

Moneyball for TV: A Model for Forecasting the Audience of New Dramatic Television Series

Starling D. Hunter III¹, Ravi Chinta², Susan Smith³, Awais Shamim¹, Alya Bawazir⁴

¹Tepper School of Business, Carnegie Mellon University Qatar, Doha, Qatar

²College of Business, Auburn University at Montgomery, Montgomery Alabama, USA

³Department of Mass Communication, American University of Sharjah, Sharjah, United Arab Emirates

⁴Vodafone Qatar, Doha, Qatar

Correspondence: Starling D. Hunter III, Tepper School of Business, Carnegie Mellon University Qatar, Doha, Qatar.

Received: May 17, 2016 Accepted: June 15, 2016 Online Published: June 29, 2016

doi:10.11114/smc.v4i2.1611

URL: <http://dx.doi.org/10.11114/smc.v4i2.1611>

Abstract

The specific objective of the present study is to develop and test an early-stage, empirical model for predicting the audience of new television series. We test our model on a sample of 107 new dramatic television series that debuted on one of the four major US television networks during the 2010-2014 seasons. In particular we examine the role of three previously untested predictors of the performance of new television shows, all of which can be known prior to the decision to greenlight the pilot script. Those three are the *originality* of the concept of the show, the *track record* of success of the show's creative team, and the *size of the conceptual network* created from the teleplay of the pilot episode.

Keywords: television, network analysis, pilot episodes, ratings, audience, dramatic series, content analysis, forecasting

“They talk in terms of ‘swings,’ ‘home runs’ and ‘batting averages,’ but there is no *Moneyball* of television, yet.”

—David Hornsby, Producer of *It's Always Sunny in Philadelphia*

1. Introduction

In the spring of 2007 an article appeared in the *New York Daily News* entitled “Writers Can't Predict Success of Network Pilots” (Littlejohn, 2007). The article included interviews with two highly experienced and well-placed TV industry sources. The first was screenwriter and producer Greg Berlanti who had already experienced success as the creator and lead writer for the adult-themed *Everwood* (2002) and the documentary-style political drama *Jack & Bobby* (2004). The article related that Berlanti was waiting to learn from the *American Broadcasting Company* (ABC) whether two new pilots that he had produced—one entitled *Eli Stone* (2008) and the other *Dirty Sexy Money* (2007)—would be included among the approximately 40 new series that would make it onto the major network's fall 2007 schedule. Both did and went on to have modestly successful first season ratings—averaging 8.3 and 7.7 million viewers per week, respectively. However, both series were cancelled after just two seasons. And while the cancellations were undoubtedly disappointing, they couldn't have come as a complete surprise. It has, after all, been well-documented that cancellation rates for new TV shows routinely exceed 60% in their first year, with most of the surviving shows cancelled after the second season (Nathanson, 2013; Stelter, 2012; Konda, 2014; Ocasio, 2012; Schwab, 2013).

The article's second informant was Katherine Pope, then head of series development at the *National Broadcasting Company* (NBC). She described one of several strategies intended help to improve the odds of success for new series. Hers was to take an “unusual approach” to familiar subject matter rather than coming up with an “original idea.” She would then choose producers and creators with “unique viewpoints” and “real vision.” Berlanti's strategy differed: his was to focus on making the pilot episode as compelling as possible. The underlying logic took into account that pilot or premiere episodes typically have a new series' largest viewership of the season. Drop-offs in ratings for less-than-compelling pilots can be as steep as 30-50% over the next few episodes. Numbers like these matter because television studios earn the bulk of their revenues by selling blocks of time to advertisers (Meyer & Hyndman, 2006). Seen in that light, highly compelling pilots are not an end in and of themselves. Rather, they are a means to end, that end being two-fold—attracting a large number of viewers and then holding on to as many of them as possible over the

ensuing weeks. But while such a strategy is eminently sensible, there is a fly in the ointment: the prices that the studios charge for advertising on new television shows are based on audience projections that must be made months before the first episode has even aired. These projections serve as the basis upon which advertisers or ad agencies purchase upwards of 75-80% of network advertising inventory for those new shows (Napoli, 2001).

Unfortunately for both sides, the predictive models of the ratings performance of new television shows have long been recognized as “notoriously inaccurate” (Rust & Eechambadi, 1989). And the financial consequences for such inaccuracy have never been higher. Consider, for example, the 2012-13 season, one in which the four major studios and cable networks were reported to have committed over \$700 million on 186 proof-of-concept pilots, an average of \$3.8 million each (Nathanson, 2013). Given that well under a third were ever aired, and then three quarters of those were off the air within two seasons, it’s understandable why some have characterized the present system as being “wasteful of time, talent, and money” (Stelter, 2012). Others have suggested that these hundreds of millions of dollars are spent as “hedgies” because of the studios’ persistent inability to garner any “predictive edge” from their collective “creative experience, development process, and top talent” (Nathanson, 2013). In an interview with the *New York Times* (Stelter, 2012), TV producer David Hornsby resorted to sports metaphors to bemoan the state of play:

Network executives test the (proof-of-concept) pilots with focus groups before they choose which ones to order. But it is a largely unscientific process...one that has far more to do with art than science. They talk in terms of “swings,” “home runs” and “batting averages,” but there is no “Moneyball” of television, yet.

