

A mathematical model of cinematic box-office dynamics with geographic effects

D. A. EDWARDS*

Department of Mathematical Sciences, University of Delaware, Newark, DE 19716-2553, USA

*Corresponding author: edwards@math.udel.edu

AND

R. BUCKMIRE AND J. ORTEGA-GINGRICH

Department of Mathematics, Occidental College, 1600 Campus Road, Los Angeles, CA 90041, USA

ron@oxy.edu ortegagingrich@oxy.edu

[Received on 9 July 2012; accepted on 8 March 2013]

A new deterministic mathematical model for North American box-office film grosses is presented. The model may be simplified to a set of non-linear ordinary differential equations describing the evolution over time of the film's gross and exhibited sites. The novel feature of this work is the inclusion of geography-based effects to model moviegoer and exhibitor behaviour. Several key regimes are identified, depending on the popularity of the film, as well as how the screens are divided among geographical regions. Analytical results are presented for several relevant cases. Numerical simulations demonstrate close agreement between the model's predictions and actual box-office data.

Keywords: box office; films; motion picture studies; deterministic models; numerical simulations; mathematical model; movies; differential equations; prediction; motion pictures.

1. Introduction

Since the publication of our first paper in this area (Edwards & Buckmire, 2001), the study of the motion picture industry has become quite a popular academic research activity. Subsequently, there have been several thorough surveys of the research literature, such as those by De Vany (2006), Walls (2008), Chisholm (2011), McKenzie (2012) and Eliashberg *et al.* (2006). De Vany and Walls can probably be considered the most prominent researchers in this area, publishing together (De Vany & Walls, 1996, 2002, 2004) and separately (De Vany & Eckert, 1991; Walls, 2005a,b,c) field-defining and classic contributions to the literature (De Vany & Walls, 2004), including two important books (De Vany, 2004; Walls, 2008). Chisholm (2011) provides a useful recitation of the early history of the motion picture industry and a clear explanation of why the topic is so attractive to analysis by economists. McKenzie (2012) has an excellent summary of the wide diversity of problems that have been investigated and includes approximately 100 references to published academic work. Eliashberg *et al.* (2006) provide an overall review of the 'motion picture studies' literature that includes nearly 150 references as they try to highlight important areas of future research that will be of practical use to industry professionals.

The unique structure of the motion picture industry provides several areas for mathematical investigation. Movie studios (and producers) put up the money to pay for the cast, marketing, production and distribution costs. The studios are more commonly called *distributors* in the parlance of the field. The *exhibitors* are the entities that actually show the film to the public in movie theatres. When the film

is released, contracts are negotiated between the exhibitors and distributors splitting the income generated by the film in specific ratios which generally favour the distributors in the early weeks of release but skew towards the exhibitors after an initial contract period typically lasting 2–4 weeks. Some of the interesting questions involving the motion picture industry which have been academically investigated include the effect on a film's box-office gross of pre-release advertising (Elberse & Anand, 2007), post-release advertising (Renhoff & Wilbur, 2011), film critics (Basuroy *et al.*, 2003), a film's rating and genre (Ravid & Basuroy, 2004), online reviews (Duan *et al.*, 2008; Yeung *et al.*, 2011), release date (Chiou, 2008), word-of-mouth (Liu, 2006; Moul, 2007), movie stars (Elberse, 2007) and Academy Award nominations or wins (Nelson *et al.*, 2001).

Much of the work in the area of modelling cinematic box-office has often involved deploying stochastic methods or statistical techniques to research these questions (Hand, 2002; Ishii *et al.*, 2012; Ravid, 1999; Terry *et al.*, 2005; Walls, 2005a,b,c). In contrast, our research has involved creating and testing deterministic models using systems of differential equations which dynamically describe the financial performance of motion pictures released in North America (Edwards & Buckmire, 2001). Our primary research goal is to be able to produce a deterministic algorithm which can be used to predict the final domestic gross of a motion picture prior to (or as soon as possible after) the film's release. In Edwards & Buckmire (2001), the mathematical description for each film depended on a multitude of parameters that were movie-dependent (and somewhat difficult to estimate accurately): favourability rating, reviews, marketing and advertising budget, etc. In this work, we reduce the number of parameters that have to be estimated for a particular movie down to just one. The other parameters in the model are taken to be independent of an individual film, and hence can be determined *a priori*.

The problem of accurately predicting a movie's final gross mathematically has long been studied (Litman, 1983), and more recently has been attacked using a variety of different modern approaches (Hennig-Thurau *et al.*, 2007; Ishii *et al.*, 2012; Lee & Chang, 2009; Sawhney & Eliashberg, 1996; Sharda & Delen, 2006; Sochay, 1994). We readily acknowledge the difficulty (some would say futility) of the elusive goal of *a priori* box-office prediction owing to the seminal work of (De Vany & Walls, 2002), who argued that the box-office gross of a film has essentially infinite standard deviation. Regardless, we believe the goal of producing a mathematical model that uses differential equations which can accurately describe the dynamics of the box-office receipts of a motion picture is a worthy one, and so we return to this topic in this paper, presenting a mathematical model which this time includes geographic effects.

The structure of the paper is as follows. In Section 2, we provide the specific details of the development of our new mathematical model for the cinematic box-office dynamics of a film released in North America. We do this by introducing the concept of a region availability function which is geography-based. In Section 3, we consider the case where the region availability function is constant and in Section 4 we consider the polynomial case. In Section 5, we consider a special case of the region availability function that would correspond to a planned economy (equal number of screens spread uniformly across the nation). In Section 6, we validate our model by providing a computational algorithm for estimating the parameter values needed to compare our mathematical model to real-world box-office data. In Section 7, we provide some brief remarks as to how the model may be used by practitioners in the real world. In the last section, we provide conclusions and suggestions for further research.

2. Governing equations

In this section of the paper, we present the equations which describe our new geography-based mathematical model of cinematic box-office dynamics.

TABLE 1 *Nomenclature. Units are listed in terms of dollars (\$), people (N), time (T) or sites (S). With the exception of $\tilde{\mu}$, if the same symbol appears both with and without tildes, the letter with a tilde has dimensions, while the symbol without a tilde is dimensionless. The equation where a quantity first appears is listed, if appropriate*

