Draft

The Downside of Experimentation: Evidence from Television Shows

Ankur Chavda

Massachusetts Institute of Technology

November 2018

Abstract

When does experimentation on new ideas improve final outcomes? Experimentation can provide an early signal of poor final outcomes, enabling decision makers to winnow out bad ideas before those outcomes are realized. But not committing to ideas and retaining right to terminate after experimentation can also be detrimental to outcomes, by for example attracting lower quality workers or motivating exploitive rather than explorative behavior. By testing theoretical models against a dataset of television shows that enables a comparison of final outcomes between experimental and non-experimental production processes, I find evidence that experimentation both handicaps worker recruitment and shifts effort away from improving final outcomes toward improving the experimental signal. This results in experimentation only improving final outcomes when it winnows out enough bad ideas, otherwise experimentation is detrimental as its benefits are unable to overcome its downside: the lack of commitment.

1. Introduction

Experimentation is a vital part of many innovative industries where outcomes are uncertain. However, theory is conflicted as to whether experimentation always improve final outcomes. Experimentation provides an early signal of likely outcomes can enable decision makers to winnow out potentially poor outcomes before they are realized, improving observed final outcomes (Bowman and Hurry, 1993). Experimentation can also enable improvement of outcomes when the design process is flexible and allows for learning (Thomke, 1997). Yet experimentation often involves a lack of commitment (Ghemawat, 1991), suggesting the existence of downsides to experimentation: experimentation might encourage competitive investment or entry (Schmalensee, 1978), make it harder to recruit high quality talent that prefers long term contracts (Holmstrom, 1983), shift worker incentives towards improving the experimental signal to the determinate of final outcomes (Baker, 1992) or encourage workers to shy away from riskier, more valuable projects (Manso, 2011). A tension therefore exists between the benefits and drawbacks to experimentation with a natural research question: when does experimentation improve final outcomes and in the cases experimentation does not, why do final outcomes worsen?

I explore this research question in the setting of television show production in the US, where television networks fund new shows through staged development, a form of experimentation (Bhattacharya, Chatterjee and Samuelson, 1986; Iansiti 2000; Kerr, Nanda and Rhodes-Kropf, 2014). Although the networks can directly order the development of a new show's first season of episodes (straight to series order), historically only the first episode, called the pilot, is ordered with the option of developing the rest of the season after evaluating the pilot. Importantly, the setting has variation in the winnowing value of experimentation: the likelihood a show passes the pilot stage is higher when a show's creators have a track record of creating award-winning shows. There also exists variation in the use of experimentation: the industry's reliance on piloting was relaxed when Netflix entered the industry in 2013, resulting in two regimes, one where pilots was almost always used and another where a mixture of pilots and straight to series orders were used. I exploit both these characteristics to empirically address my research question.

As my main result, I find the winnowing effect of removing potentially poor outcomes is critical to whether experimentation improves final outcomes. When this effect is strong, experimentation improves final outcomes, otherwise experimentation worsens final outcomes. I use two approaches to support this finding. First, employing propensity score matching and a model of the network's piloting decision, I estimate the treatment effect of piloting on Internet Movie Database (IMDb) ratings overall to be negative and specifically negative for shows with award-winning creators and positive otherwise. Since shows with award-winning creators are more likely to pass the pilot phase, this suggests the strength of experimentation's winnowing effect is critical to whether final outcomes are improved. Second, I construct a panel of each network's yearly portfolio of new shows and illustrate the average IMDb rating of these portfolios of shows improved after Netflix's entry shock triggered less piloted show development, suggesting on the margin experimentation was detrimental to final outcomes prior to the shock, consistent with my show level evidence.

As a secondary result, I find a combination of worker preferences for long term contracts and misaligned incentives can lead to overall worse final outcomes under experimentation. I model how various mechanisms behind a treatment effect of piloting would differentially affect the first episode of a show relative to the rest of a show's first season. Using predictions from these models, I provide show episode level evidence consistent with misaligned incentives. In addition, I find piloted show are less likely to have veteran actors for their main roles, suggestive of worker preference for long term, non-piloted employment. I also argue the existence of these mechanisms is supportive of my main result.

This paper's primary contribution is the first empirical study of the tension between the benefits and drawbacks of experimentation, resulting in the exposure of a heterogeneity in when experimentation is worthwhile: final outcomes only improve when experimentation sufficiently winnows out bad ideas. In addition, this paper is the first to show that in some cases the drawbacks of worker preferences for long term contracts and misaligned incentives are enough to make experimentation overall detrimental to final outcomes.

The rest of this paper proceeds as follows. Section 2 provides a review of previous literature related to my research question. Section 3 describes experimentation in television show production as well as how Netflix changed the incumbent's decision to experiment. Section 4 models the network's experimentation decision to generate testable hypotheses. Section 5 outlines the data used for this paper and lays out my empirical strategy. Section 6 reviews the paper's results and Section 7 concludes.

2. Literature Review

Few empirical papers have addressed the question of when experimentation is worthwhile, even in a broader context of experimentation outside of staged development. Thomke's (1998) setting is integrated circuit design which benefited from a new type of experimentation, computer simulation, allowing new designs to be tested without first creating a functional prototype. However, not all circuit designers benefited equally, some underlying technologies could make easier use of simulation than others. Thomke shows those with complementary technology were both more likely to use simulation and had overall lower development costs. MacCormack, Verganti and Iansiti (2001) study the development practices of Internet software companies, one development practice being the amount of new feature development after a product's initial release. When the initial release is viewed as an experiment where market demand information is learned and incorporated into the next release, their findings suggest increased reliance on the experiment to determine later feature development correlates with better outcomes, which in their case are assessments solicited from industry experts.

My paper builds on the strengths of these papers while also addressing a number of unmet empirical challenges. First, an ideal outcome measure should be independent of industry structure. For example, if projects competed against each other, increasing commitment across all projects might improve the quality of all projects without showing improvement in sales for any of them. Both Thomke (1998) and MacCormack, Verganti and Iansiti (2001) account for this and I do as well by using Internet Movie Database (IMDb) ratings my outcome measure, which Waldfogel (2017) provides evidence is invariant to changes in industry structure over time.

Second, unlike MacCormack, Verganti and Iansiti (2001) which relies on a small dataset of products as dissimilar as a web browser (Netscape 3.0) and an internet start page (My Yahoo!), my data set has enough comparable observations for statistical analysis and contains variation in experimentation. The incumbent US television networks on average produce 300 to 400 such pilots a year, enough to construct a dataset large enough for statistical analysis. Television shows are arguably comparable: they are all classifiable into either comedies or dramas, run for either a half hour or a full hour and often based on similar themes such as the anti-hero present in both *The Sopranos* and *Breaking Bad*.

Third, unlike Thomke (1998), I can track outcomes based on whether experimentation has occurred. My dataset includes information both about whether a specific show was piloted and that show's outcomes.

Fourth, unlike MacCormack, Verganti and Iansiti (2001), my dataset includes variation in the value of experimentation. Real options posits the benefit of experimentation is the information generated about potential outcomes. Experiments that are less informative about whether the project should succeed should have less beneficial impact on outcomes. Having such variation would allow an empirical study to better probe when experimentation is valuable. However uncovering for such variation is tricky when the use of experimentation itself is endogenous: in cases when an experiment would be uninformative, experimentation will not be used, so simply looking for variation in informativeness conditional on experimentation will produce a flawed analysis. Thomke (1998) skirts around this issue by arguing an obvious relationship between underlying technology and usefulness of experimentation rather than providing empirical evidence. Ideally two decision making regimes should be present, one where only experimentation is allowed to uncover informativeness and another version where experimentation varies to study the effect of experimentation itself.

Luckily my setting has these two regimes. Historically, television shows were almost always piloted, providing a fair approximation of the first regime. However, this changed after 2013, due to the unexpected success of two of Netflix's shows, *House of Cards* and *Orange is the New Black*. The success of these two shows precipitating a dramatic increase in straight to series orders by incumbent networks. I use Netflix's post entry period of 2014 to 2017 as the second regime in my analysis.

Fifth, a natural concern given my research question is selection bias: any observed improvement in outcome of a straight to series order over a piloted show could be driven by the networks selecting higher ex-ante quality shows for straight to series production. I mitigate selection on observables based bias with a propensity score based matching estimator so piloted and straight to series shows are similar in observed, ex-ante characteristics, which goes a step beyond prior research. In addition, although I cannot rule out selection on unobservables, I use a network portfolio of shows approach which is less susceptible to such bias as a complement to my matching estimator.

Finally, several mechanisms exist for how experimentation could influence outcomes. Altering competitive behavior (Schmalensee, 1978) is perhaps the canonical example of commitment improving outcomes. Experimentation could affect product values due to mis-aligned incentives (Baker, 1992): if the project team's effort to increase a positive experimentation signal detracts from the effort required to increase the value of the final product, experimentation on a project could result in lower final product quality. Worker preferences for longer term contracts (Holmstrom, 1983) can also make a committed project's outcomes better relative to experimented project by enabling the project to attract higher skilled workers if those preference tend to be uncompensated (Bonhomme and Jolivet, 2009). A larger investment in planning could be justified with commitment but not with experiments (Milgrom and Roberts, 1990), leading to more favorable outcomes (Delmar and Shane, 2003) under commitment. Commitment has the potential of fostering core capabilities (Leonard-Barton, 1992), improving long run performance of a firm. Long term incentives can be thought of as committing to agents despite early failures (Manso, 2011), leading to better outcomes (Lerner and Wulf, 2007; Azoulay, Graff Zivin and Manso, 2011).

The prior research on experimentation suggest mechanisms behind their results without providing evidence supporting the existence of mechanisms. Although this paper does not prove any specific mechanism's existence, it does however provide evidence supporting the predominance of some mechanism over others, another improvement relative to the existing literature.

3. Television Pilots as an Experiment

3.1 Staged Development in Television

As in many creative and innovative industries, uncertainty is ever present in the production of television shows. William Goldman's famous "Nobody knows anything" quote from the movie industry applies: 65% of television shows on the major networks get cancelled after their first season (Ocasio, 2012), a clear indication of failure when it takes four seasons for a show's financiers to make a return on their investment (Bunn, 2002). This failure rate exists despite a winnowing process that results in only a fraction of ideas for new shows getting aired on television. It is also not yet clear that the big data capability of entrants like Netflix and Amazon has reduced this uncertainty (Smith and Telang, 2016). To quote Netflix's Chief Content Officer: "the data just tells you what happened in the past. It doesn't tell you anything that will happen in the future" (Adalian, 2013). Amazon is shifting away from its original data driven decision approach because as one Amazon executive puts it: "we're getting chewed up" (Fritz and Flint, 2017). This leaves staged development as the main way the industry deals with uncertainty in outcomes.

Although this paper focuses on pilot stage of development as an experiment, television show production involves multiple development stages that are each a kind of experiment. Ideas for shows can come from a variety of places; a writer might come up with a plot idea for a new show or a network executive might wonder how a book would translate into a television series. The first stage of development for these ideas is drafting a log-line or synopsis (treatment) of the idea that captures the main setting and how the story might evolve over the course of a series. Treatments are pitched to network executives. However, at around five pages long, treatments are limited in do not provide a lot of information about whether the show will be successful (Luo, 2014).

Each year, a typical major network picks around a hundred treatments for the next stage of production, the creation of a script and a “bible”, a full description of all aspects of the show’s universe. Scripts typically cost on the order of $100k and provide the network with the option to fund continued development on the show idea. The script provides more information about the show idea than the treatment itself; the script order can be thought of as the first experiment done by the networks on the show idea.

Conditional on an attractive script, the networks can choose to engage in the third stage: the pilot. For each expected open slot for a new television show on their fall schedules, networks pick two to three scripts from its pool of optioned scripts to pilot. Piloting produces the first episode of the series at a cost of about two million dollars, an order or magnitude more expensive than the script. Although now somewhat controversial given Netflix’s success with straight to series orders, piloting has a long history in the television industry; in 1951 CBS had Lucille Ball make a pilot for the *I Love Lucy* show even though she had played leading roles in both film and television. The great majority of television produced today by networks other than Netflix and Starz continues to use pilots as experiments; Starz having adopted a straight to series production model long before Netflix entered the industry.

3.2 Approaches to Television Show Experimentation

Real options does not necessary apply just because an experiment is used, nor do theories about commitment apply whenever an experiment is not used. For television show production to have promise as a setting to study the tension between these framings for experiments, the basic primitives of both must be present.

Real Options

Although a definition of real options as a small investment which, depending on what is learned over time, can optionally lead to a larger, follow-on investment is accurate, it may also be overly broad since it includes objects such as financial options which are not the primary focus of strategy researchers. Bowman and Moskowitz (2001) survey the empirical work on real options in strategy and induce an operational definition based on how the concept has been applied by strategy researchers:

> A common theme in all of these decisions is that they entail the use of a two stage process: In the first stage, a small investment is made that gives the company the right to participate in the project (i.e., the company purchases the option). Some time later, when more information is known, the second stage occurs when the company faces a choice about making a larger investment in the project (i.e., the company exercises the option).

They suggest there are couple key elements to how strategy researchers have applied real options. First, there is a project being considered by a firm, ranging from investments in other firms (Kogut, 1991; Hurry, Miller and Bowman, 1992) to product development (Thomke, 1998; MacCormack, Verganti and Iansiti, 2001). Second, two distinct investment opportunities exist, an initial one that creates an option on the project and a second one that funds the project. Third, between the first and second investments, information is learned about the project which influences the decision to invest in the second stage. Empirical settings that display these characteristics will at least be consistent with how strategy researchers have operationalized real options in the past.

The production of television shows is one such setting. First, new television shows are projects under consideration for investment by firms. The incumbent television networks have historically providing the bulk of funding for new television shows by paying for the right to first broadcast the show (Blumenthal and Goodenough, 2006, p. 203). Each fall, the incumbent television networks decide which ideas for shows receive the investment necessary to produce the show. The outcome of these investments, television shows, are products in the sense of a good for sale, albeit intangible ones. Television audiences usually “pay” for watching them by consuming advertising in a bundle, but it is also common to pay for the product more directly by for example subscribing to a cable channel or purchasing a show’s DVD. Television shows are also products in the differentiated sense in contrast to commodities: although both *My Little Pony* and *Bojack Horseman* are animations about equines, a typical consumer is likely to prefer one over the other.

Second, the investment in television shows is staged as described in the previous section. Television show production follows a yearly schedule that starts with ideas and ends with a season of show episodes. The first decision point on the part of a network is whether to take an “option” on a show idea. In return for an upfront investment, an option gives the purchasing network the right of first refusal for the script based on a show idea. After the script is created, the network can make a second investment by commissioning a pilot, a show’s first episode, based on the script. Finally, the network can order the show’s first season for broadcast.

Lastly, there is uncertainty in television production that gets resolved at each stage as information about the show is learned. Ideas for shows initially are presented to the networks as treatments: a description of the show’s concept usually less than five pages long. The next stage of a show’s development, a script, typically requires an investment by the network on the order of one hundred thousand dollars and results in both the text for the first episode, running 30 to 50 pages, as well as a “bible”, a full description of all aspects of the show’s universe. The increased content gives the network more detail about the potential for the show idea. Following the script, some shows are piloted, requiring an investment of around two million dollars and resulting in the first episode of the show’s season. Beyond the network being able to now see the show idea brought to life by actors, the network is able to “test” the pilot against a live audience in a controlled environment and gain a consumer signal of the show’s quality. The networks use this additional information to inform the decision of whether to “greenlight” the show: produce the rest of the show’s first season for broadcast.

