferences to reflect that expenditure shares of housing decline in income (Albouy et al., 2016; Finlay and Williams, 2021). Specifically, we assume that preferences take the form of Price Independent Generalized Linear (PIGL) utility, a popular choice for non-homothetic preferences (Boppart, 2014; Alder et al., 2022; Eckert et

[41]We think of the food coupons provided by SNAP as equivalent to cash transfers, as is common in the literature (See e.g. Ortigueira and Siassi (2021)). Quantitatively, we will assume that all households take a maximum "shelter-cost" deduction for SNAP. In previous versions of the paper, we found similar results if we did not make this assumption.

al., 2018). Preferences can be represented by the indirect utility function

$$V_{ij} = \frac{1}{\eta}\left(Y_{dj}^{\eta} - 1\right) - \frac{\alpha_d}{\gamma}\left(r_j^{\gamma} - 1\right) + \Gamma_{ij},$$

where V_{ij} denotes household's i's indirect utility if they locate in location j, r_j is the location-specific cost of housing, η and γ are parameters that are assumed to be common across all households, α_d is a parameter which can vary by the household's demographic group, and Γ_{ij} represents the amenity utility household i receives when they live in location j.[42] This includes all non-pecuniary benefits the household receives for living in city j, including for example, the weather, restaurants, and idiosyncratic preference for living in a city.

By Roy's identity, the household's optimal expenditure share of housing conditional on living in city j is equal to

$$\frac{h_{dj}^{\star} r_j}{Y_{dj}} = \alpha_d r_j^{\gamma} Y_{dj}^{-\eta}. \tag{5}$$

From (5), we can see that the parameter γ will dictate the price elasticity of the housing share; a larger value of γ implies that, all else equal, increases in housing prices will lead to larger increases in the optimal housing share. The parameter η dictates the elasticity of the housing share with respect to expenditures. Finally, the parameter α_d determines the optimal level of the housing share. Quantitatively, we allow the parameter α_d to vary by the household's marital status and number of children, to reflect that preferences for housing relative to other goods may vary by household composition. As γ and η go to 0, the preferences become Cobb Douglas and the housing share is constant at α_d.

Amenities, Γ_{ij}, consist of a term that is common to all households of a given group, a term which measures how close the location is to an individual's birth state, and an idiosyncratic term which is unique to the individual household. We write a household's amenity utility

[42] In general, PIGL preferences do not admit a closed-form expression for the utility function except in the special limit cases discussed in Boppart (2014). As shown in Boppart (2014), this is a valid indirect utility specification if and only if $Y_{dj}^{\eta} \leq \frac{1-\eta}{1-\gamma}\alpha_d r_j^{\gamma}$. We confirm this condition holds quantitatively for all demographic groups across all equilibria we study.

for living in location j as

$$\Gamma_{ij} = \underbrace{\xi_{dj}}_{\text{Common term}} + \underbrace{f_d\left(j, Bstate_i\right)}_{\text{Distance from Birth State}} + \underbrace{\sigma_d \epsilon_{ij}}_{\text{Idiosyncratic}} . \tag{6}$$

The first term ξ_{dj} is the component of amenity in location j that is common to all households of demographic d. The next term $f_d\left(j, Bstate_i\right)$ gives the utility from living from a location near the household head's state of birth, $Bstate_i$. We parameterize $f\left(\cdot\right)$ as

$$f_d\left(j, Bstate_i\right) = \gamma_d^{hp} \mathbb{1}\left(j \in Bstate_i\right) + \gamma_d^{\text{dist}} \phi\left(j, Bstate_i\right),$$

where $\mathbb{1}\left(j \in Bstate_i\right)$ indicates that location j is within the households head's birth state, and $\phi\left(j, Bstate_i\right)$ gives the distance between the household head's birth state and location j. These parameters play an important role in our analysis as they dictate how far a household is willing to locate from their birth place to take advantage of differences in social transfers across locations. We specify $f_d\left(\cdot\right)$ as a function of the household head's state of birth, rather than city of birth, because our data only contain an individual's birth state. The model therefore does not account for the costs of relocating within one's birth state.

The term ϵ_{ij} is the idiosyncratic utility the household i receives for living in city j. We assume that ϵ_{ij} is distributed as Type 1 Extreme Value. The parameter σ_d dictates the dispersion of this idiosyncratic preference draw. The assumption of Type 1 Extreme Value idiosyncratic draws implies that the probability a given household i chooses to live in location j is given by

$$P_{ij} = \frac{\exp\left(\frac{\tilde{V}_{ij}}{\sigma_d}\right)}{\sum_{j'} \exp\left(\frac{\tilde{V}_{ij'}}{\sigma_d}\right)}, \tag{7}$$

where $\tilde{V}_{ij} = V_{ij} - \epsilon_{ij}$ denotes indirect utility less the idiosyncratic preference term. The partial equilibrium elasticity of this location choice probability with respect to expenditures is given by

$$\frac{\log P_{ij}}{\log Y_{dj}} = \frac{1}{\sigma_d} Y_{dj}^{\eta} \left(1 - P_{ij}\right).$$

We can see that a smaller value of σ_d implies that household location choices will be more responsive to changes in net income, all else equal. In the quantitative version of the model, we will assume one value of σ_d for households who have attended college, and one value for households who have less than a college education.

3.3.2 Housing Supply

Absentee landowners own plots of land which may be developed for housing. These plots of land vary in their marginal costs of development and therefore generate an upward sloping housing supply curve in each city. Let $r_j\left(H_j\right)$ be the marginal cost of producing an additional unit of housing as a function of the total amount of housing supplied in city j, H_j. We parameterize this following Kline and Moretti (2014) as

$$r_j = z_j H_j^{k_j}. \tag{8}$$

The parameter z_j is a parameter which shifts the level of housing costs in city j. A higher value of z_j implies higher costs of developing housing in city j, all else equal. The parameter k_j dictates the elasticity of the housing supply curve: a higher value of k_j implies that housing costs increase more rapidly with housing supply. We allow k_j to vary across cities to allow for differences in housing supply elasticities across cities. In particular, we let $\frac{k_j}{1+k_j} = \left(\nu_1 + \nu_2 \psi_j^{WRI}\right)$ where ψ_j^{WRI} gives a measure of the strictness of local land-use restrictions (Gyourko et al., 2008).[43] The parameter ν_1 dictates the overall level of the housing supply elasticity across cities while ν_2 dictates the extent to which a city's housing supply curve

[43] These measures are created by aggregating the measures of local land use restriction provided by Gyourko et al. (2008) to the core-based statistical area (CBSA). Similar parameterizations of the housing supply curve are also used in Diamond (2016), Piyapromdee (2019), Colas and Hutchinson (2021), and Colas and Morehouse (2022).

increases in local land-use restrictions.

We assume these landowner profits are distributed lump-sum back to households. Letting s_d denote the share of total landowner profits that are owned by a household of demographic d, and letting Π denote total landowner profits, a household's unearned income is given total landowner profits multiplied by their share of profits as $\Upsilon_d = s_d \Pi$.[44]

3.3.3 Labor Demand

In each city, perfectly competitive firms use a CES production function combining labor supplied by households from each of the following education groups: high school dropouts, high school graduates, college, and post college.[45] We index these education groups by $e \in \{e_1, e_2, e_3, e_4\}$. We assume that high school dropouts and high school graduates are perfectly substitutable and will be referred to as "unskilled labor" and that households with a college education those with post-college education are perfectly substitutable and will be referred to as "skilled labor". We allow for skilled and unskilled labor to be imperfectly substitutable.[46]

Let $L_{e1,j}$, $L_{e2,j}$, $L_{e3,j}$, and $L_{e4,j}$ give the total efficiency units of labor supplied by each of the four narrow education groups in city j. We can write the production function as

$$F_j\left(L_{e1,j}, L_{e2,j}, L_{e3,j}, L_{e4,j}\right) = A_j[(1-\theta_j)L_{Uj}^{\frac{\varsigma-1}{\varsigma}} + \theta_j L_{Sj}^{\frac{\varsigma-1}{\varsigma}}]^{\frac{\varsigma}{\varsigma-1}}, \tag{9}$$

where

$$L_{Uj} = L_{e1,j} + \theta_{Uj} L_{e2,j} \tag{10}$$

[44] We estimate share of landowner profit owned as the share of total interest, dividend, and rental income owned by each demographic group.

[45] We aggregate households with some college experience with households with a college degree. We consider an alternative specification in which households with some college are aggregated with high school graduates in Section 3.6.3.

[46] Card (2009) concludes that the "elasticity of substitution between dropouts and high school graduates is effectively infinite." Ottaviano and Peri (2012) come to a similar conclusion as they estimate an inverse elasticity of substitution between dropouts and high school graduates less than 0.04 across specifications. Their estimates are not statistically different from 0.

denotes the unskilled labor aggregate and

$$L_{Sj} = L_{e3,j} + \theta_{Sj} L_{e4,j} \tag{11}$$

denotes the skilled labor aggregate. The parameter A_j gives the city's total factor productivity, θ_j gives the skill intensity of skilled labor, and θ_{Uj} and θ_{Sj} give the factor intensity of high-school graduate labor and post-college labor, respectively. These technology parameters are allowed to differ across cities, reflecting exogenous differences in production technology across cities. Households are paid the marginal products of their labor. All else equal, households living in cities with higher values of A_j will have higher wages and therefore receive less social transfers. The parameter ς dictates how much relative wages change in response to changes in the ratio of skilled to unskilled workers.

Within education levels, demographic groups are perfect substitutes but vary in their productivity levels. Let ℓ_d give the efficiency units of labor inelastically supplied by a household of demographic d, reflecting the productivity level, hours worked, and propensity to be employed of the demographic group.[47] Total labor supply of each education level in each city is then given by sum of these efficiency units of labor. In particular, letting D_e give the sets of demographic group that classify as education level e, total labor supplied by education level e in city j is given by $L_{ej} = \sum_{d \in D_e} N_d \ell_d$.

3.4 Data and Quantification

In this section, we describe the data and estimation procedure. Details on how the production function and housing supply curves are taken to the data are included in Appendix B.2 and Appendix B.3, respectively.

[47]Importantly, we do not assume that households work full time. This is a departure from many papers using similar models, (see Colas and Hutchinson (2021), for example) who only use data on full-time workers. Full-time household receive higher income and therefore are more likely to be ineligible for transfers.

3.4.1 Data

We use the 5-year aggregated 2017 American Community Survey as our main data source (Ruggles et al., 2010). This dataset provides household-level data on respondents' location, state of birth, demographics, earned and unearned income, and housing costs. We define locations as Core Based Statistical Areas (CBSAs). Specifically, we chose the 70 CBSAs with the largest population in 1980.[48] We aggregate the remainder of locations to the nine census divisions. This gives us a total of 79 locations in our quantitative version of the model.

As discussed above, the extent to which social transfer affect a household's decisions depends on the household's demographics. In our quantification, we divide households into 128 demographic groups, differentiated by four education groups, two experience groups, marital status, number of children (0, 1, 2, and 3 or more), and race (non-minority vs. minority).[49] As we describe in Section 3.4.4, we allow household productivity levels to vary across marital status (reflecting more working adults), race, education, and experience. Conditional on income and location, transfer functions b_{dj} depend on marriage and number of children, reflecting the dependence of TANF and SNAP programs on these characteristics. All demographic groups vary in their preferences over locations, which is captured by differences in amenity values across cities. An important assumption is that these demographic characteristics are exogenous and do not depend on the social transfer system. In reality, marital status, education, and especially the number of children may be endogenous to the generosity of transfer programs.

We supplement this ACS data with data from the SIPP. In addition to data on household

[48]These 70 locations make up approximately 60% of the entire US population. We choose these CBSAs based on their 1980 populations, rather than their current populations, so that the set of locations is not affected by current transfer program generosity. In 1980, transfer generosity provided by Aid to Families with Dependent Children did not differ substantially across states and therefore the populations of these CBSAs would not be largely affected by transfer generosity.

[49]We define non-minority households as households in which the household head is white, non-Hispanic, and not an immigrant. In our baseline specification, we aggregate households into four education groups: high school dropouts, high school graduates, college (including some college), and post-college. In Section 3.6.3, we consider an alternative specification in which we instead aggregate households with some college education with high school graduates.

income and demographics, the SIPP contains detailed information on participation and transfers received from TANF, SNAP, and other programs. As we describe below, we use these data combined with the SNAP Policy Database (Economic Research Service, 2019) to estimate take-up and accessibility of social transfer programs across demographic groups and states.

3.4.2 Social Transfer Programs

The function $b_{dj}\left(I_{dj},\Upsilon_{d}\right)$ gives transfers as a function of earnings, unearned income, household demographics, and location. We assume b_{dj} consists of transfers from TANF and SNAP:

$$b_{dj}\left(I_{dj},\Upsilon_{d}\right)=b_{dj}^{T}\left(I_{dj},\Upsilon_{d}\right)+b_{dj}^{F}\left(I_{dj},\Upsilon_{d},b_{dj}^{T}\right).$$

The functions b_{dj}^{T} and b_{dj}^{F} give TANF and SNAP transfers received, respectively, taking into account a household's demographics, location, and earned and unearned income.[50]

To quantify these social transfer functions, we mostly rely on the administrative formulas for TANF and SNAP. However, there are several details of the data and of social transfer programs that are not modeled directly and need to be taken into account. First, even conditional on being eligible for transfers, take-up rates of social transfer programs in the data are often less than 100%. Second, we are not able to directly model some eligibility criteria, such as asset tests or time limits. To account for incomplete take-up rates and unmodeled eligibility criteria, we therefore supplement the administrative formulas with reduced-form estimates of the expected fraction of time a household will take-up transfers and meet the unmodeled criteria. Specifically, we model our transfer functions as the product of 1) "benefit amounts", the amount of transfers received conditional on taking up social transfers and meeting unmodeled eligibility criteria; and 2) transfer "accessibility", the reduced-form representation of the expected fraction of time a household will meet the unmodeled criteria

[50]TANF benefits are counted as unearned income for the sake of determining SNAP benefits, which is why TANF transfers b_{dj}^{T} are an argument in the SNAP function.

and take up social transfers. We write this as

$$b_{dj}^T = \tilde{b}_{dj}^T \times o_{dj}^T \quad \text{and} \quad b_{dj}^F = \tilde{b}_{dj}^F \times o_{dj}^F,$$

where $\tilde{b}_{dj}^T$ and $\tilde{b}_{dj}^F$ denote the TANF and SNAP benefit amounts, and o_{dj}^T and o_{dj}^F denote transfer accessibility.

Benefit Amounts The amount of transfers received are modeled using the institutional formulas of TANF and SNAP. The SNAP benefits formula is set federally and therefore all states in the continental US share the same benefits formula.[51] We give a brief overview of this formula here, with a detailed description in Appendix A.1. Generally speaking, SNAP benefits are equal to a "maximum allotment" minus 0.3 times "net income", given by income minus deductions. Both the maximum allotment and many of the deductions are increasing in family size. Our SNAP benefits function $\tilde{b}_{dj}^F$ follows the institutional SNAP benefits formula closely, accounting for differences in program parameters across household sizes.[52]

While the formulas determining SNAP benefits are largely a matter of federal policy, the welfare reform underpinning current TANF programs gave states wide latitude to change how TANF benefits are calculated. Conditional on eligibility, TANF benefit in most states are calculated as a benefit standard minus household income less deductions. As is the case under SNAP, benefit standards for TANF are normally increasing with household size; however in contrast to SNAP, each state sets its own benefit standard and chooses the number and size of deductions they offer. We collect data on these state- and demographic-specific parameters from the Welfare Rules Database (The Urban Institute, 2019). Further, as mentioned in Section 3.2, many states have experimented more drastically with their

[51] The formula is slightly different in Hawaii and Alaska. Our model accounts for this.

[52] The income eligibility tests can create discontinuities in the SNAP formula as a function of earnings. These discontinuities can prevent the model from converging. To deal with this, we replace the SNAP formula with a linear basis function in earnings in a small interval around these discontinuities.

TANF formulas and do not follow this same basic structure. For these state we supplement the information from the Welfare Rules Database with information from the individual state TANF manuals. Details are in Appendix A.2.[53]

Accessibility While SNAP benefits formulas are set federally, ease of use and access and some eligibility criteria vary across states, and can lead to substantial differences in SNAP enrollment rates (Currie and Grogger, 2001; Kabbani and Wilde, 2003; Ratcliffe et al., 2008; Bitler and Hoynes, 2010; Dickert-Conlin et al., 2010; Ganong and Liebman, 2018). To estimate SNAP accessibility taking into account these state-level differences in SNAP implementation, we combine household-level SIPP data on program participation, demographics and income with data on SNAP implementation across states from the USDA's SNAP Policy Database (Economic Research Service), which contains state-level data on eligibility criteria and application and certification procedures. In particular, we estimate via ordinary least squares the fraction of all months a given household in the SIPP receives SNAP benefits as a function of their demographic characteristics, the SNAP policy characteristics of the state in which they live, and their earnings as a fraction of the federal poverty level. Letting o_i^F be the fraction of months a given household receives SNAP benefits, we write our reduced-form estimating equation as

$$o_i^F = \beta^{F1}\text{Policy}_s + \beta^{F2}\frac{I_i}{FPL_{d(i)}} + \beta^{F3}\text{ABAWDWaiver}_s \times \text{ABAWD}_i + \beta^{F4}X_i^{\text{Rec}} + \varepsilon_i^F, \quad (12)$$

where Policy_s is a vector of state-specific SNAP implementation policies, $\frac{I_i}{FPL_{d(i)}}$ is household earnings as a fraction of the poverty line, ABAWDWaiver_s indicates that state s has time-limit waivers for able-bodied adults without dependents, ABAWD_i indicates household i is an able-bodied adult without children, and X_i^{Rec} is a vector of demographic control variables.

[53]Overall, we have tried to preserve as much of the state variation in TANF policy as our data allows. In situations where we cannot model a state's TANF formula exactly, we have opted to be general, using the policies which would apply to most TANF recipients most of the time. In several states, the formula for net income changes based on how long a family has received TANF benefits. For these states, we use the modal formula for net income that would apply in a majority of months.

We include in Policy_s five variables describing eligibility criteria, how often a household is required to re-certify their SNAP eligibility, and details on the application process.[54]

The estimates of (12) are displayed in Appendix C.1.[55] We find that all the policy variables have the expected sign and are quite predictive of state-level take-up rates. In particular, and consistent with Kabbani and Wilde (2003), we find that frequent re-certification requirements have a large negative effect on SNAP take-up. Using these estimates, we then calculate the SNAP accessibility measures o_{dj}^F as the predicted values of (12) for each demographic group and location.

