Essays in Behavioral Economics and Innovation

Essays in Behavioral Economics and Innovation

A dissertation presented

by

Duncan Sheppard Gilchrist

to

The Committee on Higher Degrees in Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

December 2014

Dissertation Advisors:
Professor David Cutler
Professor Andrei Shleifer

Author:
Duncan Sheppard Gilchrist

Essays in Behavioral Economics and Innovation

Abstract

This dissertation consists of three essays, two in behavioral economics and one on the economics of innovation.

The first essay, which is joint work with Emily Glassberg Sands, exploits the randomness of weather and the relationship between weather and movie-going to quantify network externalities (i.e., a preference for shared experience) in movie consumption. Instrumenting for early viewership with unanticipated and plausibly exogenous weather shocks captured in LASSO-chosen instrument sets, we find that a shock to opening weekend viewership is doubled over the following five weekends. Our estimated momentum arises almost exclusively at the local level, and varies neither with ex-post movie quality nor with the precision of ex-ante information about movie quality, suggesting the observed momentum is unrelated to learning.

The second essay, which is joint work with Michael Luca and Deepak Malhotra, asks whether higher wages elicit reciprocity and hence higher productivity. In a field experiment with 266 employees, we find that paying above-market wages, per se, does not have an effect on productivity relative to paying market wages (in a one-time job with no future employment opportunities). However, structuring a portion of the wage as a clear and unexpected gift–by offering a raise (with no additional conditions) after the employee has accepted the contract–does lead to higher productivity for the duration of the job. Targeted gifts are more efficient than hiring more workers. However, the mechanism underlying our effect makes this unlikely to explain persistent above-market wages.

Finally, the third essay examines how an incumbent's patent protection acts as an implicit subsidy towards non-infringing substitutes. I analyze whether classes of pharmaceuticals whose first entrant has a longer period of market exclusivity (time between approval and generic entry) see more innovation. Instrumenting for exclusivity using plausibly exogenous delays between patent filing and the start of clinical trials, I find that one extra year of first in class exclusivity increases subsequent entry by 0.2 drugs. The effect is stronger for drugs targeting conditions for which demand is more price-elastic, and for drugs that are lesser advances.

Contents

List of Tables

List of Figures

Acknowledgments

I am deeply grateful for the advice and encouragement I received from my committee members, David Cutler, Andrei Shleifer, and Michael Luca. Though they were not on my committee, I am grateful to Ariel Pakes and Greg Lewis for extensive support, and also to Al Roth, who has been an exceptional mentor.

My colleagues in graduate school have helped me tremendously over the years. Special thanks to Natalie Bau, Aubrey Clarke, Thomas Covert, Peter Ganong, Ben Hebert, Sabrina Howell, Stephanie Hurder, Simon Jaeger, Rohan Kekre, Ben Lockwood, Mikkel Plagborg-Moller, Guillaume Pouliot, Martin Rotemberg, Bryce Steinberg, and Thomas Wollmann.

For thoughtful comments on the first chapter, I would like to thank Victor Chernozhukov, Mark Duggan, Liran Einav, Hank Farber, Kevin Garewal, Ed Glaeser, Claudia Goldin, Christian Hansen, Larry Katz, Jeffrey Miron, Markus Mobius, Amanda Pallais, Adam Presser, Jesse Shapiro, Fanyin Zheng, as well as seminar participants at Harvard's labor and public economics lunch and five referees. The second chapter benefited from Kristiana Laugen's research support, as well as from comments from Uri Gneezy, an associate editor, and three referees. Additional acknowledgement for the third chapter is due to Philippe Aghion, Kate Baicker, Lanier Benkard, Ernie Berndt, Dan Carpenter, Michael Chernew, Liran Einav, Richard Freeman, David Friedman, Kevin Garewal, Jerry Green, Shane Greenstein, Rob Huckman, Haiden Huskamp, Michael Lanthier, Mark Lemley, Robin Lee, Jon Levin, Hong Luo, Eric Maskin, Genevieve Pham-Kanter, Benjamin Roin, Bhaven Sampat, Vicki Sato, Kevin Sharer, Ariel Dora Stern, Stuart Watt, Jeff Way, Heidi Williams, and Ali Yurukoglu, as well as seminar participants at the Harvard IO lunch, the Harvard HCP Seminar, and the NBER Productivity Seminar. Financial support from the Harvard Business School Doctoral Program is gratefully acknowledged.

Finally, I would like to thank my family for their love and support.

To my parents, Donald and Suzanne.

Chapter 1

Something to Talk About: Network Externalities in Movie Consumption[1]

1.1 Introduction

We grab a bite at a bustling restaurant and then kick back at a top-selling movie. Neither the meal nor the film is necessarily remarkable. So why do we follow the crowd?

The tendency to follow in the footsteps of others has been observed in decisions ranging from which stocks to buy to how many children to bear, which technologies to adopt, and whether or not to apply for disability insurance. Much of the existing theoretical research on crowd-following focuses on the role of information. Prominent examples include models of 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 private information about the quality of a good or experience, and so takes into account the observed choices and/or reports of others in making her own decision. In an array of contexts, observational and experimental studies have found strong empirical evidence of

[1]Co-authored with Emily Glassberg Sands

[2]See, for example, Banerjee (1992); Bikhchandani *et al.* (1992, 1998); Ellison and Fudenberg (1995); McFadden and Train (1996); Çelen and Kariv (2004).

information stories driving convergent, or herd, behavior.[3]

This paper analyzes a very different explanation for crowd-following: network externalities in consumption. Amid classic network externalities, an individual's demand for a good depends on the demands by other consumers (Becker (1991)). Although network externalities and information stories can certainly coexist (see, e.g., Choi (1997)), there is no direct role for either quality or information about quality in network externalities themselves. Rather, a good or experience is simply more useful or more enjoyable when others share in it.[4] Some goods have network externalities by construction. Facebook, Twitter, or Instagram, for example, are more useful the more peers are 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 the consumption of a major entertainment good, in-theater movies. The thought experiment is simple: holding all other characteristics of a movie fixed, does an individual's demand for the movie depend on whether others have seen it? We begin by presenting a simple theoretical model of movie-going in the presence of network externalities. The model is designed to clarify the mechanism and guide the empirical analyses that follow: consumers value share experiences, so the utility of film attendance at any point depends on both movie quality and on the film's prior viewership. These consumer are assumed to be myopic, and thus attend movies the first time the value of attendance exceeds its opportunity cost. The model illustrates how, through network externalities, sales in one weekend increase the utility of attendance, and thus viewership,

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

[4]The mechanisms underlying any preference for shared experience could vary. For example, an individual might value consuming an experience in parallel with others, or she might value having had a common experience around which she can interact with others. Alternatively, she may value simply knowing her experience is shared.

[5]A large literature in industrial organization studies the effects of network externalities in platform markets, particularly with regards to the market power they can induce through consumer lock-in. See Farrell and Klemperer (2007) for an overview. Relatedly, at the firm level Katz and Shapiro (1986) analyze technology adoption in the presence of network externalities.

in subsequent weekends. It also shows how momentum generated by network externalities need not depend on anything aside from the strength of the externalities themselves; that is, it can be independent of other movie characteristics, including both movie quality and the precision of prior information about quality.

To test for and quantify network externalities in movie consumption, our empirical strategy is to exploit the randomness of weather, and the relationship between weather and movie-going.[6] In brief, we instrument for opening weekend viewership with unanticipated and plausibly exogenous weather shocks that weekend. We then 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 is doubled over the following five weekends.

While theorists have long posited that network externalities may contribute to convergent choices across individuals, empirical evidence on network externalities is limited. The work perhaps closest to our own is an experiment by Bursztyn *et al.* (2014), which finds that investors' stock choices are influenced by network externalities. Randomizing both whether an investor's planned stock purchase decision is shared with another investor and whether that purchase is executed, they show that investors value actually owning the same stocks as others (over and above the information contained in others' planned stock purchases).[7] Quantifying the effects of network externalities with observational data, however, is generally challenging because of the simultaneity problem that plagues studies of peer effects (see, e.g., Manski (1993)). In observational data it is difficult to identify one person's causal effect on another because the outcome of interest is generally observed for both individuals at the same time. We overcome this issue by removing the simultaneity of the decision: we analyze the effect of an exogenous (weather-driven) shock to opening weekend movie viewership on demand for that movie in subsequent weekends.

---

[6] We note that Dahl and DellaVigna (2009) are the first to document that weather shifts movie sales.

[7] Recent work by Lahno and Serra-Garcia (2014) in a related setting also suggests that knowledge of a peer's risk-taking can impact behavior.

In the presence of network externalities in consumption, subsequent demand for a movie would be increasing in opening weekend viewership, but there are many reasons we might expect a movie's viewership to be positively correlated over time. First, and perhaps 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). Second, even if people had different information about quality at different points in time, a learning model could predict momentum. In work related to our own, Moretti (2011) analyzes social learning in movie-going, and Cabral and Natividad (2013) provide evidence that top-sellers in opening weekends tend to earn significantly more in subsequent weekends, primarily due to increased awareness.[8] Third, quality and information about quality aside, viewership could also be correlated over time if over the course of the movie's run people are subject to similar supply shocks (e.g., an unusually appealing movie trailer) or to similar demand shocks (e.g., a close World Series that leaves people tied to their home televisions).

By exploiting weather shocks in a movie's opening weekend as a plausibly exogenous source of variation in opening weekend viewership, our empirical specifications are designed to identify raw social multipliers purged of these potential confounders.[9] In the first stage, we instrument for opening weekend viewership with weather shocks occurring that weekend. Controlling for general seasonality, these unanticipated weather shocks are orthogonal to

[8]Moretti (2011)'s empirical analysis actually tries to rule out network externalities in movie-going; however, as we show in Appendix Section A.5, his results are not robust to the inclusion of seasonal controls. Analyzing social learning in another medium, 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.

[9]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 *et al.* (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.

unobserved demand and supply shocks, and to movie quality. In the 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 viewership in weekends subsequent to opening on contemporaneous weather.

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 weather measures is large, particularly at the national level.[10] The risk of 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 without objective rhyme or reason) make careful variable aggregation and selection methods crucial in this setting. Given the large set of potential weather instruments, we implement Least Absolute Shrinkage and Selection Operator (LASSO) methods as developed in Belloni *et al.* (2010) to estimate optimal instruments in linear IV models with many instruments. We leave the details for later but, in brief, we run a penalized least-squares regression of the first stage outcome on a large set of potential instruments; LASSO then machine-selects the instruments which are sufficiently explanatory to justify their associated penalties, and these instruments are included in the first stage. Existing research using LASSO is largely theory- and simulation-based; to our knowledge, we are the first non-methodological paper to use LASSO techniques for instrumental variable selection.

Instrumenting for opening weekend viewership with weather shocks that weekend, we find strong evidence of large and persistent momentum due to network externalities in consumption. For 100 weather-induced additional viewers opening weekend, we observe almost fifty additional viewers in the second weekend and almost thirty the third. Consis-

[10]Consider a simple Google search of "02138 weather," which yields a wealth of information including Cambridge's hourly maximum temperature, probability of precipitation, humidity, wind speed, and cloud cover. Moreover, such weather measures are available for each of thousands of weather stations.

tent with model's predictions, the impact of opening weekend viewership on viewership in subsequent weekends falls approximately exponentially over time, yet the aggregate impact of network externalities remains large: by the sixth weekend, cumulative momentum has yielded more than one subsequent viewer for each additional viewer during opening weekend. Although our preferred estimates are generated using LASSO-chosen instruments, instrumenting with intuitive, hand-selected instruments yields similar results. We additionally present evidence that our instrument is indeed exogenous, and that the observed shift is a demand-side phenomenon: prices are fixed and supply responses (e.g., adjustments to the number of screens on which the movie shows or changes in its duration in theaters) can explain little, if any, of our estimated momentum.

Do these network externalities exist predominately at the local level (e.g., through conversations among friends) or at the national level (e.g., by way of national media coverage of box office sales)? Our model does not speak specifically to this distinction but it is helpful for interpretation. Since local movie ticket sales data are typically unavailable to researchers, we proxy for local viewership with MSA-level Google search volume data, and show that the lion's share of our estimated momentum is bred within-MSA.

Although our empirical strategy is designed to isolate momentum from network externalities, we also test for evidence of an important alternative explanation: learning. The theoretical literature on learning models is large and varied in both its assumptions and its predictions. To make progress, we focus on two intuitive predictions of prominent learning models: (1) in the presence of social learning, shocks to viewership should induce stronger momentum for high quality movies than low quality movies, and (2) in the presence of observational learning, shocks to viewership should induce stronger momentum for movies with more diffuse ex-ante priors. Importantly, these predictions need not arise from a model of network externalities, and indeed are not implied by our model. Proxying for realized vertical quality with critic reviews, and for the precision of information about quality with production budgets, we find no evidence that our estimated momentum varies along either dimension.[11] Although our estimates do not rule out some role for learning, taken together the results suggest network externalities are a prominent driver of the momentum we estimate.

Finally, we analyze the economic implications of our results. We first ask where the network externality-induced viewers are coming from. Looking across all movies available in theaters simultaneously, we show that viewers are predominately substituting across movies, rather than across activities. Second, we leverage our framework to ask whether certain groups are disproportionately affected by network externalities; we find that our estimated momentum is nearly 50% stronger among viewers of child-friendly movies (MPAA rating of G or PG) than among viewers of adult-oriented movies (PG-13 or R). We conclude by discussing the absolute and relative magnitudes of the social multiplier we estimate, and the implications for firms.

The remainder of the paper proceeds as follows: We open in Section 1.2 with a simple model of movie-going in the presence of network externalities. Section 1.3 then details our national movie and weather data, our main empirical approach (including our instrument selection methods), and our first stage results. Our baseline estimates of network externalities follow in Section 1.4. In Section 1.5, we proxy for local movie going using MSA-level Google search data, and show that the observed momentum is bred principally at the local level. Section 1.6 considers learning as an alternative explanation for our results, Section 1.7 explores the economic implications of our estimated momentum, and Section 1.8 concludes.

1.2 A Simple Model of Network Externalities in Movie-Going

In this section, we present a simple model designed to illustrate the mechanism through which network externalities can shape demand for movie tickets. The model yields three

[11] We note that we cannot rule out a model of pure conformity in which choices are influenced by a preference for social esteem, as in Bernheim (1994). However, we do not expect a preference for social esteem would play much role in the decision to see a given movie, in particular since public perceptions about an individual's predispositions are unlikely to be impacted 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.

predictions, which we return to in subsequent empirical sections. Our framework is highly stylized and does not capture all features of the market, and we discuss its limitations below.

1.2.1 Preliminaries

There is a unit mass of co-located consumers who value both seeing a good movie and *sharing in the experience* of movie-going. In particular, the utility of attending a given movie is increasing both in the movie's quality and in cumulative local viewership of the movie. For simplicity, we focus here on demand in a single locality; extending the model to include multiple localities yields similar results.

The set up is as follows: Before movie j's release, each consumer i observes its quality, α_j, and her idiosyncratic valuation of viewing, ϵ_{ij}. The consumer attends the movies on her own, but each weekend she learns how many others in her locality attended in weeks prior. Thus, on any weekend of showing $w > 0$, demand for tickets is increasing in movie j's cumulative prior local viewership, denoted CPV_{jw}. (Cumulative prior viewership on opening weekend is zero.)

In any weekend a movie is showing, the consumer compares her utility of attending that movie with the opportunity cost of viewership, c, and attends if the difference is at least zero.[12] Altogether, consumer i's utility from viewing movie j on weekend w is

$$U_{ijt} = \alpha_j + \epsilon_{ij} + \lambda CPV_{jw}, \tag{1.1}$$

where the parameter $\lambda \geq 0$ captures the extent to which consumers value shared experiences. That is, λ parameterizes the strength of network externalities. If $\lambda = 0$, then the value of shared experiences drops out of the utility specification completely, while at larger λ's consumers receive additional utility from movie-going the more others have already attended.

To close the model, we assume that the idiosyncratic valuations of viewing, ϵ_{ij}, are

[12] An alternative interpretation of ϵ_{ij} is that individuals differ in the opportunity cost of movie attendance (and c is the mean opportunity cost).

uniformly distributed on $[0,1]$, that consumers receive positive utility only from their first viewing of a movie, and that consumers are myopic. These assumptions make the analysis tractable, ensure that demand for movies is decreasing in weekends since release, and rule out the possibility that consumers choose to delay viewing until more of their peers have already viewed. Finally, we assume that the parameters are such that an interior solution exists, so cumulative viewership is always at least 0 and never exceeds 1 (i.e., $\lambda < 1$).[13]

Before turning to the analysis and predictions, we note that this model is deliberately held simple. Importantly, the model assumes movie qualities are public and certain. Incorporating uncertainty in movie quality, α_j, on its own does not change our results as long as we do not also allow consumers to have private information about movie quality.[14] However, that consumers do not have heterogenous beliefs about movie quality explicitly rules out the potential for consumers to learn from one another, e.g. as in Moretti (2011). We return to learning models, and the differences in their assumptions and predictions, in Section 1.6.

1.2.2 Analysis

We solve for viewership in each weekend and then discuss the model's predictions. Opening weekend viewership for movie j is simply given by the range of consumers who receive positive utility from attendance without any utility gain from prior viewership:

$$V_{j1} = 1 - (c - \alpha_j). \tag{1.2}$$

Working forward, cumulative prior viewership in subsequent weekends $w > 1$ is then

$$CPV_{jw} = 1 - (c - \alpha_j - \lambda CPV_{j,w-1}), \tag{1.3}$$

[13]It will become clear below that if $\lambda \geq 1$, then viewership grows exponentially in weekends subsequent to opening, making an interior solution infeasible.

[14]As consumers can form expectations of quality, our results still hold even if α_j has an arbitrary distribution.

and solving the recursion yields

$$CPV_{jw} = \left(1 - (c - \alpha_j)\right) \sum_{\tau=0}^{w-1} \lambda^{\tau}. \tag{1.4}$$

Thus, viewership in weekend w is given by

$$V_{jw} = \lambda^{w-1} \left(1 - (c - \alpha_j)\right). \tag{1.5}$$

This simple model yields three main predictions.

Prediction 1 *The ratio of viewership in weekend $w > 1$ to viewership opening weekend decreases exponentially in weeks since opening:*

$$\frac{V_{jw}}{V_{j1}} = \lambda^{w-1}. \tag{1.6}$$

That is, the model predicts that viewership in any weekend w is simply a fraction λ^{w-1} of viewership opening weekend. Correspondingly, viewership in a given weekend is simply a constant fraction λ of viewership in the weekend prior. We return to this prediction in Sections 1.4 and 1.5.

Prediction 2 *The relationship between opening viewership and viewership in any subsequent weekend, w, is not influenced by movie quality.*

That $\frac{V_{jw}}{V_{j1}}$ is independent of α_j is apparent from Equation (1.6). This is not to say that viewership itself is independent of quality; viewership in any weekend w is actually increasing in α_j, and it is straightforward to show that $\frac{\partial V_{jw}}{\partial \alpha_j} > 0$. Rather, the model predicts that it is subsequent viewership *relative to opening viewership* which does not depend on quality. We return to this prediction in Section 1.6, where we compare the predictions of our model to learning models.

Prediction 3 *Stronger network externalities increase viewership in subsequent weekends relative to opening weekend:*

$$\frac{\partial}{\partial \lambda}\left(\frac{V_{jw}}{V_{j1}}\right) > 0 \textit{ for } w > 1. \tag{1.7}$$

The model predicts that the more viewers value the viewership of others – i.e., the stronger are network externalities – the larger is viewership in all subsequent weekends relative to viewership in opening weekend. We return to this prediction in Section 1.7, where we explore heterogeneity in the magnitudes of momentum from network externalities across audience demographics.

1.3 Estimating Network Externalities at the National Level

To isolate momentum from network externalities, as opposed to from unobservable movie quality or from other supply or demand shocks, we 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. In this section, we first detail our national ticket sales data and our nationally-aggregated weather measures. We then present our main empirical strategy, including our instrument selection methods, followed by our first stage results.

1.3.1 National Tickets and Nationally-Aggregated Weather Measures

Our national box office data comes from BoxOfficeMojo, a reporting service owned by IMDB, and includes daily ticket sales in the U.S. In the weeks just following release (when a movie can generally be viewed exclusively in theaters), box office data provide an excellent measure of a movie's total sales.[15] Our ticket sales sample is comprised of all movies

[15]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.

wide-released in U.S. theaters between January 1, 2002 and January 1, 2012.[16] We track audience sizes during the six weeks following the date of wide-release. To avoid truncation issues, the 19% of films that do not last at least six weeks in theaters are excluded from our main analysis.[17] 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%) of ticket sales.[18]

Figure 1.1 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 1,381 movies in our sample. Average daily ticket sales exceed one million during opening weekend, but fall off exponentially 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.[19] Our unit of observation for audiences, then, is at the opening weekend by date level. In our eleven-year sample we observe 557 opening weekends, or 1,671 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 flocking to new releases is 2.5 million viewers. The corresponding number for movies in their second weekend is just over 1.3 million; this falls to about 200,000 by the sixth weekend in theaters.

---

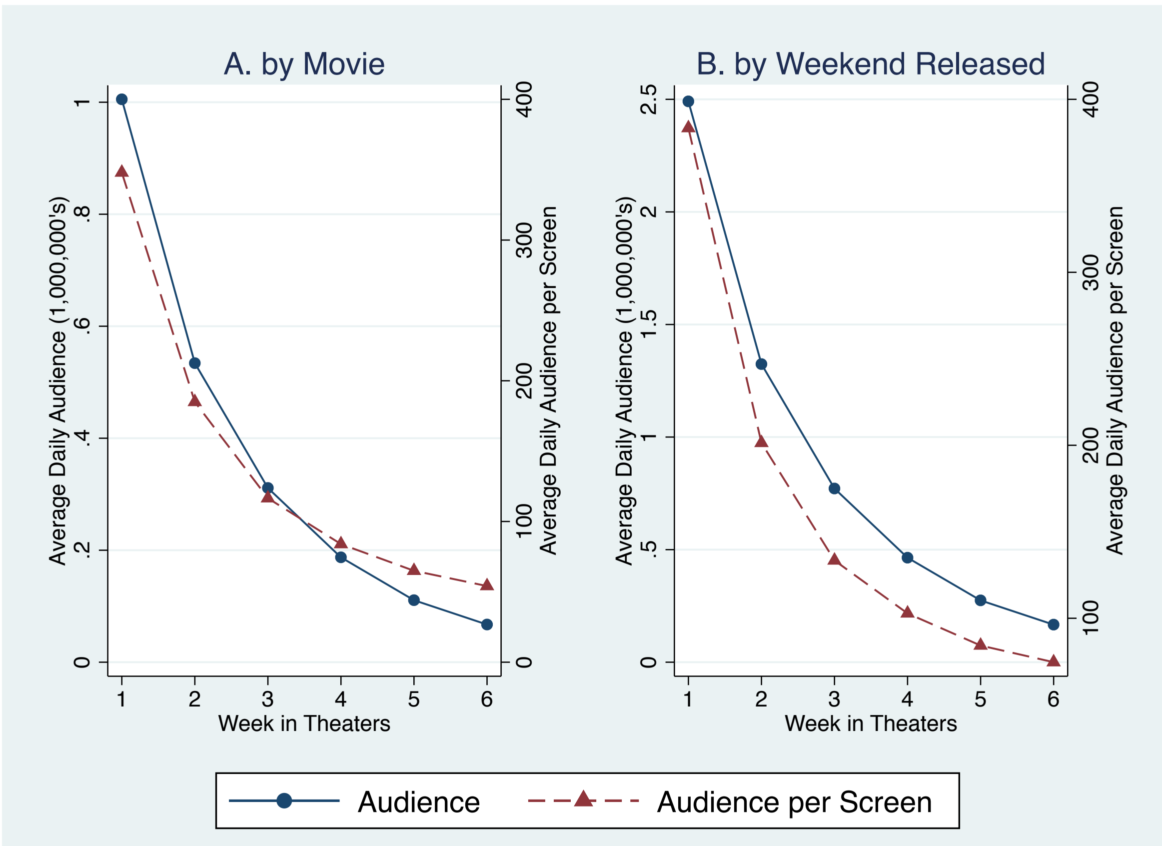
[16]We follow Corts (2001) and Einav (2007) in defining as "wide-released" any movie that ever showed on 600 or more screens, and omit from the sample the less than 1% of movies that never reached wide release. For the 20% of films in our sample that start with a limited release before reaching wide 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% of the eventual maximal number of screens for that movie. (Excluding the limited-release films from the sample does not substantively change our results.) Alhough 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.

[17]We return to them when examining supply responses in Appendix Section A.6 and show that our results are robust to their inclusion. Since BoxOfficeMojo also reports the number of screens on which the movie shows weekly, in that Appendix Section we also analyze any supply shifts that might impact our observed quantity effects. Combined with total ticket sales quantities, the number of screens (supply) facilitates an isolation of demand shifts since ticket prices are generally fixed.

[18]Restricting to weekend audiences is standard in this literature; see, e.g., Dahl and DellaVigna (2009).

[19]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 1.1: *Average Audience Sizes by Week in Theater*



Notes: For our sample of 1,381 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.

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 provider of real-time and historical weather information online, with most U.S. data coming from the National Weather Service, and we observe daily weather measures for each of 1,941 U.S. weather stations. We focus on four weather measures: maximum temperature, precipitation, and the interaction of temperature and precipitation.[20] To reduce the effect of possibly spurious outliers, we winsorize our temperature and average hourly precipitation measures at 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, as well as indicators for any snow or any rain.[21] Our nationally aggregated weather measures are simply, for each measure, the weighted average of that measure across weather stations. Weights are assigned to weather stations annually based on the percentage of total movie-theater establishments to which the weather station is matched.[22]

[20]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.

[21]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. Our weather measures are designed to capture this variation in temperature.

[22]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. 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)". 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 160 km. For the years in our sample, less than 1% of establishments fall outside a 160 km radius of any weather station.

1.3.2 Instrumenting for Viewership with Weather Shocks

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 by v_{tj}. To compute abnormal viewership during opening weekend, we first regress viewership in opening weekend, v_{t1}, on a constant and a vector of indicators for day of week, week of year, year, and holidays, which we denote F_t:[23]

$$v_{t1} = \alpha_1 + F_t' \Phi_1 + \varepsilon_{t1}. \tag{1.8}$$

We call the resulting fitted, or predicted, values $\widehat{v_{t1}}$ and define abnormal viewership opening weekend as the difference between realized and predicted viewership:

$$v\_abn_{t1} = v_{t1} - \widehat{v_{t1}}. \tag{1.9}$$

We want to instrument for this abnormal viewership opening weekend with contemporaneous weather shocks. Given the natural (and anticipated) seasonality of weather, and our desire 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

$$w_{tk} = \delta_k + F_t' \Phi_k + \varepsilon_{tk}, \tag{1.10}$$

[23]Please see Appendix A.3 for the full set of holidays.

where t again indexes the date, k indexes the particular weather measure, and the fixed effects, F_t, are as defined in Equation 1.8. We call the resulting fitted values $\widehat{w_{tk}}$ and define the weather shock w_shock_{tk} as the difference between the realized and predicted weather measure:

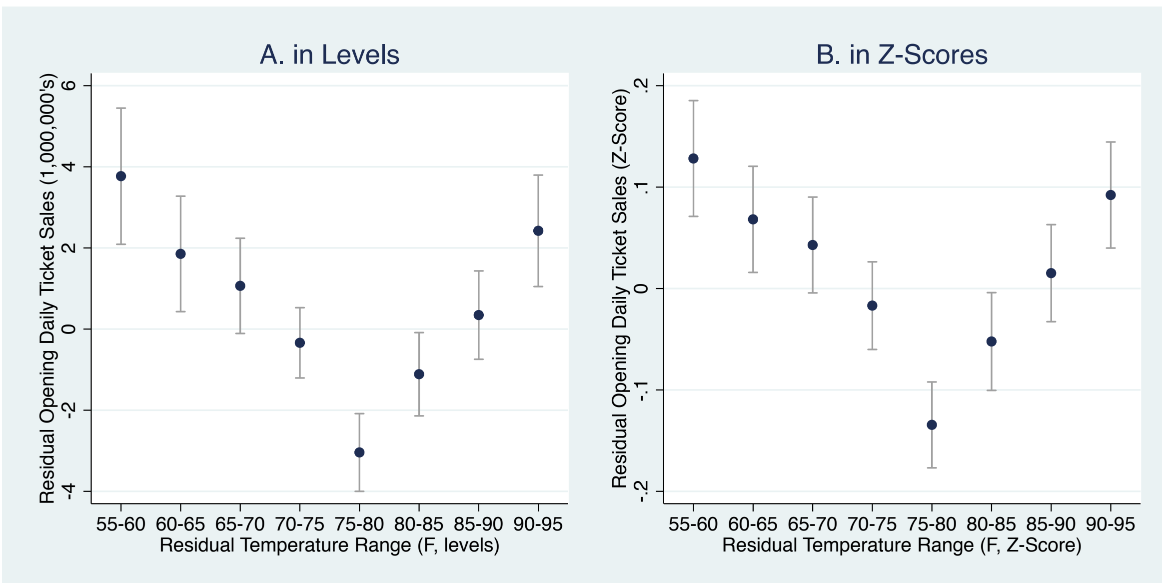
$$w\_shock_{tk} = w_{tk} - \widehat{w_{tk}}. \tag{1.11}$$

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 1.2 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 5-degree F bins. For exposition, we focus on the common summer range of 50 to 95 degrees F. Amid unexpectedly beautiful weather (especially 75 - 80 degrees F), opening weekend ticket sales are lower than would be predicted by seasonality. In the presence of 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 unexpectedly (for the time of year) in the given temperature range, and when viewership is measured in residualized ticket sales. Each plotted coefficient, then, represents estimated abnormal viewership when all (versus no) theaters are unexpectedly in that temperature range. When ten percent of theaters unexpectedly experience "ideal" temperatures (i.e., in the 75 - 80 degree F range) for example, aggregate viewership to movies opening that weekend is about 300,000 per day lower than expected (one-tenth of 3 million). With daily aggregate viewership to opening movies averaging about 2.5 million, this corresponds to more than a 10% reduction in viewership.

To facilitate comparison of effect sizes across temperature ranges, Panel B shows the corresponding results when each weather shock is normalized to be mean-zero with unit variance, and residual ticket sales are similarly normalized. This panel also serves to illustrate the sizable role a day's weather can play in that day's ticket sales. For a one

Figure 1.2: *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% confidence intervals. National weather measures are as described in the text. Panel A shows the relationship in levels: weather shocks are measured as the percentage of theaters unexpectedly in a given temperature range, and abnormal viewership is measured in number of tickets. Panel B shows the relationship when both weather shocks and residual ticket sales are measured in Z-scores. Each plotted coefficient is from a separate regression. Observations are at the date level (1,671 observations).

standard deviation increase in the percentage of movie theaters unexpectedly in the 75 to 80 degree F range, for example, we observe a 0.15 standard deviation decrease in daily viewership of opening movies. This corresponds to about 230,000 fewer tickets per day to opening movies, or about a 10% reduction on average ticket sales per day to opening movies (2.5 million).

Although weather shocks are important predictors of abnormal viewership, the large number of potential weather shock specifications makes variable selection methods appealing. We provide further detail on our 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):[24]

$$v\_abn_{t1} = \eta + \left[W_t^{LASSO}\right]' \Omega + \varepsilon_{t1}. \tag{1.12}$$

We call the instrumented abnormal viewership $\widehat{v\_abn}_{t1}$.

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:

$$v_{tj} = \alpha_j + F_t' \Phi_j + X_t' \Gamma_j + \varepsilon_{tj}. \tag{1.13}$$

The fixed effects in F_t are as defined in Equation 1.8, and X_t denotes the vector of contemporaneous (date t) weather.[25] We call the resulting fitted values $\widehat{v_{tj}}$ and define abnormal

[24]This is often referred to as post-LASSO; coefficients estimated by post-LASSO differ from those estimated via LASSO because of LASSO's shrinkage bias (i.e., the presence of the penalty in the LASSO optimization problem).

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

viewership in subsequent weekends as the difference between realized and predicted:

$$v\_abn_{tj} = v_{tj} - \widehat{v_{tj}}. \tag{1.14}$$

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

$$v\_abn_{tj} = \mu_t + \theta_j \widehat{v\_abn}_{t-7(j-1),1} + \varepsilon_{tj}. \tag{1.15}$$

The estimated momentum in the j-th week of showing is $\hat{\theta}_j$. Amid positive network externalities, we expect positive exogenous shocks to viewership opening weekend to increase viewership in later weeks.

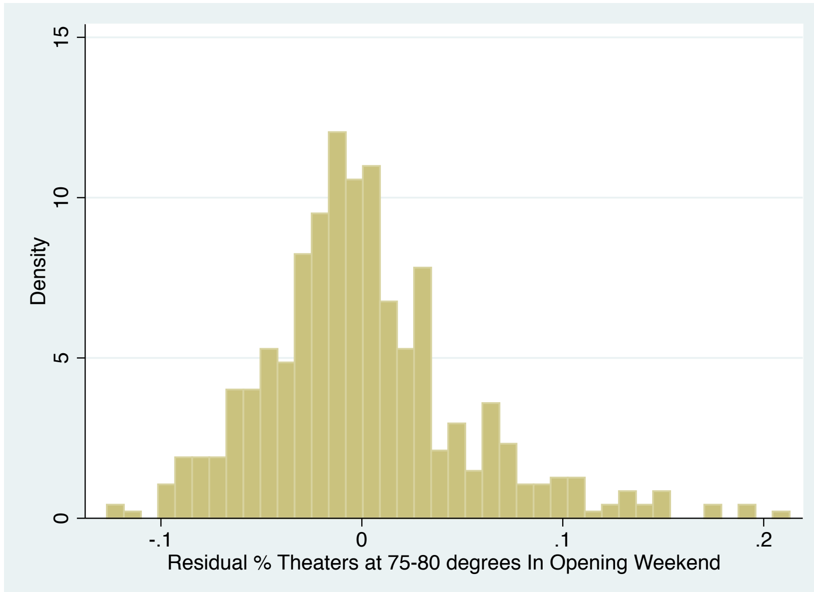
1.3.3 LASSO Instrument Selection and First Stage Results

Instrumenting with weather is non-trivial in part because the set of potential weather measures is large.[26] 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 number of potential weather instruments is effectively infinite).

Given the issues with either hand-picking a small number of instruments or naively including a large number of instruments, we implement variable selection methods. In particular, we follow Belloni *et al.* (2010), and implement Least Absolute Shrinkage and Selection Operator (LASSO) methods to estimate optimal instruments in linear IV models with many instruments. The LASSO procedure provides a principled method for instrument selection; in simulation experiments, it performs well relative to recently advocated many-instrument robustness procedures (see Belloni *et al.* (2012)). We follow Belloni *et al.* (2011)

[26] With multiple metrics of daily weather (e.g., maximum temperature, minimum temperature, precipitation, etc.) for each of thousands of weather stations, it is easy to see that the number of weather measures (or interactions and functions thereof), all of which could be used as instruments, is very large indeed.

Figure 1.3: *Histogram of the Instrument*



Notes: We plot a histogram of the abnormal percentage of movie theaters with weather in the 75 - 80 degree F range.

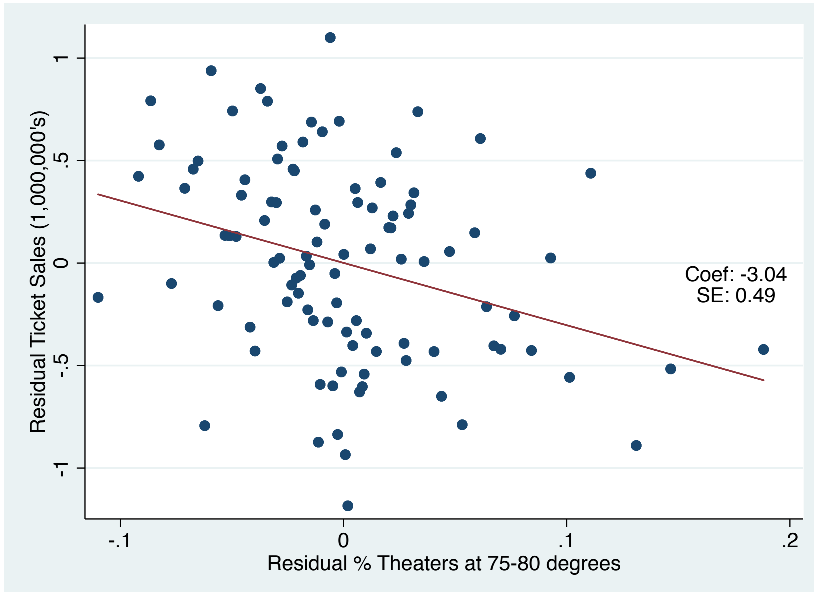
in using conventional standard errors and also in adding a constraint on the maximum number of instruments chosen.[27] Here, we present the machine-chosen instrument sets and the corresponding first stages, and refer the interested reader to Appendix Section A.4 for a brief overview of our LASSO method, which draws heavily on Chernozhukov and Hansen (2013).

We denote the final output of the LASSO methodology (i.e., the maching-chosen instrument set) by W^{LASSO}. With a single-instrument constraint, the LASSO-chosen instrument is the 75-80 degree F measure. Figure 1.3 shows a histogram of this weather measure; the mass is fairly tightly distributed between -10% and +10%. Figure 1.4 shows the corresponding first stage relationship in a binned scatterplot: the more theaters that are unexpectedly in the 75-80 degree F range, the lower is abnormal viewership.

Panel A of Table 1.1 shows the first stage results from several different LASSO-chosen instruments sets. The first row corresponds to the LASSO specification above. Here, ten

[27]Conventional standard errors are appropriate as long as the number of selected instruments is not close to the sample size – see Belloni *et al.* (2011) and Cattaneo *et al.* (2012) for more detail. We probe the instrument constraint specification choice below and show that our results are not sensitive to different instrument counts.

Figure 1.4: *First Stage Binscatter*



Notes: We plot the abnormal percentage of movie theaters with weather in the 75 - 80 degree F 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, and the corresponding robust standard error clustered by date, are included on the figure.

percent more theaters unexpectedly in the 75 - 80 degree range corresponds to about 300,000 lower daily viewership opening weekend (over 10 percent of average daily viewership for new releases). For robustness, subsequent rows show the first stage results when we instead constrain to a maximum of two or three instruments, or when we constrain to a maximum of one instrument from among a choice set of ten degree temperature bins.[28]

As of the time of writing, the LASSO-IV literature continues to evolve. While the procedure we use assumes a homoscedastic error structure, in a recent paper Belloni *et al.* (2014) develop an extension of the LASSO-IV framework that adapts the penalty parameter for the case of a clustered error structure with linear individual-specific fixed effects. In simulation experiments, they find that the new procedure performs substantially better than the former, although the differences are less distinguishable in their application (analyzing

[28]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. The full set of instruments we provide to LASSO includes these and the temperature variables for both 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.

Table 1.1: *LASSO-Chosen First Stages*

A. LASSO-Selected Instruments

Set of Potential Instruments	Count Constraint	LASSO-Chosen Instrument(s)	Coefficient (s.e.)	F-Stat on LASSO Choice
5 Degree Temp Increments	Choose 1	75-80F	-3.041*** (0.488)	38.80
	Choose 2	75-80F	-2.635*** (0.487)	25.86
		50-55F	3.419*** (0.811)	
	Choose 3	75-80F	-2.686*** (0.488)	20.95
		50-55F	3.165*** (0.837)	
		10-15F	-2.756** (1.097)	
10 Degree Temp Increments	Choose 1	70-80F	-1.253*** (0.319)	15.47

B. Hand-Selected Instruments

Instrument(s)	Coefficient (s.e.)	F-Stat
$(\text{Temperature} - 75)^2$ x (abs(Temperature - 75) ≤ 20)	0.00449*** (0.000824)	29.74
All Instruments Provided to LASSO in Base Case	-	3.804

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]), conditional on the full set of seasonal controls described in the text. From this set, LASSO is set to choose a maximum of one, two, or three instruments, respectively. In the fourth specification, a single instrument is again chosen, but the instrument choice is altered to include the analogous temperature measures in 10 degree increments instead of in 5 degree increments. Observations are at the opening weekend by date level (1,671 observations). Standard errors, clustered at the date level, are in parentheses. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

the effect of gun prevalence on crime rates). We implement their procedure in our setting and find that whether we cluster either by date or by opening weekend (as we do in computing our second stage standard errors), their new methods consistently yield the 75-80 degree F measure as the instrument of choice, i.e., the same instrument as chosen above.

