Essays in Applied Microeconomics

Dissertation Advisors:
Professor Claudia Goldin
Professor Larry Katz

Author:
Emily Glassberg Sands

Essays in Applied Microeconomics

Abstract

This dissertation contains three chapters. Each applies the tools of applied microeconomics to questions in labor economics, the economics of education, and social economics, respectively. In the first chapter, which is joint work with Amanda Pallais, we present the results of a series of field experiments in an online labor market designed to test whether workers referred to a firm by existing employees perform differently from their non-referred counterparts and, if so, why. We find that referred workers have higher performance and lower turnover than non-referred workers. We demonstrate a large role for selection: referred workers perform better and persist longer even at jobs to which they are not referred at a firm where their referrers do not work. Team production is also important: referred workers are much more productive when working with their own referrer than with someone else's referrer.

In the second chapter, I examine growth in educational attainment over the past thirty years by gender and demographic characteristics. I show that both the rise in educational attainment and the rise of the female advantage in educational attainment occurred relatively similarly across socioeconomic status (SES). I also demonstrate how a prior result showing an increased gradient of education by SES used incorrect sampling weights and is not robust to a more permanent measure of SES.

In the third chapter, which is joint work with Duncan Gilchrist, we exploit the randomness of weather, and the relationship between weather and movie-going, to test for and quantify network externalities in movie consumption. We instrument for opening weekend viewership with unanticipated weather shocks when a movie first opens and estimate the effects of early viewership shocks on later viewership. Given the large set of potential weather measures, we implement Least Absolute Shrinkage and Selection Operator (LASSO) variable selection methods. We find large momentum from network externalities in movie consumption. Neither a supply response nor information dissemination plays a significant role in our estimates. Network externalities appear to be stronger for females than males, and for youth than adults.

List of Tables

List of Figures

Acknowledgments

I have deep gratitude for Claudia Goldin and Larry Katz. Their stimulating engagement, unwavering support, and warm friendship has made this journey particularly rewarding and enjoyable. I am also grateful to Amanda Pallais for her mentorship and coauthorship, and for teaching by example.

I would like to thank Kehinde Ajai, David Autor, Felipe Barrera-Osorio, Raj Chetty, Melissa Dell, David Deming, Itzik Fadlon, Adam Guren, John Friedman, Roland Fryer, Edward Glaeser, Josh Goodman, Rick Hornbeck, Lisa Kahn, John List, Ben Schoefer, Sarah Turner, Marty West, and seminar participants at Berkeley, Booth, Kellogg, Harvard, the New York Federal Reserve, Princeton, RAND, University of British Columbia, University of Chicago, Wharton, and NBER Summer Institute 2013 Labor Studies for their many helpful comments and suggestions on the content in Chapter 1. Thanks also to John Horton and the oDesk Corporation for help running the experiments. Financial support for these experiments from the Lab for Economic Applications and Policy at Harvard is gratefully acknowledged. Additional acknowledgments are due to seminar participants at the Harvard's labor and public economics lunch for their feedback on the content in Chapter 2, and to Natalie Bau, Kevin Garewal, Edward Glaeser, Hank Farber, Peter Ganong, Mike Luca, Jeffrey Miron, Andrei Shleifer, Fanyin Zheng, and seminar participants at the Harvard's labor and public economics lunch for their feedback on the content in Chapter 3.

To Sally Glassberg, the original.

Introduction

The first chapter presents the results of three field experiments in an online labor market designed to determine whether referred workers perform better than non-referred workers and, if so, why. We first hired experienced workers and asked them to refer other workers; we then hired all referred and non-referred applicants that met our basic criteria. We find that referred workers have significantly better observable characteristics than non-referred workers, but perform substantially better and have less turnover even conditional on these characteristics.

We consider three potential explanations for this performance difference: selection (referred workers would perform better than non-referred workers even if they had not been referred), peer influence (referred workers perform better because they believe their performance will affect their referrer's employment outcomes or their relationship with their referrer) and team production (referred workers perform better because they are working with their referrer). We find that team production is important: referred workers are much more productive when working with their own referrer than with someone else's referrer. However, we also find a large role for selection: referred workers perform better and persist longer even at jobs to which they are not referred at a firm where their referrers do not work. A referral contains information about worker quality that is not present in the worker's resume. This information is more valuable (the referral performs better on average) when the referrer is more productive and when the referral is closer with her referrer.

In the second chapter, I examine changes in educational attainment between a cohort of Americans born in the early 1960's and one born two decades later, in the early 1980's. Using detailed and nationally representative longitudinal data from the National Longitudinal Surveys of Youth (NLSYs), I estimate growth by gender and by SES. In the older of these cohorts, females are at parity with males in years of schooling; by the later cohort, females surpass males in nearly all measures of educational attainment. I then pool males and females and examine the growth in educational attainment across socioeconomic status. I show that this female educational advantage rises in approximately equal magnitudes across the SES distribution

I also find that aggregate growth in attainment also occurs quite similarly across SES, pooling males and females. I demonstrate how a prior result showing an increased gradient of education by SES used incorrect sampling weights and is not robust to a more permanent SES measure.

The third chapter explores crowd-following in movie-going. Previous work on momentum in the consumption of entertainment goods like movies has highlighted the role of information dissemination and learning. In this chapter, we explore a different explanation for crowd-following: network externalities in consumption (i.e., a preference for shared experience).

We quantify the effects of network externalities in movie-going by exploiting unanticipated weather shocks when the movie first opens. These weather shocks provide a plausibly exogenous source of variation in opening weekend viewership. Given the large set of potential weather measures, we implement Least Absolute Shrinkage and Selection Operator (LASSO) variable selection methods and instrument for opening weekend viewership with the machine-chosen measures. For 100 additional viewers opening weekend, we estimate that network externalities drive 51 additional viewers the second weekend and 27 the third. By the end of the sixth weekend, network externalities have doubled the effect of the initial shock.

Testing a range of alternative explanations for the estimated momentum, we show that our results are not driven by supply shifts. They are, moreover, independent of both movie quality and the level of ex-ante information about movie quality, suggesting we are also not picking up the effects of social or observational learning. Estimating separately by target demographics, we find that network externalities are significantly larger for women than for men, and for youth than for adults. Finally, we show that most additional viewers are substituting across activities, not simply across movies.

Chapter 1

Why the Referential Treatment? Evidence from Field Experiments on Referrals [1]

1.1 Introduction

A large body of empirical literature has shown that many workers find jobs through networks (e.g., Bewley (1999); Ioannides and Loury (2004); Granovetter (1995)). A consensus estimate is that at least half of jobs are found through informal contacts (Topa (2011)). Theoretical literature (e.g., Calvo-Armengol and Jackson (2004); Montgomery (1991)) suggests that the use of referrals may disadvantage workers without labor market connections; consistent with this, empirical findings show that applicants who are not referred by current employees are much less likely than referred applicants to receive an offer (e.g., Fernandez and Weinberg (1997); Petersen, Saporta, and Seidel (2000); Brown, Setren, and Topa (2012); Burks, Cowgill, Hoffman, and Housman (2013)).

[1] Co-authored with Amanda Pallais

Yet the prevalence of referrals suggests that firms likely benefit from their use. Existing empirical work finds that referred workers have less turnover than non-referred workers (e.g., Brown, Setren, and Topa (2012); Holzer (1987); Simon and Warner (1992); Datcher (1983); Burks, Cowgill, Hoffman, and Housman (2013)). It remains divided, however, on whether referred workers are more productive. A few studies directly compare the performance of referred and non-referred workers working at the same or very similar firms: Castilla (2005) finds that referred workers perform better, Blau (1990) finds that they perform worse, and Burks, Cowgill, Hoffman, and Housman (2013) find that referred workers perform better, but only on a few metrics. Other papers use wages or promotion rates to proxy for the performance of referred and non-referred workers; their findings are similarly mixed (e.g., Dustmann, Glitz, and Schönberg (2011); Simon and Warner (1992); Brown, Setren, and Topa (2012); Pistaferri (1999); Bentolila, Michelacci, and Suarez (2010)).

We undertook three field experiments in an online labor market to identify whether referred workers perform better and have lower turnover than non-referred workers and, if so, why.[2] Our experimental approach affords us a unique opportunity to compare the performance of referred and non-referred workers without the filter of firms' hiring decisions. Most of the existing empirical literature compares the performance of referred and non-referred *hires* (where hires are a subset of all applicants). Differential selection of referred and non-referred workers into employment, however, complicates the interpretation of these results. For example, suppose a firm knows that referred applicants are on average more productive than non-referred applicants; a rational firm would incorporate this information into its hiring decisions such that, in order to be hired, a non-referred applicant would have to look relatively better on other characteristics. Indeed, the exist-

[2]There are other reasons firms might benefit from hiring referrals. For example, referrals might decrease the cost of recruiting or be a perk to existing (referring) workers. In this chapter, we focus on productivity and turnover differences between referred and non-referred workers and abstract away from other potential benefits of hiring referrals.

ing literature finds that even conditional on resume quality, referred workers are more likely than non-referred workers to be hired (e.g., Fernandez and Weinberg (1997); Burks, Cowgill, Hoffman, and Housman (2013)). Amid differential hiring by firms, hired referred workers may not perform any better than hired non-referred workers, even when referrals provide positive information about worker quality.

In our experiments, we hired workers directly so that no differential employer selection could confound our comparisons between referred and non-referred workers. To recruit our experimental sample we first hired experienced workers, asked them to complete a short task unrelated to the experimental tasks, and solicited referrals from those who complied. We then invited referred workers and a random sample of non-referred workers to apply, and hired all applicants who met our basic wage criteria. Our design thus facilitates comparisons between referred and non-referred *applicants*.

In all three experiments (the "individual," "supplemental," and "team" experiments), we find that referred workers performed better than non-referred workers. Referred workers also had less turnover. These facts hold even conditional on resume characteristics; that is, referrals contained information about worker quality that was not contained in workers' resumes. The heart of this chapter (and the motivation behind the three experiments) lies in assessing three potential explanations for these performance and turnover differences.[3] The first explanation, *selection*, says that a referred worker would perform better and stay longer at the firm even if she had not been referred. This may be, for example, because high-ability referrers also have high-ability friends (e.g., Montgomery (1991); Granovetter (1995); Rees (1966)), or because workers have information about their friends and select relatively productive and persistent contacts to refer (e.g., Beaman and Magruder (2012);

[3]There is little empirical evidence on the mechanisms underlying performance and turnover differences between referred and non-referred workers. A prior version of Burks et al. (2013), entitled "The Value of Hiring through Referrals", analyzed potential mechanisms. But this discussion has been mostly removed from the current version, which focuses on observed differences (e.g., in offer rates, turnover, and performance) between referred and non-referred workers, rather than on the mechanisms underlying these differences.

Fernandez, Castilla, and Moore (2000); Rees (1966)).[4]

The second and third explanations, in contrast, emphasize how the productivity and turnover of referred workers may be affected by on-the-job interactions with their referrers. In the second explanation, *peer influence*, a referred worker exerts more effort and stays at the firm longer because she believes her performance and persistence will affect her friend's employment outcomes and/or their relationship. Consistent with this explanation, Heath (2013)'s model suggests that referred workers work hard because if they perform poorly the firm will punish their referrers through lower wages; and Kugler (2003) model assumes referrers directly exert peer pressure on their referrals to perform well.[5]

In the third explanation, *team production*, a referred worker performs better and may enjoy her job more when working directly with her referrer. While this explanation for referrals' positive performance has not been emphasized to the same extent in the economics literature, general research on team production implies that it may be an important benefit of referrals. Bandiera, Barankay, and Rasul (2012)'s model, for example, finds that when working in teams with their friends, workers receive more utility and are less likely to free-ride. Furthermore, Bandiera, Barankay, and Rasul (2005) find that workers are more able to cooperate with their teammates when their teammates are friends; and Costa and Kahn (2003, 2010) find that Civil War soldiers were less likely to desert and were more resilient to job-related stress when more of their unit was from their own birthplace.

Our three experiments are designed to test these three explanations: the individual

---

[4]Our experiments were designed so that referring workers had no information about the job itself at the time they submitted their referrals. Our results thus speak to selection on general observable and unobservable characteristics, and not to selection on match quality (i.e., workers referring friends who would be a good fit for the particular job).

[5]The peer influence explanation is also related to group lending in microfinance wherein a worker's peers may pressure the worker to repay the loan (e.g., Bryan, Karlan, and Zinman (2012)).

experiment distinguishes between selection and peer influence, the supplemental experiment explores selection more deeply, and the team experiment isolates team production. We find that selection is important. On-the-job interactions between referrers and their referrals are also important; while we see only limited evidence of peer influence on the job, we find substantial evidence of team production.

The individual experiment distinguishes between selection and peer influence. All referred and non-referred workers in this experiment performed an individual task: testing an airline flight website by answering a few questions about the flights listed on the site every other day over the course of 12 days.[6] Referrers were simultaneously completing a different task and were randomized, along with their referrals, into one of two treatment groups. Treatment 1 was designed to maximize peer influence. For example, each referrer in this treatment received an update on her referral's performance after each day of work and the referred worker knew her referrer was receiving these updates. We implied to each referrer that her referral's performance and willingness to continue working for us would affect whether the referrer was promoted. Treatment 2, in contrast, was designed to minimize peer influence. Each referred worker in this treatment was told her referrer would never know how she performed, and referrers were told explicitly that they would be judged on their own merits, *not* on the performance of their referrals. At the end of the job, we asked each worker if she would like to continue with the firm.

From the individual experiment, we learn that selection is important. Even non-monitored (Treatment 2) referred workers performed better and stayed longer than non-referred workers. We also find that the referral provides information to employers that could not easily be obtained through observables or initial job performance: the non-monitored referred workers had better observable characteristics than non-referred

[6]The tasks for all three experiments were chosen to be similar to tasks that are common on oDesk. In particular, many jobs on oDesk require visiting websites and answering questions about them.

workers, but they outperformed and outlasted their non-referred counterparts conditional on these.[7] They even outperformed non-referred workers on the last day of the contract, controlling for their performance on all of the prior days. Comparing Treatment 1 (monitored) and Treatment 2 (non-monitored) referred workers, we do not find evidence that peer influence had large effects on workers' productivity or persistence: monitored referred workers performed slightly better, but the difference was not statistically significant. They were, if anything, slightly *less* likely to want to continue working for the firm, perhaps because they disliked being monitored.

Since even referred workers who were not monitored may have faced some subtle peer influence in the individual experiment, we ran a supplemental experiment four months later to isolate the effects of selection. We made job offers to all referred and non-referred workers from a new firm that had no affiliation with the firm from the individual experiment and had no contact with any of the referrers. The task was designed to be credibly different from that in the individual experiment, though it similarly measured diligence over time and willingness to stay on at the firm.

The supplemental experiment provides the strongest evidence that selection is a key driver in the superior performance and persistence of referred workers. Even at a firm to which they had not been referred and at which their referrers did not work, referred workers exhibited substantially higher performance and lower turnover than non-referred workers. The effects are generally large and significant, regardless of whether we restrict attention to workers who accepted our offer of employment.

Our third experiment, the team experiment, isolates the effect of team production. The task was to work with an assigned partner to create a single, shared slogan for a public service announcement (PSA). Each of the two partners was given a different information

---

[7]The online marketplace in which these experiments take place is a unique setting in that we see workers' entire resumes. Because interviews are relatively uncommon and workers and employers do not meet face-to-face, we observe most characteristics that an actual employer would observe when making its hiring decisions.

sheet containing a distinct criterion for the slogan (e.g., be written in all capital letters, be exactly three words long). We asked the partners to use the chat box provided on the site to discuss the task and then to each submit the same slogan, which should have satisfied both criteria. Workers completed three such PSA tasks, each with a different partner. Every referrer participated exactly once in each of three team types: a Type A team (where she was paired with her own referral), a Type B team (where she was paired with someone else's referral), and a Type C team (where she was paired with a non-referred worker). We measured performance in each pairing and, after all three PSA tasks had been completed, also asked which partner(s) they would want to work with again.

We find substantial evidence of team production. Referred workers outperformed non-referred workers even when both types were assigned partners they did not previously know, but referred workers performed substantially better still when paired with their own referrers. They also worked longer on the task when paired with their own referrers and were more likely to report wanting to continue working with their own referrers than with their other partners. These results suggest team production is an important benefit of hiring referrals.

Across experiments we find that referrals provide (positive) information about worker performance on top of workers' observable characteristics, but not all referrals are created equal. Workers referred by high-performers performed particularly well themselves. Part of this can be explained by a tendency among referrers to refer workers with observable characteristics similar to their own: referrers with stronger resumes on average provide referrals with stronger resumes. But even controlling for workers' observable characteristics, those referred by high-performers tended to perform better themselves. We also explore the relationship between the strength of the referrer-referral tie and the performance of the referred worker. At the time of referral, we asked the referrer three questions about her relationship with her referral: how well she knows her referral, how many friends they have in common, and how often they interact. (A caveat is that these were self-reported before the referral had been hired.) We find that when a worker refers someone with whom she is not as close (a weak tie), she tends to refer someone who looks better on paper. Nevertheless, it is the referral who has a stronger tie to her referrer who performs better, even before controlling for observable characteristics.

Finally, we use our experimental data to show that if we had only compared the performance of those referred and non-referred applicants whom employers had actually chosen to hire, we could have obtained misleading results about the information contained in a referral. We first simulate which of our applicants employers would choose to hire if they observed both resumes and referral status, assuming they knew the relationship between resumes and referral status, and performance. Because referred workers substantially outperformed non-referred workers conditional on observable characteristics, employers would hire relatively few non-referred workers, and the non-referred workers hired would be very positively selected on observables. We then compare the actual performance of the workers hired in our simulations. We show that even though the referral contained important information about worker quality, there would be no significant difference in the performance of *hired* referred and non-referred workers.

All three field experiments took place on oDesk, the largest online labor market, with over 2.5 million workers (Horton (2013)) and 35 million hours billed in 2012 (oDesk Corporation (2013)). In this context, we were able to hire workers directly, thus eliminating the concern that employers differentially selected referred and non-referred workers into employment. Equally important, the online labor market allowed us to carefully alter the parameters of the jobs and what workers observed in order to tease out the effects of selection, peer influence, and team production in ways that would be very difficult to execute effectively in brick-and-mortar firms. The trade-off is that our results come from a specific labor market. Before detailing the experiments or their results, we first describe the marketplace (Section 2). We also discuss external validity and the main way we think oDesk differs from more traditional labor markets: oDesk workers are often less strongly tied to employers than are workers in other labor markets. Selection and peer influence may thus be less important on oDesk than in other markets. For example, if oDesk workers are less concerned with remaining in good standing with their employers, they may not refer particularly talented workers (selection) or put pressure on their referrals to work hard (peer influence). Thus, given that we find that selection is important even on oDesk, it seems likely that selection is also quite important in other contexts. However, the fact that we don't find strong evidence of peer influence on oDesk does not eliminate the possibility that it is important in other contexts.

After describing the marketplace and discussing external validity, Section 2 also explains the sample selection for our experiments and provides descriptive statistics about our sample. The three subsequent sections describe the design and results from the individual experiment (Section 3), the supplemental experiment (Section 4), and the team experiment (Section 5). Section 6 analyzes how referrers' performance and the referrer-referral relationship predict referred workers' performance. Section 7 shows that the comparison of referred and non-referred workers' performance could be biased if we only observed the performance of workers employers chose to hire and Section 8 concludes.

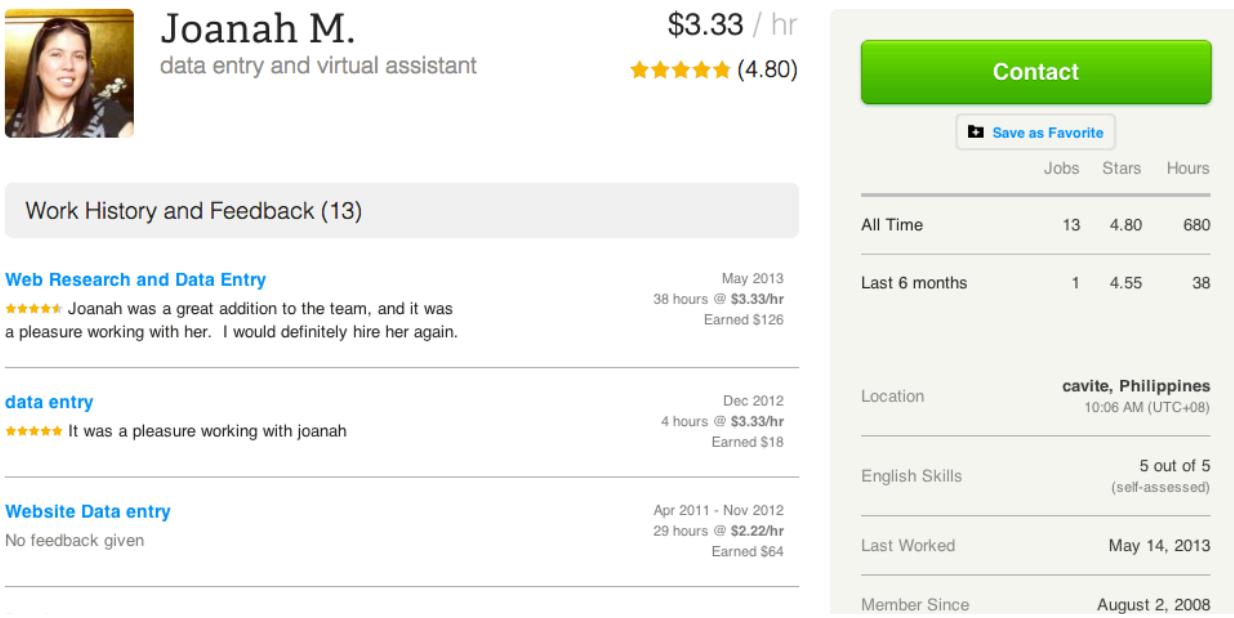
1.2 Experimental Context and Recruitment Design

1.2.1 Online Labor Market

oDesk is an online labor market where employers, mostly from the United States, hire independent contractors from all over the world for tasks ranging from software development to administrative support.

Employers post job listings and can invite workers to apply; workers, meantime, post online resumes and bid on those jobs. Resumes typically include previous oDesk jobs, a one-to-five feedback score from these jobs, and an hourly wage suggested by the worker. Many also list other qualifications, such as degrees held and oDesk tests passed. Figure 1 shows a sample oDesk resume. (This worker was not in our experiment.)

Figure 1.1: *oDesk Profile Example*



Employers decide which of the two oDesk job types they want to offer: hourly jobs or fixed-wage jobs. Hourly jobs are the more popular type of job and the type used throughout our experiments. In these jobs, workers propose an hourly wage when they apply. Employers choose which workers to hire. Hired workers track the time they are working and oDesk monitors that they are actually working during these periods by taking screenshots and analyzing keystroke volume. Workers are then paid their set hourly wage for the time worked regardless of output quality.[8]

Many workers have friends and relatives who also work on oDesk. Though there is at present no explicit referral mechanism on oDesk, employers can solicit referrals from

[8]In contrast, in fixed wage jobs workers and employers agree on a price for the entire job and employers have discretion at the job's end over how much to actually pay.

their current workers and workers can recommend people they know to their employers.

1.2.2 External Validity

Completing these experiments in an online labor market allows us to observe the performance and persistence of workers without the filter of firms' hiring decisions. It also allows us to vary parameters of the jobs workers completed to cleanly identify why referred workers perform better and have less turnover than non-referred workers. The trade-off, however, is that the results of this experiment come from one particular labor market. Perhaps the biggest difference between oDesk and other labor markets is that because oDesk jobs are relatively short and oDesk workers work for many employers, oDesk workers are less tied to any particular employer than are workers in other markets. Prior to our experiment, the average job taken on by the referrers in our sample paid $237 and lasted 81 working hours.

The fact that oDesk workers are less tied to any particular employer could mean that selection and peer influence are weaker here than in other markets. For example, suppose that selection stems from workers choosing their most talented friends to refer (as opposed to homophily in friend networks). If workers are not as tied to employers, they may be less careful to refer only their particularly talented friends. Similarly, if referrers are not as worried about their standing with the employer, they may exert less pressure on their referrals to perform well.

Because we were concerned that peer influence might not be as strong a motivator on oDesk as in other labor markets, we aimed to maximize the effect of peer influence in Treatment 1 of the individual experiment. It is hard to imagine another context wherein a worker's promotion would be so closely tied to her referral's performance. Still, in this experiment, we find limited effects of peer influence. Despite this, our findings do not rule out the fact that peer influence may be important in other markets.

However, the fact that selection might be less important in oDesk than in other labor markets is of limited concern. Given that we find strong evidence that selection is important on oDesk, it seems likely that it is important in other markets as well.

1.2.3 Hiring our Experimental Samples

We hired workers for the individual and team experiments in the same way. We first invited a random sample of oDesk workers who (1) were from the Philippines, (2) listed an hourly wage of $5 or less on their resume, (3) had earned $50 or more on oDesk, and (4) had an average job feedback score of four or higher to apply to our job. We eliminated workers with ratings below four because we only wanted referrals from workers we would actually hire; because most oDesk ratings are very positive, only 16 percent of workers who met our other criteria had ratings below four.[9] We told these workers very little about the task, only that we were hiring "for a variety of ongoing administrative support tasks of varying durations" and that we were looking for "diligent and highly-motivated individuals who are competent in the English language and interested in an ongoing relationship with our firm." We also told them that the position came with the possibility of promotion to managerial roles. We gave workers 48 hours to apply and then hired all workers who applied at an hourly wage of $3 or less.

Original hires were asked to visit our website to initialize the job. The initialization step was intended to give workers some connection to our firm and to weed out the least responsive workers. (We fired the 5 percent of workers who did not initialize.) We then asked the workers who initialized to refer up to three other oDesk workers who were "highly-qualified" and whom they thought would "do a good job and be interested in an ongoing relationship with our firm." On each referral form we included questions about

---

[9]We only included workers from the Philippines because we wanted all workers in the team task to be able to communicate easily and the Philippines is the most common country of residence for low-wage oDesk workers doing these types of jobs.

how well the referrer knew her referral, how often they interacted (remotely and/or or in person), and how many people they knew in common. We also asked if they ever worked in the same room; since referrers might have more easily monitored and/or collaborated with referrals working in the same room, we eliminated from our sample any referral who ever worked in the same room as her referrer.

We invited to our job all referred workers who listed an hourly wage of $5 or less; (all workers who were referred were located in the Philippines). We simultaneously invited to our job a random sample of oDesk workers from the Philippines with hourly wages of $5 or less.[10] We again gave workers 48 hours to apply and then hired all referred and non-referred workers who applied at an hourly wage of $3 or less.[11]

This recruiting process, used for both the individual and team experiments, produced an experimental sample with three types of workers: referred workers, non-referred workers, and "referrers" (i.e., workers who made a successful referral). Workers who did not refer anyone or who referred a worker we did not hire performed a different, shorter task and are not included in any performance results. In the supplemental experiment, we made job offers to all referred and non-referred workers from the individual experiment; no referrers were included.

[10]We eliminated from the pool of both referred and non-referred workers any workers who had already been invited as a potential referrer. We also eliminated from the team experiment anyone who had been invited to the individual experiment. As a result, referred and non-referred workers in the team experiment look worse on observables than do referred and non-referred workers in the individual experiment.

[11]We designed the recruitment process so that when referrers were submitting their referrals, they had no information about our actual tasks. The initialization step, for example, was unrelated to the tasks themselves. From their own invitation to apply and from our request for referrals, referrers did know that we were hiring "for a variety of ongoing administrative support tasks of varying durations" and that we were looking for "diligent and highly-qualified individuals who are competent in the English language and interested in an ongoing relationship with our firm." However, all referred and non-referred workers saw this same description on our job posting. Since referred workers had no private information about the job before referring, in our context there is no scope for selection on match quality.

1.2.4 Descriptive Statistics

Table 1.1 describes the characteristics of three groups of workers: (1) all referred workers, regardless of whether they met our criteria, (2) included referred workers (i.e., referred workers who met our criteria and applied at a wage of $3 or less), and (3) included non-referred workers (i.e., non-referred workers who met our criteria and applied at a wage of $3 or less).[12] Included referred workers had, on average, been on oDesk for about 15 months. Almost two thirds had prior oDesk employment; those who had been employed, had, on average, about nine jobs and earned about $1,350. Non-referred workers had been on oDesk for about four months longer, but were much less likely to have been previously hired; only 28 percent had prior experience. Referred workers also appeared to be more qualified than non-referred workers: they had higher feedback scores from prior employers, had passed more oDesk tests, and had higher self-assessed English abilities. Despite being seemingly more experienced and qualified than non-referred workers, referred workers posted wages on their resumes that were over 20 percent *lower* than those posted by non-referred workers, and they proposed significantly lower wages to our jobs. Referred workers were also much more likely to apply to our job: 68 percent of referred workers applied versus only six percent non-referred workers. (The six percent of non-referred workers who took the time to apply were themselves a very positively self-selected group.) This suggests referrals are a way to identify workers with good

[12]While we hired all referred workers who met our criteria and applied at a wage of $3 or less, only one (randomly-selected) referred worker per referrer was actually included in the team experiment. (The remaining referred workers completed the same tasks, but with different partners. Their performance data is not presented.) Thus, there are some referred workers who applied and met our hiring criteria but are not considered "included referred workers." After all three experiments, we were required by the Harvard IRB to inform all participants about the study and give them the opportunity to remove their data from our study. One worker who was referred but had been excluded from our experiments requested to have his data removed and we have done so. Removing this worker's data only affected the "All Referred Workers" column in Table 1.1 and the "Excluded Referred Workers" column in Appendix Table A.8.

resumes who are interested in the job.[13]

Table 1.1: *Descriptive Statistics*
Individual and Team Experiments

	All Referred Workers	Included Referred Workers	Included Non-Referred Workers	Difference
Has Prior Experience	0.62	0.64	0.28	**
Earnings	$1,481	$864	$353	**
Number of Previous Jobs	7.29	5.79	2.05	**
Has Feedback Score	0.53	0.55	0.24	**
Feedback Score	4.54	4.55	4.26	**
Posted Wage	$2.97	$2.58	$3.29	**
Days Since Joining oDesk	501	462	572	**
Has Portfolio	0.46	0.48	0.23	**
Number of Tests Passed	4.14	4.49	3.19	**
Has English Score	0.98	0.99	0.96	**
English Score	4.67	4.68	4.58	**
Agency Affiliated	0.18	0.12	0.06	**
Number of Degrees	1.27	1.35	1.01	**
Proposed Wage		$2.34	$2.59	**
Observations	1,854	537	274	

Notes: Each statistic in the table presents the mean of the characteristic indicated by the row for the sample indicated by the column. *All Referred Workers* denotes all workers who were referred, while *Included Referred Workers* is the subset of *All Referred Workers* who applied for our job and whom we hired for the individual or team experiment. *Included Non-Referred Workers* are non-referred workers who applied for our job and whom we hired for the individual or team experiment. *English Score* is self-reported English ability on a one-to-five scale, a *portfolio* is where a worker posts prior work, and *agency-affiliated* workers pay a fraction of their earnings to report they are part of a given group of oDesk workers (an agency). ** denotes that the means of the characteristic for *Included Referred Workers* and *Included Non-Referred Workers* are significantly different at the 5% level.

[13] Appendix Table A.1 describes the characteristics of workers whom we asked to refer. It shows that workers who referred someone look somewhat more qualified than those who did not.

1.3 Individual Experiment

1.3.1 Design: Identifying Selection and Peer Influence

The task for the individual experiment was designed to measure referred and non-referred workers' diligence when working alone on a project. We designed our task to emphasize diligence because showing up to work and completing tasks in a timely manner are key determinants of success for low-skilled workers, both in more general labor markets and on oDesk.[14] We also designed the task to measure worker turnover, since decreased turnover is emphasized in the literature as a benefit of hiring referrals (e.g., Brown, Setren, and Topa (2012); Holzer (1987); Datcher (1983)). The treatments were designed to determine the extent to which observed differences in workers' performance and turnover were driven by selection relative to peer influence.

All referred and non-referred workers completed the same task. We told them they would be doing testing for an airline flights website, and asked that they visit the site every other day for twelve days (six visits total), answering the questions on the site each day. For each worker on each day, the site displayed a table with a randomly-generated set of ten flights. Each flight was identified by a flight number and included a departure and arrival city, price, and number of available seats. Just below the flights table were six fill-in-the-blank questions (e.g., the flight number of the cheapest flight). The questions were the same each day, but the correct answers changed with the set of flights shown. Appendix Figure 1 displays a sample flights table followed by the questionnaire.

We told all referred and non-referred workers to complete the task on the assigned

[14]For example, on oDesk, Pallais (2014) finds that employers care more about whether a worker completed a data entry task by the deadline than the worker's accuracy. In more general labor markets, firms respond to absenteeism by having other employees work overtime, reassigning workers from other jobs, and/or hiring temporary workers. These adjustments are all costly and often require manager time. Moreover, the replacement workers may not be as productive as the absent workers (e.g., Herrmann and Rockoff (2010)).

day and asked, but did not require, that they complete each day's task by 11:00 am Philippine Time. We also informed all referred and non-referred workers that we would send performance updates to a manager after each working day reporting (1) whether they submitted a response on the assigned day, (2) whether they submitted a response by 11am on that day, (3) whether they answered all the questions, and (4) the percentage of working days they had met each of these three performance criteria. Appendix Figure 2 shows an example performance report.

Referrers were randomized to Treatments 1 and 2. Each referred worker was assigned the same treatment as her referrer. Appendix Table A.2 shows that the randomization produced balanced samples between Treatment 1 and Treatment 2 within both the referrer and referred worker samples. Out of 26 comparisons between the two treatments groups, only one difference is significant at the 10 percent level.[15]

Treatment 1 was designed to facilitate monitoring of the referred worker by her referrer, while Treatment 2 was designed to minimize peer influence on the referred worker. Referred workers in Treatment 1 were told that their daily performance statistics would be sent to their referrer as well as the manager. Referred workers in Treatment 2, meantime, were explicitly told that their referrer would never see their performance statistics, only the manager would.

Referrers worked on a different task. We wanted to employ them for the duration of their referrals' contracts, and we wanted them to understand the performance metrics we would be sending them about their referrals. Thus, we asked them to answer questions on a website every other day over the same time twelve day period, and we assigned them a soft deadline for submitting on each day. We did not, however, want the referrers to garner insights from their own task with which they could potentially help their referrals, so we had them work on a site that had a different login method, was focused

[15] While there are 28 comparisons in the table, by construction, there is no variation in prior experience or in having a feedback score among referrers.

on consumer products rather than flights, and asked a different set of questions; referrers also had a different soft deadline (2:00 pm Philippine Time).

To strengthen the treatment, we told all referrers before work began that they were being considered for a higher-paying management position. We implied to referrers in Treatment 1 that whether they were promoted would depend on their referrals' performance.[16] Referrers in Treatment 2 were also informed of the management position, but were assured that they would be "judged on their own merits" and that the performance of their referral would in no way influence the promotion decision. As promised, we sent the performance statistics of each referred worker in Treatment 1 to her respective referrer. We also sent referred and non-referred workers' statistics to a manager we hired.

At the end of the task, we invited all referred and non-referred workers to re-apply to continue on the same project. We use this as an (inverse) measure of worker turnover. Re-application updates by worker type and treatment mirrored performance updates. Each referred and non-referred worker was told that the manager would receive an update on whether she accepted our offer to re-apply. Referred workers in Treatment 1 were told this update would also go to their referrers, while referred workers in Treatment 2 were explicitly told their referrers would not see this information. To strengthen the treatment, when we invited referrers in Treatment 1 to apply for the management position, we told them that we had just invited their referrals to continue on with their task and hoped their referrals would accept the invitation. We invited referrers in Treatment 2 to apply for the management position as well, but made no mention at all of their referrals.

[16] All referrers were told that the management position would require being able to identify "high-ability workers interested in an ongoing relationship with our firm." When we told referrers in Treatment 1 about the position, we also said that they would receive daily performance updates on their referrals "because we care about workers' performance."