We use a similar technique to estimate o_{dj}^T, our TANF accessibility measure. In particular, we estimate the fraction of months a given household in the SIPP receives TANF as a function of their demographic characteristics, the state in which they live in, and their income as a fraction of the poverty line:

$$o_i^T = \beta_s^{T1} + \beta^{T2} \frac{I_i}{FPL_{d(i)}} + \beta^{T3} X_i^{\text{Rec}} + \varepsilon_i^T, \tag{13}$$

where β_s^{T1} is a state-specific intercept, and, as before, $\frac{I_i}{FPL_d}$ is household earnings as a fraction of the poverty line and X_i^{Rec} is a vector of demographic variables. Note that, unlike our estimation procedure for o_{dj}^F, we do not use data on state-level TANF accessibility, and instead rely on state-level fixed effects to capture differences in TANF accessibility across states. We then set the TANF accessibility as the predicted values from (13) for each demographic group and location.

[54] We use the following 5 variables, which have been previously shown to be predictive of SNAP caseload (Dickert-Conlin et al., 2010): (i) whether the state uses broad-based categorical eligibility, (ii) whether one vehicle can be excluded from asset test, (iii) whether all vehicles can be excluded from the asset test, (iv) whether the state has an online application, and (v) how often a household must re-certify their SNAP eligibility. We use SNAP Policy Database from October of 2015, the latest date with no missing data on all variables.

[55] Transfer receipts are often under-reporting in survey data (Meyer et al., 2009). We therefore multiply our estimated accessibility measures by the inverse of the reporting rates calculated in Meyer et al. (2009), which calculates the ratio of the number individuals who report received SNAP and TANF in survey data divided by the number of individuals who receive these benefits according to administrative data sources.

3.4.3 Income Taxes

The function $\tau_{dj}(\cdot)$ represents the federal and state income taxes paid by the household. We assume this is given by

$$\tau_{dj}\left(I_{dj},\Upsilon_{d}\right)=\tau_{d}^{FED}\left(I_{dj}+\Upsilon_{d}\right)+\tau_{j}^{State}\times\left(I_{dj}+\Upsilon_{d}\right),$$

where $\tau_{d}^{FED}(\cdot)$ is a function which gives the federal income tax (including credits) as a function of household demographics d and total income. We assume that state income taxes take of the form of flat taxes with tax rate τ_{j}^{State}.

We quantify $\tau_{d}^{FED}(\cdot)$ and τ_{j}^{State} using the tax simulator TAXSIM, a program which replicates the federal and state tax codes in a given year, accounting for the different tax schedules, tax deductions, and credits afforded by various demographic groups, such as by marital status or number of dependents. Specifically, to quantify $\tau_{d}^{FED}(\cdot)$, we estimate separate linear splines of federal tax burden on household yearly income for each demographic group, taking into account the number of children and marital status associated with each demographic group. Our splines include knots at every 1000 dollars of household income. To estimate the state flat tax rate, τ_{j}^{State}, we calculate the average tax rate for a married household in each state with an income of \$60,000, roughly the median household income in 2017.

3.4.4 Productivity and Wages

Note that the demographic-specific income levels can be rewritten as:

$$I_{dj}=\ell_{d}W_{ej} \tag{14}$$

where e is the education level associated with the demographic group d, and where W_{ej} represents the wage levels in city j paid for one unit of labor of education level e. Recall that demographic-specific efficiency units of labor, ℓ_d captures both differences in the probability of working and productivity and hours worked conditional on working. We therefore specify ℓ_d as the product of the probability of working and productivity conditional on working. Specifically, let $\ell_d = E_d \tilde{\ell}_d$, where E_d is the probability of working for agents of demographic group d, and $\tilde{\ell}_d$ represents the productivity conditional on working. Further, we parameterize $\tilde{\ell}_d$, the productivity level conditional on working as $\log \tilde{\ell}_d = \beta_e X_d^{\text{Prod}}$ for each education level e, where each β_e is a vector of parameters and X_d^{Prod} is a vector of demographic variables indicating the marital status, experience level, and minority status, of demographic group d.

We estimate ℓ_d is two steps. In the first step we estimate E_d, the demographic-specific employment probability, for each demographic group as the proportion of households of this group who are employed.[56] In the second step, we estimate the productivity levels conditional on working, $\tilde{\ell}_d$, and the education-specific wage levels. Let i index individual households, and let X_i^{Prod} be a vector of household-specific demographic variables for the same characteristics included in the vector X_d^{Prod}. Using data on household income with at least one employed spouse from the ACS, we estimate the following equations via ordinary least squares using household-level earnings:

$$\log I_{ij} = \hat{\beta}_e X_i^{\text{Prod}} + \gamma_{ej} + \varepsilon_i \tag{14'}$$

for each education level e, where I_{ij} gives household i's earnings, ε_i represents household level measurement error and the set of γ_{ej}, our estimates of $\log W_{ej}$ for each city, are estimated as CBSA fixed effects. The underlying assumption is that there is no selection on unobservables which affect income after controlling for the vector of household demographics, X_i^{Prod}.

The above regression provides us with estimates of the β's, which we can use to calculate productivity conditional on working, $\tilde{\ell}_d$. We can then combine this with our estimates of employment probabilities E_d to calculate demographic specific productivity levels, ℓ_d. The

[56]For married households, ℓ_d represents the efficiency units supplied by the household head and spouse. We therefore estimate ℓ_d for married demographic groups as the proportion of households with at least one working spouse.

estimates of equation 14' are displayed in Appendix C.2.

3.4.5 Household Sorting

We estimate the parameters of the household utility function using a two-step procedure in which we estimate most parameters via maximum likelihood and calibrate several parameters using estimates from the literature.[57] It will helpful to refer to the portion of the indirect utility function that is common for all households of a given demographic group as the "mean utility". The mean utility of demographic d for living in location j is given by

$$\delta_j^d = \frac{1}{\eta}\left(Y_{dj}^{\eta} - 1\right) - \frac{\alpha_d}{\gamma}\left(r_j^{\gamma} - 1\right) + \xi_{dj}.$$

Further, let the "standardized indirect utility" denote the indirect utility divided by σ_e:

$$\hat{V}_{ij} = \hat{\delta}_{dj} + \hat{\gamma}_d^{hp}\mathbb{1}\left(j \in Bstate_i\right) + \hat{\gamma}_d^{\text{dist}}\phi\left(j, Bstate_i\right) + \epsilon_{ij} \tag{15}$$

where hatted values represent a value divided by σ_e (e.g. $\hat{\delta}_{ij} = \frac{\delta_{ij}}{\sigma_e}$).[58]

In the first step of estimation, we estimate these mean utility terms $\hat{\delta}_{dj}$, and $\hat{\gamma}_d^{hp}$ and $\hat{\gamma}_d^{\text{dist}}$, the parameters which dictate the preference for living near ones' state of birth, for each demographic group via maximum likelihood. The log-likelihood function for households of demographic group d can be written as

$$\mathcal{L}_d(\gamma_d^{hp}, \gamma_d^{\text{dist}}, \boldsymbol{\delta^d}) = \sum_{i=1}^{N_d}\sum_{j=1}^{J}\mathbb{1}_{ij}\log(P_{ij}), \tag{16}$$

[57]This procedure is similar to the two-step estimation technique commonly used in the industrial organization literature to estimate demand systems (Berry et al., 2004) and employed with increasing frequency in the urban economics literature (see e.g. Diamond (2016)). The key difference is that we calibrate two parameters rather than estimating them using instrumental variables.

[58]Because of the large amount of heterogeneity we assume, there are some demographic groups which we do not observe in each location. To deal with these, we assume that there is one household of each demographic group in locations in which we do not observe any observations of a given demographic group. This allows the estimation procedure to run.

where $\boldsymbol{\delta^d}$ give the vector of mean utility across locations for households of demographic group d, $\mathbb{1}_{ij}$ is an indicator equal to one if individual i lives in location j and zero otherwise, and P_{ij} is given (7).

In the second step of estimation, we decompose the estimated mean utility into the component of indirect utility arising from net income and rent, and the component arising from amenities. We first set $\eta = 0.248$ and $\gamma = 0.390$ based on the estimates from Finlay and Williams (2021), who estimate price and expenditure elasticities of housing demand using consumption microdata from the restricted-access Panel Study of Income Dynamics.[59] Next, we choose α_d to match the share of housing of each marital status by number of children group in the data. Specifically, using data on renters from the ACS, we calculate the median share of income spent on housing for each combination of marital status by number of children group. We then numerically choose the α_d parameters such that the average housing shares of these groups are equal to those in the data.

This leaves the parameters which determine the dispersion of the idiosyncratic preference shock, σ_d. Recall that in Section 3.3.1 we showed that the elasticity of location choice with respect to net income is given by $\frac{\log P_{ij}}{\log Y_{dj}} = \frac{1}{\sigma_d} Y_{dj}^{\eta} (1 - P_{ij})$. Therefore, once η has been calibrated, this elasticity pins down the parameter σ_d. We set σ_d to match estimates of partial equilibrium location choice elasticities from previous studies. As noted in Section 3.3.1, we choose one value of σ_d for households with college experience and one value for households with less than college. In our main specification, we choose these two values to match estimates of partial equilibrium elasticities of location choice from Colas and Hutchinson (2021), who estimate location choice elasticities by creating synthetic tax instruments which generate variation in after-tax wages across cities. [60] We examine the robustness of our

[59] Finlay and Williams (2021) then combine these estimates with a spatial equilibrium model with non-homothetic preferences to quantify the role of rising income inequality on diverging location choices between skilled and unskilled households. The income elasticity estimate from Finlay and Williams (2021) is close to that estimated by Albouy et al. (2016).

[60] Given the Cobb-Douglas utility function in Colas and Hutchinson (2021), the partial equilibrium elasticity of location choice is given by $\frac{\log P_{ij}}{\log Y_{dj}} = \frac{1}{\sigma_d^{CD}} (1 - P_{ij})$, where σ_d^{CD} is the dispersion parameter of the idiosyncratic preference draw in their model with Cobb-Douglas utility. The average elasticity for

findings to alternative values of this parameter in Section 3.6.1.

Unobserved common amenities, ξ_{dj}, can then be calculated given information on net income, rents and the estimates of mean utility. We can back out the unobserved common amenities using $\xi_{dj} = \delta_{dj} - \left(\frac{1}{\eta}\left(Y_{dj}^{\eta} - 1\right) - \frac{\alpha_d}{\gamma}\left(r_j^{\gamma} - 1\right)\right)$.

Parameter Values The estimates of the key parameters of household preferences are displayed in Tables 5 and 6. We calibrate values of σ_d of 1.1 for household with a college education and 1.7 for households without a college education. We find that α_d, the parameter that dictates the strength of housing versus the tradable good, is increasing in number of children for married households, but slightly decreasing in number of children for single households.

Table 5: Parameter values of household preferences

		Estimate	Source/Target
Indirect Utility: $V_{ij} = \frac{1}{\eta}\left(Y_{dj}^{\eta} - 1\right) - \frac{\alpha_d}{\gamma}\left(r_j^{\gamma} - 1\right) + \Gamma_{ij}$			
Income Elasticity	η	0.248	Finlay and Williams (2021)
Price Elasticity	γ	0.390	Finlay and Williams (2021)
Housing Preferences	α_d	See Table 6	Median housing shares from ACS
Variance of Prefs	σ_d		Elasticities from Colas and Hutchinson (2021)
Less than College		1.7	
College Plus		1.1	

Table 6: Calibrated values of α_d

We numerically choose the α_d parameters to match the median housing shares by marital status and number of children in the estimation data.

		Single	Married
Children:	0	0.42	0.37
	1	0.41	0.41
	2	0.39	0.43
	3	0.37	0.46

households of demographic group d is then $\sum_{i \in I_d}\sum_{j \in J}\frac{1}{\sigma_d^{CD}}\left(1 - P_{ij}\right)$, where I_d is the set of households in demographic group d. The average elasticity of households of demographic d in our model is equal to $\sum_{i \in I_d}\sum_{j \in J}\frac{1}{\sigma_d}Y_{dj}^{\eta}\left(1 - P_{ij}\right)$. We choose σ_d such that the average partial equilibrium elasticity of location choice in our model matches that from Colas and Hutchinson (2021) for both households with and without college education.

The estimates of the birth state premium parameters are presented in Appendix C.3. We find that the disutility associated with locating far from one's birth place is largest for households with low education, indicating that low-education households need to pay a large utility premium to take advantage of generous welfare programs in states far from their birth place.

In Appendix C.5, we simulate general equilibrium elasticities of location choice with respect to transfers and compare our elasticities with the literature. The average elasticity for high school dropout households is 0.024 — a one percent increase in local transfers leads to a 0.024 percent increase in the population of high school dropout households. The elasticity is strongly increasing in number of children and is larger for single households than married households; single, high school dropout households with children have an elasticity of 0.081. This is consistent with the elasticities in Kennan and Walker (2010), who find that a 20% increase in benefits is associated with a 1% to 2% increase in state population of single women with dependents after 10 years, implying an elasticity of .05 to .1.

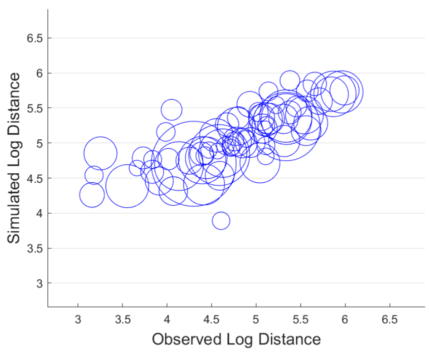
3.4.6 Model Fit

As highlighted earlier, household preferences to live close to their birth place play an important role in determining the magnitude of the generosity distortion relative to the earnings distortion. Figure 23 examines how well the model replicates households' average log distance away from their birth place by plotting the simulated and observed average log distance between a household head's birth state and chosen location for each of the four education levels. Each circle plots the average log distance for households who choose to live in a specific location. The fit is quite good.

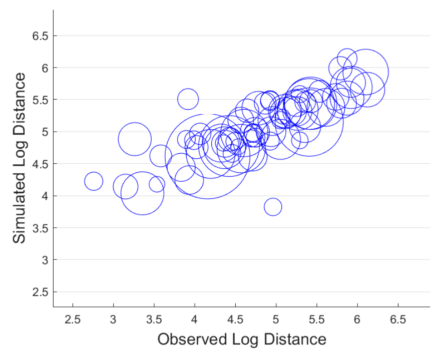
Next, we examine how location decisions vary with the generosity of transfer programs, where we measure the transfer generosity associated with a location as the amount of transfers a household with zero income would receive in this location, averaged over demographic groups. Figure 24 plots the average transfer generosity at choice location for all households

Figure 23: Model fit: log distance from birth state by education group

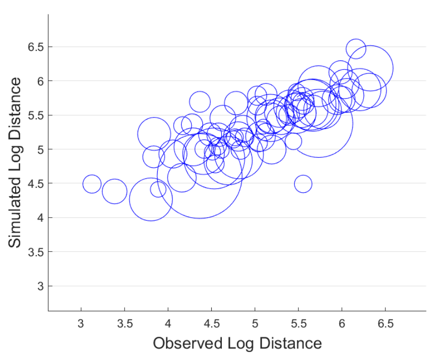
Each circle represents a CBSA, and the size of the circle is proportional to population. The X-axis of each graph gives the observed average log distance between a household's location and the birth place of the household head. The Y-axis gives the simulated average log distance. Each panel shows the fit for one of the four narrow education groups.



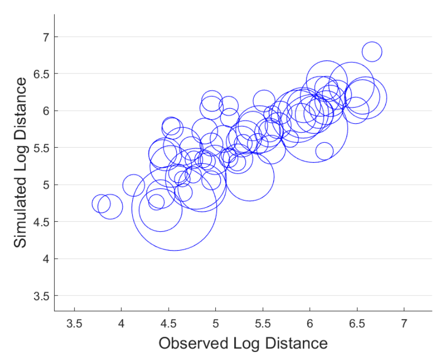
(a) HS Dropouts



(b) HS Graduates



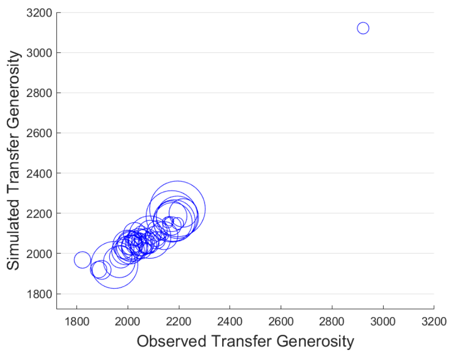
(c) College



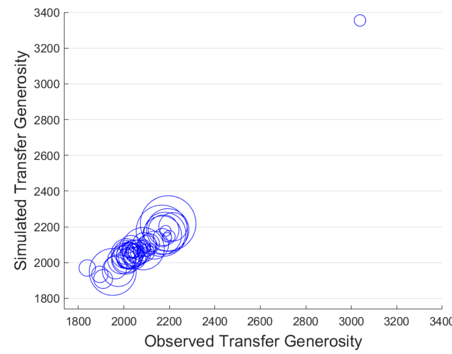
(d) Post College

Figure 24: Model fit: welfare generosity at destination by birth state and education group

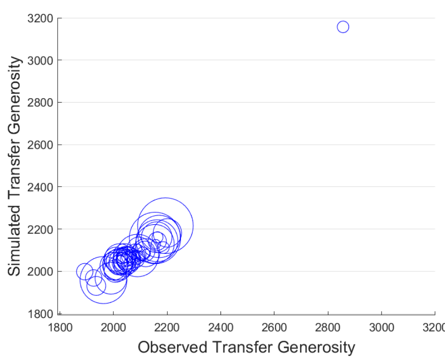
Each dot represents a CBSA, and the size of the dot is proportional to population. The X-axis of each graph gives the observed average welfare generosity at destination by all households from a given birth state. The Y-axis gives the simulated average welfare generosity. The outlier at the top right of the graph is Hawaii. Each panel shows the fit for one of the four narrow education groups.



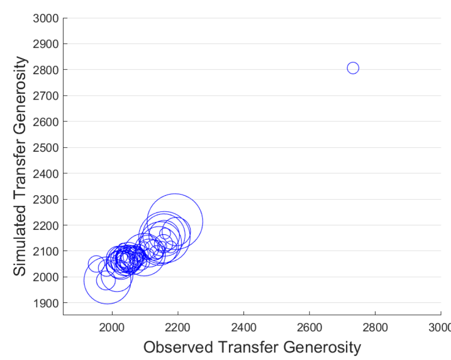
(a) HS Dropouts



(b) HS Graduates



(c) College



(d) Post College

from a given birth state and education group in the model and the data. The fit is very good.[61]

We next examine the fit with respect to housing share. First, we examine how well the model replicates average housing share by education level. Figure 25 shows the housing cost as a fraction of earnings simulated by the model and in the estimation data for each of the four education groups. The model slightly over predicts the housing share of high school dropouts, but overall the fit is good. In the simulation and data, the housing share is decreasing in education, reflecting that income is increasing in education and the income elasticity of the housing is less than one. In Appendix C.4, we show the average housing share for each of the 128 demographic groups in the model and the data.

Figure 25: Model fit: housing cost as a fraction of earnings by education group

The blue bars show the median housing cost as a fraction of earnings by each education group in the estimation data. The red bars show the mean housing cost as a fraction of earnings by each education group in the model. The model produces a mean housing share of 0.34 across all education groups.