Television show production therefore is consistent with other settings where strategy scholars have studied real options. The show it itself is a project, piloting is a separate investment stage from a full season, and piloting reveals information about whether the full season investment is worthwhile.

Commitment

Ghemawat (1991) defines commitment as earlier choices that constrain later ones. He calls out three characteristics projects must have for investments in them to be considered commitments. First, they must be durable in the sense of being long lasting and affecting future decisions. Second, the must be specialized in that not all future courses of action are enabled by them. Third, they must be untradeable; significant frictions must prevent a firm from offloading a committed investment.

A network that skips a stage of the television production process is committing to the project in a way consistent with Ghemawat's definition of commitment. Durability manifests in television when for example ordering a show straight to series by directly greenlighting a script means one less funding slot available for the shows that are piloted. This could lead to a piloted show not getting funding even if post-pilot it appears better than the previously committed show. Shows are specialized in that a committed show that attracts a small audience will not enable the network to generate as much revenue as a show attracting a large audience. Show commitments are also not tradable between networks since the content needs and audience demographics vary even between the major national networks; a diverse comedy like ABC's *Modern Family* would be an unlikely fit for CBS due to CBS's more culturally conservative audience (Canfield, 2016).

Skipping the pilot phase by ordering a show straight to series is therefore consistent with definitions of commitment. It affects a network's later development decisions, restricts the set of actions available to the network and cannot be trivially undone. How might commitment improve outcomes in television? Several arguments have been made: that it preempts similar shows on competitor networks (Andreeva, 2014), the first episode becomes better integrated with the rest of the season (Hawley, 2014), high skilled labor becomes easier to recruit (Anders, 2015), there is less incentive to improve the pilot at the expense of the rest of the season (Anonymous Emmy Nominated Producer, 2017) or the entire season benefits from an extended planning period

(Adalian, 2013). These arguments mirror existing theoretical mechanisms for how experimentation could worsen outcomes and therefore make my setting a reasonable one to investigate mechanisms in.

3.3 Netflix Changing Experimentation

Previously known as an online streaming distributor of film and television content, Netflix began funding new television production in 2011. Many have tried and failed to break into the exclusive club of prestige television show producers so by itself Netflix's funding of original constitute did not constitute a major event in the industry. For example, in 2001 A&E networks commissioned two scripted dramas which were cancelled by the end of 2002 because, in the words of one of the involved actors, "A&E was transforming from the premier intellectual cable network in America to one that airs Dog the Bounty Hunter on repeat (Farquharson, 2008)." Revealed preference is consistent with the idea that new entrants often fail: Netflix was initially only attracting show ideas that everyone else had passed on (Dawn, 2013).

The low expectations for Netflix contributed to why Netflix ordered all its shows straight to series. Netflix needed to prove their own commitment to television (Weisman, 2014); no one wanted to make a pilot for a new entrant that might quit the industry without making a single show. But low expectations were not the only reason for Netflix to order shows straight to series; technological change is often cited as another cause. Unlike traditional broadcast networks that could only broadcast a fixed number of shows each week during prime-time hours, Netflix faced no such capacity constraint. For Netflix, even a poor show was one more show that could attract subscribers on their service (Weisman, 2014); piloting has little value when a show was worth developing regardless of the information gained from the pilot. Finally, the lack of internal resources may have played a role: initially Netflix only had one person assigned to developing new shows (Weisman, 2014).

In 2013 Netflix surprised the industry when two of its shows, *Orange Is the New Black* and *House of Cards*, garnered Golden Globe and Emmy nominations. Afterwards, Netflix turned into a desirable place for creators to bring new show ideas (Adalian, 2013). As depicted in Figure 1, incumbent networks began skipping pilots for their own shows, both because Netflix's entry suggested commitment could pay off and in fear of losing shows to Netflix since show creators preferred commitment by the networks (Adalian, 2013; Brembilla, 2014). As 2013 was the year Netflix successfully broke into the prestige television industry, for the purposes of this paper I refer to 2013 as Netflix's true entry date rather than its initial funding of shows in 2011. I posit this entry constituted an unexpected shock to the decision making of the incumbent networks.

[Insert Figure 1 about here]

4. Modeling the Decision to Experiment

Why do networks pilot some shows and order others straight to series? First, piloting may not be worth it. In 2008, Jeff Zucker, NBC's CEO at the time, outlined a plan to reduce piloting because the cost of pilots was not worth the amount of information they generated (Nordyke, 2008), a trade-off that suggests the network's decision making was focused on the payoffs from experimentation. Second, the network may not have the choice to pilot. After being kicked off *Two and a Half Men* for substance abuse, Charlie Sheen secured enough independent funding to fully develop his next show, *Anger Management*. This independent funding meant networks could only order the full first season or pass on the idea, an example of when supplier bargaining power could force a network to order a show straight to series (Andreeva, 2011; Dillon, 2012).

These two possible drivers of a show's production decision, payoffs and bargaining, were also cited in how the networks reacted to Netflix's entry (Adalian, 2013). Netflix's entry could have raised beliefs about the payoff from ordering a show straight to series, reducing the rate of piloting. Alternatively, Netflix's entry could have improved supplier bargaining power, again reducing the rate of piloting.

Models of payoffs and bargaining can help interpret the observed outcomes in my data, indicating when the tradeoff between the benefits of a real options approach and commitment can be identified.

4.1 Optimal Decision to Experiment

The existing strategy literature is missing a formal model of the decision to experiment that incorporates both theories of real options and commitment. The foundational models of real options apply to sequential decisions to invest across multiple stages (Roberts and Weitzman, 1981; McDonald and Siegel, 1986; Majd and Pindyck, 1987) and therefore focus on answering questions such as when it is optimal to stop investing rather than when to commit by investing for more than one stage at once. In contrast, the existing empirical work that does investigate variation in multi-stage investments (Thomke, 1998; MacCormack, Verganti and Iansiti, 2001) lacks a model accounting for the endogenous decision to experiment and therefore provides correlations rather than causal findings.

To fill this gap, I take advantage of the experimentation approach used by Nanda and Rhodes-Kropf, (2016) in the entrepreneurial finance literature. In their paper they illustrate the use of staging investments by venture capitalists to overcome the uncertainty inherent in startups. Partially funding a startup both generates a signal of the startup's viability and provides the VC with the option to further invest at later stages. In contrast, the VC can also choose to fully fund a startup without a signal. Committing to the startup may be worthwhile when the incremental cost of staging investments is greater than the value of the generated signal. Their modeled tradeoff between real options and commitment is not unique to venture capitalists; it more generally applies to any decision to experiment as a means of resolving uncertainty.

For a particular project i, the decision maker an ex-ante expected belief θ_i about the outcome, drawn from a distribution of outcomes Θ. Executing on the project has fixed cost C associated with it, regardless of the outcome. For an additional cost e, the decision maker can generate a signal about the project which results in a posterior belief $y(\theta_i)$ about the project. The distribution of the posterior belief is dependent only on θ_i, has expected value $E[y(\theta_i)] = \theta_i$ and would always be worth doing if the cost of experimentation was zero: $E[y(\theta_i)|y(\theta_i) > C] > \theta_i$. Subject to the payoff being positive, the decision maker would therefore maximize:

$$\max_{\lambda_i \in \{0,1\}} \lambda_i(\Pr(y(\theta_i) > C)\, E[y(\theta_i) - C|y(\theta_i) > C] - e) + (1 - \lambda_i)(\theta_i - C) \tag{1}$$

This leads to two cutoffs, θ' and θ'' under some basic assumptions about C and e. θ' is the lower bound threshold for experimentation. Projects with ex-ante beliefs below θ' do not have a positive payoff after experimentation and will not be considered. θ'' is the upper bound threshold for experimentation. The experimental signal does not terminate enough projects when beliefs are above the θ'' cutoff to justify the cost of experimentation, yet these projects are profitable so they are executed directly without the experiment. All projects with ex ante beliefs between θ' and $\theta''$are experimented on and conditionally executed depending on the signal. Figure 2 plots the optimal decision to experiment for an example signal and cost parameters as well as what observed outcomes would be under the optimal decision.

[Insert Figure 2 about here]

How might Equation 1 apply to a network's decision to pilot a particular television show idea i? Each idea likely varies in the network's beliefs about outcome θ_i: a creator with a long history of successful shows may have a higher probability of success θ_i (Nussbaum, 2018). Conditional a belief that a new show is better than the network's outside option, its poorest performing existing show, the network will order a season of episodes to broadcast. Pilots cost more than an average episode to produce due to for example the shorter-term labor contracts involved (Anonymous Emmy Nominated Producer, 2017). Evaluating the pilot is another expense the networks must consider (Bunn, 2002). e represents both these production and signal evaluation costs. From network's perspective, cost C is primarily a fee payed to the creator of a show for the right to broadcast the show. I also assume show specific characteristics other than the main scripted genre categories of drama and sitcoms do not affect e or C. Networks do not consider the cost of a specific script when making pilot or series order decision for a script (Anonymous ABC Executive, 2017). Although a show on HBO will cost far more than one on ABC, variance in cost between two scripts is lower within the same network. Cost is instead involved when determining the total number of scripts to pilot or order to series; in 2006 NBC for example chose to reduce the number of scripted dramas it ordered rather than attempt to reduce the cost per drama (Barnes, 2006).

Equation 1 generalizes the model provided in Nanda and Rhodes-Kropf (2016). To include the concepts of the benefits and drawbacks of experimentation, I extend this model by added two terms in Equation 2. V_B represents the value of the benefits of experimentation over commitment, which for example can include incorporating learning from the experiment to improve the final outcome. In my setting this would be the case if piloted generated useful feedback from network executives that enables the creators of television shows to improve their shows. V_D represents the drawbacks to experimentation, if for example actors became harder to recruit, lowering final outcomes.

$$\max_{\lambda_i \in \{0,1\}} \lambda_i(\Pr(y(\theta_i) > C)\, E[y(\theta_i) + (V_B - V_D) - C | y(\theta_i) > C] - e) + (1 - \lambda_i)(\theta_i - C) \qquad (2)$$

How would this model capture the incumbent networks' reaction to Netflix's entry? Concern about the informativeness of the pilot signal has been used by network executives in the past to justify a limited number of straight to series orders (Nordyke, 2008): perhaps Netflix's entry reduced beliefs about usefulness of experimentation to identify poor ideas. Netflix could have changed how much the networks needed to compensate creators for experimenting, increasing cost e. Networks could have also decreased their beliefs about the net effect of experimentation on outcomes, either by dropping V_B or increasing V_D. Comparative statics on Equation 2 suggest any of combination of these potential consequences of Netflix's entry would have led to less piloting, see Appendix for a proof.

Proposition 1: If the signal technology y is ordered to be weakly increasing in the difference between experimentation and non-experimentation payoffs, experimentation modeled in Equation 2 is increasing in y, V_B,$-V_D$, and $-e$.

Observed outcomes for non-piloted shows under Equation 2 would be $E[\theta|\theta'' < \theta]$ and for piloted shows $E[y(\theta) + (V_B - V_D) - C|y(\theta) > C, \theta' < \theta < \theta'']$. Due to the conditional expectation form of these payoffs, the likelihood ratio order would be a natural ordering for distributions of θ. Of course, the direct interpretation of the likelihood ratio order would be that as θ_i increases for a given i, the probability i was drawn from A relative to B also increases if $A >_{LR} B$. In my television context this is also equivalent to saying a) all network's utility from a show weakly increase in θ and b) regardless of the exact shape of a network's utility function, for any given target range $[a, b]$ of show quality θ, networks would prefer to take the average draw from distribution A over B. The likelihood ratio order has two important consequences for interpreting Equation 2's empirical objects. First, if $A >_{LR} B$ then the observed average quality of committed shows drawn from A should be greater than those drawn from B: $E[\theta|\theta'' < \theta, \theta \in A] > E[\theta|\theta'' < \theta, \theta \in B]$. Second, unless the net improvement from experimenting $(V_B - V_D)$ is smaller for A than for B, the observed average quality of piloted shows drawn from A should also be greater than those drawn from B : $E[y(\theta) + (V_B - V_D)|y(\theta) > C, \theta' < \theta < \theta'', \theta \in A] > E[y(\theta) + (V_B - V_D)|y(\theta) > C, \theta' < \theta < \theta'', \theta \in B]$. Finally, if the set of observations were restricted to show to any particular subset of prior qualities, specifically those just below or above the cutoff θ'', shows drawn from A should be greater than those drawn from. Proofs are provided in the Appendix.

Proposition 2: If $A >_{LR} B$, then the expected value of non-experimental payoffs under Equation 2 is higher when θ is drawn from A then from B. If in addition the posterior belief conditional on passing the experiment phase is always weakly greater than the prior belief, the expected value of experimental payoffs is also higher when θ is drawn from A then from B whenever $V_B - V_D$ is the same for both distributions. If experimental payoffs are worse when drawn from A, $V_B - V_D$ must be lower for A than for B. This relationship holds as well for any set of outcomes shows around the cutoff in prior θ between experimented and non-experimented projects.

Comparing observed outcomes to detect whether $A >_{LR} B$ can be misleading. In the case for experimented shows, Proposition 2 states observations from B can be greater than A even when $A >_{LR} B$. This is also true for non-experimented shows. Equation 2 places the net effect $V_B - V_D$ on the experimentation payoff, but an alternate formulation might put this net effect on the non-experimentation payoff, implying non-experimented outcomes also cannot be trusted to determine whether $A >_{LR} B$. In contrast, looking only at the probability of passing the experimental stage would provide supportive evidence.

Proposition 3: Given a signaling technology that is increasing in θ, then if $A >_{LR} B$ the probability of a project passing the experimental phase is greater when θ is drawn from A then from B.

Differences in distribution of θ can be useful to uncover where V_D might be large enough to overcome the gains from experimentation so that piloting worsens outcomes. This requires an additional assumption on the signaling function that as prior beliefs about a show increases, the posterior belief about the show does not increase any faster.

Proposition 4: Given a signaling technology that maps prior θ to posterior $\psi(\theta)$ such that for all θ, $\psi(\theta) > \theta$ and for all $\theta' > \theta$, $\psi(\theta') > \psi(\theta)$ and $\psi(\theta') - \theta' < \psi(\theta) - \theta$, the differences between experimented and non-experimented outcomes is decreasing in the distribution of θ when the distributions of θ are ordered by likelihood ratio and $V_B - V_D$ is constant across distributions of θ.

Finally, what happens to the average outcomes of a portfolio of shows as a show is moved from piloting to straight to series production? A detailed proof is in the Appendix, but if there is no net benefit from commitment the portfolio of shows always worsens. The attrition effect of piloting is lost for the switched show, lowering the average outcomes for the portfolio unless the net gain from experimentation, $V_B - V_D$, is sufficiently negative.

Proposition 5: Let A be a set of experimented projects and B be a set of non-experimented projects. Assume the posterior belief conditional on passing the experiment phase is always weakly greater than the prior belief. If $V_B - V_D$ is weakly positive, the average outcome across both sets A and B can only worsen if a project is moved from A to B.

4.2 Evidence of Mechanism

Rather than model all possible mechanisms of how experimenting can affect outcomes, I characterize them into three broad categories. Those that that improve outcomes of the final stage relative the interim, experimental stage, those that affect the outcomes of both stages equality and those that improve the interim stage relative to the final stage relative. Although this limits the strength of my evidence for any particular mechanism, it does make my analysis tractable while still being insightful as to what mechanisms could be driving my main result.