1.3.2 Performance and Persistence by Worker Type and Treatment

Table 1.2 shows how monitored referred workers (Treatment 1), non-monitored referred workers (Treatment 2), and non-referred workers compare on three measures of performance and our measure of persistence with the firm (the inverse of turnover).[17] Each column presents the results of regressing an outcome on an indicator for being a referred worker in Treatment 1 (a referred worker monitored by her referrer), an indicator for being a non-referred worker, and workers' observable characteristics.[18] The omitted group is referred workers in Treatment 2 (non-monitored referred workers).

[17]Two of the performance metrics are metrics the workers were told the manager would see daily: an indicator for submitting any response on a given day and an indicator for submitting the response by 11:00 am. Workers were also told that the manager would see whether the worker answered all questions, but we exclude this metric from our analysis since 99.8 percent of submissions were complete. The final performance metric is accuracy (non-responses are marked as incorrect).

[18] The observable characteristics included in the regressions are as follows: an indicator for having any oDesk experience, total oDesk earnings, the number of previous oDesk assignments, oDesk feedback score, an indicator for not having a feedback score, the wage listed on the worker's resume, the number of days since joining oDesk, an indicator for having a portfolio, the number of oDesk tests passed, the self-reported English skill level, an indicator for not reporting an English skill level, an indicator for being affiliated with an agency of oDesk workers, and the number of degrees listed on the resume.

Table 1.2: *Performance and Persistence*
Individual Experiment: Base Group is Non-Monitored Referred Workers (Treatment 2)

	A. All Days			
	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred (Treatment 1)	0.021	0.039	0.015	-0.033
	(0.042)	(0.047)	(0.040)	(0.035)
Non-Referred	-0.127**	-0.090*	-0.102**	-0.216**
	(0.046)	(0.048)	(0.042)	(0.043)
Base Group Mean (Treatment 2)	0.757	0.563	0.640	0.953
Controls	Yes	Yes	Yes	Yes
Observations	2,610	2,610	2,610	435
R-squared	0.078	0.063	0.075	0.130
	B. Last Day Only, Controlling for Performance on First Five			
	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred (Treatment 1)	-0.055	0.000	-0.051	-0.047
	(0.044)	(0.053)	(0.039)	(0.035)
Non-Referred	-0.163**	-0.125**	-0.120**	-0.180**
	(0.048)	(0.050)	(0.043)	(0.041)
Daily Performance Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	435	435	435	435
R-squared	0.528	0.405	0.535	0.306

Notes: Each column in each panel presents the results of a separate regression of the dependent variable (indicated by the column) on an indicator for being a referred worker in Treatment 1 and an indicator for being non-referred. In the first three columns in each panel, observations are worker-days and standard errors are clustered at the worker level. Regressions in Panel A include all six days of work while Regressions in Panel B are limited to observations on workers' last day of work. In the final column, observations are workers and Huber-White standard errors are presented. All regressions include the controls for worker characteristics listed in footnote 18. Regressions in Panel B add daily performance controls: each of the first three columns includes controls for the worker's performance as measured by the dependent variable on each of the first five days. The final column includes controls for each of the three performance measures on each of the six days. *, ** denote significance at the 10% and 5% levels, respectively.

Although referred workers had more positive observable characteristics than non-referred workers, the referral still had substantial predictive power even conditional on these characteristics. Referred workers consistently outperformed non-referred workers, even when the referred workers were not monitored. For example, non-monitored (Treatment 2) referred workers submitted responses on just over three-quarters of assigned days; conditional on observable characteristics, non-referred workers were 13 percentage points less likely to submit. Referred workers were also much more likely to want to continue with our firm. While almost all referred workers in Treatment 2 (95 percent) wanted to continue working with us, non-referred workers were 22 percentage points less likely to re-apply to continue the task (conditional on observables).

Across the three performance metrics, the coefficients on the Treatment 1 dummy suggest that peer influence may have led referred workers to perform better still. Anecdotal evidence suggests that referred workers in Treatment 1 were, in fact, monitored by their referrers. Many Treatment 1 referrers replied to our daily performance reports and indicated a strong interest in their referrals' performance. They often apologized when their referrals had not completed the task on the preceding day and/or had not completed it by the soft deadline, and assured us they would encourage their referrals to do better on subsequent days. Nonetheless, all of the performance differences between referred workers in Treatments 1 and 2 appear smaller than the differences between the referred workers in Treatment 2 and non-referred workers, and none is significant. The negative (though again insignificant) coefficient on the Treatment 1 dummy in the final column suggests that referred workers in Treatment 1 were, if anything, slightly less likely to be interested in continuing with the firm, perhaps because they disliked being monitored.

Throughout the chapter, we use the covariates listed in footnote 18 as our main controls. Our results are, however, robust to adding the squares of each of the (non-binary) covariates and the interaction of each pair of covariates to the regressions (what we call "second order controls"). The first two panels of Appendix Table A.3 shows the results of replicating Panel A of Table 1.2, eliminating all the worker controls (Panel A) and adding the second order controls (Panel B). When the second order controls are added, two of the coefficients on the non-referred dummy increase and two decrease, though none changes significantly. Unsurprisingly given random assignment, adding control variables does not affect the estimated differences between monitored and non-monitored referred workers.

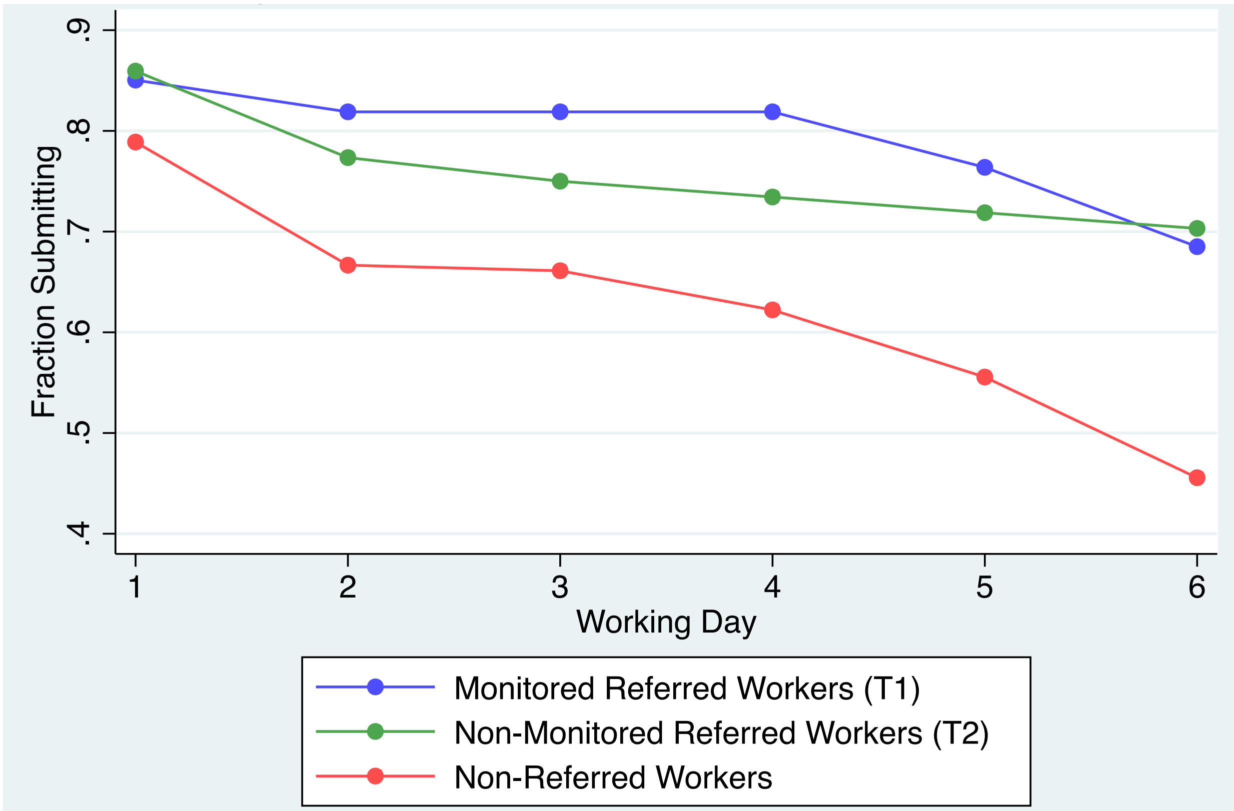
Referrals provide the firm with information about worker quality; firms might also get information about a worker's quality by hiring the worker for some trial duration and observing her performance directly. Longer trials almost certainly provide better information, but at a direct wage cost to the firm. Panel B shows that the referral still has predictive power for worker performance on the last day of the contract, conditional on worker performance on all prior days. Panel B replicates Panel A, limiting the observations to the last day of the contract. Regressions in the first three columns now additionally control for the worker's performance (on the same metric as measured by the dependent variable) on each of the first five days. All differences in performance between referred and non-referred workers remain large and significant.

The referral also provides information about worker persistence at the firm above and beyond the information provided by the worker's performance throughout the full contract. The final column of Panel B adds controls for each of our performance measures (submission, on-time submission, and accuracy) on each of the six days. Even controlling for all our performance measures on all days, referred workers were 18 percentage points more likely than non-referred workers to want to continue on with the firm.[19] Panel C of Appendix Table 1.3 shows that these results are robust to adding the second order controls.

[19]Unreported coefficients in the final column of Panel C show that workers who performed better were more likely to want to continue.

Figure 2 shows worker performance over the course of the experiment by worker type and treatment. Submission rates of referred workers were consistently higher than those of non-referred workers. Both types of workers became less diligent over time, but diligence fell off much more for non-referred workers. Thus, the performance gap between referred and non-referred workers grew with time. The performance gap between Treatment 1 and Treatment 2 referred workers was less stark. On the first day of work, before any performance reports had been sent out, monitored and non-monitored referred workers performed equivalently. The graph suggests that peer influence may have stemmed the drop-off in performance in days two, three, and four among Treatment 1 referred workers, though the differences between monitored and non-monitored referred workers on these days are not significant. By day six, however, monitored referred workers were no more likely than their non-monitored counterparts to submit.

Figure 1.2: *Submission Rates by Day*
Individual Experiment



Taken together, the results of the individual experiment suggest that selection is important. Even when referred workers were not monitored by their referrer, they performed much better than non-referred workers and were more eager to continue on with the firm. The referral, moreover, contained information that was not present on a worker's resume or in her performance on the majority of her contract. In contrast, we do not find robust evidence in favor of peer influence, though we cannot rule out the presence of peer influence, particularly at the beginning of the contract.

1.4 Supplemental Experiment

1.4.1 Design: Isolating Selection

Even though all referrers in Treatment 2 of the individual experiment were assured that they would be judged only on their own merits and all referred workers in this treatment were assured that their referrers would not see their performance statistics, these referred workers may still have been influenced by the presence of their referrers at the firm. They may have, for example, felt grateful for having been referred or faced informal pressures from their referrers, either of which could have affected their performance or persistence in the individual experiment. The supplemental experiment was designed to eliminate any such potential influences.

In the supplemental experiment, we measured the performance and persistence of referred and non-referred workers in a job to which the "referred workers" had not been referred. Four months after the individual experiment, we created a firm with a different name, location, job posting, and writing style from that of the individual experiment. None of the referrers was contacted by this firm. To minimize selection, we sought to hire the maximum possible number of referred and non-referred workers. We made direct job offers to all referred and non-referred workers from the individual experiment and sent three reminders to accept to workers who had not yet responded.

Workers who accepted were given a task that, like the task of the individual experiment, measured individual diligence over time. They were asked to visit the Twitter pages of three successful musicians and to answer a ten-question survey about those accounts every day for five consecutive days (Monday through Friday). We assured workers they needed no prior knowledge of Twitter and explained where to find the relevant information. Most of each day's task involved reporting on the Twitter activity of the artist from the day before. Although we asked workers to complete the task on the correct day, we also accepted retroactive submissions and automatically recorded the time of submissions. Appendix Figure 3 displays the site and questionnaire. After the last assigned day of work, we again invited workers to a continuation of the task and recorded whether they re-applied.

1.4.2 Pure Selection Effects

The majority (61 percent) of workers from the individual experiment accepted our offer and so were hired for the supplemental experiment; referred workers were significantly more likely to accept than non-referred workers. However, regardless of whether we include all referred and non-referred workers in the analysis (counting as not submitting work those who did not accept our employment offer) or instead analyze performance conditional on accepting our job offer, our key results remain unchanged: even working at a job for which they were not referred at a firm with which their referrers were not affiliated, referred workers outperformed non-referred workers and had less turnover.

The three performance metrics in the supplemental experiment mirror those of the individual experiment: an indicator for submitting a response for a given day, an indicator for submitting that day's response on the correct day (analogous to the soft deadline of the individual experiment in that it was requested, but not required), and the fraction of questions answered correctly. Panel A of Table 1.3 shows that unconditional on accepting our employment offer, referred workers were 9 percentage points more likely to submit a response and to submit it on the correct day, even conditional on their observable characteristics. In addition to performing better, referred workers were 12 percentage points more likely than non-referred workers to apply for a continuation of the task.

Table 1.3: *Performance and Persistence in New Firm*
Supplemental Experiment: Base Group is All Referred Workers

	A. All Workers				
	Accepted Job Offer	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.064	-0.090*	-0.088*	-0.028	-0.121**
	(0.055)	(0.048)	(0.047)	(0.026)	(0.056)
Base Group Mean	0.678	0.518	0.499	0.247	0.553
Controls	Yes	Yes	Yes	Yes	Yes
Observations	435	2,175	2,175	2,175	435
R-squared	0.126	0.130	0.132	0.100	0.132

	B. Conditional on Accepting Job Offer			
	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.104*	-0.104*	-0.026	-0.131*
	(0.057)	(0.058)	(0.033)	(0.071)
Base Group Mean	0.763	0.735	0.363	0.815
Controls	Yes	Yes	Yes	Yes
Observations	1,325	1,325	1,325	265
R-squared	0.096	0.098	0.063	0.088

Notes: Each column in each panel reports the results of a separate regression of the dependent variable (indicated by the column) on an indicator for being a non-referred worker. All regressions include the controls for worker characteristics listed in footnote 18. Panel A includes all workers to whom we made employment offers; Panel B includes only workers who accepted these offers. Observations in the first and last columns (Accepted, Re-Application) are workers; observations the middle three columns of regressions (Submission, On-Time Submission, Accuracy) are worker-days. Accepted Job Offer is an indicator for whether the worker accepted our invitation to work for the new firm created for the supplemental experiment. Huber-White standard errors are presented when observations are workers and standard errors are clustered at the worker level when observations are worker-days. *, ** denote significance at the 10% and 5% levels, respectively.

Next, we compare the performance of referred and non-referred workers who accepted our job offer. Appendix Table A.4 provides suggestive evidence of, if anything, differentially positive selection of non-referred workers into accepting. In this table, we regress an indicator for accepting our job offer in the supplemental experiment on an indicator for being non-referred, a metric of performance or persistence in the individual experiment, and the interaction of that metric and the non-referred dummy. Each column uses a different performance or persistence metric from the individual experiment. The large standard errors render many of the results statistically insignificant, but the coefficients suggest that there was positive selection of non-referred workers relative to referred workers into accepting our job.

Panel B of Table 1.3 shows the results of estimating the same regressions as in Panel A, now limited to workers who accepted our job offer. In this sample, referred workers were 10 percentage points more likely to submit work and 13 percentage points more likely to re-apply. However, given the results in Appendix Table A.4, we might expect the conditional results to slightly underestimate the true performance and turnover differences between referred and non-referred workers.

The magnitudes of these estimates are similar to those from the individual experiment. Referred workers performed about as well here as did the non-monitored (Treatment 2) referred workers in the individual experiment (both submitted 76 percent of days, for example). The performance gap between non-referred workers and referred workers is also similar to that of the individual experiment. Appendix Figure 4 shows that, as in the individual experiment, the gap between referred and non-referred workers widened over the contract, while Appendix Table A.5 shows that the results in Table 1.3 are robust to the addition of the second order controls. It also shows results without controls for worker characteristics.

1.5 Team Experiment

1.5.1 Design: Identifying Team Production

The team experiment was designed to measure whether referred workers outperform non-referred workers in a task involving teamwork and, if so, to determine how much of the performance difference between referred and non-referred workers is due to the fact that referred workers may perform particularly well when working with their referrers (team production).

The task involved brainstorming and we encouraged teamwork. Each worker was paired with three successive partners and asked to come up with a slogan for each of three different public service announcements (PSAs).[20] The first PSA was to encourage highway drivers to wear seat belts, the second was to encourage children to practice good dental hygiene, and the third was to encourage college students to get the flu vaccine. For each PSA, we asked the worker to use the chat box we provided on our site to communicate with her partner and to come up with a single slogan that both partners would submit through our online form. Appendix Figure 5 gives an example of what workers saw when they logged in to the team task site.

Though a worker could complete the task without her partner, the task was designed so that the best output necessitated teamwork. Each partner received a different sheet with information relevant to the PSA. For the first PSA, for example, one partner received information on seat belts' efficacy, while the other received information about highway drivers. The justification was that there was a lot of information to process and that by giving the partners different information, each partner would only have to read half as much. We told workers we wanted them to work with a partner to come up with their slogan because brainstorming is often more effective in teams.

[20] As with the prior tasks, we chose this task because there are many jobs on oDesk that ask low-skill workers to come up with advertisements, for example in the form of flyers, posters, and/or slogans.

Each information sheet contained a specific criterion we wanted the slogan to meet as well as a reason for that criterion. In the first round, for example, we told one partner that we wanted the slogan to be only three words long (so as not to distract drivers) and we told the other that we wanted the slogan to be in all capital letters (so drivers would be more responsive to it). In the second round, we told one partner to use an emoticon in their slogan (to make dental hygiene seem more upbeat) and the other to use the name of a real or fictitious person (since kids may respond to role models). In the third, we told each partner we wanted one of four specific words included in the PSA; one partner's word choices emphasized that getting the flu shot would be quick, the other partner's word choices emphasized that flu shots are effective. When giving workers their information sheets, we told them only that the sheets would contain information, not particular criteria for the slogans.

When workers submitted their slogans, we asked them also to answer a "team question": a multiple choice question about the slogan. Each of the three PSA assignments had a different team question (what color sign the PSA should be printed on, what type of lettering the slogan should be written in, and where the PSA should be placed). This question had no correct answer, but partners were instructed to give the same answer.[21]

For comparison with the individual and supplemental experiments, we also collected measures of individual diligence. We monitored whether each worker logged in to the site and whether she submitted work. We also asked each worker an "individual question," the answer to which was in her own information sheet (e.g., the fraction of highway drivers who wear seatbelts). Because workers were instructed that they should complete the task even if they could not make contact with their partner, workers should have logged in, submitted work, answered their individual question correctly, and used the

[21]Because we wanted to measure how effectively workers worked with their partners, we strongly encouraged each worker to complete each PSA. Unlike in the individual experiment, in which we sent workers no reminders about the task, in the team experiment we sent two reminders about each PSA to each worker who had not already submitted work.

criterion from their own information sheet in their slogan regardless of whom they were partnered with.

In the experiment, each referrer completed three different PSA tasks as part of three different types of teams: (1) a Type A team, in which she was paired with her own referral, (2) a Type B team, in which she was paired with someone else's referral, and (3) a Type C team, in which she was paired with a non-referred worker. Each referred worker worked with her own referrer when her referrer was in a Type A team and with someone else's referrer when her referrer was in a Type B team. (When her referrer was in a Type C team, she worked with another referred worker in the same position; results from this treatment are not presented.) Non-referred workers worked with referrers for all three rounds; that is, they were always in Type C teams.

Because we thought worker performance might be correlated not just between partners, but also among partners' partners, we placed workers into blocking groups. By definition, every worker in the blocking group only ever partnered with others in the same blocking group. In all analyses of the team experiment, we cluster standard errors by blocking group.

Each of the 47 blocking groups contained six referrers, their six referred workers, and two non-referred workers. The placement into blocking groups was random, except that a referrer and her referral were always in the same group.[22] Within a blocking group, the ordering of the type of team workers participated in was random. And, within team type, when relevant, workers' assigned partners were also random.

In addition to measuring worker performance, we collected a proxy for worker enjoyment of the partnered task and willingness to continue working with each partner. After the worker submitted her last slogan, we asked, "In case we have more tasks like

[22]As in the individual experiment, we hired all referred and non-referred workers who met the selection criteria. However, only one randomly-selected referral from each referrer and only 94 non-referred workers were included in this experiment.

this in the future, which if any of the partners that you've worked with would you be interested in working with again?" Workers could select all, none, or a subset of their partners.

1.5.2 Performance by Team Type

Panel A of Table 1.4 compares referred and non-referred worker performance across team types on measures that do not rely on teamwork, but may be indicative of individual diligence. These are indicators for logging in to our site to see the given PSA task, submitting work, correctly answering the question about their own individual reading, and including the criteria from their own information sheets in their slogans.[23]

[23]If a worker did not answer the question about her reading, she is marked as not answering it correctly. Similarly, if she did not submit a slogan, she is marked as not including her own criteria in the slogan.

Table 1.4: *Individual Diligence and Team Performance*
Team Experiment: Base Group is Referred Workers Paired with Someone Else's Referrer (Type B)

	A. Individual Diligence			
	Logged in	Submitted	Individual Question Correct	Own Criteria in Slogan
Referred Worker When Working with Own Referrer (Type A)	0.018	0.046**	0.053*	0.004
	(0.017)	(0.017)	(0.029)	(0.034)
Non-Referred Worker When Working with Referrer (Type C)	-0.194**	-0.229**	-0.245**	-0.087
	(0.055)	(0.053)	(0.052)	(0.058)
Base Group Mean (Type B)	0.883	0.837	0.755	0.440
Controls	Yes	Yes	Yes	Yes
Observations	846	846	846	846
R-Squared	0.188	0.187	0.147	0.066

	B. Team Performance			
	Both Submitted	Team Question Matches	Same Slogan	Same Slogan & Both Criteria
Referred Worker and Own Referrer Team (Type A)	0.099**	0.287**	0.372**	0.103**
	(0.023)	(0.028)	(0.032)	(0.024)
Non-Referred Worker and Referrer Team (Type C)	-0.206**	-0.165**	-0.108**	-0.031
	(0.053)	(0.048)	(0.044)	(0.031)
Base Group Mean (Type B)	0.730	0.496	0.337	0.142
Controls	Yes	Yes	Yes	Yes
Observations	846	846	846	846
R-Squared	0.155	0.193	0.213	0.055

Notes: Each column in each panel reports the results of a separate regression of the dependent variable indicated by the column on indicators for being in a Type A team and for being in a Type C team. Observations in Panel A are at the worker-PSA level; only referred and non-referred workers (not referrers) are included. Observations in Panel B are at a team-PSA level. All regressions include the controls for worker characteristics listed in footnote 18. Standard errors are clustered at the blocking group level. *, ** denote significance at the 10% and 5% levels, respectively.

Each outcome is regressed on an indicator for being in a Type A team (a referred worker paired with her own referrer) and an indicator for being in a Type C team (a non-referred worker paired with a referrer). Controls for the referred and non-referred worker's own characteristics are included throughout. The omitted group contains workers in Type B teams (referred workers paired with someone else's referrer). Thus, the coefficient on the Type A dummy indicates how much better referred workers perform when paired with their own referrer than with someone else's referrer; the Type C dummy indicates how much worse non-referred workers perform than referred workers when both are paired with someone else's referrer. Each observation is a partner pair, but in these diligence measures, we consider only referred and non-referred workers. Referrers' performance does not vary significantly across team types.

On average, referred workers performed well on these diligence measures. When paired with someone else's referrer, referred workers logged in 88 percent of the time, submitted work 84 percent of the time, and correctly answered their own question 76 percent of the time. Less than half (44 percent), however, included the criteria from their own information sheet in their slogan.[24] Non-referred workers, meantime, were substantially less diligent than referred workers, even when neither group was working with a partner they previously knew. As compared to referred workers in Type B teams, non-referred workers (all on Type C teams) were approximately 20 percentage points less likely to log in to the site, to submit a slogan, and to correctly answer their individual question, even conditional on their observable characteristics.

The coefficients on Type A teams show that referred workers were five percentage points more likely to submit their work and to correctly answer the question about their own reading when they were paired with their own referrer instead of with someone else's referrer. Given that these are measures of diligence more than teamwork, it could suggest that peer influence may have played a role in the team task. When working

[24]For comparison, referrers on these teams were four percentage points more likely to log in and eight percentage points more likely to use their own criteria in their slogan.

together, referrers may have put more pressure on their referrals to be diligent because in this context, their referrals' performance affected their own.

Panel B compares team performance by team type. Observations are again at the partner-pair level. It shows that, on measures of team performance, teams with a referred worker consistently outperformed those with a non-referred worker, even when the referred worker was working with someone else's referrer. For example, while half of Type B teams answered the team question (e.g., what color sign the PSA should be printed on) the same way, Type C teams were 17 percentage points less likely to do so.

While referred workers did well relative to non-referred workers even when not working with their referrers, they did particularly well when working with their referrers. Referred workers were, for example, substantially (29 percentage points) more likely to answer the team question the same way when working with their own referrers than when paired with referrers they did not know; of the Type A teams that both submitted responses, only 6 percent failed to submit the same response to the team question. The results are consistent across team performance metrics. The third column shows similar results for submitting the same slogan. Only about one-third of Type B teams submitted the same slogan. Type C teams were about a third less likely to do so; Type A teams were more than twice as likely.[25] Appendix Table A.6 replicates this table, both removing the individual controls and by adding the second order controls.

[25]One potential explanation for why referred workers performed better when working with their referrers is that a referred worker and her referrer were, on average, more similar than a randomly-selected referrer and referred worker. We find no evidence, however, that this drives our results. We create indicators for whether both partners were of the same gender (using workers' names and honorifics), whether they lived in the same city, and whether they had previously worked at the same oDesk firm; we also measure the difference between the partners' wages. Partners in Type A teams look more similar on each of these dimensions than do partners in Type B teams. But none of these similarities positively predicts performance, nor does including measures of them in the regressions affect the estimated effect of working with one's own referrer.

1.5.3 Enjoyment and Time Spent by Team Type

One potential motivation for hiring referrals is that workers might enjoy working with their friends and, thus, might be willing to spend more time on the job. Because oDesk requires workers to record the time they spend working on oDesk tasks, we can analyze the amount of time workers spent on each of the three PSAs. Panel A of Table 1.5 shows time spent on the task by team type, first for referrers and then for referred and non-referred workers.

Table 1.5: *Time Spent & Wanting to Partner Again, by Team Type*
Team Experiment: Base Group is Referred Workers Paired with Someone Else's Referrer (Type B)

	A. Time Spent (Minutes)		B. Wants to Partner Again	
	Referrers	Referred & Non-Referred Workers	Referrers	Referred & Non-Referred Workers
Referred Worker Paired with Own Referrer (Type A)	5.922**	5.142**	0.556**	0.451**
	(1.752)	(1.559)	(0.030)	(0.033)
Non-Referred Worker Paired with Referrer (Type C)	1.135	-15.532**	-0.100**	0.009
	(1.445)	(3.121)	(0.041)	(0.060)
Constant	37.482	38.723	0.406	0.477
	(1.291)	(1.492)	(0.029)	(0.031)
Controls	No	No	No	No
Observations	846	846	717	612
R-squared	0.009	0.087	0.338	0.211

Notes: Each column in each panel reports the results of a separate regression of the dependent variable indicated by the panel title on indicators for being in a Type A team and for being in a Type C team. No controls are included. Observations are at a worker-PSA level. The first regression in each panel includes only referrers while the second includes only referred and non-referred workers. Standard errors are clustered at the blocking group level. ** denotes significance at the 5% level.

When partnered with someone they did not know, referrers spent the same amount of time (around 37 minutes) on the task regardless of whether their partner was a non-referred worker or someone else's referral. When working with their own referral, however, they spent an average of six extra minutes on the task. Referred workers also spent significantly (14 percent more) time on the task when working with their referrers.

In general, workers who spent more time on the task performed better. Even controlling for the time workers spent on the task, however, Type A teams performed better than Type B teams. A separate but related reason Type A teams might have performed better is that they communicated via different methods. While each worker always had access to a chat box on the site in which she could chat live with and/or leave messages for her partner, Type A teams may have been advantaged by having other means of communicating. While Type A teams did communicate more both inside and outside of the chat box, this cannot explain their superior performance. Appendix A describes these analyses in more detail.

Panel B of Table 1.5 provides additional insight into how much workers enjoyed their work experience on each type of team. After they had completed all three tasks, workers reported which partner(s) they would be interested in partnering with again; workers could choose as many or as few partners as they wanted.[26] We find that referrers were significantly more likely to want to work again with referred workers they did not know than with non-referred workers.[27] But, referrers were more than twice as likely to want to partner again with their own referral as with someone else's referral. Similarly, referred workers were substantially more likely to want to work again with their own referrer than with someone else's referrer.

[26]Some workers (about 20 percent) did not answer the question, mostly because they did not complete the third PSA task. But for those who answered, we know whether or not they wanted to work again with each of their three partners.

[27]Referrers did not know who, besides their own referrals, had been referred to the firm.

1.6 Predictors of Referral Performance

We find across our experiments that having been referred is a powerful, positive predictor of performance: in each of our three experiments, referred workers substantially outperformed their non-referred counterparts. But not all referrals are created equal. In this section we focus on referred workers and identify predictors of their performance. We look first at a referrer's performance as a predictor of the performance of her referral and then turn to the relationship between the referrer and her referral.

1.6.1 Referrer's Performance

The first column of Table 1.6 shows that a referrer's performance is a strong predictor of her referral's performance. We regress the referred worker's performance in the individual experiment on her referrer's performance in the same experiment. (We use submission as our performance metric here, but using other performance metrics provides similar results.)

What this result does not illuminate is *why* the performance of the referring worker is a good predictor of her referral's performance. For example, referrers and referred workers may perform similarly because on any given day they experience common shocks, or because they have similar underlying ability or diligence. In fact we find that (1) even absent common shocks, workers tend to refer people who perform as they do, (2) part of this seems to be driven by the positive correlation between a worker's own observables and those of the worker she refers, and (3) even absent common shocks and controlling for the referred worker's observable characteristics, the referrer's performance still predicts her referral's performance.

Because it was executed four months after the individual experiment, the supplemental experiment allows us to disentangle the common shocks hypothesis from others. (This assumes that common shocks do not persist for four months.) In the second column of

Table 1.6, we regress the referred worker's performance in the supplemental experiment on her referrer's performance in the individual experiment four months earlier. Even absent common shocks, the referrer's performance remains a powerful predictor of the referral's performance. In fact, knowing the performance of a worker's referrer four months prior leads to almost two-thirds as much updating as knowing her own performance four months ago (Table 1.6).

Table 1.6: *Relationship between Referred Worker's Performance and Referrer's Performance Individual and Supplemental Experiments*

	Dependent Variable: Referred's Submission Rate, Individual Experiment	Dependent Variable: Referred's Submission Rate, Supplemental Experiment		
Referrer's Submission Rate, Individual Experiment	0.421**	0.246**	0.132	
	(0.066)	(0.079)	(0.082)	
Referred Worker's Submission Rate, Individual Experiment				0.409**
				(0.078)
Constant	0.456	0.331	0.222	0.201
	(0.059)	(0.065)	(0.421)	(0.063)
Controls	No	No	Yes	No
Observations	255	255	255	255
R-squared	0.192	0.034	0.184	0.087

Notes: Each column presents the results of a regression of the dependent variable indicated by the column on the independent variable indicated by the row. Each observation is a referred worker. Huber-White standard errors are in parentheses. No controls are included except in the second-to-last column, which includes controls for referred worker characteristics listed in footnote 18. ** denotes significance at the 5% level.

Only some of this can be accounted for by observables. Appendix Table A.7 shows that workers with better observable characteristics refer workers who also have better observables. Controlling for the referred worker's observables in the regression in Table 1.6 reduces the point estimate on the referrer's performance by about half. Nonetheless, the referrer's performance remains a large and positive (albeit not statistically significant) predictor of her referral's performance.[28] This suggests that higher performers refer workers who perform better than would even be expected based on their observable characteristics.

1.6.2 Strength of Referrer-Referral Relationship

We turn now to the relationship between referrers and their referrals. Appendix Table A.8 shows the distributions of the three relationship variables we have from referrers' reports at the time of the referral.[29] Referrers tended to refer workers they were close to. Among those included in the experiment, most reported knowing their referrals "extremely well" (six on a scale of one to six), while only one percent said they knew their referral "hardly at all" (one on the same scale). According to referrers, 32 percent of referrals interacted with their referrers more than once a day (in person or remotely) and another 19 percent interacted about once a day; meanwhile, only 7 percent interacted once a month or less. We also asked workers how many other people they knew in common with their referral: 48 percent of referred workers knew 20 or more people in common with their referrer.

Because each relationship variable is consistently a positive predictor of the referral's performance, we build an index of relationship strength and for parsimony focus here on the resulting estimates.[30] We exclude the five referred workers whose referrers did not

---

[28]When on-time submission is used as the performance metric instead of submission, this coefficient is significant at the five percent level.

[29]We caveat this section by emphasizing that these relationship characteristics are self-reported by referrers. Referrers could have reported being close with workers whom they thought would perform particularly well and/or whom they particularly wanted the employer to hire.

[30]In building the index, we first create dummy variables for reportedly knowing the referred worker well (responding more than three on a scale of one to six when asked how well she knew the referred worker), interacting with the referral at least once a week, and knowing at least twenty people in common. Our relationship index is defined as the standardized sum of these three binary variables. The magnitudes of the coefficients are similar if we define the index instead as the standardized average of z-scores for the three raw variables (on scales of 1 to 6 for how well the referrer knew her referral, 1 to 7 for how often they

answer all the relationship questions at the time of the referral.

Panel A of Table 1.7 shows how characteristics of the referred worker vary with the strength of her relationship with her referrer. Each column shows the results of a different observable characteristic regressed on the relationship index. The reported coefficients show that referred workers who have stronger relationships with their referrer look worse on observables. They have passed fewer oDesk tests, completed fewer assignments, and, conditional on receiving feedback, have received (insignificantly) worse feedback.[31] These results suggest that when referrers refer people with whom they have weaker ties, they refer people who look better on paper.

---

interacted, and 1 to 5 for how many people they knew in common) or of z-scores for three constructed variables (with how well they knew each other on the same scale, but with how often they interacted coded as the estimated number of days per month they interacted and with how many people they knew in common coded as the midpoint in the chosen range).

[31]They are also significantly (5 percentage points) less likely to have received any feedback, probably because they have completed fewer assignments. Point estimates suggest they also look worse on the other observable characteristics we have, but the coefficients are not generally significant.

Table 1.7: *Relationship Strength, Observable Characteristics, and Performance*
Individual, Supplemental, and Team Experiments: Referred Workers

		A. Observable Characteristics	
	Tests Passed	Number of Assignments	Feedback Score
Relationship Strength Index	-0.462**	-1.456**	-0.038
	(0.138)	(0.579)	(0.030)
Constant	4.447	5.774	4.545
	(0.125)	(0.593)	(0.041)
Controls	No	No	No
Observations	532	532	293
R-squared	0.025	0.011	0.003
		B. Performance, No Controls	
	Submission (Individual Experiment)	Submission (Supplemental Experiment)	Same Slogan (Team Experiment)
Relationship Strength Index	0.041**	0.018	0.053**
	(0.020)	(0.027)	(0.020)
Constant	0.780	0.518	0.514
	(0.021)	(0.029)	(0.022)
Controls	No	No	No
Observations	1,512	1,260	560
R-squared	0.012	0.003	0.009
		C. Performance, With Controls	
	Submission (Individual Experiment)	Submission (Supplemental Experiment)	Same Slogan (Team Experiment)
Relationship Strength Index	0.047**	0.037	0.051**
	(0.020)	(0.026)	(0.022)
Controls	Yes	Yes	Yes
Observations	1,512	1,260	560
R-squared	0.106	0.155	0.048

Notes: Each column in each panel reports the results of a separate regression of the dependent variable indicated by the column on an index for the strength of the referrer-referred worker relationship. This index is defined in Section 6 of the text and has mean zero and standard deviation one. All regressions in the table include only referred workers. Regressions in Panel A include referred workers in both the individual and team experiments. Observations are at the worker level. No controls are included; Huber-White standard errors are in parentheses. Regressions in Panel B include no controls, while regressions in Panel C include the controls for worker characteristics listed in footnote 18. The first two columns of Panels B and C include workers from only the individual and supplemental experiments, respectively. In these columns, outcomes are observed at the worker-day level and standard errors are clustered by worker. The final column of Panels B and C includes only workers from the team experiment; outcomes are observed at the worker-PSA level and standard errors are clustered by blocking group. ** denotes significance at the 5% level.