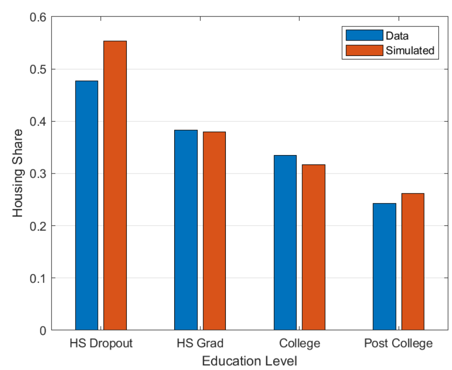
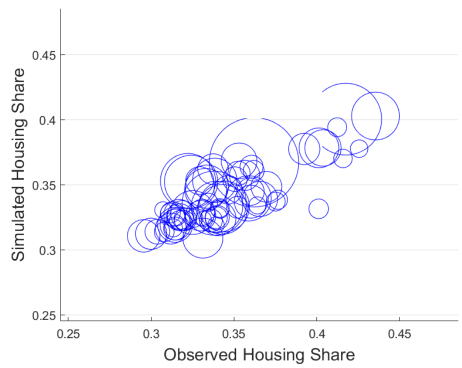


Figure 26 examines how well the model can replicate housing shares across locations. The fit is quite good, suggesting that the model does a good job of replicating how the housing share responds to differences in income levels and rents across cities. In Appendix C.4, we examine the housing shares across cities separately for each of the four education groups.

[61]The outlier in the upper right of each graph corresponds with households born in Hawaii. Hawaii has more generous SNAP parameters than the contiguous United States.

Figure 26: Model fit: housing cost as a fraction of earnings by location

The Y-axis shows the average housing share in the model and the X-axis shows the median housing share in the estimation. Circles are proportional to city size.



3.5 Results

In this section, we use the estimated model to measure the spatial distortions caused by the US social transfer system and to consider alternative systems. To visualize and quantify spatial distortions, we compare the equilibria generated by the various transfer schemes to the equilibrium when the current transfer system is replaced by lump-sum transfers. In particular, we consider an equilibrium in which all households of a given demographic group receive the same lump-sum amount, and the total amount of net transfers received by each demographic group is the same as under the current transfer system.[62]

We include additional counterfactual results in Appendices C.6 through C.10.

3.5.1 The Current US Social Transfer System

First, we quantify the distortions associated with the current TANF and SNAP programs.

Earnings Distortion As argued above, the current US transfer system incentivizes low-income households to locate in low-productivity cities. To quantify this distortion, Column A in Panel I of Table 7 gives the percentage difference in the number of households of various

[62]That is, we enforce that the total amount of transfer received minus taxes paid for each demographic group is the same as under the current transfer system. We chose this lump-sum transfer system as it does not directly transfer income across demographic groups relative to the current transfer system.

Table 7: Spatial distortions caused by current transfer programs and by alternative transfer programs

Panel I gives the percentage difference in the number of households locating in low-earnings cities compared to the equilibrium with lump-sum transfers. Low-earning locations are defined as the ten cities with the lowest average income in the data. Panel II gives the percentage difference in the number of households locating in generous-benefit locations compared to the equilibrium with lump-sum transfers. Generous-benefit locations are defined as the ten cities which provide the highest transfers to households with zero income. Deadweight loss is measured as a percent of total spending on transfer programs. Transfer spending less tax payments is held constant across counterfactuals. Column A measures the distortions of the current transfer system. Column B analyzes the case in which household earnings are indexed to average local earnings when calculating social transfers. Column C analyzes the case when transfer policies are harmonized across states. Column D analyzes the case with both the earnings index and harmonized transfers.

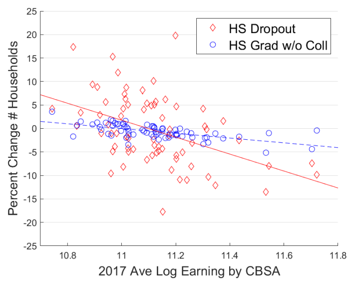
	A Baseline	B Earnings Adjustments	C Harmonize	D Earn Adj+ Harmonize
I. % Δ Low-Earning Locations				
HS Dropout	3.89	1.41	3.21	0.27
HS Grad	0.17	-0.34	0.24	-0.33
College	0.33	0.17	0.28	0.19
Post College	-0.49	-0.21	-0.50	-0.26
II. % Δ Generous-Benefit Locations				
HS Dropout	3.65	6.19	-0.08	1.33
HS Grad	-1.84	-0.76	-1.67	-0.86
College	-0.84	-0.90	-0.72	-0.86
Post College	0.22	0.13	0.26	0.26
III. Deadweight Loss	4.88	2.35	4.19	1.77

education levels choosing low-earnings cities in the equilibrium with the current SNAP and TANF programs relative to the equilibrium with lump-sum transfers. Low-earning locations are defined as the ten cities with the lowest average earnings in the data.

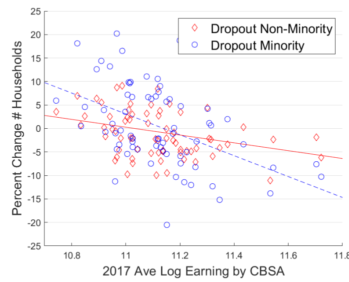
We can see that the current transfer system leads to an increase in the proportion of high school dropout households living in these cities. The first row ("HS Dropout") indicates that the number of high school dropout households who choose to locate in these low-income cities increases by 3.89% when we move from the lump-sum transfers equilibrium to the equilibrium with the current transfer programs. Households with higher education, however, are mostly unaffected, as their income levels make them less likely to be eligible to receive these transfers. These patterns are echoed in Figure 27, which shows the change in CBSA population relative to the lump-sum transfers equilibrium for various demographic groups. Across the panels,

Figure 27: Earnings distortion with baseline transfer programs: Counterfactual population relative to lump-sum transfers for current transfer system

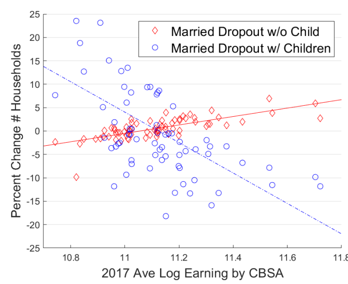
Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel (a) presents results for high school graduates (without college) and high school dropouts, Panel (b) presents results for non-minority high school dropout households compared to minority dropout households, and Panel (c) presents results for married high school dropouts with children and without children.



(a) Education Groups



(b) Minority vs. Non-Minority



(c) Children vs. Without

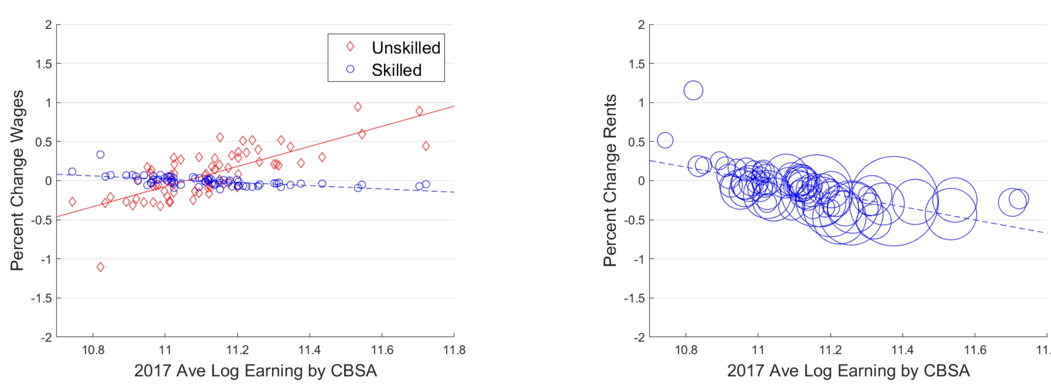
we can see an increase in the number of high school dropout households living in low-earning cities, with minority and households with children showing the largest changes. We further analyze heterogeneity in this distortion within high school dropout households in Appendix C.6. Appendix C.7 explores the consequences of the earnings distortion on average earnings across education groups.

Generosity Distortion The current system also incentivizes households to locate in states with generous transfer programs, either in the form of more generous TANF benefits or more accessible SNAP programs. We quantify this distortion in Panel II of Table 7, where we show the percentage change in the number of households living in the cities with the most generous transfer programs. To measure the transfer generosity of a location, we again calculate how much transfers a household of each demographic type with zero income would receive in this location. We then calculate the average of these zero-income transfers over demographic types. The "Generous-Benefit Locations" are defined as the ten locations with the highest average zero-income transfers across demographic groups. The current transfer system leads to a 3.65% increase in the number of high-school dropout households living in generous-benefit locations.

General Equilibrium Effects Figure 28 shows equilibrium changes in prices compared to the equilibrium with lump-sum transfers. Panel (a) shows the change in unskilled and skilled wages as a result of the current transfer programs. Unskilled wages decrease in low-income cities, reflecting the increase in the ratio of unskilled to skilled workers. Panel (b) shows the change in equilibrium rents. The transfer programs lead to an increase in rents in low-income cities, as transfer programs increase demand for living in those cities. As we show in Appendix C.6, these general equilibrium price changes can lead to "crowding out" of low-skilled households who are unlikely to receive large transfers, such as married households without children; these households are less likely to live in low-productivity locations as a result of the increase in rents and decrease in low-skilled wages.

Figure 28: Earnings distortion with baseline transfer programs: Counterfactual prices relative to lump-sum transfers for baseline transfer programs across cities

Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel (a) presents change in wages and Panel (b) presents changes in rents.



(a) Wage Changes (b) Rent Changes

Deadweight Loss To measure the efficiency cost of a given tax and transfer program, we calculate deadweight loss as the total equivalent variation of switching from the equilibrium with lump-sum taxes and transfers to the equilibrium in question.[63] We calculate equivalent variation as the household-specific lump-sum transfer that, given prices implied by the efficient equilibrium with lump-sum taxes and transfers, would provide the same utility level as the counterfactual in question. We then integrate equivalent variation over all households in the model to calculate deadweight loss. We provide additional details in Appendix B.4.

Note that there are two sources of inefficiency in the model: social transfers and income taxes. Therefore, any equilibrium allocation where both taxes and social transfers are replaced by lump-sum transfers is Pareto efficient.[64]

Our goal is to quantify the portion of deadweight loss that is caused by the transfer system alone. To this end, we calculate the additional deadweight loss caused by social transfers, on top of the deadweight loss already caused by taxes. That is, we first calculate the deadweight caused by taxes alone, by calculating the deadweight loss of an equilibrium when the current tax system remains, but the social transfer system is replaced with lump-sum transfers. We then add the distortion caused by social transfers and calculate the total deadweight loss in an equilibrium with both taxes and social transfers. The deadweight loss of social transfers is calculated as the deadweight loss caused by both taxes and transfers minus the deadweight loss caused by taxes alone. Our results focus on this additional deadweight loss caused by transfers alone.

As shown in Panel III of Table 7, the current social transfer programs lead to an additional deadweight loss equal to 4.88 percent of total transfer payments; for each dollar spent on transfers, there is a locational inefficiency of transfers equal to nearly 5 cents.

[63]This is the classic definition of deadweight loss as suggested by Mohring (1971) and Kay (1980).

[64]This relies on the assumptions that 1) all markets are competitive, and 2) there are no externalities (e.g. no agglomeration effects or endogenous amenities). See Colas and Hutchinson (2021) or Fajgelbaum and Gaubert (2020) for a proof.

3.5.2 Alternative Transfer Programs

Indexing Household Earnings to Average Local Earnings Social transfers incentivize households to live in low-productivity cities because a household's income, and therefore the transfers they would receive, depend on where they live. As a potential way to lessen this distortion, we consider indexing the earnings used to calculate transfer benefits to local average earnings levels. In this case, household earnings are measured against average earning level in a city, and therefore households are not penalized for living in cities where average earnings are higher. Formally, let $\hat{I}_{dj} = \frac{I_{dj}}{\bar{I}_{e1,j}}$ be local average earnings-adjusted household earnings, where $\bar{I}_{e1,j}$ is the average composition-adjusted earnings of high school dropout households in city j. Then transfers are calculated as $b_{dj}\left(\kappa \hat{I}_{dj}, \Upsilon_d\right)$, where κ is a parameter we choose to keep total transfers equal to their baseline levels. As $\hat{I}_{dj}$, local average earnings-adjusted household earnings, are what determines transfer receipt, households are not penalized for choosing locations where average earnings are high.

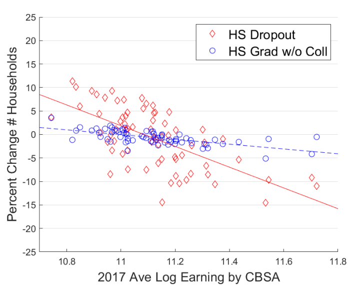
The results are displayed in Column B of Table 7. The local earnings adjustment significantly reduces the distortion towards low-income cities, as it essentially removes the incentive to locate in cities where average earnings are low. However, the generosity distortion is exacerbated: the number of high school dropout households in generous-transfer locations is 6.19% higher than in the case with lump-sum transfers compared to only 3.65% higher in the baseline. The fact that this is higher than the baseline case reflects the positive correlation between state-level earnings and transfer generosity documented in Figure 22b: locations with lower earnings also tend to have less generous transfer programs. The deadweight loss of social transfers with the earnings index is equal to 2.35% of total transfer payments, roughly 50% less than the baseline case.

Harmonizing Transfer Programs We remove the differences across locations in transfer generosity by standardizing the SNAP and TANF benefit functions across all states and setting o^T_{dj} and o^F_{dj}, TANF and SNAP accessibility, to the population-weighted average across states.[65] To keep net transfer spending constant, we additionally add lump-sum transfers so that the total transfers less taxes paid to each demographic group are the same as under the current transfer program.

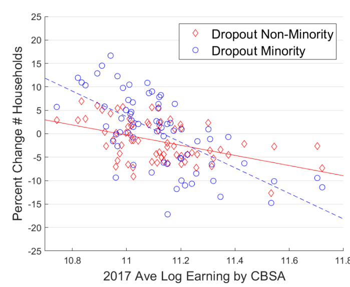
The main results are displayed in Column C of Table 7 and in Figure 29. Panel A of Table 7 shows the earnings distortion given the harmonized transfer programs. Household location choices are distorted towards low-income cities — 3.21% more high school dropout households locate in low-income cities compared to the case with lump-sum transfers.

Figure 29: Earnings distortion with harmonized transfer programs: Counterfactual population relative to lump-sum transfers for harmonizing transfer programs across cities

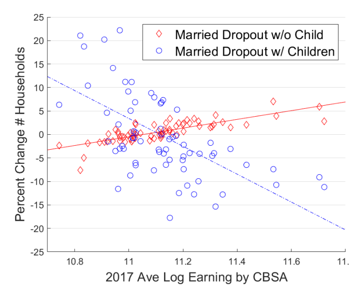
Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel (a) presents results for high school dropouts and high school graduates, Panel (b) presents results for non-minority high school dropout households compared to minority dropout households, and Panel (c) presents results for married high-school dropouts with children and without children.



(a) Education Groups



(b) Minority vs. Non-Minority



(c) Children vs. Without

Panel B of Table 7 shows the generosity distortion. The distortion towards generous states is effectively eliminated, and the proportions of household who locate in originally generous locations is similar to the equilibrium with lump-sum transfers.

All together, we find a deadweight loss of social transfers equal to 4.19% of transfer spending with the harmonized transfer system, only 14 percent less than the current system. Harmonizing transfers is significantly less effective than the earnings index at reducing deadweight loss. Taken together, these previous two counterfactuals suggest that most of the locational inefficiency arising from the current transfer system is due to the fact that

[65] Specifically, we set all TANF benefit formulas to the formula used in California, the largest state by population. Recall that Hawaii and Alaska have different parameters in their SNAP benefit function. These are standardized as well in this counterfactual.

transfer programs reward living in low-productivity cities, with a much smaller proportion due to the differences in transfer generosity across locations.

Combined Program Finally, we consider a program which targets both distortions by harmonizing transfer functions across states and indexing household earnings to local average earnings levels. The results are presented in the Column D in Table 7. We can see that the number of households in low-income cities and generous-benefit locations are relatively similar to the lump-sum transfers equilibrium, suggesting both the earnings distortion and generosity distortions are small. Further, as we show in Appendix C.7, average earnings across education groups are similar to those in the equilibrium with lump-sum transfers, implying that this policy intervention would lead to a substantial decrease in earnings inequality compared to current programs. Overall, we find a deadweight loss of social transfers equal to 1.77% of total transfer spending, a reduction of 64% from the baseline case.

3.6 Robustness

3.6.1 Alternative Parameter Values

We now calculate the distortions associated with current transfer programs using alternative values of σ_d, the parameter dictating the variance of the idiosyncratic preference draw. Details on these alternative calibrations are included in Appendix B.5.

The results are displayed Table 8. The first column displays the spatial distortions caused by current transfer programs given the baseline calibration of σ_d. The following three columns show the results when we base our calibration of σ_d on the estimates from Notowidigdo (2020), Diamond (2016), and Suarez Serrato and Zidar (2016), respectively. The results are qualitatively similar across specifications, but vary in their magnitudes. These results highlight the importance of the dispersion of idiosyncratic preferences in our quantitative results.

Table 8: Spatial distortions caused by current transfer programs under alternative model calibrations

The first column presents results with the baseline calibration. The next three columns calculate the distortions associated with current transfers program when we use alternative values of σ_d based on estimates of elasticity of location choice from other papers. See Table 7 for details.

		Alternative Estimates of Location Choice Elasticity		
	Baseline	Notowidigdo (2020)	Diamond (2016)	Suarez Serrato and Zidar (2016)
I. % Δ Low-Earning Locations				
HS Dropout	3.89	7.72	2.42	0.83
HS Grad	0.17	0.04	0.25	0.16
College	0.33	0.23	0.07	0.05
Post College	-0.49	-0.76	-0.07	-0.03
II. % Δ Generous-Benefit Locations				
HS Dropout	3.65	7.09	2.10	0.66
HS Grad	-1.84	-3.60	-1.12	-0.35
College	-0.84	-0.98	-0.19	-0.11
Post College	0.22	0.26	0.02	0.01
III. Deadweight Loss	4.88	6.43	2.15	1.05
IV. Calibrated Values of σ_d				
Less than College	1.7	0.80	3.04	10.16
College Plus	1.1	0.93	6.69	11.75

3.6.2 Elastic Labor Supply

Recall that in our baseline setting, a household of demographic group d exogenously supplied ℓ_d units of labor, regardless of where they lived and the wages they faced. We now allow for a household's labor supply to depend on an endogenous component, representing endogenously-chosen hours worked, and an exogenous component, reflecting fixed differences in labor productivity. Specifically, we assume a household's total efficiency units of labor supplied is given by $\tilde{\ell} \times \ell_d$, where $\tilde{\ell}$ denotes hours of labor that the household chooses to supply and ℓ_d is an exogenously-given productivity component. Earned income is given by total efficiency units of labor supplied multiplied by the wage rate: $I_{dj} = \tilde{\ell} \times \ell_d \times W_{ej}$.