Notation	Description	Equation
Variables and parameters		
$\tilde{A}(\tilde{t})$	Amount of money earned per week at time \tilde{t} , units $\$/ (ST)$	(2.1)
$\tilde{D}(\tilde{t})$	Demand per week to see film at time \tilde{t} , units N/T	(2.5)
$\tilde{G}(\tilde{t})$	Gross earnings of film at time \tilde{t} , units $\$$	(2.1)
P	Average ticket price, units $\$/N$	(2.6)
r	Rate relating revenue and change in sites, units $S^2/\$$	(2.3)
$\tilde{S}(\tilde{t})$	Number of sites on which film is showing at time \tilde{t} , units S	(2.1)
T	Number of geographic regions under consideration	
\tilde{t}	Time from opening day, units T	(2.1)
$\tilde{\alpha}$	Demand decay rate, units T^{-1}	(2.5)
β	Exponent in power-law form for $\tilde{\mu}$	(4.1a)
$\tilde{\kappa}$	Constant relating number of sites to change in sites, units $\$/S^2T$	(2.3)
$\tilde{\mu}(\tilde{S})$	Region availability function	(2.6)
τ	Dimensionless time from contract expiration date	(3.3)
Other notation		
c	As a subscript, used to indicate the contract period	(2.2)
max	As a subscript, denotes the maximum possible value	(2.4)
s	As a subscript, used to indicate the beginning of the saturation period	(3.6a)
t	As a subscript, used to indicate the beginning of the transition period	(4.6b)
0	As a subscript, used to indicate an initial state	(2.2)
*	As a subscript, used to indicate the transition period	(3.5)
-	Used to indicate a scaled dependent variable for the Matlab fit	(6.5)

Let $\tilde{G}(\tilde{t})$ be the gross of a film at time \tilde{t} (measured in days; for a full list of variables, see Table 1). It must evolve according to the following equation:

$$\frac{d\tilde{G}}{d\tilde{t}} = \tilde{S}\tilde{A}, \quad \tilde{G}(0) = 0, \tag{2.1}$$

where $\tilde{S}(\tilde{t})$ is the number of sites at which the movie is playing and $\tilde{A}(\tilde{t})$ is the amount of money earned per site per week. (Note that though we have only daily measurements of these quantities ([Internet Movie Data Base, 2012](#)), we consider them as being spaced closely enough to use a continuous-time model.) The innocuously simple expression given in (2.1) can be thought of as the Fundamental Equation of cinematic box-office dynamics. As the readily available data calculates the quantities in (2.1) only for the entire country, we consider them to be cumulative totals that do not depend on exhibitor location.

Our primary goal is to mathematically describe (and thus predict) the behaviour of $\tilde{G}(\tilde{t})$, so we must understand how \tilde{S} and \tilde{A} evolve with time. Most screening agreements between exhibitors and distributors have a fixed *contract period* $0 \leq \tilde{t} \leq \tilde{t}_c$ when the number of sites \tilde{S} is fixed, which locks in

a guaranteed revenue stream for the distributors (Hanssen, 2000). For the purposes of this model, we assume that \tilde{t}_c is the same for all sites, so

$$\tilde{S}(\tilde{t}) = \tilde{S}_0, \quad 0 \leq \tilde{t} \leq \tilde{t}_c, \quad (2.2)$$

where \tilde{S}_0 is the initial number of sites.

After $\tilde{t} = \tilde{t}_c$, we propose that \tilde{S} will change based upon a cost–benefit analysis. We hypothesize that $d\tilde{S}/d\tilde{t}$ is proportional to the difference between the marginal revenue \tilde{A} from the film and the marginal screening costs, which for simplicity we assume that are proportional to \tilde{S} . Hence, we have

$$\frac{d\tilde{S}}{d\tilde{t}} = r(\tilde{A} - \tilde{\kappa}\tilde{S}), \quad \tilde{t} \geq \tilde{t}_c, \quad (2.3)$$

where r and $\tilde{\kappa}$ are positive constants. Note from (2.3) that the number of screens can actually increase if \tilde{A} is high enough. This result would reflect exhibitors trying to ‘cash in’ on revenue from a film which is so popular and profitable that typical exhibition contracts would be superseded.

Note that r and $\tilde{\kappa}$ should be independent of the particular film, since they simply relate to the business decision of increasing sites for *any* movie. Also, if A_{\max} is the revenue per site if every seat is filled, and S_{\max} is the total number of available sites, then $\tilde{\kappa}$ should be chosen such that

$$\tilde{\kappa} > \frac{A_{\max}}{S_{\max}}, \quad (2.4)$$

as no more sites can be added.

In any particular week, there is a pent-up demand rate $\tilde{D}(\tilde{t})$ of people who want to see the film. However, we assume that each week, a certain fraction of those wishing to see the movie either see it or decide they no longer wish to see it. Hence, the demand decays exponentially:

$$\frac{d\tilde{D}}{d\tilde{t}} = -\tilde{\alpha}\tilde{D}, \quad (2.5)$$

where $\tilde{\alpha}$ is a positive constant. Here (2.5) uses the simplest possible model to describe a very complicated concept. For instance, in the case of a sleeper film, it is possible that the demand could increase over time with positive word-of-mouth (Liu, 2006; Moul, 2007) or post-release advertising (Renhoff & Wilbur, 2011). We may best model this case with very small values of $\tilde{\alpha}$, as shown in Section 5.

In order to close the system, we need to relate the attendance rate to the demand. We use a *geography-based model*. In particular, we break the total geographical area under consideration into T regions. We assume that each region has the same number of moviegoers, and that moviegoers in that area must attend theatres in that region. Hence, if there is no site in a region screening a particular film, the segment of the population in that region cannot see the film.

If the movie is not accessible to moviegoers in every region, the actual attendance rate will be less than \tilde{D} . Therefore, we have the following:

$$\frac{d\tilde{G}}{d\tilde{t}} = P\tilde{\mu}(\tilde{S})\tilde{D}, \quad (2.6)$$

where P is the average price of a ticket (to turn people-demand into revenue) and $\tilde{\mu}(\tilde{S})$ is the *region availability function* illustrating what fraction of the population can see the film. $\tilde{\mu}(\tilde{S})$ must have the following properties:

- (1) If no sites are showing the movie, no one can attend, so

$$\tilde{\mu}(0) = 0. \tag{2.7}$$

- (2) If the movie is showing only at one site, then the fraction of the population that can see that film is $1/T$ (by the geographic assumption). Hence,

$$\tilde{\mu}(1) = T^{-1}. \tag{2.8}$$

- (3) Everyone who wants to see the film can do so if it is playing at all the sites, so

$$\tilde{\mu}(S_{\max}) = 1. \tag{2.9}$$

This seemingly obvious fact holds only if (as we assume) the maximum number of sites has enough capacity to accommodate the maximum demand for any particular film.

- (4) Suppose that an exhibitor shows the movie at one more site. If the site is in a region that already has the movie, $\tilde{\mu}$ is unchanged:

$$\tilde{\mu}(\tilde{S} + 1) = \tilde{\mu}(\tilde{S}). \tag{2.10a}$$

If the site is in a region that does not have the movie, one additional region has the movie available, so

$$\tilde{\mu}(\tilde{S} + 1) = \tilde{\mu}(\tilde{S}) + \frac{1}{T}. \tag{2.10b}$$

If we approximate $d\tilde{\mu}/d\tilde{S}$ by $\tilde{\mu}(\tilde{S} + 1) - \tilde{\mu}(\tilde{S})$, then (2.10) may be approximated by

$$0 \leq \frac{d\tilde{\mu}}{d\tilde{S}} \leq T^{-1}, \quad \tilde{S} \geq 1, \tag{2.11}$$

so $\tilde{\mu}$ is monotone non-decreasing. The restriction that $\tilde{S} \geq 1$ in (2.11) (which allows for easier modelling later on) reflects the fact that the regime $0 \leq \tilde{S} \leq 1$ is not realistic.