Panel B investigates how a referral's performance varies with the strength of the referrer-referral relationship. For each experiment, a worker's performance on a given day (or a given PSA in the team experiment) is regressed on the relationship index. For parsimony we present only one outcome per experiment, though within experiments, the magnitudes of the coefficients are similar when performance is defined using the other metrics. In each experiment, referred workers performed better the stronger their relationship with their referrer. A referred worker with a one standard deviation stronger relationship with her referrer was four percentage points more likely to submit work in the individual experiment and two percentage points more likely to submit work in the supplemental experiment, though the latter point estimate is not statistically significant. In the team experiment a referred worker with a one standard deviation stronger relationship with her referrer was five percentage points (nearly ten percent) more likely to have her slogan match her partner's.

Panel C presents the results of these same regressions with the inclusion of controls for the referred worker's observable characteristics. Given that referred workers with stronger ties to their referrers tended to have worse observable characteristics, it is unsurprising that the coefficients on the relationship index are, on average, larger when the controls are added.[32] These results are consistent with the idea that when workers refer people they know well, they choose workers who do not look as good on paper, but who perform well in ways that would not be predicted by their observables.

[32]We do not have enough power to test the interaction of the relationship variables and treatment (in the individual experiment) or team type (in the team experiment). However, the coefficients on the interactions are not consistently signed across outcomes within each experiment.

1.7 Potential Bias from Employers' Hiring Decisions

In each experiment, we hired all applicants who met our basic hiring criteria. This ensures that employers' hiring decisions did not lead to differential selection of referred and non-referred workers into our sample. In this section, we use our experimental data to simulate how our comparisons between referred and non-referred workers might have been biased had we only observed the performance of workers an employer chose to hire. Though the results of this exercise are qualitatively similar under different assumptions, our aim in this section is not to pin down the particular bias that would be generated by an employer's hiring decisions, but rather simply to demonstrate that such a bias might exist.

We first simulate which workers employers would hire if they only observed the characteristics on workers' resumes; we then simulate whom employers would hire if they additionally observed which workers had been referred. In each hiring scenario, we assume that employers want to maximize the fraction of workers who submit a response on a given day and that they know the relationship between demographics and referral status, and performance.[33] Employers predict each applicant's performance using the information they observe and then hire the half of the applicant pool with the best predicted performance.

Table 1.8 shows the results of the simple simulations. Results in the first row simulate hiring under the assumption that employers only see workers' resumes, not who was referred. To calculate a given worker's predicted performance, we first regress the performance of all other workers (excluding herself) on their resume characteristics and then use the estimated coefficients to predict the excluded worker's own performance. We

[33]In practice, an employer may prefer to hire a referred worker over a non-referred worker who is predicted to perform slightly better either as a source of compensation to an existing employee or because the referred worker is predicted to persist longer at the firm. For simplicity and clarity, we abstract away from any such considerations here.

also use this same predicted performance as a summary measure of workers' observable characteristics. Results in the second row simulate hiring under the assumption that employers observe not only workers' resume characteristics but also who was referred. We follow the same procedure to predict workers' performance except that the regressions of worker performance on observable characteristics also include an indicator for whether the worker was referred.

Panel A shows the fraction of referred and non-referred applicants that would have been hired under each scenario. If employers only took workers' resume characteristics into account, a higher fraction of referred (58 percent) than non-referred (39 percent) workers would have been hired because referred workers had better observable characteristics. However, if employers also observed who was referred, the fraction of referred applicants that would have been hired jumps to 77 percent; meantime, only 12 percent of non-referred applicants would have been hired. Panel B displays the summary measure of the hired workers' observable characteristics: when employers observe who was referred, hired non-referred workers have substantially better observable characteristics than hired referred workers.

Panel C shows the average actual submission rates of the referred and non-referred workers that would have been hired in each scenario. If employers did not observe who was referred, hired referred workers would have been substantially (19 percentage points) more likely to actually submit work. However, this difference would have been only five percentage points (and statistically indistinguishable from zero) if employers also observed who was referred.[34] This suggests that if we had only observed the performance of hired workers and did not observe all the characteristics employers used in hiring decisions, we might have mistakenly concluded that referrals contained little to

[34]In fact, if we assume employers hired the top third or top quarter of the applicant pool (rather than the top half), hired referred workers would have performed two or three percentage points *worse* than hired non-referred workers when the employer used referral status in hiring decisions. As in the main specification, these differences are not significant.

Table 1.8: *Simulated Hiring*

	A. Fraction Hired		B. Measure of Observables		C. Actual Submission Rate		
	Referred Applicants	Non-Referred Applicants	Hired Referred Workers	Hired Non-Referred Workers	Hired Referred Workers	Hired Non-Referred Workers	Difference
Observe Characteristics Only	58%	39%	80%	79%	82%	64%	19%**
Observe Characteristics & Referral Status	77%	12%	77%	86%	81%	75%	5%
Applicant Pool Average			73%	68%	81%	75%	15%**

Notes: The first row simulates hiring under the assumption that employers observed only workers' resume characteristics, but not their referral status. The second row simulates hiring assuming employers observed workers' resume characteristics and referral status. Panel A presents the fraction of referred and non-referred workers that would have been hired under each scenario. Panel B presents a summary measure of the observables of the referred and non-referred workers who would have been hired. (This is the average predicted probability of submission, based on the observable characteristics listed in footnote 18. It is described in the text.) Panel C presents the actual submission rate of the referred and non-referred workers who would have been hired. The column labeled Difference provides the difference in average submission rates of the referred and non-referred workers who would have been hired under each scenario. ** denotes that this difference is significant at the 5% level.

Individual Experiment: Assuming Top 50 percent of Applicants Hired no information about worker performance.

1.8 Conclusion

This chapter presents the results of three field experiments in an online labor market, comparing the performance and turnover of referred and non-referred workers. Throughout, we find that even conditional on their resume characteristics, referred workers performed better and had less turnover than their non-referred counterparts. That is, referrals contained information about worker quality that was not present on workers' resumes.

Much of the performance and turnover differential between referred and non-referred workers was driven by selection. In the individual experiment, even non-monitored referred workers outperformed and outlasted non-referred workers. In the supplemental experiment, referred workers outperformed and outlasted non-referred workers, even at a job for which they were not referred at a firm at which their referrers did not work.

However, we also find strong evidence that on-the-job interactions between referred workers and their referrers drove some of the performance differential. In particular, our results suggest that team production is an important benefit of referrals. Referred workers in the team experiment performed particularly well when working with their own referrers; they were also more eager to continue working in that pairing.

We find that workers referred by high-performers and workers with strong ties to their referrers performed particularly well. High-performers tended to refer workers who looked better on observables, but who performed better than expected, even conditional on these characteristics. Referrals with strong ties to their referrers actually looked *worse* on paper than did those with weak ties. Nonetheless, it was the referrals with strong ties who performed better, even without conditioning on observable characteristics.

The existing literature finds mixed results on whether referred workers perform better than non-referred workers (e.g., Blau (1990), Burks, Cowgill, Hoffman, and Housman

(2013); Castilla (2005)). We see our results as consistent with these seemingly divergent papers. The performance data in much of this literature come from workers firms chose to hire. However, if employers incorporate referrals into hiring decisions (for example, because referrals positively predict performance and persistence), hired referred workers could perform better than, worse than, or similarly to non-referred workers even though a referral is a positive signal of productivity.

We find that selection is important in explaining why referred workers outperform and outlast non-referred workers, but we have limited evidence on why. One explanation (as in Montgomery, 1991) is that there is simply homophily among friends: productive workers have productive friends. Another explanation is that (as in Beaman and Magruder, 2012) workers have information on which of their friends are particularly productive and may choose these particularly productive workers to refer. Understanding the relative contributions of these two factors is an important question for future research.

Chapter 2

Gen X to Gen Y: Changes in Educational Attainment by Gender and Socioeconomic Status

2.1 Introduction

Currently, gender and socioeconomic status strong predictors of educational attainment. Females tend to outpace males; and individuals from high-income families on average accrue more education than their lower-income counterparts. In this paper, I compare two cohorts of Americans: one born in the early 1960's and the other born in the 1980's. I examine the rise of the female advantage in education. In the older of these cohorts, females are at parity with males in years of schooling; by the later cohort, females surpass males in nearly all measures of educational attainment. I then pool males and females and examine the growth in educational attainment across socioeconomic status.

The female advantage in education has attracted considerable interest, most recently in its relation to socioeconomic status. Bailey and Dynarski (2011) posit that the growth of the female advantage is tied to an increased gradient of education with respect to income.

They argue that the recent increase in the relative educational attainment of females is driven largely by the increase among females from higher-income families amid little change for males throughout the SES distribution. Goldin, Katz, and Kuziemko (2006), meantime, find that the reversal of the college gender gap over a slightly longer period occurred somewhat continuously throughout the socioeconomic distribution.

Using the same data as in Bailey and Dynarski (2011) (henceforth referred to as BD), but nationally representative weightings and more permanent measures of SES, I examine the relationship between the female educational advantage and socioeconomic status. I show that the recent rise of the female college advantage actually occurred quite evenly across the SES distribution. That is, females gained on males in educational attainment, but it was not due to any particular change in educational attainment by sex for the upper part of the SES distribution.

The growth of female educational attainment relative to male educational attainment has also attracted some interest in its relationship to household structure. Bertrand and Pan (2013) show that in single-mother homes, more so than in two-parent households, girls outpace boys in non-cognitive and academic skills through at least grade school. Jacob (2002) provides evidence that non-cognitive skill differences between males and females can explain part of the educational gender gap at later ages, including college completion.

Between the earlier and later cohorts examined in this paper, the likelihood that a child was raised in a single-mother household doubled. Although girls used to outpace boys slightly (insignificantly) more in single-mother homes than in intact families, however, I show that today's female advantage in education is in fact seen in approximately equal magnitudes across household structures, and is, if anything, slightly *less* pronounced among those raised in single-mother homes. Examining the female advantage by race and ethnicity, I also show that the Hispanic population exhibited the smallest female educational advantage in each cohort. The doubling of the within-cohort percentage

Hispanic over the period in question may have tempered somewhat the growth of the female educational advantage.

Educational attainment in this period also grew quite similarly throughout the socioeconomic distribution for males and females taken as a group. I reconcile my results with BD, who use the same data, but posit a rise in the gradient of education with respect to income. First, I show how sampling weight errors drove an overstatement of the growth of educational inequality in BD. I then show that with more permanent measures of SES, the increase in educational attainment was more similar still across the SES distribution. My results are loosely consistent with Chetty, Hendren, Kline, Saez, and Turner (2014), who find high levels of stability in intergenerational mobility for the 1971 to 1993 birth cohorts.

The remainder of the paper proceeds as follows. In Section 2.2, I describe the national longitudinal data on the two cohorts and present my educational attainment measures, SES proxies, and other demographic variables. In Section 2.3, I quantify the rise of the female educational advantage between the NLSY cohort born in the early 1960's and that born in the early 1980's, and show relatively equally across SES. I also estimate the rise of the female educational advantage separately by racial/ethnic group and by family household structure. My results suggest that the increase in single-mother households cannot explain much, if any, of the growth of the female advantage in education over this period, and that the concurrent growth of the Hispanic population may have tempered the rise. In Section 2.4, I pool males and females and show that overall educational attainment also increased quite similarly throughout the SES distribution, and reconcile my findings with BD, who find an increased gradient of education with respect to income over the same period. Section 2.5 concludes.

2.2 The NLSY Education and Demographic Data

I use data from the 1979 and 1997 National Longitudinal Surveys of Youth (henceforth referred to as the NLSY79 and NLSY97, respectively). Each survey follows a nationally representative sample of Americans beginning in young adulthood and tracks respondents as they move through the educational system. For respondents completing school before first interview, the surveys also include retrospective data on some educational outcomes. Importantly, the high levels of consistency between the NLSY79 and the NLSY97 facilitate cross-cohort research.

The sample restrictions imposed in most of this paper are minimal. To have a consistent sample for each of the ages at which I measure educational attainment, I exclude NLSY respondents for whom years of schooling attended and completed by ages 19 and 25, and degrees completed by age 25, are not known (19 percent of NLSY79 respondents and four percent of NLSY97 respondents). For the NLSY79, the excluded respondents are primarily those for whom, on account of both having been older than 20 at age of first interview and having incomplete retrospective data, are missing education data at age 19. In the case of the NLSY97 sample, they are primarily respondents who were still too young at the last available interview year (2009 Wave) to have been observed at age 25. From the remaining respondents, I drop only the small number for whom I have no measure of SES (less than one percent in each cohort). The resulting sample includes 10,427 respondents from the NLSY79 and 8,575 respondents from the NLSY97.[1]

2.2.1 Education Measures

The NLSY surveys provide an array of potentially informative educational measures. I focus for parsimony principally on a summary outcome measure: years of schooling

[1] To reconcile my findings with existing work, in Section 2.4 I impose additional sample restrictions in line with the paper, which I will discuss in turn.

completed by age 25. Relative to educational attainment measures like college entry and college completion, years of schooling completed is a fairly precise measure of schooling.[2]

Less precise measures of educational attainment like college entry (to two- or four-year college by age 19) and college completion (of 4-year institution by age 25) tend to follow similar trends. There is some evidence that years of schooling are more prone to misreporting than are these other, coarser measures.[3] I return to entry and completion measures in Section 2.4 when I reconcile my findings with an existing paper that uses these outcomes.

2.2.2 Proxying for SES and Other Demographic Measures

I examine whether the rise in educational attainment (and in the female advantage in educational attainment) has varied by the socioeconomic status of the household in which the respondent was raised. To capture a relatively credible and permanent measure of household SES, I proxy with the average of the z-scores for the following measures (included for each respondent when available): (1) annual net household income for each year the individual is 18 or under, and (2) the educational attainment (in years) of each biological parent as reported in the first interview (Wave 1). Summary statistics on these household income and parental educational attainment measures are displayed in Appendix Table B.1. In the paper's final section, I show educational attainment by cohort

[2]An important implicit assumption is that every year of schooling has the same value.

[3]Kane, Rouse, and Staiger (1999), for example, find that while less than ten percent of degree recipients inaccurately report degree attainment, almost half of those with 1 or 2 years of college credit inaccurately report their educational attainment. For robustness I thus also replicate the main specifications measuring educational attainment instead as college entry (by age 19) and college completion (by age 25). High school transcript data exists for most NLSY respondents; comparisons of this transcript data to self-reported highest grades attended and completed suggest that misreporting is minimal. College transcript data is only linked for a small fraction of college attendees in the sample, however, such that for most respondents completing more than 12 years, similar misreporting checks are infeasible. The results are generally similar also when the outcomes are categorical variables for degrees by age 25 (e.g. AA, BA, BS), or when the outcome is years of schooling attended (rather than completed) by age 25; both are available upon request.

and SES quartile for each of an array of different SES proxies.

I also study the rise of the female advantage by family structure and by race/ethnicity. I classify respondents by family structure based on the composition their household when young. I focus on two household structures: two-parent and single-mother. Two-parent households include all households with a mother and father figure living at home; the parental figures can be biological, adoptive, or step. Single-mother households include all households with a biological mother present and no father figure in the house.[4] For both the NLSY79 and the NLSY97 samples, household structure classification is based on retrospective respondent reports on household composition at age 12, collected during the Wave 1 interview. I classify each respondent into one of three racial/ethnic groups based on self-reported predominant race or ethnicity: Black, Hispanic, and non-Black, non-Hispanic. These are mutually exclusive and exhaustive.

Table 2.1 reports descriptive statistics for demographic variables both by NLSY cohort and by gender within NLSY cohort. Panel A reports the means of demographic variables for the NLSY79 and NLSY97 cohorts, respectively. From the earlier generation to the later, both the composition of household structures and racial diversity changed. Relative to members of the NLSY79 cohort, members of the NLSY97 cohort were much less likely to grow up in a household with both parents. Whereas over 80 percent of the earlier NLSY cohort lived with both biological parents at age twelve, 70 percent of the later cohort did. Instead, many members of the NLSY97 cohort were raised in single-mother households. The proportion of individuals raised in single-mother households more than doubled from 12 percent to 25 percent over this period. Members of the more recent NLSY cohort were also slightly (2 percentage points) more likely to be Black and more than twice as likely to be Hispanic (13 percentage points as compared to 6 percentage points).

[4]Two-parent and single-mother household categories are not mutually exhaustive; respondents raised in households headed by grandparents, single fathers, and foster parents, for example, are not included in this particular analysis.

Table 2.1: *Descriptive Statistics by NLSY Cohort*

	A. By Cohort	
	NLSY79	NLSY97
Female	0.492	0.486
Black	0.136***	0.154***
Hispanic	0.063***	0.129***
Both Bio Parents Household	0.825***	0.679***
Single Mother Houshold	0.122***	0.247***
Observations	10,261	8,575
	B. NLSY79 Cohort by Gender	
	Female	Male
Black	0.135	0.137
Hispanic	0.063	0.064
Both Bio Parents Household	0.821	0.828
Single Mother Houshold	0.127	0.118
SES	-0.004	0.004
Observations	5,107	5,154
	C. NLSY97 Cohort by Gender	
	Female	Male
Black	0.162	0.157
Hispanic	0.127	0.131
Both Bio Parents Household	0.664***	0.693**
Single Mother Houshold	0.263***	0.231***
SES	-0.013	0.011
Observations	4,172	4,403

Sources: NLSY79 and NLSY97.
Notes: Sample restrictions and variable definitions as outlined above. Sampling weights used throughout; significance of differences calculated using population standard deviations. ** and *** denote rejection of the null of equality of means for the two columns at the 5 percent and 1 percent levels, respectively.

The subsequent panels of Table 2.1 report means separately by gender for the NLSY79 cohort (Panel B) and for the NLSY97 cohort (Panel C). In the earlier cohort, there are no significant differences between boys and girls in race or in our household structure. By the later NLSY cohort, males are more likely than their female peers to have grown up in a household with both biological parents and, correspondingly, less likely to have grown up with a single mother. This is loosely consistent with Dahl and Moretti (2008), which demonstrates that a first-born daughter is less likely to be living with her father than is a first-born son.

2.3 The Rise of the Female Advantage in Education

2.3.1 Overview

The growth of female educational attainment in recent decades has garnered particular attention, in part because it involves females surpassing males in educational attainment.[5] Table 2.2 shows summary statistics on educational attainment by gender for nationally representative samples of Americans. The first columns correspond to those born in the early 1960's; the next three columns to those born in the early 1980's. The final column reports the change in the female educational advantage across cohorts.

In the NLSY cohort born in the early 1960's, females were at parity with males in most measures of educational attainment. Although females in this NLSY cohort were slightly more likely than males to enter a two- or four-year college by age 19, by age 25 they had completed just one-tenth of a year more schooling and were no more likely to have graduated college. By the next generation, however, females had surpassed males in educational attainment. In the NLSY cohort born in the 1980's, the most recent cohort

[5] See, e.g., Peter and Horn (2005); Goldin, Katz, and Kuziemko (2006); Bailey and Dynarski (2011); Quenzel and Hurrelmann (2013); DiPrete and Buchmann (2013).

available for study at age 25, females were about ten percentage points more likely than males both to enter college by age 19 and to complete college by age 25. And they on average attained one-half a year of schooling more than their male counterparts.[6]

Table 2.2: *Rise of the Female Advantage in Education*

	1959 - 1965 Birth Cohort			1979 - 1985 Birth Cohort			Δ Female Advantage
	Females	Males	$(\mu_F\text{-}\mu_M)$	Females	Males	$(\mu_F\text{-}\mu_M)$	
College Entry	0.436 (0.496)	0.405 (0.491)	0.031**	0.594 (0.491)	0.471 (0.499)	0.123***	0.092***
College Completion	0.205 (0.404)	0.206 (0.404)	-0.001	0.342 (0.474)	0.245 (0.430)	0.097***	0.098***
Highest Grade Completed	13.055 (2.199)	12.942 (2.342)	0.113**	13.614 (2.695)	12.987 (2.598)	0.627***	0.514***
Observations	5,107	5,154		4,172	4,403		

Sources: NLSY79 and NLSY97.
Notes: Sample restrictions as detailed in text. Entry is to two- or four- year college (by age 19); completion is of 4-year college (by age 25); highest grade completed is measured at 25. Standard errors listed in parentheses. *, **, and *** denote significance of the differences at the 10 percent, 5 percent, and 1 percent levels, respectively.

2.3.2 Female Educational Advantage by SES

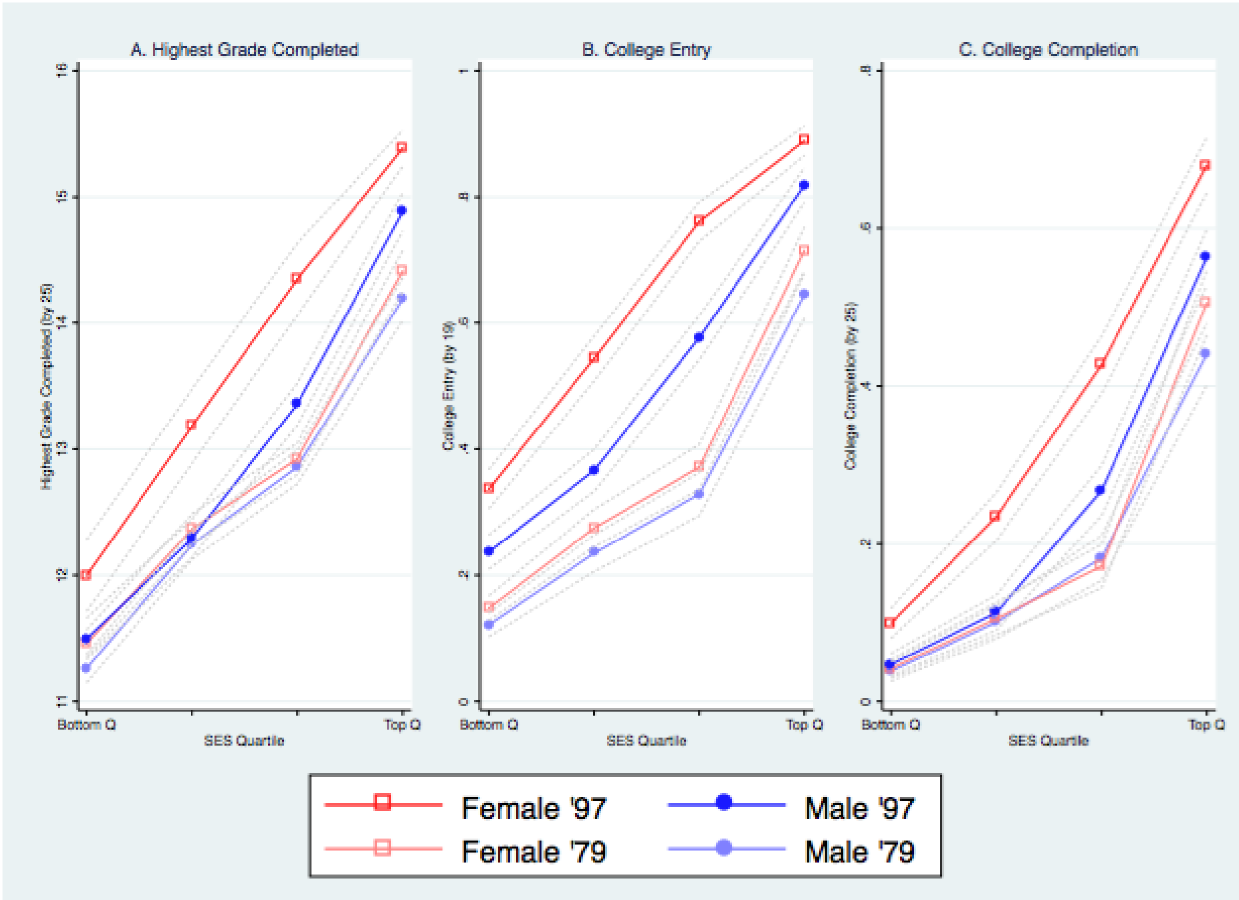
Did the rise of the female educational advantage occur approximately equally throughout the SES distribution, or did it occur more prominently within certain classes? I examine the rise of the female educational advantage by SES using nationally representative samples, consistent measures of educational attainment by ages 19 and 25, and consistent and relatively stable proxies for childhood SES.

[6]Educational attainment of females relative to males was not stagnant in the preceding periods. Early in the 20th century, females were about as likely as males to attend college. Male attendance began outpacing female attendance in the 1930's and through the 1940's. Thereafter females began to catch up again in enrollment numbers and began surpassing males as of the 1980's (Goldin, Katz, and Kuziemko (2006)).

I study the same NLSY cohorts as BD, but my methods differ from theirs in three ways. First, BD restrict the sample to a subset of NLSY respondents: those born 1961 to 1964 and those born 1979 to 1982 for the NLSY79 and NLSY97 cohorts, respectively. I only exclude respondents on whom educational data is missing at age 19 and/or 25, or for whom I have no measure of SES. Second, BD proxy for socioeconomic status with household income at the time of the first wave, when children were 15 to 18 years of age. I use an SES proxy that is arguably less noisy and provides a more permanent measure of household socioeconomic status. I include (where available) household income for each year the individual was 18 or under and the educational attainment of each biological parent (see Section 2.2 for details). Third, BD define SES quartiles within the sample rather than within the weighted sample. Since the NLSY oversamples from underpriviledged populations, and since the sampling changed across cohorts, their results are not nationally representative. In Section 2.4, I compare my findings with theirs.

With minimal sample restrictions, more permanent measure of household SES, and nationally representative weightings, I find that the rise of the female educational advantage was actually fairly similar across SES groups. Panel A of Figure 2.1 shows average highest grade completed by SES quartile for each gender in each NLSY cohort. In the earlier NLSY cohort, females in each SES quartile were about on par with their male peers. By the later NLSY cohort, educational attainment had risen significantly for females across the board while male educational attainment was stagnant in the lower quartiles and rose only slightly (insignificantly) in the upper quartiles. The results for other measures of educational attainment like college entry by age 19 and college completion by age 25 are highly comparable and are presented in Panels B and C, respectively.

Figure 2.1: *Female Educational Advantage by SES*



Sources: NLSY79 and NLSY97.
Notes: Excludes individuals for whom educational attainment at 19 and/or 25 are not observed, or for whom no SES measure is observed. Dotted lines represent 95 percent confidence bands. Sample restrictions and SES definition as detailed in text. Panel A shows highest grade completed by age 25; Panel B shows college entry (two- or four-year college) by age 19; and Panel C shows college completion (four-year college) by age 25.

Table 2.3 shows the corresponding regression results for Panel A. Each of the first four columns restricts to respondents from a single SES quartile and estimates the change in the female educational advantage in highest grade completed from one NLSY cohort to the next.[7] The coefficient on the interaction of the indicator for female and the indicator for the NLSY97 cohort captures the rise of the female educational advantage for that quartile of the SES distribution. Consistent with Figure 2.1, the rise of the female educational

[7]Recall that SES quartiles are defined *within* each NLSY cohort.

advantage is large and significant within each quartile. Though none of the differences is significant, the magnitude of the rise is slightly larger in the second and third quartiles than in the top and bottom quartiles. The underlying reasons, however, are if anything more different between the bottom half and the top half of the SES distributions. In the bottom half of the SES distribution, the rise of the female advantage came from increased educational attainment among females amid decreased educational attainment among males. In the top half of the SES distribution, educational attainment rose for both genders, though the increase among females was significantly larger than among males.

Table 2.3: *Female Educational Advantage by SES*

	By Quartile				Overall
	Bottom Q	2nd Q	3rd Q	Top Q	
Female	0.198**	0.124	0.064	0.221*	0.155***
	(0.080)	(0.085)	(0.093)	(0.120)	(0.047)
NLSY97	0.235**	0.049	0.498***	0.691***	0.370***
	(0.102)	(0.097)	(0.109)	(0.116)	(0.053)
Female x NLSY97	0.304	0.765***	0.926***	0.284*	0.561***
	(0.185)	(0.189)	(0.193)	(0.161)	(0.091)
SES					1.111***
					(0.039)
Female x SES					-0.004
					(0.053)
SES x NLSY97					0.072
					(0.057)
Female x SES x NLSY97					0.023
					(0.083)
Constant	11.261	12.243	12.858	14.190	12.631
	(0.058)	(0.059)	(0.068)	(0.087)	(0.034)
Observations	6,777	4,611	3,980	3,468	18,836
R-squared	0.009	0.020	0.051	0.044	0.202

Sources: NLSY79 and NLSY97.
Notes: NLSY79 and NLSY97 pooled for ease of exposition; population weights for both samples employed together. Estimating separately by NLSY cohort, each with own set of weights, yields similar results. Sample restrictions and variable definitions as detailed in text. Standard errors in parentheses. *, **, and *** denote significance at the 10 percent, 5 percent, 1 percent levels, respectively.

The final column of Table 2.3 shows that the similarity of the rise in the female educational advantage across SES holds also when measuring SES continuously, rather than in quartiles. Consistent with the per-quartile evidence, the coefficient on the triple interaction of an indicator for female, an indicator for the later NLSY cohort, and a continuous measure of SES is small and statistically insignificant. These results are robust also to the inclusion of controls for race and household structure (Appendix Table B.2).

2.3.3 Female Educational Advantage by Race, Household Structure

Table 2.4 shows the rise of the female educational advantage separately by predominant racial or ethnic group. The growth of the female advantage in education between the NLSY cohorts was quite similar across races and ethnicities. The magnitudes in the later cohort did, however, vary since the female advantage in education was already prominent among Blacks in the earlier NLSY cohort.

The source of the rising female advantage in education also differed by race. For Blacks, it was driven by an increase in female attainment amid only very small increases in male attainment. For non-Blacks, it arose from a large increase in female attainment combined with a relatively smaller but still substantial increase in male attainment (0.6 years for non-Hispanic, non-Black males and 0.9 years for Hispanic males). In each NLSY cohort, moreover, the female educational advantage among Hispanics was lower than for non-Hispanics, suggesting that the growth of the Hispanic population across the cohorts if anything tempered the rise of the female advantage in education over this period.

Table 2.4: *Female Educational Advantage by Race*

	Non-Hispanic, Non-Black	Hispanic	Black
Female	0.060	-0.026	0.341***
	(0.064)	(0.130)	(0.078)
NLSY97	0.579***	0.852***	0.082
	(0.074)	(0.128)	(0.121)
Female x NLSY97	0.599***	0.539***	0.545***
	(0.127)	(0.197)	(0.159)
Constant	12.773	11.507	11.992
	(0.046)	(0.092)	(0.055)
Observations	10,264	3,679	4,893
R-Squared	0.029	0.045	0.022

Sources: NLSY79 and NLSY97.
Notes: NLSY79 and NLSY97 pooled for ease of exposition; population weights for both samples employed together. Estimating separately by NLSY cohort, each with own set of weights, yields similar results. Sample restrictions and variable definitions as detailed in text. Standard errors in parentheses. *, **, and *** denote significance at the 10 percent, 5 percent, 1 percent levels, respectively.

Bertrand and Pan (2013) show that among children from single-mother homes, girls outpace boys up through at least grade school along measures of both non-cognitive skills and academic success. In related work, Jacob (2002) shows that non-cognitive skill differences between males and females can explain part of the college completion gap between males and females. Given that the percentage of children raised by single mothers doubled over the period in question, I examine the female educational advantage by household structure.

Panel A of Table 2.5 shows the rise of the female educational advantage separately by each of two types of households: two-parent and single-mother. Relative to single-mother homes, homes with two parents may have had a smaller gender gap in educational attainment in the earlier NLSY cohort, but the female educational advantage appears to have risen by somewhat more among children from two-parent households than among children from single-mother homes (0.61 years versus 0.46 years). On net, then, the magnitude of the female educational advantage in the later NLSY cohort is if anything slightly *smaller* among children from single-mother homes.

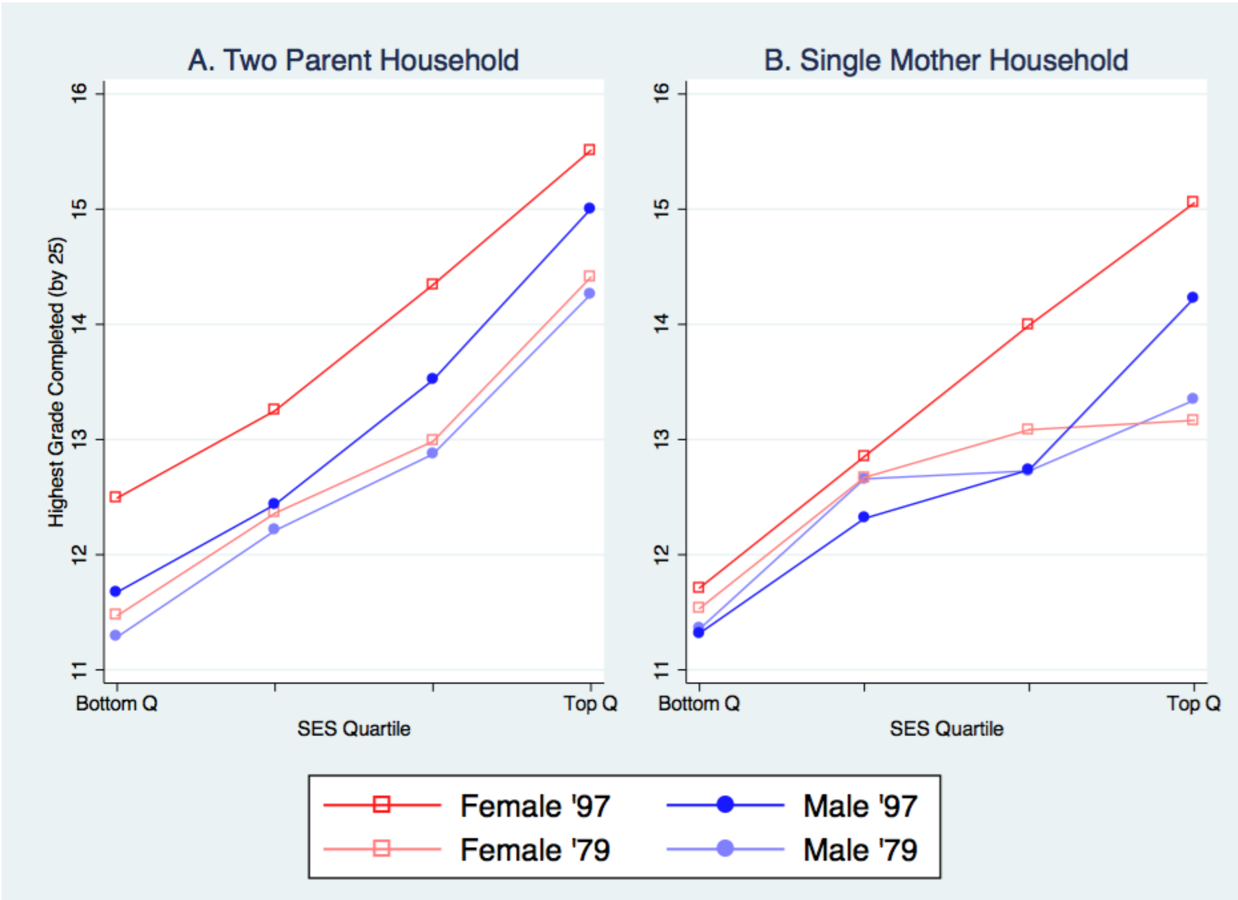
Table 2.5: *Female Educational Advantage by Household Structure and SES*

	A. Overall		B. By SES	
	Two Parent	Single Mother	Two Parent	Single Mother
Female	0.109*	0.126	0.157***	0.198
	(0.060)	(0.130)	(0.052)	(0.155)
NLSY97	0.719***	0.074	0.572***	0.110
	(0.069)	(0.135)	(0.061)	(0.142)
Female x NLSY97	0.606***	0.456**	0.582***	0.534***
	(0.109)	(0.187)	(0.103)	(0.203)
SES			1.139***	0.848***
			(0.045)	(0.090)
Female x SES			-0.019	0.063
			(0.060)	(0.141)
SES x NLSY97			0.001	0.269*
			(0.065)	(0.144)
Female x SES x NLSY97			-0.123	0.263
			(0.098)	(0.218)
Constant	12.707	12.090	12.644	12.514
	(0.043)	(0.095)	(0.037)	(0.105)
Observations	12,474	3,977	12,474	3,977
R-Squared	0.044	0.010	0.249	0.120

Sources: NLSY79 and NLSY97.
Notes: Outcome is highest grade completed by age 25. Two-parent and single-mother households are as defined in the text based on retrospective household structure at age 12 as self-reported at baseline (Wave 1) interview. NLSY79 and NLSY97 pooled for ease of exposition; population weights for both samples employed together. Estimating separately by NLSY cohort, each with own set of weights, yields similar results. Sample restrictions and variable definitions as detailed in text. Standard errors in parentheses. *, **, and *** denote significance at the 10 percent, 5 percent, 1 percent levels, respectively.