We additionally note that although, for the reasons detailed above, we focus on LASSO-selected instruments throughout this paper, our results are robust to estimation using natural sets of hand-selected instruments. One hand-selected instrument we use is motivated by the intuitive observation (visualized in Figure 1.2) that the effect of weather on movie-going is roughly quadratic with a minimum in the 75-80 degree range: the instrument is simply the squared difference in temperature from 75 degrees (with the difference set to 0 at 55 and 95 degrees to avoid assigning excess weight to outliers).[29] The second set of hand-selected instruments is just the set of all potential weather instruments we supply to LASSO. Panel B of Table 1.1 presents first stage results from these two sets of hand-selected instruments; as expected, the first yields a strong first stage with an F-statistic of nearly 30, while the second is weaker with an F-statistic just below 4. As we show in the next section, however, our second stage results importantly remain largely unchanged with these hand-selected instrument sets.

1.4 Momentum from Exogenous Viewership Shocks, at the National Level

In this section, we present our estimates of momentum at the national level, along with a series of robustness checks. We also show that, consistent with the exogeneity assumption, our instruments are uncorrelated with expected demand, and discuss results which suggest analyzing potential supply-side shifts.

[29] As with all our national instruments, we aggregate across airports using the same establishment-weighting discussed earlier, and similarly condition it day of week, week of year, year, and holiday fixed effects.

Table 1.2: *Momentum from Viewership Shocks*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
IV	0.474***	0.269***	0.188***	0.112***	0.0960***	1.139***
	(0.0474)	(0.0360)	(0.0287)	(0.0203)	(0.0162)	(0.131)
OLS	0.423***	0.235***	0.140***	0.0888***	0.0630***	0.950***
	(0.0152)	(0.0111)	(0.00721)	(0.00498)	(0.00362)	(0.0388)
R-squared	0.653	0.498	0.357	0.301	0.264	0.570

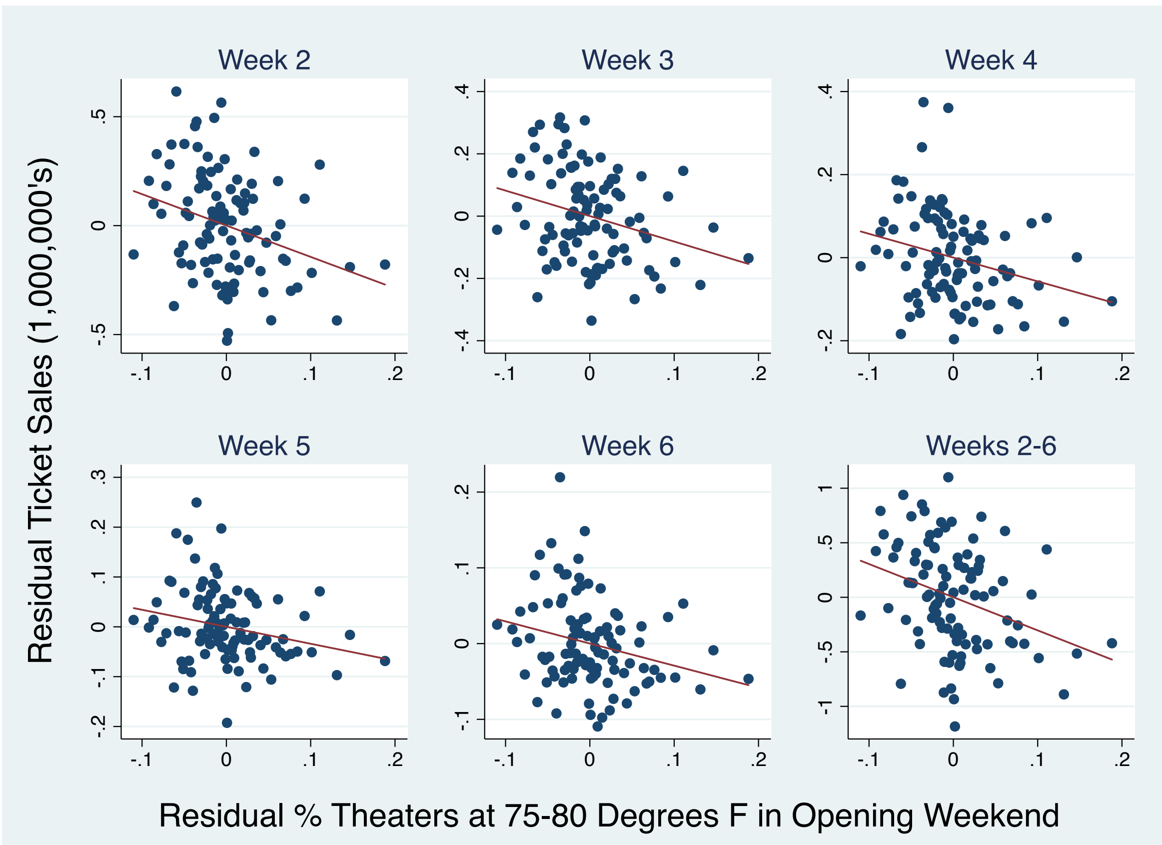
Notes: The first row reports the results of the IV regression of daily abnormal audiences in each later weekend on daily abnormal audiences opening weekend; the second row reports the corresponding OLS results. Observations are at the opening weekend by date level (1,671 observations). Standard errors, clustered at the date level, are in parentheses. */**/*** denote significance at the 10%, 5%, and 1% 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 1.1.

1.4.1 Base Case Results

The binned scatter plots in Figure 1.5 show the reduced form relationship between opening weekend weather and ticket sales in subsequent weekends. When more theaters are unexpectedly in the 75 - 80 degree F range during a movie's opening weekend, abnormal viewership of that movie is lower not only in that weekend (Figure 1.4), but also in each subsequent weekend. Implementing the second stage (Equation 1.15), we find similarly substantial momentum from exogenous shocks to opening weekend viewership. Table 1.2 presents our base case IV estimates. The first five columns report the relationship between abnormal viewership opening weekend and subsequent abnormal viewership, separately for each of weekends 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.

One-hundred additional viewers on opening weekend yields an estimated 114 additional viewers across the following five weekends. The observed momentum is largest in the weekend immediately following opening weekend, when nearly half of the total effect is

Figure 1.5: *Reduced Form Binscatters*



Notes: We plot the abnormal percentage of movie theaters with weather in the 75 - 80 degree F range against abnormal viewership in each subsequent weekend 2 through 6, and overall across those weekends. 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.

realized; an additional quarter is realized in the third weekend. Though the magnitude of the effect falls off in subsequent weeks, it remains statistically significantly above zero through each of the five subsequent weekends.

The pattern presented in these estimates is consistent with the first prediction of the model in Section 1.2: viewership in any weekend $w > 1$ is simply a fraction λ^{w-1} of viewership opening weekend. Indeed, a value of $\lambda = 0.5$ in the model predicts that each opening weekend viewer corresponds to 0.5 viewers in the second weekend, 0.25 viewers in the third, and 0.125 viewers in the fourth; by the end of the sixth weekend, this cumulatively predicts about one additional viewer for each viewer opening weekend. This pattern is remarkably similar to that observerved in the data (and reported in Table 1.2).[30]

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 of Table 1.2) closely resemble our IV estimates. If anything, our OLS estimates lie slightly below our IV estimates, but in no week is the difference between the IV and OLS estimates statistically significant. Two other factors instead likely contribute to the close alignment of our IV and OLS estimates.

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 controlled for any aspects of quality, supply, and demand that would be captured in this seasonality. Second, while our OLS estimates could be biased upward by 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.[31] 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

[30]This prediction is also consistent with the raw relationship between audience sizes and time since opening depicted in Figure 1.1 – a fact we return to in Section 1.7.

[31]Given that our LASSO selection methods yield a single instrument with an F-statistic over 38, we are not concerned about weak instruments biasing our IV estimates.

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 undecided movie-goers, suggesting that network externalities from their viewership could be stronger than from the average viewership.

1.4.2 Instruments, Clustering Level, and Other Robustness Checks

Our results are robust to different numbers of instruments chosen within LASSO and to hand-chosen instrument sets. Panel A of Appendix Table A.1 shows the corresponding second stage results with alternative LASSO specifications. Since the strength of the first stage (as measured by its F-statistic) rises when temperature is more precisely defined and when the set of chosen instruments is kept small (see Table 1.1), our base case IV estimates are derived from a single instrument chosen from among the set of five-degree temperature increments. When LASSO is instructed to choose no more than two, or even three, instruments (rather than just one), the first stage is weakened somewhat, but the second stage point estimates remain generally unchanged. They are also comparable when the potential instrument set is altered to include temperature variables in broader (ten degree) increments. Although we focus on LASSO-selected instruments throughout this paper, Panel B of Appendix Table A.1 shows that our results are robust also to using the natural sets of hand-selected instruments discussed in Section 1.3.3. Both sets of hand-selected instruments yield second stage estimates in line with our base case results.

Our results are also robust to clustering at a higher level, to larger units of observation, and to the inclusion of second stage contemporaneous weather controls in the first stage. In Panel A of Appendix Table A.2, observations are clustered at the weekend level; in Panel B, observations are defined at the opening weekend by weekend level (rather than opening weekend by day level); and in Panel C, controls for the second stage's contemporaneous weather are included in the first stage.[32] In each case, the estimated coefficients closely

[32]For example, when the second stage outcome variable is ticket sales in Week 2, Week 2 contemporaneous weather controls (as defined in Footnote 25) are included also as controls in the first stage.

resemble our base case results, and in each robustness check, estimated momentum from exogenous viewership shocks remains highly significant in each week.

1.4.3 Evidence on Exogeneity

Recall that we seek to quantify 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 we are identifying off of 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.

To test the exogeneity of our weather shocks, we ask whether they are correlated with expected demand. In Table 1.3, we follow Moretti (2011) in proxying for expected demand with the number of screens on which the movie opened and control for this measure of expected demand in our specification from before.[33] 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 insignificantly); the average change of the point estimates is on the order of 2% and each week's estimated momentum remains large and highly significant. In the third row, we define the outcome variable as abnormal viewership *per opening screen*. For comparison with 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

[33]First stage results for this specification, as well as for other specifications that extend our base case, are reported in Appendix Table A.3.

Table 1.3: *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***	0.474***	0.269***	0.188***	0.112***	0.0960***	1.139***
	(0.000)	(0.0474)	(0.0360)	(0.0287)	(0.0203)	(0.0162)	(0.131)
Tickets, Controlling for Opening Theaters (2)	1.000***	0.457***	0.257***	0.186***	0.110***	0.0977***	1.107***
	(0.000)	(0.0516)	(0.0398)	(0.0320)	(0.0227)	(0.0182)	(0.145)
Tickets per Opening Theater	1.019***	0.353***	0.248***	0.173***	0.120***	0.115***	1.010***
	(0.240)	(0.115)	(0.0775)	(0.0535)	(0.0408)	(0.0296)	(0.289)
Standardized Tickets per Opening Theater (3)	1.000***	0.346***	0.243***	0.170***	0.118***	0.113***	0.991***
	(0.238)	(0.114)	(0.076)	(0.052)	(0.040)	(0.029)	(0.286)
Differences:							
(1) - (2)	--	0.017	0.012	0.002	0.002	-0.001	0.032
		(0.0701)	(0.0537)	(0.043)	(0.0305)	(0.0244)	(0.1954)
(1) - (3)	--	0.121	0.021	0.015	-0.008	-0.019	0.129
		(0.1233)	(0.0841)	(0.0591)	(0.0447)	(0.0331)	(0.3146)

Notes: This table presents results from three different IV specifications. The first is our base case from Table 1.2; in the second row, controls for number of opening theaters, a proxy for expected demand, are added in both the first and second stages; in the third row, the outcome variable is defined as abnormal viewership per opening screen. Observations are at the opening weekend by date level (1,671 observations). Standard errors, clustered at the date level, are in parentheses. */**/*** denote significance at the 10%, 5%, and 1% 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 rows of Table 1.1 and of Appendix Table A.3.

large and statistically significant throughout. Our second stage indeed appears to be picking up viewership shocks that are orthogonal to expected demand.

In Appendix Section A.6, we analyze an additional explanation for the observed momentum: supply shifts. For example, it could be the case that opening weekend viewership shocks lead theater owners to increase screenings, which decreases the effective cost of attendance and (endogenously) increases viewership in subsequent weeks. Thus, in the Appendix Section A.6 we present a brief overview of the supply side of the market and show that our estimated momentum is not driven by supplier responses. Given that prices are almost always fixed for in-theater movies, we conclude that demand shifts must be driving the observed quantity effects.

1.5 Local vs. National Momentum

We have shown there is substantial momentum in movie ticket sales measured at the national level. In this section, we examine whether the observed momentum in movie-going is driven predominately at the local level (e.g., through conversations among friends) or at the national level (e.g., by way of national media coverage of box office sales). Since local movie ticket sales data are typically unavailable to researchers, we proxy for local viewership with MSA-level Google search data.[34]

1.5.1 Proxying for Local Movie-Going with Google Search Data

Our search data come from Google Trends, which provides a measure of search activity on Google pertaining to specific keywords and topics.[35] While search volumes do not perfectly capture movie consumption, they should provide a strong proxy. As we describe below, we find national searches to be highly predictive of national sales. This is intuitive, as many people search online for showtimes before attending a movie; Stephens-Davidowitz *et al.* (2014) similarly use local searches for movies as a proxy for local movie-going to study the impacts of Super Bowl advertising on movie consumption.

We proxy for local movie-going with Google Trends search data at the MSA by day by topic level (the most granular level at which it is made publicly available) and employ Google's topic classification engine to classify searches as pertaining to particular movies. Appendix A.7 details our data collection methods and variable definitions. In brief, our MSA by day by movie level search measure corresponds to the Z-score of search volume within that MSA. We focus on the five MSAs with the most complete Google data: New York, Los Angeles, Chicago, San Francisco, and Washington D.C.[36]

[34]Google search data has been used by researchers to proxy for behaviors ranging from turning up at the polls to maltreating children, and even sentiments such as racial animus (see, e.g., Stephens-Davidowitz (2013a,b, 2014)).

[35]Google is a dominant player in the US search market, with approximately 68% market share at present.

[36]Note that, since Google provides only unit-less search figures, we are unable to directly compare search volumes *across* MSAs and thus rely on MSA-level z-score normalizations.

1.5.2 Estimating Momentum Locally

Our methodology for estimating momentum locally is simply the local analog of our main methodology outlined in Section 1.4.1. Since our data now vary at the MSA level, we condition our MSA-level weather and MSA-level search data on MSA-level fixed effects for day of week, week of year, year, and holiday. (This is equivalent to implementing the conditioning methodology in our main specification, but separately for each MSA.) We use the same LASSO procedure for selecting weather instruments (see Section 1.3.3 for details).

One advantage of estimating momentum using local-level search data is that we can more directly control for film quality and latent demand by controlling for both (1) local abnormal searches in the weekend *before* opening weekend and (2) contemporaneous abnormal searches in distant MSAs.[37] The first of these controls is designed to capture local demand in advance of movie opening; the second is designed to capture the effects of any national-level shocks that might impact demand on opening weekend (e.g., last minute advertising pushes, breaking national news events). Another advantage of using search data is that it facilitates analysis of how local viewership in weekends subsequent to opening are affected by each of local viewership and national viewership on opening weekend (proxied for with local and national searches, respectively, that weekend).[38]

Before turning to our results, we note that we require two additional assumptions in order to compare the results presented here with the national results presented previously. First, we assume that local searches are a reasonable proxy for local movie going. Without local ticket sales data, we are not able to test this assumption directly. However, at the national level the correlation between abnormal national searches and abnormal national movie-going is quite strong (0.74).

[37]We define distant MSAs as MSAs more than 1,000 km away.

[38]This analysis requires instruments for local opening searches and for national opening searches. In the absence of a LASSO-IV procedure which explicitly allows for a design with multiple endogenous variables, we LASSO-select the instruments for each endogenous regressor separately, and then use the union of the LASSO-chosen sets in the actual estimation. As before, we cluster our standard errors at the date level. Note that we would ideally analyze the effect of opening weekend viewership in *all other* MSAs, as opposed to simply nationally, but given the sparsity of our MSA-level data (i.e., searches from four other MSAs are not representative of the entire nation), we focus on national searches.

Second, we assume that the relationship between searches and sales is not confounded by the weather, i.e., that weather shocks do not change how search behavior and ticket sales relate. We probe this assumption in Appendix Figure A.1, which shows how the residual from a regression of abnormal national ticket sales on abnormal national searches is related to national weather shocks in five-degree increments from 50 - 95 F. All estimated coefficients are relatively close to 0, and none is statistically significant, suggesting that the relationship between searches and ticket sales is quite stable relative to changes in the weather.

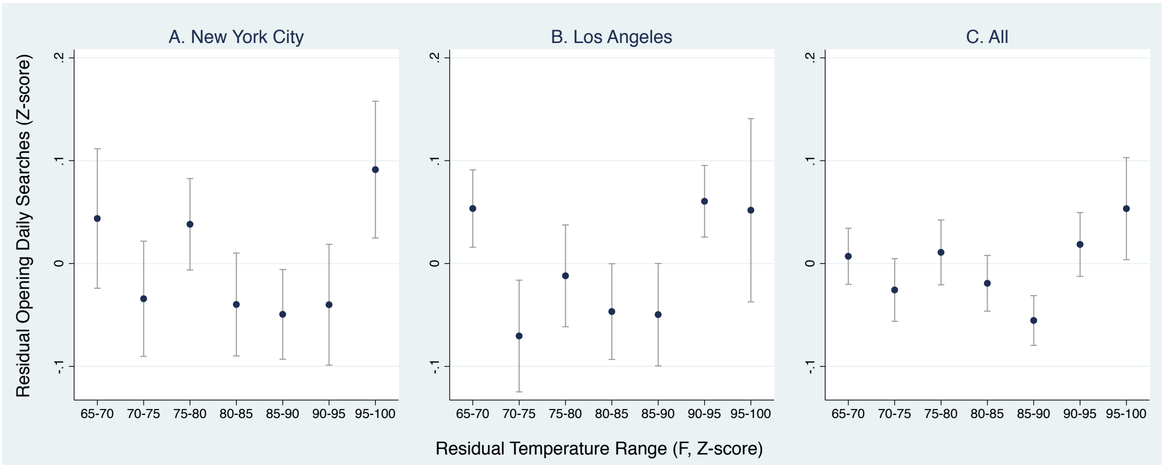
1.5.3 Local Momentum Results

The directionality of the local-level first and second stage relationships is very similar to that at the national level. Figure 1.6 (the local analog of Figure 1.2) previews a simplified version of the opening weekend relationship between local weather shocks and abnormal local searches (our proxy for abnormal local movie going), controlling for abnormal local searches in the weekend prior and for abnormal searches in MSAs more than 1,000 km away. Each coefficient comes from a separate regression of abnormal viewership on contemporaneous weather shocks in 5-degree F bins. Panel A corresponds to New York, Panel B to Los Angeles, and Panel C to all included MSAs. As at the national level, when the local weather is unexpectedly beautiful, opening weekend ticket sales tend to be lower than would be predicted by seasonality; and when local weather is unexpectedly cool or warm, local audiences tend to be higher.

With our MSA-level data, the LASSO-chosen instrument is a residualized indicator for the 85 to 90 degree F temperature range.[39] Figure 1.7 shows the corresponding first stage relationship in a binned scatterplot: when unexpectedly in the 85 to 90 degree F temperature range (“good" weather), abnormal viewership locally is lower. The binned scatter plots in Figure 1.8 show the corresponding reduced form relationship between opening weekend weather and ticket sales in subsequent weekends: when unexpectedly in the 85 - 90 degree F

[39]This is different from our main LASSO-chosen national weather instrument, but of course these five MSAs do not perfectly capture the entire country.

Figure 1.6: *The Effect of Local Weather Shocks on Local Viewership*



Notes: The local analog of Figure 1.2, each panel plots the coefficient of the regression of abnormal local Google searches on each listed weather shock, along with the corresponding 95% confidence intervals, controlling for local searches in the MSAs more than 1,000 km away. Weather shocks are measured as the z-score of the residual of the indicator for the MSA in each temperature range, and abnormal local searches are as described in the text. Panel A corresponds to just New York, Panel B to just Los Angeles, and Panel C to all five MSAs in our sample. Observations are at the movie by date by MSA level (576 observations in Panel A, 480 in Panel B, 2064 in Panel C).

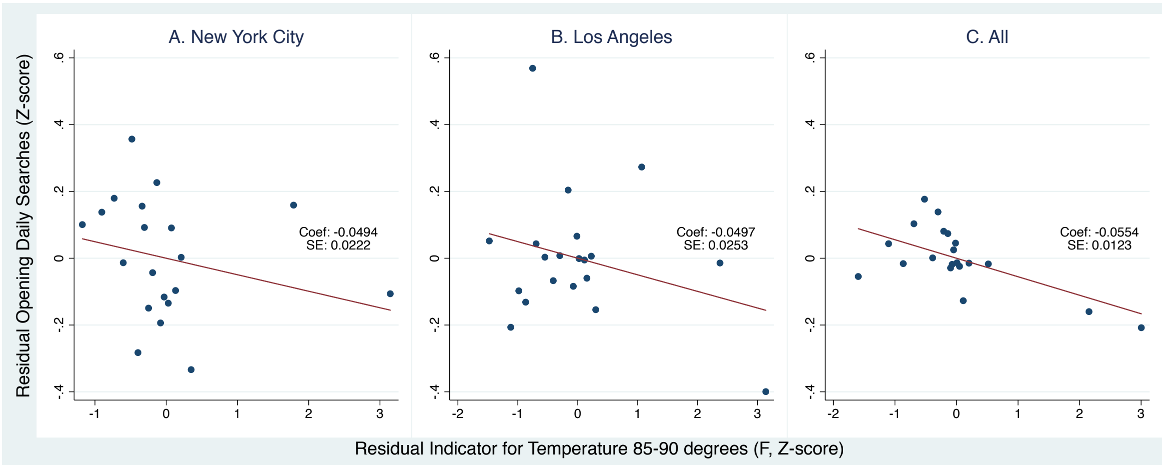
range opening weekend, abnormal local viewership is lower in each subsequent weekend.[40]

Table 1.4 reports the magnitudes of network externalities observed locally from instrumented movie-going shocks at each of the local and national levels. Throughout, the dependent variable is abnormal local searches in week w, but the independent variables vary across the panels. The first row in Panel A is the local analog of our base case estimates from Table 1.2; this row reports how local searches in opening weekend, instrumented for with local weather shocks, impact local searches in subsequent weekends.[41] The results are highly comparable to our overall estimates (Table 1.2), suggesting the momentum observed nationally may be predominantly local in nature. In the second and third rows of Panel A,

[40] As a robustness check, we also analyze whether local searches are impacted by weather shocks in MSAs at least 1000 km away. Since weather shocks at that distance are negatively correlated (e.g., a day that falls between 90 and 95F in New York City is associated with a cooler day in Los Angeles), we condition on local weather controls (10-degree temperature dummies, quarter inch precipitation dummies, and snow and rain dummies, as in our national regressions). The results, presented in Appendix Figure A.2, show a slight inverted-U shape – the inverse of our estimated relationship between local weather and local searches (Figure 1.6). We infer that our local weather controls may not be picking up all effects of local weather, but that the relationship is the inverse of the local relationship reduces our concern about the potential for cross-MSA contamination effects.

[41] See Appendix Table A.4 for the corresponding first stage results.

Figure 1.7: *Local First Stage Binscatters*



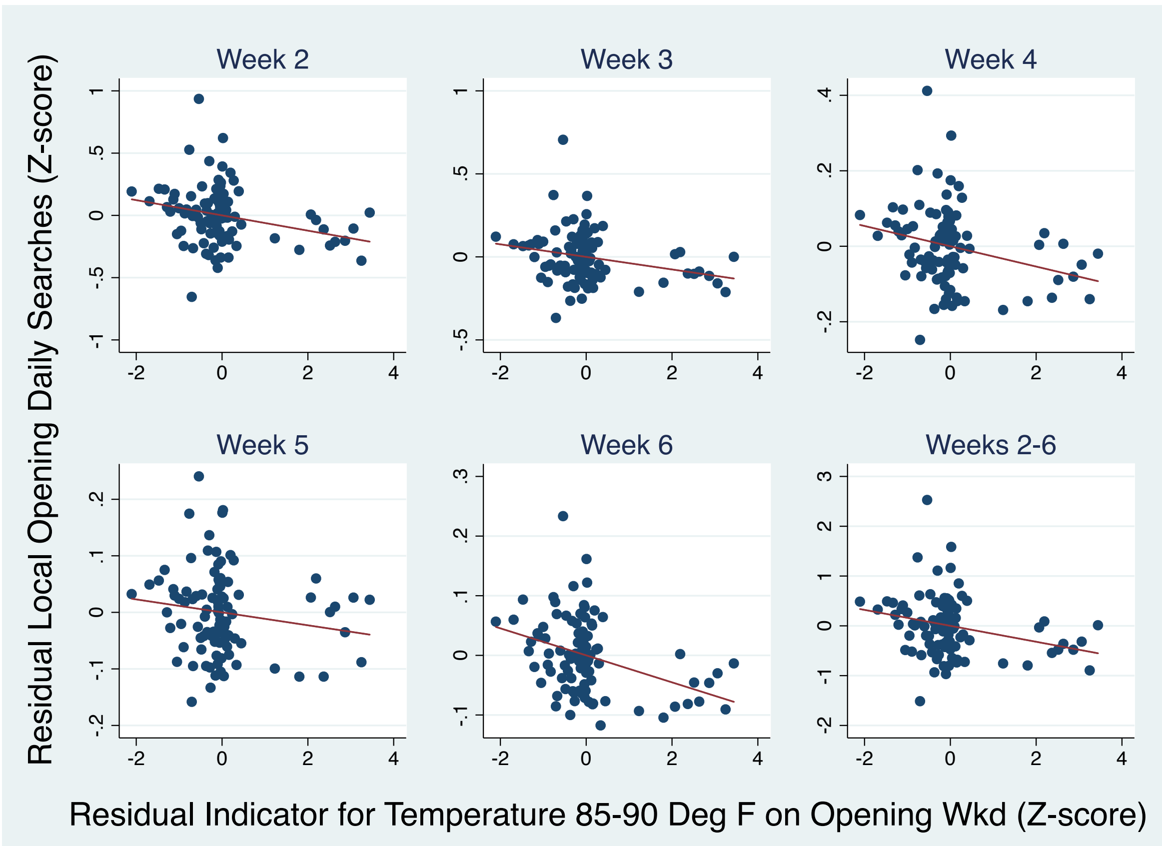
Notes: The local analog of Figure 1.4, each panel plots the abnormal local searches against the residual of the indicator for whether the MSA is in the 85 - 90 degree F range, controlling for abnormal searches in the MSAs more than 1,000 km away. For exposition, the weather shock measure is grouped into 100 equal-sized bins; each point corresponds to the mean local weather shock and local abnormal viewership within a bin. Panel A corresponds to just New York, Panel B to just Los Angeles, and Panel C to all five MSAs in our sample. The slope of the line of best fit, and the corresponding robust standard error clustered by date, are included in each Panel.

we add controls incrementally for (1) local searches in the weekend prior to opening and (2) contemporaneous (same weekend) searches in distant MSAs. These additional demand controls change the magnitude of the point estimates only slightly (and insignificantly) relative to the results presented in the first row.[42]

Panel B shows how both local and national opening weekend searches, instrumented for with local and national weather shocks, impact local searches in future weekends. Controlling for local movie-going shocks, we observe no positive effect of national movie-going shocks on local movie-going in subsequent weekends. That is, we find no evidence that the momentum is driven at the national level, for example via publicized national box office results. Instead, the lion's share of the momentum arises at the local level. These results hold also when controlling for local demand prior to release (Panel C). We conclude that our estimated momentum is driven by a social multiplier arising predominately at the

[42]If consumers are forward looking and weather forecasts are sufficiently accurate, searches in the weekend prior to opening might actually constitute a direct measure of opening weekend demand that takes weather shocks into account. However, controlling instead for searches two weeks prior does not change the results; for brevity those results are excluded here, but are available upon request.

Figure 1.8: *Local Reduced Form Binscatters*



Notes: The local analog of Figure 1.5, the panels plot the residual of abnormal local searches in each subsequent weekend 2 through 6, and overall across those weekends, against a residualized indicator for the MSA in the 85 - 90 degree F range, controlling for local searches in the MSAs more than 1,000 km away. 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.

Table 1.4: *Local Momentum from Network Externalities*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
	A. Effect of Local Opening Searches on Subsequent Local Searches					
Local Searches	0.518***	0.320***	0.227***	0.0969***	0.192***	1.354***
	(0.0801)	(0.0639)	(0.0459)	(0.0296)	(0.0348)	(0.215)
Local Searches, Controlling for Local Searches in Weekend Prior to Opening	0.551***	0.349***	0.249***	0.0975***	0.216***	1.463***
	(0.0882)	(0.0659)	(0.0494)	(0.0361)	(0.0405)	(0.231)
Local Searches, Controlling for Local Searches in Weekend Prior to Opening and Local Searches in Distant MSAs	0.736***	0.485***	0.349***	0.115*	0.330***	2.015***
	(0.166)	(0.122)	(0.0927)	(0.0626)	(0.0782)	(0.450)
	B. Effect of Local and National Opening Searches on Subsequent Local Searches					
Local Searches	0.568***	0.367***	0.263***	0.102**	0.229***	1.529***
	(0.114)	(0.0896)	(0.0680)	(0.0414)	(0.0527)	(0.322)
National Searches	-0.101	-0.0952*	-0.0728*	-0.00949	-0.0747**	-0.353*
	(0.0653)	(0.0486)	(0.0389)	(0.0232)	(0.0308)	(0.182)
	C. Effect of Local and National Opening Searches on Subsequent Local Searches, Controlling for Local Searches in Weekend Prior to Opening Weekend					
Local Searches	0.576***	0.373***	0.267***	0.100**	0.234***	1.550***
	(0.115)	(0.0905)	(0.0690)	(0.0429)	(0.0540)	(0.326)
National Searches	-0.0917	-0.0890*	-0.0684*	-0.0109	-0.0683**	-0.328*
	(0.0654)	(0.0471)	(0.0387)	(0.0231)	(0.0311)	(0.182)

Notes: This table analyzes how local searches in subsequent weekends are affected by local and national searches on opening weekend. Panel A replicates the IV results from Table 1.2 using local Google searches as a proxy for ticket sales. Panel B does the same, but investigates the relationship between national searches on opening weekend and local searches in subsequent weekends. Panel C uses local searches as the dependent variable but includes and instruments for both local searches and national searches on opening weekend, while Panel D does the same but also adds a control for abnormal searches in the weekend prior to opening weekend. The first stage results are included in Appendix Table A.4. Observations are at the movie by date by MSA level (2,064 observations). Standard errors, clustered at the date level, are in parentheses. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. The corresponding OLS estimates are presented in Appendix Table A.7.

local level.

1.6 A Role for 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 shown that the estimated momentum is not driven by supply shifts and that it is largely local in nature. The particular nature of the local demand shifts, however, remains important for interpretation. In this section, we look for evidence on whether a straightforward social or observational learning story could explain our results.

At the most basic level, a model in which momentum is generated by some type of learning differs from a model in which momentum is generated by network externalities via the role of private information. Momentum originating from network externalities need not require consumers to have private information. In our model, for example, all consumers hold the same (unchanging) views about quality, and there is nothing for consumers to learn. All information is public, and momentum is generated by the combination of heterogeneity in preferences and network externalities: some consumers find it worthwhile to view a movie irrespective of its viewership, and these consumers view the movie as soon as it opens; meantime, others view the movie only when prior viewership is sufficiently high that they find it worthwhile to view the movie themselves.

Much of the existing theoretical research on crowd-following, in contrast, focuses on the role of information and, in particular, on models of social learning and observational learning.[43] The precise 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 takes into account the observed choices and/or reports of others in making her own decision. Because the existence of private information that can somehow be passed from

[43] See, for example, Banerjee (1992); Bikhchandani *et al.* (1992, 1998); Ellison and Fudenberg (1995); McFadden and Train (1996); Çelen and Kariv (2004).

one consumer to another creates an opportunity for learning, our strategy for distinguishing momentum due to network externalities from momentum due to learning is to focus on the role of information. Before proceeding, we note that the learning literature distinguishes between models in which consumers simply observe whether or not others are consuming the good and take that as a signal of quality – typically called "observational learning" – from models in which payoffs are shared directly – typically called "social" learning. We focus first on tests for social learning before turning to tests for observational learning.

1.6.1 A Simple Test for Social Learning amid Weather Shocks

By instrumenting with shocks that are orthogonal to movie quality, we have isolated shocks to opening weekend viewership that are plausibly independent of quality. These viewership shocks should, in and of themselves, thus provide no quality signal, and need not induce quality updating among individuals considering attending the movie in later weeks. Nevertheless, we might wonder whether larger early viewership boosts later sales in part because it influences the availability of information about the movie.

Disentangling a network externalities story from a social learning story is not easy. For one, a sufficiently complex model of movie-going in the presence of network externalities could predict behavior patterns which are similar to those predicted by a learning story.[44] Moreover, as Young (2009) notes, the theoretical literature on product adoption through learning is characterized by substantial diversity in behavioral and informational assumptions; taken together, learning models can generate a large and varied set of empirical predictions making it hard to rule out a learning story fully.

To make progress, we focus first on an intuitive prediction of behavior in the presence of social learning (and address a separate prediction of behavior in the presence of observational learning in Section 1.6.2): in the presence of social learning, shocks to viewership should induce stronger momentum for higher quality movies than lower quality movies.

[44] Indeed, Moretti (2011) writes that the four main predictions of his model of movie-going in the presence of social learning could equally be predicted by network externalities in consumption.

This learning prediction arises out of a canonical normal-normal model of social learning, such as that elegantly exposited by Young (2009), in which agents are risk neutral and hold private (conjugate) priors about the normally-distributed value of viewing. In this model, the net utility of viewership has both a private and public component, and social learning occurs over time as agents view films and share their payoffs with their peers. Since weather-induced shocks to viewership on opening weekend are orthogonal to film characteristics, such shocks simply change the amount of information available after opening weekend; below, we explain why social learning predicts that shocks like these differentially affect films according to their quality.

As an illustrative example, suppose there are two movies which are identical aside from their expected (ex-ante) and realized (ex-post) mean utilities of viewing: i.e., viewers correctly expect that one film is of higher quality than the other. On opening weekend, potential viewers have only their noisy priors with which to make attendance decisions; in subsequent weekends, new potential viewers form expectations by updating their priors using the payoffs communicated to them by those who have already viewed. In early weekends, the low-quality film benefits from uncertainty about its quality as some viewers incorrectly believe it is worth attending, while the opposite is true for the high-quality film (which is initially hurt by the fact that some viewers believe it is worse than it is). Social learning reduces uncertainty over time, leading the sales paths of the two films to diverge. A shock to opening weekend viewership simply speeds the process along; in the limit, the shock leads demand for the high quality film to surge because potential viewers learn it is excellent, while demand for the low quality film plummets because no viewers remain who believe it is worth the cost of attendance.

That quality influences adoption in a learning framework has been used in the literature to distinguish social learning stories from other adoption models. For example, Young (2009)'s analysis focuses on the features that distinguish social learning stories from social influence and contagion models, and he shows that adoption curves in learning frameworks are unusually reliant on the payoffs of adoption.[45] This is a useful distinction because, as we demonstrated in Section 1.2, momentum generated from network externalities can be independent of film quality.

As a test for social learning, we ask whether momentum is stronger for higher quality films than for lower quality films. We proxy for realized 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."[46] Our motivation for using expert ratings is that they are a plausibly impartial measure of quality (as we show below, they do appear to be unrelated to weather shocks). To avoid having to make strong functional form assumptions about how quality might interact with momentum, we cut movies into quality terciles based on expert ratings.[47] That this dimension captures an important feature of demand is demonstrated by the substantial difference in ticket sales by tercile: on average, top tercile movies sell nearly 8 million tickets over the course of the first six weeks, while bottom tercile movies on average sell just 4.6 million tickets over the same period.

We find no evidence that our estimated momentum is influenced by realized film quality. Panel A of Table 1.5 shows estimated momentum separately by high quality (top-tercile) and low quality (bottom-tercile) movies. Relative to movies with low expert ratings, movies with high expert ratings experience about the same levels of momentum in early weeks and only slightly (insignificantly) more momentum in later weeks.[48]

---

[45] We note that Young (2009) actually shows in a continuous-time framework that social learning leads the adoption paths of higher payoff goods first-order stochastically dominate those of lower payoff goods.

[46] This is assuredly not the only dimension of quality that might matter, but it is readily quantified. 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." All movies in our sample have had at least a full year to accrue votes, and have been rated on average by 483 of these Top-1000 voters.

[47] Movies with an average Top-1000 voter rating of 6.3 or above fall in the top tercile, while movies with a rating of 5.6 or below fall in the bottom tercile.

[48] Appendix Table A.5 shows the corresponding OLS results.

Table 1.5: *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	0.505***	0.295***	0.201***	0.136***	0.0946***	1.232***
(obs. 705)	(0.0609)	(0.0485)	(0.0351)	(0.0250)	(0.0192)	(0.176)
Low-Rated	0.538***	0.288***	0.139***	0.0766***	0.0518***	1.093***
(obs. 825)	(0.0871)	(0.0617)	(0.0431)	(0.0265)	(0.0169)	(0.211)
Difference:						
(High - Low)	-0.033	0.007	0.062	0.060*	0.043*	0.139
	(0.105)	(0.077)	(0.055)	(0.036)	(0.024)	(0.274)
			B. By Information about Movie Quality			
High Budget	0.435***	0.224***	0.169***	0.0971***	0.0667***	0.991***
(obs. 744)	(0.0684)	(0.0451)	(0.0358)	(0.0272)	(0.0185)	(0.183)
Low Budget	0.449***	0.298***	0.136***	0.0736***	0.0368**	0.994***
(obs. 705)	(0.0602)	(0.0449)	(0.0305)	(0.0230)	(0.0152)	(0.156)
Difference:						
(High - Low)	-0.014	-0.074	0.033	0.024	0.030	-0.003
	(0.090)	(0.062)	(0.046)	(0.035)	(0.023)	(0.240)

Notes: Panel A replicates the IV results from Table 1.2 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 A.3. The final column reports the differences in the point estimates. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. The corresponding OLS estimates are presented in Appendix Table A.5.

Finally, consistent with the assumption of the exogeneity of these expert reviews, we also find no evidence that initial, weather-induced shocks to viewership are related to the overall number of votes cast by Top-1000 voters, or to the likelihood of being in the top versus bottom tercile. Appendix Table A.6 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. Weather-induced shocks to viewership do not significantly impact the number of residual votes cast by expert reviewers, nor are the level of expert ratings broadly impacted.

We note here that Moretti (2011) analyzes social learning in movie-going and attempts to rule out the possibility that there is momentum from network externalities by instrumenting for opening weekend viewership with weather. However, in Appendix A.5, we show that Moretti (2011)'s rejection of network externalities is not robust to the inclusion of seasonal controls in his specification.