Commitment improves later episodes relative to first episode

Consider an example of a mechanism that would improve the final stage relative to the interim stage, where experimentation distorts effort towards passing the interim stage and commitment removes that distortion. A basic multitasking model with two actions can represent the level of effort placed in the first episode a_1 and subsequent episodes a_2. The payoff to the show's creator the total quality of the show, here equal to the sum of effort $a_1 + a_2$, minus the cost of that effort, $a_1^3 + a_2^3$; a cubic cost function is used rather than the traditional quadratic to ensure cost grows fast enough for later modifications to have finite solutions.

$$\max_{a_1, a_2 \in \mathcal{R}^+} a_1 + a_2 - (a_1^3 + a_2^3)$$

The solution to this model is symmetric with equal levels of effort applied to a_1 and a_2. This symmetry also holds if rather than receiving the payoff $a_1 + a_2$ with certainty, the show creator only gets that payoff with probability $p \in (0,1)$; the network can decide not to broadcast the show.

$$\max_{a_1,a_2 \in \mathcal{R}^+} p * (a_1 + a_2) - (a_1^3 + a_2^3)$$

Effort levels become asymmetric when the pilot influences the probability of receiving the payoff. Assuming for the moment that $a_1 \in [0,1]$, optimizing

$$\max_{a_1,a_2 \in \mathcal{R}^+} a_1 * (a_1 + a_2) - (a_1^3 + a_2^3)$$

leads to a higher level of effort a_1 than a_2 as applying effort to the first episode not only improves the payoff $a_1 + a_2$ but also increases the probability of receiving that payoff.

Combining the two models by letting $\gamma \in [0,1]$ represent the relative importance of pilot effort to influence passing the pilot stage, the creator optimizes

$$\max_{a_1,a_2 \in \mathcal{R}^+} (p * (1 - \gamma) + a_1\gamma) * (a_1 + a_2) - (a_1^3 + a_2^3)$$

At $\gamma = 0$, effort is balanced between a_1 and a_2 but a gap grows with $a_1 > a_2$ as γ increases. When $p = 1$ and $\gamma = 0$, we can think of this model as one of commitment, with a distortion of effort towards the pilot as γ increases to represent the use of experimentation. Of course, in my setting we would naturally expect greater effort exerted in a_2 versus a_1 due to the sheer quantity of later episodes versus the first one, but we can think of the above model as representing a normalized effort that is balanced when the first episode does not influence the probability of receiving the payoff.

A more general version of this model would be of the form:

$$\max_{a_1,a_2 \in \mathcal{R}^+} p(a_1) * \pi(a_1, a_2) - c(a_1, a_2)$$

Assume the typical conditions for a solution to exist. Then shifting from a production model of commitment where $p'(a_1) = 0$ to a production model of experiment where $p'(a_1) > 0$ would correspond with a change from the symmetric solution of $a_1 = a_2$ to an asymmetric one of $a_1 > a_2$.

Commitment affects all episodes equally

In a planning model, one can imagine planning time a and an execution time b as a shares of total time spent in development.

$$\max_{a,b\in[0,1]} a + b - a^2 - b^2 \;\; s.t.\; a + b = 1$$

The optimal allocation a^*, b^* of time between these two tasks results in the best possible show. Suppose piloting a show constricts the amount of time that can be spent on planning before production to c, since the first episode must be produced according to the network's pilot schedule.

$$\max_{a,b\in[0,1]} a + b - a^2 - b^2 \;\; s.t.\; a + b = 1, a \leq c$$

If $a^* > c$, then piloting results in the suboptimal amount of time in planning. Switching to commitment would improve the quality of the show by benefiting all episodes.

A similar pattern of improvement would occur for many of the theories relating to commitment. All episodes would improve if the cast and crew of televisions shows preferred the stability of a straight to series show and were uncompensated for the disutility of piloting. Changes in the risk profile of projects principles or agents would again affect all episodes equally.

Commitment improves first episode relative to later episodes

A task model could have the first episode improve relative to the rest of the season under straight to series production if the effort was sequentially decided and creators felt effort did not influence passing the pilot phase, for example if creators believed network executives lacked the ability to judge pilot quality. Then the first episode gets lower effort under piloting because it may never get broadcast while subsequent episodes get higher effort. With a straight to series order, all episodes get this higher level of effort since creators know their work will be seen by an audience, so it's the first episode that benefits the most from the switch from real options to commitment.

$$\max_{a_1,a_2\in\mathcal{R}^+} pa_1 + a_2 - a_1^2 - a_2^2$$

A model where piloting helps creators improve the quality of the show could have the similar comparative statics if an overall improvement in all episodes from straight to series is also present. Suppose there is some characteristic about a show that is the creator's control which determines the quality of the show. In order for a show to be viewed as high quality, the level of that characteristic should match the desires of the audience, for example a drama show might need to have some action elements but too much action could detract from the plot and hurt the show's ratings.

Let θ be the level of the characteristic preferred by the audience, distributed F_{θ}. The network knows θ but the creator does not. Each creator has a belief γ about what the audience would like, distributed F_{γ}. Suppose there are two outputs, the first episode and the rest of the season, and the show's value is the sum of quadratic loss from the difference between audience taste and creator beliefs for those two outputs:

$$E[-(\gamma-\theta)^2]+E[-(\gamma-\theta)^2]$$

Now allow the network to adjust γ to equal θ by piloting with cost e. The network then maximizes whether to pilot $\lambda=1$ based on the trade-off between piloting cost and the expected loss from the rest of the season:

$$\max_{\lambda\in(0,1)} \lambda(E[-(\gamma-\theta)^2|\theta]-e)+(1-\lambda)\,(E[-(\gamma-\theta)^2|\theta]+E[-(\gamma-\theta)^2|\theta])$$

The network would decide to pilot when $E[(\gamma-\theta)^2|\theta]>e$.

In this model if the distributions are non-degenerate, the quality the episodes beyond the first will be better for piloted shows; non-piloted shows have a quality loss of $E[(\gamma-\theta)^2|\theta]$ while piloted shows have no quality loss. More interestingly, the reverse is true for the first episode. Since piloted shows have $E[(\gamma-\theta)^2|\theta]>e$, non-piloted shows must have $E[(\gamma-\theta)^2|\theta]<e$; $E[(\gamma-\theta)^2|\theta,\lambda=1]>E[(\gamma-\theta)^2|\theta,\lambda=0]$ so the first episode of non-piloted shows are better. Experimentation occurs on the relatively "bad" shows so the first episode of the piloted shows are worse. Experimentation improves those "bad" shows, so the subsequent episodes of the piloted shows are better. This paired with a benefit of commitment such as improved planning or higher quality workers could show an overall positive effect from straight to series production driven primarily by an improvement in the quality of the first episode.

4.3 Bargaining

Beyond showing that pilots were unnecessary to create a critically acclaimed television show, there is evidence that Netflix, as an entrant committed to skipping pilots, changed the bargaining relationship between incumbent networks and content creators (Adalian, 2013). However, a shift in bargaining power away from the networks doesn't necessarily mean networks are going to order shows straight to series more often. The appendix contains a more formal mathematical argument, but under a standard cooperative bargaining model everyone is always better off making decisions that maximize the total value generated by the group. Before Netflix's entry, the networks would only be selecting piloting if the expected payoff from piloting is higher than from straight to series. Because piloting generates more value and the network's decision does not affect the fraction of value allocated to itself, the networks should always prefer piloting. Netflix's entry just affects the share of surplus shared with show creators and the networks will always be better off with a share of the larger piloting surplus than the smaller straight to series surplus. Therefore, a change in bargaining power after Netflix's entry would not be able to explain an increase in straight to series production in a standard cooperative bargaining model.

But a standard cooperative bargaining model may not apply in this setting. Based on interviews with content creators, many prefer a straight to series order over a pilot order. A straight to series order ensures their work will be seen by the public whereas "if you do a pilot and it doesn't get picked up, you don't have anything", to quote *Sex and the City* director Allison Anders (Bunn, 2002). This preference for straight to series stands in contrast to the actions of the networks, which overwhelmingly chose to pilot shows. An equilibirum could have existed where all the incumbent networks piloted and left the straight to series preference of creators uncompensated (England, Farkas, Kilbourne and Dou, 1988). By committing to straight to series production, Netflix would have changed this equilibirum upon entry, increasing the rate of straight to series orders by incumbents. The networks would have been forced to compensate creators for their disutility from piloting since the networks would lose creators to Netflix without this compensation. If the piloting payoff on a particular show prior to Netflix's entry is not too much higher than the straight to series payoff, adding compensation to creators for piloting can make the network prefer the straight to series payoff after Netflix's entry.

This incremental shift in payoff is not the only mechanism that could increase straight to series orders in such a model. When a show's creator has high bargaining power, the network is already giving a large share of show rents to the creator. They may be constrained from giving an even larger share once Netflix enters to compensate for piloting; the creator's show might be considered a loss leader to draw in larger audiences for the other shows on the network. In this case, the network would switch to straight to series after Netflix's entry because simply because compensating for piloting is not possible.

More formal arguments are provided in the appendix, but this intuition suggests bargaining could be an alternative explanation to the decision process modeled in Equation 2 for observed network behavior. Depending on how bargaining power correlates with a show's quality, θ_i, Netflix's entry could have various effects on outcomes. Specifically, no correlation would explain a shift towards straight to series production without a corresponding drop in straight to series show quality.

> Proposition 5: Assume bi-literal bargaining occurs between project agents and the principles that fund the projects and that the shared payoff from experimentation is higher than from non-experimentation. Let there be a private cost to agents from experimentation that is excluded from the shared payoff calculation. Then there exists a subgame perfect equilibirum where all projects are experimented on. In addition, if one principle deviates and never experiments, agent bargaining power will determine which projects get experimented on for the other principles.

4.4 Implications for Empirics

The models above provide a set of testable hypotheses. First, according to Proposition 1, Netflix's entry could have changed beliefs about experimentation, shifted bargaining power towards creators, or perhaps done both. Regardless, the models predict the incumbent networks would have increased their use of straight to series orders as observed in Figure 1.

> Hypothesis 1: *Netflix's entry should have increased the amount of straight to series production by incumbents.*

Second, if the networks were making an optimal decision about which show to pilot and which shows to order straight to series, shows with certain ex-ante properties would have been selected into straight to series. If for example the network viewed the probability of a show's success or likelihood to pass the pilot phase to be high, then experimentation adds less value; the chance of observing a bad pilot could be too low to justify the incremental cost of creating a pilot. This would cause selection bias to be present in any comparison between observed outcomes for piloted and straight to series shows.

> Hypothesis 2: *Under the optimum decision model, networks are more likely to order shows straight to series that they believe have a higher likelihood of success.*

Third, according to Proposition 4, when the distribution of a set of potential shows is "better" in terms of likelihood ratio, the improvement piloting has in outcomes over not piloting should decrease. This could help uncover cases where piloting worsens outcomes. Based on Proposition 3, looking differences in the probability of a show passing the pilot phase is an indication of when a distribution could be better in this sense.

> Hypothesis 3: *Piloting should have a worse effect on outcomes for distributions of shows that are more likely to pass the pilot phase.*

Fourth, from Proposition 5, the selection bias mentioned above has no effect on the portfolio of shows, regardless of whether an optimal decision rule or bargaining is driving a chance in straight to series orders. This implies that if an improvement in portfolio outcomes is observed after Netflix's entry, commitment does improve show outcomes.

Hypothesis 4A: *If there is no benefit to commitment, a network cannot improve its portfolio of new shows by ordering more shows straight to series.*

Hypothesis 4B: *If a portfolio's outcomes improve after increasing straight to series production, commitment must render benefits to television show production, regardless of whether the networks are selecting ex-ante better shows for straight to series production.*

Finally, the improvement of the first episode relative to the rest of the season can provide clues as to some underlying mechanisms of how experimentation affects the distribution of outcomes.

Hypothesis 5A: *If the rest of the season improves relative to the first episode under straight to series production, piloting could be distorting incentives towards passing the pilot phase.*

Hypothesis 5B: *If the first episode improves relative to the rest of the season, piloting could lower incentives to invest in the first episode or learning from the pilot feedback could be improving later episodes.*

Hypothesis 5C: *If the first episode and later episodes are affected equally by straight to series productions, the mechanisms in 5A and 5B are less likely to be important relative to other mechanisms.*

5. Data and Measures

To empirically test the above propositions, I pool data from three sources: Film L.A., Gracenote, and the Internet Movie Database (IMDb). Film L.A. is a non-profit dedicated to facilitating film and television production in Los Angeles. They have a proprietary dataset which tracks the production of scripted US television starting at the pilot phase. Importantly, the dataset flags shows that were ordered straight to series phase, a variable crucial to this paper's analysis. Gracenote, a subsidiary of Neilson Holdings, has a dataset provided commercially to the television industry. A record is made whenever a network makes a public announcement of investment in a show idea, by for example paying a writer to produce a script. Metadata is associated with each show such as genre and creators responsible for the show's production. IMDb, a subsidiary of Amazon.com, has a public dataset which includes ratings for shows that made it to a public airing on a network.

Joining these three datasets is non-trivial because of variances in a show's title, year of production and network across the datasets. Since for example Film L.A. creates a show record earlier in the show's production history than IMDb, the title used in Film L.A. may be a working title, different from the official, release title used in IMDb. In both databases, spelling errors can exist in their title fields. Film L.A. tracks shows by their development season while IMDb records the year of a show's first broadcast; a show broadcast early in 2014 according to IMDb might be labeled as part of the 2013 development season according to Film L.A. Film L.A. tracks which network each show was developed for while IMDb's distributors for each show is often incomplete. Similar issues appear when comparing records between Film L.A. and Gracenote or Gracenote and IMDb.

To build my matched dataset, I first treat the smaller, curated Film L.A. dataset my main set of observations. I then build lists of alternative titles for each television show and match these alternative titles across the three datasets using bigram matching to allow for spelling errors. I prioritize records with exact matches across these alternative titles, year and network, but allow for deviations in year and network when exact matches are not available. Table 1 provides summary statistics for the combined dataset.

[Insert Table 1 about here]

The data is restricted to the incumbent networks that circa 2008 were consistently producing scripted television. This includes the prestige networks that would win Emmy or Golden Globe awards (ABC, NBC, FOX, CBS, HBO, FX, USA, Showtime and AMC) as well as other established networks known for creating original content (CW, Freeform, TNT, SyFy, Starz, and A&E). By restricting my dataset to these fifteen networks, my analysis is focused on the incumbents' reaction to Netflix's entry.

Since the Film L.A. dataset is only contains shows between 2008 and 2017, all observations of shows outside those years in the other datasets are not included. Data for 2017 is currently only partially available. The years used are season years; for example, the 2008 season year runs from September 2008 to August 2009.

The funnel from script to pilot to series is represented by the first few rows of Table 2. I restrict my data to show ideas that were developed in some way, either piloted or ordered straight to series. In the period prior to Netflix's entry, only 3% of shows were ordered straight to series, increasing to 15% after Netflix's entry. The genre and show length variables provide some indication of the stability of show types over my period of interest despite this change in production.

Based on Gracenote's data on the past work of a show's creators, I create a binary variable which indicates whether any of the show's creators have previously created a show that won a major Emmy or Golden Globe award. This constructed variable could be interpreted as a measure of a new show's uncertainty. A major concern for the networks is whether creators can "through skill and/or luck, manage to assemble all the right elements (cast, director, score, VFX, etc.) to perfectly execute the script they've written?" (Hawley, 2014); the creators with award-winning shows have a track record of having done so in the past.