Let indirect utility conditional on supplying $\tilde{\ell}$ units of labor and living in location j be given by

$$V_{ij}\left(\tilde{\ell}\right) = \frac{1}{\eta}\left(\left(Y_{dj}\left(\tilde{\ell}\right)\right)^{\eta} - 1\right) - \frac{\alpha_d}{\gamma}\left(r_j^{\gamma} - 1\right) + \Gamma_{ij} - \frac{\kappa_d}{\zeta}\tilde{\ell}^{\zeta}$$

where we now write net income, $Y_{dj}\left(\tilde{\ell}\right)$ as a function of hours worked, $\tilde{\ell}$, and where $\frac{\kappa_d}{\zeta}\tilde{\ell}^{\zeta}$ gives the disutility of working $\tilde{\ell}$ hours. The parameter κ_d is a parameter which governs the overall level of disutility associated with labor supply and is allowed to vary by demographic group. The parameter ζ dictates the elasticity of labor supply with respect to wages.

Calibration and Estimation With endogenous labor supply, we must calibrate the new parameters κ_d and ζ. We must also modify our the strategy through which we estimate wage levels and demographic-specific productivity levels to account for the fact that hours are chosen endogenously. Therefore differences in earnings across households and locations reflect not only productivity and wages, but also hours worked. We give a brief overview of our calibration and estimation strategy here and provide greater detail in Appendix B.6.

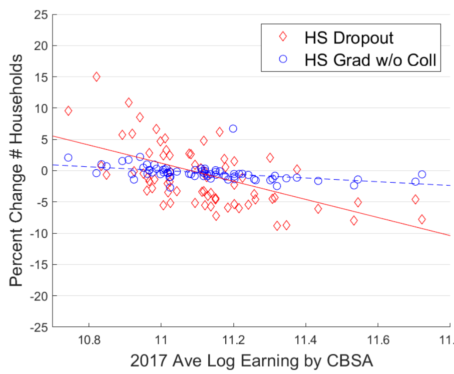
We estimate demographic-specific productivity levels and wages using a similar strategy to that outlined in Section 3.4.4. The key difference is that we use data on earnings per hour, rather than total earnings, to account for the fact that different households have endogenously chosen different amount of hours to work. We choose ζ, the parameter which dictates the elasticity of labor supply, based on the estimates of uncompensated total hours elasticities from Bargain et al. (2014). Finally, we choose κ_d to match the average hours worked nationally by each demographic group.

Results The distortions caused by the current transfer system given endogenous labor supply are shown Figure in 30 and in Table 9. The changes in the spatial distribution of low-education households are similar in magnitude to those in the baseline model. We can also see that the current transfer system leads to a decrease in labor supply of high school dropouts and high school graduates. This decrease is most pronounced in low-earning cities. As a result, the deadweight loss is considerably larger than in the case with inelastic labor supply. This is what we expect, as now social transfers lead to a distortion of both location choice and labor supply choice.

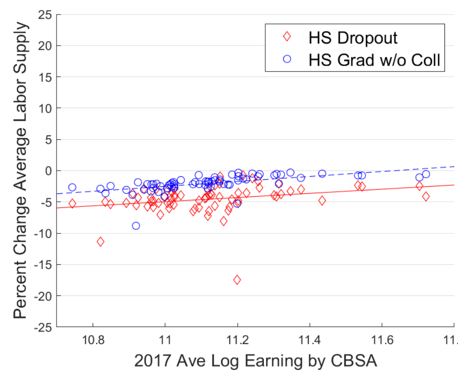
We also simulate a version of our model with elastic labor supply in which locations

Figure 30: Earnings distortion and labor supply distortion of baseline transfer programs and endogenous labor supply

Panel (a) shows counterfactual population relative to lump-sum transfers for the current transfer system for high school dropouts and high school graduates. Panel (b) shows the percent change in average hours worked under the current transfer system relative to lump-sum transfers for high school dropouts and high school graduates. Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households.



(a) Change in Population



(b) Change in Average Hours Worked

Table 9: Spatial distortions and labor supply distortions caused by current transfer programs with endogenous labor supply

The first column gives the percentage difference in the number of households locating in low-earnings and generous-benefit locations compared to the equilibrium with lump-sum transfers in the specification with inelastic labor supply. The second column calculates the percentage change in the number of households when we allow for elastic labor supply. The third column calculates the percentage change in the average labor supply of households compared to the equilibrium with lump-sum transfers. See Table 7 for details.

	Baseline	Endogenous Labor Supply	
	Population	Population	Labor Supply
I. % Δ Low-Earning Locations			
HS Dropout	3.89	3.39	-5.98
HS Grad	0.17	0.51	-3.35
College	0.33	0.04	-0.62
Post College	-0.49	-0.43	0.01
II. % Δ Generous-Benefit Locations			
HS Dropout	3.65	2.88	-6.61
HS Grad	-1.84	-0.56	-1.57
College	-0.84	-0.29	-0.17
Post College	0.22	0.14	0.01
III. Deadweight Loss	4.88	17.76	

are fixed, and therefore the transfer system only leads to a labor supply distortion, but not a geographic distortion.[66] We find a deadweight loss of social transfers equal to 13.1% of transfer spending.

3.6.3 Alternative Skill-Classification

In our baseline specification, we aggregate households into four education groups: high school dropouts, high school graduates, college (including some college), and post-college. In this section, we consider an alternative specification in which we instead aggregate households with some college education and high school graduates into a single education group. Households are thus divided into the following four education groups: high school dropouts, high school graduates and some college, college graduates, and post-college. We classify the former two groups as "unskilled labor" and the latter two as "skilled labor". We re-estimate the model given this alternative classification and recalculate the distortions caused by the current transfer programs. Note that this new specification allows for higher granularity for higher education levels at the cost of lower granularity for lower education levels.

The main results are displayed in Table 10. The effects of the current transfer program on the spatial distribution of high school dropouts are fairly similar to the baseline setting. However, location decisions of the aggregated education group of high school graduates and households with some college education are less distorted towards low-earning locations compared to high school graduates alone in the baseline specification. Therefore, the aggregated group of some college households and high school graduates should be less affected by social transfers compared to the disaggregated group of high school graduates alone. The consequence of this is that we find a lower deadweight loss of 2.83% of transfer spending when some college households and high school graduates are aggregated together. Overall, this alternative specification illustrates the importance of allowing for sufficient heterogeneity at the lower end of the income distribution, given that households with lower income levels are

[66]To fix locations, we set σ, the parameter which dictates the dispersion of the idiosyncratic preference draw, to $100,000$.

Table 10: Spatial distortions caused by current transfer programs with alternative skill classification

The first column calculates the distortions of current transfer programs using the baseline skill classification. The second column calculates the distortions of current transfer programs when households with some college education are aggregated with high school graduates. See Table 7 for details.

	Baseline	Alternative Skill Classification	
I. % Δ Low-Earning Locations			
HS Dropout	3.89	HS Dropout	4.25
HS Grad	0.17	HS Grad +Some College	0.02
College	0.33	College	-0.16
Post College	-0.49	Post College	-0.17
II. % Δ Generous-Benefit Locations			
HS Dropout	3.65	HS Dropout	3.63
HS Grad	-1.84	HS Grad +Some College	-2.12
College	-0.84	College	-0.02
Post College	0.22	Post College	-0.02
III. Deadweight Loss	4.88		2.83

most affected by social transfers.

3.7 Conclusion

In this paper, we combined a spatial equilibrium model with a detailed model of the United States social transfer system to quantify the locational inefficiency caused by these programs. We found that the current transfer program leads to deadweight loss mostly by incentivizing households to locate in cities where they have lower earnings. We also showed that simultaneously harmonizing transfer programs across state and indexing household earnings to local average earnings could reduce the locational inefficiency caused by these programs substantially while still providing mean-tested transfers.

Future work could also utilize this framework to analyze other means-tested programs. Analyzing the distortions caused by Medicaid would be interesting, as the Medicaid schedules are highly progressive and the Medicaid schedule and eligibility varies across states. It would also be interesting to analyze the distortionary effects of these programs in a dynamic setting by using a dynamic spatial equilibrium model, in the spirit of Almagro and Domınguez-Iino (2019), Colas (2019), Greaney (2019), Caliendo et al. (2019), or Giannone et al. (2020). In this setting, it would also make sense to analyze the role of borrowing constraints. We leave these questions for future research.

CHAPTER IV

DISSERTATION CONCLUSION

In this dissertation I consider three problems in applied microeconomics.

In Chapter 1, we considered grade-point averaging, and document two intrinsic features of the data-generating process leading to grade-point averages: mean convergence and combinatorics. We document that due to these two features, intuition about the comparability of students at different GPAs can be challenging. In particular, we demonstrate that the local comparability of students close to any given GPA threshold that is necessary for some common methods of causal inference can be challenged by combinatorics. We first show this in a simple model with two types of students, and then extend this model to a more general case involving continuous student ability.

In Chapter 2, we analyzed the causal effect that violent video games have on violent crime patterns in the US. Policymakers have focused attention on the role that violent video game playing has on violent crime, but we argue that selection into playing violent video games creates difficulty in infering the causal effect of violent games from purely correlational data. In order to overcome this challenge, we developed an empirical strategy that exploits two sources of variation in the ability to play violent video games: the release dates of new violent games, as well as differences in the popularity of video game consoles across space. Since violent games release for a variety of game consoles and other platforms, the prior popularity of those consoles influences the extent to which new games can be taken-up by different communities. We document this fact empirically, and then use this variation in an instrumental variables strategy to isolate the causal effect that violent games have on violent crime. We find that violent video games do not lead to increased rates of violent crime, counter to the popular narrative espoused by policymakers. Moreover, we find some evidence to support the hypothesis of incapacitation: that violent video game releases lead to decreases in violent crime due to changes in time use.

In Chapter 3, we examined variation across space in the US social safety net, documenting that the current social safety net produces two distortions in the location choice of US households. First, we find evidence for an earnings distortion, in which US families are incentivized to live in cities in which earnings are lower, in order to preserve access to transfer programs. Second, we find evidence for a generosity distortion, in which US families are incentivized to live in states with more generous transfer programs. Due to the correlation between state transfer generosity and earnings, these two distortions exist in tension for many US families. We develop a spatial equilibrium model which includes a detailed model of the US tax and transfer system in order to estimate the effects of these distortions. We use this model to estimate the impact on location choice caused by these two distortions, and to quantify the magnitude of the resulting inefficiency. We then use our model to simulate the effects of several counterfactual policy scenarios, including indexing transfer payments to local earnings levels and also standardizing transfer payments across states.

We find that the location distortions caused by the current transfer regime lead to a deadweight loss equal to 4.88% of total transfer payments. Additionally, we find that by jointly targeting both of these distortions, policymakers could decrease the associated deadweight loss by 64%.

A Institutional Details Appendix

In this section we give further details on the eligibility criteria and benefits formulas for SNAP and TANF. We also give more details on how we model these programs.

A.1 SNAP

Eligibility There are three eligibility criteria for SNAP: a gross income test, a net income test, and an assets test. Gross income is the sum of earned and unearned income, including income from other transfer programs, such as TANF. In the context of our model, this includes earnings, I_{dj}, unearned income, Υ_d, and TANF transfers, b^T_{dj}. Gross income as measured for SNAP, GI^F_{dj}, in our model is therefore given by

$$GI^F_{dj} = I_{dj} + \Upsilon_d + b^T_{dj}.$$

A household passes the gross income test if gross income is less than 1.3 times the federal poverty level. Note that the federal poverty level depends on household size and is higher for households in Hawaii and Alaska.

Net income is given by gross income less deductions. There is a deduction for a portion of earned income, a standard deduction, an excess-shelter deduction, and deductions for dependent care, medical expenses, and child support. We assume that all households take the maximum allowable excess-shelter deduction. As we do not model dependent care, medical expenses, or child support, and because these three deductions are not widely taken,[67] we set these last three deductions to 0.

[67] Only 3 percent of SNAP households claim the dependent care deduction, 2 percent claim the child support deduction, and 6% claim the medical expense deduction (on Budget and Priorities, 2017).

Net income as calculated for SNAP, NI^F, is then given by

$$NI^F_{dj} = \max\left\{0, GI^F_{dj} - \text{StandardDeduction}_{dj} - \text{Disregard} \times I_{dj} - \text{ShelterDeduct}_j\right\}.$$

$\text{StandardDeduction}_{dj}$ is a standard deduction. It is indexed by d to reflect that the standard deduction is increasing in family size, and by j to reflect that the standard deduction is larger for households living in Hawaii and Alaska. Disregard is a parameter and is equal to .2. The shelter deduction $\text{ShelterDeduct}_{dj}$ is indexed by j to reflect that the maximum shelter deduction is larger in Alaska and Hawaii. A household passes the net income test if net income is less than the federal poverty level.

The asset test requires that household assets fall below a certain limit. The details of how the asset test is implemented, such as whether vehicles are included in the asset calculation, varies across states. We do not model assets directly and therefore do not include the asset test in our eligibility criteria. We can think of our SNAP accessibility measures, o^F_{dj}, as capturing the probability at which a household of a given demographic group will pass the asset test. Note that our vector of SNAP implementation policies in (12) includes variables describing how assets are calculated and an indicator for whether or not the state relaxes the asset test through broad based categorical eligibility.

Some states also include a three-month time limit for able-bodied adults without dependents. This time limit is not modeled directly and is therefore captured by the SNAP accessibility estimates. The SNAP policy implementation vector includes a dummy for whether the household is an able-bodied adult without dependents and an indicator for whether the state waives this three-month time limit.

Benefits Benefits are calculated as a "maximum allotment" minus a constant times net income. We therefore can write:

$$\tilde{b}^F_{dj} = \text{MaxAllotment}_{dj} - \text{NetIncWeight}^F \times NI^F_{dj},$$

where we index the maximum allotment by d to reflect that the maximum allotment is increasing in household size, and by j to reflect that the maximum allotment is higher in Hawaii and Alaska. NetIncWeight^F is a parameter which is equal to 0.3. We include an F superscript to distinguish between the net income weight for SNAP denoted here, and the net income weight for TANF.

There is a minimum benefit amount for households with one or two members who are eligible for SNAP. As these minimum benefit amounts are very small ($16 per month for households in the continental US), they are ignored here.

A.2 TANF

In what follows, we first describe the general TANF structure that applies in most states. We then describe alternative TANF structures that have been implemented which do not follow this structure.

Eligibility Similar to SNAP, most states have three eligibility criteria for TANF: a gross income test, a net income test, and an assets test. As with SNAP, these tests compare some measure of income or assets against a threshold, which can vary by state and household characteristics. Unlike SNAP, though, households can be subject to two different versions of the gross and net income test: one version used for the initial application for TANF benefits, and one version used to determine continuing eligibility.[68] Both versions of these tests, though, simply compare the pertinent income measure to some threshold. We implement the more restrictive test (i.e., the lower threshold) in each location. This is based on the fact that if a household were to move between states, they would almost always have to re-apply for TANF. Since income is static in our model once households choose location, a family passing the more restrictive test implies also passing the test with the higher threshold. While uncommon, some states have also implemented tests comparing gross and net earnings alone

[68] Generally speaking, households are required to report any substantive change in monthly income which could affect their TANF benefit. As with SNAP, states also have the ability to implement recurring reporting requirements.

to some threshold. Only the gross earnings test on recipients is used in the states included in our model.

We calculate gross income for TANF as the sum of unearned and earned income:

$$GI_{dj}^{T} = I_{dj} + \Upsilon_d.$$

Net income is given by earnings minus deductions, plus unearned income. In most states, there are two deductions to earned income. The first is a deduction given in dollars, which is fixed conditional on family composition. This "dollar deduction" in most states will vary only with the number of adult workers in the household, as most states apply a portion of this deduction twice for families with two adult earners. The second deduction is a percentage of the household's remaining gross earned income after the dollar deduction is applied. This "percentage deduction" is standard across household characteristics. As with SNAP, net income cannot be negative for the purposes of TANF benefit calculation or eligibility testing. Net income can be represented then as:

$$NI_{dj}^{T} = \max\left\{0, (1 - \text{PctDeduction}_j)(I_{dj} - \text{DollarDeduction}_{dj}) + \Upsilon_d\right\}.$$

In some states, the deduction vary in size based on how long a household has received TANF. For instance, some states deduct the entirety of a household's earnings in the first month of TANF receipt and then deduct a fixed portion of earnings for all future months. More rarely, several states decrease the percentage deduction periodically as a household continues to receive TANF. Because our model does not account for time in this way, we use the modal deduction in all cases: that deduction which would apply in the most months of a household's TANF receipt.

In order to pass the asset test, household assets must fall below a certain limit. States vary in how assets of calculated. We do not model assets directly and therefore do not include the asset test in our eligibility criteria. Similar to SNAP, we can think of our TANF accessibility measures, o_{dj}^{T}, as capturing the probability at which a household of a given demographic group will pass the asset test.

States have also implemented work requirements for TANF households. In most cases, each adult parent in a TANF-eligible household must be actively working, actively seeking work, or engaged with a state-facilitated work-training program. These requirements are generally written so as to require parents to work or search for work for some minimum number of hours per week. States also have extensive rules for the number of months that a household may claim TANF. As a baseline, heads-of-household may only receive federal TANF payments for 60 months over a lifetime, by federal statute. However, most states have added additional structure around this 60-month cap. Some states have legislated shorter lifetime limits on TANF receipt. Other states have left the 60-month cap alone, but have implemented rules which allow families to claim TANF only intermittently.[69] On the other hand, some states have chosen to extend TANF benefits to families for more than 60 months using state funds.

To respond to the diverse circumstances that lead a family to be in need, each state has also formalized a large set of exceptions to both work rules and time limits. For instance, most states exempt from work requirements those parents with children under the age of two, and those parents who are physically or mentally dependent, or who care for another dependent adult in the household. A variety of circumstances will lead to the suspension of the 60-month TANF clock. For instance, the Family Violence Option provides each state with the option to stop counting months of TANF use against the 60-month cap in situations involving domestic violence.[70]

Most of the parameters which govern how work rules and time-limits impact a household's

[69] Most commonly, a household may claim TANF benefits for 12 months, but is then ineligible until the household has went without TANF benefits for some period of time.

[70] Specifically, states may suspend the 60-month clock in situations "where compliance with such requirements would make it more difficult for individuals receiving assistance under this part to escape domestic violence or unfairly penalize such individuals who are or have been victimized by such violence, or individuals who are at risk of further domestic violence." 42 U.S.C. § 602(a)(7)(A)(iii).(http://www.ncdsv.org/images/LM_FamilyViolenceOptionStateByStateSummary_updated-7-2004.pdf)

TANF eligibility fall entirely outside the scope of our modeling. As such, we capture the probability that a household would be ineligible for TANF due to work or time rules in our TANF accessibility measures.