1.6.2 A Simple Test for Observational Learning amid Weather Shocks

In a standard observational learning model, for example that set forth in Banerjee (1992), an individual holding private beliefs about the utility of viewing updates her priors by looking at the previous choices made by others. Thus, in our context, observational learning predicts that the viewership decisions of others inform new potential viewers about where their priors are situated relative to those of others.

Observational learning could take place in a variety of ways. Perhaps the most plausible observational learning mechanism in movie-going is one in which potential viewers interpret past box-office success as a signal of quality. However, the results presented in Section 1.5 suggest this mechanism is unlikely to be driving our results given that our estimated momentum is largely local in nature and box office sales are predominately reported at the national level. Alternatively, observational learning could occur locally, for example if viewers communicate their attendance decisions to their local peers (e.g., in person or through social media).

As a simple test for whether observational learning drives our results, we ask whether our estimated momentum is affected by the precision of ex-ante information about quality. This a key parameter in observational learning models because the precision of priors determines the extent to which prior viewership informs new potential viewers' beliefs about quality (see, e.g., Bikhchandani *et al.* (1992) for a formalization). For exposition, consider the adoption paths of two movies which have identical expected and realized qualities, but where consumers have private beliefs which are precise for one movie and diffuse for the other. A shock to opening weekend viewership is more meaningful for the film for which priors are diffuse because for this film each incremental ticket sale provides a strong signal of quality.[49] In the limit, ticket sales for a film for which priors are perfectly precise are unaffected by an opening weekend viewership shock.

We proxy for the precision of information available ex-ante about a movie with that movie's production budget. Since advertising budgets are generally set as a fixed percentage of production budgets, production budgets plausibly capture the extent to which moviegoers have precise prior knowledge about the film's quality.[50] Indeed, we find evidence that production budgets are negatively correlated with uncertainty. In an adaptation of the methodology in Moretti (2011), we define a film's "surprise" as that film's total ticket sales in weekends subsequent to opening not predicted by its opening weekend theaters (i.e., the residual from a regression of the log of total sales in weekends subsequent to opening on the log of opening weekend theaters). This measure reflects the portion of ticket sales which was not predicted by what was known about a film prior to its release (where the ex-ante prediction is captured by the number of theaters who, operating as rational profit-maximizers, chose to screen the film in its opening weekend). So as to focus only on those ticket sales that occur after a film's quality has been publicly realized, we exclude opening weekend sales from the specification.[51]

Consistent with a negative relationship between production budgets and the precision of information about a movie available ex-ante, we find more variance in surprise for films with lower production budgets than for films with higher production budgets. Splitting production budgets by tercile, the respective standard deviations of surprise for films in the

[49] In this simple expository framework, we are assuming viewers are unable to distinguish between ticket sales which are driven by the weather and ticket sales driven by expectations of quality. This seems reasonable since, even though they are public information, the weather shocks we identify off of are not extreme events; rather, they are marginal changes in temperature.

[50] Although production budgets do not, according to IMDB, usually include advertising costs, Einav (2007) notes that advertising budgets are generally set as a fixed percentage of production budgets. From among the 1,381 movies in our main sample, we have production budgets from IMDB for 88%.

[51] Our main results always focus on subsequent weekend sales relative to first weekend sales; subsequent weekend sales are those that, at least to some degree, incorporate realized quality and not merely expected quality. However, we note the results are similar if opening weekend sales are included.

bottom (below \$29M), middle (between \$29M and \$48M), and top (in excess of \$48M) terciles are 0.92, 0.83, and 0.75, respectively. Relatedly, Panel A of Appendix Figure A.3 shows surprise plotted against production budget, and Panel B shows separate kernel densities of surprise for movies in the top and bottom terciles of production budget. As expected, the distribution for films in the bottom tercile has less mass in the middle and more in the tails. A two-sample Kolmogorov-Smirnov test shows the distributions of surprise for top and bottom terciles are statistically significantly different at the 10% level (p-value=0.091) and Kolmogorov-Smirnov tests are known to be insensitive to differences in the tails (Conover (1999)). We infer that production budgets do serve as a meaningful proxy for uncertainty about quality prior to a film's release.

Panel B of Table 1.5 analyzes the role of uncertainty in our estimated momentum. Here we report momentum estimates separately for movies that fall in the top tercile and the bottom tercile by production budget. The final row reports the differences between the point estimates. Overall, we find no evidence that movies are differentially impacted according to the level of ex-ante uncertainty about quality. Although low-budget films exhibit slightly higher momentum in Week 2, they actually have slightly *lower* momentum in all other weeks, and in no week is the difference in estimated momentum between high- and low-budget films statistically significant.[52]

While we are unable to completely rule out learning stories – other learning models are certainly possible, and learning stories may of course co-exist with network externalities (e.g., see Choi (1997)) – given the close alignment between the estimates presented in this section split across both quality and the precision of information about quality, we infer that it is unlikely that simple social or observational learning frameworks could explain the bulk of our estimated momentum.[53]

[52]Appendix Table A.5 shows the corresponding OLS results.

[53]More complex learning models are possible, e.g. models in which consumers are risk averse. However, risk aversion should only amplify the effect that the ex-ante variance of film quality has on momentum because risk aversion heightens the cost of going to a bad movie. Given the close alignment of our results in which we categorize films according to the level of prior uncertainty about quality using production budgets, it seems unlikely that such a mechanism is driving our results.

1.7 Economic Implications

We have demonstrated that the observed change in quantity is local in origin, that it is a demand-side phenomenon, and that the momentum we estimate appears to be largely unrelated to learning. Taken together, these results suggest that our empirical strategy is predominately capturing momentum from network externalities. In this section, we explore what decision these network externalities drive, and for whom network externalities matter most. We then discuss the broader economic implications of the observed momentum from network externalities.

1.7.1 Substitution

Where are these viewers, induced by network externalities, coming from? In Table 1.6, we analyze to what extent the marginal viewers are substituting across movies versus across activities. In the first row, the endogenous regressor is abnormal viewership of new releases in week w and 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.[54] Each reported coefficient is from a separate regression. The first column simply reproduces our base case results. 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 substitution across movies: for 100 more viewers of movies opening last weekend, we see 47 more viewers of those movies this weekend and 35 fewer viewers of new movies just opening. This is consistent with network externalities increasing the utility of seeing movies that did particularly well last weekend, which in turns leads to an increase in overall demand for these films and a corresponding reduction in demand for this weekend's new movies (which experienced no such shock). The positive yet statistically imprecise point estimate in the

[54] The first stage is the same as in our base case (see Table 1.1).

Table 1.6: *Substitution across Movies and Activities*

Endogenous regressor: Audiences this week	Outome variable: Audiences next 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	0.474***	0.417***	-0.352*	0.0482
	(0.0474)	(0.0820)	(0.197)	(0.182)
Movies in 1st to 5th Week		0.460***	-0.388*	0.0531
		(0.0680)	(0.213)	(0.201)

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 $w+1$, and (4) played in week $w+1$, respectively. The corresponding first stage is in the first row of Table 1.1. 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 stages are reported in Appendix Table A.3. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

final column leaves us unable to identify conclusively whether network externalities produce momentum in aggregate movie-going, or whether the effect is entirely infra-marginal.

Since weather shocks may well engender momentum for *any* movie showing both this weekend and the next (not just movies that opened this weekend), the second row of Table 1.6 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 $w+1$.[55] With this specification, we again find strong momentum from network externalities and some evidence of substitution away from new movies released the following weekend. For 100 additional viewers in weekend w to movies showing in both w and $w+1$, we observe about 46 more viewers of those same movies in $w+1$; an estimated 80% of these would otherwise have seen one of the new releases in $w+1$. In sum, a positive, weather-induced shock this week leads next week's viewers to substitute away from next week's new releases and towards movies that are randomly popular in theaters this week.

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

1.7.2 Social Multipliers by Age

Recall that the model in Section 1.2 predicts that the more viewers value the viewership of others – i.e., the stronger are network externalities – the larger is viewership in all subsequent weekends relative to viewership in opening weekend. In this section, we explore the empirical implications of this prediction by splitting films according to the age of their target audience; to the extent that different age groups place different values on peer viewership, our model predicts we should see divergence in estimated momentum.

To examine momentum from network externalities by audience age, we classify each movie into one of two categories based on its age appropriateness according to the MPAA: (1) "Child-friendly" and (2) "Adult-oriented". 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; adult-oriented movies are those rated PG-13 (Parents Strongly Cautioned; some material parents might consider inappropriate for children under 13 years) or R (Restricted; people under 17 years may only be admitted if accompanied by a parent or guardian).[56]

Table 1.7 shows estimated momentum from network externalities separately by age suitability and Figure 1.9 plots these estimated network externality effects by week in theater. Child-friendly movies exhibit significantly stronger momentum from network externalities in early weeks: for 100 additional viewers opening weekend, child-friendly films bring in just over 70 additional viewers the second weekend, compared with just 46 additional viewers for adult-oriented movies. However, the momentum among children falls in later weeks, and by the fourth week marginal momentum in child-friendly films is about on par with adult-oriented films. Still, cumulative momentum for child-friendly movies is roughly 50% higher than that for adult-oriented movies.

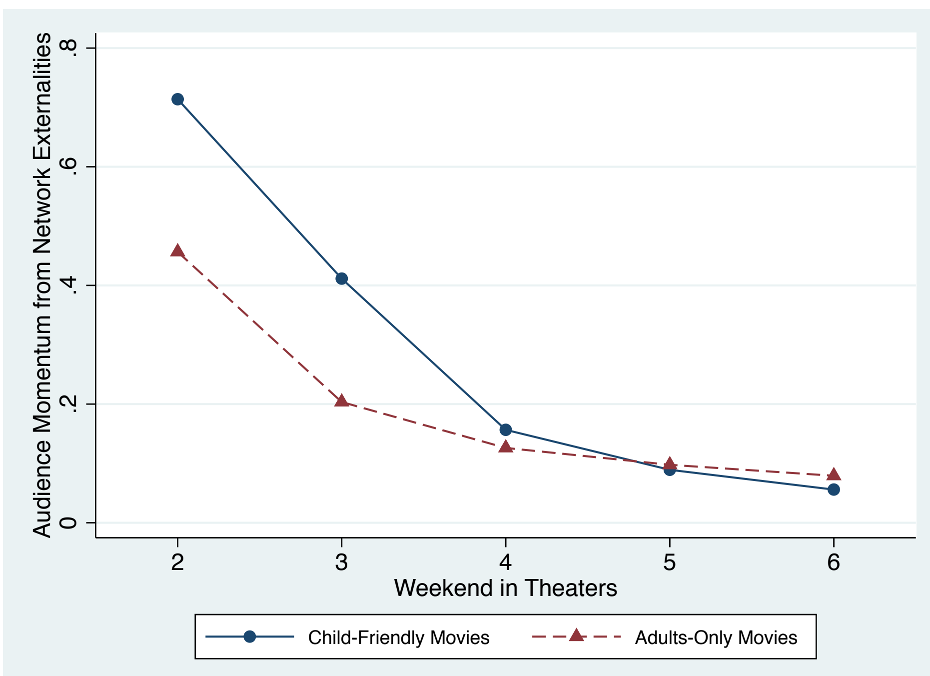
[56]Since NC-17 (No One 17 and Under Admitted, exclusively adult) likely captures a different demographic, we omit throughout the less than 1% of NC-17 films in our sample; their inclusion in the adult-oriented category, however, does not significantly change our results.

Table 1.7: *Network Externalities by Age Suitability*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Child-Friendly Movies	0.714***	0.411***	0.157***	0.0892***	0.0559**	1.427***
(obs. 802)	(0.113)	(0.0677)	(0.0467)	(0.0336)	(0.0240)	(0.209)
Adults-Only Movies	0.457***	0.204***	0.126***	0.0977***	0.0792***	0.963***
(obs. 1629)	(0.0505)	(0.0367)	(0.0274)	(0.0205)	(0.0161)	(0.135)
Differences:						
(Child - Adult)	0.257**	0.207***	0.031	-0.008	-0.024	0.464*
	(0.124)	(0.077)	(0.054)	(0.039)	(0.029)	(0.249)

Notes: This table replicates the IV results from Table 1.2 separately by film MPAA rating. Child-friendly films are those rated G or PG and adult-oriented films are rated PG-13 or R. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. The first stage results are included in Appendix Table A.3.

Figure 1.9: *Network Externalities by Movie Age Suitability*



Notes: This figure plots the coefficients from Table 1.7 for each of Weeks 2 through 6. Child-friendly films are those rated G or PG and adult-oriented films are rated PG-13 or R.

1.7.3 Aggregate Magnitudes

Altogether, our work presents two basic sets of facts: the first is that the weather has a significant influence on contemporaneous ticket sales; the second is that initial viewership shocks, such as those due to the weather, have multiplicative effects on subsequent viewership through network externalities. In what follows, we discuss the economic implications of these effects.

Movie-Going and Opening Weekend Weather

While Dahl and DellaVigna (2009) and Moretti (2011) note that the weather has an influence on ticket sales, our work quantifies the effect and shows that the magnitude of the contemporaneous response of movie-goers to weather is large. Our first stage results reveal that when one standard deviation more theaters are in the "beautiful weather" range of 75 to 80 degrees F, contemporaneous attendance to movies opening that weekend decreases on average by about 0.15 standard deviations (see Section 1.3.3). Yet this simple first stage relationship between weather and opening weekend sales understates the impact of the weather on movie-going for at least two reasons.

First, our focus on a single weather measure is sufficient for two-stage least squares analysis, but it does not capture the full impact of weather shocks on viewership. While our main measure capturing the percentage of theaters unexpectedly in the 75 to 80 degree range opening weekend does explain nearly 2% of aggregate abnormal ticket sales, all of the opening weekend weather measures we provide to LASSO together explain a full 10%.[57] This is sizable relative to other demand shifters. For example, Moretti (2011) finds that TV advertising explains 48% of total sales, suggesting the effect of opening weekend weather shocks on sales is roughly 20% that of advertising.

Second, our second stage estimates show that, through network externalities, opening weekend viewership shocks have multiplier effects which roughly double their aggregate

[57] Aggregate abnormal ticket sales are defined as abnormal ticket sales summed across opening weekend and subsequent weekends.

impact. With opening weekend ticket sales making up nearly one half of total ticket sales, this means a one standard deviation change in the 75 to 80 degrees F opening weekend weather measure leads to a 0.15 standard deviation change in *aggregate* viewership and, thus, aggregate revenue.

That a random factor like opening weekend weather can have such a substantial impact on ticket sales suggests care should be taken before interpreting sales figures as a direct measure of film quality, much less of the skill of individual cast and crew. It also points to the influence of otherwise small, random factors in determining human behavior, similar to the "butterfly effect" of chaos theory in which small changes in a system's initial conditions can yield large changes in outcomes. Here, however, the mechanism is peer effects (as opposed to the laws of physics). Weather – or other seemingly unrelated shocks – may similarly explain momentum in additional contexts, such as the "viral" spread of particular articles, songs, and apps that at times feels only tenuously related to quality, and perhaps even random.

Network Externalities and Firm Incentives

Recall that our estimates suggest that the aggregate effect of an incremental (and randomly induced) opening weekend ticket sale is doubled in subsequent weekends. This multiplier effect is in line with the approximately one-to-one relationship between opening weekend ticket sales and cumulative sales in subsequent weekends observed in the data, and helps to explain why opening weekend viewership explains 60% of the variance in ticket sales in future weekends.[58]

The multiplier effect we estimate is also consistent with the movie industry's strong focus on opening weekend viewership. This focus can be seen in the intense competition for release dates timed on major holidays, and in the fact that the vast majority of advertising

[58]Regressing abnormal opening weekend viewership on abnormal subsequent weekend viewership (where both are conditional on our baseline set of controls) yields an R-squared of 0.60.

outlays occur in advance of opening.[59] Prior research has proffered explanations for the industry's focus on opening weekend outcomes that do not take into account the social multiplier we estimate. For example, Caves (2001) argues that opening weekend viewership is cheaper to acquire in expectation because it does not rely as much on film quality. Our findings build on this literature by demonstrating that network externalities double the marginal benefit of an opening weekend viewer: if there is some reason to believe that viewership on opening weekend is cost-effective to acquire to relative to viewership in other weekends, then taking the multiplier into account makes it twice as cost-effective.[60]

A network externalities multiplier of the size we estimate may also have important implications for understanding and influencing individual choices in other contexts. Our work shows how, through network externalities, the level of local early adoption can affect subsequent (and aggregate) local adoption, even independent of quality or the precision of prior information about quality. Early adoption can be idiosyncratic in origin, as we exploit for identification, or can be crafted strategically. For example, a store might discount, or even give away, products during its Grand Opening; and at sports events, young companies often hand out samples of their energy bars and drinks in hopes of gaining early traction in a particular community.[61] Our results suggest that products and marketing campaigns designed specifically with the network externality multiplier in mind – e.g., those which leverage the subgroups among which network externalities are particularly strong (such as youth) and which make salient the component of the experience which is shared – may achieve additional gains.

[59] Elberse and Anand (2007) write that 90% of advertising occurs before opening.

[60] Although our estimates do not speak to whether the social multiplier is largest when it arises out of viewership on opening weekend versus out of viewership on subsequent weekends, our work shows that there *is* a large (opening-weekend) social multiplier, so that the aggregate contribution of any given demand shifter is larger than that its contemporaneous effect.

[61] Early adoption may of course be affected by information as well, a possibility tackled by Gill and Sgroi (2012) in a model analyzing how firms should choose pre-launch reviewers.

1.8 Conclusion

In this paper, 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. We implement LASSO variable selection methods to select from among the large number of potential weather instruments in order to generate a strong and econometrically sound first stage. We expect this approach will prove similarly useful in other settings where weather is a powerful and exogenous determinant of behavior but specifying the appropriate first stage is otherwise non-obvious.

Using our LASSO-chosen instruments, we estimate the effect of exogenous shocks to opening weekend viewership on viewership in later weekends. Consistent with our simple model of movie-going in the presence of network externalities, our results show that network externalities engender a multiplier effect: through network externalities, a shock to opening weekend viewership is doubled over the following five weekends. Almost all of this effect is observed at the local (MSA) level. Although we cannot fully reject the hypothesis that our results are at least in part generated by a learning story, the fact that our estimated momentum is largely local and that it varies neither with critic ratings nor with the precision of prior information about quality suggests the role of learning is limited.

While this paper has focused on in-theater movies, our findings may extend to other settings in which herd behavior is observed. DellaVigna *et al.* (2014) note that research examining why people vote has focused on two potential explanations: (1) that individuals vote because seek to affect the outcome of the election, or (2) that individuals vote because they believe it is the right thing to do. Our work may suggest a third: that individuals vote because they know others are voting, and they value sharing in that experience. If so, policies designed to accentuate or enhance the shared experience of voting could be particularly effective in increasing voter turnout.

Network externalities could also be an implement in the toolbox of the "economist as engineer" (Roth (2003)) to enhance otherwise largely independent experiences. With the rise of relatively solitary activities like gaming, remote work, and online learning, we expect that further research into where and how platforms might leverage the positive effects of network externalities to deepen participation and engagement would prove fruitful.

Chapter 2

When 3+1 > 4: Gift Structure and Reciprocity in the Field[1]

2.1 Introduction

Economists have long recognized that employees are often paid more than the market clearing wage, and that unemployed workers do not bid wages down to the point where supply equals demand. The neoclassical explanation for this phenomenon comes in the form of efficiency wage theories, generally arguing that employees will work harder when they receive high wages because they do not want to lose a high-paying job (Katz (1986)). This type of mechanism relies on repeated interactions between employers and employees. In one-time jobs without any consideration for future employment, the neoclassical model would argue that efficiency wages do not increase productivity.

At the same time, a robust literature based in psychology and behavioral economics demonstrates that people care about fairness, and that these fairness considerations may create incentives for reciprocation. If you give a gift to someone, that person might reciprocate – even in a one-shot game with no potential for future interaction (Berg *et al.* (1995), Fehr and Gächter (2000), Pillutla *et al.* (2003), Falk *et al.* (2008)). This principle has been

[1]Co-authored with Michael Luca and Deepak Malhotra

implemented in field settings as well. For example, Falk (2007) shows that including a small gift in a fundraising letter leads to higher donation levels.

In the context of labor economics, fairness concerns and reciprocity have been offered as an explanation for efficiency wages (Akerlof (1982), Akerlof and Yellen (1990), Fehr *et al.* (2009)). If employees view high wages as a gift, then they may reciprocate by working harder even though there is no financial incentive to do so. The thrust of this argument is that the market wage serves as a reference point, and employees will reward positive deviations from this reference point even in a one-shot employment contract with no career concerns.

Do employees work harder when they are paid more? Laboratory experiments have mostly shown that paying an unconditional bonus before the work starts causes workers to reciprocate by working harder (e.g. Fehr *et al.* (1993), Fehr and Gächter (2000), Hannan *et al.* (2002), Charness (2004)). In seminal work in the field, Gneezy and List (2006) look at two field settings, hiring roughly 40 workers to test whether paying higher than market wages increases output in a library data entry task and in door-to-door fundraising. Specifically, Gneezy and List compare the output of workers assigned to a "gift" treatment, in which workers are hired at a low wage and then offered a raise immediately before starting work, with the output of workers assigned to a "non-gift" treatment, in which workers are hired at and paid the low wage. They find that workers who receive the "gift" (i.e., the additional money) exert higher productivity for the first few hours of the task but that the effect wears off after a few hours; in their setting, the temporary increase in productivity does not justify its cost.

However, because this prior research has not varied the base wage, the set-up cannot identify whether reciprocity is triggered because wages are above the market rate (which is the argument laid out in the literature that uses reciprocity to explain observed above-market wages) or because workers unexpectedly receive the gift of a raise after having already agreed to a job (which makes the higher wages considerably more salient). In other words, wage amount (high vs. low) is confounded with wage structure (salient gift vs. no gift). In this paper, we offer a large-scale gift exchange experiment in a field setting where we vary both whether or not a worker receives an unexpected raise before starting work as well as the base wage offered to potential hires. This subtle difference allows us to differentiate between the impact of salient "gifts" on performance and the impact of above-market wages on performance.

Our results show that the way in which a wage is structured (and not simply its level) may be essential to generating reciprocity. In our study, hiring at and paying workers a high wage ($4) leads to no increase in productivity, i.e. to an economically identical and statistically indistinguishable amount of productivity relative to hiring at and paying workers a low wage ($3). However, hiring workers at a low wage and then offering them an unexpected raise ($3+$1) significantly increases performance. Common models of reciprocity in labor markets (e.g., Akerlof (1982), Akerlof and Yellen (1990)) assume that high wages alone are the determinant of reciprocation, but these models do not differentiate between our 3+1 and 4 offers. To our knowledge, field experiments have similarly not differentiated between these two treatments.

Our experiment takes place on oDesk, an online labor market with several million registered contractors. Using the oDesk platform allows us to vary wages and gifts in a setting where workers are accustomed to tasks like ours. Because we conduct a natural field experiment, the employees do not know that they are part of an experiment. Indeed, the oDesk platform is a powerful setting for an experiment like ours because using it allows us to compare and control for the entire oDesk work histories of our employees. Furthermore, the oDesk marketplace allows us to conduct targeted hiring by directly inviting workers to take up our job instead of simply posting a job publicly and waiting for applications. This means we are able to hire workers at different base wages (without individuals knowing how much others have been paid) so that we can test whether it is the base wage amount or the salience of the gift that affects performance. Normally selection would be a concern when hiring at different wages, but we are able to address the potential for selection through the combination of high take up rates (the take up rate in our study is 95% among workers with prior experience), comparison of characteristics across treatments for job takers (they are not significantly different), and a conservative robustness test.

Our experimental design, which we describe in more detail in the next section, proceeds by hiring three groups of oDesk workers for a data entry task, all of whom have requested wages of less than \$3 per hour according to their oDesk profiles. We are clear in our recruitment messages that this is a one-time job. The first group is hired at \$3 per hour (i.e., the “3" condition). The second group is also hired at \$3 per hour, but before starting work, workers in this group are told that we unexpectedly have extra money in the budget and will pay an extra \$1 per hour, so that the total they will receive is \$4 per hour (“3+1"). The third group is hired directly at \$4 per hour so the fact that we are paying them the higher (above-market) wage does not signal a “gift" in a salient way (“4"). To increase the validity of the results, we choose a data entry task (entering CAPTCHAs, to be described in more detail later) that is fairly common on online labor markets, and we only recruit workers who self-report data entry as a specialty on their oDesk profiles.

Consistent with the notion of reciprocity, we find that higher wages that include a salient gift (3+1) lead to higher and more persistent productivity across our task relative to the other two groups (3 and 4). More specifically, paying \$3+\$1 yields a 20% increase in productivity compared to paying \$4, with no additional cost. Compared to paying \$3, paying \$3+\$1 results in a 20% increase in productivity with a 33% increase in cost. Notably, we find that varying the base wage from \$3 to \$4 in the original contract has no statistically distinguishable effect on productivity – in fact, the point estimate of the effect is 0. To our knowledge, our experiment is the first to include the high base wage (4) condition, allowing us to better understand the extent to which salience (vs. mere offering) of a gift drives reciprocity.

In addition our main results, we also find suggestive (i.e., economically large but not statistically significant) evidence that the effect is largest for those workers for whom the +1 gift may have been most salient (e.g., experienced employees, who are familiar with the typical wage structure on oDesk).

Altogether, our results suggest that unexpected gifts (3+1) can increase productivity but that this effect is due to the salient nature of the gift and not simply due to the fact that a worker is receiving a higher, above-market wage. Indeed, our results suggest that reciprocity and gift giving are unlikely to explain most efficiency wages as we see them in the world around us, because high wages written into the initial wage contract do not seem to elicit any productivity increase.

Our paper proceeds as follows. In Section 2.2, we outline our experimental design. Next, in Section 2.3, we present our findings, and finally in Section 2.4 we conclude.

2.2 Experimental Design

Our experimental methodology proceeded in two steps. First, we selected a sample of oDesk workers, randomized those workers across treatments, and invited the treated workers to accept our job. Second, treated workers who responded were sent a link to a website where they could complete our data entry task. For workers in the 3+1 condition, this message also provided notification of a change–a $1 per hour increase–in their wages. This section describes the sample, the treatment conditions, and the task itself in more detail.

2.2.1 Sample

We began our sample selection by restricting the universe of oDesk workers to those who claim to specialize in data entry. In particular, we required that (a) worker profiles are returned by a keyword search for "data entry" and (b) workers classify themselves as Administrative Support workers with a Data Entry specialty.[2] Next, we further restricted the sample to workers that list a requested hourly wage of between $2.22 and $3.33. Since oDesk charges workers a 10% fee on gross earnings, this restriction amounts to requiring that workers request net wages between $2 and $3 per hour.

[2] We note that oDesk specialties are self-reported so neither of these two restrictions required a worker to actually have any experience in data entry on the oDesk platform (or elsewhere).

It is relatively easy for someone to create an oDesk profile, regardless of whether they actually plan to seek work on oDesk, so in an attempt to select only serious candidates we further restricted the sample by requiring that workers had (1) logged into oDesk within the last 30 days and (2) were listed as "oDesk ready," which means they had passed a multiple choice exam that tested their knowledge of the oDesk interface. Finally, to ensure workers were autonomous and not making decisions within a larger organization, we required that workers were listed as independent contractors, meaning they were unaffiliated with a company.

A total of 17,482 workers satisfied these joint criteria at the time of data collection. From this set of workers, we randomly selected 540 workers and allocated them randomly across our 3 treatments.[3] Panel A of Figure 2.1 presents the initial recruitment messages with which we invited selected workers to take up our task. Of the 540 workers we invited to take up our task, 266 accepted our job offers. We exclude 36 of these workers due to a technical glitch described below, so our final sample is comprised of 230 workers. Notably, among workers with prior experience, our take up rate was 95%.

2.2.2 Treatments

As we discussed above, we randomized workers into three treatment groups. Workers in two of the treatment groups were initially offered \$3 per hour while workers in the third group were initially offered \$4 per hour. (Technically, we would pay \$3.33 and \$4.44, respectively, because 10% of gross wages go to oDesk. For simplicity, we refer to wages in net terms throughout the paper.)

We refer to the two groups of workers who had been offered \$3 per hour as the "no gift" (3) and "salient gift" (3+1) treatments, and we call the \$4 per hour treatment the "gift" condition. After accepting our job offer, workers in all treatments were reminded of the task instructions and presented with a link to a webpage where the task was located. Workers in

[3]We oversampled the \$3 per hour treatment because those workers participated in a trust-building exercise independent of this experiment that took place after this experiment was completed.

Figure 2.1: *Job Offer Messages*

Panel A: Recruitment messages

All treatments

We are currently looking to hire a group of people to help with simple data entry. The job consists of looking at a photograph of a word, and typing that word into the space provided. This is a four-hour job, with the goal of entering as much data as possible while minimizing the number of mistakes. Specifically, we need as many correctly entered words as possible in four hours because we need the data for a future task and only correct entries can be used. You will have seven days to complete the task.

You will be paid $3 ($4) per hour. Therefore, your total payment for the four hours will be $12 ($16). We hope you will accept this job.

Panel B: Acceptance messages

Treatments: 3 & 4

Great, you are hired.

By the way, we want you to know that this is a one time job; we do not expect to have more work in the future.

Below, you will find a link to a page where you will do the data entry. As we mentioned, the job consists of looking at a photograph of a word, and typing that word into the space provided. Please enter words for four hours, after which you will be ineligible to receive further pay. Finally, please take no more than a week. We will not accept work done more than seven days after you receive this assignment.

Link to job: **here**

Treatment: 3+1

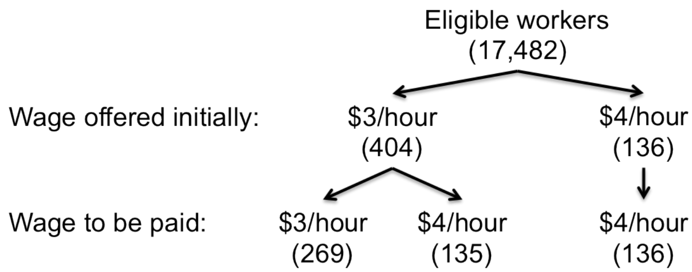
Great, you are hired. As it turns out, we have a bigger budget than expected. Therefore, we will pay you $4 per hour instead of $3 per hour, so the total you can earn is $16.

By the way, we want you to know that this is a one time job; we do not expect to have more work in the future.

Below, you will find a link to a page where you will do the data entry. As we mentioned, the job consists of looking at a photograph of a word, and typing that word into the space provided. Please enter words for four hours, after which you will be ineligible to receive further pay. Finally, please take no more than a week. We will not accept work done more than seven days after you receive this assignment.

Link to job: **here**

Figure 2.2: *Experimental Design*



the salient gift treatment (3+1) were additionally notified in the same message that we "have a bigger budget than expected ... [and that] we will pay ... $4 per hour instead of $3 per hour".[4] Panel B of Figure 2.1 presents the messages we sent to workers to let them know we had agreed to hire them, and Figure 2.2 summarizes the experimental set up.

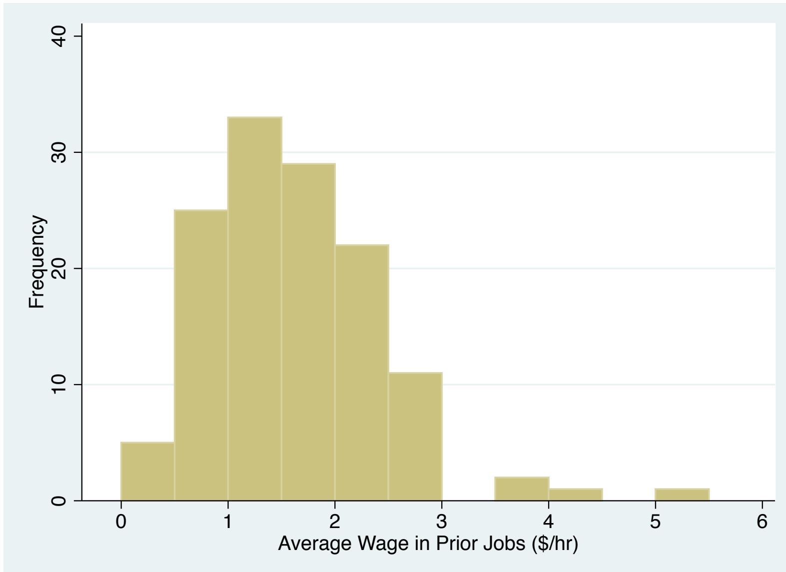
Before proceeding, we note that although it is difficult to ascertain what constitutes the market wage for a specific task at a particular point in time, even our lowest offer of $3 per hour appears generous compared with wages paid to our workers in their prior roles on oDesk. Figure 2.3 presents a histogram of the average prior wage earned by experienced workers invited to take up our task.[5] The graph shows that prior wages are roughly normally distributed, with a peak in the density just below the median of $1.52 per hour. For 97% of workers, average prior wages fall below $3 per hour. Since looking only to each worker's average prior wage might disguise variance in the underlying wages, we also compute the average percentage of prior jobs held which paid, respectively, at least $3 and $4 per hour.[6] After weighting jobs by hours worked, on average only 5.3% of prior jobs paid

[4]This message was phrased so as not to emphasize any sacrifice the firm was making in paying a higher wage, nor to emphasize our desire that a worker reciprocate. However, our design does not allow us to analyze the impact of the specific wording of this message.

[5]The picture is nearly identical for the subset of workers who accept our job: as we show below, only 5% of experienced workers do not accept our offer of employment. Our focus on prior wages instead of requested wages is motivated primarily by the fact that requested wages are self-reported and thus difficult to interpret. Moreover, there is little variation in the distribution of requested wages: requested wages in our sample are essentially bimodal, with 79% of workers requesting either $2 or $3 per hour.

[6]These percentages decrease slightly to 3.8% and 1.6%, respectively, if we compute the overall percentage of prior jobs held that pay at least $3 or $4 per hour as opposed to the within-worker statistics being averaged across workers.

Figure 2.3: *Distribution of Prior Wages*



Notes: This figure presents a histogram of average prior wages. The sample is all invited workers, restricted (by necessity) to those with prior experience.

at least $3 per hour, and only 1.7% of prior jobs paid at least $4 per hour. We infer that our task was likely highly desirable even at our lowest wage of $3 per hour; this is consistent with our high take up rates (particularly among experienced workers), discussed below.

2.2.3 Task

As we explained in our initial recruitment messages, the task we presented to workers asked them to correctly enter as many CAPTCHAs as possible in the four hours allotted. Figure 2.4 presents a screenshot of the task itself, as seen by workers. CAPTCHA is an acronym for "Completely Automated Public Turing test to tell Computers and Humans Apart," which is a system that asks you to transcribe a word or phrase that is presented to you as a picture. Many online companies use CAPTCHAs to prevent automated software from easily accessing information or making decisions without a human being involved. For example, Ticketmaster requires potential buyers to enter a CAPTCHA before purchasing tickets in order to stop a person from using a program that repeatedly buys tickets (which is something that a scalper may otherwise do). On online labor markets such as oDesk, there is a high level of demand for people to do data entry (and in fact, even specifically to enter CAPTCHAs), which means that our task would come across as a reasonably natural request. Importantly, CAPTCHA entry requires effort, but has a straightforward and simple output (correctly entered CAPTCHAs), giving employees sufficient opportunity to "repay the gift", all factors that Englmaier and Leider (2012) and Kessler (2013) find are crucial for gift exchange to occur.

2.2.4 Experimental Validity

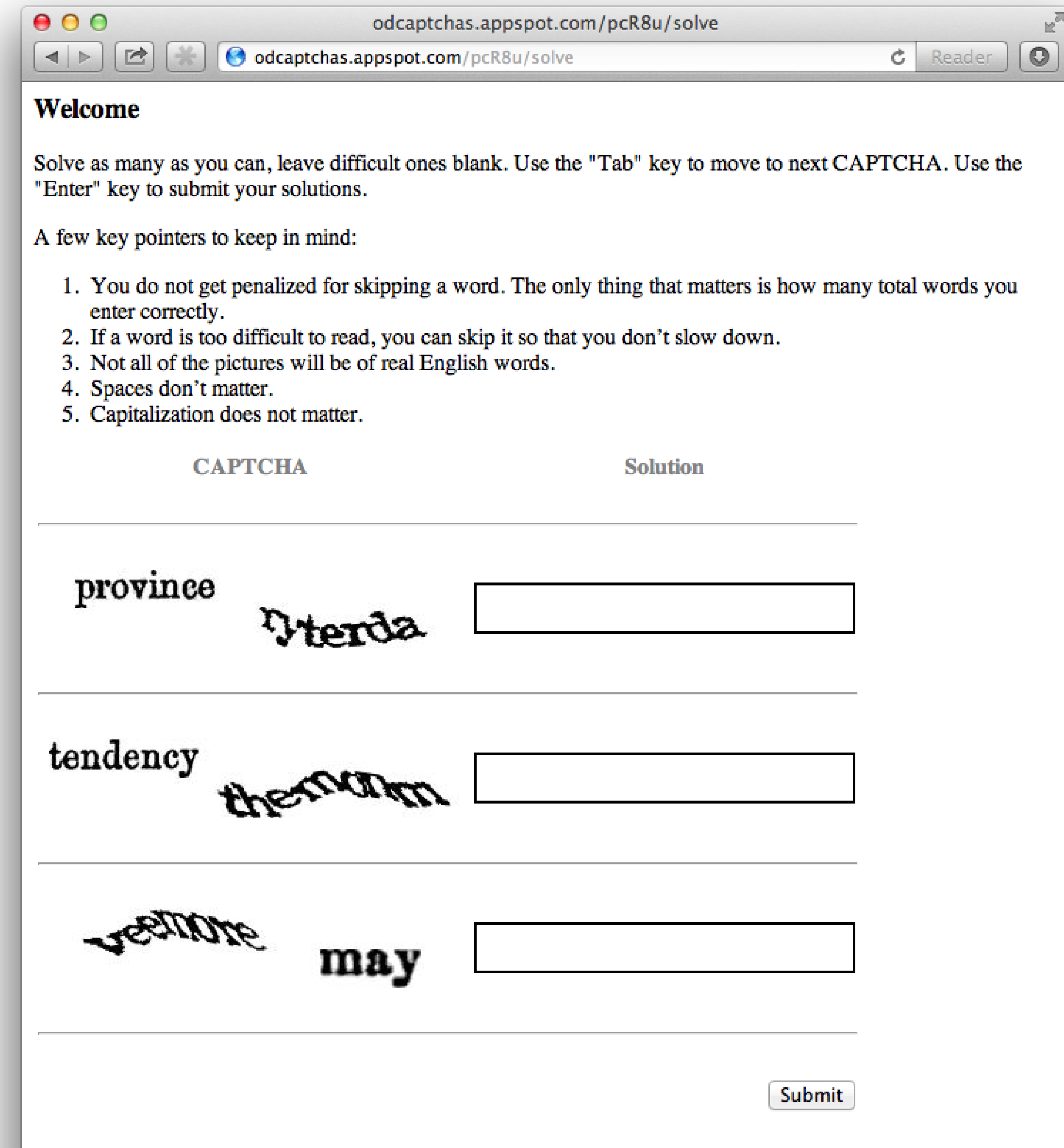
Even though our wages are generous, the nature of our experimental design–which pays different wages to different workers–in combination with the fact that we do not have 100% take up means we are concerned about the potential for selection. Thus, before proceeding to our main findings, we analyze how worker characteristics are related to treatment and job take up.[7] Throughout, we exclude results from 12 workers who did not complete the 4 hours of work, as well as 24 workers in the initial wave who were able to complete more than 4 hours of work due to a technical glitch that allowed them to exceed the time limit.[8] Thus, of the 540 workers invited to take up our task, we are left with 230, all of whom are included in our analysis.