Note that with these assumptions, $\tilde{\alpha}$ will be the only movie-dependent parameter. To complete the system, we must have an initial condition for \tilde{D} ; we equate (2.1) and (2.6) to obtain

$$\tilde{D}(0) = \frac{\tilde{S}_0 \tilde{A}_0}{P \tilde{\mu}(\tilde{S}_0)}, \tag{2.12}$$

where \tilde{A}_0 is the initial rate of money earned. We also note that \tilde{D} will also be cumulative across the country to conform with the measurable data; hence, any variations it might have across the country based upon the number of screens available are integrated away.

One can obviously quibble with some of the assumptions above as being distinctly different from the actual commerce taking place. For instance, the assumption of isolated regions works better for sparsely populated areas or limited-release films. However, the object of the model we present here is not to translate as complicated a system as the motion picture exhibitor–distributor market into mathematics

exactly. Instead, the goal is to present a model that will (re)produce the box-office dynamics which lead to a prediction of the final box-office gross of any film. Hence, the key question is whether the behaviour of the model system described above is close enough to the behaviour of the true commercial system to match the actual data (and eventually, to provide predictive power).

As is typical in the presentation of mathematical models, we now introduce scalings to simplify the model (Logan, 2006). Since our characteristic values should be independent of any particular movie, we use the maximum values as scalings:

$$\tilde{S}(\tilde{t}) = S_{\max}S(t), \quad \tilde{A}(\tilde{t}) = A_{\max}A(t), \quad \tilde{D}(\tilde{t}) = \frac{S_{\max}A_{\max}}{P}D(t), \quad (2.13a)$$

$$\tilde{t} = \frac{t}{r\tilde{\kappa}}, \quad \tilde{G}(\tilde{t}) = \frac{S_{\max}A_{\max}}{r\tilde{\kappa}}G(t), \quad (2.13b)$$

keeping in mind that $\tilde{\mu}(S_{\max}) = 1$.

Substituting (2.13) into (2.1–2.3), (2.5), (2.6) and (2.12), we obtain the following:

$$\frac{dG}{dt} = SA, \quad G(0) = 0, \quad (2.14)$$

$$S(t) = S_0, \quad 0 \leq t \leq t_c, \quad (2.15a)$$

$$\frac{dS}{dt} = \frac{A}{\kappa} - S, \quad \kappa = \frac{\tilde{\kappa}S_{\max}}{A_{\max}}, \quad t \geq t_c, \quad (2.15b)$$

$$SA = \mu(S)D, \quad (2.16)$$

$$\frac{dD}{dt} = -\alpha D, \quad \alpha = \frac{\tilde{\alpha}}{r\tilde{\kappa}}, \quad D(0) = D_0 \equiv \frac{S_0A_0}{\mu(S_0)}, \quad (2.17)$$

where $\mu(S)$ is the normalized availability function. Here (and throughout), the scaling on any variable with a subscript is the same as the variable without a subscript (e.g. $t_c = r\tilde{\kappa}\tilde{t}_c$). Equations (2.14–2.17) represent the clearest statement of the system of non-linear, coupled ordinary differential equations that we consider our geography-based mathematical model of cinematic box-office dynamics.

Some other forms of the equations in the above model will prove useful in later calculations. Combining (2.14) and (2.16), we have

$$\frac{dG}{dt} = \mu(S)D. \quad (2.18)$$

Alternatively, we may combine (2.15b) and (2.16) and simplify to obtain

$$e^{2t}\mu(S)D = \frac{\kappa}{2} \frac{d(e^{2t}S^2)}{dt}, \quad t > t_c. \quad (2.19)$$

3. Keeping the region availability function constant

In this section of the paper, we consider different versions of our model, using various assumptions about the geographic nature of the model. This is done by altering the functional form of μ , the region availability function introduced and defined in the previous section.

We first consider the two periods when the region availability function will be constant. These are the contract period $0 \leq t \leq t_c$, and any period for which S is so large that $\mu(S) = 1$. (This second scenario we refer to as *saturation periods*.)

3.1 The contract period

The solution of (2.17) is given by

$$D(t) = D_0 e^{-\alpha t}, \quad 0 \leq t \leq t_c. \tag{3.1}$$

During the contract period $0 \leq t \leq t_c$, $S \equiv S_0$ and hence $\mu \equiv \mu(S_0)$. Using this result along with (2.15a) and (3.1) in (2.18), we have

$$G(t) = \frac{S_0 A_0 (1 - e^{-\alpha t})}{\alpha}, \quad 0 \leq t \leq t_c. \tag{3.2}$$

Since the interval for the contract period is fixed, it is convenient to introduce the new variable $\tau = t - t_c$, so the initial value problem for the remaining part of the screening period reduces to standard form. Substituting these forms into (3.1), we obtain

$$D(\tau) = D_0 e^{-\alpha(\tau+t_c)} = D_c e^{-\alpha\tau}, \quad D_c = D_0 e^{-\alpha t_c}. \tag{3.3}$$

The differential operators from the previous section remain unchanged under the replacement of t by τ ; only the initial conditions are different. The initial condition for G becomes

$$G(\tau = 0) = G(t = t_c) = G_c \equiv \frac{S_0 A_0 (1 - e^{-\alpha t_c})}{\alpha}, \tag{3.4a}$$

while the initial condition for S becomes

$$S(\tau = 0) = S_0. \tag{3.4b}$$

3.2 The saturation period

Now consider the form of $\tilde{\mu}$ for large \tilde{S} . We know from (2.9) that $\tilde{\mu}(S_{\max}) = 1$. In order for $\tilde{\mu}(S_{\max} - 1) < 1$, two highly restrictive conditions must be satisfied:

- (1) There is only one region not running the film and
- (2) That region must have only one site (the only one in the country not running the film).

Hence, for most functional forms of μ , there exists an $S_* < S_{\max}$ such that

$$\mu(S) = 1 \quad \text{for } S \geq S_*. \tag{3.5}$$

Any period of time for which $S \geq S_*$ (and hence the movie appears in every region in the country) is called the saturation period. We denote any other intervals of time as *transition periods*.

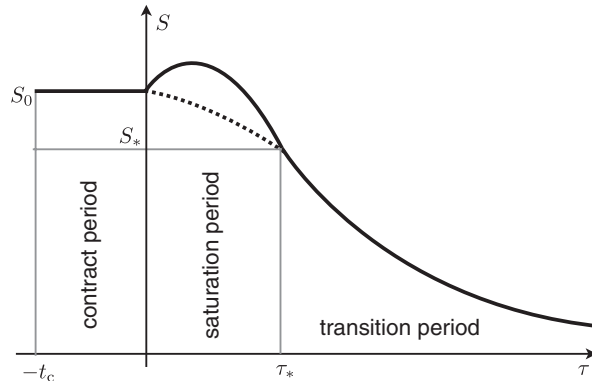


FIG. 1. Schematic of a wide-release movie ($\tau_s = 0$, $S_s = S_0$). Depending upon initial attendance, the number of sites can go up (solid) or down (dashed) during the saturation period, as shown.