The specific objective of the present study is to develop and test just such a “Moneyball” model for predicting the success of new television series. Consistent with the broad outlines of the Moneyball strategy, our model focuses on under-valued and under-examined determinants of ratings at the expense of over-valued and over-examined ones. As we argue later, the former are disproportionately and necessarily located in the earlier stages of the value chain—the pre-production or development stages. We test our model on a sample of 107 new dramatic television series that debuted on one of the four major US television networks—ABC, NBC, CBS, and Fox—during the 2010 through 2014 seasons. In particular we examine the role of three empirically-untested predictors of the performance of new television shows, each of which can be known very early on. Those three are the *originality* of the concept of the show, the *track record* of success of the show’s creative team, and the *size of the conceptual network* created from the teleplay of the pilot episode. As predicted, we find that each of these variables has a positive and statistically significant impact on ratings performance as measured by audience size.

The remainder of this paper is organized as follows. The next section contains our literature review and hypotheses. In section three we outline the methods and data that we subsequently analyze. In the fourth section we describe the results of our statistical analyses while in the final section we elaborate upon the implications of and limitations to our findings and offer a few suggestions on how the model might be profitably applied.

2. Literature Review & Hypotheses

2.1 Concept Originality

In our review of over 60 studies of predictors of television show ratings and audience size, we were unable to identify a single one in the last forty years that examined concept originality directly or, for that matter, any proxy for it. Research in the film industry, however, has quite closely examined a related question—that of the box office performance of sequels. The consensus is that sequels are associated with higher box office performance than non-sequels (Basuroy & Chatterjee, 2008). But sequels are only one of the many ways in which both television shows and movies borrow from pre-existing source material. Among those other ways are remakes & re-boots, spin-outs and prequels, and adaptations of novels, fairy tales, comic books, and short stories. Only one study of which we are aware in the field of media or cultural economics has considered this broader range of adaptations. Specifically, in an empirical study of 170 films released in 2010 and 2011, Hunter, Smith, & Singh (2016) found that adaptations of all kinds had significantly higher opening weekend box office. Thus, while the prior literature on TV ratings leaves us with little upon which to firmly ground an expectation, empirical research in a closely related field leads us to hypothesize that *all else equal, new television series adapted from prior source material will be associated with higher ratings performance.*

2.2 Track Record of Creative Team

The experience or track record of the creative team of a new series is another factor that can be known at the earliest stages in the production value chain. But again, our review of the literature on television ratings returned not one study that has examined this factor either directly or indirectly. But as with adaptation, the literature on the determinants of box office performance has done so. Specifically, the importance of creative talent has been studied exhaustively, having been included as an independent or control variable in no fewer than two dozen studies. While most of them were concerned with the impact of star actors on box office of US-released films, a few have examined the impact of

star directors on the same (Nelson & Glotfelty, 2012). Typical measures of star quality were star-meter rankings (ibid), participation in a top-ten grossing film in prior years (Basuroy, Chatterjee, & Ravid, 2003), the number of Academy Awards wins or nominations garnered in some period prior to the current film project (Litman & Kohl, 1989), or inclusion on widely-respected lists of “the 100 most powerful people in Hollywood” or of “A and A+” actors (Walls, 2005). While the above studies all found a positive and significant relationship between measures of director “star power” and box office, the effects were much less significant than that associated with the star actors and actresses.

Another recent and relevant study is by Hunter, Smith, & Singh (2016) who reported a positive and significant impact of the prior success of screenwriters on the box office receipts of their current projects. In particular they found that the higher the opening weekend box office of the screenwriter’s most recent film, the higher was the opening box office of the subsequent one. Again, while there is no prior research in television studies that directly examines the relationship between any aspect of the creative team and program ratings, all of the above leads us to conclude that a positive relationship should exist between the creative team’s track record and ratings performance of a new series. Specifically, our second hypothesis is that *the better the track record of a new series’ creative team, the greater will be the series’ ratings performance.*

2.3 Textual Properties

A third set of predictors that can be known at the earliest stages of the value chain concern objective, text-based measures of the teleplay. As with the previous two variables, we were unable to identify any prior research studies that have attempted to link any text-based or semantic properties of teleplays to television ratings. But as with the above hypotheses, we did find some guidance in the box office literature, two recent studies in particular. The first is by Eliashberg, Hui, & Zhang (2014) who linked several textual, content, and genre properties of screenplays to box office performance. Among the text-based variables they examined were two “bag-of-words” measures which captured the styles and frequencies of individual words in the text of the 300 shooting scripts. The latter measure of the pair—named LS2—was related the various styles of language in the dialogs with higher values indicating a more prevalent use of vulgarity. Higher values of LS2 were also associated with lower levels of box office. The second relevant study from that literature is by Hunter, Smith, & Singh (2016) wherein the box office performance of 170 films was shown to be strongly explained by a network-of-words measure derived from a text analysis of the films’ screenplays. Specifically, they found that the size of the text network created from selected words in each film’s screenplay was positively and significantly associated with opening weekend box office, even when controlling for several other covariates. Moreover, the strength of the effect of the network-of-words measures in that study was the strongest among all covariates.