Table 2.1 presents the results. In Panel A, we first present a base set of statistics for all workers, and then in Panel B we present an extended set of statistics which are available only for the subsample of experienced workers (i.e., those with at least one prior job). We separately focus on experienced workers because they exhibit much higher take up rates, provide more data for analyses, and have greater familiarity with tasks and wages on oDesk relative to inexperienced workers. Within each panel, the first set of three columns presents statistics on key characteristics (e.g., number of prior jobs), separated by treatment

[7] We remind the reader that we did not stratify the sample on any characteristics.

[8] There was no statistically distinguishable difference across the three treatment groups in the likelihood of an employee working for more or less than 4 hours; we exclude these workers because allowing employees to work for different lengths of time makes it harder to compare total productivity across employees. Including results for these workers does not change our baseline results, however; see Table B.1.

Figure 2.4: *CAPTCHA Task*



Notes: This figure presents a histogram of average prior wages. The sample is all invited workers, restricted (by necessity) to those with prior experience.

group, for those who did not accept our job offer, while the second set of three columns presents analogous statistics for those who did. The statistics presented are the mean of a characteristic by subgroup, and then below, in parentheses, the p-value from a T-test comparing the values of a subgroup's covariate with the same covariate of the analogous 3+1 subgroup. Our main empirical analysis focuses on the productivity of employees who accepted our job offers, so our objective in this analysis is to both verify that the treatment groups are balanced and also to verify that there is no apparent selection among employees who accepted our job offers.

Although our overall take up rate is 46%, job acceptance rates and other worker characteristics are similar across treatment groups. Overall, about two thirds of job takers are experienced, and the take up rate is 95% among workers with prior experience and 22% among inexperienced workers.[9] The only notable difference across the three groups of job takers (which is still not statistically significant) is that the number of prior jobs is lower in the 4 group than in the 3+1 and 3 groups. Probing more closely, we see that this is due in combination to the fact that an (insignificantly) larger percentage of takers in the 4 group are inexperienced and also to the fact that experienced takers in the 4 group themselves have (insignificantly) fewer prior jobs. The former could be a cause for concern but, as we show below, our results are actually most pronounced for experienced workers. Moreover, it seems unlikely that the latter feature (insignificantly fewer prior jobs among experienced workers in the 4 group) is biasing our results given that the difference is relatively small and that take up rates among experienced workers are all close to 100%. Thus, in aggregate, the data support the validity of our experimental design.[10] Finally, to further ensure that selection is

[9]In an oDesk experiment that uses a similar pool of data entry workers to analyze the impact of reviews on future hiring, Pallais (2014) also finds that experienced workers are more likely to take up jobs. However, the divergence is less extreme in her context: she respectively finds 54% and 33% of experienced and inexperienced workers accept her job offers.

[10]We note that two characteristics in the sample of experienced workers who did not take up the job are actually statistically significantly different across treatment groups. However, we believe this is not a cause for concern given this subgroup's exceptionally small sample size – each treatment in this subgroup is composed of three or fewer workers. Nevertheless, as an additional robustness check against selective take up, we repeat the analysis in Table 2.1 (in which we compare characteristics across treatments) but this time pool the 3 and 3+1 treatments since these treatments were hired at the same wage. For brevity, we exclude the results here but

Table 2.1: *Comparing Worker Characteristics Across Treatments*

	Did not take up the job			Took up the job		
	3	3+1	4	3	3+1	4
Panel A. All workers						
Experienced	0.021	0.048	0.029	0.691	0.69	0.629
	(0.301)		(0.527)	(0.987)		(0.482)
Number of prior jobs	0.162	0.238	1	8.082	7.397	4.548
	(0.881)		(0.194)	(0.719)		(0.185)
Wage requested	2.646	2.678	2.666	2.741	2.738	2.815
	(0.675)		(0.884)	(0.972)		(0.392)
N	142	63	69	110	58	62
N taker / N invited	-	-	-	44%	48%	47%
Panel B. Experienced workers						
Number of prior jobs	7.667	5	34.50	11.70	10.72	7.231
	(0.757)		(0.0231)	(0.703)		(0.236)
Wage requested	3.053	2.777	2.775	2.761	2.750	2.882
	(0.585)		(0.998)	(0.909)		(0.234)
Average prior wage	1.769	1.494	2.233	1.706	1.957	1.908
	(0.716)		(0.397)	(0.191)		(0.828)
Average prior rating	4.957	4.503	4.615	4.565	4.548	4.495
	(0.285)		(0.802)	(0.913)		(0.759)
Worked in last 30 days	0.333	0	1	0.447	0.525	0.410
	(0.314)		(0.0301)	(0.429)		(0.311)
N	3	3	2	76	40	39
N taker / N invited	-	-	-	96%	93%	95%

Notes: This table presents characteristics of workers invited to take up our job, split into workers that did not take up the job (first set of three columns) and workers that did take up the job (second set of three columns). Panel A analyzes characteristics for all workers, while Panel B examines experienced workers only. Means are presented in the first row, while the second row presents p-values from T-tests comparing respective characteristics in a given sub-sample with those of the analogous 3+1 treatment. The sample sizes in the calculation of average prior wage for takers in the 3, 4, and 3+1 treatments are respectively 66, 30, and 33, because some experienced workers had only fixed-price prior jobs. Similarly, the sample sizes in the calculation of average prior rating for takers in the 3, 4, and 3+1 treatments are respectively 72, 38, and 37, because some experienced workers' first and only contract was ongoing at the time of the experiment.

not driving results, we show below that our main results are robust to a conservative test where we code all non-takers as having completed 0 CAPTCHAs.

2.3 Results

This section presents the results of our experiment. We first discuss the main effect and its persistence, and then analyze how the effect varies with worker characteristics.

2.3.1 Main Effect: 3+1 > 4 = 3

The gift exchange literature has posited that high wages elicit reciprocity, which could in turn rationalize above-market wages even in the context of unrepeated employment. This suggests that market wages might be a reference point and that paying more would elicit higher productivity. In this case, our 4 and 3+1 conditions should elicit the same response, which would be higher than the 3 condition. To our knowledge, we are the first to include something akin to our “4 treatment", which allows us to shed light on the conditions under which we should expect workers to reciprocate high wages.

Figure 2.5 presents the main effect graphically and column (1) of Table 2.2 presents it in regression form. In particular, column (1) of Table 2.2 presents results from a regression of the total number of correctly completed CAPTCHAs on a constant and dummies for treatments 3 and 4, respectively (so all estimates are relative to the 3+1 group). The figure and the table both show that the respective numbers of CAPTCHAs correctly completed by the 3 and 4 groups are nearly identical, and we cannot reject the hypothesis that they are the same. However, the 3+1 group correctly entered 146 more CAPTCHAs relative to both groups over the course of the task (a 20% increase), and this result is statistically significant at the 5% level.

Column (2) shows the results are unchanged if the number of completed correct CAPTCHAs is specified in log form, while in column (3), we show that the treatment

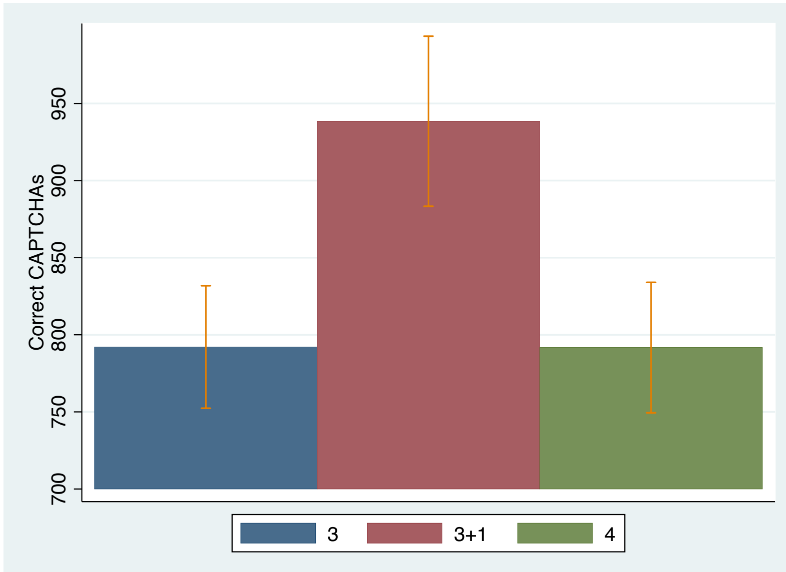
none of the differences between the pooled treatments hired at $3 per hour and the treatment hired at $4 per hour are statistically significant at the 10% level.

Table 2.2: *Performance*

	(1) Correct	(2) Log correct	(3) Percentage correct	(4) Correct per quarter
Wage = 3+1	omitted	omitted	omitted	omitted
Wage = 3	-146.4**	-0.286***	-1.224	-32.97**
	(67.93)	(0.100)	(1.433)	(16.23)
Wage = 4	-146.8**	-0.214**	-0.797	-32.44*
	(69.36)	(0.0990)	(1.335)	(16.54)
Quarter = 1				omitted
Quarter = 2				16.12*
				(8.698)
Quarter = 3				18.33*
				(11.02)
Quarter = 4				20.64
				(13.62)
Wage = 3 x Quarter 2				2.161
				(10.09)
Wage = 3 x Quarter 3				-6.182
				(13.02)
Wage = 3 x Quarter 4				-10.52
				(15.05)
Wage = 4 x Quarter 2				-7.685
				(10.16)
Wage = 4 x Quarter 3				-8.070
				(12.76)
Wage = 4 x Quarter 4				-1.880
				(15.42)
Constant	938.5	6.754	85.33	219.9
	(55.04)	(0.0565)	(0.980)	(12.87)
N	230	230	230	920
Adjusted R-squared	0.017	0.017	-0.006	0.0014
Effects of treatments 3 and 4 relative to 3+1 by quarter:				
[Wage = 3] + [Wage = 3 x Quarter 2]				-30.81
				(18.94)
[Wage = 3] + [Wage = 3 x Quarter 3]				-39.15**
				(19.77)
[Wage = 3] + [Wage = 3 x Quarter 4]				-43.49**
				(19.15)
[Wage = 4] + [Wage = 4 x Quarter 2]				-40.13**
				(18.56)
[Wage = 4] + [Wage = 4 x Quarter 3]				-40.51**
				(20.22)
[Wage = 4] + [Wage = 4 x Quarter 4]				-34.32*
				(20.50)

Notes: This table presents our main results. Column 1 analyzes the number of correct CAPTCHAs by treatment, and column 2 does the same but where the outcome is in logs instead of levels. The outcome in column 3 is the percentage of CAPTCHAs that were correct, and column 4 presents results from a panel regression where the outcome is the number of correct CAPTCHAs completed by each worker in each quarter of their work. For ease of interpretation, the bottom panel presents treatment effects relative to 3+1 by quarter. All standard errors are robust, and standard errors for column 4 are clustered at the worker level. *** p<0.01, ** p<0.05, * p<0.1.

Figure 2.5: *The Gift Leads to Higher Productivity than either High or Low Base Wages*



Notes: This figure shows our main result. Thick bars present means and confidence bands present one standard error relative to the mean.

effect measured in number of correct CAPTCHAs entered is not due to a change in the ratio of correct to incorrect entries. Overall, we infer that the gift only matters if the wage is structured in a way that the gift component is made salient (e.g., by presenting it separately, as in this study). Paying 4 elicits the same amount of productivity as paying 3, but not the same as 3+1.

Although our sample is balanced on observables and take up rates were similar across all treatments (see Table 2.1), we conduct an additional robustness check to verify that our result that 3+1>4 is not driven by selection induced by the fact that different treatments are hired at different wages. That is, a potential explanation for why 4 performs worse than 3+1 is that negative-sorting led less-able workers to refuse the job invitation for $3 per hour (into the 3+1 treatment), but those same kinds of workers did accept the job invitation for $4 per hour. Such a scenario could lead us to falsely conclude that the 3+1>4 when really selection was driving the result. To test whether negative selective take up is driving our result, we run a regression where all non-takers are coded as having completed 0 CAPTCHAs and all takers are coded normally. If the differential yield is driving our result and the incentive provided by 4 is actually equivalent to that provided by 3+1, then 4 should perform at least as well as 3+1 in this regression since this new specification estimates the average treatment effect conditional on job invitation, instead of the average treatment effect conditional on job take up. The results are reported in column (2) of Table B.1. The 3+1 treatment still performs better than 4 and the difference remains statistically significant at the 10% level. We conclude that our results are not driven by selection.

Aside from the fact that we observe 3+1>4, the extent to which our salient gift treatment outperforms the no gift treatment (3+1>3) contrasts with Gneezy and List (2006), Hennig-Schmidt *et al.* (2010), Kube *et al.* (2012), and concurrent but independent work by Esteves-Sorenson and Macera (2013), all of whom find mild to no effects of gifts when hiring undergraduate students for field experiments (although Kube *et al.* (2012) find that nonfinancial gifts do have a significant effect). As we mentioned above, one possible factor contributing to this difference is the fact that in all of these settings, workers presumably had little or no experience with the task they were assigned. As a consequence, they likely had weak prior beliefs about both the task and how wages would be presented, which may have reduced the salience–and hence, influence–of any gifts received.

We next examine the persistence of the main effect, and then in Section 2.3.3, we explore how the impact of the salient gift interacts with worker characteristics which might further accentuate the salience of the "salient gift" condition. Our results suggest that the extent to which a gift is salient – i.e., not just "what you pay", but "how you pay" – could have important implications for the extent to which gift exchange occurs.

2.3.2 Persistence of the Effect

The final column in Table 2.2 examines how the treatments impacted worker performance over the course of the 4-hour task. Our database recorded the timestamp of every CAPTCHA entry, allowing us to examine the time series of responses. We gave workers one week to complete the task, and workers were not required to complete the 4 hour task in a single (uninterrupted) session.[11] In order to analyze the persistence of performance, we break the time series of entries into four quarters by first truncating any gaps between CAPTCHA entry that are longer than 10 minutes to just 10 minutes, and then breaking the total time between the first and the last CAPTCHA entry (with truncated breaks) into four equal blocks of time.[12] Thus, each block of time represents one quarter of the time a worker spent working on our task, or roughly one hour, although the sum of the four quarters is not exactly four hours because of the way we truncate large gaps between CAPTCHA entries. The dependent variable in this regression, then, is the number of correct CAPTCHAs entered in each quarter by each worker (so observations are the worker-quarter level; we cluster standard errors by worker). The un-interacted treatment coefficients in the top rows show that that the workers in the salient gift treatment completed an average of 32 to 33 more correct CAPTCHAs in the first quarter of the task, and this difference is significant at the 5% level relative to the 3 group and at the 10% level relative to the 4 group.

To examine the effect of the salient gift over the course of the experiment, we present in the bottom panel of Table 2.2 the sum of the treatment and treatment-quarter estimates. The salient gift treatment consistently outperforms the other treatments in all quarters by 30 to 40 CAPTCHAs. We infer from these results that the salient gift treatment increased productivity across the length of the task. Levitt and List (2007) suggest that lab evidence on reciprocity may not generalize to the field because of the short-term nature of the response (as found in Gneezy and List (2006)). While our results do not speak to the longer-run persistence of the effects of a single gift, they do show that the impact of a gift is not always as ephemeral as previous evidence has suggested.

[11] oDesk's platform logged and billed employers for the hours a worker was logged into oDesk's task-specific interface. Employers were given the option of specifying a time limit, which, in our case, was set to 4 hours.

[12] We chose 10 minutes as a natural length of time to use for truncation because the oDesk platform verifies its workers' focus by taking randomly-timed screenshots, and 10 minutes is the stated approximate time between screenshots. Thus, a break longer than 10 minutes is likely to represent true time away from the task.

2.3.3 Supplementary Analysis: Exploring the Role of Worker Characteristics

Our main results show that high wages increase productivity only when it is salient that the wage is high. In this section, we explore how the impact of the salient gift varies with worker characteristics. One might expect that the impact of the salient gift would be higher for workers who have more familiarity with how wages are typically structured (i.e., without gifts) in this context. For example, workers with prior oDesk experience may be more influenced than those who have not used oDesk before. Before turning to the results, we note that the sample sizes analyzed in this section are relatively small, so while the results are suggestive, the differences-in-differences across demographic cuts are not statistically significant at conventional levels.

Table 2.3 shows the mean number of correctly completed CAPTCHAs and its standard error, organized horizontally by treatment (3, 3+1, and 4) and demographic category (e.g., experienced vs. inexperienced), and vertically by different demographic sub-groups (e.g., different levels of experience). Panel A presents results for all workers while Panel B presents results for experienced workers only, and stars denote significance of treatments 3 and 4 relative to the relevant 3+1 group. The table shows that the estimated effect is indeed most pronounced for employees with characteristics associated with strong priors: it is strongest for experienced employees and, among experienced employees, those with more prior jobs, and those who have worked most recently. The implications of these results differ from (but are not inconsistent with) the findings of Cohn et al. (forthcoming), who find gifts can have real effects on productivity, but that they are most effective for workers who perceive the base wage as unfair–since our wages are all well above historical wages (as discussed in Section 2.2.2), it seems unlikely that fairness concerns are driving our results.

The final set of estimates suggests the salient gift's generosity appears to play a role as well: the estimated effect is largest for workers with the lowest prior wages (but the difference in effect size is not statistically significant). An effect of this magnitude actually suggests that for the workers who previously earned wages below the median of our sample, the 3+1 treatment is more efficient than the 3 treatment in terms of CAPTCHA completions

Table 2.3: *Performance and Worker Characteristics*

Treatment	3	3+1	4	3	3+1	4
			Panel A. All workers			
	No experience			Experience		
Mean	834.4	861.8	734	773.1**	973.0	825.7*
Std. error	(73.74)	(76.39)	(74.58)	(47.29)	(72.13)	(50.81)
			Panel B. Experienced workers			
	Number of prior jobs < median			Number of prior jobs ≥ median		
Mean	693.9	891.0	841.2	827.7*	1055.1	807.7*
Std. error	(75.58)	(93.73)	(60.36)	(59.91)	(108.9)	(86.40)
	Did not work in last 30 days			Worked in last 30 days		
Mean	807.7	928.6	862.7	730.5**	1013.2	772.7*
Std. error	(66.35)	(76.13)	(70.93)	(67.13)	(120.2)	(70.80)
	Average prior wage < median			Average prior wage ≥ median		
Mean	722.9**	1111.6	700.8**	869.5	952.3	816.7
Std. error	(65.85)	(147.9)	(84.06)	(72.59)	(78.76)	(83.02)

Notes: This table analyzes how the number of correct CAPTCHAs varies with worker characteristics. Each set of statistics presents means and standard errors by treatment (listed at the top of the table) and sub-sample. Panel A presents results for all workers while Panel B restricts the sample to experienced workers only. The median number of prior jobs is 6 and the median prior wage is $1.50 per hour. ** and * respectively denote statistical significance at the 5% and 10% levels of the 3 and 4 treatments relative to the 3+1 treatment (in the same sub-sample).

per dollar expenditure. The $1 per hour gift (3+1) increased average CAPTCHA completion from 723 to 1112 for this group, a 54% increase in productivity, at a cost increase of only 33%. In other words, for a subset of (targetable) employees, the 3+1 wage structure is more efficient than both the 3 and the 4 wage structures. (Our main results had earlier shown that 3+1 is efficient relative to 4.)

2.4 Discussion

We find that providing employees with an unexpected pay increase can increase productivity–even when there is no prospect for future employment. However, high wages that actually look like the types of efficiency wages we usually see in the field did not have the same impact on productivity as our salient gift treatment, and in fact had no discernible effect at all. Taken in aggregate, this suggests that "how you pay" can be as important as "what you pay." In this section, we discuss limitations of our work and areas for future research.

2.4.1 Labeling Gifts

Our main results suggest that salient gifts can lead to increased productivity. Salience can stem from many factors, e.g., the unexpected nature of a gift, from it being labeled more conspicuously, or based on the amount of attention it will garner based on when and how it is presented. In an organizational context, for example, this implies that firms that are looking to be more generous to employees might benefit from labeling the high wages and other gifts that they give to employees, rather than simply assuming that employees are correctly inferring the intention behind a wage.

To see how an organization might go about labeling gifts, consider the recent decision by the popular clothing store Gap to raise its minimum wage offered to U.S. employees to a rate that exceeds local and national standards (see Greenhouse (2014)). Many employees are now earning a wage that is above the minimum, and above that offered by neighboring stores. When an employee begins, Gap could simply tell them the starting wage in the offer letter and assume that the employee knows that this is generous. An alternative option would be to include a few lines in the letter explaining to the employee that the wage being offered is in fact higher than a competitor analysis would dictate, but that Gap wants to treat its employees well and pay a fair, living wage. Our findings suggest that this type of discussion could affect performance.

Another approach suggested by our findings is that organizations may benefit from timing and targeting gifts on the basis of when the gifts are more likely to be salient. For example, a bonus that is given independently may be more powerful than one that is included in a paycheck. Likewise, the same gift may be more or less salient to certain types of employees based on their prior expectations: a gift that is quite unprecedented in size or type may have a greater effect on employees with greater tenure than on those who are new to the organization.

2.4.2 Can Gifts be Efficient?

Clearly, there is a tradeoff involved in giving a gift to employees. An organization could simply opt to hire more employees at a lower wage. This would be efficient if the impact of a gift was small relative to the productivity of hiring a marginal worker. Our study provides preliminary evidence that targeted gifts may, in some circumstances, be efficient: although the sub-sample is too small to yield statistical significance, our point estimates suggest that among experienced employees with low prior wages, hiring three employees with salient gifts was more efficient than hiring four employees with no gift (even though the total cost is the same). This implies that managers may want to consider identifying the types of employees that are most responsive to gifts, or the types of hiring contexts in which gifts are most likely to be salient, and to target gifts based on this information. Indeed, our work builds on Englmaier and Leider (2012) and Kessler (2013), who identify a host of factors that predict when reciprocity in labor markets is likely to occur (e.g., how much a manager benefits from worker effort, and the complexity of the task). An important direction for future research is to further explore the situations in which gifts are likely to be efficient.

2.4.3 Repeated Interactions

By design, our analysis focuses on a situation where an employee is hired for a single task, with no possibility of future employment. One limitation of this is that we do not observe how gifts may impact future interactions. For example, suppose we were to go back to all of our employees in the future and invite them to take a second job with us. It is possible that the treatment to which an employee was assigned might influence their decision to accept that second job, as well as their performance in it. We expect that the effects of gift exchange within a repeated context may depend on a host of factors, including both the initial and future wage, the salience of the gift, and the amount of time that has elapsed. This is an important consideration from a managerial perspective and points to the need for future research.

2.4.4 How Should Employers Structure Wages?

From a labor and behavioral economics perspective, our results reinforce prior findings that gift exchange may be unlikely to explain above-market wages, but they also show that "how" employers pay affects productivity. In this way, our results echo findings in Al-Ubaydli *et al.* (2014), about how employers can signal through compensation decisions. Consistent with Roth (2002)'s description of the "economist as engineer", understanding how to structure wages more effectively–in particular, beyond the teachings of the neoclassical model–presents a promising direction for continued future research.

Chapter 3

Patents as a Spur to Subsequent Innovation: Evidence from Pharmaceuticals[1]

3.1 Introduction

Research dating back at least as far as Nordhaus (1967) has analyzed the role of intellectual property protection in incentivizing innovation. The motivation for such protection tends to be the idea that without property rights, the free market might underinvest – relative to the socially optimal level – in the costly development of new technologies. The tradeoff is that bestowing a firm with market power leads to prices which are statically inefficient, driving a wedge between the allocation which is realized and that which is socially optimal.

Traditional models think of an incumbent's intellectual property protection, such as a

[1]The statements, findings, conclusions, views, and opinions contained and expressed herein are not necessarily those of IMS Health Incorporated or any of its affiliated or subsidiary entities. The statements, findings, conclusions, views, and opinions contained and expressed in this article are based in part on data obtained under license from the following IMS Health Incorporated or affiliate information service(s): IMS National Sales Perspectives, 1992-2013, IMS Health Incorporated. All Rights Reserved. AHFS® Pharmacologic/Therapeutic Classification used with permission. ©2014, the American Society of Health-System Pharmacists, Inc. (ASHP) The Data is a part of the AHFS Drug Information®; ASHP is not responsible for the accuracy of transpositions from the original context.

patent, as inhibiting, or at the minimum dis-incentivizing, entry in the same market by another firm.[2] In this paper, I propose a novel channel through which patents impact innovation, and show that in fact the reverse can be true: an incumbent's patent can actually spur competing innovation. In brief, if patents are narrow, so that an incumbent's patent does not inhibit a differentiated innovator's entry directly, yet it does provide market power to the incumbent by inhibiting perfect imitation, then an innovator's profitability may be tied both to the incumbent's patent protection as well as its to own. This means that, prior to entry, the innovator's incentives to enter may be increasing in the length of incumbent's remaining patent protection. In fact, if fixed costs are high and the innovator's product is not drastically more desirable than the incumbent's, then the innovator's entry decision may be almost exclusively determined by the incumbent's remaining protection.

My focus is within the pharmaceutical sector, a sector which is the single largest in terms of domestic R&D, and a sector in which advances are believed to have contributed substantial improvements to public health.[3] The channel outlined above is straightforward to articulate in this setting: the question is the extent to which the generic version of an existing treatment impedes the development of a novel competing drug. Reports in the popular press, e.g. Parker (2013), suggest that the availability of competing generics are a prominent consideration for firms directing R&D investments toward new drug development.[4]

I analyze the impact of patent characteristics on innovation through precisely this channel. In particular, focusing within pharmacologic classes of closely related drugs, I show that the length of time between the first in class drug's (FIC's) approval and its generic's entry has an economically and statistically significant causal impact on the number of new drugs that

[2]Indeed, in the theoretical literature, the incentive to innovate created by patent protection is usually specifically due to the fact that patents provide protection from from imitation. The theoretical literature on patents, discussed below, is substantial – see, e.g. Nordhaus (1967, 1969, 1972); Scherer (1972); Fudenberg *et al.* (1983); Klemperer (1990); Gilbert and Shapiro (1990); Gallini (1992); Green and Scotchmer (1995). Scotchmer (1991) provides a helpful review.

[3]The NSF reports the pharmaceutical sector makes up nearly 20% of industrial R&D (NSF (2012)); Lichtenberg (2007) estimates that drugs approved in recent years are extraordinarily cost-effective, reducing mortality at a net cost of less than $20,000 per life year saved.

[4]A specific quotation is reproduced in Appendix Section C.1.1.

subsequently come to market in that class. That is, I show that the longer the period of time for which the FIC has market exclusivity versus its generic, the more subsequent innovation occurs in that class.[5]

I begin by analyzing a simple theoretical model that motivates the setting and clarifies how an incumbent's patent impacts the investment decision of a subsequent entrant. As is the case in pharmaceuticals, in the model, patents are narrow and do not inhibit the entry of close but differentiated substitutes. Rather, the role of patents in the model is to provide firms with the agency to set their own prices; after patents expire, firms lose this ability and prices fall. The model shows how the subsequent entrant's profitability, and thus initial investment decision, is affected by the incumbent's patent. Moreover, the model predicts that the size of the effect is moderated by properties of demand and of the products themselves: the effect of the incumbent's patent on the entrant's investment decision is increasing in the price-sensitivity of consumers, and decreasing in the quality difference between the two drugs. Intuitively, the importance of the incumbent's patent for the entrant is increased if demand is particularly price-elastic, while it is mitigated if the new drug is substantially better than the incumbent's.

My empirical analysis, which makes up the remainder of the paper, analyzes the model's predictions. I focus on pharmacologic classes of new molecular entities (NMEs), groupings of pharmaceuticals which are closely related in chemical composition and in physiological effect but which are differentiated at the molecular level. This method of categorizing drugs provides me with a principled procedure with which to identify groups of drugs that are sufficiently similar to be substitutes in the eyes of prescribers and patients yet are sufficiently differentiated so that each NME requires its own costly clinical trials to be marketed and one NME's patent protection does not preclude the entry of the others.[6] I assemble a rich

[5]I refer to innovation which is not first in class as "subsequent" throughout. Although the phrase "follow-on" is perhaps more grammatically convenient, it would be misleading because it is often used to describe cumulative innovation, e.g. as in Sampat and Williams (2014). Similarly, innovation which is called "sequential" usually refers to innovation which builds on its predecessors, as in Bessen and Maskin (2009).

[6]Getting any NME to market is costly because the FDA requires proof of safety and efficacy; a commonly cited figure is that the total capitalized cost of a new drug is on the order of $1B (DiMasi *et al.* (2003)). This

dataset which includes, for NMEs belonging to classes with FIC approval in the period 1987-2011, approval dates, annual sales, market sizes, dates of clinical development, dates of filing and expiration for key patents, and, importantly, whether a generic version has been approved and, if so, when.

The model requires that sales respond to competition within class (else the entry of one drug's generic would not erode the profits of another) and that entry is sequential (else drugs would enter the market simultaneously, leading patent lifetimes to be aligned), so I first look to the data for evidence of these characteristics in the market for pharmaceuticals. Though causal identification of the first factor is difficult, I provide anecdotal and descriptive evidence consistent with the hypothesis that the generic entry of one drug in a class reduces the sales of another. Next, I show that drug development is a highly staggered process: nearly 40% of approved drugs only begin the nearly decade-long process of clinical development *after* the first in class has been approved. Finally, I show that, consistent with the model's main prediction, very few new drugs come to market after the first in class has gone generic.

I turn next to my main empirical strategy, which analyzes how the length of FIC exclusivity impacts subsequent innovation in the same class. While exclusivity typically ends after the expiry of a key patent, the specific timing of that patent's expiry relative to approval (i.e., exclusivity) is determined by many factors. Some factors are idiosyncratic, while others are market-related. For example, as Budish *et al.* (2014) insightfully leverage to study the effect of commercialization lags on clinical trial investments, exclusivities tend to be eroded by the length of clinical trials. If the length of clinical trials is correlated among drugs in the same class, then exclusivities may be as well, confounding estimates generated from regression of entry on FIC exclusivity. My analysis exploits two strategies to handle the potential for endogeneity in market exclusivity.

First, I discuss the main ways in which endogeneity might enter – market size (and

differs from the low cost of bringing a generic to market, which I return to below. In principle, my analysis could analyze drugs which are complements instead of substitutes, but I am without a principled method for categorizing complements.

profitability) and the length of clinical trials – and show that controlling for them directly does little to change the estimates. Furthermore, if market exclusivities were endogenously related to class characteristics, then they should be correlated within class, and this is not born out in the data.

Second, I conduct an instrumental variables analysis which exploits the combined facts that (1) the path of drug development can be hard to predict, yet (2) patent filing is defensive.[7] The first means firms sometimes discover molecules before they know specifically what to do with them, and the second means that firms are incentivized to file patents early. The end result is that firms sometimes file patents on a molecule substantially in advance of its progression to clinical trials (contrasting with a strategy in which firms maximize market exclusivity by filing patents as late as possible).

I use these delays between patent filing and the start of clinical review as an instrument for market exclusivity. I show empirically that they are not related to patent characteristics such as claims and citations, nor are they related to market characteristics such as ex-post annual sales. They are, however, related to factors which are indicative of initial uncertainty and disruptions in the development process, such as the length of time between when the patent was filed and when it is first transferred to another entity (e.g., in a licensing agreement), as well as whether the patent was purchased in a merger transaction. Importantly, they are not correlated within pharmacologic classes. Since a story about the endogeneity of these delays requires that they are somehow related to a class-specific unobservable, that they are uncorrelated within class means a confounding factor would have to be at once idiosyncratic to FIC drugs, unrelated to subsequent entrants, yet also related to total entry. While I cannot rule out such a story, it is not an easy one to tell.

The results, across both IV and non-IV specifications with and without controls, consistently show that a one year increase in first in class exclusivity yields a 25-30% increase in subsequent entry, or about 0.2 drugs. This is a large effect: it implies that a one standard de-

[7]A well known example of the first phenomenon is the story of Viagra's discovery. The Viagra molecule was originally discovered as a treatment for angina, but in the course of clinical review (for angina) it was discovered that men taking the drug found themselves unexpectedly erect.

viation increase in first in class exclusivity *doubles* subsequent entry. The appendix presents a a battery of robustness tests which show that the results are not particularly sensitive to sample restrictions, to specification choices, and that they survive a placebo test where I look at the effect of last in class exclusivity on prior entry.

Finally, I analyze the model's predictions that the effect of FIC exclusivity on subsequent entry should be moderated both by the price elasticity of demand as well as by the magnitude of quality improvements. I proxy for price elasticity using both patient incomes as well as a dummy for whether a drug treats a chronic condition and find, consistent with the model's prediction, that the effect is strongest among price-elastic populations. Next, I proxy for advances in quality by asking whether new drugs receive priority review status from the FDA, a status granted to molecules found to be exceptionally promising in clinical trials.[8] Again, consistent with the model's prediction, I find that FIC exclusivity has a significantly stronger impact on the number of subsequent entrants that do not receive priority review than on those that do (and I cannot distinguish the latter effect from zero).

Prior theoretical work on patents, such as the seminar papers on optimal patent characteristics by Klemperer (1990) and Gilbert and Shapiro (1990), does not allow for the effect I describe. The models vary but a common theme is that firm profits are primitives; i.e., pricing and profits are not endogenously determined in competitive equilibrium as a function of market structure. This set up rules out the potential for one drug's patent protection to affect another's entry decision through its pricing.

The empirical literature has similarly overlooked this effect. Indeed, the majority of empirical work on patents has focused on patenting as an outcome of the innovative process.[9] Taking the reverse approach, I contribute to an extraordinarily sparse body of empirical research that considers – and finds evidence for – how the characteristics of patents manifest themselves in new innovations. Evidence provided by Sakakibara and

[8]A priority review designation accelerates approval decisions – the FDA goal is to complete a priority review in 6 months. However, it does not affect the level of evidence as to a drug's safety and efficacy required for approval.

[9]Griliches (1998) provides a review and a discussion of the evidence for and against this design.

Branstetter (2001) and Lerner (2002) actually suggests patent laws have only limited impacts on innovation. With respect to research that finds evidence of the opposite, I am aware only of Moser (2005), who shows that 19th century patent laws influenced the geographic distribution of inventors.

Of course, the idea that competition in itself might affect new product introduction is not new. Indeed, it has received considerable attention specifically in the pharmaceutical setting by researchers trying to explain why the number of new drugs approved annually in the United States has fallen since 1996 despite continued growth in R&D expenditure.[10] Scannell *et al.* (2012) describe the competition factor in pharmaceutical development as the "better than the Beatles" problem: new drugs are only worth developing if they are more effective than past successes (in apparent contrast with the situation in the music industry). And other empirical work supports the link between competition and investment: Civan and Maloney (2009) show that that early stage investment is positively correlated with the price of approved drugs in the same therapeutic area, and concurrent work by Branstetter *et al.* (2014) shows generic entry in a therapeutic area is associated with declines in early stage investment. This research supports the thrust of my findings. However, to my knowledge, no existing research has articulated and analyzed the fundamental link between the characteristics of patent protection and these downstream effects.

The remainder of the paper proceeds as follows: I open in Section 3.2 with a model to formalize the effect that an incumbent's patent protection has on subsequent entry. Section 3.3 introduces the data and Section 3.4 provides descriptive analysis and background on the pharmaceutical industry. In Section 3.5, I describe my empirical strategy before presenting estimates of the impact of an incumbent's patent protection on subsequent entry. Section 3.6 analyzes how my estimates are related to class characteristics and Section 3.7 concludes.

[10]This has been called a "crisis" – see Pammolli *et al.* (2011) and the references cited therein. Cockburn (2007) also provides an elegant analysis and discussion of some of the possible forces at work. The perplexing slow down in productivity is sometimes called "Eroom's Law", contrasting with the well-known Moore's law in the semiconductor industry.

3.2 A Model

In this section, I develop a simple model to show how an incumbent's patent protection affects the entry incentives of a subsequent innovator. I first present the set up and then discuss its assumptions and its implications.

3.2.1 Preliminaries

The model has two periods $t = 0, 1$ and two players. The incumbent A is present in the market in both periods. At $t = 0$, the potential entrant B must decide whether or not to pay a fixed development cost F in order to enter at $t = 1$. At $t = 1$, prices are set in static Bertrand-Nash equilibrium and profits are realized.

There is no uncertainty and no discounting, so B's decision at $t = 0$ is simple: B enters if the value of entry, i.e., profit less development cost, is at least 0. Formally,

$$V_B^{entry} = \pi_B(p_A, p_B) - F \geq 0 \Leftrightarrow \text{B enters.} \tag{3.1}$$

If B enters, it benefits from patent protection at $t = 1$, and is able to set its price p_B at whatever level it chooses.

However, A is not necessarily on patent at $t = 1$. If A does have patent protection, then A chooses p_A optimally. If A does not have patent protection, then generic competition exogenously pushes A's price down, though it can be no lower than unit production costs c, which are symmetric for both A and B. Altogether, this means that A's price at $t = 1$ is somewhere on the interval between c and its optimal duopoly price (in the case that B enters) or monopoly price (in the case that B does not enter). For simplicity, I do not model the process of generic entry and my analysis instead takes the price of A as given; I think of the level of A's price as directly influenced by its patent protection. My focus, then, is on the value that B accords with entry and specifically how that value is influenced by A's price.

A's patent protection is modeled as sufficiently narrow to not directly impact B's ability to enter. This captures the reality in pharmaceutical development, where patents are narrow: pharmaceutical patents generally only provide protection against identical substitutes (e.g., the same molecule) and not differentiated substitutes (e.g., molecules that share features and treat the same condition, but are not identical). I do not model generic entry explicitly for the sake of simplicity, but my set up is motivated by the fact that once patents have expired, it is not costly for generic competitors to enter the market and such entry tends to happen rapidly, reducing prices paid on the order of 90% in the year or two following patent expiry.[11] The channel through which patent protection impacts competition in my model – by not directly inhibiting entry, but rather by influencing prices – distinguishes my analysis from prior research, such as Klemperer (1990) and Gilbert and Shapiro (1990).[12]

I model the demand side through a discrete choice framework. A unit mass of consumers must pick one drug to consume, and consumers differ according to their price sensitivities. Specifically, consumer i receives utility from drug j according to its quality level δ_j and price p_j:

$$u_{i,j} = \delta_j - \frac{1}{\alpha_i} p_j. \tag{3.2}$$

Price sensitivities are parametrized the α_i's, which are distributed exponentially in the population with mean λ. The parameter $\lambda > 0$ captures the dispersion of price sensitivity in the population.[13] If λ is higher, then consumers are on average less price-sensitive. The choice of exponential distribution is motivated by the skewness of real incomes. For simplicity, I assume that B's quality is exogenously higher than A's, i.e. that $\delta_B > \delta_A$, and I write the difference in qualities $\Delta = \delta_B - \delta_A$. This demand framework captures the idea that drugs are differentiated and that consumers differ in how they value quality relative to price.

To summarize, at $t = 0$, the potential entrant B must decide whether or not to pay the

---

[11] Reiffen and Ward (2005) estimate the fixed cost of generic entry is roughly \$300,000. This is very small relative to the \$1B cost of drug development cited earlier. Berndt (2002) provides a review of the evidence on generic entry.

[12] I am not aware of a paper that characterizes optimal patent characteristics when pricing is endogenous (i.e., where patent expiration of one product impacts the profitability of another). However, that patents do not preclude entry in my model means it shares features with Gallini (1992), where inventors can pay a fee to "innovate-around" patent-protected incumbents.

[13] An alternative interpretation is that all consumers have the same price sensitivity but that quality sensitivity varies.

fixed cost F to enter at $t = 1$. If B enters, it benefits from patent protection, but it competes with the incumbent A which is priced between c and its optimal level, depending on its patent protection and the ensuing extent of generic competition. The model has three parameters: p_A is A's price and reflects A's patent protection, λ reflects the extent to which consumers are insensitive to price, and Δ reflects how much higher B's quality is than A's.