Figure 2.2 shows educational attainment by household structure, gender, and SES quartile. Though the rise of the female educational advantage was also generally similar by SES among children from two-parent households, among children from single-mother families the female educational advantage rose more at higher income levels. Panel B of Figure 2.2 shows that the underlying story for single-mother homes is in fact similar to the phenomenon set forth in BD for all: among those raised in single-mother homes, the female educational advantage rose more at higher income levels than lower income levels, driven by large advances among girls at the very top of the SES distribution, while all other girls experienced little change in educational attainment and boys remained stagnant or, in the case of the bottom three quartiles, lost ground.

Figure 2.2: *Female Educational Advantage by Household Structure and SES*



Sources: NLSY79 and NLSY97.
Notes: Replicates Panel A of Figure 2.1 separately by household structure. Two-parent and single-mother households are as defined in the text based on retrospective household structure at age 12 as self-reported at baseline (Wave 1) interview.

Given that only a very small fraction of children from single-mother households fall in the top SES quartile, I also present the corresponding results when SES is defined continuously (Panel B of Table 2.5).[8] For those *not* raised in two-parent homes, the growth of the female advantage was if anything negatively correlated with socioeconomic status while for those raised in single-mother homes the growth of the female educational advantage was larger at higher levels of SES. In neither case, however, is the point estimate

[8]In the NLSY97 cohort, children from single-mother families comprised just 2 percent of children in the top SES quartile; in the NLSY79 cohort, the corresponding statistic is a mere 1.3 percent.

statistically significant.[9]

2.4 The Gradient of Education with Respect to Income, from 1980 to 2000

BD posit that the growth of the female educational advantage is tied to an increased gradient of education with respect to income. Their argument comes in two parts. First, BD present evidence of an increased gradient of education with respect to income over this period: they show more growth of educational attainment across the NLSY cohorts for children from high-income than low-income families. Second, BD present evidence suggesting that the increased gradient is seen more among females than among males.

The increased gradient of education with respect to income presented in BD is driven largely by a lack of sampling weights in defining income quartiles. With proper weighting, the estimated rise in educational attainment is much more similar across quartiles. With more permanent SES measures, the estimated rise is more similar still. Incorrectly weighted quartiles and temporary income measures overstate not only the increase in the gradient overall, but also in the growth of the female advantage in education.

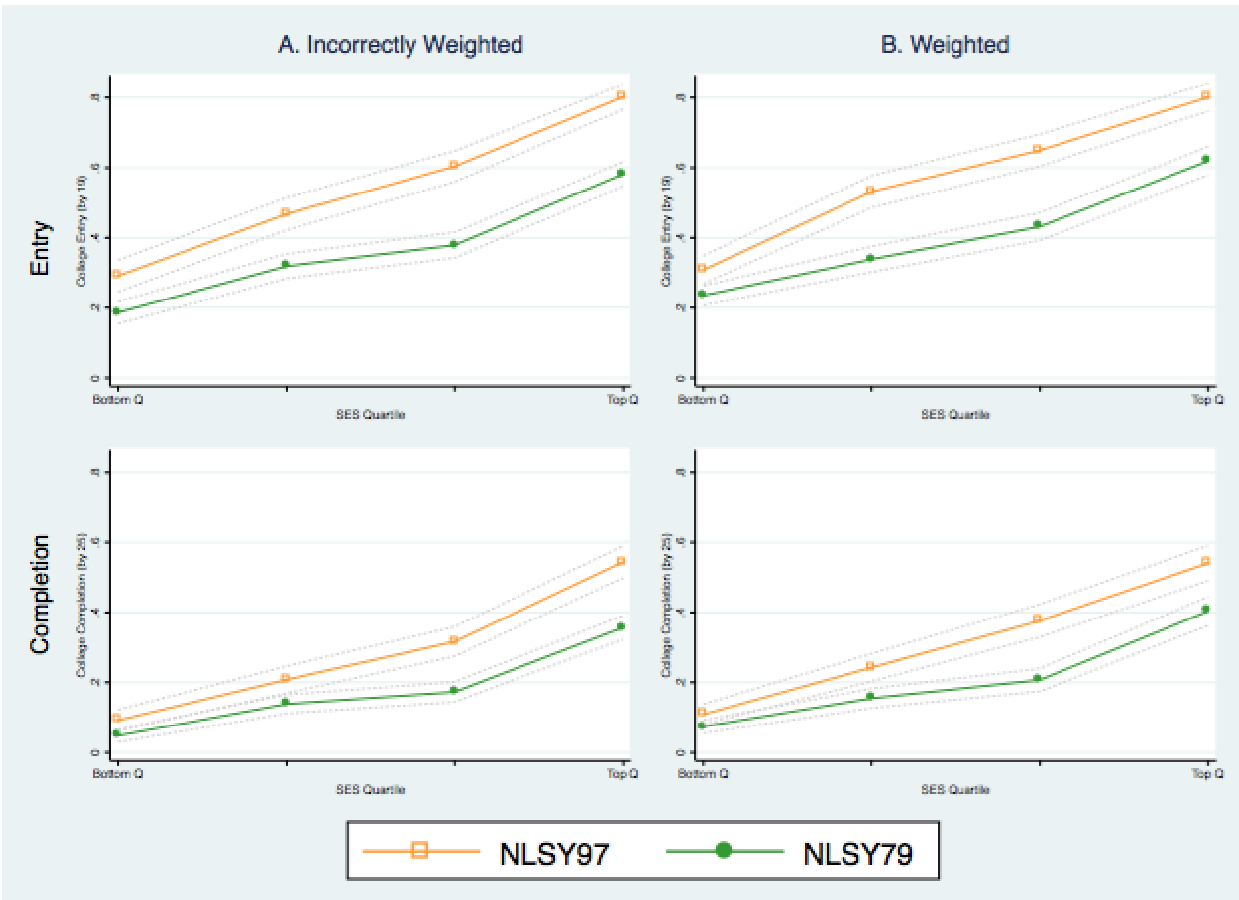
2.4.1 Implications of Weighting Errors in Bailey and Dynarski (2011)

Panel A of Figure 2.3 replicates BD's evidence of an increased gradient of education with respect to income. The first row shows college entry rates by family income quartile in the base year (Figure 2 in BD). The second row shows the corresponding college completion rates (Figure 3 in BD). BD interprets these figures as follows: "... (E)vident in Figures 2 and 3 is that the college entry rate and the college completion rate rose between the two periods. However, the increases were highly uneven, with gains largest

[9]The difference between the coefficients is just shy of significance at the ten-percent level (z=1.62).

at the top of the income distribution and smallest at the bottom ... (T)he product of this uneven growth was increased inequality in college outcomes during a period in which educational attainments became increasingly strong determinants of subsequent income."

Figure 2.3: *Gradient of Education with Respect to Income by Weighted and Incorrectly Weighted Income Quartiles*



Sources: NLSY79 and NLSY97.
Notes: The first row shows college entry rates by 19, the second shows college competion rates by 25; each is reported per quartile and per NLSY cohort. The sample restrictions in BD are used throughout. Panel A replicates Figures 2 and 3 in BD, in which sampling weights are not used in income quartile definitions. Panel B uses sampling weights in defining income quartiles.

In BD (and in Panel A), sampling weights are appropriately used in calculating the means within each quartile. However, family income quartiles are defined *without* sampling weights. That is, one-quarter of the sample – rather than of the weighted sample

– is assigned to each income quartile. I refer to these as "incorrectly weighted quartiles."[10] Since the NLSYs oversample from underprivileged populations, the estimated rates of entry and completion by incorrectly weighted quartile are not representative of national levels of entry and completion by quartile. Moreover, sampling was not consistent between the two NLSY cohorts. The top incorrectly weighted quartile in the NLSY97, for example, is a more selected sample than the top incorrectly weighted quartile the NLSY79; failing to use sampling weights in defining quartiles thus mechanically overstates growth in educational attainment at the top.

Table 2.6 shows mean family income in the base year for the incorrectly weighted quartiles and the weighted quartiles, respectively. (In all cases the means are calculated using sampling weights.) Panel A corresponds to the NSLY79 cohort, Panel B to the NLSY97 cohort. For ease of comparison, I maintain the sample restrictions in BD throughout this section: only those born 1961 - 1964 or 1979 - 1982 (and for whom education is observed at 18 and 25) are included. Because the NLSYs oversample from underprivileged households, the mean income in the incorrectly weighted quartiles is in all cases lower than in the weighted quartiles.

Table 2.6 also lists the proportion of the true population falling in each quartile. The bottom incorrectly weighted quartile in the NLSY79, for example, includes only the bottom 14.4 percent of the population. (In the weighted quartiles, the population proportion is by construction very near 25 percent in all cases, but does vary slightly since the sample restrictions are imposed after the quartiles are defined.) The final column shows the difference in population proportion between each incorrectly weighted quartile and the corresponding weighted quartile. Importantly, these differences are less stark in the later NLSY cohort than in the earlier cohort. In the earlier NLSY cohort, the bottom incorrectly weighted quartile was a more negatively selected sample than

[10]I am grateful to Martha Bailey for sending the data and code to replicate these figures.

in the later cohort and, correspondingly, the top incorrectly weighted quartile was a less positively selected sample. Since household income and educational attainment are positively correlated, using the incorrectly weighted quartile definitions thus understates educational gains between the two NLSY cohorts at lower income levels and overstates these educational gains at higher income levels.

Table 2.6: *Incorrectly Weighted and Weighted Income Quartiles*

	A. NLSY79 Cohort				
	Incorrectly Weighted Quartiles		Weighted Quartiles		Difference
	Mean Income (S.D.)	True Population Proportion	Mean Income (S.D.)	True Population Proportion	True Population Proportion
Bottom Quartile	15,213 (6,967)	14.4%	21,937 (9,776)	24.7%	-10.3%
2nd Quartile	38,718 (7,512)	23.6%	50,453 (7,513)	24.8%	-1.2%
3rd Quartile	65,274 (8,043)	28.0%	75,670 (7,589)	25.9%	2.1%
Top Quartile	117,902 (40,273)	34.0%	130,923 (40,399)	24.6%	9.4%
	B. NLSY97 Cohort				
	Incorrectly Weighted Quartiles		Weighted Quartiles		Difference
	Mean Income (S.D.)	True Population Proportion	Mean Income (S.D.)	True Population Proportion	True Population Proportion
Bottom Quartile	13,653 (7,719)	17.7%	18,022 (9,727)	24.41%	-6.7%
2nd Quartile	39,841 (8,084)	25.2%	47,630 (7,911)	26.33%	-1.1%
3rd Quartile	68,129 (9,228)	28.1%	75,833 (9,521)	25.11%	3.0%
Top Quartile	142,528 (69,128)	29.0%	152,996 (71,234)	24.16%	4.8%

Sources: NLSY79 and NLSY97.
Notes: This table shows the mean family income (in contemporary dollars) in the base year for the incorrectly weighted and weighted quartiles, respectively. Panel A corresponds to the NLSY79 cohort (i.e. 1979 family income); Panel B to the NLSY97 cohort (i.e. 1997 family income). The sample restrictions from BD as detailed in text are used throughout. The population proportion corresponds to the estimated percentage of the population based on sampling weights that falls in that "quartile." The final column reports the difference between the incorrectly weighted and the weighted population proportions for each quartile. Population standard deviations are in parentheses.

Panel B of Figure 2.3 shows college entry and completion rates by family income quartile and NLSY cohort when the quartiles are defined using sampling weights. Compared to Panel A, the purported rise in the gradient of education with respect to income is much less stark. The lack of sampling weights in BD's quartile definitions had the effect of understating the educational gains in the second and third quartiles and overstating the gains in the top quartile.

Table 2.7 compares these entry and completion rates by cohort for incorrectly weighted and weighted quartiles. Panel A reports entry rates, Panel B completion rates. The first four columns correspond to the entry and completion rates in Figure 2.3, each with the corresponding population standard deviation. The next two columns show the estimated growth in educational attainment between the earlier and later NLSY cohorts for each of the incorrectly weighted and weighted samples, respectively. The final column reports the bias on estimated growth in educational attainment due to incorrectly weighted quartiles. Consistent with the plots in Figure 2.3, the most striking bias for both entry and completion measures is the upward bias on growth in the top quartile.[11] incorrectly weighted quartiles also downwardly bias estimated growth in educational attainment within the second quartile (and, in the case of college completion, also within the third quartile). With sample-weighted quartile definitions, the estimated rise in educational attainment from across the NLSY cohorts is more equal across income quartiles.

[11]Incorrectly weighted quartiles do also result in overstating growth in the *bottom* quartile, though the magnitude of the bias is smaller.

Table 2.7: *Incorrectly Weighted Quartiles and Estimated Educational Attainment*

A. College Entry by Age 19

	NLSY79		NLSY97		(NLSY97 - NLSY79)		
	Incorrectly Weighted	Weighted	Incorrectly Weighted	Weighted	Incorrectly Weighted	Weighted	Diff.
Bottom Quartile	0.186	0.235	0.292	0.309	10.6%	7.4%	3.2%
	(0.016)	(0.014)	(0.023)	(0.020)			
2nd Quartile	0.319	0.339	0.468	0.532	14.9%	19.3%	-4.4%
	(0.018)	(0.019)	(0.023)	(0.023)			
3rd Quartile	0.379	0.432	0.604	0.650	22.5%	21.8%	0.7%
	(0.019)	(0.020)	(0.022)	(0.023)			
Top Quartile	0.582	0.619	0.803	0.801	22.1%	18.2%	3.9%
	(0.018)	(0.021)	(0.018)	(0.020)			

B. College Completion by Age 25

	NLSY79		NLSY97		(NLSY97 - NLSY79)		
	Incorrectly Weighted	Weighted	Incorrectly Weighted	Weighted	Incorrectly Weighted	Weighted	Diff.
Bottom Quartile	0.049	0.074	0.092	0.109	4.3%	3.5%	0.8%
	(0.009)	(0.009)	(0.015)	(0.015)			
2nd Quartile	0.138	0.155	0.208	0.243	7.0%	8.8%	-1.8%
	(0.014)	(0.014)	(0.020)	(0.020)			
3rd Quartile	0.173	0.207	0.317	0.377	14.4%	17.0%	-2.6%
	(0.015)	(0.017)	(0.021)	(0.024)			
Top Quartile	0.357	0.403	0.545	0.540	18.8%	13.7%	5.1%
	(0.017)	(0.021)	(0.023)	(0.025)			

Sources: NLSY79 and NLSY97.
Notes: This table shows the college entry and completion rates by NLSY cohort and quartile for the incorrectly weighted and weighted quartiles. Panel A reports college entry rates, Panel B college completion rates. The first four columns correspond to the entry and completion rates in Figure 2.3, each with the corresponding population standard deviation. The next two columns show the estimated growth in educational attainment between the earlier and later NLSY cohorts for each of the incorrectly weighted and weighted samples. The final column shows the difference in growth with weighted versus incorrectly weighted quartiles. The sample restrictions from BD as detailed in text are used throughout. Population standard deviations are in parentheses.

Panel A of Table 2.8 shows entry and completion rates over time by a continuous measure of BD's SES proxy: z-score of baseline income. The results in the first column suggest that the gradient of education with respect to family income in the base year was if anything declining over this period.

Table 2.8: *Gradient of Education with Respect to Base Year Family Income*

	A. Pooled		B. By Gender	
	Entry	Completion	Entry	Completion
NLSY97	0.1567***	0.0973***	0.1214***	0.0545***
	(0.014)	(0.013)	(0.021)	(0.018)
SES	0.1630***	0.1439***	0.1692***	0.1266***
	(0.010)	(0.010)	(0.014)	(0.015)
SES x NLSY97	-0.0298**	-0.0087	-0.0258	0.0155
	(0.013)	(0.014)	(0.019)	(0.019)
Female			0.0315*	-0.0077
			(0.019)	(0.016)
Female x NLSY97			0.0704**	0.0842***
			(0.029)	(0.026)
Female x SES			-0.0120	0.0371*
			(0.020)	(0.020)
Female x SES x NLSY97			-0.0085	-0.0511*
			(0.027)	(0.027)
Constant	0.4093	0.2122	0.3935	0.2167
	(0.009)	(0.008)	(0.013)	(0.011)
Observations	5,755	5,755	5,755	5,755
R-squared	0.1146	0.1203	0.1186	0.1239

Sources: NLSY79 and NLSY97.
Notes: This table shows the college entry and completion rates by NLSY cohort and SES, where SES is proxied for with the z-score of baseline family income. Panel A reports overall; Panel B reports by gender. Sample restrictions from BD are imposed as detailed in text.

2.4.2 Implications of Permanence of SES Measure

Recall that in BD, quartiles are defined based on family income in the base year. A similar proxy for socioeconomic status is also used in Belley and Lochner (2007), who proxy with family income when the youth were 16 to 17 in the NLSY79 and with Wave 1 income and Wave 1 net family wealth in the NLSY97. Consistent with BD, Belley and Lochner (2007) document a "dramatic increase in the effects of family income on college attendance from the NLSY79 to the NLSY97". I show that defining socioeconomic status with more permanent measures reduces the perceived increase in the gradient.

Figure 2.4 shows college entry and completion for each NLSY cohort by SES quartile. Sampling weights are used appropriately in defining quartiles (and in computing within-quartile means). In each column, SES quartile is computed based on different proxies. Moving from left to right in the figure, the proxies become arguably more permanent and less noisy measures of SES. To isolate the effect of SES measure, I throughout restrict to the sample definitions in BD (i.e. 15 to 18 years of age in the baseline survey).

Following BD, the SES measure in Panel A is simply household income in the base year (1979 for the NLSY79 cohort and 1997 for the NLSY97 cohort).[12] In Panel B, the SES measure is the average of the z-scores of household income for each year the respondent was 18 years of age or younger (i.e., a more permanent measure of income). In Panel C, the measure is the average of the z-scores of each parent's educational attainment, where available. Finally, in Panel D the SES measure is the average of the z-scores of these measures of income and parental educational attainment. Relative to each of the more permanent measures (Panels B, C, and D), the single-year income measure (Panel A) appears to overstate the rise of educational inequality. This is particularly true for college entry rates, which are plausibly more impacted by a particular year's income when the child is 15 to 18 years of age, but holds also for completion rates.

[12]Panel A of Figure 2.4 thus replicates Panel B of Figure 2.3.

Figure 2.4: *Gradient of Education by SES Proxy*

Sources: NLSY79 and NLSY97.
Notes: This figure shows college entry by 19 (row 1) and completion by 25 (row 2) for different SES proxies. Throughout, I impose the sample restrictions in BD: only those born 1961 - 1964 or 1979 - 1982 and for whom education is observed at 19 and 25. Panel A uses BD's proxy of family income in the base year. In Panel B, the SES measure is the average of the z-scores of household income for each year the respondent was 18 or under. In Panel C, the measure is the average of the z-scores of each parent's educational attainment, where available. And in Panel D, the SES measure is the average of the z-scores of the income and parental educational attainment measures. Sample weightings are used both in quartile definitions and in computing means.

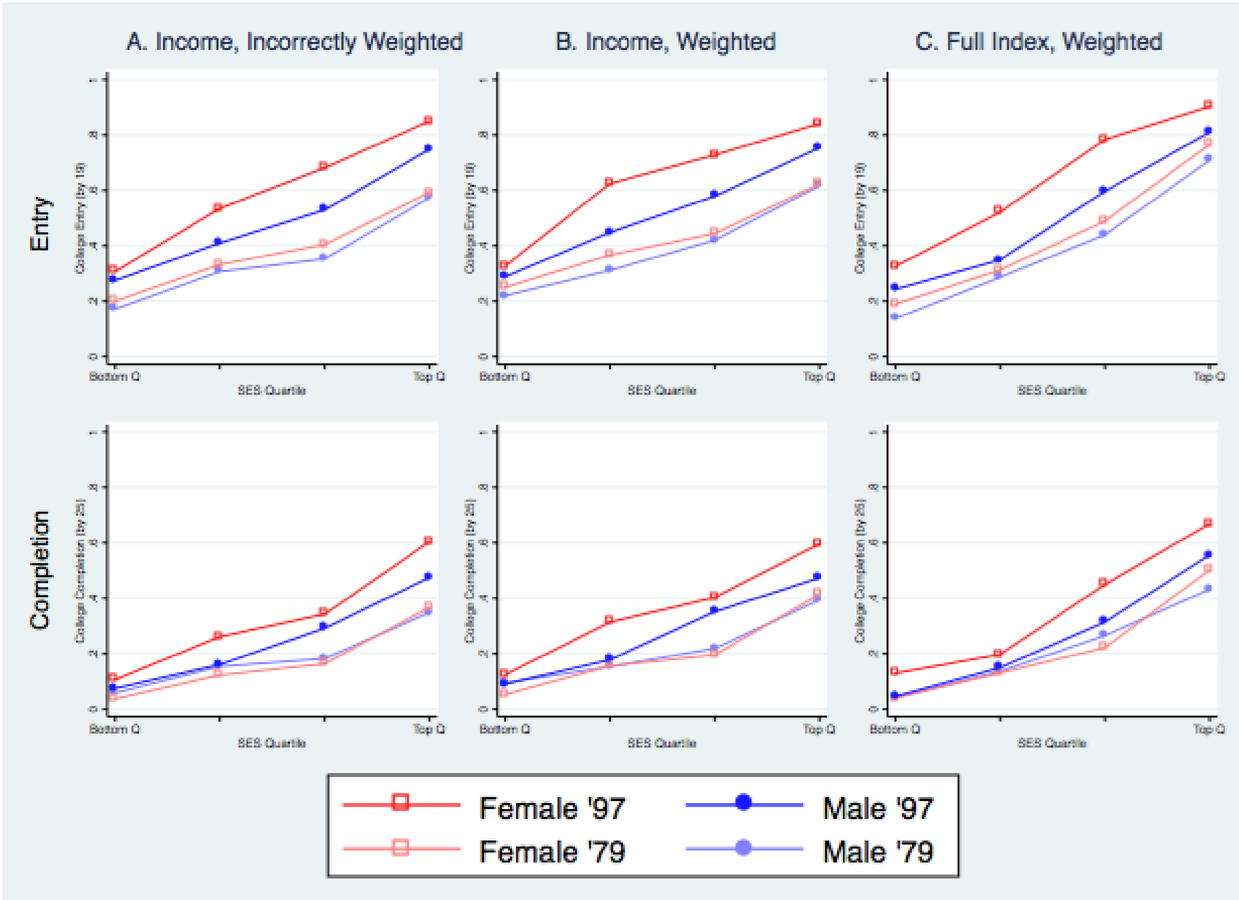
2.4.3 Female Educational Attainment by Weighting and SES Measure

In Section 2.3, I presented evidence that the increase in the female advantage between the earlier and later NLSY cohorts occurred relatively similarly throughout the SES distribution. I here return briefly to these findings to reconcile them with BD, which posits, "... (T)he increase in inequality is largely driven by the increase in college enrollment and completion among females from higher-income families."

Figure 2.5 shows college entry by 19 (row 1) and completion by 25 (row 2) by gender and NLSY cohort for each SES quartile. For comparison, I again impose the sample restrictions as in BD. Panel A uses BD's proxy of family income in the base year when quartiles are defined as in BD, i.e. *without* sampling weights. (Subtracting female from male *completion* rates in Panel A for each cohort yields the contents of Figure 5 in BD.) Panel B corresponds to the same, but with quartiles defined using sampling weights. As in the pooled analysis, when sampling weights are used in defining quartiles, growth in educational attainment is smaller in the top quartile, both for males and for females. It is also larger in the second quartile, with females in that quartile pulling away from their male peers and from both genders in the earlier cohort. Panel B of Table 2.8 shows the corresponding regression results for a continuous measure of SES: z-score of family income in the base year. The coefficients on the triple interaction of an indicator for female, an indicator for the NLSY97 cohort, and this continuous SES measure are negative, suggesting that across the distribution, the gradient of education with respect to family income if anything increased by *less* for women than for men over this period.

In Panel C, the SES measure is the average of the z-scores of the income and parental educational attainment measures (and sample weightings are used in defining quartiles). Here, as in Panel D of Figure 2.4, the rise in educational attainment – and in the female advantage in educational attainment – looks quite similar in each quartile. Whereas in Panel A (and Figure 5 in BD), the growth of the female advantage in education appears to stem from females in high-income families pulling away from their male peers.

Figure 2.5: *Female Educational Advantage by Weighting and SES Proxy*



Sources: NLSY79 and NLSY97.
Notes: This figure shows college entry by 19 (row 1) and completion by 25 (row 2) by gender and NLSY cohort for different SES proxies. Throughout, I impose the sample restrictions in BD: only those born 1961 - 1964 or 1979 - 1982 and for whom education is observed at 19 and 25. Panel A uses BD's proxy of family income in the base year where quartiles are defined *without* sampling weights. Panel B corresponds to the same, but defining quartiles with sampling weights. In Panel C the SES measure is the average of the z-scores of the income and parental educational attainment measures and sample weightings are used in quartile definitions. Sample means are used throughout in computing means.

2.5 Conclusion

This paper examines the growth in educational attainment between a cohort born in the early 1960's and one born in the early 1980's with an eye to differences by gender and socioeconomic status. I show that when proxying for SES with more permanent measures like parental educational attainment and household income over several years, the rise of the female educational advantage occurred quite similarly across the SES distribution. The increase in the female educational advantage over this period was also seen in both two-parent and single-parent households, and was if anything most pronounced among children raised in two-parent homes. The Hispanic population exhibited the smallest female educational advantage in each cohort, suggesting that its doubling over the period in question if anything tempered the growth of the female educational advantage.

I show that growth in educational attainment from the earlier to the later NLSY cohort occurred quite similarly throughout the socioeconomic distribution, pooling males and females. I reconcile my results with an existing paper that uses the same data but finds contradictory results. First, I demonstrate that Bailey and Dynarski (2011)'s failure to use sampling weights in defining income quartiles upwardly biases their estimated growth of educational inequality in this period. Sampling was not consistent between the two NLSY cohorts: the bottom incorrectly weighted quartile was a more negatively selected sample in the earlier cohort than in the later cohort and, correspondingly, the top incorrectly weighted quartile was a less positively selected sample. Since household income and educational attainment are positively correlated, failure to use sampling weights in quartile definitions mechanically understates educational gains between the two NLSY cohorts at lower income levels and overstates these educational gains at higher income levels. With appropriate weighting, the growth of educational attainment looked more similar across quartiles. Relative to proxying for SES with income in a single year, more permanent measures of socioeconomic status that include income in several years and/or parental educational attainment reveal a growth of educational attainment that was more equal still across SES quartiles.

Chapter 3

A Preference for Shared Experience: Network Externalities in Movie Consumption [1]

3.1 Introduction

We grab a bite at a bustling restaurant, and then kick back at a top-selling movie. *The popular choices are probably better*, we think. But is our crowd-following also driven by some preference for shared experience? In this chapter, we examine network externalities in consumption as a potential driver of choice convergence across individuals.

The tendency to follow in the footsteps of others has been observed in decisions ranging from which stock to buy to what books to read, how many children to bear, and whether or not to adopt a new technology. Much of the existing theoretical work on crowd-following focuses on the role of information; examples include models of

[1] Co-authored with Duncan Gilchrist

information cascades, observational learning, and social learning.[2] The exact mechanisms and contexts vary but, in brief, the individual is generally assumed to have imperfect information about the quality of a good or experience, and so relies on the observed choices and/or reports of others in making her own decision. In an array of settings, observational and experimental studies have found strong evidence of information stories driving convergent, or herd, behavior.[3]

Our interest lies in a very different potential explanation for crowd-following: network externalities in consumption. Amid classic network externalities (e.g., Becker (1991)), an individual's demand for a good is increasing in the total quantity demanded by others. A good is simply more useful, or an experience more enjoyable, the greater the number of others that share in it. Although network externalities and information stories can certainly coexist, there is no role for either quality or information about quality in network externalities themselves.[4]

Some goods have network externalities by construction. Snapchat, Facebook, Twitter, or Instagram, for example, is more useful the more peers already have it and are thus accessible through the application.[5] But even for goods with no obvious network externalities, a preference for shared experience could yield an individual demand function that is increasing in consumption by others. We explore network externalities in consumption of a major entertainment good, in-theater movies.

The thought experiment is simple: Holding all other characteristics of a movie fixed,

---

[2]Examples include Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), Ellison and Fudenberg (1995), McFadden and Train (1996), Bikhchandani, Hirshleifer, and Welch (1998), Çelen and Kariv (2004).

[3]See, for example, Scharfstein and Stein (1990), Welch (1992), Montgomery and Casterline (1996), Segrest, Domke-Damonte, Miles, and Anthony (1998), Bikhchandani and Sharma (2000), Hirshleifer and Hong Teoh (2003), Çelen and Kariv (2004), Munshi and Myaux (2006), Sorensen (2007).

[4]Choi (1997) provides an example of the coexistence of network externalilties and information stories.

[5]Relatedly, at the firm level Katz and Shapiro (1986) analyze technology adoption in the presence of network externalities.

is an individual's demand for the movie increasing in the number of others who have already seen it? Whether network externalities contribute to clustering in consumption of entertainment goods remains an open question. Moretti (2011), for example, finds strong evidence of social learning effects in this context, but no evidence of network externalities; (in our empirical specifications, we reconcile our findings with these).[6]

We focus on the relationship between opening weekend viewership of a movie and viewership of that movie in subsequent weekends. Amid network externalities in consumption, subsequent demand for a movie would be increasing in opening weekend viewership. But there are many other reasons to expect a positive correlation in a movie's viewership over time. At the most basic level, choices could be convergent across individuals over time simply because people have similar options, information, and preferences, and thus make similar decisions. We would, for example, expect an excellent movie to have higher viewership both this weekend and next if prospective viewers know it is excellent (and like excellent movies). Furthermore, even if people had different information about quality, an observational or social learning model could predict momentum. Quality and information about quality aside, viewership could also be correlated over time if people are subject to similar supply shocks (e.g., an unusually appealing movie trailer) or demand shocks (e.g., a close World Series that leaves people tied to the tube) over the course of the movie's run.

To isolate the role of network externalities, we exploit weather shocks in a movie's opening weekend as a plausibly exogenous source of variation in its opening weekend viewership.[7] In our first stage, we instrument for opening weekend viewership with

[6]In related work, Chen (2008) finds evidence of herd behavior in online book purchasing; Sorensen (2007) also tells an information story in the book market, identifying off of accidental placement on the *New York Times* bestseller list.

[7]In much of the existing literature analyzing motion picture demand, researchers deal with potentially confounding unobservables by conditioning on opening weekend audience size and then explore how things like reviews or awards shift the demand curve in later weeks (see, e.g., Prag and Casavant (1994), Mulligan and Motiere (1994), Sawhney and Eliashberg (1996), Nelson, Donihue, Waldman, and Wheaton

weather shocks during that weekend. Controlling for general seasonality, these unanticipated weather shocks are likely orthogonal to unobserved demand and supply shocks, and to movie quality. In our second stage, we estimate the effect of (instrumented) opening weekend viewership on viewership in later weekends. To account for seasonality in movie demand and supply, we throughout define viewership as audience size conditional on year, week of year, day of week, and holiday fixed effects; to account for any autocorrelations in weather, we also condition non-opening viewership on contemporaneous weather. To isolate opening weekend viewership shocks that are orthogonal to other potential demand or supply drivers, we focus on residualized viewership and instrument for opening weekend residualized viewership with plausibly exogenous weather shocks. Our second stage estimates are thus designed to capture *momentum from network externalities*, purged of potential confounders.

Using weather as an instrument is appealing in this setting – and potentially in many others – because weather is both unpredictably variable and because it has real effects on behavior. Instrumenting with weather effectively, however, is non-trivial in part because the set of potential weather measures is large.[8] Concern about either over-fitting the first-stage (e.g., by including all potential instruments) or data mining (e.g., by hand-picking some instruments and excluding others *ad hoc*) make careful aggregation and variable selection methods key in this context. To select from among a large set of potential weather instruments, we follow Belloni, Chernozhukov, and Hansen (2010) and implement Least Absolute Shrinkage and Selection Operator (LASSO) methods. We leave the details for our empirical section but, in brief, we run a penalized least-squares

(2001), and Moul (2007)). In an insightful twist, Moretti (2011) uses the number of opening theaters as a proxy for expected demand and shows differential momentum from positive and negative shocks to movie-going as evidence of social learning about film quality. Such approaches cannot, however, speak to network externalities.

[8]Consider a simple Google search of "02138 weather," which yields a deceptively simple-looking widget with a wealth of information including Cambridge's hourly maximum temperature, probability of precipitation, humidity, wind speed, and cloud cover.

regression of the first stage outcome on a large set of potential instruments and ask for machine selection of the instrument(s) that are sufficiently explanatory to be included in the first stage.

We find strong evidence of large and persistent momentum from random shocks to opening weekend viewership. For 100 weather-induced additional viewers opening weekend, we observe about 51 additional viewers in the second weekend and 27 the third. By the sixth week, cumulative momentum has yielded more than one additional subsequent viewer for each additional viewer during opening weekend.

Though our empirical strategy is designed to isolate momentum from network effects, we perform several tests of alternative explanations. We find that potential supply responses, such as adjusting the number of screens on which the movie shows or changing its duration in theaters, can explain little, if any, of our estimated momentum. Nor do we find any evidence that our estimated momentum is picking up information effects: the magnitudes and persistence of our estimates vary neither with movie quality nor with the level of ex-ante uncertainty about movie quality.

Our network externalities story is not dissimilar to the social influence model in Young (2009) in which people buy when enough other people have already bought. In related experimental research, Bursztyn, Ederer, Ferman, and Yuchtman (2013) show that social utility plays an important role in stock picking; an individual's utility from owning a stock depends directly on the possession of that stock by another individual. And DellaVigna, List, Malmendier, and Rao (2014) demonstrate that preference for shared experience may underlie the decision to vote: since voting is motivated in part by pride from telling others, people appear to be more likely to vote when they believe others will subsequently ask them whether or not they voted.[9]

[9]We think of the estimated momentum as reflecting a preference for shared experience (e.g., utility derived from being able to discuss the movie with peers). Although we cannot rule out a role of pure conformity in which choices are influenced by a preference for social esteem (as in Bernheim (1994)), we do not think a preference for social esteem would play much role in the decision to see a given movie, in

Preference for shared experience has also been studied in the sociology and psychology literature, with some exploration into demographic variation in the strength of those preferences. Survey and experimental work has found, for example, that females may have stronger preferences than males for shared experience (e.g., Barker (2009), Huberman and Rubinstein (2000), Clancy and Dollinger (1993)). Since different movies appeal to different demographics, we can apply our main empirical strategy to explore in this real and large market demographic heterogeneity in preference for shared experience. Our results suggest stronger momentum from network externalities among females than males, and among youth than adults.