To denote the interval of the saturation period, we let τ_s be the time at which S enters the saturation period and its value be defined as

$$S(\tau_s) = S_s. \quad (3.6a)$$

Moreover, let τ_* be the time at which S leaves the saturation period, so

$$S(\tau_*) = S_s, \quad \frac{dS}{d\tau}(\tau_*) < 0. \quad (3.6b)$$

With these definitions, there are three cases to consider:

Wide-release film. A film is said to be in *wide release* if $S_0 > S_*$. Hence, the film is in the saturation period directly after the contract period before moving to the transition period, as illustrated in Fig. 1. In this case, $S_s = S_0$ and $\tau_s = 0$. In common parlance, these kinds of films are often referred to as *blockbusters* or *event films*. Some examples of recent wide-release films are *The Dark Knight Rises* (2012) and *Inception* (2011).

Limited-release film. A film is said to be in *limited release* if $S_0 < S_*$, as shown in Fig. 2. In this case, the number of sites never saturates, so the film goes directly from the contract to the transition period. In common parlance, these are films that are often released in a small number of markets, usually big cities, often to ‘build buzz’ or qualify for Academy Award consideration at the end of the year. Some examples of recent limited-release films are *The Artist* (2011), *The Iron Lady* (2011) and *Lincoln* (2012).

Sleeper film. A *sleeper* is a limited-release film (so $S_0 < S_*$) with such strong attendance that there are two transition periods, as shown in Fig. 3. Hence, $S_s = S_*$ and $\tau_s > 0$. A sleeper film is a limited-release film which goes on to become a box-office hit, i.e. earn a substantial final gross. An example from our list above of recent limited-release films that is a sleeper would be *Lincoln*.

We conclude this section by deriving the exact solutions in the saturation region. Since (3.3) still holds, we have from (2.18) written in the τ variable that

$$G(\tau) = G_s + \frac{D_c(e^{-\alpha\tau_s} - e^{-\alpha\tau})}{\alpha}, \quad \tau_s \leq \tau \leq \tau_*, \quad (3.7a)$$

$$G_s = G(\tau_s). \quad (3.7b)$$

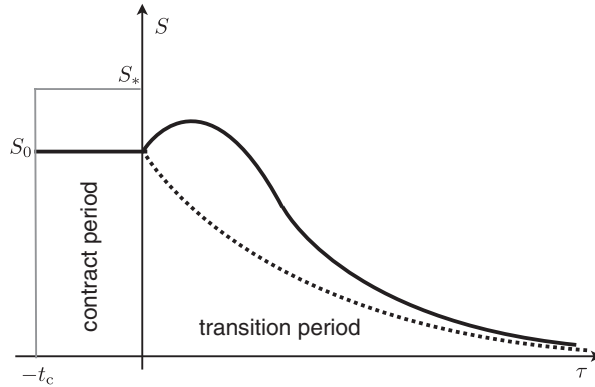


FIG. 2. Schematic of a limited-release movie ($S_0 < S_*$). Depending upon initial attendance, the number of sites can go up (solid) or down (dashed) during the transition period, as shown.

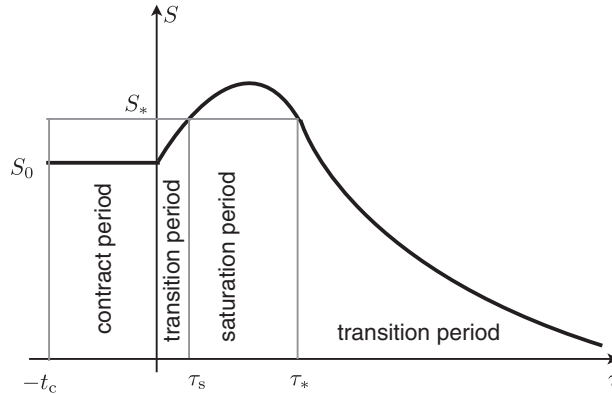


FIG. 3. Schematic of a sleeper film ($S_0 < S_*$, $\tau_s > 0$, $S_s = S_*$). The initial performance is so strong that the number of sites is increased above the saturation level.

Solving (2.19) written in the τ variable in this region subject to (3.6a), we obtain the following:

$$D_c e^{(2-\alpha)\tau} = \frac{\kappa}{2} \frac{d(e^{2\tau} S^2)}{d\tau}, \tag{3.8a}$$

$$S(\tau) = e^{-\tau} \left\{ e^{2\tau_s} S_s^2 + \frac{2D_c}{(2-\alpha)\kappa} [e^{(2-\alpha)\tau} - e^{(2-\alpha)\tau_s}] \right\}^{1/2}, \quad \tau_s \leq \tau \leq \tau_*, \tag{3.8b}$$

where τ_* is defined by letting $S(\tau) = S_*$ in (3.8b) and solving for τ .

4. Considering a polynomial form of the region availability function

In this section of the paper, we consider the implications provided by the mathematical model if we consider a functional form for the region availability function which is polynomial in nature.

To model $\tilde{\mu}(\tilde{S})$ explicitly, we use a polynomial form:

$$\tilde{\mu}(\tilde{S}) = \begin{cases} \left(\frac{\tilde{S}}{\tilde{S}_*}\right)^\beta, & \tilde{S} \leq \tilde{S}_*, \\ 1, & \tilde{S} \geq \tilde{S}_*. \end{cases} \quad (4.1a)$$

Note that this form satisfies the condition (2.7) on $\tilde{\mu}$ directly, and (2.8) is satisfied if \tilde{S}_* is defined by

$$\tilde{S}_* = T^{1/\beta}. \quad (4.1b)$$

Hence, rewriting (4.1) in dimensionless form, we have

$$\mu(S) = \begin{cases} \left(\frac{S}{S_*}\right)^\beta, & S \leq S_*, \quad S_* = \frac{T^{1/\beta}}{S_{\max}}, \\ 1, & S \geq S_*. \end{cases} \quad (4.2)$$

Satisfying (2.9) and (2.11) establishes the following bound on β :

$$\frac{\log T}{\log S_{\max}} \leq \beta \leq 1. \quad (4.3)$$

In order to plot solutions, we must have values for the dimensional parameters. The most recent value for S_{\max} is from [National Association of Theatre Owners \(2012b\)](#):

$$S_{\max} = 5561. \quad (4.4a)$$

Under our model assumptions, moviegoers in a certain region must attend theatres in that region. Therefore, to estimate the total number of regions T , it is useful to postulate a natural limiting distance beyond which moviegoers will not travel to see a film. We use a radius of 25 miles, and so, to obtain a rough estimate of T , we divide the area of the continental USA by the area of a circle with a 25-mile radius ([Wolfram|Alpha, 2012](#)):

$$T = \frac{3.12 \times 10^6 \text{ mi}^2}{\pi (25 \text{ mi})^2} = 1589. \quad (4.4b)$$

Although this estimate is necessarily crude, it agrees very closely with the actual film release data. If one examines the weekly data for the number of sites on which movies play, there is usually a clear delineation around 1500 screens between ‘wide-release’ and ‘limited-release’ films. With the values given in (4.4), the range in (4.3) corresponds to

$$8.55 \times 10^{-1} \leq \beta \leq 1. \quad (4.5)$$

A plot of the behaviour of $\mu(S)$ for various allowable values of β is shown in Fig. 4.