Again, despite there not being any direct tests of any properties of teleplay text on TV show performance, our reading of the small but relevant companion literature leads us to conclude that the same relationship should exist between text network size of teleplays and their ratings performance. But because of the widely-acknowledged primacy of pilot episodes in the development process (Nathanson, 2013), because a pilot’s job is “to set the tone for the series” (Lindauer, 2011) and “to establish the characters and situations” that will recur week after week (Anders, 2012), and because the initial ratings performance of the pilot strongly influences whether the series will get a full-season order (MacNabb, 2015; Kissell, 2015), our final hypothesis is that *all else equal, the size of the text network of the teleplay of a new series’ pilot episode will be positively associated with series’ ratings performance.*

3. Methods & Data

In order to investigate the aforementioned hypotheses, data were collected on the total number of viewers of new prime-time, hour-long television series debuting during five recently-completed broadcast seasons. Following relevant prior research, only shows debuting on the Big Four US television networks—ABC, CBS, NBC, and FOX—were included (Napoli, 2001). We used several sources to determine which shows appeared during those seasons. These include *TV Series Finale*, *TV.com*, *TV Guide*, and *Wikipedia*, particularly the latter’s series of “US Network Television Schedule” pages for each of the five seasons.

We identified a total of 115 new hour-long dramatic series that debuted between the September 1st, 2010 and August 30th 2015. Four of these shows were eliminated from consideration because they were co-produced with foreign television networks and debuted in those countries before being seen on US network television—*Welcome to Sweden* (2014), *Crossing Lines* (2013), *Taxi Brooklyn* (2014) and *The Firm* (2012).

We also eliminated four “back-door” pilots—episodes of a long-running show that serve to introduce one or more guest characters for what will become a new television show—*NCIS-New Orleans* (2014), *CSI Cyber* (2015), *Criminal Minds: Suspect Behavior* (2011), and *The Finder* (2012). Of the remaining 107 new series that had traditional pilot episodes, we were unable to locate copies of teleplays for only six of them—*Agents of SHIELD* (2013), *Once upon a Time in Wonderland* (2013), *Chicago PD* (2014), *Agent Carter* (2015), *Made in Jersey* (2012), and *Alcatraz* (2012). Of the remaining 101 new series, a further three had to be eliminated from consideration because the teleplays were not

machine-readable and thus could not be analyzed using text analysis software—*Smash* (2012), *The Undercovers* (2010), and *My Generation* (2010).

3.1 Dependent Variable

Our measure of the ratings performance was total viewership, by episode, measured in millions of viewers. We obtained the data from a number of sources including TV Series Finale, Tv.com, and the Wikipedia pages for each show, particularly the “Episodes” sub-sections which provide summary descriptions of the each episode along with the viewership numbers. Because of the highly-skewed distribution of viewership, we log-transformed those quantities and assigned the resulting variable the name LOGVIEW. As shown in Table 1, below, the average number of total viewers was 6.50 million with a standard deviation of 3.04 million. As would be expected, the total viewership of the pilot episodes was initially much higher than the average—8.39 vs. 6.50 million viewers—but then declined steadily thereafter falling to just 5.85 million by the 10th episode.

Table 1. Descriptive Statistics

	N	Mean	St. Dev.	Min	Max
Viewers (Millions)	1441	6.50	3.04	1.03	19.34
Episode 1	107	8.39	3.22	2.64	16.49
Episode 2	107	6.96	2.82	2.21	13.39
Episode 5	102	6.06	2.82	1.63	12.42
Episode 10	85	5.85	2.99	1.16	14.92
Episode 15	32	7.85	3.40	3.74	19.34
Episode 20	23	7.50	2.52	3.85	12.73
LOGVIEW	1441	6.76	0.22	6.01	7.29
ADAPTATION	107	0.38	0.49	0	1
RENEWAL	107	0.44	0.70	0	2
LOGLINK	98	1.65	0.30	1.04	2.32

Table 2. Correlation Matrix

	LOGVIEW	ADAPTATION	RENEWAL
ADAPTATION	-0.212*		
RENEWAL	0.141	0.121	
LOGLINK	0.404****	0.025	0.071

Legend: **** = $p < 0.0001$, *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$, # = $p < 0.10$, all 1-tailed

3.2 Independent Variables

3.2.1 Adaptations

We consulted the main Wikipedia page for each show in order to determine whether or not it was original or adapted from prior source material. This information was almost always provided in the first paragraph. The resulting variable that we created was named ADAPTATION which was coded “1” if the show was adapted from prior source material and “0” otherwise. Among the 40 new series in the sample that we treated as adaptations were those based on fairy tales, e.g. *Once Upon a Time in Wonderland* (2013) and comic books, e.g. *Constantine* (2014); remakes of foreign series, e.g. *Backstrom* (2015) and domestic ones— e.g. *Ironside* (2013); *spin-outs* of current new hit shows—e.g. *Chicago PD* (2014); *extensions* of long running franchises, e.g. *Law & Order: LA* (2010); and several others based on novels or short stories — e.g. *Hannibal* (2013) and *Under the Dome* (2013).