Benefits In the standard structure, benefits are calculated as a "standard of need" minus a constant times net income.[71] Benefits cannot exceed a "maximum grant" amount.[72] We therefore can write:

$$\tilde{b}_{dj}^{T} = \min \left\{ \text{MaxGrant}_{dj}, \text{StandardNeed}_{dj} - \text{NetIncWeight}_{dj}^{T} \times NI_{dj}^{T} \right\},$$

where MaxGrant_{dj} gives the maximum grant, StandardNeed_{dj} is the standard of need, and $\text{NetIncWeight}_{dj}^{T}$ gives the rate at which benefits decrease with net income. Note that all parameters are indexed by demographic d and location j to reflect that states may choose different values for these parameters within this general structure.

Exceptions Most state TANF systems follow the above standard for calculating benefits, but there are several states that have adopted alternative TANF benefit calculations that do not fit into the framework above. Note that there are other states that are not included as locations in our model which also differ from this standard TANF structure.

1. **Flat Benefits:** A handful of states have chosen to eliminate the progressive benefit structure above entirely, and instead pay flat benefits to all eligible TANF recipients, regardless of household income. The states represented in our model that have made this change are Wisconsin and Arkansas. In Arkansas, TANF benefits are flat conditional on family size, but benefits do still increase as family size increases. In Wisconsin,

[71] Many states use the term "standard of need," but terminology varies considerably between states. The term "benefit standard" has also been widely adopted. Note that some states refer to the standard of need as a "maximum benefit." This is relevant since other states have a separately codified maximum benefit in addition to the standard of need, as per the formula below.

[72] This maximum grant is set explicitely in some states, such as Delaware. In states with no separately codified maximum grant, the standard of need can be thought of as the maximum grant. This allows us to write the TANF benefit formula for most states using one equation.

every eligible family receives the same TANF payment, which was $608 in 2017. These states have also implemented several alternatives to the payment of traditional TANF benefits, such as state employment and work-training programs, which frequently fall under the authority of the same state agency that administers TANF.[73] Such forms of assistance and subsidized employment fall outside of our model, so we limit our formalization of TANF in these states to the flat benefit payments, since these are most comparable to TANF payments in general.[74]

2. **Less than 100% benefits:** Several states use the standard TANF formula above to calculate a benefit payment, but then pay less than 100% of those benefits. For instance, North Carolina pays only 50% of what the above TANF schedule would indicate.

3. **Treatment of Unearned Income:** Among those states with explicitely coded maximum benefit amounts, some will subtract unearned income from that maximum benefit amount when determining the maximum TANF payment. This matters, for instance, for families with little or no earned income but some unearned income.

4. **Intra-state Standard of Need Differences:** A handful of states have different standards of need depending on the recipient's county of residence. These seem to generally reflect cost of living differences, but are not large in size.

5. **Virginia**: In addition to the common standard of need minus net income formulation, Virginia has also established two distinct maximum grant amounts for TANF benefits, each of which is binding for a distinct set of households. The first is a set of maximum grants for different counties that are independent of household size. The second, Virginia's "standard of assistance" (SOA), does vary with household size. For households with fewer than 5 members, the state-wide maximums are larger than the appropriate

[73]E.g., Wisconsin's Community Service Jobs program.

[74]Specifically, these flat benefits are paid out under the "W-2 Transition" program, which replaced AFDC in Wisconsin.

SOA, meaning that the only binding maximum grant for these families is the SOA. For households with more than 5 members, the state-wide maximums are smaller than the appropriate standard of assistance. However, unearned income is subtracted from the SOA. This means that both maximums must be taken into account for larger assistance units in Virginia. If a household with more than 5 members has no unearned income, the SOA minus unearned income will be larger than the absolute maximum; if the unit has a high level of unearned income, the SOA minus that unearned income may be smaller than the absolute maximum.

6. **Minnesota**: Minnesota's TANF program is actually a combined cash and food aid program, in which households receive a single cash transfer every month, but a portion of that transfer may only be spent on food items.[75] Families receiving TANF in Minnesota are thus ineligible for separate SNAP benefits. The food benefits provided under this combined program are of a similar magnitude to SNAP payments in Minnesota and other states, but are not identical. To account for the fact that households do not receive SNAP when they receive TANF, we subtract TANF accessibility from SNAP accessibility in Minnesota. This solution reflects the notion that, for every portion of the year that a family receives TANF, they are ineligible to receive SNAP.

Outside Option Locations Since we include the nine census divisions as aggregate location options for households, we must also make some simplification regarding the TANF schedule for households locating there. We model TANF in these areas using the program details of the state with the largest remaining population after subtracting the 2017 population figures from each CBSA included in the model. We do the same for our measures of TANF and SNAP accessibility.

[75] This is accomplished using an electronic benefit transfer (EBT) card, as is the case with SNAP.

B Estimation and Simulation Appendix

B.1 Hedonic Rents

In order to generate comparable measures of housing rents across cities, we estimate hedonic regressions of rents on housing characteristics and CBSA fixed effects. This allows us to generate the predicted rent of a house in each city, holding housing characteristics constant.

Specifically, we estimate hedonic regressions of log gross rent on CBSA fixed effects and a vector of housing characteristics using data on renters. The vector of housing characteristics consists of the number of units in the structure containing the household, number of bedrooms, number of total rooms, and household members per room. The rent index is given by the predicted rent from the hedonic regressions using the mean values of the elements of the housing characteristics vector. This gives the predicted value of housing in each CBSA, holding housing characteristics constant.

B.2 Estimation: Production Function

Recall from (9), that the production function in location j is given by

$$F_j\left(L_{e1,j}, L_{e2,j}, L_{e3,j}, L_{e4,j}\right) = A_j[(1-\theta_j)L_{Uj}^{\frac{\varsigma-1}{\varsigma}} + \theta_j L_{Sj}^{\frac{\varsigma-1}{\varsigma}}]^{\frac{\varsigma}{\varsigma-1}},$$

where

$$L_{Uj} = L_{e1,j} + \theta_{Uj} L_{e2,j}$$

and

$$L_{Sj} = L_{e3,j} + \theta_{Sj} L_{e4,j}.$$

The parameters to estimate are the city-specific productivity, A_j; city-specific labor intensities, θ_j, θ_{Uj}, and θ_{Sj}; and the elasticity of substitution between skilled and unskilled labor

ς. We calibrate the elasticity of substitution, $\varsigma = 2$.

First, the wage ratios for narrow education groups within each skill level in city j are given by $\frac{W_{e2,j}}{W_{e1,j}} = \theta_{U1}$ and $\frac{W_{e4,j}}{W_{e3,j}} = \theta_{S1}$. We can therefore back out θ_{S1} and θ_{U1} given estimates of wages for each education level. Next, we can rewrite the wage ratio for households with college ($e = e3$) over high school dropouts ($e = e1$) in city j as

$$\log\left(\frac{W_{e1,j}}{W_{e3,u}}\right) = -\frac{1}{\varsigma}\log\left(\frac{L_{Sj}}{L_{Uj}}\right) + \log\left(\frac{\theta_j}{1-\theta_j}\right),$$

which allows us to solve for the parameter θ_j given data on wages, labor supply and the elasticity of substitution ς. Finally, we can back out A_j in each city as such that the simulated wage level are equal to the wage levels we observe in the data.

B.3 Calibration: Housing Supply

The parameters of the housing supply functions in each city are z_j for each city, and ν_1 and ν_2. We calibrate these parameters using the estimates from Colas and Hutchinson (2021). Specifically, we use estimates of ν_1 and ν_2 from this paper, which estimates housing supply elasticities using the ethnic-enclave instruments for immigrant inflows proposed by Card (2009) to instrument for housing demand. We can therefore write housing demand in city j as

$$H_j = \sum_d N_{dj} h^{\star}_{dj}, \tag{17}$$

where N_{dj} is the total number of households of demographic d living in city j. Given estimates of ν_1, ν_2, local housing rents r_j, and housing demand, we can back out the parameter z_j in each city.

B.4 Calculation of Equivalent Variation

We calculate equivalent variation as the household-specific lump-sum transfer that, given prices implied by the efficient equilibrium with lump-sum taxes and transfers, would provide the same utility level as the counterfactual in question.

More specifically, let $V_i(W, r, \mathbf{T}_i)$ give household i's maximal utility given a set of wages and rents across all locations, and the vector of transfers available to household i in each location, denoted by $\mathbf{T}_i$. Let C denote a counterfactual in question. We write household i's realized utility in counterfactual C as $V_i\left(W^C, r^C, \mathbf{T}_i^C\right)$.

Consider a vector of lump-sum transfers $\mathbf{T}_i^{LS}$ in which household i receives T_i^{LS} in each location. We calculate equivalent variation as the lump-sum transfers such that

$$V_i\left(W^C, r^C, \mathbf{T}_i^C\right) = V_i\left(W^{FB}, r^{FB}, \mathbf{T}_i^{FB} + \mathbf{T}_i^{LS}\right),$$

where FB denotes the efficient counterfactual with demographic-specific lump-sum transfers and taxes.

There is no analytical solution for the equivalent variation T_i^{LS}, because households may change their optimal location choice in response to lump-sum transfers. We therefore calculate the equivalent variation quantitatively, by repeatedly guessing values of the equivalent variation until the household's utility is equal to $V_i\left(W^C, r^C, \mathbf{T}_i^C\right)$.

Total deadweight loss is then given by the equivalent variation T_i^{LS} summed over all households i. Our results display the deadweight loss as a fraction of the total government spending on transfer payments.

B.5 Alternative Parameter Values: Calibration

In our model, the average elasticity of households of a given set $\tilde{I}$ is equal to $\sum_{i \in \tilde{I}} \sum_{j \in J} \frac{1}{\sigma_d} Y_{dj}^{\eta} (1 - P_{ij})$. We choose σ_d such that the average partial equilibrium elasticity of location choice in our model matches that from each of the papers. Below we describe the elasticities targeted from each paper.

In Suarez Serrato and Zidar (2016), the average partial equilibrium elasticity of location choice with respect to net income is equal to $\sum_{j \in J} \frac{1}{\sigma^W} (1 - P_j)$ where σ^W is the dispersion term of their Extreme-Value Type 1 idiosyncratic preference draw. We use their baseline estimate of $\sigma^W = 0.83$. Households are not differentiated by skill or education, so we choose values of σ_d such that the average elasticity over both households with less than college education and those with college education in our model are equal to the elasticity implied by these parameter estimates.

In Diamond (2016), the average partial equilibrium elasticity of location choice for workers of demographic group d is given by $\sum_{i \in I_d} \sum_{j \in J} \frac{1}{\sigma_d^D}(1 - P_{ij})$. where I_d is the set of workers of demographic group d and σ_d^D is the dispersion term of their Extreme-Value Type 1 idiosyncratic preference draw for workers in this group. We use the preferred estimates of $\sigma_d^D = \frac{1}{4.026}$ for non-college workers and $\sigma_d^D = \frac{1}{2.116}$ for college workers.

As argued by Albouy and Stuart (2020), the absolute values of the parameters σ^H and σ^L in the moving cost function of Notowidigdo (2020) are analogous to the dispersion term with Type I Extreme Value preferences. We use the baseline estimates of Notowidigdo (2020) of $\sigma^H = -0.066$ and $\sigma^L = -0.065$ for college and non-college educated households, respectively.

B.6 Elastic Labor Supply: Calibration

Note that log earnings per hours can be written as

$$\log\left(\frac{I_{dj}}{\tilde{\ell}}\right) = \log \ell_d + \log W_{ej}.$$

As in the baseline model, we parameterize $\log \ell_d = \log E_d + \beta_e X_d^{\text{Prod}}$ for each education level e, where again each β_e is a vector of parameters and X_d^{Prod} is a vector of demographic variables indicating the marital status, experience level, and minority status associated with demographic group d. As before, we estimate E_d as the proportion of households of given demographic group who are employed. Using data on employed households, we can then estimate the following equation for each education level via ordinary least squares

$$\log\left(\text{HourlyEarnings}_{ij}\right) = \hat{\beta}_e X_i^{\text{Prod}} + \gamma_{ej} + \varepsilon_i, \tag{18}$$

where HourlyEarnings are hourly earnings and γ_{ej} is our estimate $\log W_{ej}$. To calculate hourly earnings, we take a household's total earnings divided by hours worked in the previous year, which we calculate as weeks worked in the previous year multiplied by usual hours worked. As weeks worked is reported in intervals, we use the midpoint of the reported interval.

Next we calibrate ζ, the parameter which dictates the elasticity of labor supply, and κ_d, the parameter which determines each demographic group's overall disutility of labor. We choose ζ such that the average labor supply elasticity is equal to 0.165, based on the estimates of uncompensated total hours elasticities in the United States from Bargain et al. (2014).[76] We choose κ_d for each demographic group such that the average number of hours worked by households of this demographic group are equal to the national average of this demographic group in the ACS data. Specifically, to jointly calibrate these parameters, we start with a guess of ζ. Given this guess of ζ, we choose the set of κ_d to match the average hours worked by demographic group in the ACS data. We then simulate a 10% increase in wages across demographic groups and locations and calculate the average labor supply elasticity across demographic groups and locations. We repeat this process until this average elasticity is equal to 0.165.

[76] Bargain et al. (2014) estimate separate elasticities by gender and marital status. 0.165 is the simple average across these four groups.

C Results Appendix

C.1 SNAP Generosity Regressions

Table 11 presents our estimates of (12). We regress the fraction of months each household receives SNAP on a vector of demographic controls and the following 6 policy variables: (i) whether the state uses broad-based categorical eligibility, (ii) whether one vehicle can be excluded from asset test, (iii) whether all vehicles can be excluded from the asset test, (iv) whether the state has an online application, (v) how often a household must re-certify their SNAP eligibility, (vi) whether the state has time limit waivers for Able-Bodied Adults without Dependents (interacted with the household in question being less than 60 years old and having no children). We use SNAP Policy Database from October of 2015, the latest date with no missing data on all variables.

Table 11: SNAP take-up regression

	(1)
	SNAP Participation
Broad Based Categorical Eligibility	0.0370**
	(0.0148)
Can Deduct One Vehicle from Assets	0.0541
	(0.0476)
Can Deduct All Vehicles from Assets	0.0559
	(0.0488)
Has Online Application	0.0462
	(0.0300)
Average Time to Recertify	0.00765***
	(0.00200)
ABAWD Waiver	0.0439**
	(0.0186)
Constant	0.482***
	(0.0655)
Observations	12,385
R-squared	0.141
Demographic Controls	YES

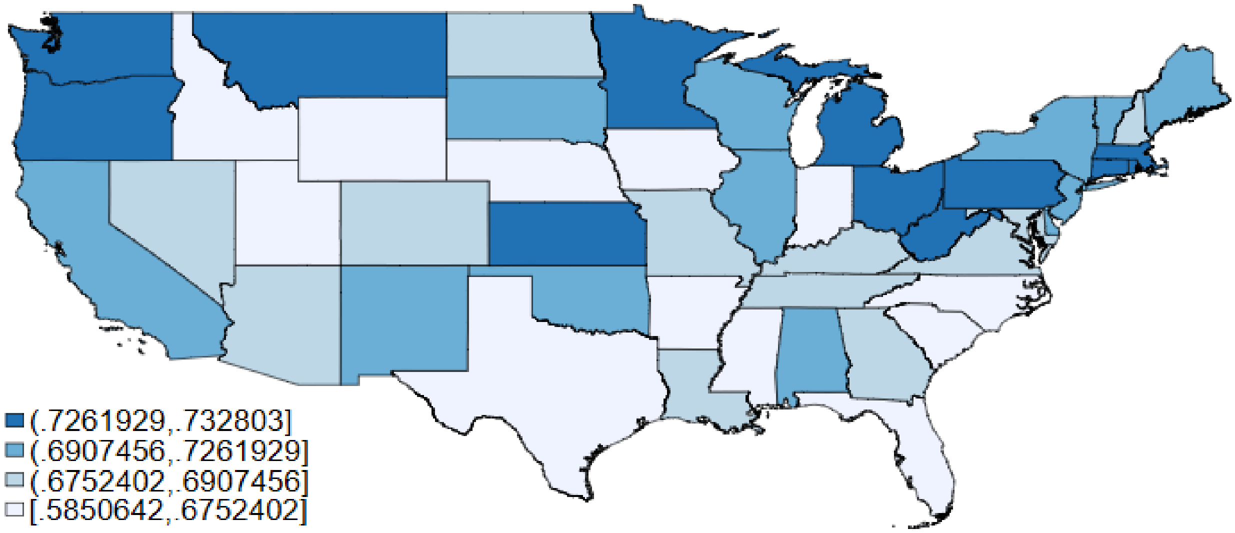
Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

To get a sense of these SNAP accessibility varies across locations, Figure 31 shows the SNAP accessibility of a single household with zero income and two children, and a with a high school dropout, white, non-immigrant head of household. We can see there is are substantial differences across states in these accessibility measures.

Figure 31: Estimated SNAP accessibility across states as measured by take-up rates predicted by state level policy variables

Measures predicted receipt rates of high school dropout with no children and single with 0 income.



C.2 Productivity Regressions

Table 12 presents the estimates of (14') each of the four education groups. Robust standard errors are displayed in parenthesis. Each regression includes a dummy for whether the household is married, has greater than 25 years of potential experience, and a dummy for the household head being a non-minority. All regressions include CBSA fixed effects.

Table 12: Estimates of Equation 14

	(1)	(2)	(3)	(4)
VARIABLES	HS Dropout	HS Grad	College	Post College
Married	0.574***	0.648***	0.789***	0.707***
	(0.00467)	(0.00228)	(0.00150)	(0.00282)
High Experience	0.146***	0.141***	0.0910***	-0.00333
	(0.00495)	(0.00234)	(0.00145)	(0.00284)
Non-minority	0.222***	0.277***	0.255***	0.126***
	(0.00644)	(0.00267)	(0.00174)	(0.00300)
Constant	9.711***	9.933***	10.20***	10.83***
	(0.0325)	(0.0101)	(0.00676)	(0.0124)
Observations	214,116	930,590	2,270,067	618,192
R-squared	0.135	0.185	0.229	0.179
CBSA FE	YES	YES	YES	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

C.3 Estimates of Birth State Premium Function

Tables 13 and 14 shows our estimates of γ_d^{hp} and γ_d^{dist}, the parameters governing the utility of location close to the household head's birth state, for all demographic groups.

