Predicting New TV Series Ratings from their Pilot Episode Scripts

Starling David Hunter[1], Susan Smith[2] & Ravi Chinta[3]

[1] Business Administration Program, Carnegie Mellon University Qatar, Doha, Qatar

[2] School of Mass Communication, American University of Sharjah, Sharjah, United Arab Emirates

[3] College of Business, Auburn University, Montgomery, Alabama, USA

Correspondence: Starling David Hunter, Business Administration Program, Carnegie Mellon University, Doha, Qatar. E-mail: starling@andrew.cmu.edu

Abstract

Empirical studies of the determinants of the ratings of new television series have focused almost exclusively on factors known *after* a decision has been made to broadcast the series. The current study directly addresses this gap in the literature. Specifically, we first develop a parsimonious model to predict the audience size of new television series. We then test our model on a sample of 116 hour-long, scripted television series that debuted on one of the four major US television networks during the 2009-2014 seasons. Our key predictor is the size of the main component of the text network developed from the script of the pilot episode of each series. As expected, this size measure strongly explains the number of viewers of the new series' first several episodes.

Keywords: television, network analysis, pilot episodes, ratings, audience, semantic network analysis, screenplay

1. Introduction

In the spring of 2011 an article appeared in the *Wall Street Journal* entitled "The Math of a Hit TV Show" (Chozick, 2011). Of particular note was the article's detailed description of the gauntlet that is the new TV show development process. In its first stage, which we are told begins in early summer, each of the big four US networks receives about 500 "elevator pitches" or loglines, each of which describes the basic idea for a new series. Next, through a process unique to each network, the review of the 500+ pitches results in about 70 scripts being commissioned. In the third stage, approximately one-third of these scripts get the green light, i.e., the show creators are given money to produce a proof-of-concept pilot episode. Once the proof-of-concept pilots have completed filming, they are subjected to varying levels and kinds of marketing research, focus-group testing being one of the most common. Depending on a variety of factors, not the least of which is the strength of a network's current slate of shows, anywhere from 4-8 of these completed pilots are given slots in the network's lineup. At this point the show's creative team ramps up its staff, particularly with writers who begin penning the next several episodes of the series. No sooner do episodes appear on the air, then their ratings and viewership are carefully scrutinized. Based on how well those episodes are received, changes may be made to the characters and story lines. Even the order of episodes may be altered. In the worst case scenario, shows falling far below ratings or audience expectations can be cancelled or replaced after as little as 2-3 weeks. The networks' post-New Year schedules always contain a raft of mid-season replacements and, like the survivors from the fall line-up, they may be subject to changes right up to and including the airing of the season's final episode. In mid-May, at the industry confab known as the "up fronts," the networks announce their lineups for the coming fall's season. A few weeks later, as the networks once again start receiving pitches, the development process for the next year begins anew.

Also of note in the article was its description of the strategies that executives, producers, and show creators use to improve their odds of first getting a show on the air and then keeping it there. One such strategy was actually spelled out in the article's subtitle—"For New Shows, Networks Try Familiar, with a Little Twist." Just such an example was *Grimm* (2011)—a cop show (familiar) with a twist (characters inspired by Grimm's fairy tales). Chief among the other strategies include focus-group testing and multiple rounds of script rewrites and revisions. Networks are also said to differ in their mix and structure of the strategies employed. For example, CBS—then and still the top-ranked network in total prime-time viewers—limits to just four the number of people providing input on new series—the CEO, the president of the entertainment division, the head of development, and the

show runner. At other studios, the process is known to involve many more people, leading some to remark upon the old adages about the number of chefs and the preparation of soup and camels being horses designed by committees. But no matter the structure of the decision process or the number of decision makers, one inarguable mathematical fact remains: the large majority of new shows fail within two seasons, thus falling very short of becoming hits (Bielby & Bielby, 1994; Nathanson, 2013).

The reason why such high failure rates matter is because television networks earn a large portion of their revenue from the sale of blocks of time to advertisers. The prices that they charge for new shows are a function of the projected audience—both in terms of size and demographics. When it comes to series that have already been on the air for one or more seasons, there is already a wealth of information upon which to base those projections (Danaher, Dagger, & Smith, 2011). But with new shows, little if any of that information is known either when key development decisions are being made or when advertisers begin buying time slots, which is shortly after the up fronts. When a new series fails to deliver the projected audience, it runs the risk of cancellation. But in addition, advertisers who bought time slots on the underperforming series are entitled to partial refunds or airtime on other shows. They are not, however, compensated for the opportunity cost of their wrong decisions. Because new shows typically comprise 20-40% or more of a network's fall lineup, they are a source of substantial uncertainty for both buyers (advertisers) and sellers (studios and networks). Those who can manage that uncertainty stand to profit handsomely. As Litman (1979) noted, "program executives who can successfully predict how viewers will respond to different types of programs can be expected to make fewer development and scheduling mistakes, hold down programming costs, and win the ratings game."