3.2.2 Track Record

Our measure of prior success focuses on the show creator(s). Specifically, we used IMDB to first determine the number of series for which the creator(s) had earned a writing credit for the teleplay of the pilot episode. From among those credits, we determined how many of the shows had been renewed, i.e. that aired for at least two seasons. We created a Likert-scaled variable (RENEWAL) where creative teams with no prior successes were assigned a score of zero ($n = 71$), those with one prior success were assigned a score of one ($n = 21$), and those with two or more prior success were assigned a score of two ($n = 12$). Notably, RENEWAL showed no significant correlation with any other variable listed in Table 2.

3.2.3 Text Network Size

Several distinct approaches exist for creating networks from texts. They differ along a number of dimensions including the level of automation, whether and how words are abstracted to higher-order categories, and the nature of the underlying relationship used to connect the words. In this study we opted for an approach that is semi-automated, that abstracts words into higher-order conceptual categories defined by common etymological root, that connects conceptual categories according to their co-occurrence in multi-morphemic compounds (MMCs), and that has already been applied to the study of screenplays (Hunter, Smith & Singh, 2016). As defined in that approach, MMCs may include,

but are not necessarily limited to, *open compounds* (middle class, attorney general), *closed compounds* (parkway, gunshot), *abbreviations* and *acronyms* (WASP, HQ, SUV), *anacronyms* (laser, scuba), *blend words* (brunch, guesstimate), *hyphenated multiword expressions* (state-of-the-art, glow-in-the-dark), *infixes* (un-bloody-believable, fan-blooming-tastic), *appositional compounds* (attorney-client, actor/model), *hyphenated compounds* (rapid-fire, wide-eyed), selected *clipped words* (internet, wi-fi), and *pseudo-compound words* (misunderstanding, overrated).

The first step in the creation of the text networks involved identifying the MMCs in each screenplay. To accomplish this we first used the *Generate Concept List* and the *Identify Possible Acronyms* commands in the CASOS Institute’s *Automap* software program to generate word lists for each screenplay (Diesner & Carley, 2004). To determine which words were MMCs, we compared each word list with the contents of a proprietary database which contains over 30,000 unique MMCs extracted from over 500 contemporary screenplays and teleplays (Hunter, 2014).

The second step involved decomposing every MMC in each screenplay into its constituent words. For example, the closed compound *gunshot* is comprised of two words—*gun* and *shot*. Next, each constituent word was assigned to a conceptual category defined by its most remote etymological root. Typically, the most remote root was Indo-European, as defined in the 3rd edition of the *American Heritage Dictionary of Indo-European Roots* (AHDIER). That source assigns over 13,000 English words to over 1,300 Indo-European (IE) roots. For example, the word *gun* descends from the IE root *gwhen-* which means “to strike, kill” while the word *shot* descends from the IE root *skeud-*, which means “to shoot, chase, throw.” Because all of the word-to-root mappings described in the AHDIER are also included in the aforementioned database, it was possible to automatically assign almost 85% of the constituent words in each screenplay to their appropriate etymological roots. The remaining 15% were instances where IE roots of constituent words either could not be identified or did not exist. In the latter case, etymological roots provided in the *American Heritage Dictionary of the English Language* were used. Most typically these were Latin, Greek, Germanic, or Old English.

The final stage was to calculate the size of the resulting network of concepts with the use of the UCINET software program (Borgatti, Everett, & Freeman, 2002). We focused on the largest cluster of mutually-reachable nodes in a network which is referred to as the “main component.” Our specific measure of the size of the text network was the number of links contained in the main component of the network. Figure 1, below, depicts a portion of the main component of the text network constructed from the teleplay of the pilot episode of CBS’s hit dramatic series *Person of Interest* (2011).

As we can see, this portion of the network is comprised of three components—one main and two minor. The main component has 38 nodes while the minor components have nine and three. As noted above, the nodes in the network are etymological roots while MMCs are associated with the links between pairs of nodes. Those MMCs took several of the aforementioned forms including *closed compounds* (hardware, gunshot, paperwork), *acronyms* (GPS, SUV, and IP), *hyphenated compounds* (upside-down, rapid-fire, heavy-set) and a *clipped word* (internet).

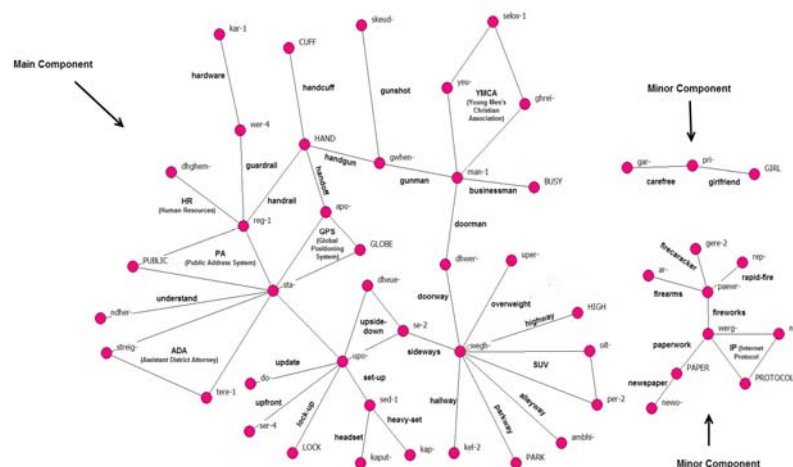


Figure 1. Portion of Text Network for *Person of Interest*

4. Results

Table 3 contains the results of several random effects, generalized least squares (GLS) regression models used to test our three hypotheses. The dependent variable in each model is LOGVIEW and the three independent variables were whether or not the series’ concept is adapted or original (ADAPTATION), the number of prior series by the creative