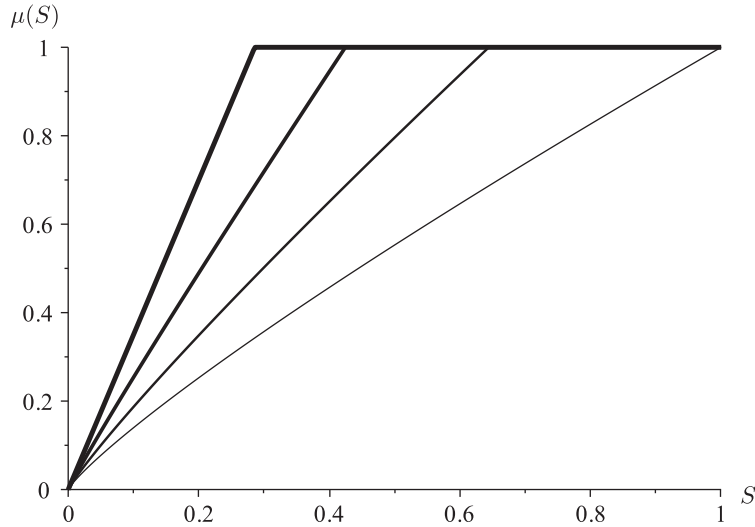


FIG. 4. $\mu(S)$ vs. S for (in increasing order of thickness) $\beta = 0.855, 0.901, 0.949, 1$.

The saturation period where $S \geq S_*$ has been discussed in Section 3.2. In the transition period where $S \leq S_*$, we may substitute (4.2) into (2.19) written in the τ variable to obtain

$$\frac{D_c e^{(2-\beta-\alpha)\tau}}{S_*^\beta} = \frac{\kappa}{2-\beta} \frac{d(e^{(2-\beta)\tau} S^{2-\beta})}{d\tau}. \tag{4.6a}$$

Recall that the special case $\beta = 2$ is not allowed by the condition on β given in (4.3). Also note that (3.8a) is a special case of (4.6a) with $\beta = 0$ and $S_* = 1$. We denote the initial condition for (4.6a) by

$$S(\tau_t) = S_t, \tag{4.6b}$$

where τ_t is the time at which the film enters the transition period. (These values may differ depending on the attendance pattern of the film; we will discuss the various values for τ_t and S_t at the end of this section.) Solving (4.6) yields

$$S(\tau) = e^{-\tau} \left\{ S_t^{2-\beta} e^{(2-\beta)\tau_t} + \frac{(2-\beta)D_c}{(2-\beta-\alpha)\kappa S_*^\beta} [e^{(2-\beta-\alpha)\tau} - e^{(2-\beta-\alpha)\tau_t}] \right\}^{1/(2-\beta)},$$

$$\tau \geq \tau_t, \quad \alpha \neq 2 - \beta. \tag{4.7}$$

Note that $S(\tau) \rightarrow 0$ as $\tau \rightarrow \infty$, as required. But there are two types of behaviour:

Screen decay dominates. If $2 - \beta - \alpha < 0$, then the braced term approaches a constant as $\tau \rightarrow \infty$, and the decay is like $e^{-\tau}$. The scaling was chosen as $r\tilde{\kappa}$, so it is the screen decay rate that dominates.

Attendance decay dominates. If $2 - \beta - \alpha > 0$, then the braced term diverges, and we have

$$S(\tau) \propto e^{-\tau} \{e^{(2-\beta-\alpha)\tau}\}^{1/(2-\beta)} = e^{-\alpha\tau/(2-\beta)}, \quad \tau \rightarrow \infty,$$

and the decay is governed by the attendance decay rate (reflected in α).

Substituting (4.7) into (4.2) and recalling that we are in the transition region where $S \leq S_*$, we obtain an expression for μ . Using this result in (2.18) rewritten in the τ variables, we obtain the following:

$$\frac{dG}{d\tau} = \frac{D_c e^{-(\alpha+\beta)\tau}}{S_*^\beta} \left\{ S_t^{2-\beta} e^{(2-\beta)\tau_t} + \frac{(2-\beta)D_c}{(2-\beta-\alpha)\kappa S_*^\beta} [e^{(2-\beta-\alpha)\tau} - e^{(2-\beta-\alpha)\tau_t}] \right\}^{\beta/(2-\beta)},$$

$$\tau \geq \tau_t, \quad (4.8)$$

where we have used (3.3). Equation (4.8) must be integrated numerically.

From the discussion in Section 3.2, we see that there are two possibilities for $\{\tau_t, S_t\}$.

Transition after saturation. Wide-release and sleeper movies have a transition period that occurs after a saturation period. Hence, we have

$$\tau_t = \tau_*, \quad S_t = S_*. \quad (4.9)$$

Substituting this result into (4.7) and (4.8) yields

$$S(\tau) = S_* e^{-\tau} \left\{ e^{(2-\beta)\tau_*} + \frac{(2-\beta)D_c}{(2-\beta-\alpha)\kappa S_*^2} [e^{(2-\beta-\alpha)\tau} - e^{(2-\beta-\alpha)\tau_*}] \right\}^{1/(2-\beta)}, \quad (4.10a)$$

$$\frac{dG}{d\tau} = D_c e^{-(\alpha+\beta)\tau} \left\{ e^{(2-\beta)\tau_*} + \frac{(2-\beta)D_c}{(2-\beta-\alpha)\kappa S_*^2} [e^{(2-\beta-\alpha)\tau} - e^{(2-\beta-\alpha)\tau_*}] \right\}^{\beta/(2-\beta)}, \quad \tau \geq \tau_*. \quad (4.10b)$$

Transition after contract. Limited-release and sleeper movies have a transition period that occurs right after the contract period with no saturation period. Hence, we have the following:

$$\tau_t = 0, \quad S_t = S_0. \quad (4.11)$$

Substituting this result into (4.7) and (4.8), we obtain

$$S(\tau) = e^{-\tau} \left\{ S_0^{2-\beta} + \frac{(2-\beta)D_c}{(2-\beta-\alpha)\kappa S_*^\beta} [e^{(2-\beta-\alpha)\tau} - 1] \right\}^{1/(2-\beta)}, \quad (4.12a)$$

$$\frac{dG}{d\tau} = \frac{D_c e^{-(\alpha+\beta)\tau}}{S_*^\beta} \left\{ S_0^{2-\beta} + \frac{(2-\beta)D_c}{(2-\beta-\alpha)\kappa S_*^\beta} [e^{(2-\beta-\alpha)\tau} - 1] \right\}^{\beta/(2-\beta)}, \quad \tau \geq 0. \quad (4.12b)$$