Unfortunately for both the networks and their advertisers, predictive models of the ratings performance of new television series are both scarce and inaccurate (Napoli, 2001). Further complicating matters is the dearth of empirical models for predicting new series performance—either from the pre-production stage or after the fact. There is, however, a small but burgeoning literature in a closely-related field—that of cultural economics—that has developed models for predicting box office revenues using only information known during pre-production. Of particular importance in those models is the utilization of variables derived from what appears on the page, i.e., what is found through the textual and content analysis of the scripts. These are factors that have gone all but unnoticed in the academic literature on television ratings. To bridge this gap, we draw upon the film studies and cultural economics literature to develop an early-stage model for predicting total network viewership of new dramatic television series. In particular, we focus on a text-analytical measure developed by Hunter, Smith, & Singh (2016). As predicted, we find that even when controlling for the track record of success of a new show's creators, the originality of that show's concept, and the television network on which the show appears, this text-analytical measure is a statistically significant predictor of audience size through at least the first five episodes of a new series' first season.

The remainder of this paper is organized as follows. In the next section we summarize the academic literature on television ratings and argue for the adaptation of its models to the study of new television show ratings. In the third section we describe our data and statistical methods. The four sections contain a discussion of the results of the analysis while in the final section we discuss implications of the same.

2. Literature Review

As noted in the introduction, whether measured in absolute or relative terms, the failure rate of new television shows is very high and has been for decades (Bielby & Bielby, 1994). Despite the fact that such failure rates are costly to industry participants—particularly the television networks and the advertisers—the determinants of success and failure remain poorly understood (Littlejohn, 2007). Compounding this problem is the fact that empirical research in the area of television studies has provided few, if any, useful insights or solutions. But as also noted in the introduction, there is small body of relevant work in a companion literature—the field of cultural economics—that sheds light upon the prediction of television ratings. Specifically, we refer to three recent studies that explain variation in box office revenues using only variables that are known during the pre-production stages.

The first of the three is authored by Goetzman, Ravid, & Sverdlove (2013) who investigated the "forward looking" nature of prices paid by movie studios for screenplays (p. 277). The authors predicted and found that prices were positive and highly significant predictors of the ensuing film's box office receipts. Because the screenplays were purchased very early in the development process, the implication is that price serves as a "signal for the perceived quality of the subsequent project" (p. 297).

The second of the relevant studies is one by Eliashberg, Hui, & Zhang (2014) who used four groups of variables derived through textual and content analysis of screenplays. The first of these four groups was the film's *genre*,

i.e., whether the film was a comedy, western, action, drama, etc. The second pertained to *story line* or *content* variables such as the likability of the protagonist, early exposition of information about the same, the presence of a surprise ending, or an unambiguous resolution to the central conflict. The third group consisted of *semantic* features of the text such as the total number of scenes and the average length of dialogs. The fourth and final group of predictors were two "bag-of-words" measures that captured styles and frequencies of individual words in the text. The authors used both human coders and computational methods to determine the levels of these variables in a sample of 300 shooting scripts of films released between 1995 and 2010. As predicted, they found that one or more variables in all four categories were strongly predictive of box office revenues.

A third study attempting early-stage prediction of box office revenue is one by Hunter, Smith, & Singh (2016). Similar to Eliashberg et al., they relied on text-derived variables in their analysis of the screenplays of 170 US-produced feature films released in 2010 and 2011. The specific method they employed was "network text analysis," a software-supported approach for constructing networks of interconnected concepts from documents. Consistent with research in the fields of educational psychology and socio-linguistics, they found that the size of the text network created from selected words in each film's screenplay was positively and quite significantly associated with opening weekend box office, even when controlling for several other covariates.