I only observe outcomes for shows that were ordered to series, either by first being piloted or directly through a straight to series order. IMDb provides show ratings at both the show level and episode level for all broadcast shows. My primary outcome variable is an average of episode ratings for the show's first season. Table 2 shows correlation between IMDb ratings and two other measures of show performance: the renewal decision made by the networks and breaking into the Nielson Top 30. Renewals have historically been a strong indicator that a show met the network's internal metrics for success. Unlike raw viewership numbers, renewals factor in the value of reaching a specific demographic and the strength of a show's timeslot competition on rival networks. The connection between renewal, Nielson viewership and a show's success is not completely mechanical; sometimes a show will be considered a success if it for example increases cable subscriptions despite having relatively low viewership and lacking plans for renewal (Thaxton, 2017; O'Connell, 2018). Unlike viewership numbers from Nielson Media Research or a network decision to renew a show for another season, IMDb ratings are less mechanically linked to competitor outcomes and therefore relatively independent of market structure (Waldfogel, 2017). Yet IMDb ratings are still meaningful to a show's success; IMDb ratings strongly correlate with both viewership and renewal decisions across my broader dataset.

[Insert Table 2 about here]

6. Empirical Framework

6.1 OLS estimate

An OLS approach could be used to estimate the relationship between piloting and outcomes.

$$FirstSeasonRating_{int} = \alpha_n + \delta_t + \beta Piloted_i + X_i + \varepsilon_{int} \quad (3)$$

In Equation 3, $FirstSeasonRating_{int}$ is the average IMDB rating of a show's first season episodes, X_i includes any show level controls while α_n and δ_t are network and year fixed effects. β is the coefficient of interest that indicates the improvement in show quality from piloting, the experiment in my context.

Based on the model defined in Equation 2, Equation 3 would estimate

$$\beta = E[y(\theta_i)|y(\theta_i) > C, \theta_i < \theta'', Z_{int}] - E[\theta_i|\theta_i > \theta'', Z_{int}] + V_B - V_D \tag{4}$$

with Z_{int} representing the controls a_n, δ_t and X_i. The selection bias manifests itself in the conditional terms $\theta_i \gtrless \theta''$; ex-ante higher θ_i shows would be produced straight to series resulting in an inability to even sign the benefit of experimentation, as predicted in Hypothesis 2. However, if the controls Z_{int} are sufficient for the conditional independence assumption to hold, Equation 3 then estimates

$$\beta = E[E[y(\theta_i)|y(\theta_i) > C, \theta_i] - \theta_i]|Z_{int}] + V_B - V_D \tag{5}$$

In Equation 5, the first term is positive by the definition of the signaling technology used for experimentation, representing the attrition bias from the lack of observed outcomes for low quality pilots not getting funded to series. The attrition represents the improvement in outcomes from experimentation; to properly measure the effect of experimentation on outcomes, it should be included in my estimate. β therefore measures the average treatment effect in changing production: how much are broadcast show outcomes improved by switching from piloting to a straight to series production model? Since this first term is positive, the estimate of β can only be negative when the net effect of experimentation on outcomes, $V_B - V_D$, is sufficiently negative.

However, Equation 5 predicts a heterogeneous effect depending on the distribution of θ. β is predicted to be smaller when piloting is less likely to weed out bad show ideas, see Hypothesis 3. However, distinguishing such shows may not trivial. Equation 2 suggests estimating the probability a show would pass the pilot phase based on a set of covariates X_i could be biased due to the selection of shows into piloting:

$$\widehat{p_i} = E[p_i|X_i, \theta_i < \theta''] \neq E[p_i|X_i]$$

This bias could be limited by looking the period between 2008 and 2013, before Netflix entered, when almost all shows were piloted.

$$\widehat{p_i} = E[p_i|X_i, \theta_i < \theta'', t \leq 2013] \approx E[p_i|X_i, t \leq 2013] = E[p_i|X_i]$$

Therefore, a natural way to uncover variation in the probability of passing the pilot is to estimate for shows i that were piloted in the period t before Netflix's entry.

$$SeriesOrdered_{int} = \alpha_n + \delta_t + \beta X_i + \varepsilon_{int} \tag{6}$$

Although passing the pilot phase does not necessarily mean a positive signal was observed (or a negative signal when show does not pass the pilot phase), I assume passing the pilot phase is positively correlated with a positive signal.

[Insert Figure 3 about here]

Figure 3 suggests likely candidate for a strong covariate X_i in Equation 6: it plots the share of piloted shows that pass the pilot phase by whether any of the show's creators had a previous award-winning show. The data is restricted to the period prior to Netflix's entry when almost all shows were piloted. Shows with these award-winning creators had a greater chance of passing the pilot phase. Therefore, I use Equation 7 as a way to operationalize Hypothesis 3, expecting if piloting can worsen outcomes, I am most likely to observe that outcome using an interaction of piloting and shows with award-winning creators.

$$\begin{aligned} FirstSeasonRating_{int} = \alpha_n + \delta_t + \ \beta_S Piloted_i + \beta_A CreatorAward_i \\ + \beta_{SA} Piloted_i * \ CreatorAward_i + \varepsilon_{int} \end{aligned} \tag{7}$$

The two coefficients of interest are β_S and β_{SA}. β_S provides an estimate of Equation 4 when the benefit of piloting should be relatively higher. β_S should be greater than the average treatment effect measured in Equation 3. Conversely, β_{SA} estimates the change in Equation 4 when the benefit of piloting is lower: experimentation less useful in weeding out bad ideas. The sum $\beta_S +$

β_{SA} represents the overall effect of piloting on shows with creators having award-winning shows which should be lower than Equation 3's β. It is possible that β_S is positive while $\beta_S + \beta_{SA}$ is negative. Such heterogeneity in treatment effect would result in the practical recommendation that experiments are useful depending on uncertainty; when a manager is doubtful about the quality of a new product idea, experimentation is optimal. However, as uncertainty is reduced, a real options approach to experimentation could be detrimental. A manager executing on a product idea that is like other successful past ideas is better off skipping experimentation to instead capitalize on the benefits of commitment.

6.2 Matching estimate

Using Equations 3 and 7 to estimate β has two main drawbacks. First, it may not be reasonable to assume conditional independence to hold; there could be unobserved variables causing correlation between the observed variables and error term. Second, even if conditional independence holds, the linear assumption of OLS could be problematic. Saturating the control variables is not possible with my dataset's number of observations which can lead to selection bias. For example, suppose individual genres were included in Equation 3's X_i term but not combinations of genres and that sci-fi comedies were only observed as committed shows with award executives. Any estimate of β_{SA} could be picking up the fixed effect of a show being both a sci-fi and a comedy rather than measuring the return to experimentation.

Matching estimators can mitigate the second issue by looking at shows that are similar. The example's observations of sci-fi dramas would in theory be removed from a matching estimator because they would lead to imbalance in the share of sci-fi and comedy shows between the piloted and straight to series groups.

I use a propensity score based estimator for the average treatment effect (Hirano, Imbens and Ridder, 2003) since coarsened methods (Iacus, King and Porro, 2012) require a more favorable ratio of observations to covariates than exists in my dataset. Propensity score matching requires a two-stage estimator, first estimating how observables affect the probability a script will be piloted versus ordered straight to series:

$$StraightToSeries_{int} = \alpha_n + \delta_t + \gamma X_i + \varepsilon_{int} \tag{8}$$

Then observations are restricted to those where the predicted $\widehat{StraightToSeries}_{int}$ lies in a region of common support. Finally, Equations 3 and 7 are estimated within those common support observations, weighted by the inverse propensity score. The weights skews the estimates toward piloted shows that were predicted most similar to straight to series shows and straight to series shows most similar to piloted shows.

Intuitively, my matching estimator using incumbent network observations from 2014 to 2017 circumvents selection bias due to the uncertainty around optimality of experimentation following Netflix's entry. Some network executives like Kevin Reilly at Fox embraced straight to series production more than others like Nina Tassler at CBS (Adalian, 2013; Andreeva, 2014), resulting in close counterfactual shows in my dataset. The area of common support constitutes the set of shows that had the possibility of either being piloted or ordered straight to series, depending on the network executives in place at any time. The estimator measures the average treatment effect of piloting on shows within the area of common support.

6.3 Portfolio Approach

The matching estimator leaves selection on unobservables unmitigated as a source of bias. The television networks could be picking better shows for straight to series production on based on characteristics missing from my dataset so any observed improvement in outcomes from straight to series orders could be generated by the selection of ex-ante higher quality shows for pilots. I therefore use a difference in difference estimator at the network's portfolio level as my first specification:

$$
\begin{aligned}
&AverageFirstSeasonRating_{nt} \\
&\qquad = \beta_S ShareStraightToSeries_{nt} + \beta_{SP} ShareStraightToSeries_{nt} \\
&\qquad * Post2013_t + \delta_t + \alpha_n + \epsilon_{nt}
\end{aligned} \tag{9}
$$

Equation 9 approaches my dataset as a panel of network observations. For each year t, each network n releases a set of new shows. The average first season rating for those shows is my outcome variable $AverageFirstSeasonRating_{nt}$. Some of those new shows were ordered straight to series, while others were the result of a piloted, staged development process. The share of the network's shows that year that were ordered straight to series is reflected in the dependent variable $ShareStraightToSeries_{nt}$. Netflix's entry into the industry increased the incumbents share of straight to series orders, this time shock is represented by the $Post2013_t$ variable. Fixed effects for networks α_n control for the differences in average show quality across networks; for example, on average HBO shows tend to be rated higher than shows on the big four networks. Fixed effects for year δ_t control for time trends in show quality, if for example the industry is overall getting better at producing higher quality shows at the end of my time period relative to the beginning.

Hypothesis 4B asserts that if β_{SP} is positive, straight to series orders must have some net benefit to the quality of shows regardless of whether networks are selecting ex-ante better shows for straight to series production.

6.4 Big Four Networks

ABC, NBC, CBS and Fox are different from the rest of the U.S. networks. They each fund over a dozen pilots per year, have a long track record of airing original programming, share a history of transitioning from over-the-air broadcasting with affiliate networks to reaching most of their audience through cable, and directly compete for prime-time advertising spend. They also have a much higher rate of using creators with previous award-winning shows, as depicted in Figure 4.

[Insert Figure 4 about here]

If as expected in Equation 7 that shows with creators having award-winning shows will have a different benefit from experimentation, we should see a corresponding difference in outcomes from a shift towards straight to series between the big four networks and other networks. To do so I extend Equation 9 to a triple difference estimator.

$$\begin{aligned} AverageFirstSeasonRating_{nt} \\ &= \beta_S ShareStraightToSeries_{nt} + \beta_{BP} BigFour_n * Post2013_t \\ &+ \beta_{SP} ShareStraightToSeries_{nt} * Post2013_t \\ &+ \beta_{BSP} BigFour_n * ShareStraightToSeries_{nt} * Post2013_t \\ &+ \delta_t + \alpha_n + \epsilon_{nt} \end{aligned} \tag{10}$$

In Equation 10, the depended variable $BigFour_n$ is introduced, set to 1 for the big four networks and 0 otherwise. β_{BSP} captures the change in big four network outcomes relative to the other networks.

6.5 Mechanism behind value of commitment

There are several mechanisms that could cause an improvement in outcomes when experimentation is avoided. One way of divining which mechanisms are more likely present in my setting is to break out the effect of ordering a show straight to series by each broadcast episode, as outlined in Hypotheses 5A to 5C.

$$\begin{aligned} EpisodeRating_{eint} \\ &= \lambda_e EpisodeNumber_e + \beta_e EpisodeNumber_e \times Piloted_i \\ &+ \alpha_n + \delta_t + \varepsilon_{eint} \end{aligned} \tag{11}$$

In Equation 11, i still represents a show but now e represents one of the show's episodes, specifically an episode that was part of a show's first season order: the follow-on series order if a show was originally piloted or the initial straight to series order if show was not piloted.

$EpisodeNumber_e$ is the ordinal number of an episode's broadcast. λ_e picks up a trend for how the IMDB rating of a piloted show evolves over its episodes while β_e, the coefficient of interest, is how the quality of straight to series shows evolve differently than piloted shows.

7. Empirical Results

Column 1 of Table 3 estimates Equation 3 without any fixed effects or covariates. Overall, we see piloting correlates with worse outcomes in my data, consistent with commitment being important determining for final outcomes after experimentation. However, as observables are added in Column 2 the relationship attenuates. Using Oster's (2016) approach to adjusting estimates for selection on unobservable based on the amount of selection on observables, the bias adjusted point estimate relationship between piloting and ratings is positive, suggesting my OLS estimates suffer from strong selection bias.

[Insert Table 3 about here]

To mitigate this selection bias, I use a propensity score matching estimator. Table 4 shows the propensity score estimate of Equation 8. The strongest predicator of straight to series is whether the show received independent funding. Such funding would enable a show to bypass the network's usual funding process of first piloting a show so it's natural for these shows to be ordered straight to series. The intuition behind my matching estimator is evident in the difference between Columns 1 and 2 as year and network fixed effects are added and the estimate's pseudo R^2 increases: in the period after Netflix's success, the industry had not converged on a reevaluation of straight to series show production. Some network executives were exploratory in increasing straight to series production more than others, leading to similar shows on different networks in different years receiving different types of production orders. The propensity score estimator is used to find a region of common support: shows that were not certain to be piloted or ordered straight to series but could have gone either way depending on the circumstances of the network at the time.

[Insert Table 4 about here]

Figure 5 provides a visualization of which observations are being dropped for common support based on this propensity score estimate. The matching estimator focuses on the difference in outcomes between straight to series shows that were predicated to be piloted and piloted shows that control for those straight to series shows. Table 5 shows how the covariates used in the propensity score logistic prediction equation differ between the full sample and matched sample. Overall balance is improved and although differences between straight to series and piloted shows remain, none of the differences in means is statistically significant at the 10% level.

[Insert Table 5 about here]

[Insert Figure 5 about here]

Returning to the matching results of Table 3, Column 3 again estimates Equation 3 without any fixed effect or covariates, now using inverse probability weights within the region of common support. As in Column 1, a negative correlation between piloting and IMDb ratings is exposed. However, unlike in Column 2, Column 4 shows this relationship strengthens as fixed effects and covariates are added, suggesting selection on unobservables is less of concern within this matched sample.

Equation 7's exploiting of a show's association with creators having prior award-winning shows is supported in Column 1 of Table 6, which display a strong, positive relationship between these creators and passing the pilot phase.

[Insert Table 6 about here]

Interpreting this correlation as indicating pilots and therefore experiments are less valuable for these creators, Column 3 of Table 7 estimates Equation 7 using the matching estimator. The relationship between piloting and outcomes presented in Column 1 is revealed to be primarily driven by creators with award-winning shows. Piloting shows without such creators improves outcomes, consistent with the winnowing effect of experimentation dominating any drawbacks from the lack commitment. In contrast, piloting shows with strong creators lowers show ratings by a full standard deviation. Based on the point estimate of Table 2, the treatment effect of piloting a show with creators with previous award-winning shows lowers the chances of renewal by 10%, material to the incumbent networks since on average only about 55% of new shows are renewed between 2014 and 2017.

[Insert Table 7 about here]

Column 1 of Table 8 estimates Equation 9. According to Hypothesis 4B, a positive coefficient *Share StS * Post 2013* would indicate that straight to series production improves outcomes. The results in Column 1 lack statistical significance, indicating any overall affect from ordering shows straight to series is too small to be detectable in my data.