team that were renewed for second seasons (RENEWAL), and the log of the number of links in the text network created from the teleplay of the pilot episode of the series (LOGLINK). The regression models test the effects of two or more of the three variables under different conditions, specifically the number of episodes. Note that Model 1 and Model 2 have “All Episodes” as their headings. The underlying models differ, however. Model 1 examines only the impact of ADAPTATION and RENEWAL on the audience size (LOGVIEW). As we can see, the model is statistically significant (Wald- $\chi^2 = 9.56, p < 0.01$) but explains only 5.3% of the variance. But interestingly, the results support the second but not the first hypothesis. Recall that our first hypothesis was that ADAPTATION would be positively associated with audience size. The results show precisely the opposite of our expectation. To wit, the beta coefficient is significant and positive ($\beta = 0.087, p < 0.01$) where we expected it to be negative. Thus, in our sample of 107 new dramatic television series, those with original concepts performed significantly better than those that were based on other source material. The same pattern of results is seen to hold in Models 2-5, as well.

Our second hypothesis was that the creative team’s track record matters in a positive sense for audience size. Support for this hypothesis would see positive and statistically significant beta coefficients for RENEWAL in Models 1-5. What we observe is somewhat supportive: the coefficients are positive and modestly significant ($0.05 < p < 0.10$) in Models 1-4 and only positive but not statistically significant in Model 5.

Our third hypothesis was that there should exist a positive and significant relationship between the size of the text network of the teleplay of the pilot episode and the total audience of the series. As we can see in Models 2-5, the variable LOGLINK is both positive and highly statistically significant— $p < 0.01$ in Models 4 and 5, $p < 0.001$ in Model 2, and $p < 0.0001$ in Model 3. This finding suggests the size of the text network of the pilot episodes’ teleplay—something that can be known in the earliest stages of the production value chain—strongly predicts the total audience of the entire first season of the new series.

That LOGLINK is the most significant predictor among three variables is something that was not anticipated. That LOGLINK is less significant in Model 2 (All Episodes, = 0.221, $p < 0.001$), Model 4 (Episodes 6-10, = 0.206, $p < 0.01$), and Model 5 (Episodes 11-15, = 0.275, $p < 0.01$) tells us that LOGLINK is much better at predicting the total audience of the first five episodes than that of the later ones. Finally, it is worth noting that Model 3, which concerns the first five episodes only, explains the greatest proportion of the variance ($R^2 = 24.3\%$).

Table 3. Results of Random Effects Generalized Least Squares (GLS) Regression

	Model 1: All Episodes	Model 2: All Episodes	Model 3: Episodes 1-5	Model 4: Episodes 6-10	Model 5: Episodes 11-15
ADAPTATION	-0.097** (2.57)	-0.102** (2.62)	-0.108** (3.04)	-0.109* (2.39)	-0.128* (2.00)
RENEWAL	0.049* (1.89)	0.042# (1.50)	0.044# (1.70)	0.050# (1.60)	0.025 (0.60)
LOGLINK		0.221*** (3.52)	0.238**** (4.16)	0.206** (2.87)	0.275** (2.86)
Sample size (n)	1441	1334	481	430	263
Number of Teleplays	107	98	98	91	71
Wald- χ^2	9.6**	22.0****	30.1****	17.1***	12.5**
R ² (between)	8.4%	19.0%	24.3%	16.5%	15.8%
R ² (overall)	5.3%	15.5%	20.9%	16.2%	14.8%
ΔR^2 -overall vs. baseline	--	12.2%	15.6%	10.9%	9.5%

Legend: **** = $p < 0.0001$, *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$, # = $p < 0.10$, all 1-tailed

5. Discussion & Conclusion

We began this study by taking note of the fact that neither conventional wisdom nor empirical research has proven capable of identifying the determinants of success of new television shows. Upon hearing this, one could be forgiven for recalling Academy Award winning screenwriter William Goldman’s infamous quip from the 1980’s concerning the then conventional wisdom concerning the prediction of box office—“nobody knows anything” (Goldman, 1983). What Caves claims this remark meant was that “while “producers and executives know a great deal about what has succeeded commercially in the past and constantly seek to extrapolate that knowledge to new projects...their ability to predict at an early stage the commercial success of a new film project is almost nonexistent” (Caves, 2003). Substitute “new film project” with “new television project” and much the same still holds. But that it does hold, doesn’t mean that it must. In fact, the express purpose of this study was to show that the performance of new television shows could, in fact, be predicted in advance. Toward that end, we developed and tested three hypotheses concerning the effect of the originality of a show’s concept, the track record of its creative team, and text-based properties of the pilot episode’s teleplay.