Taken together, these three studies show that there are reliable predictors of box office performance whose values can be known or reasonably inferred at very early stages in the film development process. While film and television development are not identical, they are similar enough in both structure and intent—and sometimes personnel—that it is not unreasonable to examine whether the same relationships might also hold between textual properties of television scripts and subsequent performance. That said, one thing about television development and production that does not have its analog in film is the critical importance placed upon the pilot episode (Littlejohn, 2007). In particular, pilot episodes are supposed to "set the tone for the series" (Lindauer, 2011) and "to establish the characters and situations" that will recur episode after episode (Anders, 2012). And because the initial performance of the pilot so strongly impacts whether the series will get either an early cancellation notice or a full-season order (MacNabb, 2015; Kissell, 2015), we believe that the text-derived properties of the pilot episode in particular will be determinative of the ratings performance of the all-important, first several episodes of a new series. Specifically, we thus hypothesize that: *all else equal, the size of the text network of the teleplay of a new series' pilot episode will be positively associated with the series' initial ratings performance.*

3. Methods & Data

In order to investigate the aforementioned hypothesis, data were collected on the total number of viewers of new prime-time, hour-long television series debuting during six recently-completed broadcast seasons (2009-2014). Following Napoli (2001) only shows debuting on the Big Four US television networks—ABC, CBS, NBC, and FOX—were included. We used several sources to determine which shows appeared during those seasons. These included *TV Series Finale*, *TV.com*, *TV Guide*, and *Wikipedia*, particularly the latter's series of "US Network Television Schedule" pages for these seasons.

We identified a total of 136 new, hour-long, dramatic series that debuted in prime-time as part of the 2009-2014 television seasons. Six of these shows were eliminated from consideration because they were (co-)produced with or produced by foreign television networks and debuted in those countries before being seen on US network television. They were *Rookie Blue* (2009) and *The Bridge* (2009), both Canadian police dramas; *Welcome to Sweden* (2014), an American-Swedish co-production; *Crossing Lines* (2013)—a Franco-German-Belgian co-production; *Taxi Brooklyn* (2014)—a French-American action-comedy; and *The Firm* (2012)—a Canadian-produced adaptation of John Grisham's novel by the same name.

We also eliminated five "back-door" pilots, i.e., episodes of long-running shows that introduced one or more guest characters for what would become a new series in the next television season. The back-door pilots were *NCIS- Los Angeles* (2009), *NCIS-New Orleans* (2014), *CSI Cyber* (2015), *Criminal Minds: Suspect Behavior* (2011), and *The Finder* (2012). Of the remaining 125 new series that had traditional pilot episodes, we were unable to locate copies of teleplays for 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 119 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. These were *Smash* (2012), *The Undercovers* (2010), and *My Generation* (2010). Our primary sources for the scripts were TV Writing (sites.google.com/site/tvwriting/us-drama/pilot-scripts), *Write-to-Reel* (writetoreel.com), *Simply Scripts* (simplyscripts.com) and *Script Fly* (scriptfly.com).

3.1 Dependent Variable

Our measure of the ratings performance for the 116 new series was the total broadcast viewership, as measured in millions of viewers, in each of the first five episodes of the first season. We obtained the data from a number of sources including *TV Series Finale, Tv.com, TV by the Numbers*, 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, LOGVIEW averaged 6.79 for the first five episodes of the 116 new series in the sample. This value corresponds to 6.17 million viewers. The minimum and maximum values of LOGVIEW were 6.20 and 7.22, respectively. These correspond to the 1.58 million viewers who tuned in for the fourth episode of the ill-fated medical drama *Do No Harm* (2013) and the 16.5 million viewers of the pilot episode of the cyber-themed, action-adventure series *Intelligence* (2014).

Table 1. Descriptive statistics

Variable	N	Mean	Min	Max
LOGVIEW				
• Episode 1	116	6.89	6.42	7.22
• Episode 2	116	6.81	6.34	7.14
• Episode 3	114	6.77	6.23	7.14
• Episode 4	113	6.75	6.20	7.11
• Episode 5	112	6.74	6.21	7.12
ADAPT	571	0.35	0	1
RECORD	571	0.41	0	2
LOGLINK	571	1.66	1.04	2.34

3.2 Independent Variable

Several distinct approaches exist for creating networks from texts (Nerghes, Lee, Groenewegen, & Hellsten, 2015). They differ along a number of dimensions including the level of automation, whether and how words are abstracted to higher-order conceptual categories, and the nature of the underlying relationship used to connect the words or concepts. In this study we opted for Hunter's (2014) morpho-etymological approach, one which is semi-automated, which abstracts words into higher-order conceptual categories defined by common etymological root, and which connects conceptual categories according to their co-occurrence in "multi-morphemic compounds" (MMCs). 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), *blend words* (brunch, biopic, 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). As Hunter (2014) noted in his study of a sample of Academy Award-nominated original screenplays, MMCs are highly complex, as measured by the number of characters, and consist largely of unique, context-specific terminology, conceptual vocabulary, jargon, or lexicon of the kind that distinguishes film genres from one another.