Column 2 of Table 8 estimates Equation 10, effectively breaking out the *Share StS * Post 2013* variable of Column 1 by whether the network was one of the big four networks. Column 2's *Share StS * Big Four * Post 2013* is both positive and significant, suggesting the big four networks fared far better than the other incumbent networks after Netflix triggered an increase in straight to series production. Since the big four networks more often used creators with previous award-winning shows as depicted in Figure 4, this is supportive a causal relationship between such shows and higher IMDb ratings when ordered straight to series as estimated in Table 3.

[Insert Table 8 about here]

However triple difference estimators like Equation 10 have point estimates that can be challenging to interpret due to the multiple interactions involved. Figure 6 provides some intuition behind Column 2 of Table 8's results by plotting the average rating of each network's new show portfolio by year. For most of the networks in my dataset, the post 2013 shift towards additional straight to series production traces out a slight drop in average portfolio quality. But for the big four networks, the shift increased average portfolio quality. The strong, positive point estimate in Column 2 of Table 8 is indicative of this shift, which if traced out to a full portfolio of straight to series shows suggests an improvement similar to the matching point estimate of Table 3.

[Insert Figure 7 about here]

Figure 7 plots the change in episode quality from piloting using a matching estimator based on Equation 11 that sets the first episode as the base quality level. The later episodes seem to increase in quality over the season, suggesting the mechanism behind the downside of experimentation may involve distorting effort. Piloting could incentivize a show's cast and crew to focus on passing the pilot phase rather than make the highest quality show possible. This trend also is suggestive that the earlier estimates of the difference in show or portfolio outcomes due to piloting are not driven solely by selection bias; selection into piloting based on ex-ante show attributes would likely affect all shows equality without such a heterogenous effect between episodes.

The trend depicted in Figure 7 could help explain why the incumbent television networks did not make more use of straight to series before Netflix's entry, given that increasing straight to series production seems to have improved outcomes. Since the early days of television, advertisers have pre-paid for slots on new television shows, the pre-payment being used to fund the development of those shows. In return, the networks guarantee to deliver a certain number of viewers to those advertisers. When a new show falls short, the networks must provide advertisers with "make goods"; free advertising on their network until the gap is closed (Vogel, 2015). A new show that underperforms therefore eats away at a network's revenue by forcing the network to give away advertising it would otherwise charge for, leading the network to kill shows early that miss targets (Angwin and Vranica, 2006). The networks bias their evaluation of a broadcasted show's success towards the outcome of the show's first episode and therefore miss the benefits commitment provides in improving later episodes. Even if this gain from commitment was known to the networks' prior to Netflix's entry, they may have lacked the ability to revise relational contracts with advertisers to measure quality later in the season and realize those gains from commitment (Gibbons and Henderson, 2012).

Finally, Figure 8 presents the relationship between the background of a show's creators, whether a show was piloted, and whether at least one of the two main actors on the show had previous experience as a main actor. The development process of television shows typically follows a timeline where the decision to pilot or order a show straight to series is done before casting. Figure 8 illustrates that creators with stronger backgrounds attract more of these veteran actors, supportive of anecdotal evidence that veteran actors are preferred in television show production (Anders, 2015). Figure 8 also shows piloted shows in general have a harder time attracting such talent relative to straight to series orders, which supports anecdotal evidence that actors prefer for long term contracts (Robinson, 2018). This preference for long term contracts is another potential mechanism behind the drop in IMDb ratings when shows with strong creators are piloted.

8. Conclusion

In this paper I use piloting process used in the television show industry to study the value of experimentation through staged development. Using an estimator that matches piloted shows with shows that were not piloted, my main result is that skipping experimentation is positively correlated with outcomes for shows with creators that had strong track records. Netflix's entry as shock to the decision to pilot, I corroborate this result with a network level analysis, suggesting my main result may not be driven solely by selection bias and a causal interpretation could be warranted. Furthermore, I provide evidence on two mechanism behind such results. Improved outcomes from skipping experimentation may derive from changed incentives: a show's creators no longer overinvest in the pilot to ensure their show passes the pilot phase. Or the improved outcomes could derive from being able to attract higher quality talent: actors prefer long term contracts over short ones.

These results can apply broadly to several areas within strategy. One is the inherent tension between the strategic approaches of real options and commitment. As small investments which, depending on what is learned over time, can optionally lead to a larger, follow-on investments (Bowman and Hurry, 1993), pilots can be thought of as a real option on a television show. Real options theory predicts pilots would unconditionally improve observed outcomes as potentially poor outcomes are filtered out at early stages: having been developed in financial contexts, real options approaches lacks a mechanism whereby the act of experimenting can itself affect outcome (Adner and Levinthal, 2004). Theories describing the consequences of commitment (Ghemawat, 1991) fill this gap, predicting piloting can the adversely affect outcomes. My empirical results show this tension is not just a theoretical one: neither real option nor commitment approaches to strategic decision making should be applied without considering their joint consequences.

Whether to evaluate research projects over short or long time horizons is a key question for research on the innovation production function (Manso, 2011; Azoulay, Graff Zivin and Manso, 2011). When funding an idea for a new television show is cast in terms of a short term pilot investment versus a long term full season investment, my results contribute to this innovation literature. One of the mechanisms I uncover in my setting has a close parallel in existing work on research environments: short term incentives can distort the actions taken by researchers (Manso, 2011). My other mechanism, worker preference for long term contracts, may be an important but less understood part of the innovation production function.

Entrepreneurship is a third area where my results apply, where early stage venture capital investments are thought of as experiments (Kerr, Nanda and Rhodes-Kropf, 2014). My results suggest venture capitalists should fund startups created by an unproven groups of recent college grads for short periods of time while funding experienced serial entrepreneurs for a longer period. Although there is some potential loss in the lack of experimentation on serial entrepreneurs, this loss could be outweighed by incentivizing the entrepreneur to focus on the startup's long-term success rather than the requirements to meet its next short-term funding goals.

This paper's focus on commitment is timely as real options approaches espoused through movements like Lean Startup and implemented with anecdotal success in the technology industry begin to get transferred to other industries. The preference for procuring a real option before making large investments could be shifting funding away from necessary innovations that require long term investments which preclude experimentation, as in the clean energy industry (Gaddy, Sivaramand O'Sullivan, 2016; Burger, Murray, Kearney and Ma, 2018) and the pharmaceutical industry (Budish, Roin and Williams, 2016). The preference could also be forcing experimenting that is harmful to final outcomes, by for example starting new schools under the expectation that failure is expected, with long term consequences to the students enrolled in those schools (Duane, 2018). Real options theory may have well served the technology industry with its low cost of experimentation and perhaps low upside to commitment. However, these parameters may not exist in other industries. The use of approaches like Lean Startup need to be evaluated on a case by case basis as to not overlook the downside of experimentation: the lack of commitment.

9. References

Adalian, J., 2013. Straight to Series: The Networks' Big-Money Bet to Skip Pilots. *Vulture*. November 27th.

Adner, R. and Levinthal, D.A., 2004. What is not a real option: Considering boundaries for the application of real options to business strategy. *Academy of management review*, *29*(1), pp.74-85.

Anders, C., 2015. The Whole Crazy Process of Creating A TV Show, From Pitch To Pilot. *Gizmodo*. January 23rd.

Andreeva, N., 2011. FIRED! Charlie Sheen Axed From 'Two And A Half Men', He Fires Back & Vows To Sue. *Deadline Hollywood*, March 7th.

Andreeva, N., 2014. ABC Orders 'Designated Survivor' Series Starring Kiefer Sutherland From David Guggenheim, Mark Gordon, Simon Kinberg. *Deadline Hollywood*, December 14th.

Andreeva, N., 2014. CBS' Nina Tassler: "Pilot Season Is Not Perfect But Works For Us". *Deadline Hollywood*, January 15th.

Azoulay, P., Graff Zivin, J.S. and Manso, G., 2011. Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, *42*(3), pp.527-554.

Baker, G.P., 1992. Incentive contracts and performance measurement. *Journal of political Economy*, *100*(3), pp.598-614.

Barnes, B., 2006. Reality TV: NBC Universal to Slash Costs In News, Prime-Time Programs. *The Wall Street Journal*. October 19th.

Bhattacharya, Sudipto, Kalyan Chatterjee, and Larry Samuelson, 1986, Sequential research and the adoption of innovations, *Oxford Economic Papers* 38, 219–243.

Blumenthal, H.J. and Goodenough, O.R., 2006. *This business of television*. Billboard Books.

Bonhomme, S. and Jolivet, G., 2009. The pervasive absence of compensating differentials. *Journal of Applied Econometrics*, *24*(5), pp.763-795.

Bowman, E.H. and Hurry, D., 1993. Strategy through the option lens: An integrated view of resource investments and the incremental-choice process. *Academy of management review*, *18*(4), pp.760-782.

Brembilla, P., 2014. Straight-to-Series in Broadcast Television: Causes, Issues and Consequences.

Budish, E., Roin, B.N. and Williams, H., 2016. Patents and research investments: Assessing the empirical evidence. *American Economic Review*, *106*(5), pp.183-87.

Bunn, A., 2002. The Wasteland. *The New York Times Magazine*. June 23rd.

Burger, S., Murray, F., Kearney, S., and Ma, L.. The Investment Gap that threatens the planet. *Stanford Social Innovation Review.* Winter 2018.

Canfield, David. "How CBS's 2016 Lineup Catered to Trump's America." *Slate Magazine*, Slate, 29 Nov. 2016.

Dawn, C.C, 2013. Ted Sarandos upends Hollywood with Netflix revolution, *Los Angeles Times*. August 25th.

Delmar, F. and Shane, S., 2003. Does business planning facilitate the development of new ventures?. *Strategic management journal*, *24*(12), pp.1165-1185.

Dillon, N., 2012. #Won! Charlie Sheen's comeback show nears pickup, *New York Daily News*. August 10th

Duane, Daniel 2018. “How the Startup Mentality Failed Kids in San Francisco.” *Wired*, Conde Nast, 10 July

England, P., Farkas, G., Kilbourne, B.S. and Dou, T., 1988. Explaining occupational sex segregation and wages: Findings from a model with fixed effects. *American Sociological Review*, pp.544-558.

Farquharson, V., 2008. Whole lot of Chaykin going on; The actor is more than kind to his new sitcom, but he's not so fond of a certain other Canadian series. *National Post*. August 21

Fontenay, C.C. and Gans, J.S., 2014. Bilateral bargaining with externalities. *The Journal of Industrial Economics*, *62*(4), pp.756-788.

Fritz, B. and Flint, J. 2017. Where Amazon Is Failing to Dominate: Hollywood. *Wall Street Journal*. 6 Oct

Gaddy, B., Sivaram, V. and O'Sullivan, F., 2016. Venture Capital and Cleantech: the wrong model for clean energy innovation. *In MITEI-Working Paper No. 2016-06*.

Ghemawat, P., 1991. *Commitment*. Simon and Schuster.

Hawley, Noah. (2014) “'Fargo' Showrunner on a Straight-to-Series Order: It's the 'Holy Grail' (Guest Column).” *The Hollywood Reporter.* 20 Feb.

Holmstrom, B., 1983. Equilibrium long-term labor contracts. *The Quarterly Journal of Economics*, pp.23-54.

Iansiti, Marco, “How the Incumbent Can Win: Managing Technological Transitions in the Semiconductor Industry,” Management Science, Vol. 46, No. 2, February 2000, pp. 169-185.

Kerr, W.R., Nanda, R. and Rhodes-Kropf, M., 2014. Entrepreneurship as experimentation. *Journal of Economic Perspectives*, *28*(3), pp.25-48.

Kogut, B., 1991. Joint ventures and the option to expand and acquire. *Management science*, *37*(1), pp.19-33.

Leonard-Barton, D. 1992. Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal*, 13, 111-125.

Lerner, Joshua, and Julie Wulf. 2007. "Innovation and Incentives: Evidence from Corporate R&D." *Review of Economics and Statistics* 89(4): 634-644.

Luo, H., 2014. When to sell your idea: Theory and evidence from the movie industry. *Management Science*, *60*(12), pp.3067-3086.

MacCormack, A., Verganti, R. and Iansiti, M., 2001. Developing products on "Internet time": The anatomy of a flexible development process. *Management science*, *47*(1), pp.133-150.

Milgrom P, Roberts J (1990) The economics of modern manufacturing: Technology, strategy, and organization. *American Economic Review.* 80(3):511–528.

Nanda, R. and Rhodes-Kropf, M., 2016. Financing Entrepreneurial Experimentation. *Innovation Policy and the Economy*, *16*(1), pp.1-23.

Nordyke, Kimberly, 2008. "NBC's Zucker Outlines New Approach to TV Programming." *Reuters*, 30 Jan.

Nussbaum, Emily, 2018. "How Ryan Murphy Became the Most Powerful Man in TV." *The New Yorker*, The New Yorker, 8 May

Ocasio, A. 2012. "TV Success Rate: 65% Of New Shows Will Be Canceled (& Why It Matters)" *Screen Rant.*

O'Connell, Michael, 2018. "Showtime Boss Talks Subscriber Growth, 'Twin Peaks' and Mark Halperin." The Hollywood Reporter. January 6th.

Roberts, K. and Weitzman, M.L., 1981. Funding criteria for research, development, and exploration projects. *Econometrica: Journal of the Econometric Society*, pp.1261-1288.

Robinson, T., 2018. What Keeps You Alive shows the importance of art in horror films. *The Verge*. Aug 24rd.

Schmalensee, R., 1978. Entry deterrence in the ready-to-eat breakfast cereal industry. *The Bell Journal of Economics*, pp.305-327.

Shaked, M. and Shanthikumar, J.G., 2007. Stochastic orders. Springer Science & Business Media.

Smith, M.D. and Telang, R., 2016. *Streaming, sharing, stealing: big data and the future of entertainment*. Mit Press.

Thaxton, Ryan, 2017. "Twin Peaks Breaks Showtime Subscription Records, Yet Notches Mediocre Ratings." *Pastemagazine.com*. May 23rd.

Thomke, S.H., 1998. Managing experimentation in the design of new products. Management Science, 44(6), pp.743-762.