Recall that in the first of these hypotheses we predicted that adapted concepts would be positively associated with total audience. To our surprise, we not only failed to find support for this hypothesis, we actually found support for the opposite condition. Specifically, we found that adapted concepts were associated with significantly lower total audiences. This result stands in direct opposition to the conventional wisdom concerning the efficacy of adapted concepts over originals. To make sure that different risk-return tradeoffs between them didn't explain the result, we even conducted a post hoc test of the variance around audience size associated with each alternative. Neither straight tests of variance nor a more sophisticated method—multiplicative heteroscedastic regression—revealed any difference between original and adapted premises. Specifically, in our sample of new dramatic television series airing during the 2010 through 2014 seasons, the first few episodes of shows with adapted concepts had about 22% smaller average audiences and no difference in the variances around the averages.

Our second hypothesis concerned the role of the creative team's prior successes on the total audience of new television shows. As predicted, we found that the relationship was positive and significant. While ours is the first study of which we are aware to publish such a finding, it's important to note that there may be better ways to operationalize the creative team's track record of success or reputation. For example, instead of counting the number of prior shows that were renewed, we could have, for example, constructed a measure based on the average total audience of their last several or just their most recent series. Future research should attempt to determine whether measures such as these more accurately capture track record and/or reputation effects.

Our third and final hypothesis posited that the size of the largest component of the text network was a positive and significant predictor of total audience. This is what we found and the finding was extremely robust, holding as it did across all conditions that we tested—the entire season and various 5-episode groupings. During the first two episodes, the total audience of the 25 series with network sizes in the top quartile was 62% larger than series whose networks were in the bottom 75%. Specifically that was 10.0 million viewers versus 6.18 million. One thing we did not expect to find was that this variable explained about 3-5 times as much variance in audience as did the other two variables combined. Despite this highly significant result, it's worth noting that the size of a text network indicates nothing about the teleplay's content. Rather, it appears to be a measure of the teleplay's cognitive complexity (Hunter, Smith, & Singh, 2016). Future research might consider whether network measures uncorrelated with size, e.g. the position within the network of genre-specific words and concepts, can also predict total audience.

6. Application

A final word about the potential contribution of our approach to practice is in order. Recall that the decision process by which development executives greenlight pilots has been characterized as intuitive and ineffective, the latter being operative under the assumption that keeping failure rates and developments costs low is an explicit goal. Taylor among others, has suggested that research that can systematically de-bias the process should be brought to bear as early in the value chain as possible (Taylor, 2013). To that end, we offer the following suggestions on how the three variables tested in this study could be profitably employed.

The first step begins with the recognition that when the three variables are dichotomized there are eight possible combinations thereof. At the opposite ends of the spectrum, so to speak, are a null condition—an *adaptation* from a creative team with *little or no track record of success* and a teleplay with a *small text network*—and a positive condition—*original concept, two or more prior successes*, and a *large text network*. Over the past five television seasons, twenty six of the 98 shows were in the null condition. Of these, twenty-two were not renewed for a second season—a failure rate of 85%. In alphabetical order the failed shows were *666 Park Avenue* (2012), *Allegiance* (2015), *American Odyssey* (2015), *Astronaut Wives Club* (2015), *Backstrom* (2015), *Betrayal* (2013), *Constantine* (2014), *Cross Bones* (2014), *Do No Harm* (2013), *Dracula* (2013), *Gracepoint* (2014), *Hostages* (2013), *Ironside* (2013), *Killer Women* (2014), *Law & Order: LA* (2010), *Lucky 7* (2013), *Rake* (2014), *Red Band Society* (2014), *Red Widow* (2013), *Terra Nova* (2011), *Wayward Pines* (2015), and *The Whispers* (2015). Some failed more quickly than others. For example, there were some that received cancellation notices even before the season was completed, e.g. *Red Band Society* (2015) and *Rake* (2010), others that ended with episodes still unaired, e.g. *Ironside* (2013) and *Killer Women* (2014), and one that can lay the claim to being the lowest rated fall drama premiere in its network's history—*Lucky 7* (2013).

Of the four that survived were *Mistresses* (2013), *Resurrection* (2014), *Sleepy Hollow* (2013) and *Secrets & Lies* (2015). Not one could be even remotely classified as a hit. The former, now in fourth season, has 13-episode seasons, all of which run in the late spring and summer on Monday nights with the total broadcast audience always averaging under 4 million viewers per episode. *Resurrection* was renewed after a modestly successful first season of only 8 episodes but was cancelled after the end of its second season after a total of only 21 episodes—fewer episodes than a hit show gets in one season. The final episode had a total audience of 3.73 million viewers, a 73% drop from its pilot audience. *Sleepy*

Hollow, now in its fourth season, is in terminal ratings decline. While its first season average was a respectable 7.46 million viewers over 13 episodes, the second season saw that number decline 39% to 4.57 million. The third season saw an additional 29% decline to just 3.25 million viewers. The fate of *Secrets & Lies*, whose second season is yet to begin airing, remains to be determined. Its first season of 10 episodes averaged on 5.66 million viewers and critical reviews have been mixed, the show having received only a 37% rating from Rotten Tomatoes and a 48/100 from Metacritic.