Table 13: Estimates of birth state premium parameters for all demographic groups with less than college education

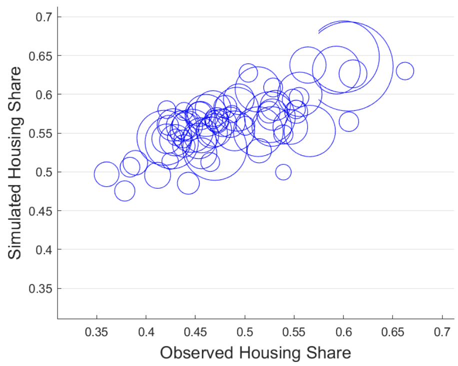
Education	Marital Status	# Children	Experience	Minority	γ_d^{hp}	Standard Error	γ_d^{dist}	Standard Error
Dropout	Single	0	Not Experienced	Minority	3.24	.03	-.93	.03
Dropout	Single	0	Not Experienced	Non-Minority	3.35	.03	-.83	.03
Dropout	Single	1	Not Experienced	Minority	3.47	.05	-1.21	.06
Dropout	Single	1	Not Experienced	Non-Minority	3.48	.04	-.77	.05
Dropout	Single	2	Not Experienced	Minority	3.19	.05	-1.43	.06
Dropout	Single	2	Not Experienced	Non-Minority	3.35	.04	-.91	.05
Dropout	Single	3	Not Experienced	Minority	3.13	.04	-1.53	.05
Dropout	Single	3	Not Experienced	Non-Minority	3.15	.05	-1.29	.06
Dropout	Single	0	Experienced	Minority	3.14	.02	-1.21	.02
Dropout	Single	0	Experienced	Non-Minority	3.13	.01	-.85	.01
Dropout	Single	1	Experienced	Minority	3.2	.03	-1.36	.04
Dropout	Single	1	Experienced	Non-Minority	3.15	.03	-.93	.03
Dropout	Single	2	Experienced	Minority	3.18	.05	-1.2	.05
Dropout	Single	2	Experienced	Non-Minority	3.2	.05	-.86	.06
Dropout	Single	3	Experienced	Minority	3.07	.06	-.97	.06
Dropout	Single	3	Experienced	Non-Minority	3.09	.08	-.99	.09
Dropout	Married	0	Not Experienced	Minority	2.74	.09	-.86	.09
Dropout	Married	0	Not Experienced	Non-Minority	3.13	.06	-.84	.07
Dropout	Married	1	Not Experienced	Minority	2.52	.08	-1.15	.09
Dropout	Married	1	Not Experienced	Non-Minority	3.06	.05	-.93	.06
Dropout	Married	2	Not Experienced	Minority	2.78	.07	-.8	.06
Dropout	Married	2	Not Experienced	Non-Minority	3.03	.04	-1.15	.05
Dropout	Married	3	Not Experienced	Minority	2.43	.05	-1.04	.05
Dropout	Married	3	Not Experienced	Non-Minority	2.69	.03	-1.39	.05
Dropout	Married	0	Experienced	Minority	2.99	.03	-1.21	.03
Dropout	Married	0	Experienced	Non-Minority	3.03	.02	-.93	.02
Dropout	Married	1	Experienced	Minority	3	.04	-.9	.04
Dropout	Married	1	Experienced	Non-Minority	3.16	.03	-.88	.03
Dropout	Married	2	Experienced	Minority	3	.06	-.81	.05
Dropout	Married	2	Experienced	Non-Minority	3.18	.04	-.8	.04
Dropout	Married	3	Experienced	Minority	2.9	.06	-.7	.05
Dropout	Married	3	Experienced	Non-Minority	2.36	.04	-1.45	.06
HS Grad	Single	0	Not Experienced	Minority	3.2	.01	-.91	.01
HS Grad	Single	0	Not Experienced	Non-Minority	3.38	.01	-.7	.01
HS Grad	Single	1	Not Experienced	Minority	3.42	.02	-1.04	.03
HS Grad	Single	1	Not Experienced	Non-Minority	3.39	.02	-.88	.02
HS Grad	Single	2	Not Experienced	Minority	3.37	.02	-1.11	.03
HS Grad	Single	2	Not Experienced	Non-Minority	3.37	.02	-.86	.02
HS Grad	Single	3	Not Experienced	Minority	3.25	.03	-1.31	.03
HS Grad	Single	3	Not Experienced	Non-Minority	3.33	.03	-.87	.03
HS Grad	Single	0	Experienced	Minority	3.18	.01	-1.05	.01
HS Grad	Single	0	Experienced	Non-Minority	3.28	.01	-.78	.01
HS Grad	Single	1	Experienced	Minority	3.2	.02	-1.12	.02
HS Grad	Single	1	Experienced	Non-Minority	3.29	.01	-.78	.01
HS Grad	Single	2	Experienced	Minority	3.26	.03	-1.2	.04
HS Grad	Single	2	Experienced	Non-Minority	3.37	.02	-.79	.03
HS Grad	Single	3	Experienced	Minority	3.1	.05	-.99	.05
HS Grad	Single	3	Experienced	Non-Minority	3.38	.05	-.68	.05
HS Grad	Married	0	Not Experienced	Minority	2.85	.03	-.45	.03
HS Grad	Married	0	Not Experienced	Non-Minority	3.36	.02	-.48	.02
HS Grad	Married	1	Not Experienced	Minority	2.89	.03	-.73	.03
HS Grad	Married	1	Not Experienced	Non-Minority	3.3	.02	-.8	.02
HS Grad	Married	2	Not Experienced	Minority	2.75	.03	-.89	.02
HS Grad	Married	2	Not Experienced	Non-Minority	3.36	.01	-.89	.02
HS Grad	Married	3	Not Experienced	Minority	2.71	.03	-.99	.03
HS Grad	Married	3	Not Experienced	Non-Minority	3.23	.02	-.91	.02
HS Grad	Married	0	Experienced	Minority	3.03	.02	-.93	.02
HS Grad	Married	0	Experienced	Non-Minority	3.2	.01	-.94	.01
HS Grad	Married	1	Experienced	Minority	3.12	.02	-1.02	.02
HS Grad	Married	1	Experienced	Non-Minority	3.31	.01	-.86	.01
HS Grad	Married	2	Experienced	Minority	3	.03	-.9	.03
HS Grad	Married	2	Experienced	Non-Minority	3.36	.01	-.82	.02
HS Grad	Married	3	Experienced	Minority	3	.04	-.9	.04
HS Grad	Married	3	Experienced	Non-Minority	3.24	.02	-.86	.03

Table 14: Estimates of birth state premium parameters for all demographic groups with college and greater education

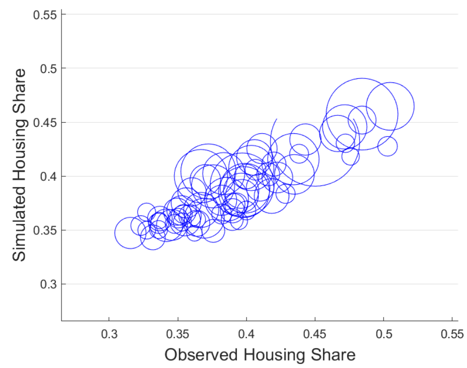
Education	Marital Status	# Children	Experience	Minority	γ_d^{hp}	Standard Error	γ_d^{dist}	Standard Error
College	Single	0	Not Experienced	Minority	2.74	.01	-.6	.01
College	Single	0	Not Experienced	Non-Minority	2.96	0	-.52	0
College	Single	1	Not Experienced	Minority	3.16	.01	-.84	.01
College	Single	1	Not Experienced	Non-Minority	3.19	.01	-.66	.01
College	Single	2	Not Experienced	Minority	3.13	.02	-.94	.02
College	Single	2	Not Experienced	Non-Minority	3.17	.01	-.72	.01
College	Single	3	Not Experienced	Minority	3.13	.02	-1	.02
College	Single	3	Not Experienced	Non-Minority	3.09	.02	-.73	.02
College	Single	0	Experienced	Minority	2.87	.01	-.79	.01
College	Single	0	Experienced	Non-Minority	2.88	0	-.62	0
College	Single	1	Experienced	Minority	2.93	.02	-.91	.02
College	Single	1	Experienced	Non-Minority	2.92	.01	-.66	.01
College	Single	2	Experienced	Minority	2.92	.03	-.92	.03
College	Single	2	Experienced	Non-Minority	2.93	.02	-.66	.02
College	Single	3	Experienced	Minority	2.9	.05	-.86	.04
College	Single	3	Experienced	Non-Minority	2.87	.03	-.65	.03
College	Married	0	Not Experienced	Minority	2.47	.02	-.56	.01
College	Married	0	Not Experienced	Non-Minority	2.91	.01	-.52	.01
College	Married	1	Not Experienced	Minority	2.57	.02	-.67	.01
College	Married	1	Not Experienced	Non-Minority	3.05	.01	-.61	.01
College	Married	2	Not Experienced	Minority	2.6	.01	-.73	.01
College	Married	2	Not Experienced	Non-Minority	3.01	.01	-.73	.01
College	Married	3	Not Experienced	Minority	2.6	.02	-.73	.01
College	Married	3	Not Experienced	Non-Minority	2.91	.01	-.79	.01
College	Married	0	Experienced	Minority	2.53	.01	-.82	.01
College	Married	0	Experienced	Non-Minority	2.76	0	-.74	0
College	Married	1	Experienced	Minority	2.6	.02	-.87	.01
College	Married	1	Experienced	Non-Minority	2.89	.01	-.71	.01
College	Married	2	Experienced	Minority	2.61	.02	-.7	.02
College	Married	2	Experienced	Non-Minority	2.93	.01	-.65	.01
College	Married	3	Experienced	Minority	2.46	.03	-.8	.03
College	Married	3	Experienced	Non-Minority	2.86	.01	-.66	.01
Post-College	Single	0	Not Experienced	Minority	2.18	.02	-.51	.01
Post-College	Single	0	Not Experienced	Non-Minority	2.45	.01	-.45	.01
Post-College	Single	1	Not Experienced	Minority	2.73	.04	-.74	.03
Post-College	Single	1	Not Experienced	Non-Minority	2.73	.02	-.5	.02
Post-College	Single	2	Not Experienced	Minority	2.69	.05	-.78	.04
Post-College	Single	2	Not Experienced	Non-Minority	2.7	.03	-.57	.02
Post-College	Single	3	Not Experienced	Minority	2.93	.07	-.62	.06
Post-College	Single	3	Not Experienced	Non-Minority	2.73	.04	-.57	.04
Post-College	Single	0	Experienced	Minority	2.53	.02	-.67	.02
Post-College	Single	0	Experienced	Non-Minority	2.48	.01	-.54	.01
Post-College	Single	1	Experienced	Minority	2.65	.05	-.74	.04
Post-College	Single	1	Experienced	Non-Minority	2.5	.02	-.53	.02
Post-College	Single	2	Experienced	Minority	2.71	.09	-.66	.07
Post-College	Single	2	Experienced	Non-Minority	2.53	.04	-.46	.04
Post-College	Single	3	Experienced	Minority	2.61	.18	-.81	.16
Post-College	Single	3	Experienced	Non-Minority	2.19	.1	-1	.1
Post-College	Married	0	Not Experienced	Minority	2.01	.03	-.52	.02
Post-College	Married	0	Not Experienced	Non-Minority	2.44	.01	-.49	.01
Post-College	Married	1	Not Experienced	Minority	2.21	.03	-.63	.02
Post-College	Married	1	Not Experienced	Non-Minority	2.55	.01	-.59	.01
Post-College	Married	2	Not Experienced	Minority	2.27	.02	-.53	.02
Post-College	Married	2	Not Experienced	Non-Minority	2.57	.01	-.61	.01
Post-College	Married	3	Not Experienced	Minority	2.14	.03	-.7	.03
Post-College	Married	3	Not Experienced	Non-Minority	2.51	.01	-.72	.01
Post-College	Married	0	Experienced	Minority	2.21	.03	-.67	.02
Post-College	Married	0	Experienced	Non-Minority	2.33	.01	-.64	.01
Post-College	Married	1	Experienced	Minority	2.15	.04	-.78	.03
Post-College	Married	1	Experienced	Non-Minority	2.34	.01	-.62	.01
Post-College	Married	2	Experienced	Minority	1.9	.06	-.77	.04
Post-College	Married	2	Experienced	Non-Minority	2.33	.02	-.61	.02
Post-College	Married	3	Experienced	Minority	1.95	.1	-.75	.08
Post-College	Married	3	Experienced	Non-Minority	2.37	.03	-.56	.03

Figure 32: Model fit: housing cost as a fraction of earnings by location

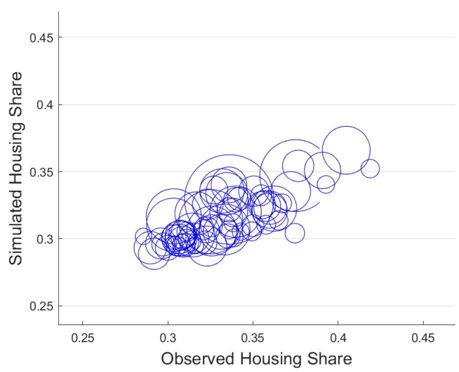
The Y-axis shows the average housing share in the model and the X-axis shows the median housing share in the estimation. Circles are proportional to city size. Each panel shows the fit for one of the four narrow education groups.



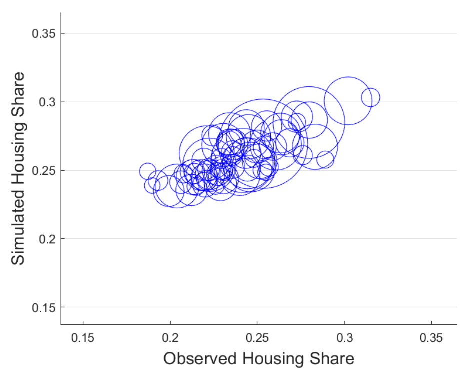
(a) HS Dropouts



(b) HS Graduates



(c) College



(d) Post College

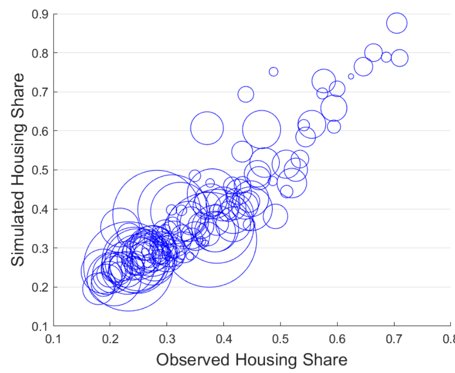
C.4 Additional Model Fit

Figure 32 examines how well the model can replicate housing shares across locations seperately for each of the four education groups. The Y-axis shows the average housing share in the model and the X-axis shows the median housing share in the estimation.

Figure 33 shows housing expenditure for each demographic group in the model and the data. The 128 demographic groups differ in their education level, experience level, race, marital status, and number of children.

Figure 33: Model fit: housing cost as a fraction of earnings by demographic group group

The Y-axis shows the average housing share in the model and the X-axis shows the median housing share in the estimation. Circles are proportional to population of each demographic group.



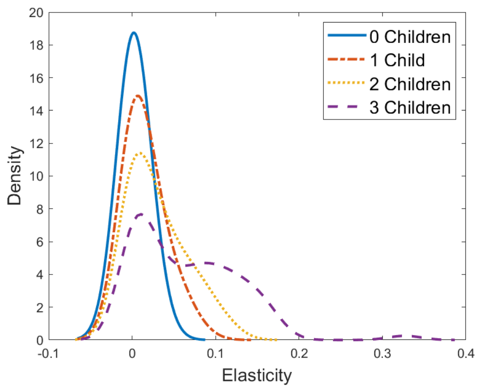
C.5 Simulated Elasticities of Location Choice w.r.t. Social Transfers

To understand what our quantification implies for the responsiveness of household location choice with respect transfers, we now simulate the effect of a ten percent increase in transfer generosity in a given city. Specifically, we simulate an equilibrium in which we increase transfers in a given city j by ten percent, such that households who live in j receive $1.1 \times b_{jd}(\cdot)$. We then calculate the percentage change in location j's population relative to the equilibrium with the baseline transfer function. We calculate the elasticity with respect to transfers as the percent change in population divided by the percent change in benefits (10%). We repeat this exercise for all 79 locations in our model. Note that this represents a general equilibrium elasticity, and therefore includes not only the direct effect of the transfer itself, but also the effect of general equilibrium changes in wages in rents. In fact, for some household who do not receive benefits, the elasticities are negative—reflecting that these general equilibrium price changes effectively crowd them out of a location when transfers become more generous.

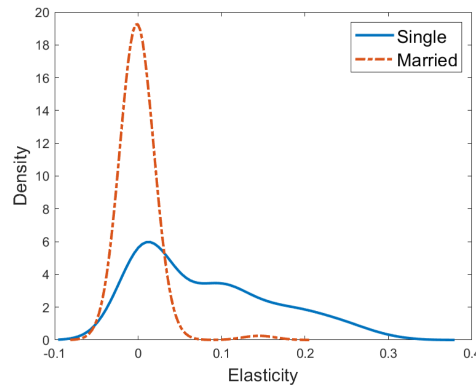
We present the simulated elasticities of selected demographic groups in Figure 34. Panel (a) presents the distribution of elasticities across the 79 simulations for high school dropout households who vary in their number of children. We find that the migration elasticities are strongly increasing in number of children, reflecting that households with children receive larger transfer amounts, all else equal. Panel (b) presents the distribution of elasticities across the 79 simulations for high school dropout households who vary in their marital status. A single, high school dropout household with children has an elasticity of 0.081.

Figure 34: Simulated elasticities of location choice with respect to social transfers for high school dropout households

We simulated increasing social transfers in a given location j by one percent and calculate the percentage increase in location j's population. We repeat the exercise for all 79 locations in the model. The histogram shows the density of elasticities over all 79 simulations. Panel (a) shows the density for high school dropout households who vary in number of children. Panel (b) shows the density for high school dropout households who vary in their marital status.



(a) HS Dropouts by Number of Children



(b) HS Dropouts by Marital Status

C.6 Heterogeneous Effects of Transfer Programs

Table 15 analyzes heterogeneity in the spatial distortions of each transfer system within high school dropout households. Panel I describes the distribution of high school dropout households divided by marital status, the presence of children and minority status. The magnitude of the distortions are highly heterogeneous across demographic groups.

Distortions are larger for minority relative to non-minority households because minority households generally have lower income levels and therefore are more likely to be receive transfers. The distortions are larger for households with children than those without, as transfers are generally more generous for households with children. The number of single dropout households with children in low-earning locations, for example, increases by 6.91%

compared to the counterfactual of lump-sum transfers. There is also evidence of general-equilibrium effects at play: the number of married dropout households without children *decreases* in low-earning cities, as these households are effectively "crowded out" by households more likely to receive transfers.

C.7 Changes in Average Earnings From Current Transfer Program

Table 16 shows the effects of each transfer program programs on average earnings across education groups. Each row shows the percent change in average earnings of a given group across all cities compared to the lump-sum equilibrium. The current program leads to an increase in earnings inequality: earnings of high school dropouts decrease by 0.29% relative to the equilibrium with lump-sum transfers as high school dropouts are more likely to locate in lower-productivity cities.

C.8 Decomposition: TANF vs. SNAP

In the main body we analyzed the distortions caused by the current social transfer programs and considered several alternative programs aimed at minimizing these distortions. In this subsection, we decompose the distortions into those caused by TANF and those caused by SNAP.