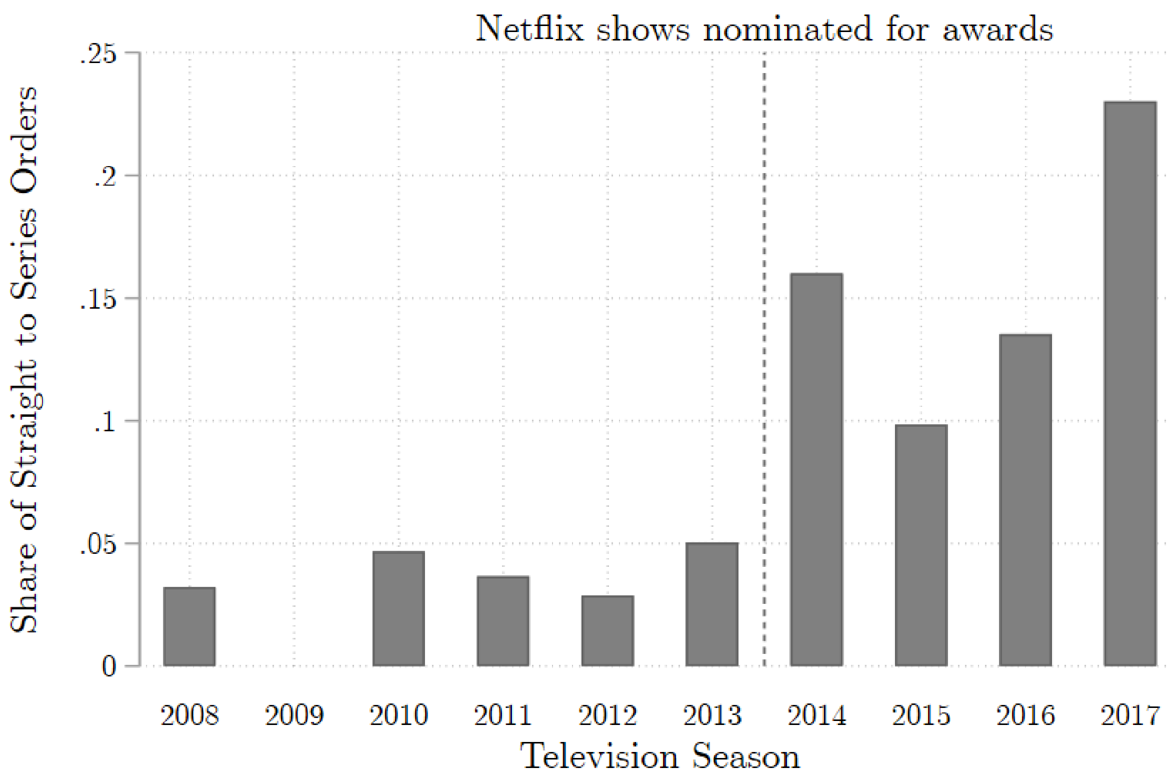
Waldfogel, J., 2017. The random long tail and the golden age of television. *Innovation Policy and the Economy*, *17*(1), pp.1-25.

Weisman, A., 2014. The 3 Reasons Netflix Decided On Its Straight-To-Series Viewing Model. *Business Insider*. June 25th.

10. Figures and Tables

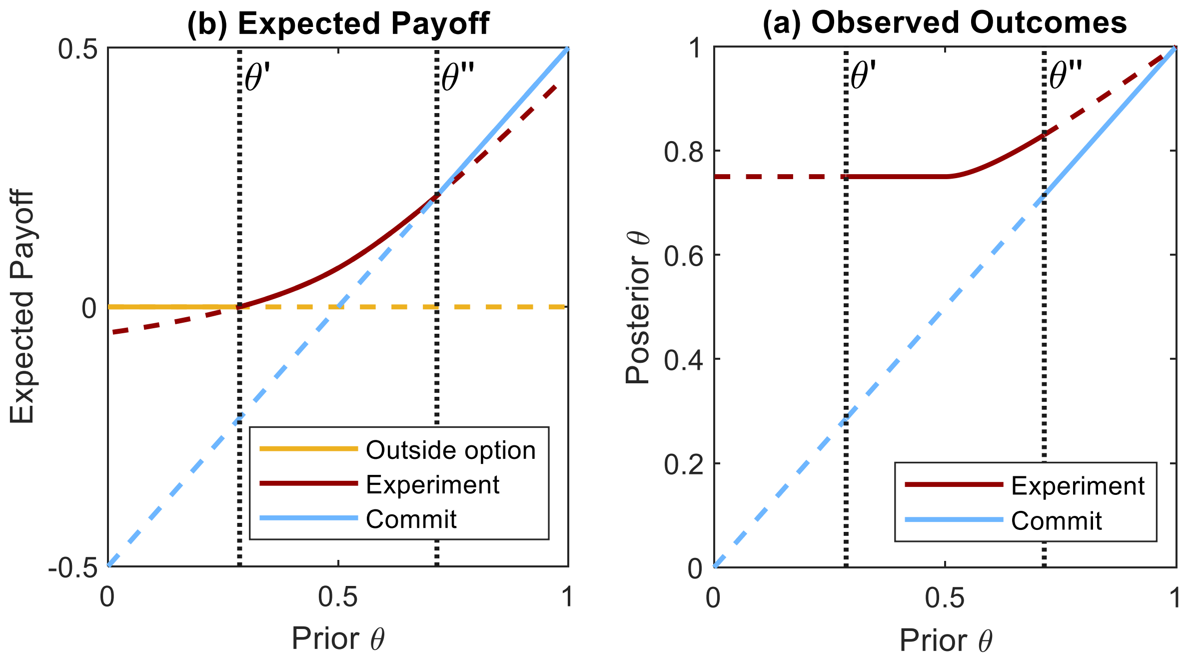
10.1 Figures

Figure 1. Share of Shows Ordered Straight to Series



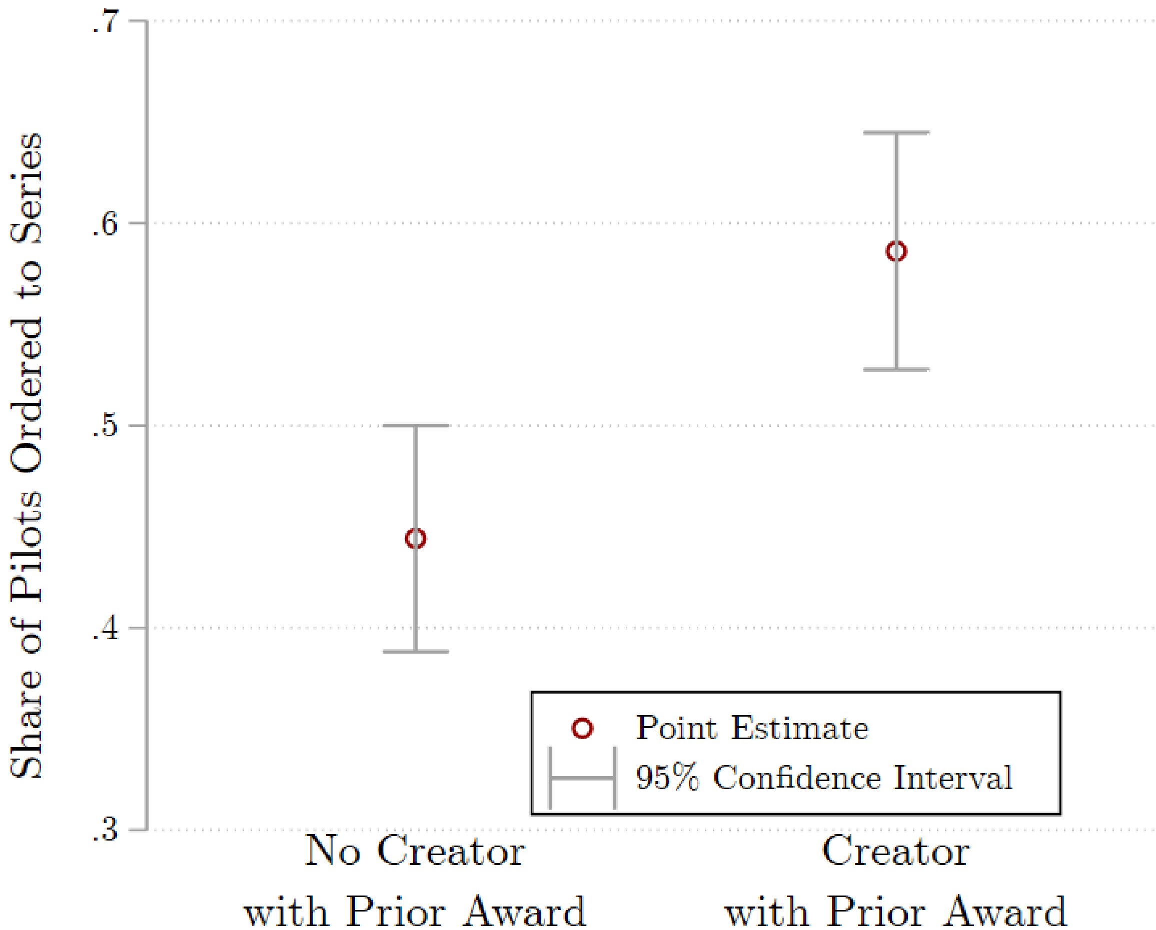
Scripted shows on incumbent networks that were either piloted or ordered directly to series.

Figure 2. Example Payoffs from Experimental Decision Making



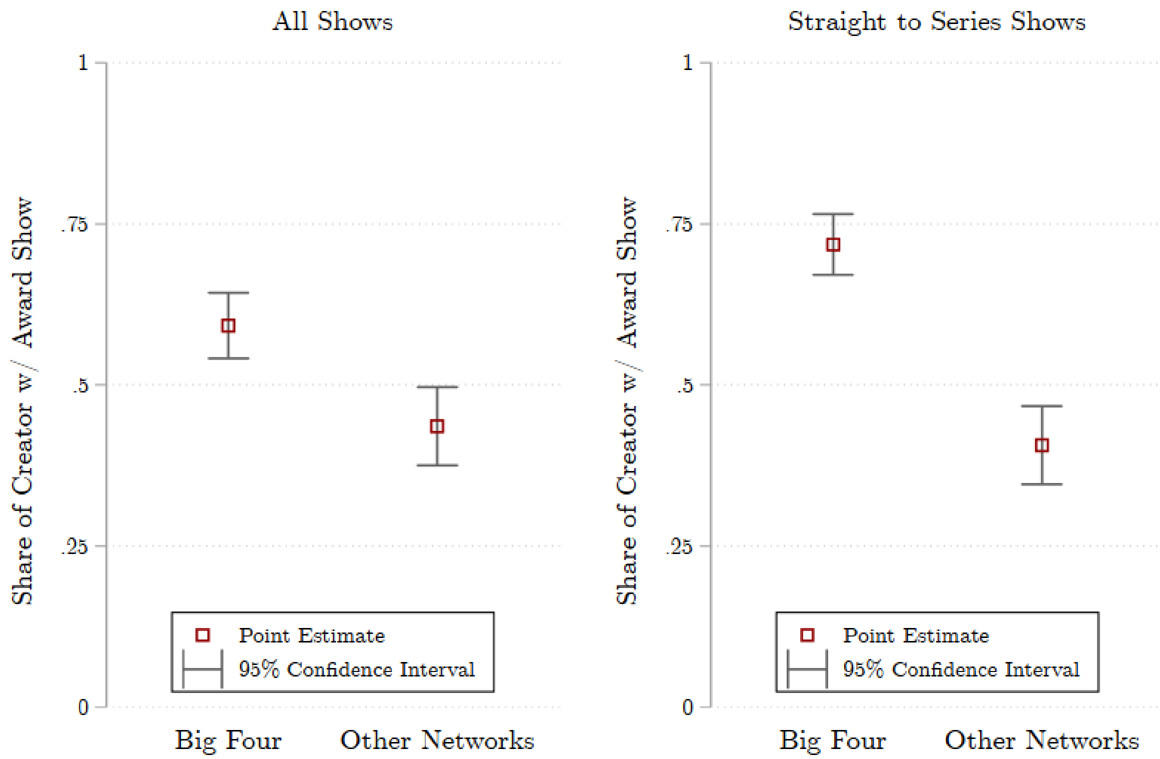
Assumes θ domain is $[0,1]$, project cost $C = 0.5$, experimental cost $e = 0.05$, $V_B - V_D = 0$ and a signal technology whose posterior distribution is mixture of uniform densities in ranges $[0, \theta]$ and $[\theta, 1]$ with mixture weights set so unconditional expected value of signal is θ. Panel (a) plots expected payoff conditional on prior belief θ. Optimum decision envelope represented by solid line. Panel (b) plots the observed outcomes under each production process with the optimum decision areas represented by solid lines.

Figure 3. Rate of Shows Passing Pilot Phase



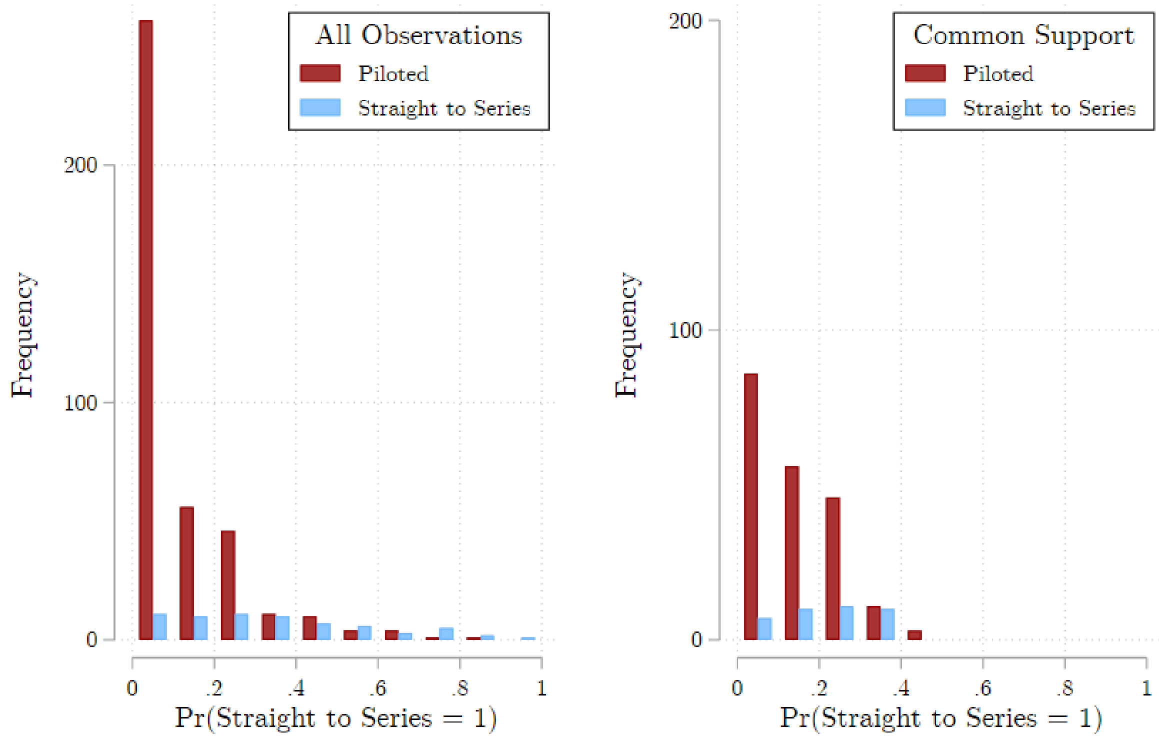
Piloted shows on incumbent networks from 2008 to 2013. Shows having creator with prior award have at least one creator with a previous show that one a major Emmy or Golden Globe award.