Finally, we ask whether the network-externality induced viewers are merely substituting across movies or whether they are, instead, substituting away from alternative (non-movie-going) activities. We find only small and statistically insignificant levels of substitution across movies. Instead, most of the observed momentum from network externalities arises from an arguably more dramatic choice: substitution across activities.

The remainder of the chapter proceeds as follows: Section 3.2 details our movie and weather data, and our procedure for aggregating weather to the national level. In Section 3.3 we describe our empirical approach, including our instrument selection methods, and present our first stage results. Our baseline estimates of momentum follow in Section 3.4. In Section 3.5 we demonstrate that any supply-side adjustments have little bearing on our estimates. Given the large observed quantity effects and the fixed-price nature of movies, this suggests demand shifts are at play. In Section 3.6 we show that our estimates are independent of both movie quality and the level of information about movie quality, suggesting an information story is not driving our results. Section 3.7 explores how the magnitudes of network externalities vary with gender and age. Finally, Section 3.8

particular since public perceptions about an individual's predispositions seem unlikely to be impacted notably by the individual's movie choices. A notable exception could be "rebel" viewerships, such as adolescent viewership of an R-rated movie, but such viewerships are by definition rare.

examines from where the network-induced viewers are coming, and shows that most are substituting not simply movies, but rather across activities and Section 3.9 concludes.

3.2 Measuring Movie-Going and Weather

As set-up for the empirical section that follows, in this section we describe our movie data and our nationally-aggregated weather measures.

3.2.1 Data on the In-Theater Movie Market

Box office data provide an excellent measure of a movie's total market in the weeks just following release (when a movie can generally be viewed exclusively in theaters).[10] Our box office data comes from BoxOfficeMojo, a reporting service owned by IMDB, and includes both consumption quantities and supply levels at the movie level. U.S. ticket sales are reported daily; the number of screens on which the movie shows is reported weekly. The latter facilitates an analysis of any supply shifts that might impact our observed quantity effects. This, combined with total ticket sales quantities, facilitates an isolation of demand shifts since ticket prices are generally fixed.

Our sample is comprised of movies wide-released in U.S. theaters between January 1, 2002 and January 1, 2012.[11] We track audience sizes during the six weeks following the

[10]Though a few distributors have tried experimenting with simultaneous release in theaters and in home video, the vast majority do not release on home video until months (usually three to four) after the end of the theatrical release. Additionally, although we do not observe viewership of pirated movies, as long as an individual's demand for the pirated version does not fall the more others have seen the movie in theaters, then at worst our estimated network externalities would be biased downward.

[11]We follow Einav (2007) and Corts (2001) in defining as "wide-released" any movie that ever showed on 600 or more screens, and omit from the sample the less than 1 percent of movies that never reached wide release. For the 20 percent of films in our sample that start with a limited release before reaching wide release, we again follow Einav (2007) in defining the wide release date as the first date on which the movie is shown on more than the maximum of 400 screens and 30 percent of the eventual maximal number of screens for that movie. Though box office data is available for earlier years, we focus on the post-2001 period because for earlier years most ticket sales data are reported only at the movie by week level.

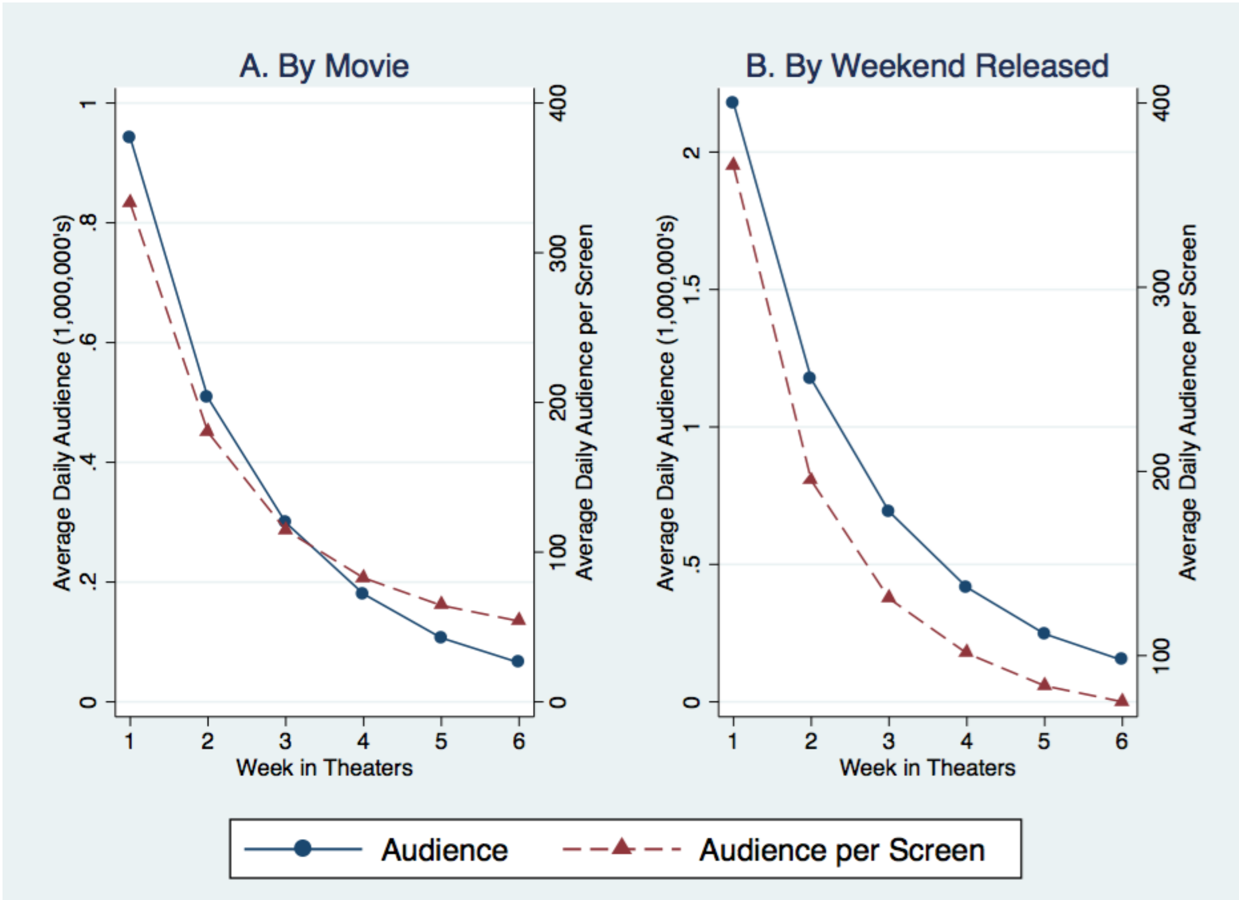
date of wide-release. To avoid truncation issues, the 19 percent of films that do not last at least six weeks in theaters are excluded from our main analysis. (We return to them when examining supply responses in Section 3.5 and show that our results are robust to their inclusion.) We focus throughout on weekend (Friday, Saturday, and Sunday) audiences since these are most responsive to weather shocks and weekend audiences account for the vast majority (over 75 percent) of ticket sales.[12]

Figure 3.1 shows average daily ticket sales and average daily ticket sales per screen for each of the first six weekends in theaters. Panel A plots averages across the 1,245 movies in our sample. Average daily ticket sales approach one million during opening weekend, but fall off quickly in subsequent weeks. The modal number of new movies per weekend is two, though some weeks have no new releases and others have as many as five. Since our weather instruments are at the daily level, in our analyses we group movies by the weekend on which they were released.[13] Our unit of observation for audiences, then, is at the opening weekend by date level. In our eleven-year sample we observe 538 opening weekends, or 1,614 opening weekend days. Panel B plots the average of daily ticket sales (and ticket sales per screen) at the release weekend level. The average audience for new releases is just over 2 million tickets. The corresponding number for movies in their second weekend is just over 1 million; this falls to 200,000 by the sixth weekend in theaters.

[12] In related work on movie audiences, Dahl and DellaVigna (2009) similarly restrict to weekend audiences.

[13] Almost all movies are released on Fridays; a few are released on Wednesdays. For Wednesday releases, we omit the first two daily observations, thereby treating the first Friday after opening as the opening date. Grouping by opening weekend is thus equivalent to grouping by opening date or by opening week.

Figure 3.1: *Average Audience Sizes by Week in Theater*



Notes: For our sample of 1,245 movies, in Panel A we plot average daily ticket sales (in 1,000,000's) and average daily ticket sales per screen for each of the first six weeks in theaters. In Panel B, we sum across movies released in the same weekend and report average daily ticket sales and average daily ticket sales per screen for each of the first six weeks after release. Here, and throughout our analysis, we restrict to weekend (Friday, Saturday, Sunday) audiences.

3.2.2 Nationally-Aggregated Weather Measures

Our nationally aggregated weather measures reflect the percentage of movie theaters in the country experiencing a particular type of weather. The raw data are from Weather Underground, a commercial weather service provider of real-time and historical weather information online; most U.S. data at Weather Underground comes from the National Weather Service. From Weather Underground, we observe daily weather measures for each of 1,941 U.S. weather stations.

We focus on four daily weather measures: maximum temperature, precipitation, and the interaction of temperature and precipitation.[14] To reduce the effect of possibly spurious outliers, we first winsorize our temperature and average hourly precipitation measures at the one-percent levels. Then, to facilitate national aggregation, we create temperature dummies in five degree bins and precipitation dummies in quarter-inch per hour bins. We also create indicators for any snow or any rain. Our motivation for dummying out before aggregating is perhaps best shown by example. Suppose the population lived in equal numbers in two cities, Los Angeles and Boston. On a particular summer day, Los Angeles had a maximum temperature of 105 degrees Fahrenheit (F) while Boston had a maximum of 55. If we aggregated nationally by simply taking the weighted average across cities, we might erroneously conclude that the country experienced a beautiful (80 degree) day when in fact half the country was cold and half was hot.

For each of our weather indicators (maximum temperature in five degree increments, average hourly precipitation in quarter-inch increments, any snow, any rain), we take weighted averages across weather stations. From the U.S. Census' annual ZIP Code Business Patterns data, we observe for each year from 2002 to 2011, inclusive, the number of theater establishments in each ZIP code.[15] We match each ZIP code (and all its movie-theater establishments) to the weather station that is nearest in great-circle distance to the ZIP code's center, conditional on that distance being no greater than 100 miles.[16] Weights

[14]We use maximum temperature (rather than minimum temperature) because we expect much of weather's impact on movie-going to be driven by its effect on alternative afternoon activities, and afternoons are generally the warmest time of day. Evening substitutes for movies are activities like dinners and indoor parties that are not heavily weather dependent. Afternoon substitutes like barbecues and pool-time, in contrast, are more weather dependent.

[15]Since the 2012 data was not available at the time of writing, we proxy for the 2012 establishment numbers with those from 2011. Though the "movie industry" spans across multiple six-digit NAICS codes, we include only establishments with NAICS code 512131, i.e."Motion Picture Theaters (except Drive-Ins)".

[16]For the years in our sample, less than 1 percent of establishments fall outside a 100 radius of any weather station.

are assigned to weather stations annually based on the percentage of total movie-theater establishments to which the weather station was matched.

3.3 Empirical Methodology

To isolate momentum from network externalities, as opposed to from unobservable movie quality or from other supply or demand shocks, we first instrument for viewership opening weekend with weather shocks that same weekend; we then estimate the effect of (instrumented) opening weekend viewership on viewership in subsequent weekends. This section details our empirical strategy, including our instrument selection methods, and presents our first stage results.

3.3.1 Estimating Momentum from Network Externalities: An IV Approach

Given the indoor nature of movie-going, it is perhaps not surprising that a day's weather is an excellent predictor of viewership. When it's beautiful out, there are generally fewer movie-goers; when the weather is less ideal, ticket sales tend to be higher. That is not to say, however, that the observed relationship is causal. As Einav (2007) demonstrates, the seasonality of viewership is driven by seasonality in both underlying demand *and*, since the supply side takes into account expected demand in timing releases, in the number and quality of movies available in theaters. Because seasonality is an important component of both the demand and supply, we throughout condition viewership on year, week of year, day of week, and holiday fixed effects and refer to the resulting residuals as "abnormal" viewership.

Denote the viewership on date t of movies that are in their jth-week of showing $v_{t,j}$. To compute abnormal viewership during opening weekend, we first regress viewership in opening weekend, $v_{t,1}$, on a constant and a vector of indicators for day of week, week of year, year, and holidays, which we denote F_t:[17]

$$v_{t,1} = \alpha_1 + F_t'\Phi_1 + \varepsilon_{t,1} \tag{3.1}$$

We call the resulting fitted, or predicted, values $\widehat{v_{t,1}}$ and define abnormal viewership opening weekend as the difference between realized and predicted viewership:

$$v\_abn_{t,1} \equiv v_{t,1} - \widehat{v_{t,1}} \tag{3.2}$$

We want to instrument for this abnormal viewership opening weekend with contemporaneous weather shocks. Given the natural (and anticipated) seasonality of weather, and to capture the unanticipated component, we throughout condition each of our weather measures on the same fixed effects as above. That is, for each weather measure w_k, $k \in \{1, \ldots, p\}$, we estimate

$$w_{t,k} = \delta_k + F_t'\Phi_k + \varepsilon_{t,k} \tag{3.3}$$

where t again indexes the date, k indexes the particular weather measure, and the fixed effects, F_t, are as defined in Equation 3.1. We call the resulting fitted values $\widehat{w_{t,k}}$ and define the weather shock $w\_shock_{t,k}$ as the difference between the realized and predicted weather measure:

$$w\_shock_{t,k} = w_{t,k} - \widehat{w_{t,k}} \tag{3.4}$$

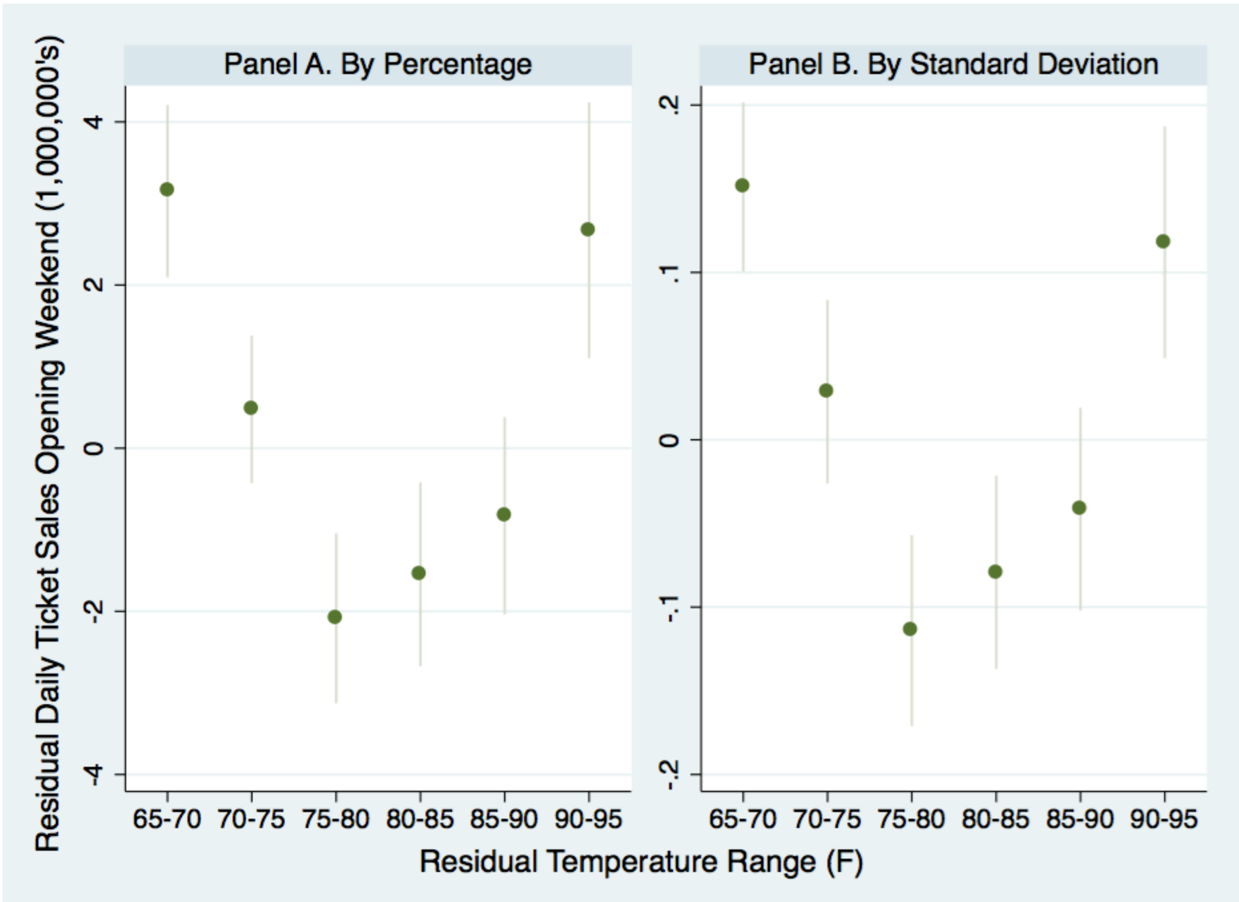
With our controls for seasonality and time trends in both weather and viewership, these weather shocks are plausibly orthogonal to movie characteristics as well as to other demand and supply shocks.

Figure 3.2 previews a simplified version of the relationship between weather shocks and abnormal viewership during opening weekend. Each coefficient is the result of a separate regression of abnormal viewership on contemporaneous weather shocks in

[17]Please see Appendix C.2 for the full set of holidays.

5-degree F bins. For exposition, we focus on the common summer range of 65 to 95 degrees F. Amid unexpectedly beautiful weather (70 to 90 degrees F broadly speaking, but most especially 75 - 80 degrees F), opening weekend ticket sales are lower than would be predicted by seasonality. Amid weather that is unexpectedly a bit too cool or too warm, in contrast, audiences are larger. Panel A shows the estimated magnitudes when weather is measured as the percentage of movie theaters in the given temperature range. The plotted coefficient represents estimated abnormal viewership when all (versus no) theaters are unexpectedly in that temperature range. When ten percent of theaters unexpectedly experience temperatures in the 65 to 70 degree range, for example, viewership rises by about 320,000 (or one-tenth of 3.2 million). To facilitate comparison of effect sizes across temperature ranges, Panel B shows the corresponding results when each weather shock is normalized to zero mean and unit variance. For a one standard deviation increase in the percentage of movie theaters unexpectedly in the the 65 to 70 degree F range, we estimate an additional 150,000 viewers of new releases. Relative to average daily ticket sales to all new releases, this corresponds to about a five percent (or 0.1 standard deviation) increase.

Figure 3.2: *The Effect of Weather Shocks on Viewership*



Notes: We plot the coefficient of the regression of abnormal viewership on each listed weather shock, along with the corresponding 95 percent confidence intervals. Each plotted coefficient is from a separate regression. Observations are at the opening weekend by date level (1,614 observations). National weather measures are as described in the text; in Panel A, these are measured in percentage of theaters in that temperature range, in Panel B they are normalized to zero mean and unit variance.

Although weather shocks are important predictors of abnormal viewership, the large number of potential weather shock specifications makes variable selection methods appealing. We detail our motivation and methods for instrument selection in the following subsection; for now let us take as given the machine-chosen instrument set, which we denote W^{LASSO}. To obtain the first stage, we run OLS on the LASSO-selected instrument(s):[18]

$$v\_abn_{t,1} = \eta + W_t^{LASSO\prime}\Omega + \varepsilon_{t,1} \tag{3.5}$$

We call the instrumented abnormal viewership $\widehat{v\_abn}_{t,1}$.

In the second stage, we estimate the relationship between this weather-induced abnormal viewership opening weekend and abnormal viewership in subsequent weekends. We define abnormal viewership in subsequent weekends as viewership conditional on year, week of year, day of week, and holiday fixed effects; given the potential for autocorrelation in weather shocks, we condition also on contemporaneous weather. That is, separately for each week $j > 1$, we first regress viewership on the set of fixed effects and contemporaneous weather:

$$v_{t,j} = \alpha_j + F_t'\Phi_j + X_t'\Gamma_j + \varepsilon_{t,j} \tag{3.6}$$

The fixed effects in F_t are as defined in Equation 3.1 and X_t denotes the vector of contemporaneous (date t) weather.[19] We call the resulting fitted values $\widehat{v_{t,j}}$ and define abnormal viewership in subsequent weekends as the difference between realized and predicted:

$$v\_abn_{t,j} \equiv v_{t,j} - \widehat{v_{t,j}} \tag{3.7}$$

Finally, to estimate the impact of abnormal viewership opening weekend on abnormal viewership $j-1$ weeks after opening, we run the the second stage separately for each $j > 1$:

$$v\_abn_{t,j} = \mu_j + \theta_j \widehat{v\_abn}_{t-7(j-1),1} + \varepsilon_{t,j} \tag{3.8}$$

The estimated momentum from network externalities in Week j is $\hat{\theta}_j$. Amid positive

---

[18]This is often referred to as post-LASSO and because of LASSO's shrinkage bias (i.e., the presence of the penalty in the LASSO optimization problem) it tends to perform better in terms of prediction and bias than does LASSO.

[19]X_t includes maximum temperature in ten-degree increments as well rain, snow, and average precipitation in quarter inches per hour.

network externalities, we expect positive (negative) exogenous shocks to viewership opening weekend to increase (decrease) viewership in later weeks.

3.3.2 LASSO Instrument Selection and First Stage Results

Instrumenting with weather is non-trivial in part because the set of potential weather measures is large. One approach might be to hand-pick a select set of instruments that together yield a strong first-stage and look “reasonable” ex post; with instruments and weighting chosen in such a seemingly arbitrary manner, however, robustness of the resulting estimates would almost certainly remain in question. A very different approach might be to include a wider array of instruments. The latter is the approach taken by Moretti (2011) in work closely related to ours. He instruments for opening weekend movie ticket sales with the maximum and minimum temperatures on opening day and the day prior to opening day for each of seven major metropolitan areas. This large set of correlated instruments, however, yields a weak first stage: the F-statistic on the excluded instruments is less than four in each of his specifications. The second stage results are neither statistically nor economically informative. It is in this context that Moretti finds no evidence of network externalities in movie-going.

Given the issues with either hand-picking a small number of instruments or including a large number of instruments, we find variable selections methods to be appealing. We follow Belloni, Chernozhukov, and Hansen (2010), and implement Least Absolute Shrinkage and Selection Operator (LASSO) methods to estimate optimal instruments in linear IV models with many instruments. In simulation experiments, the LASSO procedure performs well relative to recently advocated many-instrument robustness procedures (see Belloni, Chen, Chernozhukov, and Hansen (2012)). We here first provide a brief overview of our LASSO methods, drawing heavily from Chernozhukov and Hansen (2013), and then present the machine-chosen instrument sets and the corresponding first stages.

Consider our single endogenous variable, $v\_abn_{t,1}$, and our large set of potential weather instruments, $W_t = [w\_shock_{t,1}, \ldots, w\_shock_{t,p}]$. The econometrician would ideally select instruments by solving the standard ordinary leased squares problem subject to a binary penalty function that penalizes the inclusion of the instruments. That is, with n observations of the outcome, $v\_abn_{t,1}$, and the vector of potential instruments $W_t = [w\_shock_{t,1}, \ldots, w\_shock_{t,p}]$, we would minimize a binary integer criterion function:

$$\frac{1}{n}\sum_{t=1}^{n}[v\_abn_{t,1} - W_t'\beta]^2 + \lambda||\beta||_0, \text{ where } ||\beta||_0 = \sum_{k=1}^{p} 1\{\beta_{0k} \neq 0\}. \tag{3.9}$$

This is an NP-Hard problem so it is generally not tractable. The LASSO approach, originally due to Frank and Friedman (1993) and Tibshirani (1996), is to replace the L_0-norm in the problem above with the L_1-norm, thus minimizing

$$\frac{1}{n}\sum_{t=1}^{n}[v\_abn_{t,1} - W_t'\beta]^2 + \lambda||\beta||_1, \text{ where } ||\beta||_1 = \sum_{k=1}^{p} |\beta_{0k}|. \tag{3.10}$$

This problem is globally convex so it is straightforward to solve using conventional methods. Because the penalty function is kinked, the solution typically has many zeros; that is, the estimator only includes the set of covariates that are sufficiently explanatory to justify the penalty associated with their inclusion.[20] Bickel, Ritov, and Tsybakov (2009) show that the rate-optimal choice of the penalty parameter λ is

$$\lambda = 2\sigma\sqrt{2\log(pn)/n}, \tag{3.12}$$

where p is the number of potential instruments, n is the sample size, and σ is the standard

[20]LASSO is particularly appealing as a method for instrument selection in circumstances where the number of potential instruments is large, potentially even larger than the number of observations. Under regularity conditions, the rate of convergence is bounded by

$$||\hat{\beta}_{LASSO} - \beta_0|| \leq \sigma\sqrt{s\log(n \vee p)/n}, \tag{3.11}$$

which is close to the oracle rate $\sqrt{s/n}$. Notably, p only shows up through $\log(p)$.

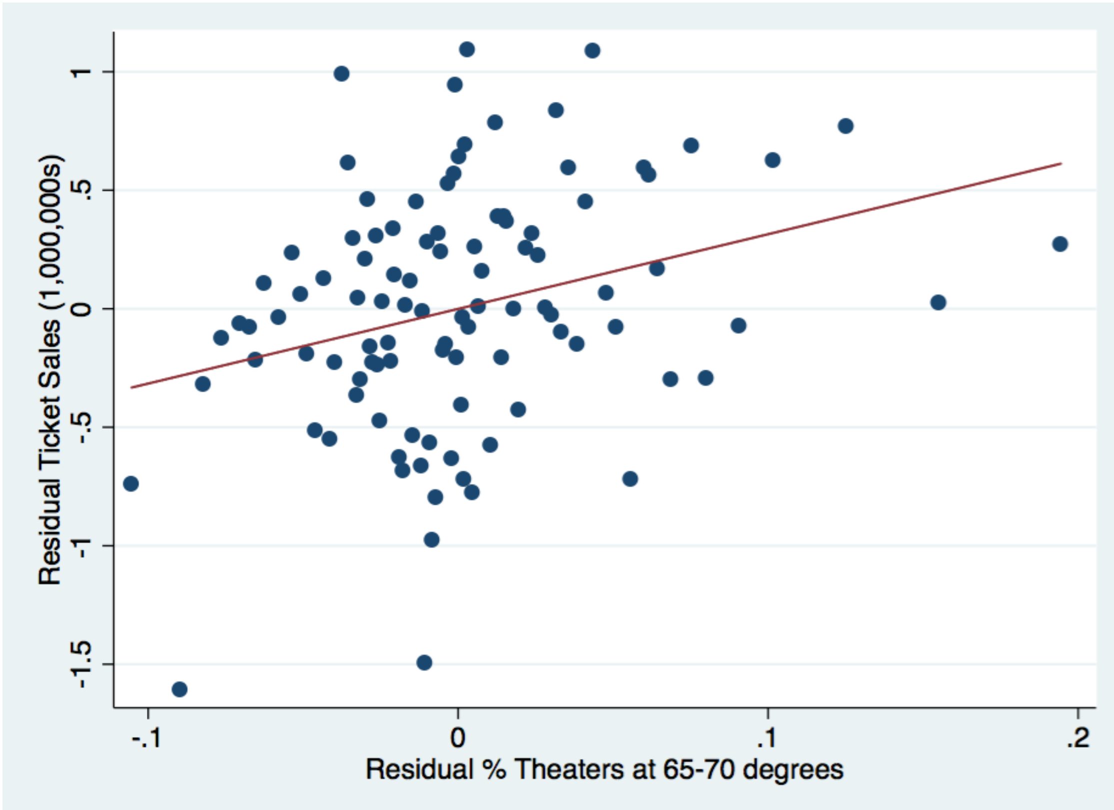
deviation of the residuals. σ is not known a-priori, so we estimate σ following the iterative methods of Belloni, Chen, Chernozhukov, and Hansen (2012). We follow Belloni, Chernozhukov, and Hansen (2011) in using conventional standard errors and also in adding a constraint on how many instruments are chosen.[21] We denote the final output of the LASSO methodology W^{LASSO}; this is the machine-chosen instrument set.

With the single-instrument constraint, the LASSO-chosen instrument is the 65 - 70 degree F measure. Figure 3.3 shows the corresponding first stage relationship in a binned scatterplot; the more theaters that are unexpectedly in this temperature range, the higher is abnormal viewership. Table 3.1 shows the first stage results from several different LASSO-chosen instruments sets. The first row corresponds to the LASSO specification above; ten percent more theaters unexpectedly in the 65 - 70 degree range (not quite warm enough for a barbecue) corresponds to about 320,000 additional daily viewership opening weekend (or about 15 percent of average daily viewership for new releases). For robustness, subsequent rows show the first stage results when we instead constrain to two or three instruments, or when we constrain to one instrument from among a choice set of ten degree temperature bins.[22]

[21]Conventional standard errors are fine as long as the number of selected instruments is not close to the sample size – see Belloni, Chernozhukov, and Hansen (2011) and Cattaneo, Jansson, and Newey (2012) for more detail. We probe the instrument constraint specification choice below and show that our results are robust to different instrument counts.

[22]In all cases we include our snow, rain, and average precipitation in quarter inches per hour variables in the set of potential instruments, though in our baseline specifications none is ever chosen by LASSO. We include these and the temperature variables both for Saturday and Sunday, but in our baseline specifications LASSO always chooses Sunday temperatures. This is consistent with a high volume of daytime (weather-dependent) movie-going on Sundays.

Figure 3.3: *First Stage Binscatter*



Notes: We plot the percentage of movie theaters with weather shocks in the 65 - 70 degree range against abnormal viewership. For exposition, the weather shock measure is grouped into 100 equal-sized bins; each point corresponds to the mean weather shock and abnormal viewership within a bin. The slope of the line of best fit is 3.151 with a standard error of 0.538.

Table 3.1: *LASSO-Chosen First Stages*

Set of Potential Instruments	Count Constraint	LASSO-Chosen Instrument(s)	Coefficient (s.e.)	F-Stat on LASSO Choice
5 Degree Temp Increments	Choose 1	65-70F	3.151*** (0.538)	34.37
	Choose 2	65-70F	3.431*** (0.547)	23.31
		90-95F	3.041*** (0.796)	
	Choose 3	65-70F	3.282*** (0.562)	16.61
		90-95F	2.565*** (0.840)	
		75-80F	-0.813 (0.540)	
10 Degree Temp Increments	Choose 1	60-70F	1.402*** (0.378)	13.78

Notes: This table presents first stage results for a variety of LASSO specifications. In the first three, the instrument choice set is as follows: national aggregates of maximum temperature indicators in 5 degree increments (on the interval [10F,100F]), indicator for snow, indicator for rain, precipitation indicators in 0.25 inches per hour increments (on the interval [0,1.5]), conditioned on year and week of year. From this set, the LASSO approach is set to choose one, two, or three instruments, respectively. In the fourth specification, a single instrument is again chosen, but the instrument choice is altered to instead include the analogous temperature measures in 10 degree increments. Observations are at the opening weekend by date level (1,614 observations). Standard errors, clustered at the date level, are in parentheses. */**/*** denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

3.4 Momentum from Exogenous Shocks to Viewership

In this section, we present our main results followed by a discussion of robustness and a test of exogeneity.

3.4.1 Main Results

Implementing the second stage (Equation 3.8), we find substantial momentum from exogenous shocks to opening weekend viewership. Table 3.2 presents our base case estimates. The first five columns report the relationship between abnormal viewership opening weekend and subsequent abnormal viewership, separately for each weekend, two through six; the final column reports the corresponding aggregate relationship, where the outcome is summed across those weekends. In the first row, we instrument for abnormal viewership during opening weekend with contemporaneous weather shocks. Since the first stage is generally stronger when temperature is more precisely defined and when the set of chosen instruments is kept small (see Table 3.1), for parsimony we focus on IV estimates derived from a single instrument chosen from among the set of five-degree temperature increments.

One-hundred additional viewers opening weekend yields an estimated 107 additional viewers at some point in the following five weekends. The observed momentum is largest in the weekend immediately following opening weekend: just over half of the total effect is realized in the second weekend; an additional quarter is realized in the third weekend. Though the magnitude of the effect falls off in subsequent weeks, it remains relatively large and statistically significant.

Table 3.2: *Momentum from Viewership Shocks*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
IV	0.511***	0.270***	0.130***	0.0968***	0.0586***	1.066***
	(0.0513)	(0.0379)	(0.0319)	(0.0230)	(0.0176)	(0.134)
OLS	0.434***	0.257***	0.156***	0.0987***	0.0636***	1.009***
	(0.0159)	(0.0108)	(0.00734)	(0.00515)	(0.00369)	(0.0400)
R-squared	0.690	0.581	0.457	0.374	0.284	0.628

Notes: The first row reports the results of the IV regression of daily abnormal audiences in each later weekend on daily abnormal audiences opening weekend; the second row reports the corresponding OLS results. Observations are at the opening weekend by date level (1,614 observations). Standard errors, clustered at the date level, are in parentheses. */**/*** denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively. National weather shock instruments are chosen using the LASSO approach described in the text; the first stage results are included in the first row of Table 3.1.

While the naive prior might be that OLS would tend to overestimate causal momentum from network externalities because other unobserved shocks would likely be positively correlated over time, our OLS estimates (presented in the second row) closely resemble our IV estimates. In some weeks, the OLS estimates lie slightly above, in other weeks slightly below, but in no week is the difference between the the IV and OLS estimates statistically significant. Given that our LASSO selection methods yield a single instrument with an F-statistic over 34, we are not concerned about weak instruments biasing our IV estimates upward.[23]

Two other factors instead likely contribute to the close alignment of our IV and OLS estimate. First, our OLS results are already purged of many major potential confounders. Since we have conditioned on year, week of year, day of week, and holiday fixed effects, we have already controlled for any aspects of quality, supply, and demand that would be captured in seasonality. Second, while our OLS estimates could be biased upward by

[23]Any inter-temporal substitution that might occur, for example if some people go to the movie on opening weekend because of the weather shock and subsequently do not go to that movie in later weekends, would if anything bias our estimates downward.

additional factors not captured by this set of controls, our IV estimates may still approach them in magnitude because each is identifying off of a different composition of viewers. Whereas the OLS estimates capture the average momentum effect across all abnormal viewers, our IV estimates pertain specifically to abnormal viewers *whose viewership choice was driven by a weather shock*. Amid homophily these marginal viewers – whose choice to see a movie was quite literally thrown to the wind – may be more likely to have friends who are also marginal movie-goers, suggesting that network externalities from their viewership could be stronger than network externalities from the average viewership. If marginal viewers are also more social (e.g., more likely to have alternative activities with friends as outside options to movies), the difference between network externalities from their viewership and from viewership by the average viewer could be larger still.

Our base case IV results are robust to different numbers of instruments chosen within LASSO and also to larger units of observation. Appendix Table C.1 shows the corresponding second stage results with alternative LASSO specifications. The effects are generally unchanged when LASSO is instructed to choose two, or even three, instruments (rather than just one). They are also comparable when the potential instrument set is altered to include temperature variables in broader (ten degree) increments. For robustness, in Appendix Table C.2 we also present the corresponding results when observations are defined at the opening weekend by weekend level (rather than opening weekend by day) level. Here, the estimated coefficients change only slightly and, although the standard errors are somewhat larger, estimated momentum from exogenous viewership shocks remains highly significant in each week.