The first step in the creation of the text networks involved identifying the MMCs in each script. To accomplish this we first used the *Generate Concept List* and the *Identify Possible Acronyms* routines in the CASOS Institute's *Automap* software program (Carley & Diesner, 2005) to generate word lists for each script in the sample. Each word list was analyzed by a pair of the authors with the intent of identifying all of the MMCs contained therein. Then each pair, in conjunction with the corresponding author, reconciled all differences in coding choices. Across the 116 scripts in the sample, we identified 5861 unique MMCs in the sample appearing a total of 14013 times. That makes for an average of almost 121MMCs per script or about 2-3 per page.

The second step involved decomposing every MMC in each list into its constituent words. For example, the closed compound *policeman* is comprised of two words—*police* and *man*. 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 *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 *police* descends from the IE root *pele-3* which means "citadel, fortified high place," while the word

man descends from the IE root *man-1*, which means "man." This stage of the analysis was software-supported. Specifically, we first created a database containing the entire contents of the AHDIER. It maps over 13,000 unique words to nearly 1300 different roots whose descendants co-occur in tens of thousands of MMCs, many of which were in contained in our sample. We then automatically assigned over 83% of the constituent words to one of 752 Indo-European roots. The remaining 17% were instances where etymological roots of constituent words were not Indo-European or did not exist. In the former 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. In the latter case, where words had no known etymological root, the base form of the word was used.

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). In social network analysis, the largest cluster of mutually-reachable nodes in a network is referred to as the "main component." Our measure of the size was the number of links contained in the main component of the text network constructed from the MMCs contained in the script of a new series' pilot episode. Figure 1, below, depicts a portion of the main component of the text network constructed from the script of the pilot episode of Fox's dramatic series *The Following* (2013), as well as several of the network's minor components. The main component has 28 nodes while the six minor components have a total of twenty-five nodes among them, with a range of 2-6 nodes apiece. As noted above, the nodes in the network are etymological roots while MMCs are associated with the links between pairs of nodes. The MMCs in the displayed portion of the text network of *The Following* included closed compounds (*madman*, *classroom*, *bloodbath*, *courtroom*, *Nevermore*), acronyms (*GPS*, *SUV*, *BAU*, *CNN*), a hyphenated compound (*college-aged*) and two clipped words (*ethernet* and *internet*). As noted in the descriptive statistics displayed in Table 1, the average of the log of the number of links in the 116 text networks is 1.65, a value that corresponds to an average size of 56 links. The variable that contains the measures of network size for the series in the sample is named LOGLINK. From this variable we created a categorical variable, TOPLINK, which was coded "1" if the value of LOGLINK was in the 25th percentile and coded "0" otherwise.
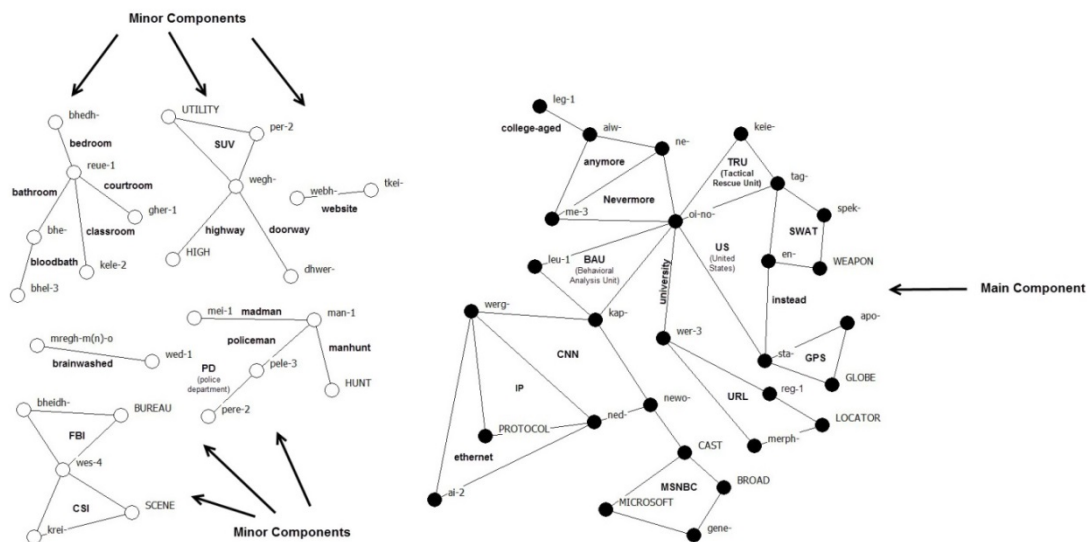

Figure 1. Portion of text network for *The Following*

3.3 Control Variables

3.3.1 Concept Originality

Our review of the television studies literature was unable to identify any prior empirical research that evaluated the influence of concept originality on a television series' ratings—be that series new or ongoing. That said, research in film studies, most notably that concerned with the determinants of box office revenues, has examined a closely-related question. Specifically, research has found that sequels in particular (Basuroy & Chatterjee, 2008) and adapted premises more generally (Hunter, Smith, & Singh, 2016) are associated with higher box office performance. As such, we include as a control variable in our statistical model a dummy variable named ADAPT

which was coded "1" if the show was adapted from prior source material, be that a novel, a comic book, a film or film franchise, another TV series (past or present), a stage play, etc. and coded "0" otherwise.