Figure 4. Share of Shows with Creators having previous Award-Winning Show



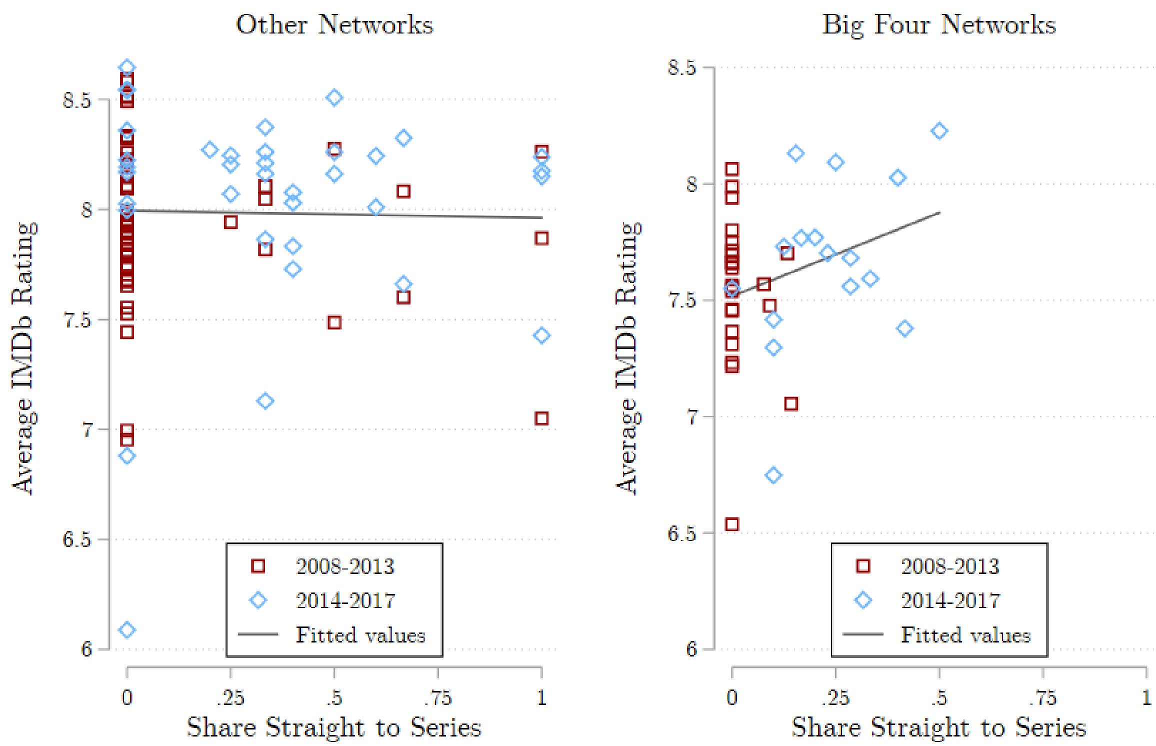
Point estimates and confidence intervals for the share of ordered shows with a creator that had a previous show which won a major Emmy or Golden Globe award. Big four networks are ABC, NBC, CBS and Fox. Observations restricted to US incumbent networks from 2008 to 2017.

Figure 5. Prediction of Straight to Series for Inverse Probability Weighting



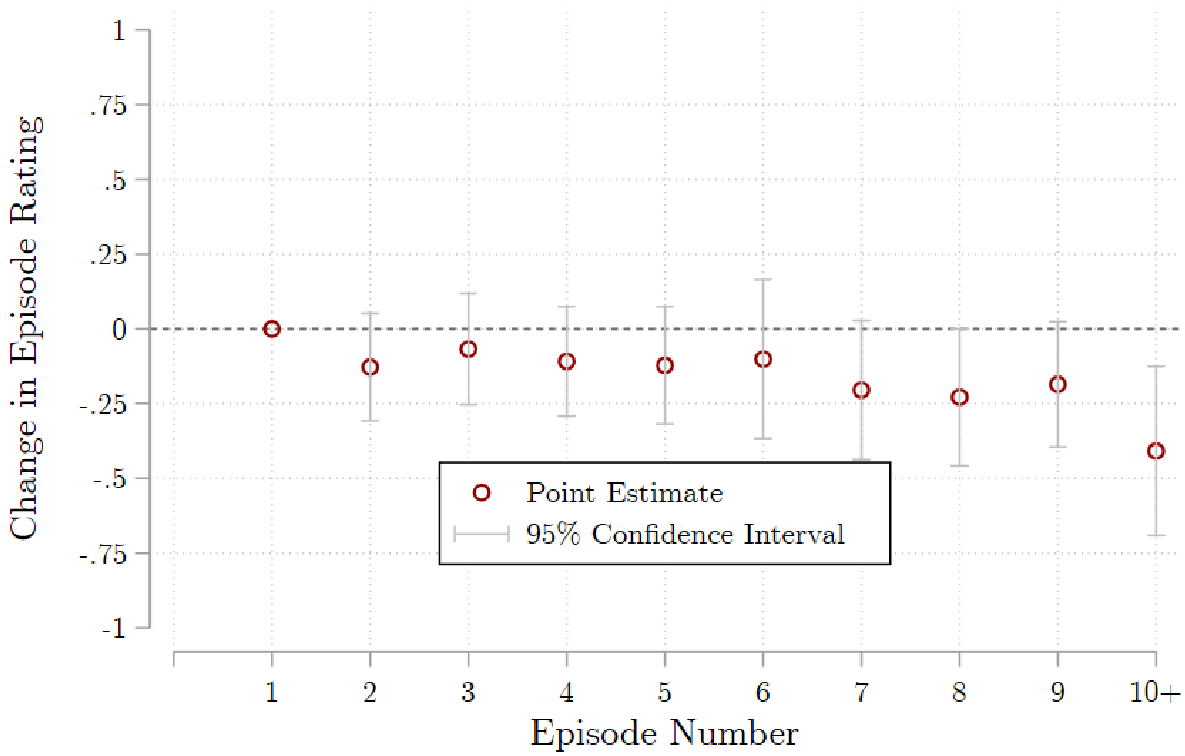
Distribution of predicted probability a piloted show will be ordered straight to series, based on incumbent network observations from 2014 to 2017. Common support cutoffs set to 5$^{\text{th}}$ percentile of straight to series predicted probabilities and 95$^{\text{th}}$ percentile of piloted predicted probabilities.

Figure 6. Portfolio Relationship between Straight to Series and IMDb Ratings



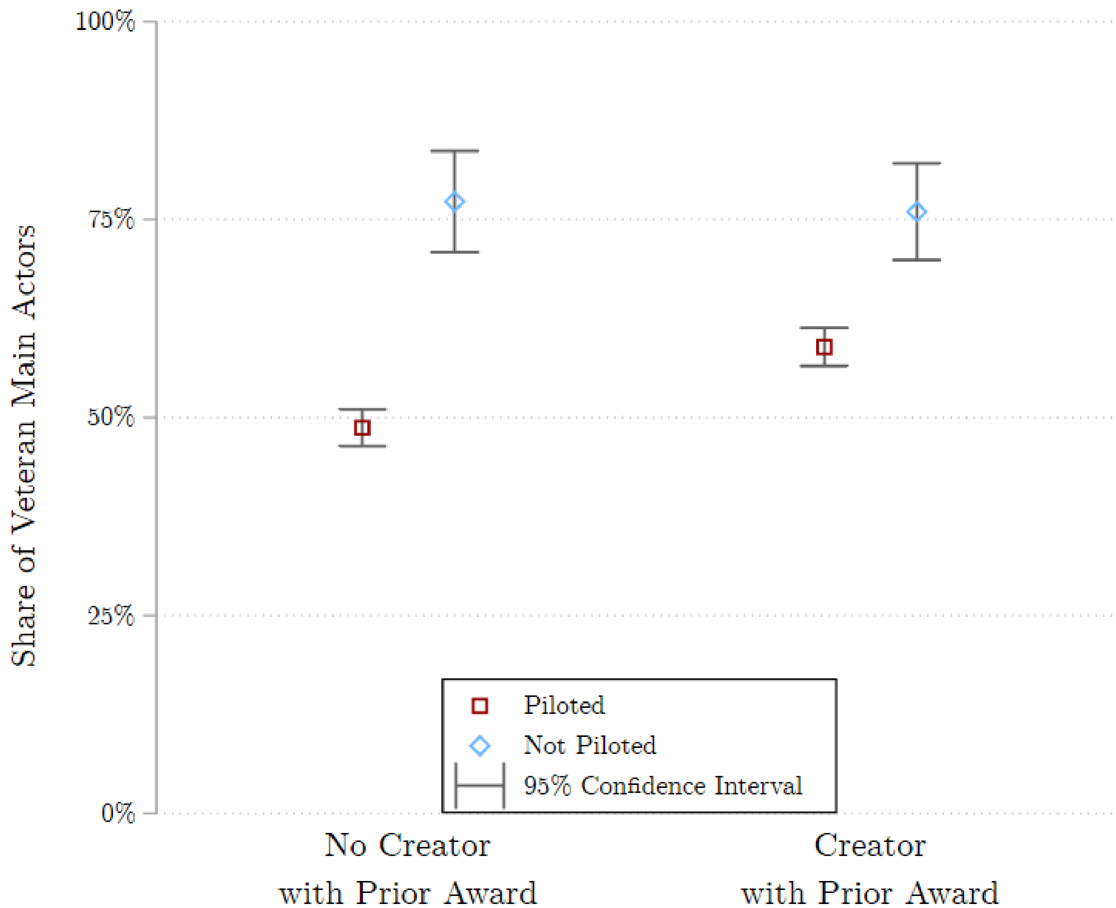
Plots of each network's yearly new show average IMDb rating against share ordered straight to series. Big four networks are ABC, NBC, CBS and Fox. Observations restricted to US incumbent networks for 2008 to 2017.

Figure 7. By Episode Relative Change in IMDb Ratings for Piloted Shows



Difference between piloted and straight to series shows episode with first episode set as baseline. Observations of ordered shows between 2014 and 2017 on incumbent networks weighted by inverse propensity score to match straight to series and piloted shows over area of common support, clustered at the show level.

Figure 8. Relationship between Veteran Actors, Piloting and Creators w/ Awards



Piloted shows on incumbent networks from 2008 to 2013. Shows having creator with prior award have at least one creator with a previous show that one a major Emmy or Golden Globe award.

10.2 Tables

Table 1. Summary Statistics Before and After Netflix's Entry

	2008-13	2014-17	Overall
Piloted or ordered shows	620	502	1122
Average per year	103.3	125.5	112.2
Piloted shows	599	425	1024
Average per year	99.8	106.3	102.4
Piloted shows ordered to series	317	200	517
Average per year	52.8	50.0	51.7
Straight to series ordered shows	21	77	98
Average per year	3.5	19.3	9.8
Is a drama	58%	57%	57%
Hour long show	59%	60%	59%
Has creator with prior award	47%	49%	48%
Ordered shows	338	277	615
Average per year	56.3	69.3	61.5
Mean IMDb first season rating	7.7	7.8	7.8
Standard deviation in IMDb first season rating	0.7	0.8	0.7
Renewed past initial order	64%	56%	60%

Table 2. Correlation between IMDb Rating and Other Outcome Measures

	(1) Renewed or Extended	(2) Nielson Top 30 Show
1st Season IMDb Rating	0.181***	0.0579***
	[0.0262]	[0.0167]
Constant	-0.802***	-0.356***
	[0.206]	[0.126]
Network Portfolios (N)	522	309
Adj. R-Squared	0.122	0.0760

OLS estimation with robust standard errors. Star levels: * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$.
Data is restricted to shows from 2008 to 2016 on incumbent networks, since renewal and Nielson data missing for 2017. Includes year and network fixed effects. 1st Season Rating is the average episode IMDb rating for a show's initial series order. Renewed or Extended is an indicator for whether a show was either renewed for another season or extended past its original order. Neilson Top 30 Show indicates whether the show broke into the Nielson Top 30 highest viewed programs for the season. Renewal correlation is for all incumbent networks while Nielson correlation is just for the big four networks since they are the only networks with enough viewership to potentially enter the top 30.

Table 3. Estimators of Treatment of Piloting on IMDb Ratings

	OLS		Matching	
	(1)	(2)	(3)	(4)
	1st Season Rating	1st Season Rating	1st Season Rating	1st Season Rating
Piloted	-0.229**	-0.169*	-0.188	-0.232*
	[0.0950]	[0.0953]	[0.131]	[0.116]
Creator w/ Award		0.0630		0.287**
		[0.0917]		[0.129]
Constant	7.989***	7.658***	8.024***	7.667***
	[0.0769]	[0.161]	[0.140]	[0.226]
Year FE		X		X
Network FE		X		X
Bias Adj. Piloted		1.243		-0.340
Shows (N)	263	263	144	144
Deg. of Freedom	51	51	43	43
Adj. R-Squared	0.016	0.115	0.009	0.214

Robust standard errors in brackets. Star levels: * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$.
Clustered by network * year. Data is restricted to shows on incumbent networks from 2014 to 2017. Matching columns use an inverse propensity score based weighting over a region of common support to match similar straight to seriers and piloted shows. Bias adjusted piloted estimate assumes unobservables relate to piloted similarly to FE and Creator w/ Award observables as in Oster (2016).

Table 4. Logistic Prediction of Straight to Series Decision

	(1) Straight to Series	(2) Straight to Series
Funded independently of major studios	2.611***	3.137***
	[0.630]	[0.753]
Funded by broadcasting network	0.408	0.272
	[0.362]	[0.411]
Creator with Prior Show on Same Network	0.236	0.192
	[0.346]	[0.381]
Filmed outside Los Angeles	-0.0360	-0.609
	[0.627]	[0.723]
Genre: SciFi	0.985*	1.194**
	[0.520]	[0.557]
Genre: Comedy	-0.495	-0.867*
	[0.456]	[0.506]
Genre: Crime	0.905*	0.922*
	[0.482]	[0.513]
Genre: Supernatural	0.670	0.999*
	[0.513]	[0.549]
Genre: Thriller	0.649*	0.577
	[0.387]	[0.466]
Genre: Action	-0.941	-0.989
	[1.086]	[1.049]
Year FE		X
Network FE		X
Shows (N)	460	438
Pseudo R^2	0.158	0.243

Robust standard errors in brackets. Star levels: * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$.
Logistic estimation of straight to series used in IPW matching. Data is restricted to shows on incumbent networks from 2014 to 2017.

Table 5. Result of Matching on Observables for Incumbent Shows 2014-2017

All Observations

	Piloted		Straight to Series		
	N	Mean	N	Mean	Diff
Funded independently of major studios	426	0.09	77	0.38	0.29***
Funded by broadcasting network	426	0.62	77	0.53	-0.08
Creator with Prior Show on Network	426	0.73	77	0.65	-0.08
Filmed outside Los Angeles	426	0.52	77	0.71	0.19***
Genre: Science Fiction	421	0.04	75	0.13	0.09***
Genre: Comedy	421	0.47	75	0.20	-0.27***
Genre: Crime	421	0.05	75	0.12	0.07**
Genre: Supernatural	394	0.05	66	0.11	0.06*
Genre: Thriller	421	0.09	75	0.19	0.09**
Genre: Action	421	0.04	75	0.03	-0.02

Common Support

	Piloted		Straight to Series		
	N	Mean	N	Mean	Diff
Funded independently of major studios	202	0.12	38	0.21	0.09
Funded by broadcasting network	202	0.61	38	0.66	0.05
Creator with Prior Show on Network	202	0.74	38	0.68	-0.06
Filmed outside Los Angeles	202	0.74	38	0.74	0.00
Genre: Science Fiction	202	0.06	38	0.08	0.02
Genre: Comedy	202	0.28	38	0.18	-0.09
Genre: Crime	202	0.09	38	0.11	0.02
Genre: Supernatural	202	0.08	38	0.05	-0.03
Genre: Thriller	202	0.15	38	0.21	0.06
Genre: Action	202	0.04	38	0.03	-0.02

Star levels for t-tests: * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Table 6. Logistic Prediction of Piloting Decision

	(1) Pilot Ordered to Series	(2) Pilot Ordered to Series
Creator w/ Award	0.615***	
	[0.200]	
Funded independently of major studios	-0.334	-0.335
	[0.443]	[0.440]
Funded by broadcasting network	-0.158	-0.158
	[0.200]	[0.198]
Creator with Prior Show on Same Network	0.266	0.467**
	[0.223]	[0.213]
Filmed outside Los Angeles	0.139	0.122
	[0.237]	[0.238]
Genre: SciFi	1.113	1.136
	[0.802]	[0.820]
Genre: Comedy	-0.334	-0.382*
	[0.229]	[0.230]
Genre: Crime	0.448	0.400
	[0.335]	[0.339]
Genre: Thriller	0.343	0.620
	[0.804]	[0.845]
Genre: Action	0.753	0.580
	[0.608]	[0.582]
Constant	-0.612	-0.561
	[0.400]	[0.401]
Shows (N)	545	545
Pseudo R^2	0.0794	0.0668

Robust standard errors in brackets. Star levels: * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$.
Logistic predictor of piloting. Network and year fixed effects included. Data is restricted to pilots on incumbent networks from 2008 to 2013.