3.2.2 Entry Incentives and Implications

I first solve for B's best-response pricing and profits before analyzing the properties of B's value function. Looking to demand, consumer i prefers B to A if $\delta_B - \frac{1}{\alpha_i} p_B > \delta_A - \frac{1}{\alpha_i} p_A$, i.e. $\alpha_i > \frac{p_B - p_A}{\Delta}$. This means aggregate demand for drug B is

$$D_B(p_A, p_B) = \int_{\frac{p_B - p_A}{\Delta}}^{\infty} \frac{1}{\lambda} \exp\left(-\frac{\alpha}{\lambda}\right) d\alpha = \exp\left(-\frac{p_B - p_A}{\lambda\Delta}\right). \tag{3.3}$$

B's first order condition then yields:

$$\begin{aligned} p_B^{BR} &= \lambda\Delta + c \\ \Rightarrow V_B^{entry} &= (\lambda\Delta + c) \exp\left(\frac{p_A - c}{\lambda\Delta} - 1\right) - F. \end{aligned} \tag{3.4}$$

This simple models yields the following three intuitive predictions.

(1) B's entry value is increasing in A's price:

$$\frac{\partial V_B^{entry}}{\partial p_A} > 0. \tag{3.5}$$

The first comparative static is quite natural: the higher A's prices, the higher B's profits. This implies that B's incentives to enter are higher when A does not face generic competition. The majority of my empirical results, presented in the following sections, focus precisely on analyzing how subsequent entry is affected by the incumbent's patent protection.

The second and third comparative statics analyze the extent to which price sensitivity (parametrized by λ) and quality differences (parametrized by Δ) affect B's entry value.

(2) The effect of A's price on B's entry value is decreasing in λ:

$$\frac{\partial^2 V_B^{entry}}{\partial p_A \partial \lambda} < 0. \tag{3.6}$$

The second comparative static shows that the influence of A's price on B's profits is moderated by population price sensitivities. As price sensitivities become more dispersed (i.e., as λ grows), the influence of A's price on B's profits decreases. This is intuitive: the more price-inelastic consumers are, the more willing they are to pay for a higher quality product, which lessens the impact of A's price on B's profitability. This implies that an incumbent's patent protection has less impact on subsequent entry when consumers are less responsive to pricing.

(3) The effect of A's price on B's entry value is decreasing in Δ:

$$\frac{\partial^2 V_B^{entry}}{\partial p_A \partial \Delta} < 0. \tag{3.7}$$

The third comparative static shows an analogous pattern is true for quality differences: the larger the quality gap between A and B (as measured by Δ), the less the influence of A's price on B's profits. This too is intuitive – drugs that are significant clinical advances should not be highly impacted by an incumbent's patent protection because the difference in quality increases demand for the new drug. I analyze the empirical implications of these last two predictions in the penultimate section of this paper.

With this expository framework and its predictions in mind, I next discuss my data. My empirical analysis follows.

3.3 Data

I analyze how the number of drugs approved in a pharmacologic class is influenced by the first in class drug's market exclusivity. In the following, I briefly explain each of the main data sources and its contribution, before providing an overview of the sample's composition and some summary statistics. The interested reader may turn to Appendix Section C.2 for an extended discussion of the data.

I begin with data from the FDA (see Lanthier *et al.* (2013)) that classifies all NMEs approved in the United States between 1987 and 2011 into pharmacologic classes.[14] For all drugs, I then obtain the date of first approval and first generic entry (if generic entry has occurred) from the FDA's Drugs@FDA database and the FDA's Orange Book, respectively. Only 38% of drugs in the sample have gone generic as of the end of 2013, so I supplement realized exclusivities with a measure of expected exclusivity. I do this by identifying, for each drug, the single patent most likely to inhibit generic entry; I then proxy for the date of expected generic entry using this patent's expiry date.[15] Altogether, of the 293 drugs in my sample which belong to 156 classes, I have a measure of exclusivity for 237, of which 127 are first in class. Next, I collect data on the timing of clinical trials, market sizes, and sales:

(1) Timing of Clinical Trials: Details on when pharmaceuticals begin development is generally proprietary, but a prerequisite for a patent extension to make up time lost in clinical trials is that firms publicly certify the beginning and end of clinical development.[16] I collect these data from the Federal Register.

(2) Market Size: My main measure of market size is inspired by Acemoglu and Linn (2004)'s analysis of the effect of market size on pharmaceutical development. In particular, using data from the Medical Expenditure Panel Survey, I identify the total annual US prevalence of each class's primary ICD-9 condition code. I call this prevalence market size.[17]

---

[14]I focus exclusively on non-biologic drugs approved in the U.S., and in order to ensure data completeness, I exclude classes in which the FIC was approved prior to 1987. The U.S. is not the only market for pharmaceuticals but it is by far the largest, accounting for over 50% of global sales in 2000 (WHO (2004)).

[15]Additional details are given in the appendix, but my methodology for identifying this patent is inspired by Hemphill and Sampat (2012) and exploits the fact that the Hatch-Waxman Act of 1984 allows firms to extend a single patent to make up for up to half the time spent in clinical trials; I use the expiry date of the extended patent. Empirically, the expiration of extended patents is highly predictive of generic entry: for the 81 drugs in my sample that have gone generic and have a patent for which an application for extension was filed, the R^2 from a regression of realized exclusivity on exclusivity predicted by the extended patent is 0.95.

[16]The term *clinical* denotes development in humans, as opposed to laboratory or computational-based development. These data were first analyzed by Keyhani *et al.* (2006).

[17]As I note in the Appendix, I have experimented with other measures of market size and find they have little influence on my results. Budish *et al.* (2014) similarly find their results are robust to conditioning on different measures of market size.

(3) Sales: To capture the price dimension of profitability, I use data from IMS Health's National Sales Perspective which detail total annual U.S. sales for the 1000 top selling drugs each year from 1992-2012.

Before turning to the summary statistics, I note that the categorization of "first in class" I use differs slightly from the FDA's formal categorization. This is motivated by the fact that, in some cases, multiple drugs are approved very early on and sometimes one of the early entrants is considerably more effective than the others. In these cases, the technical definition of first in class does not coincide with the market's focal early entrant. A poignant example is given by the class of atypical antipsychotics: The first entrant, Clozaril, was approved in 1989 but Clozaril's mass-market appeal was severely limited because it produces a potentially fatal side effect in 1-2% of patients (Alvir *et al.* (1993)). The second atypical antipsychotic, Risperdal, had no such severe side effects and was granted priority review status; after its approval in 1993, Risperdal rapidly became the standard of care, with annual sales that were nearly 50% higher than Clozaril's just one year after approval (Finkel (2012)). Though Clozaril was technically the FIC, it was largely irrelevant (as a competitor) for future entrants. Thus, in cases in which multiple drugs entered early on in short succession, my strategy is to focus on what I call the *effective* FIC, which is the authentic FIC unless the authentic FIC is immediately followed by a priority review drug, in which case it is that drug. The effective FIC differs from the authentic FIC in 32 cases and is approved an average of 3.0 years after the authentic FIC. I probe the sensitivity of my results to this definition in Appendix Section C.5.1; it should come as no surprise that my results are strongest for classes in which the definitions of authentic and effective FIC coincide. For brevity, I refer to the effective FIC simply as the FIC.

To summarize, the sample consists of all non-biologic drugs that belong to classes with first approval in the period 1987-2011. Altogether, the sample consists of 156 classes representing 293 drugs. In practice, since my data on market exclusivities and the dates marking the beginning of clinical development (used to compute the instrument) are not complete, I restrict my analysis to the 111 classes for which I have that data, representing

252 drugs (although this too I probe in Appendix Section C.5.1).

I present summary statistics on the sample in Table 3.1. The table consists of three panels which respectively present statistics by class (Panel A), then by drug for FIC drugs only (Panel B), and finally by drug but for all drugs (Panel C). Looking first to the by-class figures in Panel A, the first row shows that class sizes are generally small but that they are skewed – the mean number of entrants subsequent to the FIC is 0.73 (and the median 0) while the maximum is 7. Mean clinical development times, in the second row, have a mean of 8.31 years and standard deviation of 3.81 years. Mirroring class size, market size and maximum annual revenue are highly skewed: the mean market size is about 8 million, while the largest is 45.5 million.[18] The by-drug figures in Panels B and C show similar patterns; mean exclusivities in both panels are just below 12 years and average times in clinical development are about 8 years. I forgo discussion of the length of time between patent filing and the beginning of clinical development (the instrument for market exclusivity) until Section 3.5.2.

3.4 Facts About Competition and Entry

This section provides descriptive analysis on entry in the pharmaceutical industry. I first probe two of the model's main assumptions: that generics in class impact the sales of other branded drugs, and that pharmaceutical development is sequential and not characterized by a single race to market. I then show that, consistent with the model's first prediction, subsequent entry is substantially less likely to occur after the FIC has gone generic.

3.4.1 Generic Competition in Class

Although others have documented the fact that competition within a class of drugs – measured, e.g., by the number of distinct molecules – impacts sales (see, e.g. Lichtenberg and Philipson (2002)), to my knowledge there is no evidence on the specific impact of generic competition in class on profitability. However, that the availability of one drug's

[18] There are fewer than 111 observations in the fifth row because my revenue data only contains revenues for one or more drugs in 88 of the 111 classes.

Table 3.1: *Summary Statistics*

	Mean	Median	Std Dev	Min	Max	N
	Panel A: By Class					
	Mean	Median	Std Dev	Min	Max	N
Number of Subsequent Entrants in Class	0.73	0	1.39	0	7	111
Mean Time in Clinical Development (yrs)	8.31	7.53	3.81	0.78	26.97	111
Market Size (1,000,000s)	7.88	3.22	10.51	0.08	45.55	111
Maximum Annual Revenue ($1,000,000s)	1,000.12	364.96	1,665.66	26.84	8,558.62	88
	Panel B: By Drug, First in Class Only					
	Mean	Median	Std Dev	Min	Max	N
Market Exclusivity (yrs)	11.81	12.65	3.16	5.00	18.32	111
Time in Clinical Development (yrs)	7.98	6.98	4.06	0.78	26.97	111
Patent Filing to Clinical Development (yrs)	5.25	4.43	3.87	0.00	20.39	111
Maximum Annual Revenue ($1,000,000s)	708.79	245.35	1,227.84	24.51	7,064.03	85
	Panel C: By Drug, All					
	Mean	Median	Std Dev	Min	Max	N
Market Exclusivity (yrs)	11.61	12.66	3.25	5.00	18.32	196
Time in Clinical Development (yrs)	8.21	7.24	3.88	0.78	26.97	192
Patent Filing to Clinical Development (yrs)	4.95	4.35	3.62	0.00	20.39	189
Maximum Annual Revenue ($1,000,000s)	841.37	287.82	1,418.35	13.80	8,558.62	166

Notes: This table presents summary statistics for the main sample of 111 drug classes as described in Section 3.3. Panel A presents statistics by drug class, Panel B by drug for FIC drugs only, and finally Panel C presents statistics for all drugs. There are fewer than 111 observations in the rows describing maximum annual revenues because I only observe revenues for 88 classes and 85 drugs of the 111 classes analyzed. Panel C also faces sample restrictions: of the 252 drugs in the sample, I only observe market exclusivity for 196, the timing of clinical development for 192, the value of the instrument for 189 observations, and finally revenues for 166.

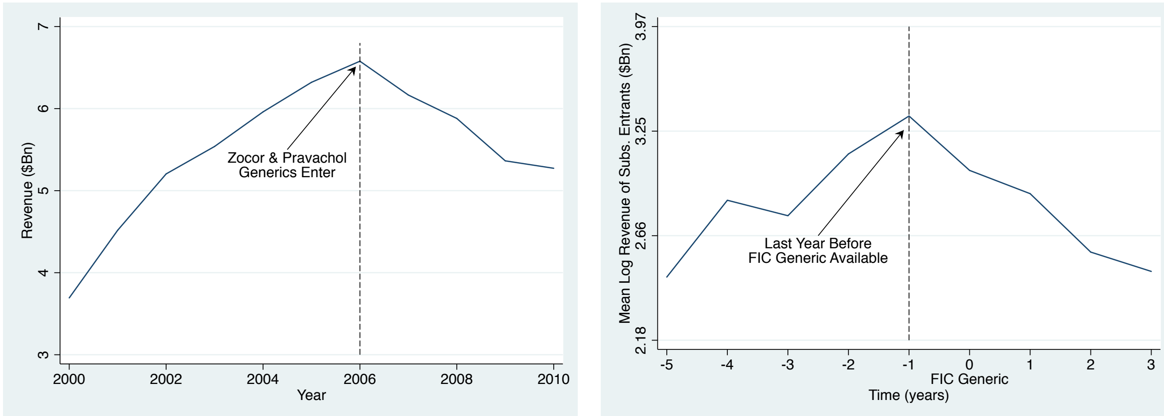
generic negatively affects the sales of different drugs in the same class is crucial for my argument; if these drugs are not close substitutes, the mechanism exposited by the model could not be rationalized by the data.

Causally estimating the impact of one drug's generic on the sales of others is difficult. Although exploiting an event-study framework around the time of generic entry might seem appealing at first glance, it is not clear that such a framework is appropriate because the timing of generic entry is not random.[19] Thus, my strategy in this section is to provide descriptive and anecdotal evidence.

I begin with a case study. Lipitor was the fifth drug approved in the class of statin drugs (technically, HMG-CoA reductase inhibitors), which primarily treat conditions associated with cardiovascular disease. Lipitor came to market in 1996 and quickly achieved blockbuster status – By the time it went generic in November of 2011, Lipitor had sold more than any

[19]Even on a relatively short timescale (e.g., one year in advance), generic entry may be anticipated in important ways (e.g., through changes in marketing strategy of the type analyzed by Ellison and Ellison (2011)).

Figure 3.1: *Generic Entry and Sales of Subsequent Entrants*



(a) *Sales of Lipitor*

(b) *Sales of All Subsequent Entrants*

Notes: These figures respectively show plots of the annual sales over time. Panel A plots annual US sales for Lipitor, where the data come from IMS Health's publicly available Top-Line Market Data. Panel B shows the mean of log annual sales of subsequent entrants relative to the time of the FIC's generic entry. The sample is as described in Section 3.3.

other medicine, with global sales of $125B over 14.5 years (Associated Press (2011)). However, Lipitor's sales did not grow monotonically over the period of its branded life; instead, Lipitor sales peaked in 2006, when two of its closest competitors, the statins Zocor and Pravachol, went generic. The press expected this to be a turning point for Lipitor's sales (see Appendix Section C.1.2), and indeed it was, as visualized in Figure 3.1a.[20] Though it may not be causal, the relationship between Lipitor's sales and the generic entry of its competitors is striking, with sales falling 25% from peak from 2006-2010. Lipitor was the best-selling drug of all time; presumably if one drug was to be able to defend itself against its competitors' generics, it would have been Lipitor.

An analogous relationship is present across drug classes. In Figure 3.1b, I plot the log of annual sales against the timing of first in class generic entry (so that year 0 denotes the year of FIC generic entry), averaged across classes.[21] Here, we see a similar downward trend that starts when the FIC goes generic. While this is not a causal test, and other events affect the

[20]The source for this sales data is IMS Health's publicly available Top-Line Market Data, available at `www.drugs.com`.

[21]I take logs before averaging because (as described in Section 3.3) drug sales show substantial skew.

results (e.g., some subsequent entrants themselves go generic three years after the FIC), the sales data do suggest that generic competition in a class impacts the sales of other drugs.

3.4.2 The Timing of Development Decisions

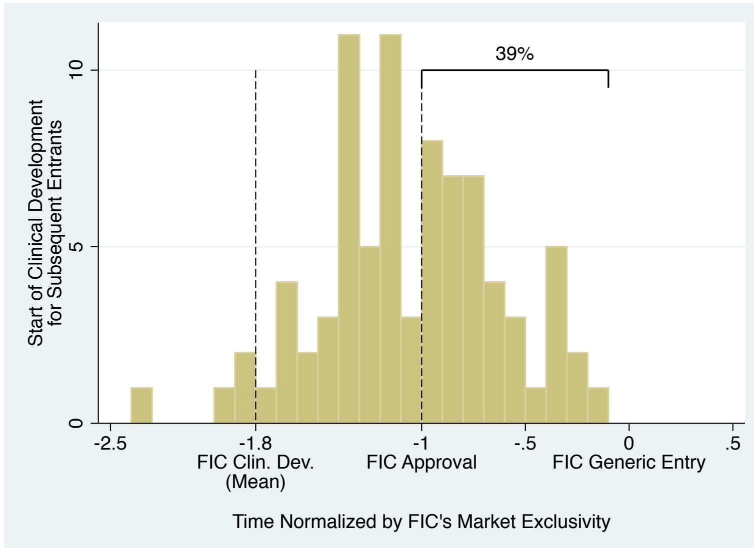
The model analyzed earlier assumed that the second entrant's entry decision was made after the first entrant had sunk its own fixed costs and entered – indeed, the model even considered the possibility that the first entrant had gone generic by the time the second entrant was approved. That development decisions are sometimes sequential and not simultaneous is important for the theory advanced in this paper; otherwise, it would be implausible that one drug's generic would deter the entry of another. My data, which detail the dates that drugs entered clinical trials for the first time, speaks to this point.

Figure 3.2 presents a histogram which examines the timing of the start of clinical development for subsequent entrants. The horizontal axis is normalized so that development start times are relative to the FIC's approval (at time = -1) and FIC generic entry (at time = 0). The chart shows a large proportion – nearly 40% – of subsequent entrants do not begin clinical development until the FIC is approved. This is important because although the start date of clinical development is public after approval for approved drugs (because of the application for patent extension), the extent to which pre-approval development decisions are public information is not clear; in the period before FIC approval, market participants may have believed they were all pursuing approval in tandem.[22] However, approval is highly publicized so firms starting trials then could not have mistakenly believed they could be first to market.

The figure also shows that no new drugs begin development after the FIC has gone generic (there is no mass to the right of time = 0). This is consistent with the model's prediction that the incumbent's generic reduces incentives for new firms to enter, as well

[22]Development decisions are often described as competitive intelligence, although at some level once a human subject pool has been recruited and used in testing it is harder to keep secret, especially since investors may require notification. Databases like Adis R&D Insight, Pharmaprojects, and Thomson Reuters Cortellis purport to collect exactly this type of information, though the extent to which their data are collected contemporaneously versus after the fact it is not clear, nor is the extent to which they are complete.

Figure 3.2: *Timing of Clinical Development for Subsequent Entrants*



Notes: This figure shows a histogram of the timing of the start of clinical development for subsequent entrants relative to FIC generic entry, where the x-axis is denominated by the FIC's market exclusivity so that -1 corresponds to FIC approval and 0 corresponds to FIC generic entry. The start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. The sample is as described in Section 3.3.

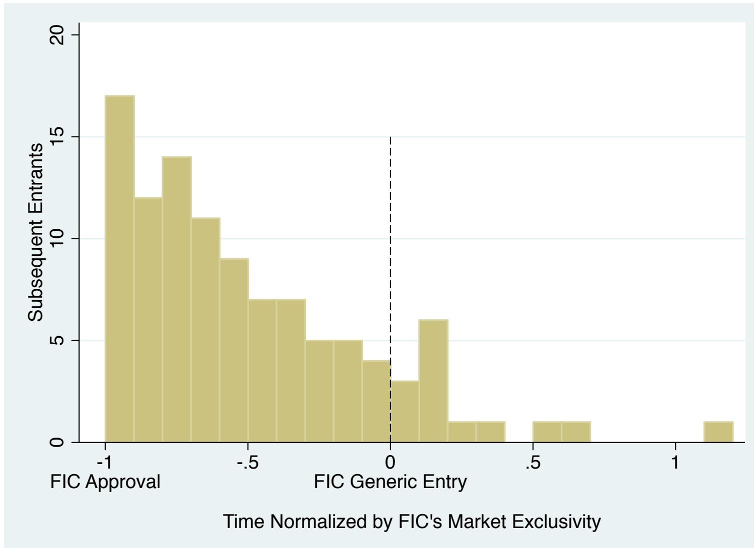
as with what Kevin Sharer, the former CEO of Amgen, shared with me: "Once we enter generic land, there's no incentive [to develop new drugs]" (quoted with permission, April 15, 2014).

3.4.3 The Timing of Subsequent Entry

If new drugs tend to be less profitable once the FIC has gone generic, then the long run profitability of any given drug should be lower the later it enters (as the clock tick's on the FIC's exclusivity). All else equal, this should create a downward gradient in new drug approvals over time, with few approvals after the FIC has lost exclusivity.[23] I analyze this prediction in Figure 3.3, which shows a histogram of the timing of non-FIC approvals relative to the FIC's generic entry, where as before the horizontal axis is normalized within

[23]Though I do not do so because it complicates the exposition slightly, it is trivial to extend the model so that B has patent protection for T periods, t of which are periods of duopoly (while both A and B are on patent), and $T - t$ periods in which B competes with A's generic. Since B's profits are strictly greater when A is on patent, B's entry incentives are increasing in t.

Figure 3.3: *Timing of Subsequent Entry*



Notes: This figure shows a histogram of the timing of subsequent entry relative to FIC generic entry, where the x-axis is denominated by the FIC's market exclusivity so that -1 corresponds to FIC approval and 0 corresponds to FIC generic entry. The sample is as described in Section 3.3. The relationship visualized here is analyzed empirically in Appendix Section C.4.

classes so that time = -1 corresponds to when the FIC enters and time = 0 corresponds to when the FIC goes generic. The plot shows that entry timing is negatively correlated with FIC exclusivity, with nearly two thirds of non-FIC drug approvals occurring before just half of the FIC's branded lifetime has expired. The right half of the histogram is almost empty: very few new drugs come to market in classes in which the FIC has gone generic. In Appendix Section C.4, I analyze this relationship in regression form and find that the correlation between the number of new entrants and remaining FIC exclusivity is statistically significant.[24]

[24] I do not have an instrument for the time remaining on the FIC's patent, so I do not interpret this analysis causally. The ideal instrument for this context would be something that unpredictably and randomly invalidates a branded incumbent's patents. Such a shock would enable an event-study analysis where the strategy would be to ask if entry is reduced after an incumbent's patents are randomly invalidated. One possibility for an instrument which has been exploited in prior research analyzing the financial impact of generic entry (see Panattoni (2011) and Jacobo-Rubio (2014)) is the outcomes of Paragraph IV lawsuits, which are sometimes precursors to generic entry. However, the sample of such lawsuits is small and neither the timing nor the outcomes of Paragraph IV lawsuits appear to be random – they tend to occur shortly after a drug's active ingredient patent has expired and often result in generic entry.

3.5 Main Results

I have presented evidence that drugs in the same class appear to be substitutes, that entry decisions are made sequentially, and that subsequent entry rarely occurs once the FIC has gone generic. These findings lay the foundation for this paper's thesis, which is that longer FIC exclusivity increases subsequent entry. This section presents my main results. I first describe my empirical framework before discussing potential confounders, my instrument, and the results.

3.5.1 FIC Exclusivity and Subsequent Entry

My empirical strategy examines how the exclusivity of FIC drugs impacts subsequent innovation. Formally, with j indexing drug classes, I estimate regressions of the form,

$$SubsEntrants_j = \alpha + \beta FICExcl_j + \gamma' X_j + \varepsilon_j, \tag{3.8}$$

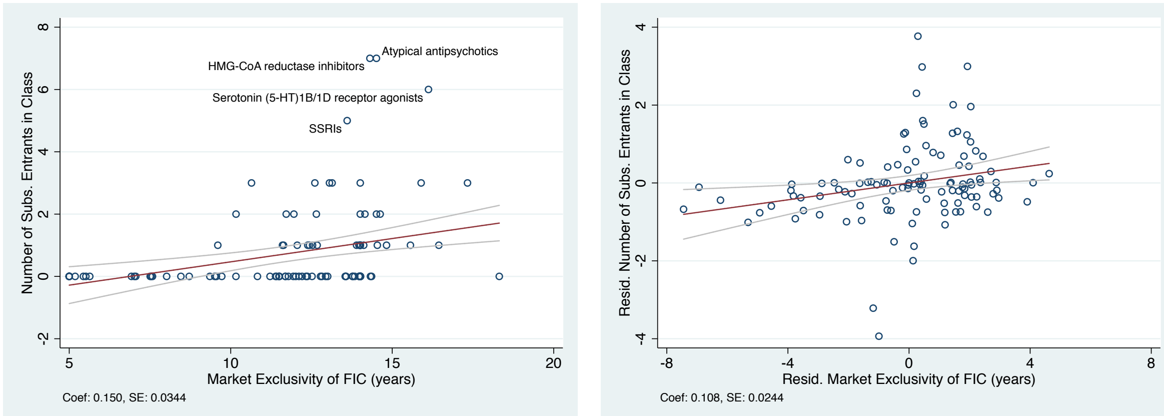
where $SubsEntrants_j$ is the number of entrants subsequent to the FIC, $FICExcl_j$ is the FIC's market exclusivity, and X_j is a vector of controls. Given the count nature of drug launches, I estimate quasi-maximum likelihood (QML) Poisson models, and since my data are as of 2011 and some classes started earlier than others, I include FIC approval year fixed effects throughout.[25]

Figure 3.4a illustrates the raw relationship between subsequent entry and FIC exclusivity without controls and shows that there is indeed a strong positive correlation. However, the slope of that graph should not necessarily be interpreted as causal for reasons explained below.

Understanding specifically what determine a drug's market exclusivity is helpful for understanding where biases could enter. Figure 3.5 visualizes the typical sequence of events

[25] Some classes have had very few years to allow for subsequent entry, so in Appendix Section C.5.4 I show that restricting the sample to only those classes that have had at least 10 years after FIC entry does not affect the results. I also note here that including fixed effects in a non-linear model typically leads to an incidental parameters problem. In Poisson models, this can be corrected using Hausman *et al.* (1984) transformation, which I apply throughout.

Figure 3.4: *Subsequent Entry and First in Class Exclusivity*



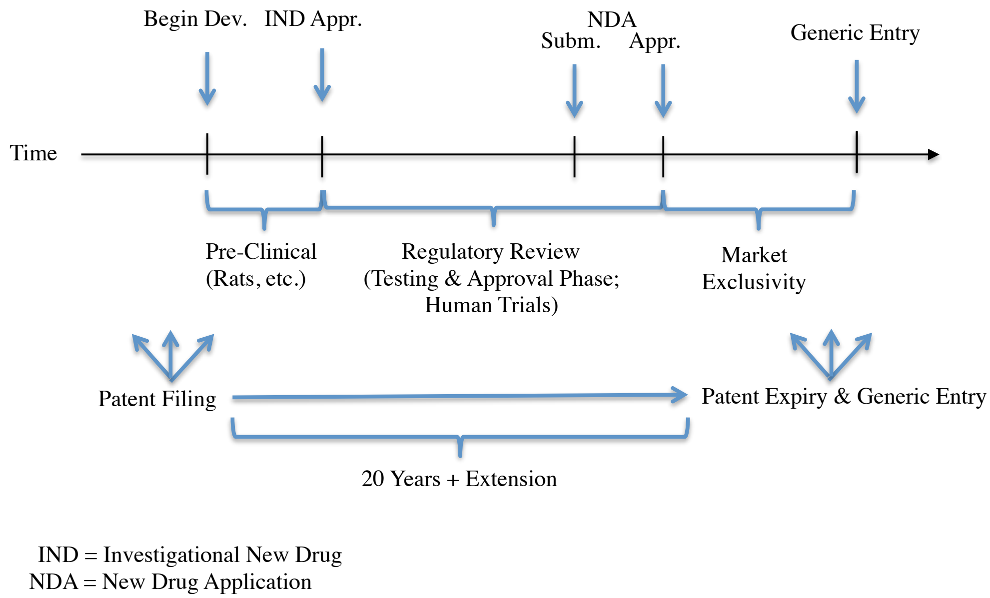
(a) *Unresidualized* **(b)** *Residualized*

Notes: Panel A figure shows a plot of the number of entrants subsequent to the FIC drug against the FIC drug's market exclusivity. Classes with at least 5 subsequent entrants are labeled. Panel B shows a plot of the residualized number of entrants subsequent to the FIC against the FIC's residualized market exclusivity. Residuals are conditional on year of FIC approval fixed effects, mean development time, and market size, and the specification of the controls and sample are as described in Section 3.3. The slope of the line of best-fit and its associated robust standard error are presented in the bottom left of the figure. The sample is as described in Section 3.3.

in drug development for a drug that progresses from discovery all the way to patent expiry. A prospective drug starts in pre-clinical development and testing, which usually takes place in the lab. After an Investigational New Drug (IND) Application is approved by the FDA, the drug progresses into the period of regulatory review characterized by human clinical trials. If trials are viewed as successful in proving safety and efficacy, an NDA is submitted to the FDA. In the case of approval, the drug is marketed exclusively by its developer until the first generic enters the market.[26] Federal law guarantees that NMEs are protected from generic entry for a minimum of five years after a drug is first approved, but in general entry is restricted for longer because of patent protection. As a legal matter, patents pertaining to active ingredients generally need to be filed before the beginning of clinical trials – once the drug has been administered to humans, its novelty (required for enforceable patenting) can be questioned.

[26]Though I focus only on approved drugs, the likelihood that any individual drug candidate progresses all the way from the start of clinical review to approval is not high – on average, Hay *et al.* (2011) estimate it is less than 10%.

Figure 3.5: *Timeline of Drug Development*



Notes: This figure shows the typical timeline of drug development. In particular, it depicts how market exclusivity, which is defined as the time between NDA approval and generic entry, is impacted by the timing of patent filing.

Altogether, there are three main factors that influence the length of exclusivity. The first factor pertains to the timing of approval, in the middle of Figure 3.5: As Budish *et al.* (2014) point out, since patent lengths are fixed at 20 years (plus applicable extensions, which tend to be minor), an important component of market exclusivity is simply the time a drug spends in trials ("commercialization lag", in their verbiage). From the start of clinical trials, the clock is ticking toward patent expiry, so long trials tend to reduce exclusivity. Thus, to the extent that trial lengths are correlated within classes, this could bias β in the estimation of Equation (3.8).[27] This means it is important that I control for the length of clinical review.

The second factor influencing exclusivity is the timing of generic entry, which is, in principle, manipulable. This factor pertains to the end point on the right-hand side of Figure 3.5. For example, the innovator may be able to file new auxiliary (yet enforceable)

[27]For example, if trials in class j are short, then drugs in class j all have long exclusivities, so that $FICExcl_j$ is larger. Keeping with the example, long exclusivities in class j presumably mean it is more profitable, so if fixed costs are similar across classes then total entry in j is higher, meaning $SubsEntrants_j$ is larger. Then the positive relationship I observe between $FICExcl_j$ and $SubsEntrants_j$ could be due to the fact that classes with shorter trials are more attractive.

patents, e.g., on the molecule's salt forms or the manufacturing process, which make it difficult for generics to enter even after the AI patent has expired. In some cases, the firm selling a branded drug might even contract with generic entrants to stay out of the market in what is called a "reverse payment settlement."[28] Since the extent of this confounder is likely related to profitability, I control for market sizes and additionally present an analysis which controls for sales.

Third and finally, the timing of the AI patent's filing relative to the start of clinical development is not always the same. Variation in the filing date affects the left-hand side of Figure 3.5. As I explain below, scientific uncertainty leads the gap between patent filing and the start of trials to exhibit substantial variation which is orthogonal to other class characteristics. This provides the basis for my instrument.

However, before proceeding to the IV analysis, I present results which control for the first two factors: time in trials and market size. Figure 3.4b shows that conditioning on these controls, as well as fixed effects for the year of FIC approval (since some classes are older than others), does not remove the positive relationship between subsequent enter and first in class exclusivity.[29]

As an additional test for endogeneity in market exclusivity, I ask whether market exclusivities are positively related within classes. The motivation underlying this test is that if some other factor confounds the estimated relationship between $FICExcl_j$ and $SubsEntrants_j$, then presumably that factor would also influence the exclusivity of other drugs in class j. This only be operationalized for classes in which there is at least one subsequent entrant, but it still provides insight into how much exclusivity is determined by factors that are idiosyncratic to a drug versus common to a class. I implement this test by estimating an OLS model where the outcome is the exclusivity of drug k and the

[28]Hemphill (2007, 2009) provide extensive description and analysis of reverse payment settlements.

[29]The relationship is slightly weakened in the linear model presented in the figure, but it is actually strengthened in the corresponding Poisson estimates which are presented, with controls added sequentially, in the first three columns of Table 3.3. However, I forgo an extended discussion of these estimates until I take up all of my main results in Section 3.5.3.

independent variables are the exclusivity of drug l, as well as my baseline controls for market size, time in trials, and year of FIC approval fixed effects, for all pairwise combinations of drugs k and l belonging to the same class. The coefficient estimate on the exclusivity of drug l is -0.0089 with standard error 0.0366, where the standard error is robust and clustered by class.[30] Overall, this suggests that after conditioning on time in development and market size, the potential for endogeneity in naive estimation of Equation (3.8) is likely to be limited. Nevertheless, I pursue clean identification in the IV analysis below.

3.5.2 IV Strategy

I have shown that there is a strong relationship between FIC exclusivity and subsequent entry, and that this relationship is not diminished by conditioning on time in trials and market size. I have also shown that exclusivity is not significantly correlated among drugs within the same class; this means an endogeneity story would require some external factor which impacts both subsequent entry and FIC exclusivity but not the exclusivity of other drugs. However, clean identification in this setting is possible through use of an instrument which exploits the third principal factor determining market exclusivity: the timing of patent filing. This section explains the logic underlying this instrument before presenting an empirical analysis of the instrument itself.

The Timing of Patent Filing

Patent lifetimes are fixed, so the date of a drug's AI patent filing has a strong influences on the drug's exclusivity. So what determines the timing of patent filing?

Firms face incentives to file patents the day before they begin clinical trials (i.e., as late as possible) because doing so maximizes future exclusivity. However, the reality is that patents are sometimes filed substantially in advance of the start of clinical trials. The summary

[30] Appendix Figure C.1 visualizes a similar relationship by plotting the exclusivities of subsequent entrants against the exclusivities of FIC drugs (not all pairwise exclusivities). Panel A shows that without controls the relationship is weak and statistically insignificant. Conditioning on mean development time in class, market size, and year of FIC approval fixed effects leads to a weak negative association, as visualized in Panel B.

statistics presented in Table 3.1 show that, for FIC drugs, patents are filed at median 4.43 years before clinical trials begin.[31] Interviews with patents experts and R&D directors suggest that this is no surprise. They point me towards three primary factors that contribute to it.

First, there can be agency problems within firms. In particular, the career motives of scientists and the motives of the larger organization are not necessarily aligned. Scientists receive acclaim through research presented in papers and in conferences. Firms support these endeavors, but they place a high value on intellectual property and tend to require that patents be filed before new results are made public. Together, this means patents may be filed at a time that suits the scientist – before the firm is prepared to undertake clinical review.

Second, there can be larger organizational frictions. The pharmaceutical industry has been characterized by a tremendous amount of merger and acquisition activity in recent decades (see, e.g., Deloitte (2009)). A major organizational shock like an acquisition presumably complicates strategic decision making and might slow the progression from the lab to clinical review. Below, I show empirically that mergers are associated with increases in delay from patent filing to the start of trials.

Third, and perhaps most importantly, the science of drug development is highly uncertain while the patent system forces firms to act defensively. That is, the trajectory of a drug's development is not necessarily linear – a molecule is sometimes initially thought to treat one condition when it later turns out to be better at treating another.[32] Yet the owner of a patent is whichever firm filed it first, and the discounted gain of waiting an extra year to file, i.e. an extra year of exclusivity twenty years in the future, may be low relative to the loss incurred if another firm files the patent first. This means firms file patents as soon as it

[31] The date on which intellectual property protection is first sought is actually called a patent's priority date. Priority dates are sometimes called "effective filing dates" and determine expiration dates. More detail on the differences between priority and filing dates are found in Appendix Section C.2.4. For simplicity, I refer to the priority date throughout as the filing date.

[32] Recall the case of Viagra in Footnote 7.

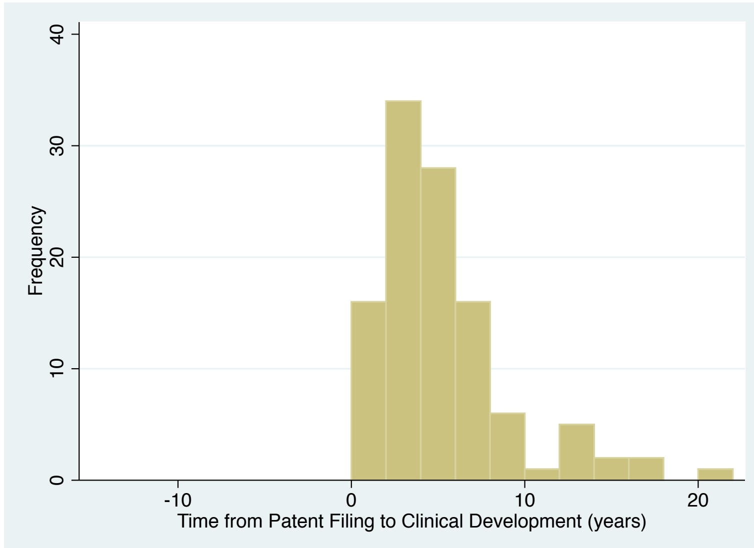
is known that a molecule *may* be of commercial interest, but before it is necessarily clear what that commercial interest will be.

Consider, e.g., the path to approval of Arava, a drug presently marketed for rheumatoid arthritis. The active ingredient in Arava was discovered and patented in the late 1970s by researchers searching for new agricultural pesticides (Oh and O'Connor (2013)). In 1985, researchers experimenting with rats believed Arava might have potential for use in transplantation (Hoi and Littlejohn (2005)). However, it was not until 1993 that the drug first entered clinical development – for arthritis. By the time Arava was approved in 1998, its patent was already about to expire. The FDA's exclusivity minimums kept generics out for almost 7 years, leading Arava to end up with about half of the mean exclusivity of almost 12 years in my sample. However, that Arava's development was delayed well past its initial discovery does not appear to be the result of low sales expectations: at the time of approval, financial analysts estimated that Arava's annual sales could exceed $800 million by 2003 (WSJ (1998)). Consistent with the thesis of this paper, Arava has now been off patent for almost a decade and there has been no entry aside from Arava in its class.

Identification Strategy

My IV strategy exploits gaps between when patents are first established on a molecule and the beginning of clinical trials. Specifically, my instrument for market exclusivity is the time between the date on which intellectual property protection was first sought and the date marking the beginning of clinical development. I denote the instrument for drug *i* by $z_i = t_{DevStart,i} - t_{PatentFiling,i}$. Large, positive values of z_i reflect molecules that are discovered substantially before they are commercialized. For example, the gap between Arava's patent filing date and the start of clinical development was 15 years (1978 through 1993), and its market exclusivity was subsequently just 7 years. Figure 3.6 shows a histogram of the instrument for FIC drugs. The majority of the mass is just above 0, reflecting patent filing

Figure 3.6: *Histogram of the Time from Patent Filing to Clinical Development*



Notes: This figure shows a histogram of the time from patent filing to the start of clinical development for FIC drugs. The start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. The sample is as described in Section 3.3.

shortly before the beginning of trials.[33] However, there is also a right tail in which drugs like Arava are found.

Formally, identification requires that $Corr(z_j, \varepsilon_j) = 0$. That is, identification requires that delays between patent filing and the start of clinical development for FIC drugs are unrelated to other factors that affect entry in the class. Although this cannot be tested directly, in the following two sections I provide two pieces of evidence to its effect. First, I ask descriptively how the instrument relates to patent characteristics. I find no evidence that delays between patent filing and the start of development signal anything about salient patent characteristics like claims or citations, but they are related to factors like merger activity. Second, to dispel concerns that the instrument might be picking something up about unobserved profitability, I ask whether the instrument is related to sales in class,

[33] In fewer than 4% of cases, calculation of the instrument yields a negative value, reflecting a filing date that occurs after the beginning of clinical trials. Since this might reflect that the patent identified by my methodology is not an active-ingredient patent, I am hesitant to use these values and set the value of the instrument in these cases to 0. However, leaving the instrument at its original values or dropping these observations altogether does not affect the main results. I note also that the mass in the histogram would be located closer to 0 if I focused on true filing dates, as these are always weakly greater than priority dates.