3.3.2 Track Record

Again, our review of the television ratings literature returned no empirical studies examining the impact of the track record of the creative team on a series' ratings performance. But as above, research on the determinants of box office has addressed a closely-related matter. Particularly, Nelson & Glotfelty (2012) reported that the star-power of directors has a positive and statistically significant impact on box office revenue of film projects to which they are attached. Further, Hunter, Smith, & Singh (2016) reported that the box office of a screenwriter's last film project is a positive and statistically significant predictor of the box office revenues associated with their current film project. In the present study, we developed a measure of prior success of the new series' creator(s). Specifically, we used the International Movie Databse (IMDb) to first determine the number of series for which the creator(s) had earned a writing credit for the script of the pilot episode. From among those credits, we determined how many of the shows had been renewed, i.e., that had aired for at least two seasons. We created a Likert-scaled variable, RECORD, where creative teams with no prior successes were assigned a score of zero (n = 80), those with one prior success were assigned a score of one (n = 25), and those with two or more prior success were assigned a score of two (n = 11).

3.3.3 Broadcast Network

Following Napoli (2001), we created dummy variables to capture unexplained heterogeneity among the four networks whose series are the object of this analysis. Because CBS has been the ratings and audience leader among the four networks over the entire span of our observations, we created three dummy variables representing the three other networks. Specifically, the first we created was named "ABC" to represent the American Broadcasting Company. It was coded "1" if the new series appeared on ABC and coded "0" otherwise. The two variables representing Fox and the NBC were constructed in an analogous fashion. Table 2, below, contains the correlation matrix for all of the aforementioned variables. Excluded, however, are correlations of the network dummy variables with another.

Table 2. Correlation matrix

	LOGVIEW	ADAPT	RECORD	TOPLINK
ADAPT	-0.163#			
RENEW	0.165#	0.089		
TOPLINK	0.470****	-0.025	0.085	
ABC	-0.026	0.091	-0.095	-0.056
CBS	0.361****	-0.082	-0.004	0.233*
FOX	-0.122	0.042	0.120	-0.124
NBC	-0.190*	-0.055	-0.002	-0.045

4. Results

Table 3, below, contains the results of two random effects, generalized least squares (GLS) and five ordinary least squares (OLS) regression models used to test our hypothesis. In all models the dependent variable is LOGVIEW, the log of the total broadcast viewing audience while the independent variable is TOPLINK. In both models, data from only the first five episodes of the new series' first season are used. The random effects regression is the appropriate choice in the first model because independent and control variables are time-invariant. That is to say, their values don't change across the five episodes. In the latter models, an OLS regression is appropriate because only one episode is considered at a time and thus, there is no question of (in) variance of across episodes.

Table 3. Results of Generalized (GLS) & Ordinary Least Squares (OLS) regressions for first five episodes

	Random Effects Models		*Episode-Specific Models*				
	Episodes 1-5	Episodes1-5	Episode1	Episode2	Episode3	Episode4	Episode5
ADAPTED [a]	-0.064*	-0.064*	-0.048	-0.059*	-0.081**	-0.071*	-0.064*
	(-2.13)	(-2.15)	(-1.64)	(-2.02)	(-2.69)	(-2.09)	(-1.73)
RECORD [b]	0.048*	0.048*	0.038*	0.049*	0.054*	0.046#	0.053*
	(2.17)	(2.19)	(1.76)	(2.24)	(2.40)	(1.82)	(1.96)
TOPLINK [c]	0.148****	0.149****	0.158****	0.173****	0.149****	0.134****	0.132***
	(4.32)	(4.38)	(4.73)	(5.25)	(4.37)	(3.50)	(3.18)
Episode = PREMIERE		0.128****					
		(18.98)					
Network = ABC	-0.104**	-0.105**	-0.074*	-0.096*	-0.098*	-0.115**	-0.129*
	(-2.49)	(-2.53)	(-1.82)	(-2.38)	(-2.34)	(-2.45)	(-2.56)
Network = FOX	-0.133**	-0.134**	-0.117**	-0.118*	-0.127**	-0.141**	-0.148**
	(-2.76)	(-2.82)	(-2.50)	(-2.54)	(-2.63)	(-2.59)	(-2.53)
Network = NBC	-0.198****	-0.199****	-0.165****	-0.183****	-0.196****	-0.214****	-0.234****
	(-4.79)	(-4.84)	(-4.09)	(-4.50)	(-4.77)	(-4.62)	(-4.66)
# Observations	571	571	116	116	114	113	112
# of Series Pilots	116	116	116	116	114	113	112
Model χ^2-statistic	64.9****	426.0****					
R^2-within	0.0%	44.5%					
R^2-between	37.3%	36.5%					
R^2-overall	31.9%	38.0%					
Model F-statistic			10.1****	11.3****	11.6****	9.0****	7.9****
Model -R^2			34.7%	38.4%	39.4%	33.7%	31.3%
Adjusted-R^2			31.1%	35.0%	36.0%	30.0%	27.4%