Table 7. Heterogeneity in Treatment Effect of Piloting by Creators w/ Awards

	(1) 1st Season Rating	(2) 1st Season Rating	(3) 1st Season Rating
Piloted	-0.231**	-0.232*	0.221
	[0.114]	[0.116]	[0.169]
Creator w/ Award		0.287**	0.596***
		[0.129]	[0.165]
Piloted * Creator w/ Award			-0.773***
			[0.268]
Constant	7.837***	7.667***	7.435***
	[0.197]	[0.226]	[0.269]
Shows (N)	144	144	144
Deg. of Freedom	43	43	43
Adj. R-Squared	0.188	0.214	0.279

Robust standard errors in brackets. Star levels: * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$. Includes year and network FE. Clustered by network * year. Data is restricted to shows on incumbent networks from 2014 to 2017. All estimates use an inverse propensity score based weighting over a region of common support to match similar straight to seriers and piloted shows as well as account for attrition in piloted shows not ordered to series.

Table 8. Effect of straight to series on network's portfolio of new shows

	(1) Average IMDB rating	(2) Average IMDB rating
Share Straight to Series	-0.147	-0.173
	[0.156]	[0.177]
Share StS * Post 2013	0.622	0.543
	[0.391]	[0.429]
Share StS * Big Four		-1.079
		[0.750]
Big Four * Post 2013		-0.318
		[0.251]
Share StS * Big Four * Post 2013		2.773**
		[1.016]
Constant	7.544***	7.559***
	[0.0831]	[0.0862]
Network Portfolios (N)	131	131
Deg. of Freedom	14	14
Adj. R-Squared	0.153	0.171

Robust standard errors in brackets. Star levels: * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$.
Observations are of each network's yearly porfolio of new shows. Includes year fixed effects. Fixed effect estimator used clustered at network level. Data is restricted to shows from 2008 to 2017 on incumbent networks.

11. Appendix

11.1 Mathematical proofs

Experimentation Decision Comparative Statics

Consider the solution to the decision problem:

$$\max_{\lambda \in \{0,1\}} \lambda(\Pr(y(\theta) > C)\, E[y(\theta) + (V_B - V_D) - C | y(\theta) > C] - e) + (1 - \lambda)(\theta - C)$$

Define the usefulness of a signal technology in terms of the highest e for which experimentation is worth doing:

$$u(y,\theta|C) \equiv \Pr(y(\theta) > C)\, E[y(\theta) + (V_B - V_D) - C|y(\theta) > C] - (\theta - C)$$

Order the signal technology such that for a given C and all θ, $y'|C > y|C \Leftrightarrow u(y',\theta|C) \geq u(y,\theta|C)\ \forall\, \theta$ with the inequality strict for some θ. Then using the monotone maximization theorem, experimentation ($\lambda = 1$) is increasing in y, V_B,$-V_D$, and $-e$:

- The space of possible choices, $\{0,1\}$, is trivially a lattice.
- The space of possible parameter values $y|\theta$, C, V_B, V_D and $e \in \mathcal{R} \times \mathcal{R}^+ \cup \{0\} \times \mathcal{R}^+ \cup \{0\} \times \mathcal{R}^+ \cup \{0\} \times \mathcal{R}^+ \cup \{0\}$ is a poset under the usual $\mathcal{R}^N$ order.
- The set of possible choices is trivially non-decreasing in parameter values, as it is always $\{0,1\}$.
- The objective function is trivially supermodular in the set of choice variables, given there is a single choice variable.
- The objective function displays increasing differences in λ, y, V_B, $-V_D$ and $-e$.

Likelihood Ratio Results

The first statement of Proposition 2, $E[\theta|\theta'' < \theta, \theta \in A] > E[\theta|\theta'' < \theta, \theta \in B]$ when $A >_{LR} B$, derives from the fact that $A >_{LR} B$ implies the distribution $A|\theta'' < \theta$ is greater than $B|\theta'' < \theta$ under the likelihood ratio ordering, see 1.C.6 in Shaked and Shanthikumar (2007). This in turn means $A|\theta'' < \theta$ first order dominates $B|\theta'' < \theta$, which implies $E[\theta|\theta'' < \theta, \theta \in A] > E[\theta|\theta'' < \theta, \theta \in B]$.

For the second statement, define

$$\psi(\theta) \equiv E[y(\theta) + (V_B - V_D)|y(\theta) > C]$$

Note that because of my assumptions on the signal technology y, $\psi(\theta)$ is an increasing function. This means if $A >_{LR} B$ then $E[y(\theta) + (V_B - V_D)|y(\theta) > C, \theta \in A]$ is greater than then $E[y(\theta) + (V_B - V_D)|y(\theta) > C, \theta \in B]$ under the likelihood ratio ordering, see 1.C.8 in Shaked and

Shanthikumar (2007). Applying 1.C.6 from Shaked and Shanthikumar (2007) once more yields the stated result of:

$$E[y(\theta) + (V_B - V_D)|y(\theta) > C, \theta' < \theta < \theta'', \theta \in A]$$

$$> E[y(\theta) + (V_B - V_D)|y(\theta) > C, \theta' < \theta < \theta'', \theta \in B]$$

Finally, the last result follows from a slight modification of the two above results. When observing outcomes in the region of prior shows $[a < \theta'' < b]$ and $A >_{LR} B$, 1.C.6 in Shaked and Shanthikumar (2007) also implies that

$$E[\theta|\theta'' < \theta < b, \theta \in A] > E[\theta|\theta'' < \theta < b, \theta \in B]$$

and

$$E[y(\theta) + (V_B - V_D)|y(\theta) > C, a < \theta < \theta'', \theta \in A]$$

$$> E[y(\theta) + (V_B - V_D)|y(\theta) > C, a < \theta < \theta'', \theta \in B]$$

Combining these two inequalities implies the outcomes from A will have a higher expected value than observations from B within prior beliefs $[a < \theta'' < b]$.

For Proposition 3, $A >_{LR} B$ implies $A >_{FOSD} B$, 1.C.6 in Shaked and Shanthikumar (2007). For any increasing function $\psi(\theta)$, this means $\psi(A) >_{FOSD} \psi(B)$ and $\Pr(\psi(\theta) > a) > \Pr(\psi(\theta) > a)$ for any a, see 1.A.3 and 1.A.1 in Shaked and Shanthikumar (2007). Since the assumed signaling technology is increasing, this means when $A >_{LR} B$ we should see the proabilty of passing the pilot phase to be higher for θ drawn from A than from B.

For Proposition 4, the difference in outcomes is $E[E[y(\theta)|y(\theta) > C] - \theta] + (V_B - V_D)$. The assumption defines $\psi(\theta) = E[y(\theta)|y(\theta) > C]$ and assumes $\psi(\theta) - \theta$ is a decreasing function. Therefore $E[E[y(\theta)|y(\theta) > C] - \theta]$ is decreasing in distribution of θ when the distributions of θ are ordered by the likelihood ratio ordering, see 1.C.8 in Shaked and Shanthikumar (2007).

Portfolio of Outcomes

Let project i be a project moved from experimentation to non-experimented production. Suppose afterwards, there are m projects still developed through experimentation. Let A be the average value of these project. For the non-experimented projects, let there be n projects other than i, with an average value B. The portfolio of shows then worsens if

$$\frac{1}{m+n+1}[mA+nB+E[y(\theta_i)+(V_B-V_D)|y(\theta_i)>C]]>\frac{1}{m+n+1}[mA+nB+\theta_i]$$

This simplifies to $E[y(\theta_i)|y(\theta_i)>C]+(V_B-V_D)>\theta_i$. Given my signaling technology, this will be true unless V_B-V_D is sufficiently negative. Moving a project from experimentation to non-experimentation production can only improve outcomes when the drawbacks from experimentation outweigh the benefits from experimentation: when commitment has value.

In the case where the experimented project did not pass the experimentation phase, outcomes would worsen whenever:

$$\frac{1}{m+n+1}[mA+nB+\theta_i|\theta_i<C]<\frac{1}{m+n}[mA+nB]$$

Since a project would only be in set A or B if the expected outcome was greater than cost C, we know:

$$\frac{1}{m+n+1}[mA+nB+C]\leq\frac{1}{m+n}[mA+nB]$$

Since the specific project θ_i did not make the cut into experimentations as post experiment the posterior belief was less than C, we can complete the proof:

$$\frac{1}{m+n+1}[mA+nB+\theta_i|\theta_i<C]<\frac{1}{m+n+1}[mA+nB+C]\leq\frac{1}{m+n}[mA+nB]$$

Bargaining

First, I argue the entry of a network has no bargaining power effect on the decision to experiment under standard cooperative bargaining models. I model bargaining through Shapely value, which is appropriate for bi-lateral relationships (Fontenay and Gans, 2014). Based on interviews with show creators, writers approach the networks serially with an idea. Writers rank networks based on fit with their vision and support for their show conditional on the other shows in the network's portfolio. The number of networks in their consideration set is usually small; perhaps two or three networks are a strong fit for a particular show in a particular season. Writers then approach each network serially. When negotiations fail with a network, writers need to tweak their idea for the next network; an aggressive cop drama originally pitched to a cable network like HBO would be moderated for a broadcast network like ABC. Bi-lateral bargaining between agents in a graph is a reasonable model for this type of linear negotiation process.

The value allocated to each player i in a collation game is a weighted sum of

$$v(S \cup \{i\}) - v(S)$$

where S is a set of players not including i and $v(S)$ is the highest total sum of payoffs that can be generated by S through cooperation. Because there is only one bilateral contract at a time that can exist between the writer and a network, we can write $v(S \cup \{i\})$ as

$$\max[v(S), \max_{\lambda^i \in \{1,0\}} (\lambda^i \hat{V}_p^i + (1 - \lambda^n)\hat{V}_s^i)]$$

Network i selects whether or not to pilot, λ^i. The expected payoff from piloting for network i is $\hat{V}_p^i$ and the payoff from deciding to skip the pilot is $\hat{V}_s^i$.

A subgame perfect equilibirum is reached if all networks always pick the higher of $\hat{V}_p^i$ or $\hat{V}_s^i$. When one or both of $\hat{V}_p^i$ and $\hat{V}_s^i$ is above $v(S)$, picking the larger one results in the highest payoff. If both are below $v(S)$, network i received no value from the term so picking the larger is still weakly preferred. So regardless of the number of other networks or the payoffs available to those networks, firm i will only consider whether $\hat{V}_p^i$ or $\hat{V}_s^i$ is larger. If the network is picking piloting over skipping the pilot prior to entry, it suggests the network views the piloting payoff as higher. Furthermore, the network should continue picking the piloting after entry since it remains optimal to do so if the coalition value is not affected by entry.

Next, what about if bargaining in television deviates from the standard assumptions, by for example exhibiting uncompensated differentials? Let V_E be expected joint payoff from experimentation, inclusive of costs of production and V_C be the joint payoff from commitment. Let $V_E > V_C$, so that experimentation results in a larger overall payoff. The payoff is split with $1-\alpha$ going to the network and α going to the show creator. To model uncompensated differentials, let κ represent a private cost to experimentation born by the show creator that is not included in V_E. Before Netflix's entry, all the networks prefer to experiment since $V_E > V_C$. Each creator's outside option is the payoff from experimentation on another network, so a subgame perfect equilibirum is for the networks to always experiment by funding pilots.

Once Netflix enters and only commits, the decision calculation of the incumbent networks changes to reflect that the creator's outside option may include a payoff V_C that crucially does not cause them to incur private cost κ. Now when experimenting, the incumbent networks need to provide a larger share of payoff V_E to the creator when experimenting, represented by Δ.

$$\max_{\lambda\in\{0,1\},\Delta\in[0,1-\alpha]} \lambda(1-\alpha-\Delta)V_E + (1-\lambda)(1-\alpha-\Delta)V_C$$

$$s.t.\ \lambda\big((\alpha+\Delta)V_E-\kappa\big) + (1-\lambda)(\alpha+\Delta)V_C \geq \alpha V_C$$

When the network chooses to commit, Netflix as an outside option is not superior from the perspective of the creator so Δ can be 0. However, under experimentation, Δ may need to be positive so that its rational for the creator to prefer the network over Netflix. In the case when the optimal $\Delta^* \in [0,1-\alpha]$, this means the private cost κ gets factored into the decision making of the network, so the optimization problem becomes:

$$\max_{\lambda\in\{0,1\}} \lambda(V_E-\kappa) + (1-\lambda)V_C$$

This does not solve the identification issue since variation in bargaining power is not affected the network's decision. However, when α is high its possible that $\Delta^* > 1 - \alpha$, making it not possible for the network to sufficiently compensate a creator to experiment: the network would be forced to commit instead. This would lead to shows with the same expected payoff to have different production processes due to the level of creator bargaining alone.

11.2 Robustness Checks

Table 8 provides an alternate version of Table 3's matching estimators that also account for the attrition basis included in Equations 3 and 7. Column 2 of Table 4 rather than Column 1 is used for probability weights since the *Creator w/ Award* variable is used in the main specification. The empirical object measured is no longer the difference between experimenting and not experimenting on final outcomes, its more the pure benefit of experimenting. Neither Columns 3 or 4's point estimates change when this bias is accounted for. Inspecting Figure 5 points to why: the common support used in Table 3's matching estimator focuses on piloted shows that were unlikely to ordered straight to series which seem to correlate highly with piloted shows that were unlikely to pass the pilot phase. In other words, the set of observations used to calculate the matching average treatment effect are similar to the kinds of shows that would normally be subject to attrition bias.

Table 8. Matching Estimates Adjusted for Attrition Bias in Piloted Shows

	OLS		Matching	
	(1)	(2)	(3)	(4)
	1st Season Rating	1st Season Rating	1st Season Rating	1st Season Rating
Piloted	-0.173*	0.0506	-0.242*	0.189
	[0.0959]	[0.130]	[0.125]	[0.174]
Creator w/ Award		0.372**		0.607***
		[0.170]		[0.175]
Piloted * Creator w/ Award		-0.413**		-0.752***
		[0.205]		[0.278]
Constant	7.685***	7.492***	7.997***	7.605***
	[0.156]	[0.169]	[0.182]	[0.247]
Shows (N)	263	263	144	144
Deg. of Freedom	51	51	43	43
Adj. R-Squared	0.117	0.127	0.158	0.222

Robust standard errors in brackets. Star levels: * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$.
Include show length, year and network FE. Clustered by network * year. Data is restricted to shows on incumbent networks from 2014 to 2017. Matching uses an inverse propensity score based weighting over a region of common support. Matching columns use an inverse propensity score based weighting over a region of common support to match similar straight to seriers and piloted shows as well as account for attrition in piloted shows not ordered to series.