3.4.2 Evidence on Exogeneity

Recall that we seek to estimate network externalities, i.e., how demand for a movie varies with how many others have seen the movie, all else equal. To isolate momentum arising out of preference for shared experience, we need the shocks off of which we are identifying to be orthogonal to all other demand drivers. Our intention in both (1) defining our endogenous regressor as abnormal audiences, and (2) instrumenting with plausibly exogenous weather shocks, is to isolate viewership shocks that are orthogonal to other potential demand-drivers like a movie's quality, distributor, or the intensity with which it was advertised. Table 3.3 shows that, consistent with the exogeneity of the opening weekend shocks, controlling for expected demand has little bearing on our results.

We follow Moretti (2011) in proxying for expected demand with the number of screens on which the movie opened. As Moretti notes, the number of screens is set by profit-maximizing theater owners who have strong incentives to accurately predict opening weekend demand; it should thus summarize well all the information the market has up to the release date about how well the movie will do. In the first row, we reproduce the results of our main specification for ease of comparison. The second row shows the results when adding in controls for the number of screens on which the movie opened. Controlling for expected demand, the estimated momentum falls only slightly and each week's estimates remain large and highly significant. In the third row, we define the outcome variable as abnormal viewership *per opening screen*. For comparison to our base case, in the final row we standardize the coefficients so that the first weekend's coefficient is one. Our estimates again fall only slightly (insignificantly) relative to the base case and remain large and statistically significant through the fourth weekend. Our second stage, then, indeed appears to be picking up viewership shocks orthogonal to expected demand.

Table 3.3: *Momentum per Opening Screen from Exogenous Viewership Shocks*

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Tickets (1)	1.000***	0.511***	0.270***	0.130***	0.0968***	0.0586***	1.066***
	(0)	(0.0513)	(0.0379)	(0.0319)	(0.0230)	(0.0176)	(0.134)
Tickets, Controlling for Opening Theaters (2)	1.000***	0.500***	0.252***	0.111***	0.0891***	0.0542**	1.006***
	(0)	(0.0635)	(0.0474)	(0.0402)	(0.0287)	(0.0218)	(0.167)
Tickets per Opening Theater	1.307***	0.664***	0.312***	0.145**	0.0747	0.0362	1.231***
	(0.244)	(0.134)	(0.0956)	(0.0650)	(0.0505)	(0.0377)	(0.336)
Standardized Tickets per Opening Theater (3)	1.000***	0.508***	0.239***	0.111**	0.0572	0.0277	0.942***
	(0.187)	(0.103)	(0.073)	(0.050)	(0.039)	(0.029)	(0.257)
Differences:							
(1) - (2)	--	0.011	0.018	0.019	0.008	0.004	0.060
		(0.000)	(0.082)	(0.061)	(0.051)	(0.037)	(0.028)
(1) - (3)	--	0.003	0.031	0.019	0.0397	0.0309	0.124
		(0.115)	(0.082)	(0.059)	(0.045)	(0.034)	(0.290)

Notes: This table presents results from three different IV specifications. The first is our base case from Table 3.2; in the second row, controls for number of opening theaters (proxy for expected demand) are added in the second stage; in the third row, the outcome variable is defined as abnormal viewership per opening screen. Observations are at the opening weekend by date level (1,614 observations). Standard errors, clustered at the date level, are in parentheses. */**/*** denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively. National weather shock instruments are chosen using the LASSO approach described in the text. The first stage results are included in the first row of Table 3.1.

3.5 A Role for Supply Shifts?

In the preceding section, we demonstrated that positive (negative) exogenous shocks to viewership opening weekend cause positive (negative) shocks to viewership in subsequent weekends. Our interest lies specifically in isolating the momentum effects of network externalities, a demand-side phenomenon. In this section, we present a brief overview of the supply side of the market and test for any supplier responses to our weather-induced viewership shocks.

3.5.1 In-Theater Movie Supply: Institutional Background

The three major categories of players on the supply side are the producers, the distributors, and the exhibitors. In brief, the producer makes the movie, the distributor decides when and how it gets released, and the exhibitor shows the movie to audiences.[24] Distributors and exhibitors contract on where the movie will be shown and how the revenues will be shared.[25] After release, the supply-side has two major margins along which it can adjust: intensive-margin adjustments occur when an exhibitor changes the number of screens on which the theater shows the film; extensive-margin adjustments occur when a distributor withdraws the movie from all exhibitors altogether (often replacing it with a new and different movie).[26]

Distributor-exhibitor contracts are designed to discourage screen adjustments by the exhibitors. Since revenues tend to drop sharply after the initial few weeks, exhibitors usually prefer shorter tenures, all else equal; having paid high fixed costs upfront, however, distributors prefer longer tenures. Some contracts thus require that the exhibitor play the film for a minimum number of weeks. More commonly, however, the exhibitor can drop the movie at will at any point after release, but is discouraged from doing so by both reputational and monetary considerations. Reputationally, an exhibitor that drops a movie early may have reduced access to future movies from that distributor.

---

[24]Major studios increasingly both produce and distribute themselves.

[25]Most commonly, exhibitors pay upfront both some advance to the distributor for the movie and their own direct-to-local-consumer advertising costs, and contracts are usually made well in advance of the release date, in part to give exhibitors time to advertise the movie to local audiences.

[26]Though advertising has also been found to play an important role in audience size (see, e.g., Prag and Casavant (1994), Zufryden (1996), Elberse and Eliashberg (2003)), prior work suggests that potential post-release adjustments in advertising intensity are relatively small. First, advertising budgets are generally set before a movie's production, thereby limiting scope of adjustment (Moul (2007)). Second, the vast majority (ninety percent) of a movie's advertising budget is already spent pre-release, limiting post-release adjustments further still (Elberse and Anand (2007), Vogel (2011)). Moretti (2011) also shows empirically that the endogenous response of advertising to surprise sales is small both because (1) "only a small amount of advertising is at risk of being affected by the surprise" and (2) "the elasticity of advertising to first-week surprise is small."

Monetarily, the trajectory of the revenue split is designed to incentivize exhibitors to keep the movie up: for a major motion picture, for example, it is common for just ten percent of the first three week's revenues (net exhibitor overhead costs) to go to the exhibitor, but thereafter the exhibitor's share rises dramatically to fifty or even seventy percent.[27] Despite institutional factors that incentivize exhibitors to stay the course, there is some evidence (as in Krider, Li, Liu, and Weinberg (2005)) that exhibitors do monitor box office sales and respond with screen allocation decisions. There is also evidence (as in Elberse and Eliashberg (2003)) that the number of screens showing the film in a given week influences that week's audience sizes.

3.5.2 Testing for a Supply-Side Response

In the following analysis, we first show that both intensive and extensive margin responses to our opening weekend viewership shocks are rare. We then show that accounting for any such responses has little effect on our estimated momentum.

Panel A of Table 3.4 reports estimated supply responses to our weather-induced viewership shocks. The first row shows the intensive margin response by exhibitors, i.e., the relationship between abnormal viewership opening weekend and the number of screens on which the movie is shown each week. The second row shows the corresponding extensive margin response by distributors, i.e., the probability in each week that the movie is withdrawn from theaters. The empirics are loosely similar to those of our main analysis, but differ in three key ways. First, since supply changes occur at most weekly, observations are at the opening weekend by weekend level (with abnormal viewership summed across weekend days).[28] Second, while the endogenous regressor continues to

[27]According to Moul (2007), for less major movies, a common rental schedule is 60 percent of net opening week revenues to the distributor, then 50 percent the second week, 40 percent the third, 35 percent the fourth, and 30 percent thereafter.

[28]For additional discussion of the supply change decision see Swami, Eliashberg, and Weinberg (1999);

be abnormal viewership opening weekend, the outcome variable is abnormal number of screens (Row 1) or abnormal probability of being withdrawn (Row 2). (Each of these is similarly conditional for year, week of year, and holiday fixed effects.) Third, viewership is measured in 10,000's for ease of exposition.

Before examining supply-side responses, we first address the relationship between viewership shocks opening weekend and the number of screens on which the movie opens. The Week 1 estimates suggest that amid weather shocks, movies opening on more screens may experience slightly (insignificantly) larger viewership shocks. One possible explanation is that consumers are responding to availability.[29] The point estimate, however, is statistically insignificant and small in magnitude; it suggests that 10,000 additional viewers from weather shocks corresponds to 2.26 additional screens. On average for new releases in our sample, each screen has about 370 viewers such that the additional screens would mechanically explain just 8 percent of the viewership opening weekend.

The similarly small and generally insignificant coefficients in subsequent weekends, moreover, indicate that exhibitors do *not* respond to positive (negative) shocks to opening weekend viewership by substantially increasing (decreasing) the number of screens on which the movie shows. The magnitudes of the point estimates suggest that any intensive-margin adjustments can explain little, if any, of the observed momentum. Movies that sold an additional 10,000 tickets opening weekend showed on just 1.44 screens second weekend; they also sold about 5,100 more tickets that weekend (see Table 3.2). Since the average tickets sold per screen for second weekend showings is 200, the additional

any change in number of screens of withdrawal of movies each week almost always occurs on the first day of the movie-industry week (Friday).

[29]Suppose that the weather incentivizes additional people to go to the movies, but that the decision then of which movie to attend is a function in part of the convenience of the available showings. A movie opening on more screens could be more likely to be showing at a convenient time and/or place, suggesting a potentially positive relationship between number of screens and (instrumented) opening weekend audiences.

screens would mechanically account for less than 6 percent of the observed viewership effect that weekend. The relative size of the mechanical effects are similarly small in other weeks, ranging from at most 12 percent (Week 4) to just 2 percent (Week 6).

To estimate any extensive-margin response from suppliers, we add truncated movies back into our sample. The reported IV estimates in the second row show that the relationship between abnormal audiences (in 10,000's) and the abnormal probability in each week of being withdrawn from theaters is similarly weak. The estimates suggest that weather-induced viewership shocks do not effect withdrawal probabilities in the short run. Though still statistically insignificant, point estimates in later weeks do suggest that movies with positive (negative) abnormal viewership opening weekend could be slightly less (more) likely to be withdrawn.

Table 3.4: *Supply-Side Adjustments*

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
			A. Supply-Side Adjustments				
Number of Screens	2.264	1.443	1.658	1.626	1.984**	1.086	7.797
(obs. 538)	(1.914)	(1.907)	(1.535)	(1.208)	(0.971)	(0.875)	(5.377)
Probability Dropped	--	3.51e-05	-0.000243	-0.000698	-0.00107	-0.00217	-0.00225
(obs. 557)	--	(0.000136)	(0.000310)	(0.000536)	(0.000780)	(0.00150)	(0.00165)
			B. Main Effects with Intensive-Margin Supply Adjustments (1,614 obs.)				
Tickets	1***	0.511***	0.270***	0.130***	0.0968***	0.0586***	1.066***
	(0)	(0.0513)	(0.0379)	(0.0319)	(0.0230)	(0.0176)	(0.134)
Tickets per Screen	1.307***	0.674***	0.358***	0.197***	0.125**	0.0809**	1.435***
	(0.244)	(0.132)	(0.0893)	(0.0610)	(0.0485)	(0.0376)	(0.309)
Standardized Tickets per Screen	1***	0.516***	0.274***	0.151***	0.0956***	0.0619**	1.098***
	(0.187)	(0.101)	(0.068)	(0.047)	(0.037)	(0.029)	(0.236)
			C. Main Effects with Supply Adjustments, Includes Dropped Movies (1,671 obs.)				
Tickets	1***	0.552***	0.273***	0.0937*	0.0818**	0.0380	1.133***
	(0)	(0.0750)	(0.0546)	(0.0505)	(0.0351)	(0.0274)	(0.176)
Tickets per Screen	1.291***	0.700***	0.335***	0.0888	0.0153	-0.110	1.490***
	(0.313)	(0.171)	(0.126)	(0.102)	(0.0870)	(0.0922)	(0.396)
Standardized Tickets per Screen	1***	0.542***	0.259***	0.0688	0.0119	-0.0852	1.154***
	(0.242)	(0.132)	(0.098)	(0.079)	(0.067)	(0.071)	(0.307)

Notes: The first row of Panel A reports the results of IV regressions of abnormal viewership opening weekend (in 10,000's, summed across weekend days) on abnormal number of screens showing the movie each week in our main sample. The second row reports the results of IV regressions of abnormal audiences opening weekend (in 10,000's) on the abnormal probability of being dropped each week and includes truncated observations. The first stage results are included in Appendix Table C.3. In Panel B, the first row is simply our base case from Table 3.2, reproduced here for ease of comparison; the second results are from the same specification, but with the outcome variable defined as tickets (in 10,000's) per screen; the final row shows these results scaled down by the Week 1 coefficient so that the first weekend's results are standardized to one, and later weekends' results can be compared in magnitude to our base case estimates. The first stage results are included in the first row of Table 3.1. Panel C replicates Panel B but with the sample expanded to include truncated observations; (any dropped movie is assigned a ticket sales number of zero for that and all subsequent weekends). The first stage results are included in Appendix Table C.3. Throughout, standard errors, clustered at the weekend level, are in parentheses. */**/*** denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively. National weather shock instruments are chosen using the LASSO approach described in the text.

Despite the relatively small and generally insignificant relationships between viewership shocks opening weekend and either the number of screens showing the movie or the probability that the movie is dropped, in Panels B and C we for robustness show our main effects when accounting for any supply adjustments. Panel B accounts for intensive-margin adjustments only, before considering any extensive margin response; here, as in our main analysis, truncated films are not included. For ease of comparison, the first row simply reproduces our main estimates from Table 3.2; the second row shows the corresponding results when the outcome variable is instead defined as tickets (in 10,000's) per screen; the final row shows these results scaled down by the first weekend's coefficient so that later weekends' results can be compared in magnitude to our base case estimates. Comparing the first and final columns of Panel B, we find that intensive-margin responses can explain little, if any, of the observed quantity effects. For each weekend, the point estimates from our main specification and from the per screen specification (scaled) differ by at most 15 percent; in some weekends the per screen results are just above, while in other weekends they are just below, the momentum estimated in our main specification, and in no week is the difference statistically significant.

Although extensive-margin supply shifts also look small in magnitude relative to the total observed quantity effects, and those shifts are only observed weeks into the run (at which point the majority of the quantity effects have already been observed), we also present specifications and samples that account for both any intensive *and* any extensive margin supply adjustments. Panel C replicates Panel B with the sample expanded to include truncated films. Once closed, a movie is assigned an audience size of zero for all subsequent days. Relative to results on our main sample (Panel B), inclusion of movies with truncated demand slightly increases our estimates in early weekends and slightly decreases them in later weekends. This is consistent with the finding in Panel A that movies with positive (negative) abnormal viewership opening weekend may be slightly less (more) likely to be taken out of theaters. None of the differences, however, are significant. Moreover, accounting in this sample for any intensive-margin responses again increases estimated momentum slightly in some weeks and decreases it slightly in others, with the differences not significant in any week. Taken together, then, the results in Table 3.4 suggest that supply-side adjustments can explain little, if any, of the observed quantity effects.

3.6 A Role for Social Learning?

We have demonstrated a strong relationship between opening weekend abnormal viewership of a movie and abnormal viewership of that movie in subsequent weekends, even when the former was driven by exogenous shocks orthogonal to movie quality. We have also provided evidence that supply shifts on neither the intensive nor the extensive margins can explain much if any of the observed quantity effects. Given the fixed-price nature of in-theater movie-going, these findings suggest that demand shifts are driving the observed quantity effects. The particular nature of the demand shifts, however, remains important for interpretation. In this section, we examine whether an information story could be at play.

By instrumenting with shocks that are orthogonal to movie quality, we sought to isolate shocks to viewership opening weekend that were independent of quality. Those viewership shocks should thus in and of themselves provide no quality signal and should not induce quality updating among individuals considering attending the movie in later weeks. Nonetheless, we might wonder whether larger early viewership boosts later sales, in part because the individual either has, or believes that she has, better *information* on quality. First, if after having seen the movie, people disseminate information about whether it was good or bad, then the implications of that early viewership would vary with movie quality. Amid an information dissemination story like this, we would expect stronger momentum for higher-quality than lower-quality movies. Or second, in an observational learning story in which people with imperfect information infer movie quality from the observed movie-going of others, and are unable to identify perfectly the component of movie-going driven by exogenous (quality-independent) shocks, we would expect stronger momentum from an initial viewership shock for movies about which there was more ex-ante uncertainty about quality. In this section, we examine these two testable predictions and conclude that such information stories are not driving our observed demand shifts.

We first examine whether, consistent with the information dissemination story, our estimated momentum is stronger for higher-quality than lower-quality movies. We proxy for movie quality with ratings by expert reviewers. The ratings come from IMDB's Top-1000 voters, a group characterized by IMDB as "the 1000 people who have voted for the most titles in [their] ratings poll."[30] Movies in our sample have been rated on average by 483 of these Top-1000 voters. We cut movies into terciles by rating. Movies with a Top-1000 voter assigned average rating of 6.3 or above fall in the top third, while movies with a rating of 5.6 or below fall in the bottom third. Panel A of Table 3.5 shows momentum effects separately by high quality (top tercile) and low quality (bottom tercile) movies. Relative to movies with low ratings, movies with high ratings experience about the same momentum in early weeks and only slightly (insignificantly) more momentum in later weeks, suggesting that a social learning story is not driving our observed momentum.[31]

[30]IMDB notes that they "don't disclose the number of votes required for a person to make this list nor can [they] confirm or deny who is on the list."

[31]Appendix Table C.4 shows the corresponding OLS results.

Table 3.5: *Momentum by Movie Quality and Information about Movie Quality*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
			A. By Movie Quality			
High-Rated	0.426***	0.271***	0.154***	0.119***	0.0729***	1.044***
(obs. 705)	(0.0810)	(0.0609)	(0.0391)	(0.0315)	(0.0218)	(0.224)
Low-Rated	0.465***	0.245***	0.137***	0.0610**	0.0452***	0.953***
(obs. 825)	(0.0752)	(0.0538)	(0.0405)	(0.0244)	(0.0159)	(0.187)
Difference:						
(High - Low)	-0.039	0.026	0.017	0.058	0.028	0.091
	(0.111)	(0.081)	(0.056)	(0.040)	(0.027)	(0.364)
			B. By Information about Movie Quality			
High Budget	0.416***	0.209***	0.147***	0.0945***	0.0682***	0.935***
(obs. 744)	(0.0669)	(0.0442)	(0.0332)	(0.0257)	(0.0181)	(0.178)
Low Budget	0.457***	0.304***	0.137***	0.0658***	0.0324**	0.996***
(obs. 705)	(0.0648)	(0.0476)	(0.0313)	(0.0236)	(0.0160)	(0.162)
Difference:						
(High - Low)	-0.0410	-0.0950	0.0100	0.0287	0.0358	-0.0610
	(0.093)	(0.065)	(0.046)	(0.035)	(0.024)	(0.241)

Notes: Panel A replicates the IV results from Table 3.2 separately by for high versus low rated movies, defined as the top third and bottom third in ratings, respectively. Top-1000 voters are the 1000 people who have voted for the most titles in IMDB ratings polls; high rated here corresponds to 6.3 and above; low rated is 5.6 and below. The final column reports the differences in the point estimates. Panel B does the same separately for movies in the top and bottom third by production budget, respectively. The first stage results are included in Appendix Table C.3. The final column reports the differences in the point estimates. */**/*** denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively. The corresponding OLS estimates are in Appendix Table C.4.

Note that initial shocks to viewership also do not appear to vary with reviews. Appendix Table C.5 reports the number of voters and the likelihood of being characterized as high-rated and low-rated as a function of (instrumented) opening weekend ticket sales. First, weather-induced shocks to viewership do not significantly impact the number of residual votes cast by expert reviewers.[32] Second, ratings also do not seem to be broadly affected. Movies that experience a one-million viewer shock are no more likely to be rated in the top third by expert reviewers; movies that experience large positive shocks may be slightly less likely to be low quality movies, though the difference is not significant.

To test for an observational learning story, we also examine whether movies are differentially impacted according to the ex-ante level of uncertainty about quality. Under an observational learning story, we would expect stronger estimated momentum among films about which there is less good information. We proxy for the ex-ante level of information about a movie with that movie's production budget. Though the production budget does not, according to IMDB, usually include advertising costs, there is likely to be more aggregate uncertainty in quality for high versus low-budget films; consistent with this, Einav (2007) notes that advertising budgets are generally set as a fixed percentage of production budgets. From among the 1,245 movies in our main sample, we have production budget from IMDB for nearly ninety percent. Panel B of Table 3.5 reports our momentum estimates separately for movies that fall in the top third (in excess of $49M) and the bottom third (below $29M) in production budget. The final row reports the differences between the point estimates. Although low-budget films exhibit slightly higher momentum in Weeks 2 and 3, they actually have slightly *lower* momentum in Weeks 4, 5, and 6, and in no week is the difference in estimated momentum between high- and low-budget films statistically significant.[33]

[32] All movies in our sample have had at least a full year to accrue votes.

[33] Appendix Table C.4 shows the corresponding OLS results.

3.7 Magnitudes of Network Externalities by Demographic

We have demonstrated that the observed change in quantity is a demand-side phenomenon and that the momentum we estimate appears to be unrelated to learning; this suggests that our empirical strategy is indeed capturing momentum from network externalities. In this section, we explore for whom these network externalities matter most. We estimate differential network externalities, first between males and females, and then between youth and adults. Though there exists a relatively rich literature in sociology and psychology (and to a lesser extent also in political science) comparing preference for shared experience across demographic groups, we are to our knowledge the first to analyze the relative impact of network externalities in a large market setting.

3.7.1 By Gender

In an array of settings, the psychology and sociology literature has found that females have stronger preference than males for group belonging and social connectedness. In survey work, Barker (2009) finds that females are more likely to report high positive collective self-esteem, which they argue can partially explain higher usage levels of social network sites among females than among males. In experimental work, Huberman and Rubinstein (2000) find that females are more likely than males to choose to side with the majority, and Clancy and Dollinger (1993) finds that females exhibit higher levels of self-connectedness while males are more likely to emphasize the quality of separateness in their self-definitions.

To explore gender differences in network externalities, we classify movies by predominant audience gender based on the predicted gender composition of all IMDB voters using the leave-one-out method. We first regress the proportion of female voters on genre and release year dummies for all movies but i. We then use the resulting estimates to predict movie i's proportion of female voters.[34] Finally, we classify movies within release year as female movies if they fall in the top third in predicted proportion voters female and as male movies if they fall in the bottom third.

We classify based on *predicted* audience gender because the proportion of actual female IMDB voters may be endogenously determined. For example, suppose females experience stronger network externalities than males. If IMDB votes are approximately proportional to viewership, then a given movie might be more likely to be categorized as a female movie if it experienced a shock with positive network effect implications. This would upwardly bias any estimated differences between male-audience and female-audience movies.

Table 3.6 shows estimated network externalities by predominant movie gender when movies are classified using the leave-one-out method. The final row reports the difference in network externalities between female audience and male audience movies. The estimated network externalities for female movies exceed those for male movies in Weeks 2 through 5.[35] Consistent with a stronger preference for shared experience among females than males, momentum from network externalities is about twice as high for female movies as for male movies.[36] Panel A of Figure 3.4 plots these estimated network externalities by gender for each week in theater; the gender differences in network

---

[34]IMDB lists 61 genre categories, many of which are quite rare; in our sample, for example, there is only one "Western Comedy" movie. We condense these 61 genre categories into 19 more general genres: Action, Adventure, Comedy, Drama, Family, Fantasy, Foreign, Historical, Horror, Musical, Sci-Fi, Sports, Thriller, War, Western, Romance. When the IMDB-listed genre contains one of more of the above key genre words, we assign the movie to the first listed genre (e.g., we classify the one "Western Comedy" as a Western movie and the one "Sports Action" as a Sports movie. Because the aggregate proportion of IMDB votes by female voters has grown by about ten percent over the decade, we include year fixed effects so as to avoid confounding our gender results with time trends. Finally, to avoid including mothers' and fathers' ratings of movies targeted to children, we omit all G- and PG-rated movies.

[35]The corresponding estimates for the omitted group of movies, i.e., movies with a highly mixed gender composition of IMDB voters, fall between female movies and male movies estimates but align more closely with those of female movies.

[36]Panel A of Appendix Table C.6 shows the corresponding OLS results.

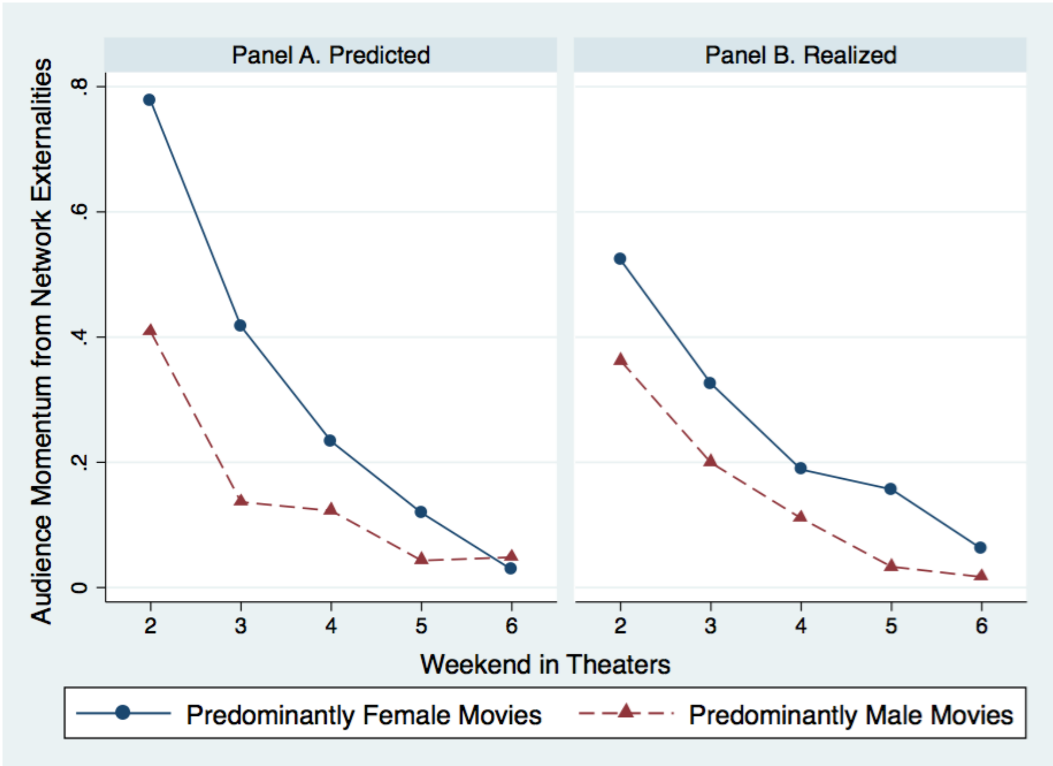
externalities are particularly marked in the weeks immediately following release.

Table 3.6: *Network Externalities by Predicted Gender Demographic*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Female Movies	0.777***	0.417***	0.233***	0.119***	0.0287	1.575***
(obs. 732)	(0.0928)	(0.0590)	(0.0432)	(0.0354)	(0.0435)	(0.207)
Male Movies	0.409***	0.137**	0.123***	0.0431	0.0484**	0.759***
(obs. 813)	(0.0700)	(0.0606)	(0.0383)	(0.0304)	(0.0214)	(0.195)
Difference:						
(Female - Male)	0.368***	0.280***	0.110*	0.076	-0.020	0.816***
	(0.116)	(0.085)	(0.058)	(0.047)	(0.050)	(0.284)

Notes: This table replicates the IV results from Table 3.2 separately by predicted gender demographic using the leave-one-out method described in the text. */**/*** denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively. The first stage results are included in Appendix Table C.3.

Figure 3.4: *Network Externalities by Predicted Gender Demographic*



Notes: Panel A plots the coefficients from Table 3.6 for each of Weeks 2 through 6. Panel B shows the corresponding results when classified based on the realized gender composition of voters as detailed in the text.

For robustness, we can also classify movies based on the *realized* gender composition of voters. "Female" movies are those in the top third that year in percentage of votes female (at least 20.1 percent); "Male" movies are those in the bottom third (i.e., the top third in percentage of IMDB voters male, or at least 82.5 percent male votes).[37] Relative to the predicted method, gender classifications change for only a small number of movies and the results, plotted in Panel B of Figure 3.4 and reported in Appendix Table C.7, are highly comparable.[38].

[37]Because the aggregate proportion of IMDB votes by female voters has grown by about ten percent over the decade, we categorize within year so as not to confound our gender results with time trends. To avoid including mothers' and fathers' ratings of movies targeted to children, we again omit all G- and PG-rated movies from this analysis.

[38]The corresponding OLS results are also similar and are shown in Panel B of Appendix Table C.6

3.7.2 By Age

We also find differences in estimated network externalities by age. We classify each movie into one of three categories based on its age appropriateness according to the MPAA: (1) "Child-friendly," (2) "Teen-friendly," and (3) "Adults-only." Child-friendly includes all films with a G (General Audiences, all ages admitted) or PG (Parental Guidance Suggested, some material may not be suitable for children) MPAA rating; teen-friendly films are those rated PG-13 (Parents Strongly Cautioned; some material parents might consider inappropriate for children under 13 years); and adults-only films are those rated R (Restricted; people under 17 years may only be admitted if accompanied by a parent or guardian).[39] Table 3.7 shows estimated network externalities separately by age suitability. The last two rows report the difference in network externalities between child-friendly and adults-only movies and between teen-friendly and adults-only movies, respectively. Suggestive of higher preference among youth for shared experience, child- and teen-friendly movies exhibit larger network externalities than do adults-only movies.[40]

[39]Since NC-17 (No One 17 and Under Admitted, exclusively adult) likely captures a different demographic, we omit throughout the less than 1 percent of NC-17 films in our sample; their inclusion in the Adults Only category, however, does not significantly change our results.

[40]Appendix Table 3.7 shows the corresponding OåLS results; though the network externalities estimated by IV were much larger for youths than for adults, the more general momentum estimated by OLS is quite similar across age suitabilities.

Table 3.7: *Network Externalities by Age Suitability*

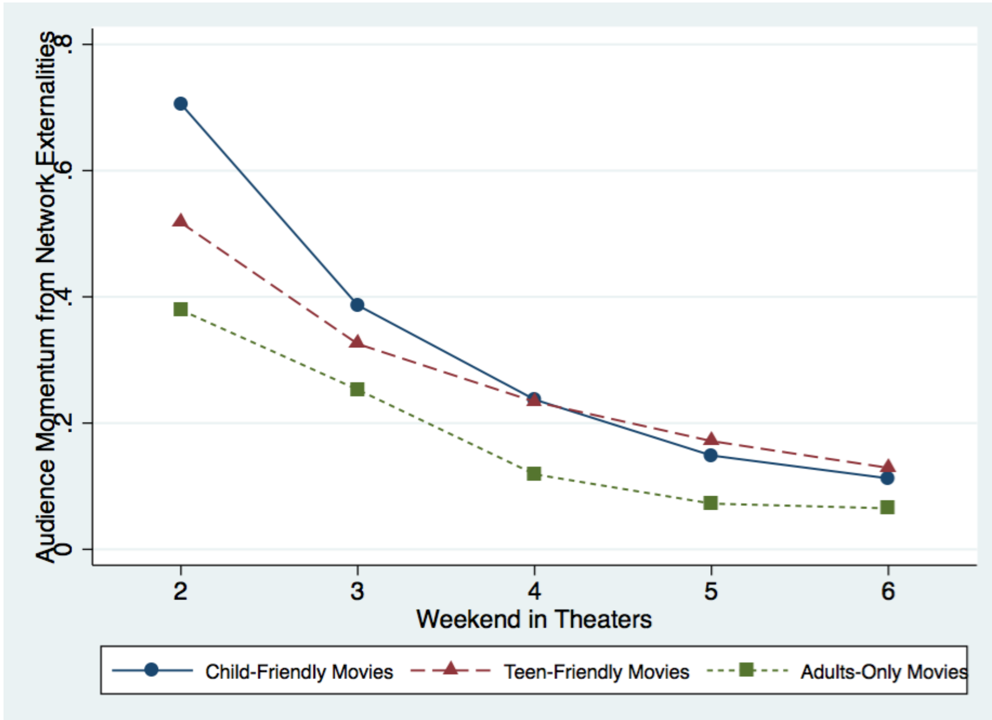
	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Child-Friendly Movies	0.706***	0.387***	0.237***	0.149***	0.112***	1.591***
(obs. 688)	(0.0896)	(0.0518)	(0.0403)	(0.0327)	(0.0279)	(0.203)
Teen-Friendly Movies	0.518***	0.325***	0.234***	0.172***	0.129***	1.377***
(obs. 1217)	(0.0767)	(0.0608)	(0.0519)	(0.0384)	(0.0313)	(0.244)
Adults-Only Movies	0.379***	0.253***	0.119***	0.0725**	0.0649**	0.888***
(obs. 909)	(0.0735)	(0.0585)	(0.0452)	(0.0355)	(0.0319)	(0.210)
Differences:						
(Child - Adult)	0.327***	0.125	0.118*	0.077	0.047	0.703***
	(0.116)	(0.078)	(0.061)	(0.048)	(0.042)	(0.292)
(Teen - Adult)	0.139	0.072	0.115*	0.100*	0.064	0.489
	(0.106)	(0.084')	(0.069)	(0.052)	(0.045)	(0.322)

Notes: This table replicates the IV results from Table 3.2 separately by film MPAA rating. Child-friendly films are those rated G or PG; teen-friendly films are rated PG-13; and adult-only films are rated R. */**/*** denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively. The first stage results are included in Appendix Table C.3.

Figure 3.5 shows the estimated network externalities for each of the age suitability categories by week in theater. Child-friendly movies exhibit the largest momentum from network externalities in early weeks; for 100 additional viewers opening weekend, child-friendly films bring in 70 additional viewers the second weekend, compared to only 50 additional viewers for teen-friendly films and 40 for adults-only films. The network effects momentum among children, however, is least persistent: by the third weekend marginal network externalities in child-friendly films is about on par with teen-friendly films.[41]

[41] We have estimated larger network externalities among females than males, and among children and teens than adults. The interested reader may review Appendix C.3 in which we explore variation in the magnitude of network externalities by adult ages; there we find that network externalities are larger among the oldest available age category (45 plus) than among mid-life adults (either 18 to 29 or 30 to 44).

Figure 3.5: *Network Externalities by Movie Age Suitability*



Notes: This figure plots the coefficients from Table 3.7 for each of Weeks 2 through 6. Child-friendly films are those rated G or PG; teen-friendly films are rated PG-13; and adult-only films are rated R.

3.8 Substitution and Aggregate Viewership

We have thus far been agnostic as to the sources of the network-externality induced viewerships. Were these viewerships simply from individuals substituting across movies? Or do they instead represent a more dramatic substitution across activities?

The results in Table 3.8 suggest that most of our estimated viewerships from network externalities come from substitution across activities. In the first row, the endogenous regressor is abnormal viewership of new releases in week w; the outcomes are abnormal viewership in $w + 1$ of (1) those same movies (i.e. our base case results), (2) all movies showing in both w and $w + 1$, (3) new movies opening in $w + 1$, and (4) all movies showing in $w+1$, respectively.[42] Each reported coefficient is from a separate regression. The first column simply reproduces our base case results for comparison; the second shows that (unsurprisingly) shocks to opening weekend viewership are correlated with higher viewership in $w+1$ of all movies that played both weekends. The first and third columns taken together provide suggestive evidence of some (statistically insignificant) substitution across movies: amid 100 more viewers of movies opening last weekend, we see 51 more viewers of those movies this weekend and 16 fewer viewers of new movies just opening. This is consistent with a partial substitution story in which network effects in consumption increase the utility from seeing movies that did particularly well last weekend, thus reducing demand for new movies this weekend (which experienced no such shock). The large and statistically significant coefficient in the final column, moreover, suggests that even accounting for substitution away from new movies, network externalities produce significant momentum in aggregate movie-going. That is, network externalities drive many who would otherwise have engaged in a different activity to attend the movies instead.