Table 3.2: *Patent Characteristics and the Time between Patent Filing and Clinical Development*

Dependent Variable is	*Number of Patent Claims*	*Number of Patent Citations*	*Number of Patents Referenced By*	*(1/0): Any Patent Reassign.*	*(1/0): Reassigned In Merger*	*Time from Filing to First Reass. (yrs)*
	(1)	(2)	(3)	(4)	(5)	(6)
Patent Filing to Clinical Dev. (yrs)	0.00167	-0.0139	-0.0395	-0.00822	0.0179*	0.353**
	(0.0183)	(0.0184)	(0.0245)	(0.00738)	(0.00942)	(0.141)
Estimation	Poisson	Poisson	Poisson	OLS	OLS	OLS
Mean of Dependent Variable	18.22	76.85	11.82	0.870	0.182	5.299
N	208	208	208	208	181	181

Notes: This table examines the relationship between patent characteristics and the instrument, which is the time between patent filing and clinical development. The start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. Columns (1)-(3) respectively present Poisson models of the relationship between the number of patent claims, number of patent citations, and number of other patents that reference the given patent and the instrument. Columns (4)-(6) examine whether the instrument is related to reassignment by estimating OLS models. Column (4) regresses a dummy for any reassignment on the instrument, while column (5) regresses a dummy for whether that patent was ever reassigned in a merger transaction on the instrument, and column (6) regresses the time from filing to first reassignment on the instrument. The sample is the full sample of classes for which a first in class was approved between 1987 and 2011 and for which I observe the start of clinical development. Note that specifications (5) and (6) are conditional on any reassignment, so the number of observations drops from 208 to 181. Poisson models are estimated by quasi-maximum likelihood and robust standard errors are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

and, in a test like that used in my analysis of market exclusivities, I ask also whether the instrument is significantly correlated within class. I find no evidence for either.

The Instrument and Patent Characteristics

I analyze the relationship between the instrument and patent characteristics in Table 3.2. Columns (1)-(3) respectively ask whether the instrument is correlated with the patent's number of claims, its number of citations, and the number of other patents that reference it. The sample is the full sample of 208 drugs for which an FIC was approved in the period 1987-2011 and for which I observe the date on which clinical development begins (and thus am able to compute a value for the instrument). Reflecting the count nature of these outcome variables, I estimate QML Poisson models. The point estimates are neither economically nor statistically significant in all cases. This suggests these patents were not in some sense less "important" ex-ante, as judged by claims and citations, nor ex-post, as judged by the number of other patents referencing the patent in question.

Next, in columns (4)-(6), I ask whether the instrument is related to reassignment of patent ownership, which is tracked by Thomson Innovation.[34] Column (4) estimates an OLS model of a dummy for reassignment on the instrument and shows that the instrument is not statistically nor economically related to reassignment. In column (5), I ask whether the instrument is related specifically to reassignment involving a merger, and this time the estimate is positive and statistically significant at the 10% level. Finally, in column (6) I analyze how the instrument relates to a proxy for the degree of initial uncertainty which is the length of the time between filing and first reassignment. The estimated coefficient is strong and positive.[35] Altogether, Table 3.2 shows that drugs with large z_i's appear to have patents that are similar to those of other drugs, except that they are more likely to have been involved in a merger and that they took longer to make it from the filer's hands into those new of a owner. These results suggest the z_i are more likely to originate from organizational changes and scientific uncertainty than from particularities of the patent itself.

The Instrument and Class Characteristics

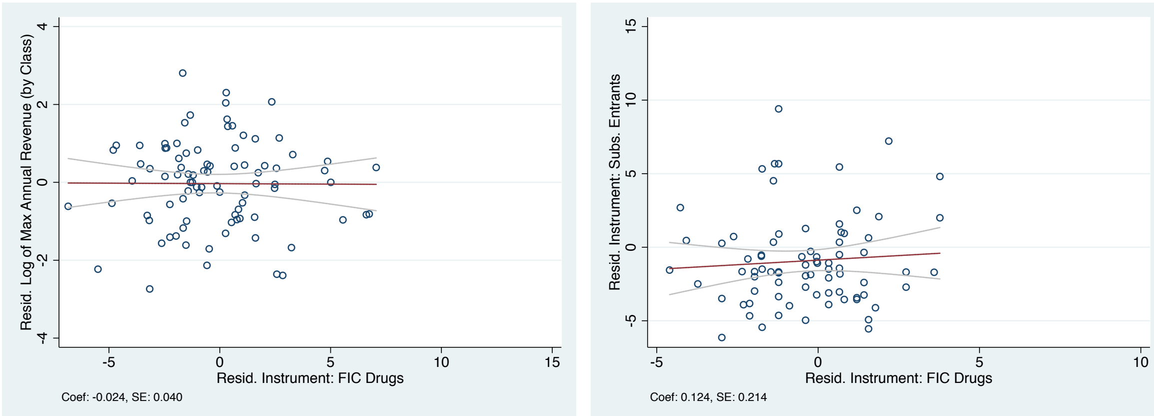
The evidence analyzed above provides context for the instrument but it does not speak directly to the question of exogeneity: the instrument might still reflect something about a class that influences both FIC exclusivity and total entry. For example, capacity-constrained firms might put lower priority on less profitable markets, or it may simply be more difficult to develop new drugs in some classes than others.

I examine the first possibility in Panel A of Figure 3.7 where I proxy for profitability using the log of the maximum annual revenue earned by any drug in a class, and ask how that relates to the instrument for FIC drugs. (I condition both on my baseline controls: FIC approval year fixed effects, market size, and mean development time in class.) The figure

[34]Technically, I examine changes in patent assignee. Such changes can represent licensing agreements in addition to changes in ownership.

[35]In columns (5) and (6), the sample is conditional on any reassignment, so the number of observations drops from 208 to 181.

Figure 3.7: *Evidence on the Validity of the Instrument*



(a) *Comparing Class Sales with the Time from Patent Filing to Clinical Dev. for FIC Drugs*

(b) *Comparing the Time from Patent Filing to Clinical Development for Subs. and FIC Drugs*

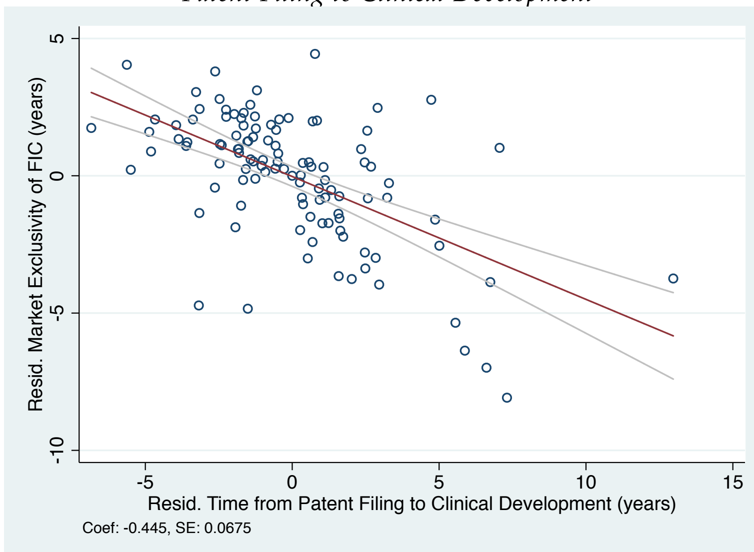
Notes: These figures respectively plot class sales against the instrument for FIC drugs and the instrument for drugs subsequent to the FIC against the instrument for FIC drugs. The instrument is the time between patent filing and clinical development and all measures are conditional on year of FIC approval fixed effects, mean development time, and market size. The measure of class sales is the log of the maximum of annual sales for all drugs in the class. The specification of the controls and sample are as described in Section 3.3 and the slope of the line of best-fit and its associated robust standard error clustered by drug class are presented in the bottom left of each figure.

shows there is no significant relationship between between class sales and the time between patent filing and the start of development for the FIC.[36]

Though the instrument is not related to class sales does not mean it is not related to other class-specific unobservables, and this possibility is difficult to rule out completely. To address it, I implement a test similar to that presented earlier: I ask whether the instrument is correlated within classes. In particular, I estimate an OLS model where the outcome is z_k and the independent variables are z_l, as well as my baseline controls, for all pairwise combinations of drugs k and l belonging to the same class. The estimated coefficient on the market exclusivity of drug l is -0.00375 with standard error 0.0471, where the standard error is robust and clustered by class. In Panel B of Figure 3.7, I also plot the instrument for subsequent entrants against the instrument for FIC drugs, where both are again conditional on my baseline controls. The figure shows the estimated relationship is economically

[36]I choose the maximum annual revenue earned by any drug in the class as a proxy for profitability because this figure reflects whether a given class could possibly be a top seller. However, the results are similar if I instead proxy for profitability with the maximum of the FIC drug's annual sales. I additionally show in Appendix Section C.5.5 that controlling directly for sales in my main regressions does not affect the results.

Figure 3.8: *Relationship Between Market Exclusivity and Time from Patent Filing to Clinical Development*



Notes: This figure shows a plot of the residualized first-stage relationship between a FIC drug's market exclusivity and the time between its patent filing and start of clinical development. Residuals are conditional on year of FIC approval fixed effects, mean development time, and market size, and the specification of the controls and sample are as described in Section 3.3. The slope of the line of best-fit and its associated robust standard error are presented in the bottom left of the figure.

and statistically insignificant. I conclude that the instrument is unlikely to be picking up class-specific unobservables, and next present the first stage before turning to the results.

The First Stage

Figure 3.8 presents the first stage, where both axes are conditional on my baseline controls. The relationship between the instrument and FIC exclusivity is strong – a one year change in the value of the instrument is associated with 0.45 fewer years of exclusivity – and the plot shows that the relationship remains strong across the support of the instrument. The associated F-statistic is 43.52.

3.5.3 Results

In the previous sub-sections, I explained the sources of variation underlying a drug's market exclusivity and I presented evidence analyzing the potential for endogeneity in exclusivities. I then proceeded to explain the logic behind my instrumental variable and presented evidence analyzing its relationship with patent and market characteristics, before finally presenting the first stage. This section presents my main results on how FIC exclusivity affects subsequent entry.

Table 3.3 presents the estimates, with results from Poisson models in the first set of three columns and from Poisson-IV models in the second set of three columns. Because of the Poisson's non-linearity, I implement the IV using 2-stage residual inclusion.[37] All standard errors are robust and the second stage IV standard errors are corrected for the two stage design using the methodology of Murphy and Topel (1985) and Hardin (2002) All specifications include fixed effects for the year of FIC approval. The first specification in each set of columns includes only the year of FIC approval fixed effects, and then in subsequent columns I incrementally add my baseline controls for mean time in development and market size. Corresponding first stage estimates are reported below the IV results.

In all cases, I estimate a strong, positive and statistically significant effect of the FIC's exclusivity on subsequent entry. The point estimates remain relatively consistent as controls are added. The Poisson-IV results are slightly larger in magnitude than the Poisson results, but they are also less precise, and the differences between the coefficients are not statistically significant. Overall, the estimates show a one-year increase in FIC exclusivity yields a 25-30% increase in subsequent entry in class. Poisson models estimate proportional effects, so to ease interpretation of the results, I present in the fifth row of the table the main exclusivity coefficient multiplied by the mean of the dependent variable. These estimates, which I refer to as "scaled", are in units of subsequent drugs per year of first in class exclusivity. The estimates presented in columns (3) and (6) show one additional year of first in class exclusivity yields approximately 0.18 to 0.25 more subsequent drugs. This is a relatively large effect. For example, It implies that if FIC exclusivities were just one standard deviation shorter (about 3 years), the number of subsequent entrants in the average class would be reduced to zero. Or, in the other direction, the estimates imply that increasing FIC

[37] The Poisson's non-linearity means that simply plugging predicted values from the first-stage into the second-stage yields inconsistent estimates. I implement a control function methodology as detailed in Wooldridge (2002) and Imbens and Wooldridge (2007).

Table 3.3: *Estimates of the Effect of First in Class Exclusivity on Subsequent Entry*

	Dependent Variable is Number of Subsequent Entrants in Class (mean = 0.73)					
	QML Poisson Models			QML Poisson-IV Models		
	(1)	(2)	(3)	(4)	(5)	(6)
Market Exclusivity of FIC (years)	0.229***	0.243***	0.251***	0.411**	0.335***	0.337***
	(0.0604)	(0.0631)	(0.0613)	(0.181)	(0.128)	(0.128)
Mean Development Time	no	yes	yes	no	yes	yes
Market Size	no	no	yes	no	no	yes
Year of FIC Approval Fixed Effects	yes	yes	yes	yes	yes	yes
Exclusivity Coef x Mean of Dep Var	0.167	0.177	0.183	0.300	0.245	0.246
Estimates from the First Stage:						
Patent Filing to Clinical Dev. (years)	-	-	-	-0.284***	-0.442***	-0.445***
				(0.0653)	(0.0669)	(0.0675)
F-Statistic	-	-	-	18.84	43.55	43.52
N	111	111	111	111	111	111

Notes: This table presents estimates of the effect of first in class exclusivity on subsequent entry. Columns (1)-(3) present results from quasi-maximum likelihood Poisson models which incrementally add controls for mean development time in class and market size. Columns (4)-(6) replicate columns (1)-(3) but instrument for FIC market exclusivity using the time between patent filing and the start of clinical development, where the start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. All models are conditional on year of FIC approval fixed effects. The IV models are implemented using a control function and first stage estimates are presented in the final rows of the table. The sample and controls are specified as described in Section 3.3. Robust standard errors are reported in parentheses and standard errors for the IV models are corrected for the two-stage design as described in the text. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

exclusivities by one standard deviation would double the number of subsequent entrants.

Appendix Section C.5 presents five sets of robustness checks. First, Appendix Section C.5.1 shows the main Poisson results remain similar when the sample restrictions described in Section 3.3 are relaxed. Second, Appendix Section C.5.2 shows that the main Poisson analysis survives placebo tests which analyze the role of the exclusivities of non-FIC drugs on entry. Third, I investigate the robustness of the main IV estimates in Appendix Section C.5.3, where I present the reduced form of the second stage and show the IV estimates are robust to a linear specification. Fourth, since some classes in my data have had only a few years to accumulate new entrants, in Appendix Section C.5.4, I show Table 3.3 is unchanged when the sample is restricted to only classes which have had at least 10 years for subsequent to entry to occur. Fifth, to ensure my results are not confounded by class profitability, I show in Appendix Section C.5.5 that Table 3.3 is unchanged when I control for sales.

3.6 Price Elasticities and Quality Levels

Having shown that FIC exclusivity drives subsequent entry, I turn to the last two predictions of the model: that the FIC's patent should have a greater impact on subsequent entry when demand is more elastic and when the entrant is a less significant clinical advance.

3.6.1 Interaction with Price Elasticity

I first analyze how the effect of FIC exclusivity on subsequent entry is related to price elasticity. I do not directly observe price elasticities and they are difficult to infer directly from the data without imposing strong structural assumptions; pharmaceutical pricing is the product of a complex multi-agent bargaining game and demand system whose players include health insurers, pharmacy benefit managers, pharmacies, hospitals, physicians, and patients. Thus, my strategy is simply to use two proxies for price elasticity. The first proxy is a dummy that captures whether a drug treats a chronic condition and the second is the mean household income of patients. Denoting the price elasticity proxy for class j by D_j, I then estimate Poisson models of the form,

$$SubsEntrants_j = \alpha + \beta FICExcl_j + \psi D_j + \phi FICExcl_j \times D_j + \gamma' X_j + \varepsilon_j. \tag{3.9}$$

The following provides more detail on my measures and then describes the results.

Chronic Conditions

My first price elasticity proxy is an indicator variable that captures whether a class of drugs treats a chronic condition (as opposed to one that is acute). This is is motivated by prior evidence that demand for acute care tends to be inelastic.[38] This is intuitive: the marginal value of a single, one-time treatment for a chronic condition is likely lower (or at least less salient) than for an acute condition. Moreover, treatment for chronic conditions takes place over longer periods of time, giving patients more opportunity to seek out lower cost options.

To determine whether a given class treats a chronic condition, I augment my data with the Chronic Condition Indicator distributed by the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project (HCUP). This specifies whether a given ICD-9 condition code and its sub-codes pertain to a chronic condition. Since I only observe 3-digit ICD-9 codes for each class (and not their sub-codes), I create a dummy for each class which designates whether any of the ICD-9 sub-codes treated by that class are chronic conditions. For example, the code 185 designates prostate cancer and is classified as chronic, while 041 designates a bacterial infection and is not chronic.[39] Overall, 83% of classes are classified as treating chronic conditions.

The results are reported in columns (1) and (2) of Table 3.4, Panel A, with the first column presenting Poisson estimates and the second column presenting Poisson-IV estimates.[40]

[38]Newhouse and the Insurance Experiment Group (1993) presents results from the RAND experiment; Ringel *et al.* (2002) reviews this literature.

[39]Technically, ICD-9 185 is "malignant neoplasm of prostate" and 041 is "bacterial infection in conditions classified elsewhere and of unspecified site."

[40]I now have two endogenous variables, so I need two instruments. Since the instrument is not binary, I follow convention and include as the second instrument the square of the first.

The results show that, consistent with the theory, the estimated effect of FIC exclusivity on subsequent entry is significantly stronger for chronic conditions – in fact, the effect seems to be largely driven by chronic conditions, as the non-interacted effect is small and statistically indistinguishable from 0.

Household Income

My second proxy for price elasticity is the mean household income of patients. The idea is simple: consumers with higher incomes are likely to be less sensitive to prices. That price elasticities are related to income has a long tradition in the industrial organization literature and is straightforward to derive out of a standard demand framework.[41]

To perform this analysis, I compute the average household income of patients in each ICD-9 code using data from MEPS.[42] The resulting measure appears consistent with existing evidence on the incidence of specific conditions by socio-economic status. For example, ICD-9 code 250 designates diabetes mellitus and has a mean household income of about \$50,000, while 692 designates eczema and has a mean household income of about \$80,000.[43]

I then estimate models which are analogous to those estimated for chronic conditions except that they use the log of mean household income as D_j. The results are reported in columns (3) and (4) of Table 3.4, Panel A, again with the first column presenting Poisson estimates and the second column presenting Poisson-IV estimates. Consistent with the theory, the interaction effect is significantly negative: i.e., the effect of FIC exclusivity on subsequent entry is significantly stronger for conditions more prevalent among lower income patients.[44] Indeed, the estimates suggest that the effect is moderated substantially by income: the effect on subsequent entry of an extra year of FIC exclusivity is 13% in classes

[41]See, e.g. Berry *et al.* (1995) for a derivation and Ackerberg *et al.* (2007) for a review.

[42]Appendix Section C.2.5 provides additional detail.

[43]Brancati *et al.* (1996) find that diabetes is more prevalent among low SES populations while Fu *et al.* (2014) find the opposite is true of eczema. ICD-9 692 technically refers to "contact dermatitis and other eczema."

[44]It is also worth noting that the coefficient on the log of mean household income is positive: intuitively, classes in which patients are relatively higher income see more entry (conditional on market size).

Table 3.4: *Demand Elasticities and Priority Review Designations*

Panel A: Interaction Effects
Dependent Variable is Number of Subsequent Entrants in Class

	(1)	(2)	(3)	(4)
Market Exclusivity of FIC (years)	-0.00510	0.0694	16.49**	16.66***
	(0.120)	(0.134)	(6.654)	(6.050)
Chronic	-2.702	-3.616*		
	(1.853)	(1.975)		
Chronic x Market Exclusivity of FIC (years)	0.251*	0.322**		
	(0.136)	(0.148)		
Log Income			19.00**	18.98**
			(8.174)	(7.379)
Log Income x Market Exclusivity of FIC (years)			-1.472**	-1.480***
			(0.601)	(0.544)
Mean Development Time	yes	yes	yes	yes
Market Size	yes	yes	yes	yes
Year of FIC Approval Fixed Effects	yes	yes	yes	yes
Estimation	Poisson	Poisson-IV	Poisson	Poisson-IV
F-Statistic from the First Stage	-	21.22	-	20.49
N	111	111	111	111

Panel B: By Priority Review Designation

	Dependent Variable is Number of Subs. Priority Entrants in Class		*Dependent Variable is Number of Subs. Non-Priority Entrants in Class*	
	(1)	(2)	(3)	(4)
Market Exclusivity of FIC (years)	0.101	0.168	0.298***	0.425***
	(0.0761)	(0.256)	-0.0779	(0.157)
Mean Development Time	yes	yes	yes	yes
Market Size	yes	yes	yes	yes
Year of FIC Approval Fixed Effects	yes	yes	yes	yes
Mean of Dependent Variable	0.09	0.09	0.64	0.64
Exclusivity Coef x Mean of Dep Var	0.009	0.015	0.191	0.272
Estimation	Poisson	Poisson-IV	Poisson	Poisson-IV
F-Statistic from the First Stage	-	43.52	-	43.52
N	111	111	111	111

Notes: Panel A replicates the base analysis in Table 3.3 but analyzes, in columns (1) and (2), the interaction of FIC exclusivity with a dummy for whether an ICD-9 code is a chronic condition, and in columns (3) and (4), the interaction of FIC exclusivity with the log of mean income of patients in that ICD-9. Panel B replicates the base analysis in Table 3.3 but splitting new drugs by Priority-Review status, an expedited review status granted by the FDA for drugs that are expected to present significant clinical advances. Results are from Poisson and Poisson-IV models estimated by quasi-maximum likelihood. The instrument, which is the difference in time between the patent's filing date and the start of clinical development, is implemented as a control function. The sample and controls are specified as described in Section 3.3. Robust standard errors are reported in parentheses and standard errors for the IV models are corrected for the two-stage design as described in the text. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

where patient incomes are 10% higher than average, while it is 43% in classes where patient incomes are 10% lower.

Overall, I infer that price elasticities influence how much the FIC's exclusivity affects subsequent entry.

3.6.2 Quality

The final comparative static signed by the model is the interaction of quality with patent protection: the model predicts an incumbent's exclusivity should have a lesser effect on drugs which are substantially better than those already on the market.

As Kesselheim *et al.* (2013) write, quantifying differences in quality between drugs is difficult; prior work has looked at a variety of measures including the number of citations a drug receives in the medical literature, the number of citations its main patent receives, or simply its sales. I take a direct approach, which is to ask whether a drug received a priority review designation from the FDA.[45]

I perform this analysis by estimating Equation (3.8) once more, but this time I separate subsequent entrants according to whether they received priority review status. The results are reported in Panel B of Table 3.4. The first two columns respectively report Poisson and Poisson-IV estimates analyzing the effect of FIC market exclusivity on the number of subsequent entrants which are priority review, while the third and fourth columns do the same for non-priority review drugs. The results in the fourth to last row, which reports the exclusivity coefficients scaled by the mean of the dependent variable, are consistent with the theory: the estimated effect of FIC exclusivity on subsequent priority review entrants is roughly one-twentieth of that on subsequent non-priority review entrants. The differences between these effects are significant at the 1% level. In other words, the FIC's exclusivity

[45] I am not the first to proxy for important advances using priority review status: among others, Lanthier *et al.* (2013) and Dranove *et al.* (2014) implement similar approaches. A different approach would be to estimate changes in consumer surplus using a full structural model, e.g., with a framework as in Petrin (2002) who evaluates the welfare impact of the introduction of the minivan. Others have tried to estimate the welfare impact of drug introductions (e.g., Arcidiacono *et al.* (2013)) but it is difficult to capture horizontal differences across drugs – and the value of such differences to consumers – with the data that are available.

only impacts subsequent entry for non-priority review drugs.

3.7 Conclusion

This paper shows that patents have important effects on the direction of pharmaceutical innovation. Specifically, I show that the length of an incumbent's patent protection has a positive impact on subsequent entry: an extra year of FIC exclusivity leads to an estimated 25-30% increase in subsequent drug approvals in the same class, which equates to about 0.2 drugs. The effect is stronger for classes for which demand is more elastic and for drugs which are lesser clinical advances.

My work has implications for how researchers and policy makers weigh the benefits and disadvantages of patents. A simple framework that determines the optimal level of patent protection by comparing a single would-be innovator's incentives with the (subsequent) social cost imposed by that innovator's patent protection fails to account for the dynamic incentives that the initial innovator's patent protection provides to subsequent innovators. Properly weighing the pros and cons of patent protection means taking into account not only how a single firm's patents impacts its own incentives but also how those incentives are shaped by other firms' patents.

Prior work has not taken this second effect into consideration. For example, Kremer (1998) proposes to eliminate the monopoly distortion created by patenting by having the government buy out important patents through an auction mechanism. But doing away with patents means shutting down precisely the effect I describe.[46] My results show that implementing such a mechanism would drastically change the incentives of subsequent entrants.

Of course, given the high fixed costs of drug development, it is not clear that a reduction in entry would decrease welfare. My analysis focuses on an outcome of direct relevance to consumers (the number of alternatives available within a class of drugs), but I do not

[46] If Kremer (1998)'s mechanism were implemented, FIC drugs would be available immediately after approval at marginal cost, so their market exclusivities would be effectively set to zero.

attempt to quantify how patents impact social surplus because characterizing the welfare effects of a change in patent length is difficult without strong assumptions about supply and demand. However, that my findings are less pronounced for priority review drugs suggests the effect I identify may be driven in part by incentives for business stealing.[47] But this statement requires a strong caveat: this is an industry known for its outliers and history shows it is not true that drugs which were neither first in class nor priority review – i.e., the subsequent entrants I find are most affected by FIC exclusivity – are necessarily unimportant, at least with respect to sales. For example, Singulair and Celexa did not receive priority review and were, respectively, fourth and fifth to market in their classes, yet both drugs' annual sales surpassed $4B.[48]

More generally, this paper has implications for how intellectual property is characterized. Traditional research identifies two independent levers in patent design: length – how long a patent remains in force – and breadth – the extent to which a patent "covers" a field. This paper shows that this characterization is too simple because it fails to account for the fact that length and breadth are linked through elasticities of substitution. For example, if increasing the length of an incumbent's patent protection increases the value of entry for differentiated substitutes, then shorter patents are, effectively, more broad.[49] Deeper understanding of this competition-induced connection between length and breadth may yield new insights into how intellectual property protection should be designed by the "economist-as-engineer" (Roth (2002)) to shape innovation.

The focus in this paper has been on small-molecule drugs and the paper's logic hinges on the fact that generics are chemically equivalent to approved drugs yet they are cheap to introduce once patents have expired. However, biologic drugs (which are more considerably

---

[47]Bloom *et al.* (2013) provide evidence that business stealing does play an important role in the market valuations of pharmaceutical manufacturers.

[48]This figure comes from publicly available IMS Health data available at `www.drugs.com`.

[49]This line of reasoning bears some resemblance to that articulated by O'Donoghue *et al.* (1998), who argue that in certain cases, what matters is not true patent life but rather the length of time a product has on patent until a new product is developed which completely displaces the old product. They call this length of time, "effective" patent life. However, they do not consider how the development of the new product might actually be deterred by the expiration of the old product's patent – i.e., the subject of this paper.

more complex at the molecular level) are an increasingly important part of the pharmaceutical market and the precise notion of a “generic" biologic drug – called a biosimilar – has not yet been established by regulators. My work suggests there is nuance to this definition because the extent of biosimilar competition is likely to have important present and future impacts on biopharmaceutical innovation.

Finally, the fundamental question considered in this paper is the extent to which cheap older generations of products can undermine incentives to create new products. Such a situation is unlikely to occur in contexts in which innovation progresses rapidly, or is even accelerating. But that new generations of pharmaceuticals are cannibalized by generic alternatives suggests a need for a broader understanding of the relationship between the design of intellectual property protection and the *speed* of innovation.

References

Acemoglu, D. and Linn, J. (2004). Market size in innovation: Theory and evidence from the pharmaceutical industry. *Quarterly Journal of Economics*, **119** (3), 1049–1090.

Ackerberg, D., Benkard, L. C., Berry, S. and Pakes, A. (2007). Econometric tools for analyzing market outcomes. *Handbook of Econometrics*, **6**, 4171–4276.

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, **97** (4), 543–569.

— and Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics*, **105** (2), 255–283.

Al-Ubaydli, O., Andersen, S., Gneezy, U. and List, J. A. (2014). Carrots that look like sticks: Toward an understanding of multitasking incentive schemes. *Southern Economic Journal, forthcoming*.

Alvir, J. M. J., Lieberman, J. A., Safferman, A. Z., Schwimmer, J. L. and Schaaf, J. A. (1993). Clozapine-induced agranulocytosis. *The New England Journal of Medicine*, **329** (3), 162–167.

Arcidiacono, P., Ellickson, P. B., Landry, P. and Ridley, D. B. (2013). Pharmaceutical followers. *International Journal of Industrial Organization*, **31** (5), 538–553.

Associated Press (2011). Lipitor becomes world's top-selling drug. *Crain's New York Business*, **December 28**.

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

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

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

—, Chernozhukov, V. and Hansen, C. (2010). Lasso methods for gaussian instrumental varriables models. *arXiv preprint arXiv:1012.1297*.

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

—, —, — and Kozbur, D. (2014). Inference in high dimensional panel models. *mimeo*.

Berenson, A. (2006). Merck loses protection for patent on zocor. *The New York Times*, **June 23**.

Berg, J., Dickhaut, J. and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, **10** (1), 122–142.

Berndt, E. R. (2002). Pharmaceuticals in u.s. health care: Determinants of quantity and price. *The Journal of Economic Perspectives*, **16** (4), 45–66.

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

Berry, S., Levinsohn, J. and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, **63** (4), 841–890.

Bessen, J. and Maskin, E. (2009). Sequential innovation, patents, and imitation. *RAND Journal of Economics*, **40** (4), 611–635.

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

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

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

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

Bloom, N., Schankerman, M. and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, **81** (4), 1347–1393.

Brancati, F. L., Whelton, P. K., Kuller, L. H. and Klag, M. J. (1996). Diabetes mellitus, race, and socioeconomic status: A population-based study. *Annals of Epidemiology*, **6** (1), 67–73.

Branstetter, L., Chatterjee, C. and Higgins, M. J. (2014). Killing the golden goose or just chasing it around the farmyard?: Generic entry and the incentives for early-stage pharmaceutical innovation. *Working paper*.

Budish, E., Roin, B. N. and Williams, H. (2014). Do firms underinvest in long-term research? evidence from cancer clinical trials. *Working paper*.

Bursztyn, L., Ederer, F., Ferman, B. and Yuchtman, N. (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, **82**, 1273–1301.

Cabral, L. and Natividad, G. (2013). Box-office demand: The importance of being #1. *mimeo*.

Cattaneo, M. D., Jansson, M. and Newey, W. K. (2012). Alternative asymptotics and the partially linear model with many regressors. *mimeo*.

Caves, R. E. (2001). *Creative Industries: Contracts Between Art and Commerce*. Cambridge, MA: Harvard University Press.

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

Charness, G. (2004). Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics*, **22** (3), 665–688.

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

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

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.

Civan, A. and Maloney, M. T. (2009). The effect of price on pharmaceutical r&d. *The B.E. Journal of Economic Analysis & Policy*, **9** (1), 1–22.

Cockburn, I. M. (2007). Is the pharmaceutical industry in a productivity crisis? In J. Lerner and S. Stern (eds.), *Innovation Policy and the Economy*, MIT Press, vol. 7, pp. 1–32.

Conover, W. J. (1999). *Practical Nonparametric Statistics*. Wiley, 3rd edn.

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.

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

DellaVigna, S., List, J. A., Malmendier, U. and Rao, G. (2014). *Voting to Tell Others*. Working Paper 19832, National Bureau of Economic Research.

Deloitte (2009). *Acquisitions versus Product Development: An Emerging Trend in Life Sciences*. Tech. rep., Deloitte, Inc.

DiMasi, J. A., Hansen, R. W. and Grabowski, H. G. (2003). The price of innovation: New estimates of drug development costs. *Journal of Health Economics*, **22** (151–185).

Dranove, D., Garthwaite, C. and Hermosilla, M. (2014). Pharmaceutical profits and innovation: Evidence from medicare part d and the biotechnology sector. *Working paper*.

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

Elberse, A. and Anand, B. (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.

— and Eliashberg, J. (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 Ellison, S. F. (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics*, **3** (1), 1–36.

— and Fudenberg, D. (1995). Word-of-mouth communication and social learning. *Quarterly Journal of Economics*, **110** (1), 93–125.

Englmaier, F. and Leider, S. (2012). Managerial payoff and gift exchange in the field. *Working paper*.

Esteves-Sorenson, C. and Macera, R. (2013). Revisiting gift exchange: Theoretical considerations and a field test. *Working paper*.

Falk, A. (2007). Gift exchange in the field. *Econometrica*, **75** (5), 1501–1511.

—, Fehr, E. and Fischbacher, U. (2008). Testing theories of fairness – intentions matter. *Games and Economic Behavior*, **62** (1), 287–303.

Farrell, J. and Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of Industrial Organization*, **3**, 1967–2072.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, **14** (3), 159–181.

—, Goette, L. and Zehnder, C. (2009). A behavioral account of the labor market: The role of fairness concerns. *Annual Review of Economics*, **1**, 355–384.

—, Kirchsteiger, G. and Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *Quarterly Journal of Economics*, **108** (2), 437–459.

Finkel, R. (2012). The atypical history of atypical antipsychotics. *Drug Information and Side Effects Database*, **September 9, 2012**.

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

Fu, T., Keiser, E., Linos, E., Rotatori, R. M., Sainani, K., Lingala, B., Lane, A. T., Schneider, L. and Tang, J. Y. (2014). Eczema and sensitization to common allergens in the United States: A multiethnic, population-based study. *Pediatric Dermatolology*, **31** (1), 21–26.

Fudenberg, D., Gilbert, R. J., Stiglitz, J. E. and Tirole, J. (1983). Preemption, leapfrogging and competition in patent races. *European Economic Review*, **22** (1), 3–31.

Gallini, N. T. (1992). Patent policy and costly imitation. *RAND Journal of Economics*, **23** (1), 52–63.

Gilbert, R. and Shapiro, C. (1990). Optimal patent length and breadth. *RAND Journal of Economics*, **21** (1), 106–112.

Gill, D. and Sgroi, D. (2012). The optimal choice of pre-launch reviewer. *Journal of Economic Theory*, **147**, 1247–1260.

Gneezy, U. and List, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, **74** (5), 1365–1384.

Green, J. R. and Scotchmer, S. (1995). On the division of profit in sequential innovation. *RAND Journal of Economics*, **26** (1), 20–33.

Greenhouse, S. (2014). Gap to raise minimum hourly pay. *The New York Times*, **February 19**.

Griliches, Z. (1998). Patent statistics as economic indicators: A survey. In Z. Griliches (ed.), *R&D and Productivity: The Econometric Evidence*, University of Chicago Press, pp. 287–343.

Hannan, R. L., Kagel, J. and Moser, D. (2002). Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences, and effort requests on behavior. *Journal of Labor Economics*, **20** (2), 923–951.

Hardin, J. W. (2002). The robust variance estimator for two-stage models. *The Stata Journal*, **2** (3), 253–266.

Hausman, J. A., Hall, B. H. and Griliches, Z. (1984). Econometric models for count data with an application to the patents-r&d relationship. *Econometrica*, **52** (4), 909–938.

Hay, M., Rosenthal, J., Thomas, D. and Craighead, J. (2011). *BIO / BioMedTracker: Clinical Trial Success Rates Study*. Tech. rep., BioTechnoloy Industry Association.

Hemphill, C. S. (2007). Drug patent settlements between rivals: A survey. *Working paper*.

— (2009). Paying for delay: Pharmaceutical settlement as a regulatory design problem. *New York University Law Review*, **1553**, 101–167.

— and Sampat, B. N. (2012). Evergreening, patent challenges, and effective market life in pharmaceuticals. *Journal of Health Economics*, **31**, 327–339.

Hennig-Schmidt, H., Sadrieh, A. and Rockenbach, B. (2010). In search of workers' real effort reciprocity – a field and a laboratory experiment. *Journal of the European Economic Association*, **84** (4), 817–837.

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

Hoi, A. A. and Littlejohn, G. (2005). Leflunomide. In R. O. Day, D. E. Fürst, P. L. C. M. van Riel and B. Bresnihan (eds.), *Antirheumatic Therapy: Actions and Outcomes Progress in Inflammation Research*, Birkhäuser Basel, pp. 199–219.

Imbens, G. W. and Wooldridge, J. M. (2007). What's new in econometrics? control function and related methods. *NBER Lectures*, **1**, 1–27.

Jacobo-Rubio, R. (2014). The private value of entry and deterrence in the us pharmaceutical industry. *Working paper*.

Katz, L. (1986). A behavioral account of the labor market: The role of fairness concerns. *Working paper*.

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

Kesselheim, A. S., Wang, B. and Avorn, J. (2013). Defining "innovativeness" in drug development: A systematic review. *Clinical Pharmacology & Therapeutics*, **94** (3), 336–348.

Kessler, J. (2013). When will there be gift exchange? addressing the lab-field debate with a laboratory gift exchange experiment. *Working paper*.

Keyhani, S., Diener-West, M. and Powe, N. (2006). Are development times for pharmaceuticals increasing or decreasing? *Health Affairs*, **25** (2), 461–8.

Klemperer, P. (1990). How broad should the scope of patent protection be? *RAND Journal of Economics*, **21** (1), 113–130.

Kremer, M. (1998). Patent buyouts: A mechanism for encouraging innovation. *Quarterly Journal of Economics*, **113** (4), 1137–1167.

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

Kube, S., Maréchal, M. A. and Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review*, **102** (4), 1644–1662.

Lahno, A. M. and Serra-Garcia, M. (2014). Peer effects in risk taking. *CESifo Working Paper Series No. 4057*.

Lanthier, M., Miller, K. L., Nardinelli, C. and Woodcock, J. (2013). An improved approach to measuring drug innovation finds steady rates of first-in-class oharmaceuticals, 1987-2011. *Health Affairs*, **32** (8), 1433–1439.

Lerner, J. (2002). Patent protection and innovation over 150 years. *NBER Working Paper 8977*.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world. *Journal of Economic Perspectives*, **21** (2), 153–174.

Lichtenberg, F. R. (2007). The impact of new drugs on us longevity and medical expenditure, 1990-2003: Evidence from longitudinal, disease-level data. *American Economic Review*, **97** (2), 438–443.

— and Philipson, T. J. (2002). The dual effects of intellectual property regulations: Within- and between- patent competition in the us pharmaceuticals industry. *NBER Working Paper 9303*.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, **60** (3), 531–542.

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

Montgomery, M. R. and Casterline, J. B. (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.

Moser, P. (2005). How do patent laws influence innovation? evidence from nineteenth-century world fairs. *American Economic Review*, **95** (1214–1236).

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 Motiere, L. (1994). The market for first-run motion pictures. *mimeo*.

Munshi, K. and Myaux, J. (2006). Social norms and the fertility transition. *Journal of Development Economics*, **80** (1), 1–38.

Murphy, K. M. and Topel, R. H. (1985). Estimation and inference in two step econometric models. *Journal of Business and Economic Statistics*, **3** (4), 370–379.

Nelson, R. A., Donihue, M. R., Waldman, D. M. and Wheaton, C. (2001). What's an oscar worth? *Economic Inquiry*, **39** (1), 1–6.

Newhouse, J. P. and the Insurance Experiment Group (1993). *Free for all? Lessons from the RAND Health Insurance Experiment*. Harvard University Press.

Nordhaus, W. D. (1967). *The Optimal Life of a Patent*. Cowles Foundation Discussion Papers 241, Cowles Foundation for Research in Economics, Yale University.

— (1969). An economic theory of technical change. *American Economic Review*, **59** (2), 18–28.