To illustrate the two cases of the decay rate, we specialize to the case of transition after contract governed by (4.11). We need a bound on $\tilde{\kappa}$ before plotting, for which we need an estimate for A_{\max} . We reason as in Edwards & Buckmire (2001). In particular, we assume that each site shows the film an average of 18 times per week with an average of 250 seats, regardless of the number of screens on which the film is showing. With an average ticket price of \$7.89 in 2010 (National Association of Theatre Owners, 2012a), we have

$$A_{\max} = 35505 \implies \tilde{\kappa} > 6.385, \quad (4.13)$$

where we have used (2.4) and (4.4a).

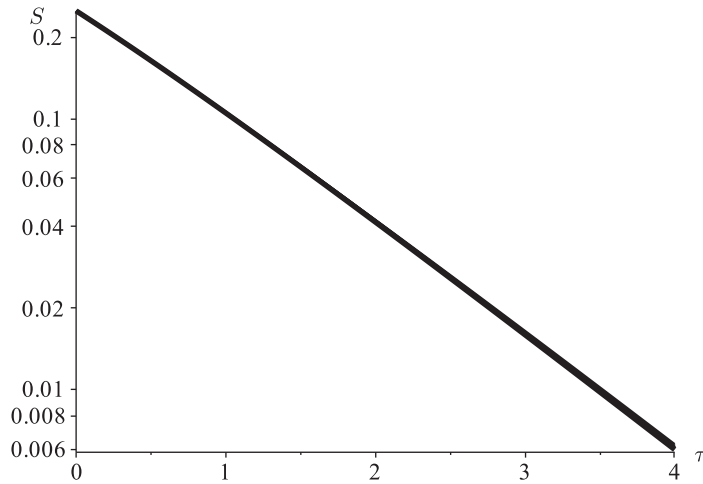


FIG. 5. S vs. τ (on log scale) for transition after contract. Parameters are $\alpha = 1.5$ and (in increasing order of thickness) $\beta = 0.855, 0.901, 0.949, 1$, so the screen decay rate dominates. The fact that the curves are indistinguishable indicates that the solutions' decay rate is the same in this regime.

In order that we are always in the case where $S_0 < S_*$, we take

$$S_0 = 0.25; \tag{4.14a}$$

we also take

$$A_0 = 0.8, \quad \kappa = 4, \tag{4.14b}$$

to ensure that $\dot{S} < 0$. (Note that κ has been scaled, so its lower bound is 1.) Finally, we take

$$t_c = 1, \tag{4.14c}$$

which allows us to determine D_c from (3.3).

We begin by looking at the case where the screen decay rate dominates. In this case, we take an α value large enough so that we always have $2 - \beta - \alpha < 0$:

$$\alpha = 1.5. \tag{4.15a}$$

The case is illustrated in Fig. 5. Note that the decay rate for all values of β is the same, illustrating that the solution decays like $e^{-\tau}$ for all choices of β .

Next we examine the case where the attendance decay dominates. In this case, we take an α value small enough so that we always have $2 - \beta - \alpha > 0$, namely

$$\alpha = 0.1. \tag{4.15b}$$

The case is illustrated in Fig. 6. Although the decay rate starts out the same (corresponding to the $e^{-\tau}$ behaviour, which dominates for small τ), for larger τ the decay rate for all the various cases is different. This illustrates that the solution decays like $e^{-\alpha\tau/(2-\beta)}$.

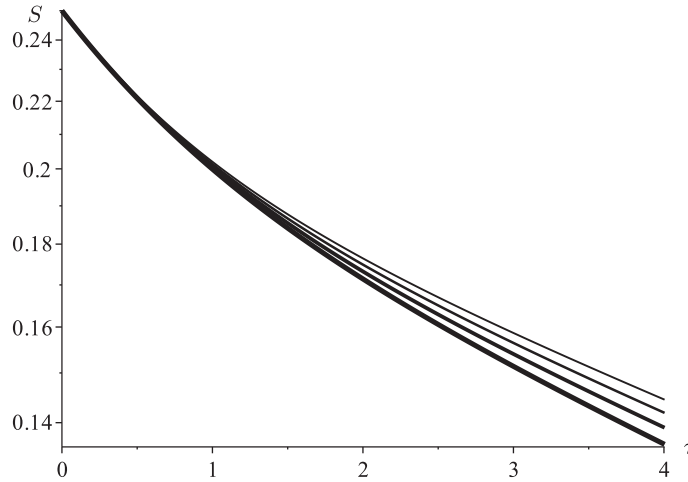


FIG. 6. S vs. τ (on log scale) for transition after contract. Parameters are $\alpha = 0.1$ and (in increasing order of thickness) $\beta = 0.855, 0.901, 0.949, 1$, so the attendance decay rate dominates.

5. Considering a planned economy

In this section of the paper, we consider a special case of the mathematical model which represents a *planned economy*. In this case, the geographic distribution of the film is such that no individual geographic region is allowed to have more than one site showing the film until every single region has the film showing at exactly one site. Clearly, this is an example of a rigidly controlled distribution and exhibition system reminiscent of a theoretically planned economy.

Figure 4 shows that $\mu(S; \beta = 1) \geq \mu(S; \beta \neq 1)$, and so from (2.18) we have that dG/dt is maximized in the transition period(s) for $\beta = 1$. Therefore, it is instructive to examine this special case. When $\beta = 1$, we are always in the situation described by (2.10b), so no region is allowed to have more than one screening site until all regions have one.