[42] The first stage is the same as in our standard base case (see Table 3.1).

Table 3.8: *Substitution across Movies and Activities*

	Outome variable: Audiences next week			
Endogenous regressor: Audiences this week	Movies in 2nd Week	Movies in 2nd to 6th Week	Movies in 1st Week	Movies in 1st to 6th Week
Movies in 1st Week	0.511***	0.474***	-0.159	0.332*
	(0.0513)	(0.0970)	(0.192)	(0.202)
Movies in 1st to 5th Week		0.468***	-0.139	0.327*
		(0.0734)	(0.187)	(0.193)

Notes: Each reported coefficient is from a separate regression. In the first row, the endogenous regressor is abnormal daily tickets sales weekend w to movies that opened in weekend w; the outcome variables are abnormal daily ticket sales in weekend $w+1$ to movies that (1) opened in week w, (2) played in both w and $w+1$, (3) opened in week w, and (4) played in week $w+1$, respectively. The corresponding first stage is in the first row of Table 3.1. In the second row, the endogenous regressor is abnormal daily ticket sales in weekend w to movies they played in both w and $w+1$; the corresponding first stage is Appendix Table 3.1. */**/*** denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Since weather shocks may well engender momentum for *any* movie showing both that weekend and the next (not just movies that opened that weekend), the second row of Table 3.8 shows the corresponding results when the endogenous regressor is defined as ticket sales in week w for all movies that showed in both w and $w+1$.[43] With this specification, we again find (1) strong momentum from network externalities; (2) some (statistically insignificant) evidence of substitution away from new movies released the following weekend; and (3) large and significant substitution across activities the following weekend. For 100 additional viewers in weekend w to movies showing in both w and $w+1$, we observe about 47 more viewers of those same movies in $w+1$; just under one-third of these would otherwise have seen one of the new releases in $w+1$, but the majority (70 percent) were drawn into theaters in $w+1$ by positive network externalities in consumption.

[43]The instrument is the same as in our base case specifications; the first stage is reported in Appendix Table C.3

3.9 Conclusion

In this chapter, we exploit the randomness of weather, and the relationship between weather and movie-going, to test for and quantify network externalities in movie consumption. In the first stage, we instrument for opening weekend viewership with unanticipated and plausibly exogenous weather shocks that weekend. Given the large number of potential instruments, LASSO variable selection methods are key in generating a strong and econometrically sound first stage. We expect this approach will prove similarly fruitful in other settings where weather is a powerful and exogenous determinant of behavior, but specifying the optimal first stage is otherwise non-obvious.

Previous work on momentum in entertainment goods has highlighted the role of learning; we find that network externalities are also important. Using our LASSO-chosen instruments, we estimate the effect of exogenous shocks to opening weekend viewership on viewership in later weekends. Our results show that network externalities engender a multiplier effect: a shock to opening weekend viewership on average yields at least as many viewers again in the following five weekends. The effects appear to be stronger for females than males, and for youth than adults. Overall, these network externalities have non-trivial impacts on behavior: most of our estimated momentum comes from substitution across activities.

We've found that people follow in the consumptive footprints of others. They do so even when suppliers are not pushing the fads, and even when those footprints provide no information about quality. The powerful and prolonged effects of network externalities in this context suggest potentially important implications of network externalities in behaviors with more obvious social welfare impacts (e.g., school attendance, fertility, smoking). Moreover, amid the rise of potentially solitary activities like gaming, remote work, and online learning, further research into where and how network externalities might deepen participation and engagement could also prove fruitful.

References

Bailey, M. J., and S. M. Dynarski (2011): "Inequality in Postsecondary Education," in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, ed. by G. Duncan, and R. Murnane. Russell Sage Foundation: New York, New York.

Bandiera, O., I. Barankay, and I. Rasul (2005): "Social preferences and the response to incentives: Evidence from personnel data," *The Quarterly Journal of Economics*, 120(3), 917–962.

——— (2012): "Team incentives: Evidence from a firm level experiment," *CEPR Discussion Paper No. DP8776*.

Banerjee, A. V. (1992): "A simple model of herd behavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

Barker, V. (2009): "Older adolescents' motivations for social network site use: The influence of gender, group identity, and collective self-esteem," *Cyber Psychology & Behavior*, 12(2), 209–213.

Beaman, L., and J. Magruder (2012): "Who gets the job referral? Evidence from a social networks experiment," *The American Economic Review*, 102(7), 3574–3593.

Becker, G. S. (1991): "A note on restaurant pricing and other examples of social influences on price," *The Journal of Political Economy*, pp. 1109–1116.

Belley, P., and L. Lochner (2007): "The Changing Role of Family Income and Ability in Determining Educational Achievement," *Journal of Human Capital*, 1(1), 37–89.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 80(6), 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen (2010): "LASSO methods for gaussian instrumental variables models," *arXiv preprint arXiv: 1012.1297*.

——— (2011): "Inference on treatment effects after selection amongst high-dimensional controls," *arXiv preprint arXiv:1201.0224*.

Bentolila, S., C. Michelacci, and J. Suarez (2010): "Social contacts and occupational choice," *Economica*, 77(305), 20–45.

Bernheim, B. D. (1994): "A theory of conformity," *The Journal of Political Economy*, pp. 841–877.

Bertrand, M., and J. Pan (2013): "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior," *American Economic Journal: Applied Economics*, 5(1), 32–64.

Bewley, T. F. (1999): *Why wages don't fall during a recession*. Harvard University Press.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009): "Simultaneous analysis of LASSO and Dantzig selector," *The Annals of Statistics*, 37(4), 1705–1732.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A theory of fads, fashion, custom, and cultural change as informational cascades," *The Journal of Political Economy*, pp. 992–1026.

——— (1998): "Learning from the behavior of others: Conformity, fads, and informational cascades," *The Journal of Economic Perspectives*, 12(3), 151–170.

Bikhchandani, S., and S. Sharma (2000): "Herd behavior in financial markets," *IMF Staff papers*, pp. 279–310.

Blau, G. (1990): "Exploring the mediating mechanisms affecting the relationship of recruitment source to employee performance," *Journal of Vocational Behavior*, 37(3), 303–320.

Brown, M., E. Setren, and G. Topa (2012): "Do informal referrals lead to better matches? Evidence from a firm's employee referral system," *Evidence from a Firm's Employee Referral System (August 1). FRB of New York Staff Report*, (568).

Bryan, G. T., D. Karlan, and J. Zinman (2012): "You can pick your friends, but you need to watch them: Loan screening and enforcement in a referrals field experiment," *NBER Working Paper 17883*.

Burks, S., B. Cowgill, M. Hoffman, and M. Housman (2013): "The facts about referrals: Toward an understanding of employee referral networks," *Available at SSRN 2253738*.

Bursztyn, L., F. Ederer, B. Ferman, and N. Yuchtman (2013): "Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions," *Available at SSRN 2101391*.

Calvo-Armengol, A., and M. O. Jackson (2004): "The effects of social networks on employment and inequality," *The American Economic Review*, 94(3), 426–454.

Castilla, E. J. (2005): "Social networks and employee performance in a call center," *American Journal of Sociology*, 110(5), 1243–1283.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2012): "Alternative Asymptotics and the Partially Linear Model With Many Regressors," *Working Paper*.

Çelen, B., and S. Kariv (2004): "Distinguishing informational cascades from herd behavior in the laboratory," *American Economic Review*, pp. 484–498.

Chen, Y.-F. (2008): "Herd behavior in purchasing books online," *Computers in Human Behavior*, 24(5), 1977–1992.

Chernozhukov, V., and C. Hansen (2013): "Econometrics of high-dimensional sparse models," *NBER Lectures*.

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner (2014): "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," Discussion paper, National Bureau of Economic Research.

Choi, J. P. (1997): "Herd behavior, the "penguin effect," and the suppression of informational diffusion: An analysis of informational externalities and payoff interdependency," *The Rand Journal of Economics*, pp. 407–425.

Clancy, S. M., and S. J. Dollinger (1993): "Photographic depictions of the self: Gender and age differences in social connectedness," *Sex Roles*, 29(7-8), 477–495.

Corts, K. S. (2001): "The strategic effects of vertical market structure: Common agency and divisionalization in the US motion picture industry," *Journal of Economics & Management Strategy*, 10(4), 509–528.

Costa, D. L., and M. E. Kahn (2003): "Cowards and heroes: Group loyalty in the American Civil War," *The Quarterly Journal of Economics*, 118(2), 519–548.

——— (2010): "Health, wartime stress, and unit cohesion: Evidence from Union Army veterans," *Demography*, 47(1), 45–66.

Dahl, G., and S. DellaVigna (2009): "Does movie violence increase violent crime?," *The Quarterly Journal of Economics*, 124(2), 677–734.

Dahl, G. B., and E. Moretti (2008): “The demand for sons,” *The Review of Economic Studies*, 75(4), 1085–1120.

Datcher, L. (1983): “The impact of informal networks on quit behavior,” *The Review of Economics and Statistics*, 65(3), 491–495.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2014): “Voting to tell others,” Working Paper 19832, National Bureau of Economic Research.

DiPrete, T. A., and C. Buchmann (2013): *The rise of women: The growing gender gap in education and what it means for American schools*. Russell Sage Foundation: New York, New York.

Dustmann, C., A. Glitz, and U. Schönberg (2011): “Referral-based job search networks,” *IZA Discussion Paper 5777*.

Einav, L. (2007): “Seasonality in the US motion picture industry,” *The Rand Journal of Economics*, 38(1), 127–145.

Elberse, A., and B. Anand (2007): “The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market,” *Information Economics and Policy*, 19(3), 319–343.

Elberse, A., and J. Eliashberg (2003): “Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures,” *Marketing Science*, 22(3), 329–354.

Ellison, G., and D. Fudenberg (1995): “Word-of-mouth communication and social learning,” *The Quarterly Journal of Economics*, 110(1), 93–125.

Fernandez, R. M., E. J. Castilla, and P. Moore (2000): “Social capital at work: Networks and employment at a phone center,” *American Journal of Sociology*, 105(5), 1288–1356.

Fernandez, R. M., and N. Weinberg (1997): "Sifting and sorting: Personal contacts and hiring in a retail bank," *American Sociological Review*, pp. 883–902.

Frank, I. E., and J. H. Friedman (1993): "A statistical view of some chemometrics regression tools," *Technometrics*, 35(2), 109–135.

Goldin, C., L. F. Katz, and I. Kuziemko (2006): "The Homecoming of American College Women: The Reversal of the College Gender Gap," *The Journal of Economic Perspectives*, 20(4), 133.

Granovetter, M. (1995): *Getting a job: A study of contacts and careers*. University of Chicago Press, Chicago.

Heath, R. (2013): "Why do firms hire using referrals? Evidence from Bangladeshi garment factories," *Working Paper*.

Herrmann, M. A., and J. E. Rockoff (2010): "Worker absence and productivity: Evidence from teaching," *NBER Working Paper 16524*.

Hirshleifer, D., and S. Hong Teoh (2003): "Herd behaviour and cascading in capital markets: A review and synthesis," *European Financial Management*, 9(1), 25–66.

Holzer, H. J. (1987): "Hiring procedures in the firm: Their economic determinants and outcomes," *NBER Working Paper 2185*.

Horton, J. J. (2013): "The effects of subsidizing employer search," *Working Paper*.

Huberman, G., and A. Rubinstein (2000): "Correct belief, wrong action and a puzzling gender difference," *Tel Aviv University Working Papers*, (00-17).

Ioannides, Y. M., and L. D. Loury (2004): "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature*, 42(4), 1056–1093.

Jacob, B. A. (2002): "Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education," *Economics of Education Review*, 21(6), 589–598.

Kane, T. J., C. E. Rouse, and D. Staiger (1999): "Estimating returns to schooling when schooling is misreported," Discussion paper, National Bureau of Economic Research.

Katz, M. L., and C. Shapiro (1986): "Technology adoption in the presence of network externalities," *The Journal of Political Economy*, pp. 822–841.

Krider, R. E., T. Li, Y. Liu, and C. B. Weinberg (2005): "The lead-lag puzzle of demand and distribution: A graphical method applied to movies," *Marketing Science*, 24(4), 635–645.

Kugler, A. D. (2003): "Employee referrals and efficiency wages," *Labour Economics*, 10(5), 531–556.

McFadden, D. L., and K. E. Train (1996): "Consumers' evaluation of new products: Learning from self and others," *The Journal of Political Economy*, 104(4), 683–703.

Montgomery, J. D. (1991): "Social networks and labor-market outcomes: Toward an economic analysis," *The American Economic Review*, 81(5), 1408–1418.

Montgomery, M. R., and J. B. Casterline (1996): "Social learning, social influence, and new models of fertility," *Population and Development Review*, 22, 151–175.

Moretti, E. (2011): "Social learning and peer effects in consumption: Evidence from movie sales," *The Review of Economic Studies*, 78(1), 356–393.

Moul, C. C. (2007): "Measuring word of mouth's impact on theatrical movie admissions," *Journal of Economics & Management Strategy*, 16(4), 859–892.

Mulligan, J. G., and L. Motiere (1994): "The Market for First-Run Motion Pictures," *University of Delaware Department of Economics Working Paper*, 12.

MUNSHI, K., AND J. MYAUX (2006): "Social norms and the fertility transition," *Journal of Development Economics*, 80(1), 1–38.

NELSON, R. A., M. R. DONIHUE, D. M. WALDMAN, AND C. WHEATON (2001): "What's an Oscar worth?," *Economic Inquiry*, 39(1), 1–6.

ODESK CORPORATION (2013): *Slide deck with statistics on the oDesk marketplace*.

PALLAIS, A. (2014): "Inefficient hiring in entry-level labor markets," *American Economic Review*, forthcoming.

PETER, K., AND L. HORN (2005): "Gender Differences in Participation and Completion of Undergraduate Education and How They Have Changed Over Time. Postsecondary Education Descriptive Analysis Reports. NCES 2005-169.," *US Department of Education*.

PETERSEN, T., I. SAPORTA, AND M.-D. L. SEIDEL (2000): "Offering a job: Meritocracy and social networks," *American Journal of Sociology*, 106(3), 763–816.

PISTAFERRI, L. (1999): "Informal networks in the Italian labor market," *Giornale degli Economisti e Annali di Economia*, pp. 355–375.

PRAG, J., AND J. CASAVANT (1994): "An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry," *Journal of Cultural Economics*, 18(3), 217–235.

QUENZEL, G., AND K. HURRELMANN (2013): "The growing gender gap in education," *International Journal of Adolescence and Youth*, 18(2), 69–84.

REES, A. (1966): "Information networks in labor markets," *The American Economic Review*, 56(1/2), 559–566.

SAWHNEY, M. S., AND J. ELIASHBERG (1996): "A parsimonious model for forecasting gross box-office revenues of motion pictures," *Marketing Science*, 15(2), 113–131.

Scharfstein, D. S., and J. C. Stein (1990): "Herd behavior and investment," *American Economic Review*, pp. 465–479.

Segrest, S. L., D. J. Domke-Damonte, A. K. Miles, and W. P. Anthony (1998): "Following the crowd: Social influence and technology usage," *Journal of Organizational Change Management*, 11(5), 425–445.

Simon, C. J., and J. T. Warner (1992): "Matchmaker, matchmaker: The effect of old boy networks on job match quality, earnings, and tenure," *Journal of Labor Economics*, 10(3), 306–330.

Sorensen, A. T. (2007): "Bestseller lists and product variety," *The Journal of Industrial Economics*, 55(4), 715–738.

Swami, S., J. Eliashberg, and C. B. Weinberg (1999): "SilverScreener: A modeling approach to movie screens management," *Marketing Science*, 18(3), 352–372.

Tibshirani, R. (1996): "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society*, pp. 267–288.

Topa, G. (2011): "Labor markets and referrals," *Handbook of Social Economics*, pp. 1193–1221.

Vogel, H. L. (2011): *Entertainment industry economics*. Cambridge University Press.

Welch, I. (1992): "Sequential sales, learning, and cascades," *The Journal of Finance*, 47(2), 695–732.

Young, H. P. (2009): "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning," *American Economic Review*, 99(5), 1899–1924.

Zufryden, F. S. (1996): "Linking advertising to box office performance of new film releases: A marketing planning model," *Journal of Advertising Research*, 36, 29–42.

Appendix A

Appendix to Chapter 1

A.1 Communication in the Team Experiment

Here we discuss communication differences across team types. In light of our findings, for robustness we also re-estimate performance differences by team type with the inclusion of several additional controls.

Panel A of Appendix Table 9 shows how the team types differed in their communication methods. We regress each communication outcome on indicators for being in Type A and Type C teams; as before, the base group is Type B teams. Controls for the characteristics of referred and non-referred workers are included throughout. The first column considers chat box use, defined as both partners typing at least one message in the chat box. The second column considers the total number of messages sent by both partners during the task and is limited to teams that used the chat box. Because we directly observe what is written in the chat box, both of these measures are known for all teams and do not rely on worker reports.

The last two columns consider communication outside the chat box, such as on Skype. When workers submitted their slogans for each task, we asked if they had used other forms of communication. We code teams as using other forms of communication if at least one partner reported doing so. The third column addresses selection into answering this question. Here we regress a dummy for whether at least one teammate answered this question on team type. In the final column, we regress an indicator for reporting using other forms of communication on team type. This final specification includes only teams that answered the communication question.

Type A teams communicated the most, both in and out of the chat box. Relative to Type B teams, Type A teams were slightly, though insignificantly, more likely to use the chat box. When they did use the chat box, Type A teams sent about one-third more messages. The biggest difference between the communication of Type A and Type B teams, however, is in the frequency with which they used other forms of communication. While 38 percent of Type B teams reported using other forms of communication, Type A teams were twice as likely to do so; the magnitude of this difference implies that the difference itself cannot be driven by the small difference in the likelihood of answering this question.[1]

On the other hand, Type C teams were significantly less likely to use the chat box. This is not surprising since the chat box was on the site and non-referred workers were significantly less likely ever to log in.

We have observed that, relative to Type B teams, Type A teams communicated more both in and out of the chat box. They also spent more time on the task (Table 5). Panel B of Appendix Table 9 provides evidence that, even controlling for communication and time spent, Type A teams still outperformed Type B teams. We replicate the main team performance specifications with controls for referred and non-referred workers' characteristics (from Panel B of Table 4), adding as controls an indicator for using the chat box, the number of messages sent in the chat box, an indicator for using other methods

[1] At least one partner answered this question in 95 percent of Type B teams; Type B teams were slightly more likely to answer this question than either of the other team types.

of communication, and the number of minutes spent by each partner.[2]

Unsurprisingly, more communication and more time spent both led to better outcomes. For example, teams that sent the median number of messages in the chat box (eight) were 21 percentage points more likely to answer the team question the same way and 15 percentage points more likely to provide the same slogan than were teams that did not use the chat box, all else equal. Teams in which each partner spent an additional five minutes each on the task were, all else equal, three percentage points more likely to have their team question match and two percentage points more likely to submit the same slogan.

Even conditioning on the type of communication used, number of messages sent, and minutes spent by each partner, however, Type A teams remained 14 percentage points more likely to provide the same answer to the team question and 23 percentage points more likely to submit the same slogan than Type B teams. Type C teams, meanwhile, remained substantially less likely to do either (11 percentage points and seven percentage points, respectively).

[2] If neither partner answered the question about using other forms of communication, we set the indicator for having reported communication outside the chat box to zero. Thus, this dummy also directly captures the effect of having at least one partner submit work.

A.2 Supplementary Tables

Table A.1: *Descriptive Statistics*
Individual and Team Experiments, Workers Asked to Refer

	Referred Someone	Referred No One	Difference	Included Referrers
Has Prior Experience	1.00	1.00		1.00
Earnings	$2,917	$2,397	**	$2,932
Number of Previous Jobs	12.58	11.07	**	12.35
Has Feedback Score	1.00	1.00		1.00
Feedback Score	4.80	4.80		4.81
Posted Wage	$2.84	$2.77	*	$2.85
Days Since Joining oDesk	689	709		666
Has Portfolio	0.69	0.61	**	0.69
Number of Tests Passed	5.80	5.60		5.84
Has English Score	1.00	0.99	**	1.00
English Score	4.79	4.79		4.77
Agency Affiliated	0.25	0.24		0.21
Number of Degrees	1.40	1.35		1.41
Proposed Wage	$2.50	$2.51		$2.51
Observations	1,246	1,867		455

Notes: Each statistic in the table presents the mean of the characteristic indicated by the row for the sample indicated by the column. *Referred Someone* denotes workers who referred at least one other worker to our firm, whether or not we hired that worker. *Referred No One* denotes workers who referred no workers to our firm. *Included Referrers* is a subset of *Referred Someone* and includes only those workers whose referral we hired. *English Score* is self-reported English ability on a one-to-five scale, a *portfolio* is where a worker posts prior work, and *agency-affiliated* workers pay a fraction of their earnings to report they are part of a given group of oDesk workers (an agency). *, ** denotes that the means of the characteristic for *Referred Someone* and *Referred No One* are significantly different at the 10% and 5% levels, respectively.

Table A.2: *Randomization Assessment*
Individual Experiment

	Referrers		Referred Workers	
	Treatment 1	Treatment 2	Treatment 1	Treatment 2
Has Prior Experience	1.00	1.00	0.73	0.75
Earnings	$2,919	$2,996	$1,396	$1,379
Number of Previous Jobs	12.78	13.09	8.28	10.14
Has Feedback Score	1.00	1.00	0.62	0.64
Feedback Score	4.80	4.76	4.66	4.59
Posted Wage	$2.78	$2.85	$2.68	$2.72
Days Since Joining oDesk	645	676	489	566
Has Portfolio	0.64	0.68	0.47	0.50
Number of Tests Passed	5.78	5.78	4.98	5.31
Has English Score	1.00	1.00	0.98	1.00
English Score	4.84	4.79	4.75	4.66
Agency Affiliated	0.08	0.08	0.05*	0.10*
Number of Degrees	1.50	1.36	1.34	1.51
Proposed Wage	$2.53	$2.53	$2.40	$2.37
Observations	86	87	127	128

Notes: Each cell presents the mean of the characteristic indicated by the row for the sample indicated by the column. Only workers in the individual experiment are included. *English Score* is self-reported English ability on a one-to-five scale, a *portfolio* is where a worker posts prior work, and *agency-affiliated* workers pay a fraction of their earnings to report they are part of a given group of oDesk workers (an agency). * denotes the Treatment 1 and Treatment 2 group means are statistically different at the 10% level.

Table A.3: *Performance and Persistence, with Different Controls*
Individual Experiment: Base Group is Non-Monitored Referred Workers (Treatment 2)

A. All Days, No Controls

	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred (Treatment 1)	0.036	0.053	0.034	-0.032
	(0.042)	(0.047)	(0.039)	(0.030)
Non-Referred	-0.132**	-0.079*	-0.101**	-0.225**
	(0.042)	(0.045)	(0.039)	(0.038)
Constant	0.757	0.563	0.640	0.953
	(0.031)	(0.034)	(0.028)	(0.019)
Controls	No	No	No	No
Observations	2,610	2,610	2,610	435
R-squared	0.027	0.013	0.020	0.085

B. All Days, with Second Order Controls

	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred (Treatment 1)	0.012	0.043	0.015	-0.027
	(0.042)	(0.046)	(0.040)	(0.039)
Non-Referred	-0.143**	-0.080	-0.109**	-0.184**
	(0.050)	(0.052)	(0.046)	(0.054)
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Observations	2,610	2,610	2,610	435
R-squared	0.190	0.160	0.180	0.270

C. Last Day Only, with Second Order Controls

	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred (Treatment 1)	-0.045	0.023	-0.032	-0.041
	(0.049)	(0.057)	(0.043)	(0.039)
Non-Referred	-0.161**	-0.104*	-0.090*	-0.125**
	(0.054)	(0.057)	(0.046)	(0.050)
Daily Performance Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Observations	435	435	435	435
R-squared	0.617	0.510	0.625	0.452

Notes: Panel A replicates Panel A of Table 2, eliminating the controls for worker characteristics. Panels B and C replicate Panels A and B, respectively, of Table 2 with additional control variables. In addition to the characteristics listed in footnote 18, these regressions all include Second Order Controls: the square of each non-binary characteristic in footnote 18 and the interaction of each pair of characteristics in footnote 18. *, ** denote significance at the 10% and 5% levels, respectively.

Table A.4: *Selection into Accepting Job Offer in the Supplemental Experiment*
Data from the Individual and Supplemental Experiments

	Dependent Variable: Accepted Job Offer in Supplemental Experiment			
	Submission	On-Time Submission	Accuracy	Re-Application
Performance in Individual Experiment × Non-Referred	0.131	0.133	0.219	0.008
	(0.129)	(0.123)	(0.135)	(0.143)
Performance in Individual Experiment	0.164*	0.113	0.159	0.217*
	(0.096)	(0.081)	(0.099)	(0.119)
Non-Referred	-0.122	-0.114	-0.165*	-0.026
	(0.106)	(0.086)	(0.098)	(0.136)
Controls	Yes	Yes	Yes	Yes
Observations	435	435	435	435
R-squared	0.154	0.144	0.161	0.149

Notes: Each column presents the results of a separate regression of an indicator for a worker accepting our job offer in the supplemental experiment on a measure of her performance in the individual experiment, an indicator for being a non-referred worker, and the interaction of these two indicators. Each column uses a different performance measure indicated by the column heading. Observations are workers; only referred and non-referred workers are included. Each regression contains controls for the individual characteristics listed in footnote 18. * denotes significance at the 10% level.

Table A.5: *Performance and Persistence in New Firm, with Different Controls*
Supplemental Experiment: Base Group is All Referred Workers

	A. All Workers, No Controls				
	Accepted Job Offer	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.167**	-0.182**	-0.178**	-0.079**	-0.236**
	(0.047)	(0.042)	(0.042)	(0.022)	(0.047)
Constant	0.678	0.518	0.499	0.247	0.553
	(0.029)	(0.029)	(0.029)	(0.015)	(0.031)
Controls	No	No	No	No	No
Observations	435	2,175	2,175	2,175	435
R-squared	0.029	0.033	0.031	0.019	0.055

	B. All Workers, with Second Order Controls				
	Accepted Job Offer	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.042	-0.073	-0.073	-0.020	-0.097
	(0.064)	(0.055)	(0.055)	(0.029)	(0.064)
Controls	Yes	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes	Yes
Observations	435	2,175	2,175	2,175	435
R-squared	0.268	0.242	0.239	0.201	0.292

	C. Conditional on Accepting Job Offer, with Second Order Controls			
	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.119*	-0.118*	-0.047	-0.198**
	(0.067)	(0.070)	(0.038)	(0.090)
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Mean of Dependent Variable				
Base Group (Referred Workers)	0.763	0.735	0.363	0.815
Observations	1,325	1,325	1,325	265
R-squared	0.247	0.25	0.192	0.358

Notes: Panel A replicates Panel A of Table 3, eliminating the controls for worker characteristics. Panels B and C replicate Panels A and B, respectively, of Table 3 with additional control variables. In addition to the characteristics listed in footnote 18, these regressions all include Second Order Controls: the square of each non-binary characteristic in footnote 18 and the interaction of each pair of characteristics in footnote 18. *, ** denote significance at the 10% and 5% levels, respectively.

Table A.6: *Performance in Team Experiment, with Different Controls*
Team Experiment: Base Group is Referred Workers Paired with Someone Else's Referrer (Type B)

	A. Individual Diligence, No Controls			
	Logged in	Submitted	Individual Question Correct	Own Criteria in Slogan
Referred Worker When Working with Own Referrer	0.018	0.046**	0.053*	0.004
	(0.017)	(0.017)	(0.028)	(0.033)
Non-Referred Worker When Working with Referrer (Type	-0.294**	-0.312**	-0.287**	-0.138**
	(0.051)	(0.048)	(0.049)	(0.052)
Controls	No	No	No	NO
Observations	846	846	846	846
R-Squared	0.124	0.134	0.102	0.018
	B. Individual Diligence, with Second Order Controls			
	Logged in	Submitted	Individual Question Correct	Own Criteria in Slogan
Referred Worker When Working with Own Referrer	0.018	0.046**	0.053*	0.004
	(0.018)	(0.018)	(0.030)	(0.035)
Non-Referred Worker When Working with Referrer (Type	-0.131**	-0.176**	-0.193**	-0.044
	(0.058)	(0.056)	(0.057)	(0.051)
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Observations	846	846	846	846
R-Squared	0.420	0.388	0.309	0.213
	C. Team Performance, No Controls			
	Both Submitted	Team Question Matches	Same Slogan	Same Slogan & Both Criteria
Referred Worker and Own Referrer Team (Type A)	0.099**	0.287**	0.372**	0.103**
	(0.023)	(0.028)	(0.032)	(0.024)
Non-Referred Worker and Referrer Team (Type C)	-0.280**	-0.206**	-0.142**	-0.053*
	(0.048)	(0.047)	(0.041)	(0.029)
Controls	No	No	No	No
Observations	846	846	846	846
R-Squared	0.117	0.164	0.194	0.031
	D. Team Performance, with Second Order Controls			
	Both Submitted	Team Question Matches	Same Slogan	Same Slogan & Both Criteria
Referred Worker and Own Referrer Team (Type A)	0.099**	0.287**	0.372**	0.103**
	(0.024)	(0.030)	(0.034)	(0.025)
Non-Referred Worker and Referrer Team (Type C)	-0.162**	-0.075	-0.032	0.026
	(0.061)	(0.054)	(0.054)	(0.031)
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Observations	846	846	846	846
R-Squared	0.314	0.327	0.311	0.170

Notes: Panels A and B replicate Panel A of Table 4 eliminating the controls for worker characteristics (Panel A) and adding Second Order Controls (Panel B). Panels C and D replicate Panel B of Table 4, eliminating the controls for worker characteristics (Panel C) and adding Second Order Controls (Panel D). The Second Order Controls are the square of each non-binary characteristic in footnote 18 and the interaction of each pair of characteristics in footnote 18. *, ** denote significance at the 10% and 5% levels, respectively.

Table A.7: *Relationship between Referrer Characteristics and Referred Worker Characteristics*
Individual and Team Experiments

	Dependent Variable: Referred Worker Characteristic										
	Earnings	Number of Previous Jobs	Feedback Score	Posted Wage	Days Since Joining oDesk	Has Portfolio	Number of Tests Passed	English Score	Agency Affiliated	Number of Degrees	Proposed Wage
Referrer Characteristic	0.060*	0.095*	0.204	0.231**	0.264**	0.272**	0.157**	0.285**	0.317**	0.075*	0.367**
	(0.034)	(0.052)	(0.180)	(0.049)	(0.049)	(0.044)	(0.043)	(0.068)	(0.050)	(0.044)	(0.052)
Constant	688.6	4.637	3.562	1.928	289.95	0.296	3.580	3.319	0.060	1.238	1.425
	(113.3)	(0.685)	(0.867)	(0.146)	(33.337)	(0.035)	(0.277)	(0.333)	(0.011)	(0.069)	(0.134)
Observations	537	537	296	537	537	537	537	533	537	537	537
R-Squared	0.014	0.009	0.005	0.036	0.060	0.064	0.032	0.057	0.146	0.006	0.103

Notes: Each column reports the results of regressing the value of an observable characteristic for a referred worker on the value of the same characteristic for her referrer. Each column corresponds to a different characteristic, indicated by the column header. All 537 hired referred workers from the individual and team experiments are included, although *Feedback Score* and *English Score* are missing for some workers. *English Score* is self-reported English ability on a one-to-five scale, a *portfolio* is where a worker posts prior work, and *agency-affiliated* workers pay a fraction of their earnings to report they are part of a given group of oDesk workers (an agency). Huber-White standard errors are in parenthesis. *, ** denote significance at the 10% and 5% levels, respectively.

Table A.8: *Characteristics of the Referrer-Referred Worker Relationship*

	Workers	Workers
How Well Referrer Knows Referral		
1 (Hardly at all)	1%	2%
2	2%	2%
3	5%	3%
4	14%	9%
5	20%	19%
6 (Extremely Well)	57%	65%
Observations	535	1,314
Frequency of Interaction		
Less than Once a Month	2%	4%
About Once a Month	5%	4%
Less than Weekly, More than Monthly	8%	4%
About Once a Week	13%	9%
Less than Daily, More than Weekly	21%	14%
About Once a Day	19%	16%
More than Once a Day	32%	47%
Observations	533	1,311
Number of People Known in Common		
0 to 4	21%	18%
5 to 9	16%	16%
10 to 19	16%	18%
20 to 29	11%	10%
30 or more	37%	39%
Observations	535	1,314
Sometimes Work in Same Room	0%	44%
Observations	537	1,317

Notes: This table presents the distributions of referrers' responses to questions about their relationships with their referrals for two different samples, indicated by the column headings. *Included Referred Workers* are referred workers we hired in either the individual or team experiment. *Excluded Referred Workers* are workers who were referred to us, but who were not included in any experiment.