— (1972). The optimal life of a patent: Reply. *American Economic Review*, **62** (3), 428–431.

NSF (2012). *Science and Engineering Indicators 2012*. National Science Foundation.

O'Donoghue, T., Scotchmer, S. and Thisse, J.-F. (1998). Patent breadth, patent life, and the pace of technological progress. *Journal of Economics & Management Strategy*, **7** (1), 1–32.

Oh, J. and O'Connor, P. W. (2013). An update of teriflunomide for treatment of multiple sclerosis. *Therapeutics and Clinical Risk Management*, **9**, 177–90.

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

Pammolli, F., Magazzini, L. and Riccaboni, M. (2011). The productivity crisis in pharmaceutical r&d. *Nature Reviews Drug Discovery*, **10** (6), 428–38.

Panattoni, L. E. (2011). The effect of paragraph iv decisions and generic entry before patent expiration on brand pharmaceutical firms. *Journal of Health Economics*, **30** (1), 126–45.

Parker, I. (2013). The big sleep. *The New Yorker*, **December 9**.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, **110** (4), 705–729.

Pillutla, M., Malhotra, D. and Murnigha, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, **39** (5), 488–455.

Prag, J. and Casavant, J. (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.

Reiffen, D. and Ward, M. R. (2005). Generic drug industry dynamics. *The Review of Economics and Statistics*, **87** (1), 37–49.

Ringel, J. S., Hosek, S. D., Vollaard, B. A. and Mahnovski, S. (2002). *The Elasticity of Demand for Health Care: A Review of the Literature and its Application to the Military Health System*. Tech. rep., RAND National Defense Research Institute.

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, **70** (4), 1341–1378.

— (2003). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, **70** (4), 1341–1378.

Sakakibara, M. and Branstetter, L. (2001). Do stronger patents induce more innovation? evidence from the 1998 Japanese patent law reforms. *RAND Journal of Economics*, **32** (1), 77–100.

Sampat, B. N. and Williams, H. (2014). How do patents affect follow-on innovation: Evidence from the human genome. *Working paper*.

Sawhney, M. S. and Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, **15** (2), 113–131.

Scannell, J. W., Blankley, A., Boldon, H. and Warrington, B. (2012). Diagnosing the decline in pharmaceutical r&d efficiency. *Nature Reviews Drug Discovery*, **11** (March), 191–200.

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

Scherer, F. M. (1972). The optimal life of a patent. *American Economic Review*, **62** (3), 422–427.

Scotchmer, S. (1991). Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives*, **5**, 29–41.

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

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

Stephens-Davidowitz, S. (2013a). Unreported victims of an economic downtown. *mimeo*.

— (2013b). Who will vote? ask google. *mimeo*.

— (2014). The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Labor Economics*, **119**, 26–40.

—, Varian, H. and Smith, M. D. (2014). Super returns? the effects of ads on product demand. *mimeo*.

Swami, S., Eliashberg, J. and Weinberg, C. B. (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.

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.

WHO (2004). *The World Medicines Situation*. World Health Organization.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

WSJ (1998). Fda advisory panel recommends approving hoechst arthritis drug. *The Wall Street Journal*, **August 7**.

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

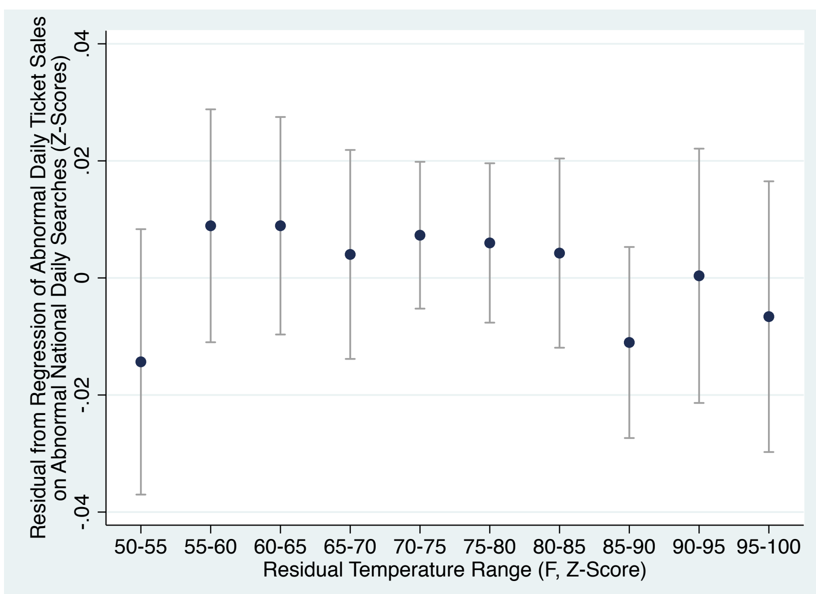
Zufryden, F. S. (1996). Linking advertising to box office performance of new film releases: A marketing planning model. *Journal of Advertising Research*, **36**, 29–42.

Appendix A

Appendix to Chapter 1

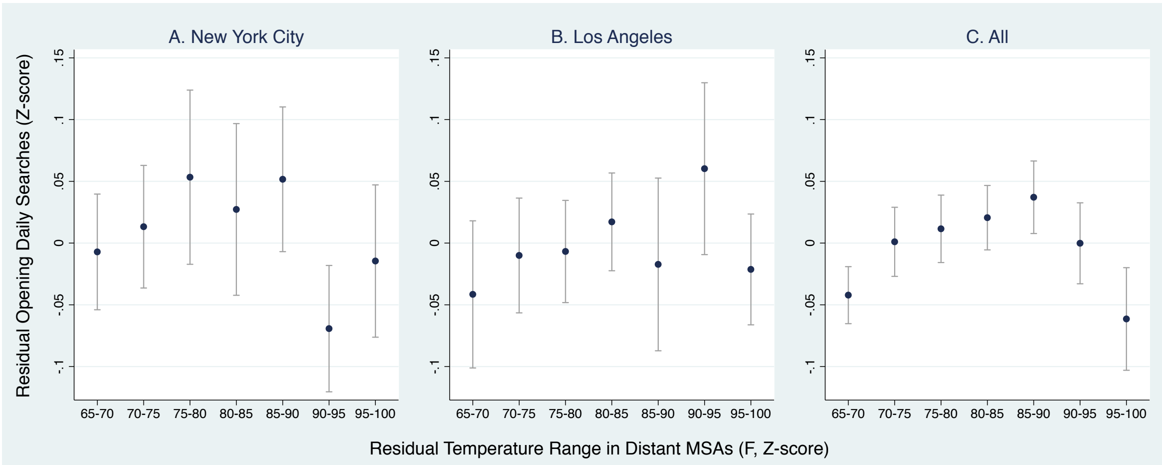
A.1 Supplementary Figures

Figure A.1: *Ticket Sales, National Searches, and the Weather*



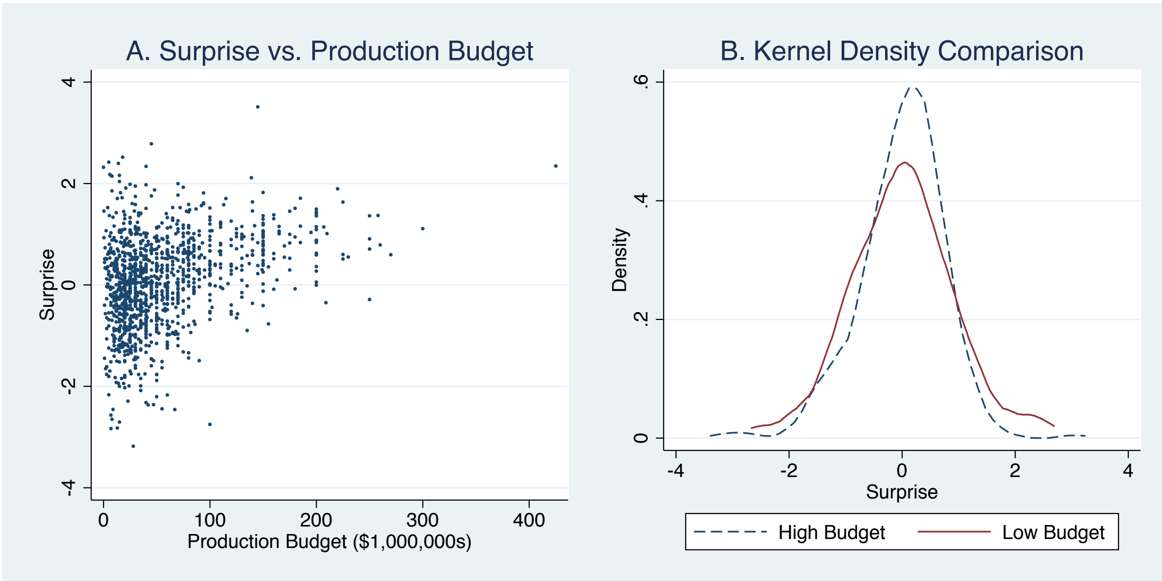
Notes: We analyze how the relationship between ticket sales and national searches varies with the weather to test our assumption that the relationship between searches and sales is not confounded by the weather, i.e., that weather shocks do not change how search behavior and ticket sales relate. First, we regress abnormal national daily ticket sales on abnormal national daily searches (by movie). We then regress the residual from this regression on each listed weather shock, and plot the coefficient along with the corresponding 95% confidence interval. Ticket sales, search data, and weather measures are as described in the text, and all are converted into Z-scores. Each plotted coefficient is from a separate regression. Observations are at the date by movie level (8,040 observations).

Figure A.2: *The Effect of Weather Shocks Elsewhere on Local Viewership*



Notes: This figure is the analog to Figure 1.6, but analyzes the relationship between local viewership and weather shocks occurring in the MSAs more than 1,000 km away. Each panel plots the coefficients of regressions of abnormal local Google searches on each listed weather shock, along with the corresponding 95% confidence intervals, controlling for local searches in the MSAs more than 1,000 km away as well as the same set of weather controls used in the national analysis (temperature in 10 degree increments, precipitation dummies in quarter-inch increments, snow, and rain dummies). Weather shocks are measured as the Z-score of the residual of the indicator for the MSA in each temperature range, and abnormal local searches are as described in the text. Panel A corresponds to just New York, Panel B to just Los Angeles, and Panel C to all five MSAs in our sample. Observations are at the movie by date by MSA level (576 observations in Panel A, 480 in Panel B, 2064 in Panel C).

Figure A.3: *Uncertainty by Production Budget*



Notes: This figure visualizes the level of quality uncertainty signaled by a film's production budget. Panel A plots surprise, which is the residual from a regression of the log of total sales in weekends subsequent to opening on the log of opening weekend theaters, against production budget. Panel B plots the kernel densities of surprise separately for movies in the top (in excess of $48M) and bottom (below $29M) terciles of production budget. A two-sample Kolmogorov-Smirnov test shows the difference in distribution of surprise between top and bottom terciles is statistically significant at the 10% level (p-value=0.091). The sample is the 1,228 movies for which we have production budgets.

A.2 Supplementary Tables

Table A.1: *LASSO and Instrument Robustness Checks*

A. LASSO Instrument Selection

Set of Potentials, Count Constraint	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
5 Degree, Choose 1	0.474***	0.269***	0.188***	0.112***	0.0960***	1.139***
	(0.0474)	(0.0360)	(0.0287)	(0.0203)	(0.0162)	(0.131)
5 Degree, Choose 2	0.524***	0.308***	0.174***	0.115***	0.0859***	1.206***
	(0.0411)	(0.0346)	(0.0248)	(0.0162)	(0.0124)	(0.113)
5 Degree, Choose 3	0.549***	0.313***	0.180***	0.127***	0.0914***	1.261***
	(0.0405)	(0.0329)	(0.0237)	(0.0166)	(0.0123)	(0.110)
10 Degree, Choose 1	0.627***	0.353***	0.216***	0.113***	0.0986***	1.408***
	(0.0878)	(0.0623)	(0.0488)	(0.0305)	(0.0256)	(0.228)

B. Hand-Selected Instruments

Instrument(s)	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
$(\text{Temperature} - 75)^2$ x	0.547***	0.308***	0.126***	0.0794***	0.0402**	1.101***
(abs(Temperature - 75) $\leq$ 20)	(0.0509)	(0.0401)	(0.0290)	(0.0204)	(0.0171)	(0.128)
All Instruments Provided	0.475***	0.269***	0.164***	0.121***	0.0932***	1.122***
to LASSO in Base Case	(0.0242)	(0.0223)	(0.0167)	(0.0132)	(0.0103)	(0.0739)

Notes: This table presents second stage results for a variety of LASSO-IV and IV specifications. Panel A varies the parameters provided to LASSO. 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 a maximum of one, two, or three instruments, respectively. In the fourth specification, a single instrument is again chosen, but the instrument choice is altered; temperature indicators are in 10 degree F increments. In Panel B, the instruments are not machine-chosen. The first specification in Panel B presents results where the instrument is specified to be the abnormal squared difference between temperature and 75 degrees multiplied by a dummy to ensure that temperatures are within 20 degrees of 75. Finally, the second specification in Panel B presents results obtained by including in the first stage all of the potential instruments which are normally provided to LASSO. First stage F-statistics for the specifications in Panel A are presented in Table 1.1; the F-statistic for the first specification in panel B is 29.74 while the F-statistic for the second specification is 3.80.

Table A.2: *Momentum from Viewership Shocks, Robustness Checks*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
			A. with Observations Clustered at the Weekend Level			
IV	0.474***	0.269***	0.188***	0.112***	0.0960***	1.139***
	(0.0617)	(0.0517)	(0.0423)	(0.0315)	(0.0247)	(0.188)
OLS	0.423***	0.235***	0.140***	0.0888***	0.0630***	0.950***
	(0.0164)	(0.0128)	(0.00869)	(0.00637)	(0.00475)	(0.0443)
R-squared	0.653	0.498	0.357	0.301	0.264	0.570
			B. with Observations at the Weekend Level			
IV	0.471***	0.275***	0.178***	0.111***	0.102***	1.137***
	(0.0756)	(0.0582)	(0.0464)	(0.0347)	(0.0277)	(0.208)
OLS	0.425***	0.241***	0.137***	0.0900***	0.0633***	0.956***
	(0.0180)	(0.0127)	(0.00928)	(0.00689)	(0.00531)	(0.0457)
R-squared	0.647	0.513	0.352	0.295	0.257	0.566
			C. with Inclusion of Second Stage Weather Controls in First Stage			
IV	0.491***	0.271***	0.184***	0.112***	0.0970***	
	(0.0469)	(0.0346)	(0.0271)	(0.0201)	(0.0164)	
First Stage						
75-80F	-3.014***	-3.122***	-3.183***	-3.133***	-3.071***	
	(0.496)	(0.483)	(0.493)	(0.503)	(0.496)	
F-Stat	36.92	41.87	41.70	38.72	38.32	

Notes: Panel A replicates Table 1.2 but with observations clustered at the weekend level, and Panel B replicates Table 1.2 but with observations defined at the opening weekend by weekend level (556 observations). In each case, the first stage results are included in Appendix Table A.3. Panel C replicates Table 1.2 but with the addition of second stage "contemporaneous" weather controls in the first stage. For example, when the second stage outcome variable is ticket sales in Week 2, Week 2 contemporaneous weather controls (as defined in 25) are included also as controls in the first stage. The corresponding first stage coefficients and F-statistics are included in the final rows.

Table A.3: *Additional First Stages*

Sample	Instrument	Coefficient	F-Stat
Controling for Opening Theaters	75-80F	-2.728*** (0.470)	33.74
Weeks 1 through 5	75-80F	-2.763*** (0.479)	33.3
Includes Truncated	75-80F	-2.941*** (0.486)	36.70
Top-1000 High-Rated	80-85F	-3.569*** -0.873	16.70
Top-1000 Low-Rated	55-60F	-1.803*** (0.529)	11.61
High Production Budget	90 -95F	3.971*** (0.947)	17.58
Low Production Budget	95-100F	4.602*** (0.924)	24.82
Main Sample, Clustered at Weekend Level	75-80F	-3.041*** (0.785)	15.01
Main Sample, Obervations at Weekend Level	75-80F	-8.597*** (2.353)	13.35
Child-Friendly	0-0.25 in	-1.943*** (0.665)	8.544
Adults-Only	75-80F	-2.618*** (0.492)	28.37

Notes: This table presents first stage results from all additional IV specifications in the paper along with the corresponding F-statistic. 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]).

Table A.4: *Local First Stages*

	First Stage Outcome	Local Instrument: 85-90F	National Instrument: 75-80F	F-Stat
Panel A	Local Searches	-0.118***		19.05
		(0.0271)		
	Local Searches, Controlling for Searches in Weekend Prior to Opening	-0.0974***		21.57
		(0.0210)		
	Local Searches, Controlling for Searches in Weekend Prior to Opening and Searches in Distant MSAs	-0.0572***		20.38
		(0.0127)		
Panel B	Local Searches	-0.131***	-0.316***	19.47
		(0.0275)	(0.0707)	
	National Searches	-0.0841	-0.649***	11.38
		(0.0619)	(0.137)	
Panel C	Local Searches	-0.105***	-0.197***	19.24
		(0.0216)	(0.0483)	
	National Searches	-0.0449	-0.465***	9.057
		(0.0518)	(0.110)	

Notes: This table presents first stage results from the local IV specifications in Table 1.4 along with the corresponding F-stat. In each case, the instrument of choice was chosen with LASSO methods described in the text from the following choice set: MSA-level 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]).

Table A.5: *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	0.416***	0.220***	0.128***	0.0765***	0.0492***	0.891***
(obs. 705)	(0.0324)	(0.0203)	(0.0128)	(0.00870)	(0.00615)	(0.0766)
R-squared	0.729	0.574	0.432	0.308	0.195	0.612
Low-Rated	0.412***	0.229***	0.142***	0.0777***	0.0447***	0.905***
(obs. 825)	(0.0151)	(0.0124)	(0.00960)	(0.00605)	(0.00339)	(0.0430)
R-squared	0.742	0.596	0.494	0.380	0.348	0.657
			B. by Information about Movie Quality			
High Budget	0.391***	0.215***	0.126***	0.0824***	0.0575***	0.872***
(obs. 744)	(0.0268)	(0.0185)	(0.0120)	(0.00875)	(0.00645)	(0.0694)
R-squared	0.676	0.517	0.418	0.344	0.276	0.579
Low Budget	0.427***	0.257***	0.143***	0.0767***	0.0374***	0.941***
(obs. 705)	(0.0193)	(0.0146)	(0.00945)	(0.00700)	(0.00464)	(0.0512)
R-squared	0.715	0.561	0.419	0.240	0.094	0.580

Notes: This table presents the OLS results that correspond to Table 1.5.

Table A.6: *Opening Weekend Viewership Shocks and Ratings*

	IV Estimates	OLS Estimates
Number of Votes	-10.87	1.385
	(16.48)	(2.311)
R-Squared	--	0.001
High Rating	-0.0364	0.0103***
	(0.0276)	(0.00343)
R-Squared	--	0.016
Low Rating	0.00831	-0.0155***
	(0.0264)	(0.00366)
R-Squared	--	0.031
Difference:		
(High - Low)	-0.04471	0.025***
	(0.0382)	(0.005)

Notes: This table shows the relationship between opening weekend sales (in millions) and the film's number of voters and likelihood of being high rated (top third) and low rated (bottom third). Both 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. The first stage results are included in Appendix Table A.3.

Table A.7: *OLS Estimates of Local Momentum*

	Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
	A. Effect of Local Opening Searches on Subsequent Local Searches					
Local Searches	0.327***	0.173***	0.118***	0.0830***	0.0662***	0.768***
	(0.0247)	(0.0176)	(0.0122)	(0.00764)	(0.00564)	(0.0666)
Local Searches, Controlling for Local Searches in Weekend Prior to Opening	0.299***	0.163***	0.112***	0.0747***	0.0568***	0.706***
	(0.0306)	(0.0219)	(0.0149)	(0.00911)	(0.00649)	(0.0811)
Local Searches, Controlling for Local Searches in Weekend Prior to Opening and Local Searches in Distant MSAs	0.324***	0.180***	0.123***	0.0794***	0.0640***	0.770***
	(0.0203)	(0.0166)	(0.0102)	(0.00761)	(0.00586)	(0.0542)
	B. Effect of Local and National Opening Searches on Subsequent Local Searches					
Local Searches	0.281***	0.148***	0.104***	0.0721***	0.0596***	0.665***
	(0.0343)	(0.0206)	(0.0140)	(0.00987)	(0.00667)	(0.0822)
National Searches	0.0326	0.0175	0.0101	0.00776	0.00463	0.0726
	(0.0249)	(0.0139)	(0.0101)	(0.00665)	(0.00471)	(0.0586)
	C. Effect of Local and National Opening Searches on Subsequent Local Searches, Controlling for Local Searches in Weekend Prior to Opening Weekend					
Local Searches	0.261***	0.141***	0.0996***	0.0657***	0.0521***	0.619***
	(0.0418)	(0.0244)	(0.0166)	(0.0111)	(0.00761)	(0.0984)
National Searches	0.0299	0.0166	0.00950	0.00694	0.00365	0.0666
	(0.0235)	(0.0136)	(0.00996)	(0.00643)	(0.00448)	(0.0563)

Notes: This table presents the OLS results that correspond to Table 1.4.

A.3 Holiday Controls

Our holiday indicators are exactly those of Dahl and DellaVigna (2009) and are similarly motivated by the fact that (1) holidays impact movie audience sizes (usually positively), (2) the effect varies across holidays, and (3) 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 Years (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.

A.4 LASSO Method Overview

In this section, we provide a brief overview of our LASSO method, which draws heavily on Chernozhukov and Hansen (2013).

Consider our single endogenous variable, $v\_abn_{t,1}$, and our large set of potential weather instruments, $W_t = [w\_shock_{t,1}, \ldots, w\_shock_{t,p}]$. In order to avoid having a weak first stage, the econometrician would ideally select instruments by solving the standard ordinary least 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:

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

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

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

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.[1] Bickel *et al.* (2009) show that the rate-optimal choice of the penalty parameter λ is

$$\lambda = 2\sigma\sqrt{2\log(pn)/n}, \tag{A.4}$$

where p is the number of potential instruments, n is the sample size, and σ is the standard deviation of the residuals. σ is not known a-priori, so we estimate σ following the iterative methods of Belloni *et al.* (2012). We denote the final output of the LASSO methodology by W^{LASSO}; this is the machine-chosen instrument set.

A.5 Reconciling Results with Moretti (2011)

In this section, we reconcile our findings with Moretti (2011), which tells a story about social learning in movie-going. Moretti develops a social learning model with four key empirical predictions, and then shows that each broadly holds true in the data. His analysis focuses on how opening weekend "surprise", defined as the residual of a regression of log

[1] 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

$$||\hat{\beta}_{LASSO} - \beta_0|| \leq \sigma\sqrt{s\log(n \vee p)/n}, \tag{A.3}$$

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

opening weekend sales on log opening weekend screens, is related to future sales. In the penultimate section, he notes that network externalities can in principle generate each of the four pieces of evidence he presents, and that they are difficult to rule out completely. To bolster his case for social learning, he then presents a test for network externalities that is very similar in spirit to our analysis: he instruments for opening weekend “surprise" using opening weekend weather and analyzes how instrumented surprise affects future sales. While we estimate a strong, persistent social multiplier, Moretti (2011) estimates an imprecise multiplier that is much smaller than his main, non-instrumented, effect. His instrumented multiplier sometimes even changes sign across specifications, and he infers that network externalities are unlikely to explain the large effects he documents in his main analysis.

Moretti has kindly shared his code and data with us, which has facilitated our analysis of the differences between his work and ours. The main differences are (1) his instruments, (2) his data (he focuses on the years 1982-2000, while our data covers 2002-2012), and (3) his functional form. In this section, we show that the difference in findings, however, is driven almost exclusively by the difference in functional form. In particular, adding season controls to his specification – controls we argue are only natural given that weather is highly seasonal – reverses his result and suggests that network externalities generate most, if not all, of the estimated effect.

Our analysis comparing our work to Moretti (2011) proceeds in the following three steps. First, maintaining our functional form and our data, we show that, although our LASSO-chosen instruments produce a much stronger first stage than the hand-selected instruments in Moretti (2011), we can replicate our main results using Moretti’s instruments. This suggests the difference in instruments does not account for the difference in findings.[2] Second, maintaining our functional form, but using both the data and instruments in Moretti

[2] At the time of our initial writing, we did not have Moretti’s code or data and did not attempt to replicate Moretti’s work directly. Using LASSO to select weather instrument allowed us to estimate a much stronger first stage than Moretti, and we posited that the differences in our findings could be due different instrument sets. However, implementing our methods on our data with the instruments in Moretti (2011) in fact produces little change in our main estimates.

(2011), we can again replicate our main results. This suggests the difference in data also does not account for the difference in findings.

Third, we compare our functional form to that used by Moretti. Altough it is not evident from the text of Moretti (2011), Moretti's code reveals that the model used in the IV estimation does not correspond to that used in the OLS estimation. In particular, it does not include movie fixed effects and so does not include controls for seasonality.[3] This omission is particularly important given seasonal variation in weather (the instrument). We show that the addition of seasonal fixed effects (the same rich set of seasonal controls we use in our base analysis) produces alignment between Moretti's IV and OLS results, suggesting the multiplier observed in Moretti (2011) may in fact be driven largely by network externalities.

A.5.1 Our Framework and Data, Moretti's Instruments

We begin this exercise by replicating our main analysis using Moretti's instruments. Before our original submission, we had not implemented Moretti's set of instruments in our framework because we found in early exploratory analysis that similar sets of instruments yielded first stage F-statistics on the order of 3 or 4, which we felt were too low to use in a 2SLS design. Our prior was thus that Moretti's results appeared to be different due to his many weak instruments. As we discuss Section 1.3.3, this motivated our decision to implement LASSO, and at the time we did not attempt to replicate our IV-LASSO analysis using non-machine selected instruments.[4]

Moretti's instruments are a set of four weather measures (maximum temperature, minimum temperature, precipitation, and snowfall) for each of opening day and the day before opening in seven metropolitan areas (New York, Boston, Chicago, Denver, Atlanta, Kansas City, and Detroit). He implements four separate specifications: (1) just the maximum and minimum temperature, specified linearly, (2) just the maximum and minimum temperature,

[3]The omission of movie fixed effects should come as no surprise since they cannot be included if opening weekend weather is used as an instrument.

[4]We have since done so, however, and Section 1.4.1 shows that hand-selected instrument sets yield highly similar results.

specified as a quadratic, (3) all four weather measures, specified linearly, and (4) all four weather measures, specified as a quadratic (see page 386 of Moretti (2011) for more detail).

Panel A of Appendix Table A.8 presents the replication of our main analysis using Moretti (2011)'s instrument sets. The top row shows our base case result, using our LASSO-selected weather instrument; the following four rows show the same IV analysis implemented using each of Moretti's four instruments sets. The F-statistic from Moretti's first instrument set is relatively low, but climbs as we add more weather measures and include the quadratic terms. While the point estimates are impacted somewhat by the choice of instruments, they are in all cases quite positive and, with the exception of Weeks 5 and 6 using the weakest instrument set, significantly different from zero. The magnitudes, moreover, are not dissimilar to those in our own base case results. Taken together, these results suggest that the differences in instrument choices between our paper and Moretti (2011)'s is not driving the divergence of results.

A.5.2 Our Framework, Moretti's Data and Instruments

Next, we conduct the same analysis but using Moretti's data instead of our own. Moretti's data differs from ours in that it covers an earlier period and is at the weekend (as opposed to daily) level.[5] As in our main analysis, we condition abnormal sales in weekends subsequent to opening on that weekend's own weather; since these data are at the weekend level we condition on weather for Friday, Saturday, and Sunday of that weekend using the same measures as in Section 1.4.1.[6]

We present our results in Panel B of Appendix Table A.8. As in Panel A, each row shows regression estimates generated using each of Moretti's four instrument sets. The F-statistics for these specifications are quite low (just above 1), and the estimated coefficients are larger than those estimated on our sample, but the qualitative results remain quite

[5] In the earlier period, daily level data was generally not made available. We have no evidence, however, that weekend-level data should be problematic; recall that we show in Section 1.4.1 that our main results are robust to estimation at the weekend level.

[6] The results are largely unchanged if we omit contemporaneous weather controls.

similar: weather-induced shocks to viewership opening weekend yield more viewership in subsequent weekends. Thus, implementing our methods on Moretti's sample with Moretti's instruments does not change our main conclusions.

A.5.3 Probing Moretti's Framework

Finally, we replicate Moretti (2011)'s full network hypothesis test on his data.

The IV model estimated in Moretti (2011) (Section 7 and Table 9 in Moretti (2011)) differs in a couple of important ways from ours. For simplicity of comparison, we adopt Moretti's notation here. Letting j index movies, Moretti estimates movie j's "surprise" S_j by regressing the log of movie j's opening weekend sales on movie j's log number of opening weekend screens. He calls the resulting residual, representing sales not predicted by screens, surprise. In some regressions, he also makes surprise a binary variable which is 1 if $S_j > 0$ and 0 otherwise.

In his main analysis, Moretti regresses the log of sales in weekend t on a linear time trend in weekend, the interaction of the surprise and the time trend, and a movie fixed effect:[7]

$$\log(y_{jt}) = \alpha_0 + \alpha_1 t + \alpha_2 (t \times S_j) + d_j + u_{jt}. \tag{A.5}$$

The coefficient of interest is α_2, the coefficient on the interaction of surprise and the time trend. This model has the flavor of a difference-in-difference framework: α_2 represents the extent to which the rate of decline in sales in future weekends is moderated by opening weekend surprise. Moretti shows in a range of OLS specifications that α_2 is robustly positive. This is his paper's main empirical result, which he interprets as evidence of learning behavior. We note that the framework forces log sales to be linear in the interaction of time and surprise. That is, sales are assumed to be exponential in the interaction of time and surprise.

With this framework in hand, Moretti describes his test for network externalities as one in which he estimates Appendix Equation A.5 by 2SLS, instrumenting for S_j with

[7] Equation 15, on page 372 of Moretti (2011); results presented in Table 4 of Moretti (2011).

the weather measures described above. Notice, however, that the movie fixed effects in Appendix Equation A.5 cannot be included in a standard IV model where surprise is instrumented with opening weekend weather because opening weekend weather varies at the movie level. Although this is not made explicit in the text or tables of Moretti (2011), the code and log files Moretti kindly shared with us reveal that the IV model in fact drops the movie fixed effects. The code and log files also reveal that the IV model additionally includes a linear surprise covariate, S_j:[8]

$$\log(y_{jt}) = \beta_0 + \beta_1 t + \beta_2 S_j + \beta_3 (t \times S_j) + v_{jt}. \tag{A.6}$$

Put otherwise, while Moretti's main specifications are within-movie estimates, Moretti's IV estimates are derived from a regression that includes no movie fixed effects. In fact, aside from the "surprise" variable, no movie characteristics are included at all. Of particular note, this model does not include controls for seasonality.

Panel A of Appendix Table A.9 shows Moretti's estimates of Appendix Equation A.6 (also found in Table 9 of Moretti (2011)). Specification (1) presents the OLS estimates, and specifications (2)-(5) present the IV estimates with each of his four different instrument sets. Specifications (2) and (3) use only the maximum and minimum temperature as instruments, either linearly or as a quadratic, respectively; specifications (4) and (5) expand the instrument set to also include precipitation and snowfall, again either linearly or as a quadratic, respectively. Each coefficient results from a separate regression: the top row analyzes the coefficient on surprise interacted with time since opening, while the second row analyzes the coefficient on an indicator for positive surprise interacted with time since opening. Moretti's basic network externalities result is contained here: the estimated IV coefficients are significantly smaller than the OLS coefficients and the coefficients switch sign across specifications.

For exposition, we add specification (0) in the left-most column. This column reports

[8]Moretti estimates Appendix Equation A.6 by simultaneously instrumenting for both S_j and $t \times S_j$ using his chosen set of instruments along with each instrument's interaction with t.

Moretti's estimates of Appendix Equation A.5, i.e., his main OLS regression, which includes movie fixed effects. Comparing specification (1) with specification (0) suggests that movie fixed effects do little to change the OLS coefficients. From this the reader might infer that controlling for movie characteristics and seasonal variation is of little importance. However, while this may be in an OLS specification, it is not necessarily true when instrumenting for surprise with weather, especially since weather itself is highly seasonal. Moreover, there is no reason not to include seasonal controls in the cross-sectional IV regressions; at the very least, doing so should improve the precision of the results. (As we show, controlling for seasonality also changes the results and implications meaningfully.)

Panel B shows the results from estimating Appendix Equation A.6 with a set of seasonal controls. For each week of showing t, we include week of year, year, and holiday fixed effects.[9] The resulting OLS estimates change barely at all, as expected since comparison of specification (0) with (1) in Panel A reveals that movie fixed effects have little bearing on the estimates. However, inclusion of seasonality controls moves the IV estimates markedly. With seasonal controls, the IV estimates are now all positive, significant in seven of the eight specifications, and quite closely aligned with the corresponding OLS estimates.

In sum, the differences between ours results and those in Moretti (2011) are generated by the specifics of his estimation strategy. We have shown that implementing our framework on our data with the instrument sets used in Moretti (2011) produces results largely in line with our base case estimates, and that implementing our framework on Moretti's data using his instruments produces estimates that are larger in magnitude but also generally in line with our main results. Probing the specifics of Moretti's framework on his data and using his instruments, we find that his rejection of network externalities is reversed when seasonal controls are included, an arguably very natural modification given the importance of seasonality to both movie-going and weather.

---

[9] In other words, our fixed effects are week of year by t, year by t, and holiday by t. This is precisely the same set of controls we use in our main analysis, where we separately conditioned sales (and weather) in each week of showing t on our set of seasonal dummies.

Table A.8: *Our Framework, Moretti's Instruments*

Panel A. Estimating our Model on our Data with Moretti's Instruments

Instrument(s)	First-Stage F-Statistic	Estimated Effect of Opening Weekend Ticket Sales on Ticket Sales in Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
LASSO-selected	38.80	0.474***	0.269***	0.188***	0.112***	0.0960***	1.139***
		(0.0474)	(0.0360)	(0.0287)	(0.0203)	(0.0162)	(0.131)
Min and max temp in 7 cities for Fri and Sat of opening weekend, linear	1.910	0.332***	0.127***	0.0919***	0.0233	0.0180	0.592***
		(0.0453)	(0.0362)	(0.0261)	(0.0203)	(0.0152)	(0.125)
Min and max temp in 7 cities for Fri and Sat of opening weekend, quadratic	2.886	0.403***	0.195***	0.123***	0.0585***	0.0426***	0.822***
		(0.0323)	(0.0243)	(0.0179)	(0.0138)	(0.0117)	(0.0826)
Min temp, max temp, precipitation, and snowfall in 7 cities for Fri and Sat of opening weekend, linear	4.158	0.407***	0.224***	0.131***	0.0851***	0.0597***	0.907***
		(0.0248)	(0.0183)	(0.0143)	(0.0118)	(0.00903)	(0.0668)
Min temp, max temp, precipitation, and snowfall in 7 cities for Fri and Sat of opening weekend, quadratic	6.487	0.444***	0.239***	0.133***	0.0888***	0.0602***	0.965***
		(0.0176)	(0.0127)	(0.0100)	(0.00802)	(0.00616)	(0.0451)

Panel B. Estimating our Model on Moretti's Data with Moretti's Instruments

Instrument(s)	First-Stage F-Statistic	Estimated Effect of Opening Weekend Sales on Sales in Week 2	Week 3	Week 4	Week 5	Week 6	Weeks 2 - 6
Min and max temp in 7 cities for Fri and Sat of opening weekend, linear	1.029	0.689***	0.504***	0.360***	0.261***	0.202***	2.063***
		(0.110)	(0.0867)	(0.0835)	(0.0734)	(0.0596)	(0.360)
Min and max temp in 7 cities for Fri and Sat of opening weekend, quadratic	1.066	0.642***	0.479***	0.320***	0.237***	0.177***	1.890***
		(0.0721)	(0.0650)	(0.0601)	(0.0515)	(0.0410)	(0.244)
Min temp, max temp, precipitation, and snowfall in 7 cities for Fri and Sat of opening weekend, linear	1.192	0.635***	0.381***	0.300***	0.221***	0.158***	1.715***
		(0.0680)	(0.0619)	(0.0635)	(0.0524)	(0.0432)	(0.244)
Min temp, max temp, precipitation, and snowfall in 7 cities for Fri and Sat of opening weekend, quadratic	1.437	0.605***	0.419***	0.285***	0.205***	0.154***	1.712***
		(0.0458)	(0.0462)	(0.0454)	(0.0365)	(0.0301)	(0.172)

Notes: Panel A replicates our base case analysis on our sample using Moretti's instruments. The first row of estimates use our LASSO-selected instrument and are the same as our base case estimates found in Table 1.2, while the four rows that follow implement the different sets of instruments used by Moretti. Panel B then replicates our base case analysis on Moretti's sample using Moretti's instruments. Data are daily in Panel A (n=1,671) and weekly in Panel B (n=748). Standard errors are clustered by date in Panel A and by saturday date in Panel B. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A.9: *Robustness of Moretti's Test for Network Externalities*

	OLS, Main	OLS, NE Test	2SLS, NE Test			
			Instruments are minimum temperature, maximum temperature		Instruments are minimum temperature, maximum temperature, precipitation, snowfall	
A. Moretti's Published Estimates						
	(0)	(1)	(2)	(3)	(4)	(5)
t × surprise	0.463***	0.413***	–0.107	0.085	–0.039	0.118*
	(0.016)	(0.014)	(0.122)	(0.083)	(0.093)	(0.067)
t × positive surprise	0.616***	0.643***	–0.277	0.123	–0.129	0.181*
	(0.022)	(0.025)	(0.235)	(0.152)	(0.184)	(0.125)
F test: first-stage coefficient = 0			3.54	3.07	2.82	2.60
N	39,936	31,528	31,528	31,528	30,320	30,320
Quadratic in weather			N	Y	N	Y
B. Including Week of Year, Year, and Holiday Fixed Effects						
		(1)	(2)	(3)	(4)	(5)
t × surprise		0.463***	0.401**	0.595***	0.336**	0.467***
		(0.0153)	(0.200)	(0.137)	(0.135)	(0.0946)
t × positive surprise		0.693***	0.446	0.668***	0.482**	0.660***
		(0.0245)	(0.323)	(0.214)	(0.226)	(0.150)
F test: first-stage coefficient = 0			0.93	1.47	1.14	1.58
N		31,528	31,528	31,528	30,320	30,320
Quadratic in weather			N	Y	N	Y

Notes: This table presents results generated by first replicating and then probing Moretti (2011)'s IV analysis. First, Panel A presents Moretti (2011)'s published results. Specification 0 includes movie fixed effects and corresponds to Appendix Equation A.5, while the other specifications do not and correspond to Appendix Equation A.6. Panel B is the same as Panel A but adds seasonal controls, which are week of year, year, and holiday fixed effects included separately for each week of showing. Standard errors are clustered by movie throughout, and */**/*** denote significance at the 10%, 5%, and 1% levels, respectively.

A.6 A Role for Supply Shifts?

In this section we first present a brief overview of the supply side of the market and show that our estimated momentum is not driven by supplier responses. Our motivation for this test is to rule out a story in which opening weekend viewership shocks lead to increased movie availability, thereby decreasing the effective cost associated with attending that movie and making an increase in future sales mechanical in nature.

A.6.1 In-Theater Movie Supply: Institutional Background

We first provide some institutional context. 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.[10] Distributors and exhibitors contract on where the movie will be shown and how the revenues will be shared.[11] 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).[12]

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. On the other hand, having paid high fixed costs

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

[11]Most commonly, exhibitors pay upfront some advance to the distributor for the movie as well 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.

[12]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."

upfront, 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. 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.[13] Despite institutional factors that incentivize exhibitors to stay the course, there is some evidence (as in Krider *et al.* (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.