In the transition-after-saturation case governed by (4.9), the system (4.10) collapses to

$$S(\tau) = S_* e^{-\tau} \left\{ e^{\tau_*} + \frac{D_c}{(1-\alpha)\kappa S_*^2} [e^{(1-\alpha)\tau} - e^{(1-\alpha)\tau_*}] \right\}, \tag{5.1}$$

$$G = D_c \left\{ \frac{e^{-(\alpha+1)\tau_*} - e^{-(\alpha+1)\tau}}{\alpha + 1} \left[e^{\tau_*} - \frac{D_c}{(1-\alpha)\kappa S_*^2} e^{(1-\alpha)\tau_*} \right] + \frac{D_c(e^{-2\alpha\tau_*} - e^{-2\alpha\tau})}{2\alpha(1-\alpha)\kappa S_*^2} \right\} + G_*, \tag{5.2a}$$

$$G_* = G(\tau_*). \tag{5.2b}$$

Alternatively, in the transition-after-contract case governed by (4.11), the system (4.12) reduces to the following:

$$S(\tau) = e^{-\tau} \left\{ S_0 + \frac{D_c}{(1-\alpha)\kappa S_*} [e^{(1-\alpha)\tau} - 1] \right\}, \tag{5.3}$$

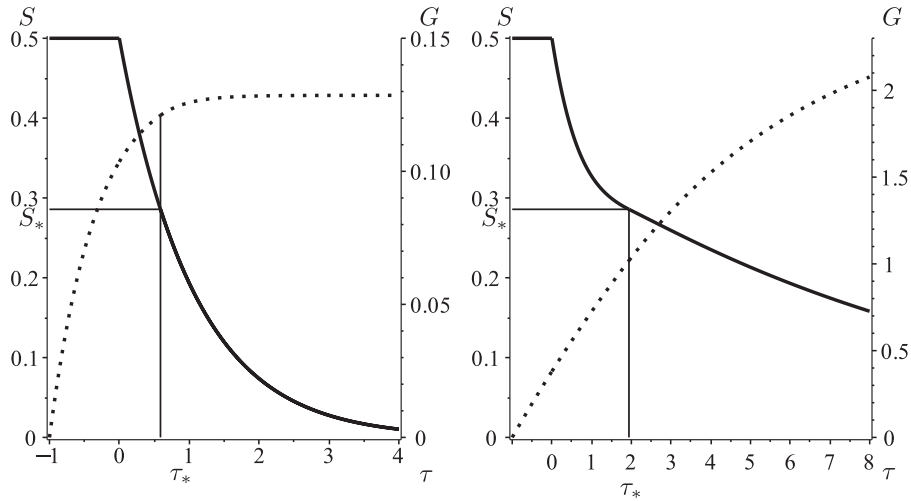


FIG. 7. S (solid line) and G (dotted line) vs. τ for a wide-release film. Left: $\alpha = 1.5$. Right: $\alpha = 0.1$.

$$G = \frac{D_c}{S_*} \left\{ \frac{1 - e^{-(\alpha+1)\tau}}{\alpha + 1} \left[S_0 - \frac{D_c}{(1 - \alpha)\kappa S_*} \right] + \frac{D_c(1 - e^{-2\alpha\tau})}{2\alpha(1 - \alpha)\kappa S_*} \right\} + G_c, \quad \tau \geq 0. \tag{5.4}$$

With $\beta = 1$, we see from (2.17) that

$$D_0 = A_0 \max\{S_0, S_*\}. \tag{5.5}$$

Using our parameters from the previous section, we have that

$$S_* = 2.86 \times 10^{-1}. \tag{5.6}$$

We begin by looking at the case of a wide-release movie, as shown in Fig. 7. In this case, we take

$$S_0 = 0.5 > S_*, \tag{5.7}$$

and use the parameters in (4.14b). Hence, we have $D_0 = 4$ from (5.5). For α , we show results for both values in Section 4: $\alpha = 1.5$, which corresponds to a rapid decay in demand, and $\alpha = 0.1$, which corresponds to a slower demand decay.

When the decay rate is large, most of the gross is made during the contract period ($-1 \leq \tau \leq 0$; see left graph). This is consistent with the fact that D decays exponentially with τ , and hence is highest during the initial contract period. Note that a small fraction of the gross is made during the saturation period $0 \leq \tau \leq \tau_*$ before the graph plateaus during the final transition period $\tau > \tau_*$.

In contrast, when the decay rate is small, the gross continues to increase throughout the run (see right graph). Note the larger scale in both τ (as larger demand causes the screens to decay more slowly) and G (as the larger demand causes a higher total gross).

Next we look at the case of a limited-release film. Here, we use the value of S_0 given by (4.14a), which implies that $D_0 = 0.223$. We know that there are two cases: either the screens continue to drop off (the regular case), or high demand forces the screens to temporarily increase (the sleeper case).

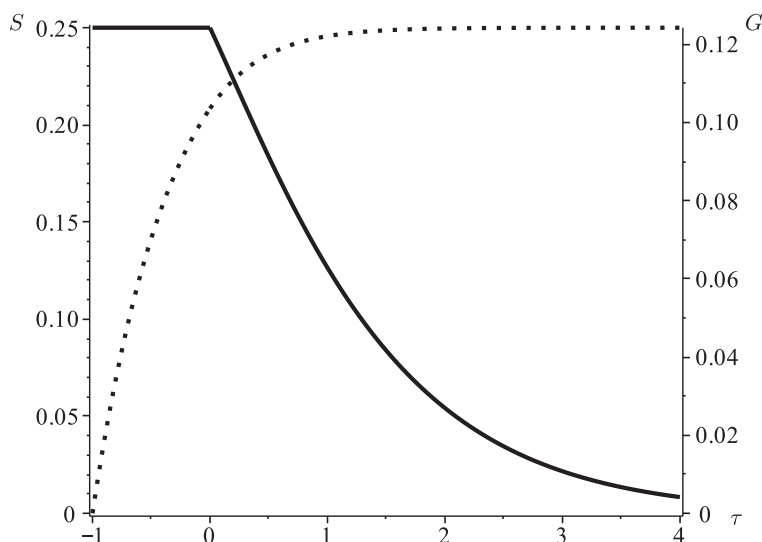


FIG. 8. S (solid line) and G (dotted line) vs. τ for a small-release film ($\alpha = 1.5$).

In order to ensure that attendance can indeed drive the screens higher, we take a lower value of κ , namely $\kappa = 1.5$.

For the non-sleeper film, we use the choice of α in (4.15a), which corresponds to steep fall-off of demand. The results are shown in Fig. 8. Unsurprisingly, we see that most of the gross is made during the contract period, with only a small fraction made during the transition period $\tau > 0$.

Lastly, we look at the case of a sleeper film, as shown in Fig. 9. Here, we use the value of α in (4.15b) corresponding to slow decay in the demand. In this case, most of the film's gross is made during the saturation period ($\tau_s \leq \tau \leq \tau_*$), though the gross continues to accumulate after that.

6. Validating the model

In this section of the paper, we present numerical simulations of the mathematical model. To validate the model, we must compare its results against real-world data. Thus, here we provide a straightforward algorithm for estimating the various parameters in our system from actual box-office data. This algorithm is used merely to validate the model by testing it on the data from a set of movies that has already been released. Instructions on how to use the model in practice are given in Section 7.

We use the computer program Matlab to actually perform the necessary computations.