In the all positive condition, there were just two shows—*The Following* (2013) and *Touch* (2012). Having just completed its third and final season, the former averaged 10.42 and 11.18 million viewers in its first two seasons, respectively. The final season premiere was down almost 57% from season 2, however. *Touch* was cancelled after its second 13-episode season, a total run only slightly more than the traditional full season of 22-24 episodes. Thus, while there is a 100% success rate associated with the all-positive condition, given the small numbers, we must be careful not to over-generalize from such a small sub-sample. The more conservative position is to state that, at present, the three variables appear to be far more useful in their null condition where they appear to excel at picking out potential ratings strike outs rather than hits or homeruns.

That said, we must also note that these facts might not have been this obvious at the time decisions were being made as to the series' fates. This is particularly true with regard to the strongest predictor—our measure of network size. To know whether a text network of any teleplay is large or small requires having a sample with which to make the comparison. An analysis done in 2010 would have had to use data from years prior and at that time, what looks small now might not have appeared so then. Thus, another way to test this decision-making strategy is to apply it to the current crop of network series.

As of the first week of December 2015, we were able to identify only one new dramatic series that met the null conditions—adapted concept, little or no prior track record of success, and a below average text network. That series is the *Minority Report* (2015), an adaptation based on both a Philip K. Dick short story by the same name published in 1956 and a movie from 2002, also of the same name and directed by Steven Spielberg. The show creator, Max Borenstein, has no prior experience writing as a show creator, and thus no track record of success. He is, however, credited as the screenwriter for the 2014 box office smash *Godzilla*. But in our analysis, this is not considered relevant experience. Nor does the fact that Stephen Spielberg is listed as an Executive Producer for the series. As for the text network of *Minority Report's* pilot episode, it has only 24 links, an amount that places it squarely in the bottom quartile in our sample. Its premiere broadcast audience was a mere 3.1 million viewers and the next four episodes have averaged a paltry 2.13 million viewers. Based on data available in the first week of 2015, our model would have predicted an average audience of 4.30 million viewers over the first five episodes—an amount double of what has actually been observed. More importantly, the draft upon which we based our analysis is the “Revised 2nd Network Draft” and is dated 08 January 2015. This means that in the first week of January 2015—just one day before it got a pilot order and a whole four months before it was formally added to Fox's Fall 2015 line-up—our model could have predicted that the show's average audience in its first five episodes would be 14% lower than the show it replaced on Monday nights—*Sleepy Hollow* (2013)—an amount far less than what was either anticipated or needed (Hibberd, 2015; Taylor, 2015).

7. Limitations

One limitation of our study concerns the representativeness of our sample. As noted previously, it is very representative in certain respects and less so in others. On the positive side, it is clearly quite representative of new dramatic television series: we found data on originality and track record for all 107 series and pilot episode teleplays for almost 92% of those. That said, because we limited our examination to hour-long programs, it is not clear if our results extend to comedies or any other scripted genres. Similarly, because we limited our analysis to the top four TV networks, it's also not clear whether our results would be the same for new series appearing on established cable networks like HBO, on smaller broadcast networks like the CW, or streamed over Amazon, Hulu, and Netflix (Tryon, 2015).

Also of concern are the pilot episodes that were burned, i.e. the ones that were greenlit by the networks but that never aired. At present we have no way of knowing whether or how those pilots differ from those of the series that made it onto the air. Also unclear is whether the burned pilots had, on average, smaller text networks than those that were aired. If they do, it suggests that knowingly or not, development executives are selecting for the same thing that our text-based measure captures. And hence, our measure is also a predictor not only of broadcast audience size, but equally a predictor of which pilots get aired and which get burned, and possibly of which teleplays become proof-of-concept pilots. On the other hand, if there are no differences in text network size of aired and unaired teleplays, then it suggests that the current decision process does not knowingly or unknowingly take size into account. This means that the application of our approach could influence how key development decisions are made.

Because the text-derived measure described herein is novel, we believe that no proxy for it is being used in the current

decision process. As such, we further believe that the more likely scenario is that there are no differences in text network size between aired and unaired teleplays. But ultimately this is an empirical question, one that can only be ascertained by further empirical research directed specifically toward this end.