Note. **** $p < 0.0001$, *** $p<0.001$** $p <0.01$*$p<0.05$ # $p<0.10$, all 1-tailed tests. (Note 1)

In short, the results of the regression analyses show very strong support for our hypothesis. In each model the coefficients associated with our independent measure, TOPLINK, are positive and highly significant statistically. The first model of the seven described in Table 3 specifies a random effects model of all five episodes. In this model, coefficients were estimated on a sample of 571 observations from all 116 new dramatic television series, an average of about 4.9 episodes per series. While all variables in the model are significant, one of the two most highly so is TOPLINK (= 0.152, p < 0.0001, 1-tailed test), and in the predicted direction. The same holds true for the remaining five, single-episode, OLS regression models. In each instance, the coefficient for TOPLINK is positive and highly, statistically significant ($0.134 < \beta < 0.173$, $0.0001 < p < 0.001$, 1-tailed test). Two notable trends are evident. First, there is a substantial increase in the coefficient value, statistical significance, and proportion of variance explained (R^2) between the model for Episode 1 and 2. From that point on, all of these values monotonically decrease. When we recall that almost without exception the first episode's audience is the largest of the season, then it suggests that the inclusion of an additional dummy variable to distinguish initial episodes from others might be in order. As shown in the second GLS model, which includes just such a variable, this supposition is confirmed. Specifically, the beta coefficient of that variable was very highly significant (p <0.0001) and the overall R^2 of the model increased from 32.7% to 38.9% while the significance of all other covariates stayed the same or improved. And because this variable is, by definition, not the same for all episodes, the within-sample R^2 value climbed from 0% in the first model to 44.5% in this one.

More generally, it can be observed that all other covariates were significant at or above the p< 0.05-level (1-tailed test). With one notable exception, the sign of the coefficients was in the expected direction. That exception was associated with the dummy variable ADAPT. Instead of the expected positive relationship, like that found in film studies between adapted concepts and box office, our results showed that new series with adapted concepts had significantly *smaller* initial audiences than did new series with original concepts. Also noteworthy is the significance of the dummy variables presenting the three broadcast networks—ABC, FOX, and NBC. They are each negative and highly significant in every model that we specified. This result confirms that in comparison to CBS, which was the reference category, new series on these three networks had much smaller initial audiences.

5. Discussion & Conclusion

The results presented strongly support our hypothesis that there should be a positive and significant relationship between the size of the text network of the pilot episode of a new series and the size of the initial audience of the

series. There are several important implications of these findings that merit further discussion. First, recall that unlike research on the determinants of box office, there is a dearth of research attempting to explain ratings or other measures of performance using only factors known during the early stages of production. This study is the first one of which we are aware that addresses this gap in the literature. Like those studies of box office performance, this one emphasizes that both characteristics of the creative team and properties of the text of the script itself are significant predictors of performance, especially the latter. What this study adds to that small but burgeoning literature is further confirmation of the decision to model performance using a limited number of early-stage factors.