The results are displayed in Table 17. As before, Column A shows the distortions caused by the combination of the current TANF and SNAP programs. In the column B, we remove the SNAP program and analyze the distortions caused by TANF alone. In both counterfactuals we provide demographic-specific lump-sum transfers such that total spending on transfers less taxes is the same as in the baseline case. When we remove SNAP, the earnings distortion is reduced substantially but there is still a generosity distortion: high school dropout households are 3.71% more likely to locate in states with generous benefits compared to the efficient equilibrium with lump-sum transfers. However, the efficiency costs are relatively small, as total deadweight loss is only 0.31% of total spending on transfer programs.

The following column (C) instead removes the TANF program and analyzes the distortion caused by SNAP. There is still a substantial earning distortion in this case. However, households are less likely than the baseline case to choose states with generous transfers, reflecting that much of the differences in transfer generosity across locations are driven by TANF. The deadweight loss is only slightly less than the baseline case, at 4.63% of total transfer spending. We conclude that the majority of the deadweight loss from the current transfer programs is caused by SNAP, and only a small proportion is caused by TANF.

C.9 Additional Counterfactual Results

In this Appendix, we display additional results for the counterfactuals from Section 3.5. In particular, while our main counterfactual results focused on differential sorting patterns by education, race, and the presence of children, this Appendix also explores different dimensions of heterogeneity and shows equilibrium price changes.

Baseline Figure 35 shows changes in sorting patterns going from the equilibrium with lump-sum transfers, to the equilibrium given the current SNAP and TANF schedules. Panel A shows sorting patterns of college-educated households compared to post-college-educated households, Panel show shows experienced compared to less-experienced households, and Panel C shows single households compared to married households.

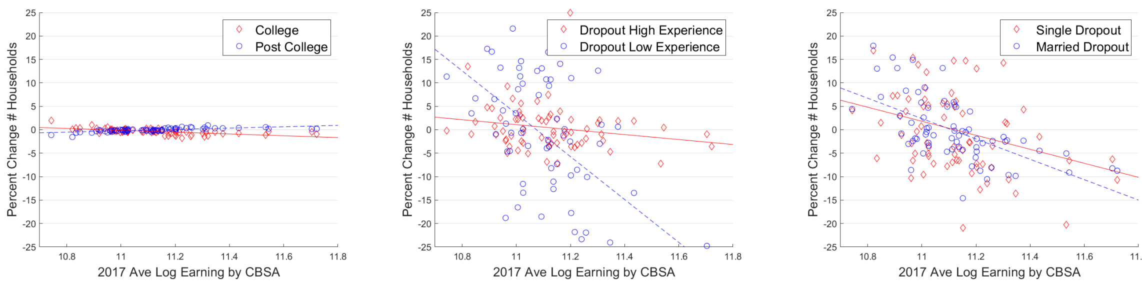
Harmonized Transfer Programs Figures 36 through 37 present additional results for the counterfactual in which we harmonize transfer schedules across all states.

Earnings Index Figures 38 through 40 present additional results for the counterfactual in which we index earnings to local average earnings levels.

Earnings Index and Harmonized Transfer Programs Figures 41 through 43 present additional results for the counterfactual in which we both index earnings to local average

Figure 35: Earnings distortion with baseline transfer programs: Counterfactual population relative to lump-sum transfers for baseline transfer programs across cities

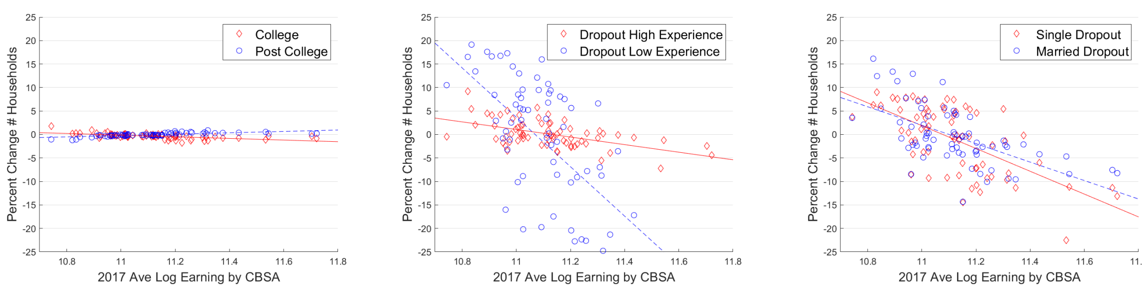
Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel A presents results for college and post-college educated, Panel B presents results for experience and less experienced dropouts households, and Panel C presents results for single dropout households compared to married dropout households.



Panel A Panel B Panel C

Figure 36: Earnings distortion with harmonized transfer programs: Counterfactual population relative to lump-sum transfers for harmonized transfer programs across cities

Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel A presents results for college and post-college educated, Panel B presents results for experience and less experienced dropouts households, and Panel C presents results for single dropout households compared to married dropout households.



Panel A Panel B Panel C

earnings levels and harmonize transfer programs across states. As we can see, the distribution of households across locations are similar to those in the equilibrium with lump-sum transfers.

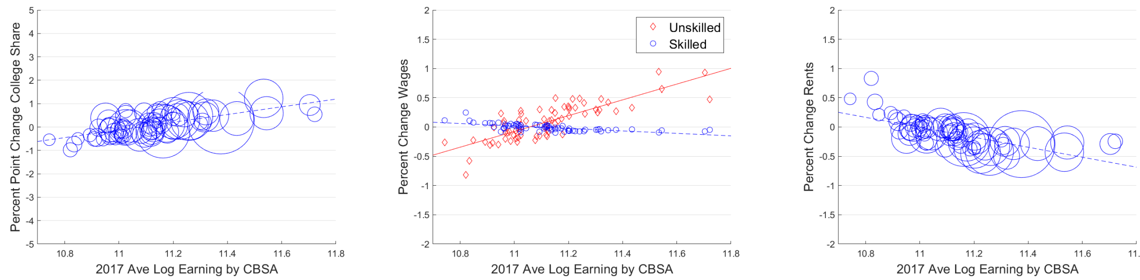
C.10 Cost-of-living Adjustments

In this section, we consider indexing earnings to local cost-of-living, such that benefits are based on real income, rather than nominal income.[77] As prices are generally higher in

[77]This adjustment was suggested by Albouy (2009) to reduce the spatial distortion caused by the federal income tax program.

Figure 37: Earnings distortion with harmonized transfer programs: Counterfactual population relative to lump-sum transfers for harmonized transfer programs across cities, other results

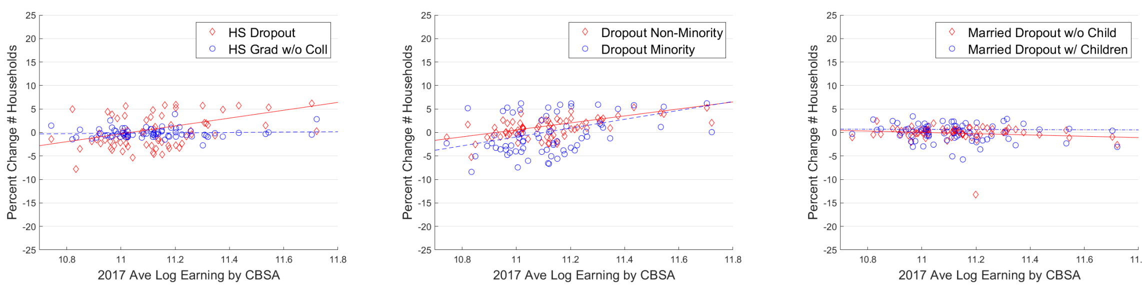
Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel A presents change in college share, Panel B presents change in wages, and Panel C presents changes in rents.



Panel A Panel B Panel C

Figure 38: Earnings distortion with earnings index: Counterfactual population relative to lump-sum transfers with earnings index

Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel (a) presents results for high school dropouts and high school graduates, Panel (b) presents results for non-minority high-school dropout households compared to minority dropout households, and Panel (c) presents results for married high-school dropouts with children and without children.



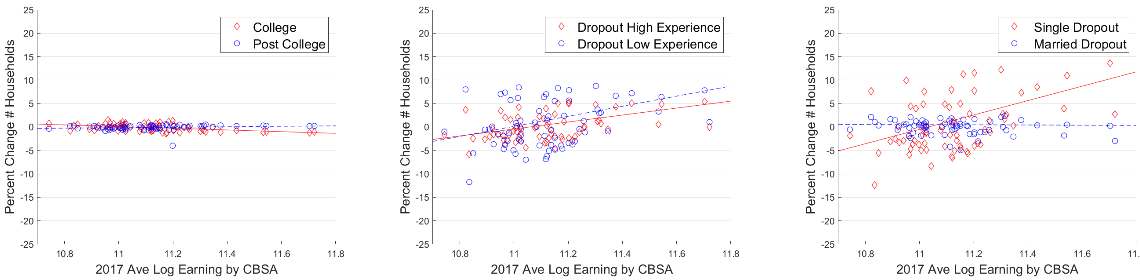
Panel A Panel B Panel C

high wage cities, this increases the amount of transfer households receive if they live in high rent, high wage cities and potentially reduces the distortion towards low-wage cities.

Let $\tilde{I}_j^d = \frac{I_{dj}}{\kappa_j}$ be cost-of-living adjusted household earnings, where κ_j is the price of a market basket in city j. We calculate the cost of the market basket as $\kappa_j = \bar{h}r_j + \bar{c}$, where $\bar{h}$ and $\bar{c}$ are the average quantities of housing and the consumption good consumed across all households. Transfers are calculated as $b_{dj}\left(\kappa \hat{I}_j^d, \Upsilon_d\right)$, where again κ is a parameter we choose to keep total transfers equal to their baseline levels.

Figure 39: Earnings distortion with earnings index: Counterfactual population relative to lump-sum transfers with earnings index, other results

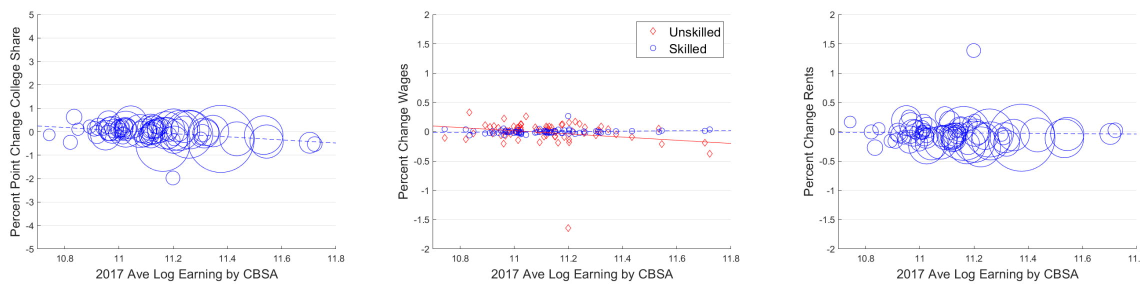
Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel A presents results for college and post-college educated, Panel B presents results for experience and less experienced dropouts households, and Panel C presents results for single dropout households compared to married dropout households.



Panel A　　Panel B　　Panel C

Figure 40: Earnings distortion with earnings index: Counterfactual population relative to lump-sum transfers with earnings index, more results

Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel A presents change in college share, Panel B presents change in wages, and Panel C presents changes in rents.



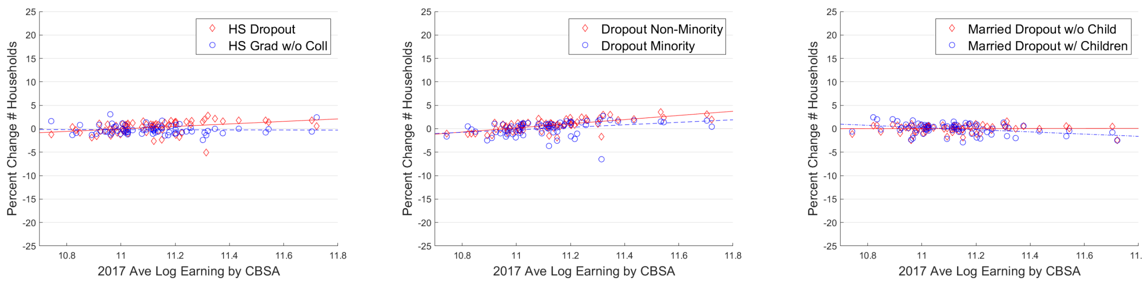
Panel A　　Panel B　　Panel C

The results are displayed in Table 18. The first column shows the distortion caused by the current transfer programs, for reference.[78] The next column shows the results with only the cost-of-living adjustments, and the final column shows the effects of both the cost-of-living adjustment and harmonizing transfer programs across states. Overall, the results are fairly similar to those with the local earnings indexing, as average rents and earnings are strongly correlated in the data. However, the deadweight loss with the cost-of-living adjustment is larger than that with the earnings index.

[78]This is the same information that is included in the "Baseline" column of Table 7.

Figure 41: Earnings distortion with earnings index and harmonized transfer programs: Counterfactual population relative to lump-sum transfers with earnings index and harmonized transfer programs

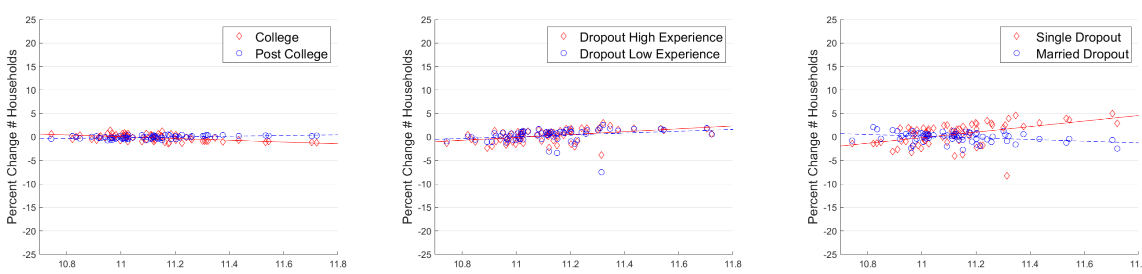
Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel (a) presents results for high school dropouts and high school graduates, Panel (b) presents results for non-minority high-school dropout households compared to minority dropout households, and Panel (c) presents results for married high-school dropouts with children and without children.



Panel A Panel B Panel C

Figure 42: Earnings distortion with earnings index and harmonized transfer programs: Counterfactual population relative to lump-sum transfers with earnings index and harmonized transfer programs, other results

Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel A presents results for college and post-college educated, Panel B presents results for experience and less experienced dropouts households, and Panel C presents results for single dropout households compared to married dropout households.



Panel A Panel B Panel C

Table 15: Spatial distortions for high school dropouts with various demographic characteristics

Panel I gives the percentage difference the number of households locating in low-earnings cities compared to the equilibrium with lump-sum transfers. Low-earning locations are defined as the ten cities with the lowest average income in the data. Panel II gives the percentage difference the number of households locating in generous-benefit locations compared to the equilibrium with lump-sum transfers. Generous-benefit locations are defined as the ten cities which provide the highest transfers to households with zero income. Deadweight loss is measured as a percent of total spending on transfer programs. See text for details on each counterfactual. Transfer spending less tax payments is held constant across counterfactuals.

	A Baseline	B Earnings Adjustments	C Harmonize	D Earn Adj+ Harmonize
I. % Δ Low-Earning Locations (HS Dropouts Only)				
By Race:				
Non-Minority	2.11	-0.35	2.21	-0.28
Minority	4.42	1.93	3.51	0.43
Single:				
No Children	2.87	-0.36	1.99	-1.06
With Children	6.91	4.40	5.94	1.25
Married:				
No Children	-1.44	-0.11	-1.26	0.17
With Children	3.89	0.93	3.35	0.57
II. % Δ Generous-Benefit Locations (HS Dropouts Only)				
By Race:				
Non-Minority	2.29	4.12	-0.21	1.30
Minority	3.75	6.34	-0.07	1.33
Single:				
No Children	2.95	5.62	-1.23	0.91
With Children	14.57	15.66	3.54	2.91
Married:				
No Children	0.03	-0.96	0.77	0.42
With Children	-1.75	2.28	-1.72	0.83

Table 16: Percentage change in average earnings

Each row shows the percent change in average earnings of a given education group across all cities relative to the lump-sum equilibrium.

	A Baseline	B Earnings Adjustments	C Harmonize	D Earn Adj+ Harmonize
HS Dropout	-0.29	0.06	-0.31	0.03
HS Grad	0.02	-0.02	0.03	-0.02
College	-0.04	-0.04	-0.03	-0.04
Post College	0.04	0.04	0.04	0.04

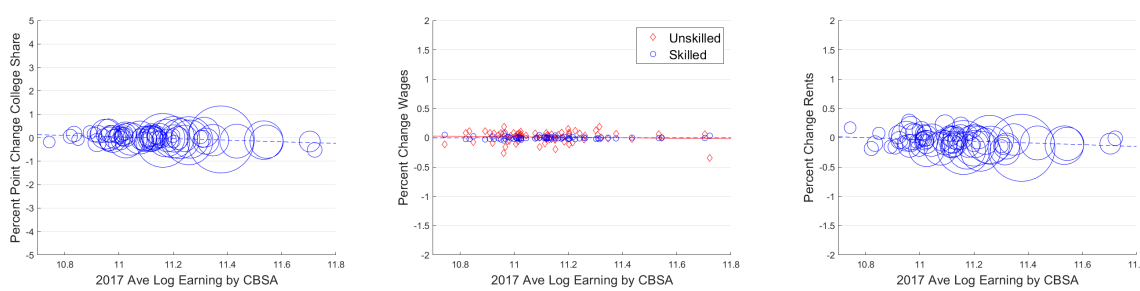
Table 17: Spatial distortions caused by SNAP and by TANF

Transfer spending less tax payments is held constant across counterfactuals. See Table 7 for details.

	A Baseline	B No SNAP	C No TANF
I. % Δ Low-Earning Locations			
HS Dropout	3.89	1.54	2.46
HS Grad	0.17	0.00	0.19
College	0.33	-0.04	0.37
Post College	-0.49	-0.03	-0.47
II. % Δ Generous-Benefit Locations			
HS Dropout	3.65	3.71	0.63
HS Grad	-1.84	-0.17	-1.67
College	-0.84	-0.09	-0.77
Post College	0.22	-0.05	0.25
III. Deadweight Loss	4.88	0.31	4.63

Figure 43: Earnings distortion with earnings index and harmonized transfer programs: Counterfactual population relative to lump-sum transfers with earnings index and harmonized transfer programs, more results

Each dot represents a CBSA. The horizontal axis is the 2017 log mean earnings for all households. Panel A presents change in college share, Panel B presents change in wages, and Panel C presents changes in rents.