References

- Anders, C. (2012, August 22). How to Tell from a Pilot if a TV Show is Going to be Any Good. *io9*. Retrieved from <http://io9.com/5936264/how-to-tell-from-a-pilot-if-a-tv-show-is-going-to-be-any-good>
- Basuroy, S., & Chatterjee, S. (2008). Fast and frequent: Investigating box office revenues of motion picture sequels. *Journal of Business Research*, 61(7), 798-803. <http://dx.doi.org/10.1016/j.jbusres.2007.07.030>
- Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing*, 67(4), 103-117. <http://dx.doi.org/10.1509/jmkg.67.4.103.18692>
- Caves, R. (2003). Contracts between art and commerce. *Journal of Economic Perspectives*, 17(2), 73-84. <http://dx.doi.org/10.1257/089533003765888430>
- Diesner, J., & Carley, K. (2004). Using network text analysis to detect the organizational structure of covert networks. *Proceedings of the North American Association for Computational Social and Organizational Science (NAACSOS) Conference*, 2004.
- Eliashberg, J., Hui, S., & Zhang, Z. (2014). Assessing box office performance using movie scripts: A kernel-based approach. *IEEE Transactions on Knowledge and Data Engineering*, 26(11), 2639-2648. <http://dx.doi.org/10.1109/TKDE.2014.2306681>
- Goldman, W. (1983). Everything you always wanted to know about stars+ selection from Adventures in the Screen Trade. *American Film*, 8(5), 57-64.
- Hibberd, J. (2015, January 9). Minority Report TV Show gets Fox pilot order. *Entertainment Weekly*. Retrieved from <http://www.ew.com/article/2015/01/09/minority-report-fox>
- Hunter, S. (2014). A Novel Method of Network Text Analysis. *Open Journal of Modern Linguistics*, 4(2), 350-366. <http://dx.doi.org/10.4236/ojml.2014.42028>
- Hunter, S., Smith, S., & Singh, S. (2016). Predicting Box Office from the Screenplay: A Text Analytical Approach. *Journal of Screenwriting*, 5(2), 135-154. http://dx.doi.org/10.1386/josc.7.2.135_1
- Kissell, R. (2015, October 9). NBC's Blindspot Fall's First New Series to Get Full-Season Order. *Variety*. Retrieved from <http://variety.com/2015/tv/news/nbcs-blindspot-full-season-order-1201614944/>
- Konda, K. (2014, May 12). Exactly how many of the major networks' new 2014/2015 shows are doomed to fail? *We Minored in Film*. Retrieved from <http://weminoredinfilm.com/2014/05/12/exactly-how-many-of-the-major-networks-new-shows-are-doomed-to-fail>
- Lindauer, S. (2011, August 7). First Impressions Count: The Importance of TV Pilots. *Pop Matters*. Retrieved from <http://www.popmatters.com/feature/142651-tv-pilots-set-tone-for-what-is-yet-to-come/>
- Litman, B., & Kohl, L. (1989). Predicting financial success of motion pictures: The '80s experience. *Journal of Media Economics*, 2(2), 35-50. <http://dx.doi.org/10.1080/08997768909358184>
- Littlejohn, J. (2007, May 10). Writers Can't Predict Success of network Pilots. *New York Daily News*. Retrieved from <http://tinyurl.com/Littlejohn2007>
- MacNabb, M. (2015, September 29). Blindspot: NBC Orders More Scripts After Premiere Ratings Success. *Screen Rant*. Retrieved from <http://screenrant.com/blindspot-nbc-premiere-ratings-season-1-episodes/>
- Meyer, D., & Hyndman, R. (2006). The accuracy of television network rating forecasts: The effects of data aggregation and alternative models. *Model Assisted Statistics and Applications*, 1(3), 147-155.
- Napoli, P. (2001). The unpredictable audience: An exploratory analysis of forecasting error for new prime-time network television programs. *Journal of Advertising*, 30(2), 53-60. <http://dx.doi.org/10.1080/00913367.2001.10673637>
- Nathanson, J. (2013, October 17). The Economics of a Hit Show. *Priceonomics*. Retrieved from <http://priceonomics.com/the-economics-of-a-hit-tv-show/>
- Nelson, R., & Glotfelty, R. (2012). Movie stars and box office revenues: An empirical analysis. *Journal of Cultural Economics*, 36(2), 141-166. <http://dx.doi.org/10.1007/s10824-012-9159-5>
- Ocasio, A. (2012, May 17). TV Success Rate: 65% of New Shows Will be Canceled (& Why it Matters). *Screen Rant*.

Retrieved from <http://screenrant.com/tv-success-rate-canceled-shows-aco-172162/>

- Rust, R., & Eechambadi, N. (1989). Scheduling network television programs: A heuristic audience flow approach to maximizing audience share. *Journal of Advertising*, 18(2), 11-18.
<http://dx.doi.org/10.1080/00913367.1989.10673147>
- Schwab, C. (2013, October 8). Rating Prediction: How to forecast the success of new TV shows. *Media Forecast*, Retrieved from <https://mediaforecast.wordpress.com/2013/10/08/rating-prediction-forecast-tv-shows/>
- Stelter, B. (2012, May 13). Few TV Shows Survive a Ruthless Proving Ground. *New York Times*. Retrieved from <http://www.nytimes.com/2012/05/14/business/media/few-tv-shows-survive-a-ruthless-proving-ground.html>
- Taylor, B. (2013, March 8). How Moneyball Can Change TV Too. *Huffington Post*. Retrieved from http://www.huffingtonpost.com/brian-taylor/moneyball-for-tv_b_2832414.html
- Taylor, S. (2015, October 11). Fox Cuts Minority Report Season to 10 Episodes: Cancelled Soon?, *MStar News*. Retrieved from <http://www.mstarz.com/articles/87734/20151011/fox-cuts-minority-report-season-10-episodes-canceled-soon.htm>
- Tryon, C. (2015). TV Got Better: Netflix's Original Programming Strategies & Binge Viewing. *Media Industries*, 2(2), 104-116.
- Walls, D. (2005). Modeling movie success when nobody knows anything: Conditional stable-distribution analysis of film returns. *Journal of Cultural Economics*, 29(3), 177-190. <http://dx.doi.org/10.1007/s10824-005-1156-5>



This work is licensed under a [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/).