Secondly, while our model does explain a high proportion of the variance in initial audience size, we should make clear the distinction between explaining variance with a model constructed from a sample of shows from the past and using that model to predict the audience size of shows currently appearing on network television. The results in Table 4, below, present our model's predictions for the seven series debuting in the fall of 2015 for which we were able to locate pilot scripts by December 1st, 2015—*Blindspot, Code Black, Limitless, Minority Report, Quantico, Rosewood,* and *Wicked City.* The second column contains the model's predictions of average viewership of the first five episodes. The third column reports the observed viewership numbers. The fourth column contains the difference between the observed and predicted amounts expressed as a percentage. The smallest differences were for *Limitless* (+0.45%) and *Quantico* (+3.7%). The largest differences in absolute terms were also negative. The two series in question were *Wicked City* (-57.8%) and *Minority Report* (-50.7%). Both under-performed the modest audience sizes projected for them. The fifth and sixth columns indicate the percentile ranking in the 2009-14 sample of the predicted and observed audience numbers. For example, the 8.98 million viewer prediction for *Limitless* falls just below the 80th percentile of initial audience figures for the 116 series in the sample. *Minority Report*'s observed average of 2.32 million viewers is in the lower 5% of that sample. The final column provides information about each series' status as of December 20th, 2015. Here we can see how early within the season important decisions were made concerning the fate of the shows. After just their third episodes, *Wicked City* was cancelled and *Minority Report's* initial order was reduced from 13 episodes to 10. Industry observers are taking the latter to be a signal of impending cancellation (Surette, 2015; Wagmeister, 2015; Piester, 2015). Both of these shows were performing well below even the modest audiences predicted for them by our model. On the positive side, *Quantico*, *Rosewood*, and *Limitless* received full season orders after their third, fourth, and fifth episodes respectively while *Blindspot* was renewed for a second season after its eighth. Further, after its fourth episode, *Code Black* received an order for six additional scripts. Notably, all five of these shows were performing above the audience levels that our model predicted. Taken together, the information contained in this table underscores the value in being able to anticipate a new series' early audience size and performance.

Table 4. Model predictions of average audience of first five episodes of selected 2015 series

Show (Network)	Predicted Viewership (millions)	Actual Viewership (millions)	(Actual-Predicted)	Predicted Rank in Sample (%)	Actual Rank in Sample (%)	Status
Blindspot (NBC)	6.59	9.01	+37%	<55%	<80%	Renewal announced after 8th episode.
Code Black (CBS)	6.39	7.09	+11%	<50%	<60%	Six additional scripts ordered after 4th episode; five more ordered after 8th episode. Renewed after end of season.
Limitless (CBS)	8.98	9.02	+0.45%	<80%	<80%	Full season order announced after 5th episode. Cancelled at end of season.
Minority Report (Fox)	4.71	2.32	-50.7%	<30%	<5%	After 3rd episode, the series order was cut from 13 episodes to 10. Cancelled at end of season.
Quantico (ABC)	5.83	6.05	+3.7%	<45%	<50%	Full season order (19 episodes) after 3rd episode; three more ordered after 6th episode.
Rosewood (Fox)	5.46	5.97	+9.3%	<40%	<45%	Full season order announced after 4th episode. Renewed for full season after 15th episode.
Wicked City (ABC)	5.83	2.46	-57.8%	<45%	<5%	Cancellation announced after 3rd episode.

Third, recall that very little empirical work, if any, has examined the series development process and the decision-making occurring in or across its early stages. And while we are certainly not suggesting that an audience estimate by our model would be the *only* factor taken into account in key decisions such as whether and when to air a new series, to order new episodes, to cancel, or to renew, we very much mean to suggest that such estimates could be taken into account. Specifically, we maintain that our sample of new series, taken as a whole, represent an expanded set of "comps", i.e., a set of TV scripts that are comparable on many dimensions, one of which is our measure of text network size. Both predicting future performance and evaluating observed performance could potentially be informed by the estimates of our model.

Fourth, we might further add that in media landscape that saw a reported 409 scripted series in 2015—and number which is up 9% from 2014 and up 100% since 2009—the quality of scripts is a very real and pressing concern (Littleton, 2014). As Hibberd (2015) of *Entertainment Weekly* recently opined,

The problem is that as the number of shows increase, the typical audience size for each show declines… (and) at a certain point, it will theoretically be impossible for networks to keep making higher and higher quality shows for an audience that's increasingly divided.

That comment came on the heels of FX Networks chief John Landgraf's recent lament that "there is simply too much television… (and that)…there is too much competition… (too) hard to find good shows…and…impossible to maintain quality control" (Littleton, 2015). These issues of quality and competition suggest another possible use for the analytical approach outlined herein. Particularly, the possibility exists to use it in that stage of the development process where decisions are being made about which pilots should have proof-of-concepts commissioned. Given the results of this study, it is certainly possible that the application of our method could identify the potentially weaker scripts in a sample. It might also help to quantify differences in re-writes and revisions of scripts as the move through the development process.

Finally, this matter of quality may also be applied in the post-broadcast period as well. Recall that the decision to broadcast a series is typically made on the basis of the proof-of-concept pilot episode. Only after that decision is made does the creative team—assisted by a team of newly-hired writers—begin work on penning the next several episodes needed to complete the network order. This study made no use of these subsequent episodes. There are two reasons why. First of all, because of the focal importance of pilot episodes and their scripts, and because those scripts are so widely available, we rightly focused our attention there. Secondly, subsequent episodes are almost never found online or made available for sale. Thus, our analysis was hampered by the lack of availability of scripts for later episodes. If, however, we were able to obtain them, it would be a relatively straightforward task of comparing both the size—and even the content—of the pilot and later episodes, and thus their quality. Differences in the sizes might also explain some of the variation within and across series' viewership. We anticipate undertaking research in the near future that directly examines this question.