Table A.9: *Team Communication and Performance Controlling for Communication & Time Spent*
Team Experiment: Base Group is Referred Workers Paired with Someone Else's Referrer (Type B)

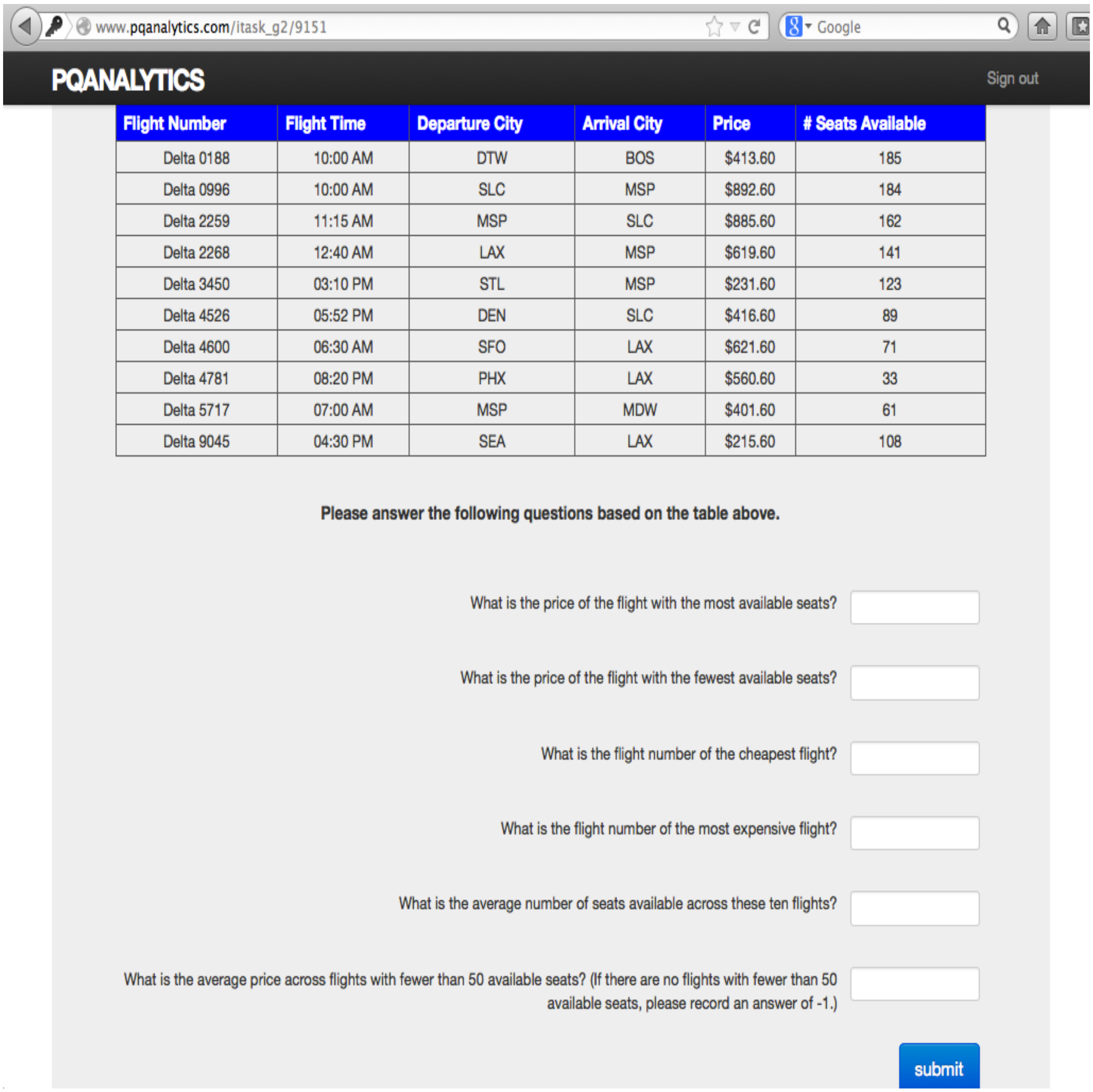
	A. Communication by Team Type			
	Chat Box Use	Total Chat Messages (Conditional on Use)	Answered Communication Question	Reported Outside Communication
Referred Worker and Own Referrer Team (Type A)	0.025	4.346*	-0.028*	0.376**
	(0.042)	(2.544)	(0.016)	(0.040)
Non-Referred Worker and Referrer Team (Type C)	-0.090*	0.805	-0.032	-0.043
	(0.047)	(2.706)	(0.027)	(0.038)
Controls	Yes	Yes	Yes	Yes
Mean of Dependent Variable Base Group (Type B)	0.408	13.522	0.947	0.378
Observations	846	307	846	778
R-Squared	0.047	0.062	0.017	0.193

	B. Team Performance Controlling for Communication and Time Spent			
	Both Submitted	Team Question Matches	Same Slogan	Same Slogan & Both Criteria
Referred Worker and Own Referrer Team (Type A)	-0.028	0.140**	0.225**	0.025
	(0.026)	(0.031)	(0.037)	(0.029)
Non-Referred Worker and Referrer Team (Type C)	-0.144**	-0.112**	-0.066*	-0.014
	(0.035)	(0.041)	(0.038)	(0.029)
Used Chat Box	0.223**	0.167**	0.116**	0.020
	(0.028)	(0.037)	(0.041)	(0.041)
Total Chat Messages	0.001*	0.005**	0.006**	0.003*
	(0.001)	(0.001)	(0.001)	(0.002)
Used Outside Communication	0.236**	0.320**	0.336**	0.173**
	(0.030)	(0.033)	(0.042)	(0.031)
Minutes Spent by Referrer	0.003**	0.002**	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.000)
Minutes Spent by Referred or Non-Referred Worker	0.005**	0.003**	0.002**	0.001**
	(0.001)	(0.001)	(0.000)	(0.000)
Controls	Yes	Yes	Yes	Yes
Mean of Dependent Variable Base Group (Type B)	0.730	0.500	0.337	0.142
Observations	846	846	846	846
R-Squared	0.497	0.478	0.467	0.171

Notes: Each column in each panel reports the results of a separate regression of the dependent variable indicated by the column on indicators for being in a Type A team and for being in a Type C team. Observations are at the worker-PSA level. Chat Box Use is an indicator for whether each partner typed at least one message in the chat box. Total Chat Messages is the aggregate number of messages sent between the two partners, and is conditional on chat box use. Answered Communication Question is an indicator for whether at least one partner responded to the question at the end of that task about how the partners had communicated. Reported Outside Communication is an indicator for whether either partner reported communicating using methods other than the chat box and is conditional on at least one partner having answered the communication question. Regressions in both panels control for the characteristics of referred and non-referred workers listed in footnote 18. Regressions in Panel B also control for whether the team used the chat box, the number of chat messages sent, whether either partner reported using other forms of communication, and (separately) the number of minutes spent by both partners. Standard errors are clustered at the blocking group level. *, ** denote significance at the 10% and 5% levels, respectively.

A.3 Supplementary Figures

Figure A.1: *Individual Experiment Task Site*
Referred and Non-Referred Workers



www.pqanalytics.com/itask_g2/9151

PQANALYTICS

Sign out

Flight Number	Flight Time	Departure City	Arrival City	Price	# Seats Available
Delta 0188	10:00 AM	DTW	BOS	$413.60	185
Delta 0996	10:00 AM	SLC	MSP	$892.60	184
Delta 2259	11:15 AM	MSP	SLC	$885.60	162
Delta 2268	12:40 AM	LAX	MSP	$619.60	141
Delta 3450	03:10 PM	STL	MSP	$231.60	123
Delta 4526	05:52 PM	DEN	SLC	$416.60	89
Delta 4600	06:30 AM	SFO	LAX	$621.60	71
Delta 4781	08:20 PM	PHX	LAX	$560.60	33
Delta 5717	07:00 AM	MSP	MDW	$401.60	61
Delta 9045	04:30 PM	SEA	LAX	$215.60	108

Please answer the following questions based on the table above.

What is the price of the flight with the most available seats?

What is the price of the flight with the fewest available seats?

What is the flight number of the cheapest flight?

What is the flight number of the most expensive flight?

What is the average number of seats available across these ten flights?

What is the average price across flights with fewer than 50 available seats? (If there are no flights with fewer than 50 available seats, please record an answer of -1.)

submit

Figure A.2: *Performance Report Example*

Me
Feb 01

Dear [redacted],

We are writing to update you, as promised, on the performance of a colleague you referred.

Yesterday,
[redacted] did not submit their work.
[redacted] did not submit their work by 11AM PHT.
[redacted] did not answer all of the questions.

Cumulative Performance :
* Percentage of days work was submitted : 80%
* Percentage of days work was submitted by 11AM PHT : 60%
* Percentage of days all questions were answered : 80%

Sincerely,
The Hiring Team

Figure A.3: *Supplemental Experiment Task Site*

We would like you to track the Twitter activity of three artists. Your three artists are (1) Justin Bieber, (2) Taylor Swift, and (3) Miley Cyrus.

Before we begin, what is today's date?

Now, check out how Justin Bieber is trending. The link to his profile is https://twitter.com/justinbieber and for Twitter users, his official handle is @justinbieber.

How many followers does Justin have right now?

How many tweets did Justin himself post yesterday? Please do NOT count retweets.

How many hashtags were there in Justin's own tweets from yesterday? (Recall that "hashtag" refers to the # symbol.)

Next, take a look at Taylor Swift's Twitter profile. The link to her profile is https://twitter.com/taylorswift13 and her official handle is @taylorswift13.

How many followers does Taylor have right now?

How many tweets did Taylor herself post yesterday? Please do NOT count retweets.

How many hashtags were there in Taylor's own tweets from yesterday? (Recall that "hashtag" refers to the # symbol.)

Then look at Miley Cyrus's Twitter profile. The link to her profile is https://twitter.com/MileyCyrus and her official handle is @MileyCyryus.

How many followers does Miley have right now?

How many tweets did Miley herself post yesterday? Please do NOT count retweets.

How many hashtags were there in Miley's own tweets from yesterday? (Recall that "hashtag" refers to the # symbol.)

Figure A.4: *Submission Rates by Day*
Supplemental Experiment

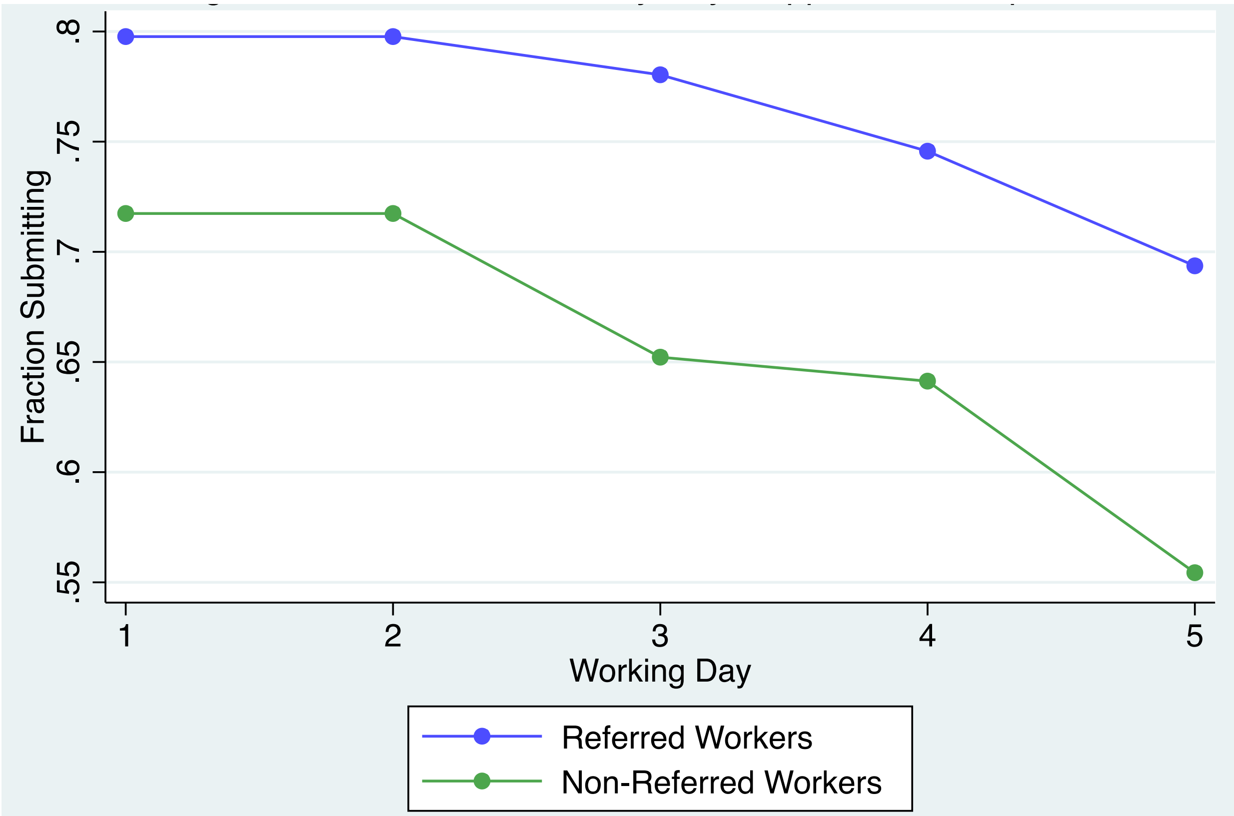
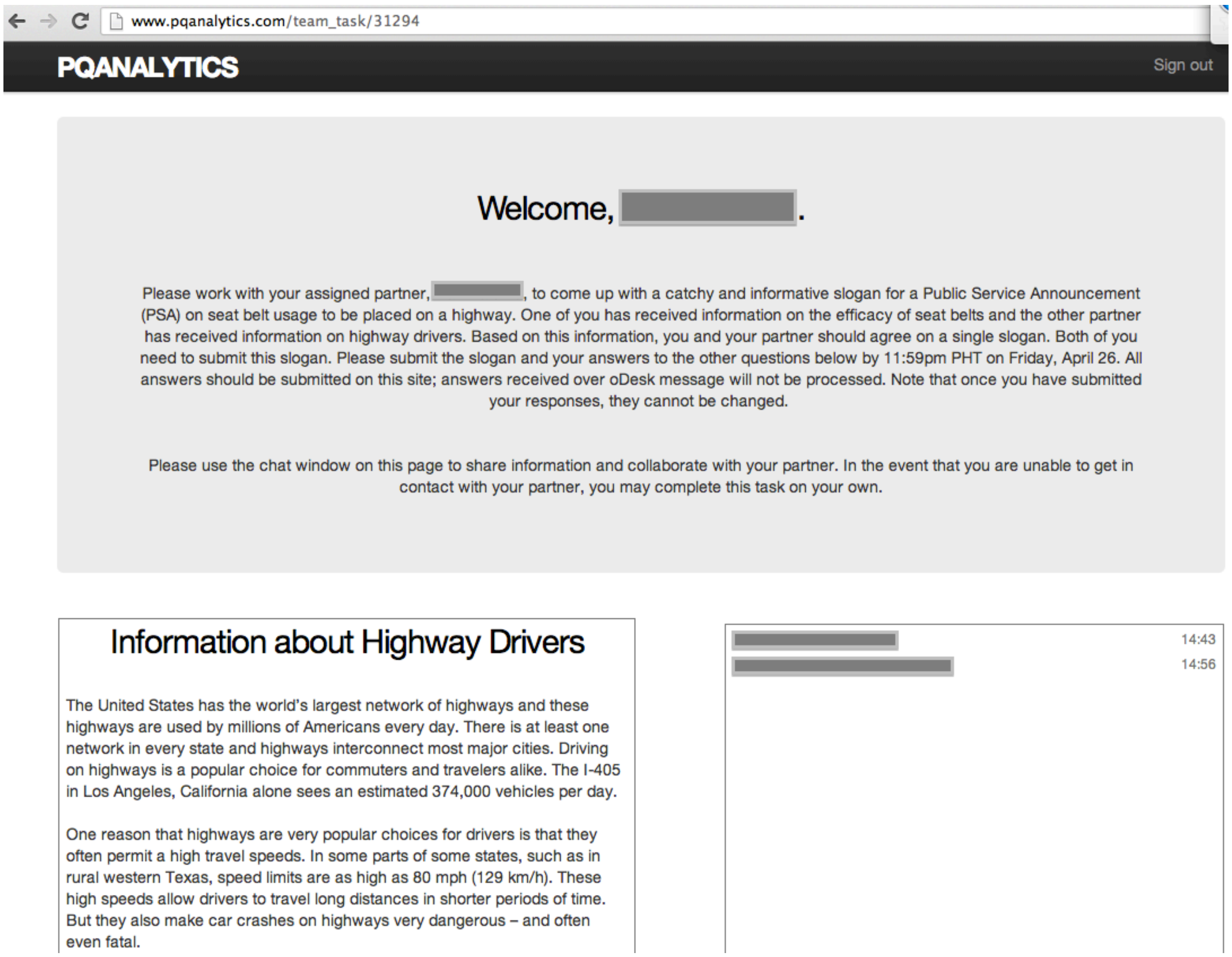


Figure A.5: *Team Experiment Task Site*



Appendix B

Appendix to Chapter 2

B.1 Supplementary Tables

Table B.1: *Descriptive Statistics on Variables Underlying SES Proxy*

		NLSY79		NLSY97	
		Obs.	Mean	Obs.	Mean
Educational Attainment (Years)	Father-Figure	9,027	11.86	7,279	13.00
	Mother-Figure	9,905	11.60	8,234	12.87
Annual Household Income (Nominal Dollars)	Wave 1	5,122	20,538	6,342	52,351
	Wave 2	3,839	22,999	1,053	65,210
	Wave 3	2,674	24,812	1,071	60,019
	Wave 4	1,628	26,226	856	56,188
	Wave 5	651	25,234	783	64,837
	Wave 6	--	--	554	72,843

Notes: Sample restrictions and variable definitions as outlined above. Population weights used throughout.

Table B.2: *Rise of the Female Advantage Similar across SES Quartiles*

	By Quartile				Overall
	Bottom Q	2nd Q	3rd Q	Top Q	
Female	0.206***	0.122	0.063	0.194	0.155***
	(0.080)	(0.085)	(0.093)	(0.120)	(0.047)
NLSY97	0.317***	0.079	0.548***	0.763***	0.417***
	(0.104)	(0.098)	(0.110)	(0.115)	(0.053)
Female x NLSY97	0.300	0.771***	0.923***	0.320**	0.571***
	(0.185)	(0.189)	(0.193)	(0.159)	(0.091)
SES					1.082***
					(0.040)
Female x SES					-0.009
					(0.052)
SES x NLSY97					0.063
					(0.056)
Female x SES x NLSY97					0.021
					(0.082)
Constant	10.758	12.099	12.691	13.314	12.186
	(0.127)	(0.351)	(0.512)	(0.297)	(0.148)
Observations	6,777	4,611	3,980	3,468	18,836
R-squared	0.015	0.021	0.054	0.064	0.206

Notes: This table reproduces Table 2.3 with the inclusion of indicators for race/ethnicity (Black, Hispanic) and for childhood household structure (single-mother household, two-parent household).

Appendix C

Appendix to Chapter 3

C.1 Supplementary Tables

Table C.1: *LASSO Robustness Checks*

Set of Potentials, Count Constraint	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
5 Degree, Choose 1	0.511***	0.270***	0.130***	0.0968***	0.0586***	1.066***
	(0.0513)	(0.0379)	(0.0319)	(0.0230)	(0.0176)	(0.134)
5 Degree, Choose 2	0.457***	0.277***	0.160***	0.109***	0.0660***	1.069***
	(0.0430)	(0.0313)	(0.0258)	(0.0239)	(0.0178)	(0.121)
5 Degree, Choose 3	0.458***	0.278***	0.163***	0.110***	0.0711***	1.081***
10 Degree, Choose 1	0.526***	0.227***	0.0687	0.0545	0.0417	0.918***
	(0.0757)	(0.0580)	(0.0551)	(0.0372)	(0.0283)	(0.213)

Notes: This table presents second stage results for a variety of LASSO specifications. In the first three specifications, the instrument choice set is as follows: national aggregates of maximum temperature indicators in 5 degree F increments (on the interval [10F,100F]), indicator for snow, indicator for rain, precipitation indicators in 0.25 inches per hour increments (on the interval [0,1.5])). From this set, the LASSO approach is set to choose one, two, or three instruments, respectively. In the fourth specfication, a single instrument is again chosen, but the instrument choice is altered; temperature indicators are in 10 degree F increments. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

Table C.2: *Momentum from Viewership Shocks, by Weekend*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
IV	0.493***	0.280***	0.0878	0.102***	0.0630**	1.026***
	(0.0918)	(0.0656)	(0.0584)	(0.0380)	(0.0298)	(0.239)
OLS	0.440***	0.258***	0.154***	0.101***	0.0636***	1.018***
	(0.0173)	(0.0124)	(0.00943)	(0.00723)	(0.00552)	(0.0450)
R-squared	0.674	0.549	0.431	0.366	0.270	0.609

Notes: This table replicates Table 3.2 but observations are defined at the opening weekend by weekend level (528 observations). */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. The first stage results are included in Appendix Table C.3.

Table C.3: *Additional First Stages*

Sample	Instrument	Coefficient	F-Stat
Child-Friendly	0.5-0.75 in	15.87*** (3.601)	19.43
Teen-Friendly	80-85F	-2.075*** (0.614)	11.43
Adults-Only	0.25-0.5 in	-4.149*** (1.100)	14.22
Female	10-15F	3.106*** (0.689)	20.30
Male	0.25-0.5 in	-4.750*** (1.248)	14.47
Predicted Female	Snow	-1.668*** (0.453)	13.58
Predicted Male	75-85F	-1.943*** (0.572)	12.53
Age 45+	35-40F	-2.375*** (0.718)	10.94
Age 30-44	65-70F	1.670*** (0.474)	12.41
Age 18-29	85-90F	3.163*** -0.938	11.38
Predicted Age 45+	35-40F	-2.439*** (0.650)	14.07
Predicted Age 30-44	25-30F	3.528*** (0.929)	13.02
Predicted Age 18-29	90-95F	4.170*** (1.220)	11.68
Top-1000 High-Rated	90-95F	5.165*** (1.410)	13.42
Top-1000 Low-Rated	55-60F	-1.764*** (0.528)	11.17
All High-Rated	65-70F	4.570*** (0.740)	38.11
All Low-Rated	55-60F	-2.798*** (0.615)	20.73
Includes Truncated	65-70F	2.285*** (0.607)	14.17
Main Sample, Weekly	65-70F	9.001*** (2.536)	12.60
High Production Budget	95 - 100F	4.164*** (1.006)	17.15
Low Production Budget	95 - 100F	4.917*** (1.023)	23.09
Weeks 1 through 5	65-70F	6.394*** (1.271)	25.32

Notes: This table presents first stage results from all addditional IV specifications in the paper, along with the corresponding F-statatistic on the excluded instrument(s). In each case, the instrument of choice was chosen with LASSO methods described in the text from the following choice set: national aggregates of maximum temperature indicators in 5 degree F increments (on the interval [10F,100F]), indicator for snow, indicator for rain, precipitation indicators in 0.25 inches per hour increments (on the interval [0,1.5]). */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

Table C.4: *OLS Estimates of Momentum by Movie Quality and Information about Movie Quality*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
			A. by Movie Quality			
High-Rated	0.411***	0.216***	0.126***	0.0754***	0.0492***	0.878***
(obs. 705)	(0.0327)	(0.0204)	(0.0129)	(0.00875)	(0.00624)	(0.0771)
R-squared	0.711	0.548	0.416	0.296	0.190	0.593
Low-Rated	0.412***	0.229***	0.143***	0.0777***	0.0445***	0.906***
(obs. 825)	(0.0152)	(0.0125)	(0.00955)	(0.00616)	(0.00338)	(0.0433)
R-squared	0.746	0.596	0.499	0.380	0.346	0.659
			B. by Information about Movie Quality			
High Budget	0.375***	0.205***	0.122***	0.0802***	0.0564***	0.839***
(obs. 744)	(0.0267)	(0.0184)	(0.0122)	(0.00880)	(0.00657)	(0.0696)
R-squared	0.637	0.483	0.397	0.327	0.264	0.547
Low Budget	0.426***	0.256***	0.145***	0.0771***	0.0378***	0.942***
(obs. 705)	(0.0197)	(0.0147)	(0.00938)	(0.00703)	(0.00466)	(0.0516)
R-squared	0.710	0.558	0.424	0.243	0.095	0.578

Notes: This table presents the corresponding OLS results from Table 3.5. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

Table C.5: *Opening Weekend Viewership Shocks and Ratings*

	IV Estimates	OLS Estimates
Number of Votes	-9.483	11.40*
	(48.31)	(6.493)
R-Squared	--	0.006
High Rating	-0.0102	0.0240*
	(0.0829)	(0.0111)
R-Squared	--	0.009
Low Rating	-0.0803	-0.0427***
	(0.0875)	-0.0117
R-Squared	--	0.024
Difference:		
(High - Low)	0.0701	0.0667***
	(0.875)	(0.118)

Notes: This table shows the relationship between opening weekend sales (in millions) and the film's number of voters. It also shows the relationship between opening weekend sales (in millions)and the film's likelihood of being high rated (top third) and low rated (bottom third). The outcome variables and the endogenous regressor are conditional on week of year, year, and holiday fixed effects. Observations are at the opening weekend level. In aggregating across films that open in the same weekend, we weight each film's rating by the number of screens on which it opened. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. The first stage results are included in Appendix Table C.3.

Table C.6: *OLS Estimates of Network Externalities by Predicted and Realized Gender Demographic*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
			A. Predicted Demographic			
Female Movies	0.413***	0.225***	0.131***	0.0768***	0.0521***	0.897***
(obs. 732)	(0.0578)	(0.0336)	(0.0206)	(0.0129)	(0.00939)	(0.130)
R-squared	0.734	0.585	0.470	0.335	0.244	0.641
Male Movies	0.397***	0.228***	0.135***	0.0879***	0.0585***	0.906***
(obs. 813)	(0.0238)	(0.0166)	(0.0114)	(0.00853)	(0.00662)	(0.0640)
R-squared	0.754	0.644	0.527	0.445	0.367	0.687
			B. Realized Demographic			
Female Movies	0.371***	0.191***	0.112***	0.0633***	0.0411***	0.779***
(obs. 777)	(0.0539)	(0.0292)	(0.0193)	(0.0118)	(0.00809)	(0.117)
R-squared	0.693	0.518	0.369	0.234	0.170	0.564
Male Movies	0.403***	0.217***	0.122***	0.0743***	0.0452***	0.861***
(obs. 882)	(0.0276)	(0.0150)	(0.00942)	(0.00636)	(0.00395)	(0.0604)
R-squared	0.754	0.644	0.527	0.445	0.367	0.687

Notes: Panel A presents the corresponding OLS results from Table 3.6; Panel B presents the corresponding OLS results from the specifications in Appendix Table C.7. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

Table C.7: *Network Externalities by Realized Gender Demographic*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Female Movies	0.523***	0.326***	0.188***	0.157***	0.0617**	1.256***
(obs. 777)	(0.0692)	(0.0634)	(0.0427)	(0.0432)	(0.0270)	(0.194)
Male Movies	0.362***	0.199***	0.111***	0.0333	0.0169	0.722***
(obs. 882)	(0.0813)	(0.0510)	(0.0377)	(0.0281)	(0.0188)	(0.190)
Difference:						
(Female - Male)	0.161	0.127	0.077	0.124***	0.0448	0.534**
	(0.107)	(0.081)	(0.057)	(0.052)	(0.033)	(0.272)

Notes: This table replicates Table 3.2 separately by the movie's realized gender demographic. Female movies fall in the top third in percentage of IMDB voters that are female, male movies in the bottom third (i.e., top third in percentage of IMDB voters that are male). */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. The first stage results are included in Appendix Table C.3.

Table C.8: *OLS Estimates of Network Externalities by Age Suitability*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Child-Friendly Movies	0.475***	0.302***	0.181***	0.111***	0.0720***	1.141***
(obs. 688)	(0.0188)	(0.0127)	(0.00909)	(0.00614)	(0.00483)	(0.0442)
R-squared	0.764	0.735	0.599	0.508	0.416	0.77
Teen-Friendly Movies	0.407***	0.221***	0.133***	0.0826***	0.0531***	0.897***
(obs. 1217)	(0.0228)	(0.0149)	(0.0101)	(0.00732)	(0.00513)	(0.0572)
R-squared	0.696	0.541	0.438	0.324	0.241	0.595
Adults-Only Movies	0.448***	0.257***	0.150***	0.0880***	0.0540***	0.996***
(obs. 909)	(0.0169)	(0.0134)	(0.00880)	(0.00624)	(0.00527)	(0.0486)
R-squared	0.799	0.641	0.509	0.375	0.25	0.675

Notes: This table presents the corresponding OLS results from Table 3.7. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

C.2 Holiday Controls

Our holiday indicators are exactly those of Dahl and DellaVigna (2009), and are similarly motivated by the fact that holidays impact movie audience sizes (usually positively), the effect varies across holidays, and audience sizes are often also impacted in the days just around each holiday. We include indicators for Martin Luther King Day, President's Day, Memorial Day, Labor Day, Columbus Day, Independence Day, Veteran's Day, Easter, Thanksgiving Day, Christmas Eve, Christmas, New Year's Eve, New Year's Day, St. Patrick's Day, Valentine's Day, Halloween, Cinco de Mayo, and Mother's Day. We also include separate indicators for the Friday, Saturday, and Sunday before each of MLK Day, President's Day, Memorial Day, Labor Day, and Columbus Day; for the Friday and Saturday before Easter; for the Wednesday before Thanksgiving and for the weekend after; for the three days before Christmas Eve (December 20 - 23) and the four days after Christmas (December 26 - 30); and for the two days after New Year's Day (January 2 - 3). Finally, for Independence Day, Veteran's Day, Christmas, New Year's, and Valentine's Day, we include an indicator for whether each falls on a Saturday or Sunday. Several of these indicators drop out when we restrict our sample to movie weekends (Friday, Saturday, Sunday) only.

C.3 Network Externalities by Adult Audience Age

We here discuss differences in observed network externalities by adult ages. IMDB provides adult voter ages in the following categories: 18 to 29, 30 to 44, and 45 and up. As with gender classifications, we first classify all PG-13 and R movies by predominant audience age based on the *predicted* age composition of IMDB voters using the same leave-one-out method; we then estimate network externalities by predicted age demographic.

Among adults, we observe the largest network externalities among the oldest demographic. Appendix Table C.9 shows estimated adult network externalities for each of the three predicted age audiences: Age 45 Plus, Age 30 to 44, and Age 18 to 29. The final two rows report the difference in network externalities between age 45-plus movies and movies in each of the two younger age categories. Movies categorized as age 30 to 44 are estimated to have network externalities almost as large as movies categorized as age 45 and up, but the difference between movies for the youngest group (18 to 29) and those for the oldest (45 and up) are large and generally significantly different at the 1% level. When an unexpected weather shock opening weekend causes an additional 100 people to attend a movie popular with the 45 plus age demographic, more than 180 additional viewers attend in the next five weekends; for movies popular with the 20-somethings, though, an additional 100 in opening weekend audience size results in only about half as much total momentum (less than 90 additional viewers in later weekends).

Table C.9: *Network Effects by Predicted Adult Age*

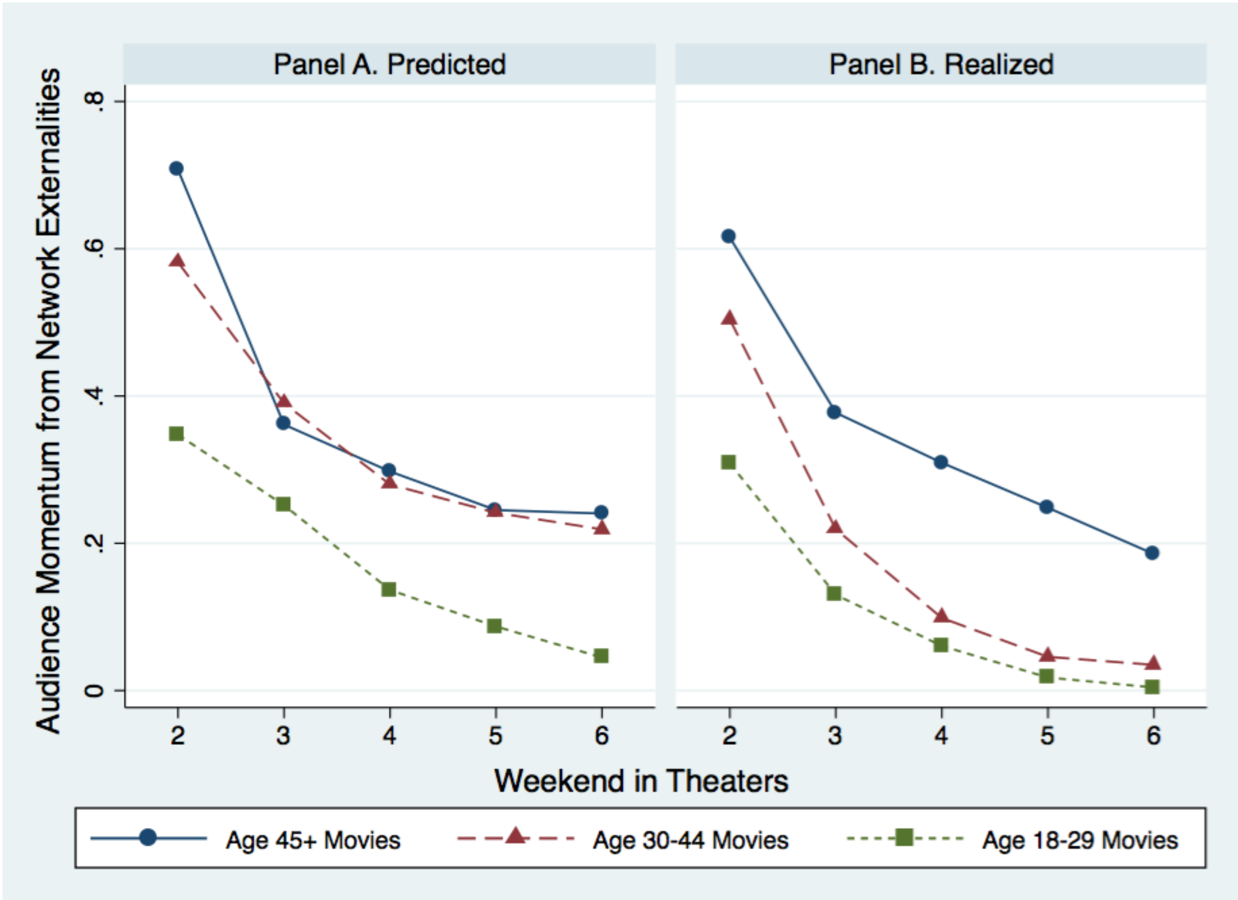
	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Age 45 Plus Movies	0.707***	0.361***	0.298***	0.245***	0.240***	1.851***
(obs. 744)	(0.104)	(0.0605)	(0.0538)	(0.0508)	(0.0567)	(0.268)
Age 30 to 44 Movies	0.582***	0.390***	0.280***	0.241***	0.219***	1.712***
(obs. 756)	(0.0618)	(0.0536)	(0.0534)	(0.0487)	(0.0479)	(0.244)
Age 18 to 29 Movies	0.347***	0.252***	0.137***	0.0870***	0.0449***	0.868***
(obs. 771)	(0.0768)	(0.0585)	(0.0380)	(0.0262)	(0.0171)	(0.203)
Differences:						
(45 Plus - 30 to 44)	0.125	-0.029	0.018	0.004	0.021	0.139
	(0.121)	(0.081)	(0.076)	(0.070)	(0.074)	(0.362)
(45 Plus - 18 to 29)	0.360***	0.109	0.161***	0.158***	0.195***	0.983***
	(0.129)	(0.084)	(0.066)	(0.057)	(0.059)	(0.336)

Notes: This table replicates Table C.7 except that gender demographics are here predicted using the leave-one-out method described in the text. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. The first stage results are included in Appendix Table C.3.

Panel A of Appendix Figure C.1 shows the estimated network externalities by week in theater for each of the three adult age categories; movies with older predicted demographics exhibit larger momentum effects in each week. We can also classify movies by predominant audience age based on the actual age composition of the movie's voters on IMDB. [1] Under this classification, we similarly observe the highest network externalities in the oldest demographic group and the lowest network externalities in the youngest demographic, but the estimated network effects for "Age 30 to 44" movies then fall more directly in between (Panel B).

[1]Because the age distribution of IMDB voters has changed over the decade duration of our sample, we define these variables within release year so as not to confound our age results with time trends. There are far fewer older voters than younger voters. To qualify for the 45 plus age group, a movie needs only 2% of voters in that age range; for the 30 to 44 age group, the corresponding number is 16%. Meanwhile, only movies with more than the majority of voters aged 18 to 29 (53%) fall in the top third in percentage voters in that age range.

Figure C.1: *Network Externalities by Predicted and Realized Adult Age Demographic*



Notes: Panel A plots the coefficients from Panel A of Table C.9 for each of Weekends 2 through 6; Panel B plots the coefficients from Panel A of Appendix Table C.10

Appendix Table C.10 shows the corresponding regression results. When an unexpected weather shock opening weekend causes an additional 100 people to attend a movie popular with the 45 plus age demographic, more than 170 additional viewers attend in the next five weekends. The number for movies popular among 30 to 44 year olds is just over half that (90 viewers). And for movies popular with the 20-somethings, an additional 100 in opening weekend audience size results in only about 50 additional viewers in later weekends.

Table C.10: *Network Externalities by Realized Adult Age*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Age 45 Plus Movies	0.615***	0.377***	0.309***	0.248***	0.185***	1.736***
(obs. 747)	(0.0673)	(0.0505)	(0.0558)	(0.0538)	(0.0441)	(0.233)
Age 30 to 44 Movies	0.504***	0.219***	0.0986***	0.0458*	0.0349*	0.902***
(obs. 705)	(0.0550)	(0.0441)	(0.0343)	(0.0270)	(0.0182)	(0.161)
Age 18 to 29 Movies	0.308***	0.130**	0.0606	0.0177	0.00417	0.521**
(obs. 783)	(0.0785)	(0.0604)	(0.0463)	(0.0359)	(0.0288)	(0.228)
Differences:						
(45 Plus - 30 to 44)	0.111	0.158***	0.210***	0.202***	0.150***	0.834***
	(0.087)	(0.067)	(0.065)	(0.060)	(0.048)	(0.283)
(45 Plus - 18 to 29)	0.307***	0.247***	0.248***	0.230***	0.181***	1.213***
	(0.103)	(0.079)	(0.073)	(0.065)	(0.053)	(0.326)

Notes: This table replicates Table 3.2 separately by the movie's actual adult age demographic. Age 45 plus movies fall in the top third in percentage of IMDB voters aged 45 and up; similarly for age 30 to 44, and for age 18 to 29 movies. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. The first stage results are included in Appendix Table C.3.