A.6.2 Testing for a Supply-Side Response

In the following analysis, we show that intensive and extensive margin responses to the viewership shocks we are identifying off are essentially non-existent. We then show that our estimated momentum is not substantively affected if we account for any such responses.

Panel A of Table A.10 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 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

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

weekend days).[14] Second, while the endogenous regressor continues to be abnormal viewership opening weekend, the outcome variable is abnormal number of screens (first row) or abnormal probability of being withdrawn (second row). (Each of these is similarly conditional for year, week of year, and holiday fixed effects.) Third, viewership is measured in 100,000'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 a movie opens. The Week 1 estimates suggest that movies with higher viewership due to weather on opening weekend also open on slightly (but insignificantly) more screens. One possible explanation is that consumers are responding to availability.[15] The point estimate, however, is statistically insignificant and small in magnitude; it suggests that 100,000 additional viewers from weather shocks opening weekend corresponds to about 1.5 additional screens. On average for movies in our sample, each screen has about 1,100 viewers per opening weekend; the additional screens, then, would mechanically explain less than 2.5% of the additional opening weekend viewings.

The coefficients in subsequent weekends indicate that exhibitors do respond to shocks to opening weekend viewership by changing the number of screens on which the movie shows. However, the magnitudes of the point estimates suggest that these intensive-margin adjustments can explain only a small fraction of the observed momentum. For example, movies that sold an additional 100,000 tickets opening weekend showed on just 11 additional screens second weekend; they also sold about 47,000 more tickets that weekend (see Table 1.2). Since the average tickets sold per screen in the second weekend is 600, the additional screens would mechanically account for 14% of the observed viewership effect that weekend. The relative size of the mechanical effects are similarly small in other weeks, ranging from

[14]For additional discussion of the supply change decision see Swami *et al.* (1999); any change in number of screen on which a movie is shown almost always occurs on the first day of the movie-industry week (Friday).

[15]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.

at most 18% (Week 5) to 11% (Week 3).

To examine any extensive-margin response from suppliers, we add truncated movies back into our sample. The reported IV estimates in the second row of Table A.10 show that the relationship between abnormal audiences (in 100,000's) and the abnormal probability in each week of being withdrawn from theaters is similarly weak. The estimates are all small and statistically insignificant, suggesting that weather-induced viewership shocks do not have substantive effects on withdrawal probabilities in the short run.

In Panels B and C, we show our main effects when accounting for supply adjustments. As expected given the relatively relatively weak relationships between viewership shocks opening weekend and both the number of screens showing the movie and the probability that the movie is dropped shown in Panel A, accounting for intensive and/or extensive-margin supply adjustments has little if any bearing on the estimated momentum.

Panel B accounts for intensive-margin adjustments by examining viewership per screen. 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 1.2. The second row shows the corresponding results when the outcome variable is instead defined as tickets per screen; and the third 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 rows of Panel B, we find that intensive-margin responses can explain little, if any, of the observed quantity effects. Overall, the point estimates from our main specification and from the per screen specification (scaled) differ by less than 15 percent. In some weeks the per screen results are just above and in other weeks just below, but 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 has little if any effect on estimated momentum. This is consistent with the finding in Panel A that movies with positive abnormal viewership opening weekend are not significantly less likely to be taken out of theaters. Moreover, accounting in this sample for intensive-margin responses increases estimated momentum slightly in some weeks and decreases it slightly in others. Taken together, the results in Table A.10 suggest that supply-side adjustments can explain little, if any, of the observed quantity effects.

Table A.10: *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	1.580	10.50*	8.117*	8.423**	7.340**	7.552***	41.94**
(obs. 557)	(6.112)	(5.867)	(4.612)	(3.834)	(3.180)	(2.431)	(16.36)
Probability Dropped	--	-0.000202	0.000413	0.00138	-0.000208	0.00110	0.00258
(obs. 568)	--	(0.000473)	(0.000828)	(0.00124)	(0.00150)	(0.00158)	(0.00472)
			B. Main Effects with Intensive-Margin Supply Adjustments (1,671 obs.)				
Tickets (1)	1.000***	0.474***	0.269***	0.188***	0.112***	0.0960***	1.139***
	(0.000)	(0.0474)	(0.0360)	(0.0287)	(0.0203)	(0.0162)	(0.131)
Tickets per Screen	1.019***	0.317***	0.244***	0.188***	0.130***	0.120***	0.999***
	(0.240)	(0.113)	(0.0713)	(0.0496)	(0.0396)	(0.0341)	(0.269)
Standardized Tickets per Screen (2)	1.000***	0.311***	0.239***	0.184***	0.127***	0.117***	0.980***
	(0.235)	(0.110)	(0.069)	(0.048)	(0.038)	(0.033)	0.263***
Difference:							
(1) - (2)	--	0.163	0.030	0.004	-0.016	-0.022	0.159
		(0.120)	(0.078)	(0.055)	(0.043)	(0.037)	(0.294)
			C. Main Effects with Supply Adjustments, Includes Dropped Movies (1,698 obs.)				
Tickets (3)	1***	0.463***	0.262***	0.188***	0.111***	0.0972***	1.125***
	(0)	(0.0485)	(0.0375)	(0.0304)	(0.0218)	(0.0167)	(0.139)
Tickets per Screen	0.747***	0.178	0.177**	0.163***	0.108***	0.124***	0.744***
	(0.231)	(0.118)	(0.0736)	(0.0511)	(0.0407)	(0.0341)	(0.279)
Standardized Tickets per Screen (4)	1.000***	0.238	0.236**	0.218***	0.144***	0.165***	0.995***
	(0.309)	(0.157)	(0.097)	(0.068)	(0.053)	(0.045)	(0.373)
Difference:							
(3) - (4)	--	0.225	0.025	-0.030	-0.034	-0.069	0.129
		(0.165)	(0.104)	(0.074)	(0.057)	(0.048)	(0.398)

Notes: The first row of Panel A reports the results of IV regressions of abnormal number of screens showing the movie each week in our main sample on abnormal viewership opening weekend (in 100,000's, summed across weekend days). The second row reports the results of IV regressions the abnormal probability of being dropped each week on abnormal audiences opening weekend (again in 100,000's) and includes truncated observations. The first stage results are respectively included in Table 1.1 and Appendix Table A.3. In Panel B, the first row is simply our base case from Table 1.2, reproduced here for ease of comparison; the second results are from the same specification, but with the outcome variable defined as tickets (in 100,000's) per screen; the final row shows these results divided 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 1.1. 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 A.3. Throughout, standard errors, clustered at the weekend level, are in parentheses. */**/*** denote significance at the 10%, 5%, and 1% levels, respectively. National weather shock instruments are chosen using the LASSO approach described in the text.

A.7 Google Trends Search Data as Proxy for Viewership

We use the Google Trends search data at the day by MSA by topic level, the most granular level at which it is made publicly available. Although Good Trends data are available for specific queries, we use the topic classification engine, which classifies searches as pertaining to particular movies.[16]

The raw data consist of integral figures from 0 to 100, where 0 and 100 are the lowest and highest points, respectively, in any single data export. Daily data can only be exported in three-month windows, and if volumes are sufficiently low for the entire period of intended export then daily data are not available. In this case, Google provides weekly data, or if volumes still remain too low, monthly data, but even weekly data is insufficient for our analysis because the weeks are measured from Sunday to Saturday, which does not allow us to distinguish between weekends since release.

Since Google censors to 0 any observation for which the total search volume falls below some (undisclosed) threshold, we are restricted in the number of movies and cities for which we can undertake the local analysis of network externalities.[17] We begin with the top 500 movies by U.S. gross ticket sales that were released May through September of 2004 through 2013.[18] For each movie, we collect local search volumes in each of the 10 largest MSAs (as well as national search volumes) on each day in a 3-month period beginning two weeks before the movie's release. Of the 5,000 MSA by movie combinations, nearly 4,000 are

[16]Without a topic classification system it can be difficult to determine which queries relate to specific movies. For example, a simple search for "Superman" could pertain to one of the many Superman movies, comic books, or other Superman merchandise. Thus, we use Google's topic classification service to classify searches as pertaining to specific movies, such as the 2006 film, "Superman Returns."

[17]While Stephens-Davidowitz (2014) develops an algorithm to circumvent some of Google's censoring, the algorithm relies on the fact that Stephens-Davidowitz (2014) is focused in volumes relating to specific queries. Our data exports, meanwhile, rely on Google's topic-classification system, which renders the algorithm ineffectual.

[18]We restrict to the top 500 movies because, since Google does not provide an API for data access, collection of large amounts of data from Google Trends is cumbersome. We focus on May through September because the summer release season is associated with significant variance in the instrument of choice in our main analysis (the percentage of establishments with maximum temperature unexpectedly between 75 and 80 degrees F), and do not include 2014 because at the time of writing Google's topic classification of searches pertaining to movies released in 2014 was largely incomplete.

censored, leaving 1,000 MSA by movie combinations that come disproportionately from the largest MSAs.[19] Our analysis is conditional on a rich set of fixed effects so it is important that we observe search volumes for the same weekend in the same city over multiple years; given this, we restrict our data set to the five MSAs with the most observations: New York, Los Angeles, Chicago, Washington D.C., and San Francisco. This corresponds to 67% of the original MSA by movie sample.

Since the data are normalized so that the highest point in any given data export is 100 and data exports are limited to four topics for a 90-day period, we separately export search volume for each movie in combination with the "Harvard University" topic. We then use the trend in Harvard searches to standardize movie search data across time.[20] Finally, we convert our search measure to the Z-score of search volume within each MSA. Note that because Google provides only unit-less search figures, we are unable to directly compare search volumes across MSAs.

[19]The extent of the censoring and the selection it might induce appears to be largely related to idiosyncratic factors affecting the extent to which Google's topic classification system is able to successfully classify searches. For example, "Ironman", which grossed over $300M in the US, is censored throughout, while "Ironman 2", which grossed nearly as much, is not.

[20]We use Harvard-related searches as an aggregator because searches for Harvard always show substantial volume, while individual movies tend to have non-zero volume only in the weeks surrounding release.

Appendix B

Appendix to Chapter 2

Table B.1: *Robustness*

	(1) Correct	(2) Correct
Wage = 3+1	omitted	omitted
Wage = 3	-161.6**	-133.4**
	(64.20)	(56.98)
Wage = 4	-156.9**	-115.2*
	(66.22)	(62.43)
Constant	932.9	497.5
	(52.05)	(48.78)
Sample	taker + excluded	taker + excluded + non-taker
N	266	540
Adjusted R-squared	0.022	0.009

Notes: This table analyzes the potential for selection in our analysis. Column 1 repeats the baseline analysis presented in the first column of Table 2, but includes results for the 36 workers who did not complete exactly 4 hours of work due to a technical glitch. Column 2 does the same but additionally includes the 274 workers who did not accept our job offer (these workers are coded as having completed 0 CAPTCHAs). Robust standard errors are presented in parentheses, ** p<0.05, * p<0.1.

Appendix C

Appendix to Chapter 3

C.1 Anecdotal Evidence

This section reproduces two quotes from the popular press.

C.1.1 Belsomra vs. Ambien

The following quote from *The New Yorker* reproduces the explanation given by a Merck neuroscientist in response to a question about how Merck evaluated the decision to pursue the development of Belsomra, a competitor to the blockbuster insomnia drug, Ambien. The quote suggests Merck was concerned about the fact that a generic version of Ambien would be available shortly.

> The perception at that time was, "You have a lot of medications available – should we be working on this? How large was the population of insomniacs poorly served by Ambien? *Should Merck invest in a market dominated by a drug that, within a few years, would become a cheap generic?*"
>
> Parker (2013), emphasis added

C.1.2 Lipitor vs. Zocor

The following quote from *The New York Times* suggests Lipitor was expected to lose market share to Zocor's generic, when it became available.

> Today, Merck's cholesterol-lowering drug Zocor loses its United States patent protection, becoming the largest-selling drug yet to be opened to cheap generic competition.
>
> That change will cost Merck billions of dollars a year. *But it could be nearly as damaging to Pfizer, whose rival cholesterol drug, Lipitor, is the world's most popular medicine, with global sales last year of $12 billion.*
>
> Berenson (2006), emphasis added

C.2 Data Appendix

This section provides more detail as to how I compute various components of the data set described in Section 3.3.

C.2.1 Dates of Approval and Generic Entry

Dates of first approval come from New Drug Applications (NDAs) in the Drugs @ FDA database. Next, I obtain the date of first generic entry, if available, by matching drugs by active ingredient to Abbreviated New Drug Applications (ANDAs) in the FDA's Approved Drug Products with Therapeutic Equivalence Evaluation (the Orange Book). Following Hemphill and Sampat (2012), I include only generic alternatives with therapeutic equivalence ratings of "A" to ensure that the date of generic entry used corresponds to the date a generic alternative became available that is chemically equivalent to the branded drug.

C.2.2 Expected Exclusivity

Firms sometimes file dozens of patents on a single drug so identifying key patents by hand is difficult. My strategy is to exploit the fact that the Hatch-Waxman Act of 1984 allows pharmaceutical manufacturers to extend a single patent for as much as half of the time a drug spent in development. As Hemphill and Sampat (2012) note, patents selected for extension typically pertain to active ingredients – i.e., the extended patent is typically a drug's primary patent. Indeed, Hemphill and Sampat (2012) codify patents for drugs that went generic in the 2000's and find that the extended patent represents the patent on the drug's active-ingredient in 79% of drugs analyzed. I obtain a list of extended patents from the USPTO, and the dates of filing and expiration for each patent from Thomson Innovation. I additionally note here that the FDA provides a floor of 5 years of market exclusivity for any new molecular entity, so I impute a market of 5 years for the 18 drugs that receive less than 5 years of exclusivity from the patent extension. This is conservative, since the floor of five years is often supplemented by additional extensions for pediatric indications and for orphan drugs.

C.2.3 Market Size

My measure of market size is derived from the total number of patients afflicted by the primary condition treated by class of drugs. Using the Medical Expenditure Panel Survey's (MEPS) Prescribed Medicines data, I first determine the ICD-9 code that is most commonly associated with each class of drugs. Cleaning the Prescribed Medicines data is cumbersome because the same drug often is often listed with multiple different names, so I aggregate the MEPS data by matching prescribed medications listed in MEPS with an extensive list of drug names and NDC codes provided by the American Society of Health-System Pharmacists (AFHS). The match rate is roughly 95%. The Prescribed Medicines files do not include records for drugs in 38 classes, likely because they are drugs prescribed in inpatient settings. For these classes, I identify primary ICD-9s from FDA approval documents.

Then, using the MEPS Medical Conditions data, I compute the national prevalence of that ICD-9 in each year. My main measure of market size is then the mean market size for that ICD-9, averaged over the period 1996-2011 for which MEPS data are available. I use the average of market size to capture market sizes over the entire period; using market sizes in the year of approval, or year before approval, does not affect the main results.

C.2.4 Patent Start and Expiry Dates

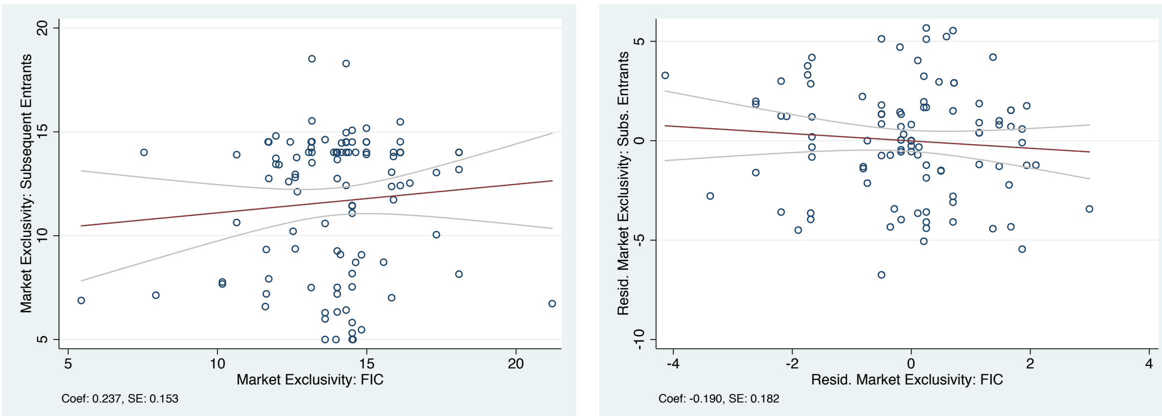
The patent priority date, sometimes called the "effective filing date" is the earliest date date any claim listed in the patent was filed. A patent's priority date is always no later than its filing date. If all claims are submitted for the first time in the application for the patent at issue, then the priority date is the same as its filing date. However, if some claims were filed in other patent applications, then the priority dates can be substantially earlier than the filing date of the patent at issue. This can occur, for example, because the patent filer seeks to delay publication of claims, or simply because the patent filer keeps adding claims to subsequent patent applications. It can also occur if a firm files a patent in one country and then later files the same patent in another – the earlier date is the priority date. I focus on priority dates because they reflect the first time intellectual property protection was sought and cannot be manipulated.

C.2.5 Incomes

Household incomes come from the Full Year Consolidated data files, which I match to ICD-9 codes in the Medical Conditions files. To avoid the endogeneity of demand choices, I use incomes from the year before the first in class was approved. Since the MEPS data only begin in 1996, I use 1996 incomes for classes in which the first in class was approved before 1996. Additionally, I deflate incomes to constant 2000 dollars using the GDP deflator provided by the St. Louis Fed.

C.3 Appendix Figures

Figure C.1: *Market Exclusivities Within Class*



(a) *Unresidualized* **(b)** *Residualized*

Notes: These figures respectively show plots of the market exclusivities of entrants subsequent to the FIC against the market exclusivities of the FIC. Panel A shows the relationship without controls, and Panel B shows the relationship conditional on mean development time, market size, and year of FIC approval fixed effects. The specification of the controls and sample are as described in Section 3.3. The slope of the line of best-fit and its associated robust standard error clustered by drug class are presented in the bottom left of each figure.

C.4 Panel Analysis of Entry and Remaining FIC Exclusivity

This section presents a panel analysis of entry timing relative to FIC exclusivity. In particular, I analyze how entry in class j in year t relates to the first in class exclusivity remaining in that year, conditional on time since first in class launch fixed effects and class fixed effects. Formally, the model is:

$$NumEntry_{jt} = \alpha + \beta FICExclRem_{jt} + \gamma_t + \zeta_j + \epsilon_{jt}. \tag{C.1}$$

The independent variable $FICExclRem_{jt}$ takes the same value as the independent variable of interest in my main analysis, $FICExcl_j$, in the year in which the first in class enters ($t = 0$) but then counts down until first in class generic entry occurs; thereafter, the variable takes on a value of 0. Since I am conditioning on time since FIC launch and class fixed effects, the estimate of β tells us how entry relates to first in class exclusivity, conditioning on average entry across classes in year t and average entry within class j.

The sample is the same as the main sample described in Section 3.3, aside from the fact that since I am not conducting an IV analysis, I do not restrict to only those classes for which I have data on the instrument. Thus, I have 127 classes. It is important to note the panel constructed with these data is not balanced because I do not observe the same number of years since FIC launch for all classes. For example, although I observe whether any entry occurs 10 years after FIC launch for classes that experienced FIC launch in 2000, I do not for classes that experienced FIC launch in 2005. In Appendix Table C.1, presented below, I present results from regressions that handle the censoring problem in a variety of different ways.

The first column presents results from a Poisson model estimated on the unbalanced panel, while the second does the same but restricts the panel to only those classes that experienced FIC launch no later than 2001. This restriction reduces the number of classes to 82 but also makes the sample more balanced. Columns (3) and (4) replicate the analysis in columns (1) and (2) but estimate OLS regressions. Finally, in the fifth column, I handle the censoring problem formally by estimating a Tobit (censored) regression. The results from this model need to be interpreted with caution, however, because Tobit models are non-linear so they suffer from the incidental parameters problem in the presence of fixed-effects. Altogether, although effect magnitudes vary slightly across regressions, the results show that entry tends to occur when the FIC has more exclusivity remaining. This is consistent with the evidence visualized in Figure 3.3.

C.5 Robustness and Placebo Tests

C.5.1 Sample Definitions

It is useful to remind the reader of how my data are constructed. First, recall that my baseline measure of market exclusivity is a constructed measure: it is the realized market exclusivities for the 56 first in class drugs which have gone generic, and for the others it

Table C.1: *Remaining First in Class Exclusivity and the Timing of Subsequent Entry*

	Dependent Variable is Number of Subsequent Entrants in Class in Year				
	QML Poisson Models		OLS Models		Tobit Model
	(1)	(2)	(3)	(4)	(5)
Remaining Market Exclusivity of FIC (years)	0.375*	0.375*	0.0158***	0.0159***	0.0374***
	(0.208)	(0.208)	(0.00597)	(0.00595)	(0.00741)
Drug Class Fixed Effects	yes	yes	yes	yes	yes
Time Since Launch Fixed Effects	yes	yes	yes	yes	yes
Sample Restriction	Uncensored	Uncensored & FIC Approval ≤ 2001	Uncensored	Uncensored & FIC Approval ≤ 2001	All
Mean of Dependent Variable	0.067	0.073	0.067	0.073	0.036
Number of Classes	127	82	127	82	127
N	1,639	1,376	1,639	1,376	3,048

Notes: This table presents panel model results from regressions of the number subsequent entrants approved in a given year on the remaining exclusivity of the FIC drug conditional on drug class fixed effects and time since launched fixed effects. Observations are censored if an observation's associated year takes place after the sample ends in 2011. Columns (1) and (2) present Poisson models estimated by quasi-maximum likelihood. Column (1) includes only uncensored observations and column (2) truncates the sample further to include only uncensored observations for which at least 10 years of entry data post-FIC launch are observed. Columns (3) and (4) replicate columns (1) and (2) using OLS models. As an additional robustness check, column (5) presents results from a Tobit model that includes all observations. Note that the Tobit model suffers from the incidental parameters problem so estimates are inconsistent and should be interpreted accordingly. Robust standard errors clustered at the drug class level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

is expected market exclusivity computed using patent expiry dates. Second, recall that although my data includes this measure for 127 classes, I restrict my baseline analysis to the 111 classes for which I additionally have values of the instrument. Third, recall that I focus on the effective FIC, not the authentic FIC.

Panel A of Appendix Table C.2 probes these three choices, one after the other. I focus here on Poisson (not Poisson-IV) specifications because the sample in my main analysis is limited by data on the instrument; it is only by focusing on non-IV specifications that I am able to relax the sample restrictions. For the same reason, these specifications do not include controls aside from year of FIC approval fixed effects. For ease of comparison, the first column repeats my base case analysis from specification (1) of Table 3.3, and the third and second to last rows respectively report the mean of the dependent variable (which changes with the sample) and the exclusivity coefficient multiplied by the mean of the dependent variable. Recall that since these are Poisson regressions, it is this last number which is comparable across columns because it is in units of subsequent entrants per year of FIC exclusivity.

I begin with the measure of market exclusivity and the sample restrictions. Column (2) reports results from the same regression as in column (1), but where the sample is restricted to only those observations for which I observe realized market exclusivity. Next, column (3) expands the baseline sample to additionally include the 16 classes for which I have a measure of exclusivity but no value for the instrument. In both cases, the coefficient on exclusivity moves slightly but remains positive and statistically significant at the 5% level. In column (4) I conduct an additional test where I expand my sample to include all 156 classes, replacing the data points for which I have no measure of FIC exclusivity with the 14-year threshold suggested by Keyhani *et al.* (2006).[1] The scaled estimate in column (4) has a smaller magnitude (0.121 versus 0.167 in my baseline case), but the relationship remains

[1]This is motivated by the fact that 14 years is the maximum exclusivity obtained through a patent extension, so drugs that already have 14 or more years of exclusivity would not receive an extension and would thus not have an exclusivity measure in my data. Using 14 years in place of the missing exclusivities represents a worst case scenario as these drugs presumably have longer exclusivities – to the extent that exclusivity and subsequent entry are positively related, censoring exclusivity at 14 years should attenuate the results.

Table C.2: *Robustness of Poisson Estimates to Sample Specification and Placebo Tests*

Panel A: Robustness of Sample Restrictions
Dependent Variable is Number of Subsequent Entrants in Class

	(1)	(2)	(3)	(4)	(5)	(6)
Market Exclusivity of FIC (years)	0.229***	0.137**	0.162**	0.165**	0.0500	0.322***
	(0.0604)	(0.0651)	(0.0633)	(0.0682)	(0.0503)	(0.103)
Year of FIC Approval Fixed Effects	yes	yes	yes	yes	yes	yes
Sample Restriction or Exclusivity Measure	Base	Realized Exclusivities Only	All Expected Exclusivities	Col (3) and Missing Exclusivity is 14 yrs	Authentic FIC & Restr. to Classes with Priority Next	Authentic FIC & Excluding Classes with Priority Next
Mean of Dependent Variable	0.730	1.500	0.827	0.731	2.632	0.719
Exclusivity Coef x Mean of Dep Var	0.167	0.206	0.134	0.121	0.132	0.232
N	111	56	127	156	19	96

Panel B: Placebo Tests
Dependent Variable is Num of Subs. Entrants Following

	Any Drug (mean = 1.07)		*The FIC (mean = 2.37)*	
	(1)	(2)	(3)	(4)
Market Exclusivity (years)	0.0701	-0.0237		
	(0.0482)	(0.0520)		
Market Exclusivity of FIC (years)		0.202***		0.108
		(0.0687)		(0.0678)
Market Exclusivity of LIC (years)			-0.0196	0.00942
			(0.0470)	(0.0441)
Year of FIC Approval Fixed Effects	yes	yes	yes	yes
Sample Restriction	All Drugs		Have Exclusivity for Last in Class	
Excl. Coef x Mean of Dep Var	0.075	-0.021	-	-
FIC Excl. Coef x Mean of Dep Var	-	0.216	-	0.256
LIC Excl. Coef x Mean of Dep Var	-	-	-0.045	0.021
N	196	196	35	35

Notes: This table presents robustness checks on the sample and definition of FIC in Panel A, and placebo tests in Panel B. All regressions include year of FIC approval fixed effects. Panel A, column (1) repeats the base analysis from column (1) of Table 3.3, where the sample is restricted to include only observations for which the time of the start of clinical trials is known. The sample in column (2) is further restricted to only observations for which generic entry is actually observed. Column (3) includes the 16 additional observations for which the data include a measure of FIC exclusivity but not the timing of the start of clinical trials. Column (4) codes FIC exclusivity as 14 years for observations that are missing it. Column (5) restricts the sample to only those classes for which a priority review drug immediately succeeded the authentic FIC drug and codes the FIC exclusivity as the authentic FIC's exclusivity, and column (6) restricts the sample to only those classes for which a priority review drug did not immediately succeeded the authentic FIC drug. Panel B, column (1) repeats the base case analysis but includes one observation for all drugs (so long as that class's FIC exclusivity is known), and column (2) repeats column (1) but includes a separate control for the FIC's exclusivity. Columns (3) and (4) restricts the sample to classes for which the data include the last in class's exclusivity. Robust standard errors are presented in parentheses and standard errors in Panel B, columns (1) and (2) are clustered at the class level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

positive and statistically significant. Overall, I infer that my main effects are robust to alternative definitions of exclusivity.

The last two columns of Appendix Table C.2, Panel A analyze my definition of first in class. In column (5), I restrict the sample to the classes for which my definition of FIC and authentic FIC do not coincide, and analyze how the authentic FIC's exclusivity relates to subsequent entry. As expected, the exclusivity of these drugs has a weaker relationship with subsequent entry: the scaled estimate is positive (0.132) but it is smaller than that of the base case, and it is not statistically significant (note the sample is very small). However, column (6) presents results from classes in which effective and authentic FIC drugs are one and the same. Here the scaled estimate is larger (0.232) and highly significant (though it is not significantly larger than that in column (5)).[2] I infer that while my focus on effective FIC drugs appears to accentuate the results, it is not driving them.

C.5.2 Placebo Tests

I conduct two sets of tests which probe the extent to which the exclusivities of non-FIC drugs are related to subsequent entry.

In the first test, I replicate my main analysis, but instead of including only one observation per class, I include one observation for each of the 196 drugs for which I have a measure of exclusivity. In other words, for each drug j, the independent variable is drug j's exclusivity and the outcome is the number of entrants subsequent to drug j.[3] Intuition predicts that this regression should yield a positive relationship but that it should not be as strong as that estimated using only FIC exclusivities (because the first generic should have an outsize effect on prices and thus entry incentives). The results, presented in column (1) of Appendix Table C.2, Panel B, show a small, positive, but statistically imprecise relationship between subsequent entry and exclusivity for all drugs. However, in column (2) I add a

[2]The number of observations in columns (5) and (6) do not add up to 127 because I do not have exclusivity measures for 12 authentic FIC drugs. It is not surprising these data are missing: these drugs are less commercially important, so generics may be less likely to be aggressive in pursuing entry.

[3]I cluster standard errors by drug class for these regressions.

control which is the exclusivity of the FIC for class j. In column (2), the FIC's exclusivity shows up as highly significant and the exclusivity of non-FIC drops (the point estimate is actually negative). I infer that FIC exclusivities are indeed most important in determining subsequent entry.

The second test investigates how the exclusivity of the last drug to enter in class relates to the number of entrants subsequent to the FIC. That is, this test is the same as my baseline analysis except that instead of focusing on the exclusivity of the first in class, I focus on the exclusivity of the last in class (LIC). Clearly, the exclusivity of the LIC cannot directly affect the number drugs in the class, so the coefficient should be zero; a non-zero coefficient would suggest endogeneity. For this test, I restrict the sample to classes that have at least two drugs in them, else the LIC is the same as the FIC. Only 50 classes satisfy this requirement, of which I only have an exclusivity measure for the LIC for 35. The results are presented in column (3). As expected, the scaled coefficient on LIC exclusivity (-0.045) is small (in fact, negative), and statistically insignificant. To make sure that this result is not an odd artifact of this sample, in column (4) I estimate the same regression but this time include also the first in class's exclusivity. The scaled estimate on FIC exclusivity (0.256) is close to that estimated in my main sample, but the scaled estimate on LIC exclusivity remains close to zero (0.021). Both are relatively imprecisely estimated but the sample is small. I interpret this as additional evidence that endogeneity is limited in this context.

C.5.3 Robustness of IV Estimates

In this section, I probe the robustness of my IV results. I first provide evidence that they are strongest for classes in which the delay between patent filing and the start of trials is large. This is reassuring since it seems unlikely that marginal changes in the date of patent filing for classes in which the date of patent filing already takes place immediately before the start of trials should have a substantive impact on exclusivity and thus subsequent entry. Second, I show that there is a strong relationship in the reduced form of the second stage, and third that the results remain similar in linear, as opposed to Poisson, regression specifications.

Table C.3 presents the estimates. All regressions include my full set of controls: development time, market size, and year of FIC approval fixed effects. The first column shows my baseline result from column (6) of Table 3.3, and then the second column includes only classes for which the value of the instrument is greater than its median of 4.43 years. The scaled coefficient in the second column is approximately 50% larger than that in the first column, suggesting the marginal effect of FIC exclusivity on subsequent entry is larger for drugs for which delay between filing and development is larger (though I note the difference in coefficients is not statistically significant). Next, the third column presents results from the reduced form of the second stage (for my baseline sample of 111 classes): I estimate a Poisson regression where the outcome is the number of subsequent entrants in class and the independent variable is the instrument. The estimated coefficient is significant at the 10% level and implies that a one year delay between patent filing and the beginning of development for the FIC is associated with a 7% decrease in subsequent entry.

To ensure my IV results are not due to the functional form of the Poisson model, in columns (4)-(6) I replicate the analysis in columns (1)-(3) but using a linear framework. Since the outcome is a count variable and is highly skewed (which motivates my use of the Poisson framework), the outcome in these regressions is the log of 1 + the number of subsequent entrants. Columns (4) and (5) present results from linear IV models, while column (6) presents the reduced form of the second stage estimated by OLS. Although the scaled estimates move slightly, all remain statistically significant.[4]

[4]My non-IV results are also robust to a linear specification – see Figure 3.4.

Table C.3: *Robustness of IV Estimates*

	Dependent Variable is Number of Subsequent Entrants in Class			*Dependent Variable is Log(1+Number of Subsequent Entrants in Class)*		
	QML Poisson-IV		QML Poisson	Linear IV		OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Market Exclusivity of FIC (years)	0.337***	0.482***		0.0583**	0.0988**	
	(0.128)	(0.184)		(0.0277)	(0.0394)	
Patent Filing to Clinical Dev. (years)			-0.0692*			-0.0261*
			(0.0357)			(0.0140)
Mean Development Time	yes	yes	yes	yes	yes	yes
Market Size	yes	yes	yes	yes	yes	yes
Year of FIC Approval Fixed Effects	yes	yes	yes	yes	yes	yes
Sample Restriction	Base	Instr. > Median	Base	Base	Instr. > Median	Base
Mean of Dependent Variable (levels)	0.73	0.78	0.73	1.73	1.78	1.73
Exclusivity Coef x Mean of Dep Var	0.246	0.376	-	0.100	0.174	-
F-Statistic from the First Stage	43.52	19.68	-	24.94	15.99	-
N	111	56	111	111	56	111

Notes: Columns (1) and (2) present results from Poisson-IV models estimated by quasi-maximum likelihood and using a control function for the instrument, which is the difference in time between the patent's filing date and the start of clinical development. The first column repeats the analysis from column (6) of Table 3.3, while the second column restricts the sample so that the value of instrument must be greater than its median of 4.43 years. The third column presents results from a Poisson regression of the second stage outcome variable on the instrument. Next, columns (4)-(6) repeat the analysis in columns (1)-(3) but in a linear-IV/OLS framework, where the outcome is the log of 1+ the number of subsequent entrants. All regressions include controls for mean development time in class, market size, and year of FIC approval fixed effects. The sample and controls are specified as described in Section 3.3. All standard errors are robust, and standard errors in the Poisson-IV models (which are estimated by two-stage residual inclusion) are corrected for the two-stage design as described in the text. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

C.5.4 Base Case with Sample Restricted to Classes Starting No Later than 2001

In this section, I repeat the base analysis presented in Table 3.3 but restrict to the 62 classes that started no later than 2001. The motivation for the restriction is to ensure that all classes included in the analysis have had sufficient time to mature and see subsequent entry. The results are presented in Appendix Table C.4, below.

Although the restricted sample is substantially smaller than that used in the main analysis, and this weakens the first stage of the IV analysis and generally decreases precision, the point estimates presented here are remarkably comparable to in my main analysis. Estimates in columns (3) and (6), which are conditional on my full set of controls, suggest respectively that an extra year of FIC exclusivity is associated with a 27% and 37% increase in subsequent entry. Relative to my full sample, the classes analyzed in this table are on average larger, so the scaled coefficients are also slightly larger than those presented in Table 3.3. Altogether, I conclude that, if anything, including classes that started after 2001 dampens my main results.

C.5.5 Base Case Controlling for Sales

In this section, I show my main results do not appear to be driven by drug or class profitability. I do this by repeating the base analysis presented in Table 3.3 but controlling for sales, for which I employ two measures. The first is the log of the maximum annual revenue ever earned by the first in class, and the second is the log of the maximum annual revenue ever earned by any drug in given class. These measures respectively capture the extent to which the first in class is a top-earner as well the extent to which a class is attractive from a market size perspective.

The results are presented in Appendix Table C.5, with Panel A analyzing the control for FIC revenue and Panel B analyzing the control for class revenue. I do not have these measures for all classes, so I present first my base case results (without the control) on the sample for which I do have the measures. In particular, the first set of two columns in each panel present Poisson results, while the second set of two columns present results from

Table C.4: *Base Analysis with Sample Restricted to Classes Starting No Later than 2001*

	Dependent Variable is Number of Subsequent Entrants in Class for Classes such that FIC Approval Occured No Later than 2001 (mean =1.18)					
	QML Poisson Models			QML Poisson-IV Models		
	(1)	(2)	(3)	(4)	(5)	(6)
Market Exclusivity of FIC (years)	0.224***	0.241***	0.248***	0.473**	0.403**	0.402**
	(0.0660)	(0.0699)	(0.0690)	(0.206)	(0.161)	(0.160)
Mean Development Time	no	yes	yes	no	yes	yes
Market Size	no	no	yes	no	no	yes
Year of FIC Approval Fixed Effects	yes	yes	yes	yes	yes	yes
Exclusivity Coef x Mean of Dep Var	0.264	0.284	0.293	0.558	0.476	0.474
F-Statistic from the First Stage	-	-	-	11.96	18.91	20.01
N	62	62	62	62	62	62

Notes: This table repeats the base analysis presented in Table 3.3 but restricts to the 62 classes that started no later than 2001. Columns (1)-(3) present results from quasi-maximum likelihood Poisson models which incrementally add controls for mean development time in class and market size. Columns (4)-(6) replicate columns (1)-(3) but instrument for FIC market exclusivity using the time between patent filing and the start of clinical development, where the start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. All models are conditional on year of FIC approval fixed effects. The IV models are implemented using a control function and first stage estimates are presented in the final rows of the table. The sample and controls are specified as described in Section 3.3 aside from the sample restriction. Robust standard errors are reported in parentheses and standard errors for the IV models are corrected for the two-stage design as described in the text. *** p<0.01, ** p<0.05, * p<0.1.

Poisson-IV models, and the first column in each set does not include the control while the second does. Altogether, including the control for sales does not have a statistically nor economically significant effect on the point estimates. I infer that my results are not related to class profitability.

Table C.5: *Base Analysis Controlling for Sales*

Dependent Variable is Number of Subsequent Entrants in Class

Panel A: Controling for Max Annual Revenue of FIC

	QML Poisson Models		QML Poisson-IV Models	
	(1)	(2)	(3)	(4)
Market Exclusivity of FIC (years)	0.224***	0.170**	0.380**	0.374**
	(0.0808)	(0.0840)	(0.173)	(0.171)
Max Annual Revenue of FIC	no	yes	no	yes
Mean Development Time	yes	yes	yes	yes
Market Size	yes	yes	yes	yes
Year of FIC Approval Fixed Effects	yes	yes	yes	yes
Mean of Dependent Variable	0.871	0.871	0.871	0.871
Exclusivity Coef x Mean of Dep Var	0.195	0.148	0.331	0.326
F-Statistic from the First Stage	-	-	17.59	18.16
N	85	85	85	85

Panel B: Controling for Max Annual Revenue in Class

	QML Poisson Models		QML Poisson-IV Models	
	(1)	(2)	(3)	(4)
Market Exclusivity of FIC (years)	0.200***	0.229**	0.279*	0.261
	(0.0767)	(0.0972)	(0.159)	(0.188)
Max Annual Revenue in Class	no	yes	no	yes
Mean Development Time	yes	yes	yes	yes
Market Size	yes	yes	yes	yes
Year of FIC Approval Fixed Effects	yes	yes	yes	yes
Mean of Dependent Variable	0.886	0.886	0.886	0.886
Exclusivity Coef x Mean of Dep Var	0.177	0.203	0.247	0.231
F-Statistic from the First Stage	-	-	17.61	17.38
N	88	88	88	88

Notes: This table repeats the base analysis presented in columns (3) and (6) of Table 3.3 but adds controls for sales. In Panel A, the control for sales is the log of the maximum annual revenue ever earned by the FIC, and in Panel B, the control for sales is the log of the maximum annual revenue ever earned by any drug in that class. The sample is restricted to those classes for which the sales measure is available throughout each panel. Columns (1) and (2) present results from quasi-maximum likelihood Poisson models and columns (3) and (4) replicate columns (1) and (2) but instrument for FIC market exclusivity using the time between patent filing and the start of clinical development, where the start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. All models are conditional on year of FIC approval fixed effects, mean time in development, and market size. The IV models are implemented using a control function and first stage estimates are presented in the final rows of the table. The sample and controls are specified as described in Section 3.3. Robust standard errors are reported in parentheses and standard errors for the IV models are corrected for the two-stage design as described in the text. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.