Panel A Panel B Panel C

Table 18: Spatial distortions caused by current transfer programs and by alternative programs with cost-of-living adjustments

Panel I gives the percentage difference the number of households locating in low-income cities compared to the equilibrium with lump-sum transfers. Low-earning locations are defined as the ten cities with the lowest average income in the data. Panel II gives the percentage difference the number of households locating in generous-benefit locations compared to the equilibrium with lump-sum transfers. Generous benefit locations are defined as the ten cities which provide the highest transfers to households with zero income. Deadweight loss is measured as a percent of total spending on transfer programs. Column A measures the distortions of the current transfer system. Column B analyzes the case in which household earnings are indexed to local cost of living when calculating social transfers. Column C analyzes the case with both the cost-of-living adjustment and harmonized transfers.

	A Baseline	B COLA Adjustments	C COLA+ Harmonize
I. % Δ Low-Earning Locations			
HS Dropout	3.89	2.49	1.66
HS Grad	0.17	-0.08	-0.06
College	0.33	0.52	0.47
Post College	-0.49	-0.34	-0.39
II. % Δ Generous-Benefit Locations			
HS Dropout	3.65	13.28	7.16
HS Grad	-1.84	-0.81	-0.76
College	-0.84	-0.51	-0.50
Post College	0.22	-0.09	0.09
III. Deadweight Loss	4.88	2.62	1.85

References

Albouy, David, "The unequal geographic burden of federal taxation," *Journal of Political Economy*, 2009, *117* (4), 635–667.

_ **and Bryan A Stuart**, "Urban population and amenities: the neoclassical model of location," *International economic review*, 2020, *61* (1), 127–158.

_ **, Gabriel Ehrlich, and Yingyi Liu**, "Housing demand, cost-of-living inequality, and the affordability crisis," Technical Report, National Bureau of Economic Research 2016.

Alder, Simon, Timo Boppart, and Andreas Muller, "A theory of structural change that can fit the data," *American Economic Journal: Macroeconomics*, 2022, *14* (2), 160–206.

Almagro, Milena and Tomás Domınguez-Iino, "Location sorting and endogenous amenities: Evidence from amsterdam," *NYU, mimeograph*, 2019.

Anderson, Craig A., "An update on the effects of playing violent video games," *Journal of Adolescence*, 2003, *27*, 113–122.

_ **, Akira Sakamoto, Douglas A. Gentile, Nobuko Ihori, Akiko Shibuya, Shintaro Yukawa, Mayumi Naito, and Kumiko Kobayashi**, "Longitudinal effects of violent video games on aggression in Japan and the United States," *Pediatrics*, 2008, *122* (5), 1067–1072.

_ **and Arlin J. Benjamin**, "Violent Video Games: Specific Effects of Violent Content on Aggressive Thoughts and Behavior," *Advances in Experimental Social Psychology*, 2004, *36*, 199–249.

_ **and Brad J. Bushman**, "Human aggression," *Annual Review of Psychology*, 2002, *53* (1), 27–51.

_ **and** _ , "Violent Video Games and Hostile Expectations: A Test of the General Aggression Model," *Personality and Social Psychology Bulletin*, 2002, *28* (12), 1679–1686.

_ **and Karen E. Dill**, "Video Games and Aggressive Thoughts, Feelings, and Behavior in the Laboratory and in Life," *Journal of Personality and Social Psychology*, 2000, *78* (4), 772–790.

Bandura, Albert, *Social Learning Theory*, Prentice-Hall, 1977.

Bargain, Olivier, Kristian Orsini, and Andreas Peichl, "Comparing labor supply elasticities in europe and the united states new results," *Journal of Human Resources*, 2014, *49* (3), 723–838.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?*, W.E. Upjohn Institute, 1991.

Becker, Gary, "A theory of the allocation of time," *Economic Journal*, 1965, *75* (299), 493–517.

Bell, Elizabeth, "Does Free Community College Improve Student Outcomes? Evidence From a Regression Discontinuity Design," *Educational Evaluation and Policy Analysis*, 2021, *43* (2).

Berry, Steven, James Levinsohn, and Ariel Pakes, "Differentiated Product Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Bitler, Marianne and Hilary W Hoynes, "The state of the safety net in the post-welfare reform era," Technical Report, National Bureau of Economic Research 2010.

Blanchard, Olivier Jean and Lawrence F. Katz, "Regional Evolutions," *Brookings Papers on Economic Activity*, 1992, *1.*

Blank, Rebecca M, "The effect of welfare and wage levels on the location decisions of female-headed households," *Journal of Urban Economics*, 1988, *24* (2), 186–211.

Bleemer, Zachary and Aashish Mehta, "Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major," *American Economic Journal: Applied Economics*, 2020.

Blundell, Richard, Monica Costa Dias, Costas Meghir, and Jonathan Shaw, "Female labor supply, human capital, and welfare reform," *Econometrica*, 2016, *84* (5), 1705–1753.

Bonilla, Sade, Thomas S. Dee, and Emily K. Penner, "Ethnic Studies Increases Longer-Run Academic Engagement and Attainment," *Proceedings of the National Academy of Sciences*, 2021, *118* (37).

Boppart, Timo, "Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences," *Econometrica*, 2014, *82* (6), 2167–2196.

Borjas, George J, "Immigration and welfare magnets," *Journal of labor economics*, 1999, *17* (4), 607–637.

Brualdi, Richard A, *Introductory Combinatorics*, Pearson, 2009.

Bureau of Labor Statistics, "Median Weekly Earnings by Education Second Quarter 2020," Technical Report, U.S. Department of Labor 2020.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and labor market dynamics: General equilibrium analysis of the china trade shock," *Econometrica*, 2019, *87* (3), 741–835.

Card, David, "Immigration and Inequality," *American Economic Review: Papers and Proceedings*, 2009, *99* (2), 1–21.

Chan, Marc K, "A dynamic model of welfare reform," *Econometrica*, 2013, *81* (3), 941–1001.

Coen-Pirani, Daniele, "Geographic mobility and redistribution," *International Economic Review*, 2021, *62* (3), 921–952.

Colas, Mark, "Dynamic responses to immigration," *Opportunity and Inclusive Growth Institute*, 2019.

_ **and John M Morehouse**, "The environmental cost of land-use restrictions," *Quantitative Economics*, 2022, *13* (1), 179–223.

_ **and Kevin Hutchinson**, "Heterogeneous workers and federal income taxes in a spatial equilibrium," *American Economic Journal: Economic Policy*, 2021, *13* (2), 100–134.

Cunningham, Scott, Benjamin Engelstätter, and Michael R. Ward, "Violent Video Games and Violent Crime," *Southern Economic Journal*, 2016, *82* (4), 1247–1265.

Currie, Janet and Jeffrey Grogger, "Explaining recent declines in food stamp program participation," *Brookings-Wharton papers on urban affairs*, 2001, pp. 203–244.

Dahl, Gordon and Stefano DellaVigna, "Does Movie Violence Increase Violent Crime?," *The Quarterly Journal of Economics*, 2009, *124* (2), 677–734.

Davidovici-Nora, Myriam and Marc Bourreau, "Les marchés à deux versants dans l'industrie des jeux vidéo (Two-sided markets in the video game industry)," *Réseaux*, 2012, *173-174*, 97–135.

Derdenger, Timothy, "Technological tying and the intensity of price competition: An empirical analysis of the video game industry," *Quantitative Marketing and Economics*, 2014, *12*, 127–165.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *American Economic Review*, 2016, *106* (3), 479–524.

Dickert-Conlin, Stacy, Katie Fitzpatrick, Brian Stacy, and Laura Tiehen, "The Downs and Ups of the SNAP Caseload: What Matters?," *Applied Economic Perspectives and Policy*, 2010.

Duranton, Gilles and Diego Puga, "The economics of urban density," *Journal of economic perspectives*, 2020, *34* (3), 3–26.

Eckert, Fabian, Michael Peters et al., "Spatial structural change," *Unpublished Manuscript*, 2018.

Economic Research Service, U.S. Department of Agriculture, "SNAP Policy Database," Technical Report 2019. data retrieved from U.S. Department of Agriculture (USDA), `https://www.ers.usda.gov/data-products/snap-policy-data-sets/`.

Eeckhout, Jan and Nezih Guner, "Optimal Spatial Taxation: Are Big Cities Too Small?," 2017. CEMFI Working Paper 1705.

Enchautegui, Maria E, "Welfare payments and other economic determinants of female migration," *Journal of Labor Economics*, 1997, *15* (3), 529–554.

Engelstätter, Benjamin and Michael R. Ward, "Strategic timing of entry: evidence from video games," *Journal of Cultural Economics*, 2018, *42*, 1–22.

Fajgelbaum, Pablo D and Cecile Gaubert, "Optimal spatial policies, geography, and sorting," *The Quarterly Journal of Economics*, 2020, *135* (2), 959–1036.

_ , Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar, "State taxes and spatial misallocation," *The Review of Economic Studies*, 2019, *86* (1), 333–376.

Feenberg, Daniel and Elisabeth Coutts, "An introduction to the TAXSIM model," *Journal of Policy Analysis and management*, 1993, *12* (1), 189–194.

Ferguson, Christopher J., "Blazing Angels or Resident Evil? Can Violent Video Games Be a Force for Good?," *Review of General Psychology*, 2007, *14* (2), 68 – 81.

_ , "The Good, The Bad and the Ugly: A Meta-analytic Review of Positive and Negative Effects of Violent Video Games," *Psychiatric Quarterly*, 2007, *78*, 309–316.

Finlay, John and Trevor Williams, "Housing demand, inequality, and spatial sorting," *Available at SSRN 3635035*, 2021.

Frandsen, Brigham, "Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable is Discrete," in Matias Cattaneo, ed., *Regression Discontinuity Designs: Theory and Applications*, Vol. 38, Emerald Publishing Ltd., 2017, pp. 281–315.

Ganong, Peter and Jeffrey B Liebman, "The decline, rebound, and further rise in SNAP enrollment: Disentangling business cycle fluctuations and policy changes," *American Economic Journal: Economic Policy*, 2018, *10* (4), 153–76.

Gelbach, Jonah B, "Migration, the life cycle, and state benefits: How low is the bottom?," *Journal of Political Economy*, 2004, *112* (5), 1091–1130.

Gelman, Andrew and Guido Imbens, "Why High-Order Polynomials Should Not be Used in Regression Discontinuity Designs," *Journal of Business and Economic Statistics*, 2019, *37*.

Gentile, Douglas A., Patrick K. Bender, and Craig A. Anderson, "Violent video game effects on salivary cortisol, arousal, and aggressive thoughts in children," *Computers in Human Behavior*, 2017, *70*, 39–43.

_ , Paul J Lynch, Jennifer Ruh Linder, and David A Walsh, "The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance," *Journal of Adolescence*, 2004, *27* (1), 5–22.

Gerbner, George, Larry Gross, Michael Morgan, and Nancy Signorielli, *Growing up with television: The cultivation perspective*, Lawrence Erlbaum Associates, Inc, 1994.

Giannone, Elisa, Qi Li, Nuno Paixao, and Xinle Pang, "Unpacking Moving," Technical Report, Working Paper 2020.

Glaeser, Edward L. and Joshua D. Gottlieb, "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, 2009, *47* (4), 983–1028.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

Greaney, Brian, "The Distributional Effects of Uneven Regional Growth," Technical Report, Working Paper.[5] 2019.

Greenwood, Jeremy, Nezih Guner, and John A Knowles, "Women on welfare: A macroeconomic analysis," *American Economic Review*, 2000, *90* (2), 383–388.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura, "Child-related transfers, household labour supply, and welfare," *The Review of Economic Studies*, 2020, *87* (5), 2290–2321.

Gyourko, Joseph, Albert Saiz, and Anita A. Summers, "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 2008, *45* (3), 693–729.

Heusmann, L. Roweel, "An information processing model for the development of aggression," *Aggressive Behavior*, 1988, *142*, 13–24.

Imbens, Guido and Karthik Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *The Review of Economic Studies*, 2012, *79*, 933–959.

Jaeger, David A., Theodore J. Joyce, and Robert Kaestner, "Does Reality TV Induce Real Effects? On the Questionable Association Between 16 and Pregnant and Teenage Childbearing," *IZA Working Paper 10317*, 2016.

_ , _ , and _ , "A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?," *Journal of Business & Economic Statistics*, 2020, *38*.

Kabbani, Nader S and Parke E Wilde, "Short recertification periods in the US Food Stamp Program," *Journal of Human Resources*, 2003, pp. 1112–1138.

Kane, Thomas J., "A Quasi-Experimental Estimate of the Impact of Financial Aid On College-Going," Working Paper 9703, National Bureau of Economic Research May 2003.

Kay, John A, "The deadweight loss from a tax system," *Journal of Public Economics*, 1980, *13* (1), 111–119.

Keane, Michael P and Kenneth I Wolpin, "The role of labor and marriage markets, preference heterogeneity, and the welfare system in the life cycle decisions of black, hispanic, and white women," *International Economic Review*, 2010, *51* (3), 851–892.

Kearney, Melissa S. and Phillip B. Levine, "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing," *American Economic Review*, 2015, *105.*

Kennan, John and James R Walker, "Wages, welfare benefits and migration," *Journal of Econometrics*, 2010, *156* (1), 229–238.

Kersten, Riccarda and Tobias Greitemeyer, "Why do habitual violent video game players believe in the cathartic effects of violent video games? A misinterpretation of mood improvement as a reduction in aggressive feelings," *Aggressive Behavior*, 2021, *48* (2), 219–231.

Kline, Patrick and Enrico Moretti, "People, places, and public policy: Some simple welfare economics of local economic development programs," *Annual Review of Economics*, 2014, *6* (1), 629–662.

Kolesár, Michal and Christoph Rothe, "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, August 2018, *108* (8), 2277–2304.

Konečni, Vladimir and Anthony N. Doob, "Catharsis through displacement of Aggression," *Journal of Personality and Social Psychology*, 1972, *23* (3), 379–387.

Lee, David S. and David Card, "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 2008, *142*, 655–674.

Levine, Phillip B and David J Zimmerman, "An empirical analysis of the welfare magnet debate using the NLSY," *Journal of Population Economics*, 1999, *12* (3), 391–409.

Lindo, Jason M., Isaac D. Swenson, and Glen R. Waddell, "Effects of violent media content: Evidence from the rise of the UFC," *Journal of Health Economics*, 2022, *83.*

Low, Hamish, Costas Meghir, Luigi Pistaferri, and Alessandra Voena, "Marriage, labor supply and the dynamics of the social safety net," Technical Report, National Bureau of Economic Research 2018.

Maltz, Michael D. and Joseph Targonski, "A Note on the Use of County-Level UCR Data," *Journal of Quantitative Criminology*, 2002, *18*, 297–318.

Markey, Patrick M., Charlotte N. Markey, and Juliana E. French, "Violent Video Games and Real-World Violence: Rhetoric Versus Data," *Psychology of Popular Media Culture*, 2015, *4* (4), 277–295.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Meyer, Bruce D, *Do the poor move to receive higher welfare benefits?*, Institute for Policy Research, Northwestern University, 1998.

_ , Wallace KC Mok, and James X Sullivan, "The under-reporting of transfers in household surveys: its nature and consequences," Technical Report, National Bureau of Economic Research 2009.

Mohring, Herbert, "Alternative welfare gain and loss measures," *Economic Inquiry*, 1971, *9* (4), 349.

Möller, Ingrid and Barbara Krahé, "Exposure to violent video games and aggression in German adolescents: a longitudinal analysis," *Aggressive Behavior*, 2009, *35* (1), 75–89.

National Archive of Criminal Justice Data, *National Incident-Based Reporting System, 2006: Extract Files* Inter-university Consortium for Political and Social Research 2011-04-19.

_ , *National Incident-Based Reporting System, 2007: Extract Files* Inter-university Consortium for Political and Social Research 2018-09-19.

_ , *National Incident-Based Reporting System, 2008: Extract Files* Inter-university Consortium for Political and Social Research 2018-09-19.

_ , *National Incident-Based Reporting System, 2009: Extract Files* Inter-university Consortium for Political and Social Research 2018-09-27.

_ , *National Incident-Based Reporting System, 2010: Extract Files* Inter-university Consortium for Political and Social Research 2018-10-01.

_ , *National Incident-Based Reporting System, 2011: Extract Files* Inter-university Consortium for Political and Social Research 2018-10-15.

Notowidigdo, Matthew J, "The incidence of local labor demand shocks," *Journal of Labor Economics*, 2020, *38* (3), 000–000.

Olson, Cheryl K., Lawrence A. Kutner, and Dorothy E. Warner, "The Role of Violent Video Game Conent in Adolescent Development," *Journal ofAdolescent Research*, 2008, *23* (1), 55–75.

on Budget, Center and Policy Priorities, "A Quick Guide to SNAP Eligibility and Benefits," Technical Report 2017. ", https://www.cbpp.org/research/a-quickguide-to-snap-eligibility-and-benefits".

Ortigueira, Salvador and Nawid Siassi, "The US tax-transfer system and low-income households: Savings, labor supply, and household formation," *Review of Economic Dynamics*, 2021.

Ost, Ben, Weixiang Pan, and Douglas Webber, "The Returns to College Persistence for Marginal Students: Regression Discontinuity Evidence from University Dismissal Policies," *Journal of Labor Economics*, 2018, *36* (3).

Ottaviano, Gianmarco I.P. and Giovanni Peri, "Rethinking the effect of immigration on wages," *Journal of the European Economic Association*, 2012, *10* (1), 152–197.

Piyapromdee, Suphanit, "The impact of immigration on wages, internal migration and welfare," Technical Report, Working paper 2019.

Ratcliffe, Caroline, Signe-Mary McKernan, and Kenneth Finegold, "Effects of food stamp and TANF policies on food stamp receipt," *Social Service Review*, 2008, *82* (2), 291–334.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Rosen, Sherwin, "Wage-based Indexes of Urban Quality of Life," in Peter Mieszkowski and Mahlon Straszheim, eds., *Current Issues in Urban Economics*, Baltimore, MD: John Hopkins Univ. Press, 1979.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman, "Cognitive hubs and spatial redistribution," Technical Report, National Bureau of Economic Research 2019.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek, "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," 2010. Minneapolis: University of Minnesota.

Rysman, Mark, "The Economics of Two-Sided Markets," *Journal of Economic Perspectives*, 2009, *23*, 125–143.

Serrato, Juan Carlos Suarez and Owen Zidar, "Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms," *American Economic Review*, 2016, *106* (9), 2582–2624.

Suziedelyte, Agne, "Is it only a game? Video games and violence," *Journal of Economic Behavior and Organization*, 2021, *188*, 105–125.

The Urban Institute, "Welfare Rules Database," Technical Report 2019. `https://wrd.urban.org/`.

van der Klaauw, Wilbert, "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach," *International Economic Review*, 2002, *43* (4).

Walker, James R, *Migration among low-income households: Helping the witch doctors reach consensus*, Institute for Research on Poverty, University of Wisconsin–Madison, 1994.

Ward, Michael R., "Video Games and Adolescent Fighting," *The Journal of Law & Economics*, 2010, *53* (3), 611–628.

— , "Cutting class to play video games," *Information Economics and Policy*, 2018, *42*, 11–19.

Williams, Dmitri, "Structure and Competition in the U.S. Home Video Game Industry," *The International Journal on Media Management*, 2002, *4*, 41–54.