References

Anders, C. (2012). 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

Bielby, W. T., & Bielby, D. D. (1994). "All Hits Are Flukes": Institutionalized Decision Making and the Rhetoric of Network Prime-Time Program Development. *American Journal of Sociology,* 1287-1313. http://dx.doi.org/10.1086/230412

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). Ucinet for Windows: Software for social network analysis. *Discover the world's research.*

Carley, K. M., & Diesner, J. (2005). AutoMap: Software for network text analysis. Center for Computational Analysis of Social and Organizational Systems, Carnegie Mellon University, Pittsburgh, PA, USA.

Chozick, M. (2011). The Math of a Hit TV Show. *Wall Street Journal.* Retrieved from http://www.wsj.com/articles/SB10001424052748703864204576315240324571266

Danaher, P. J., Dagger, T. S., & Smith, M. S. (2011). Forecasting television ratings. *International Journal of Forecasting, 27*(4), 1215-1240. http://dx.doi.org/10.1016/j.ijforecast.2010.08.002

Eliashberg, J., Hui, S. K., & Zhang, Z. J. (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

Goetzmann, W. N., Ravid, S. A., & Sverdlove, R. (2013). The pricing of soft and hard information: economic lessons from screenplay sales. *Journal of Cultural Economics, 37*(2), 271-307. http://dx.doi.org/10.1007/s10824-012-9183-5

Hibberd, J. (2015). 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, 7*(2), 135-154. http://dx.doi.org/10.1386/josc.7.2.135_1

Kissell, R. (2015). 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/

Lindauer, S. (2011). 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. R. (1979). The television networks, competition and program diversity. *Journal of Broadcasting & Electronic Media, 23*(4), 393-409. http://dx.doi.org/10.1080/08838157909363949

Littlejohn, J. (2007). Writers Can't Predict the Success of Network Pilots. New York Daily News. Retrieved from http://tinyurl.com/Littlejohn2007

Littleton, C. (2014). How Many Scripted Series Can the TV Biz—and Viewers—Handle? Variety. Retrieved from http://variety.com/2014/tv/news/new-television-fall-season-glut-of-content-1201306075/

Littleton, C. (2015). FX Networks Chief John Landgraf: There Is Simply Too Much Television. Variety. http://variety.com/2015/tv/news/tca-fx-networks-john-landgraf-wall-street-1201559191/

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/

Napoli, P. M. (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). The Economics of a Hit TV Show. *Priceonomics.* Retrieved from http://priceonomics.com/the-economics-of-a-hit-tv-show/

Nelson, R. A., & 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

Nerghes, A., Lee, J. S., Groenewegen, P., & Hellsten, I. (2015). Mapping discursive dynamics of the financial crisis: a structural perspective of concept roles in semantic networks. *Computational Social Networks, 2*(1), 1-19. http://dx.doi.org/10.1186/s40649-015-0021-8

Piester, L. (2015). How TV Networks are Secretly Canceling Shows Without You Knowing. E! Online. Retrieved from http://www.eonline.com/news/710893/how-tv-networks-are-secretly-canceling-shows-without-you-knowing

Surette, T. (2015). Minority Report Looks Like it Will be the First Canceled Show of the Season. TV.com. Retrieved from http://www.tv.com/shows/minority-report/community/post/minority-report-cancellation-144415085612/

Travers, B. (2014). Are There Too Many TV Shows? With 350 Airing This Year, Some Say Yes. Indie Wire. Retrieved from http://www.indiewire.com/article/television/are-there-too-many-tv-shows-some-producers-argue-yes

Wagmeister, E. (2015). Fox Cuts Minority Report Order to 10 Episodes. Variety. Retrieved from http://variety.com/2015/tv/news/minority-report-cancelled-10-episodes-fox-1201615074/

Notes

Note 1: [a] Dummy variable where yes = "1" if the show/series concept is adapted from other source material and "0" otherwise. [b] A Likert-scaled variable where the number of prior shows/series lasting two or more seasons for which some member of the creative team(co-) wrote the pilot episode. If the number of prior shows is 0, the variable was scored "0"; it was scored "1" if there was only one such prior success; and it was scored "2" if there

were two or more such prior successes. [c] Dummy variable coded "1" if the value of LOGLINK was in the top quartile of its distribution and "0" otherwise.

Copyrights