6.1 Equations to fit

At first glance, it may seem easier to work with the dimensional data reported online (cf. [Internet Movie Data Base, 2012](#)). Hence, we first analyse the unscaled equations. Substituting the dimensional form of (3.1) into (2.6), we obtain

$$\frac{d\tilde{G}}{d\tilde{t}} = \frac{\tilde{\mu}(\tilde{S})}{\tilde{\mu}(S_0)} \tilde{S}_0 \tilde{A}_0 e^{-\tilde{\alpha}\tilde{t}}. \quad (6.1)$$

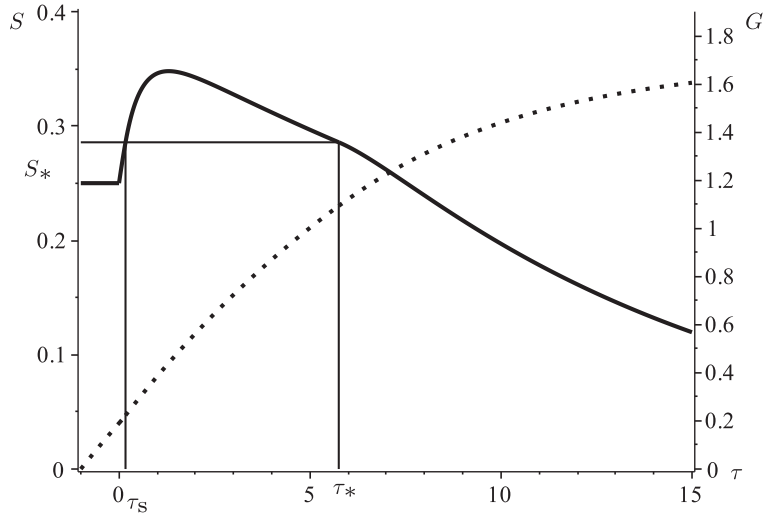


FIG. 9. S (solid line) and G (dotted line) vs. τ for a small-release film ($\alpha = 0.1$).

But from (6.1) evaluated at $\tilde{t} = 0$, we have

$$\frac{d\tilde{G}}{d\tilde{t}}(0) = \tilde{S}_0 \tilde{A}_0. \tag{6.2}$$

Indeed, this is how attendance rates are calculated, since the true data are given in terms of \tilde{G} and \tilde{S} . Hence, in order to calculate \tilde{A}_0 , we estimate $d\tilde{G}/d\tilde{t}(0)$ from the data using the simplest forward difference formula. Thus, we have

$$\frac{d\tilde{G}}{d\tilde{t}} = 7 \frac{\tilde{\mu}(\tilde{S})}{\tilde{\mu}(\tilde{S}_0)} \tilde{G}(1 \text{ day}) e^{-\tilde{a}\tilde{t}}, \tag{6.3}$$

where the coefficient 7 converts the rate to a weekly measurement, and we have used the fact that $\tilde{G}(0) = 0$.

To use Matlab, we also need an equation for $d\tilde{S}/d\tilde{t}$. By combining (2.1) and (2.3), we obtain the following:

$$\frac{d\tilde{S}}{d\tilde{t}} = \frac{r}{\tilde{S}} \frac{d\tilde{G}}{d\tilde{t}} - r\tilde{\kappa}\tilde{S}, \quad \tilde{t} \geq \tilde{t}_c. \tag{6.4}$$

Equation (6.4) is particularly convenient to implement as it contains the previously computed $d\tilde{G}/d\tilde{t}$. For $\tilde{t} < \tilde{t}_c$, we just set the right-hand side of (6.4) equal to zero.

However, using the raw data does not work well. When curve fitting, Matlab minimizes the fit error across *all* measured quantities. This is a problem because \tilde{G} and \tilde{S} typically have values that differ by several orders of magnitude. The values of \tilde{G} can be $O(10^8)$, while $\tilde{S} = O(10^3)$; hence during the optimization any (absolute) errors in the \tilde{S} fit would be swamped by errors in the \tilde{G} fit. Thus, the raw-data fits for \tilde{S} are poor.

Hence, we introduce scaled dependent variables as follows:

$$\bar{G} = \frac{\tilde{G}}{\tilde{G}(\tilde{t}_{\max})}, \quad \bar{S} = \frac{\tilde{S}}{\tilde{S}_0}, \quad (6.5)$$

where \tilde{t}_{\max} corresponds to the closing day of the movie. With these choices, both dependent variables typically vary between 0 and 1, and in the exceptional case of a sleeper where \bar{S} can be greater than 1, the variables will remain at the same order of magnitude. Hence, errors in both \bar{G} and \bar{S} will be of the same order in the minimization.

Obviously the scaling in (6.5) requires *a priori* knowledge of the final gross, and as such is useful only for model validation. A similar scaling suitable for estimates of a movie still in release is presented in Section 7. Note that in contrast to our theoretical work, it is acceptable to choose movie-dependent parameters to scale the dependent variables since it is unnecessary to normalize the time scale.

Substituting (6.5) into (6.4), we obtain

$$\frac{d\bar{S}}{d\tilde{t}} = r \left[\frac{1}{\bar{S}} \frac{\tilde{G}(\tilde{t}_{\max})}{\tilde{S}_0^2} \frac{d\tilde{G}}{d\tilde{t}} - \tilde{\kappa} \bar{S} \right], \quad \tilde{t} \geq \tilde{t}_c. \quad (6.6)$$

Substituting (6.5) into (6.3) and (4.1a), we have

$$\frac{d\tilde{G}}{d\tilde{t}} = 7 \frac{\tilde{\mu}(\tilde{S})}{\tilde{\mu}(1)} \tilde{G} (1 \text{ day}) e^{-\tilde{\alpha}\tilde{t}}, \quad \tilde{\mu}(\tilde{S}) = \min \left\{ \frac{(\tilde{S}_0 \tilde{S})^\beta}{T}, 1 \right\}. \quad (6.7)$$

6.2 Four-parameter fit

With the proper equations derived, we now focus on testing our model. We choose 10 successful films as a baseline test set, and fit the data for them (collected from [Internet Movie Data Base, 2012](#)) using Matlab's `lsqcurvefit` command. (It should be noted that the number of movies selected is perfectly arbitrary; we have available data on more movies but feel that the 10 selected are enough to illustrate the most important aspects of the model.) Initially, we fit all four parameters $\{\tilde{\alpha}, \tilde{\kappa}, r, \beta\}$ separately for each film. The results are shown in Roman type in Tables 2 and 3. We note from Table 2 that with one exception (namely *Harry Potter 7*), the values for $\tilde{\kappa}$ are very close to the minimum value specified in (4.13). In the case of *Harry Potter 7*, the number of screens increased in a particular week, which is not the normal case.

Fitting the raw data is difficult due to the fact that \tilde{S} will normally stay the same for an entire week (7 data points) before dropping down discontinuously to the next level. This discontinuous behaviour cannot be captured well by the model. Therefore, as a next step we used the seven-day rolling average of the data for parameter fitting. These results are shown in italics in Tables 2 and 3. This smoothed the discontinuities and improved the fit for *Harry Potter 7*, as shown in Fig. 10. Note that the improvement extended beyond that attributable just to the data smoothing (in particular, see the better fit of the final number of screens in Fig. 10).

The key test of any model that purports to mathematically describe cinematic box-office dynamics should be how well it estimates the final gross of the film, which is equivalent to determining how close $\tilde{G}(\tilde{t}_{\max})$ is to 1. Note from Table 3 that our estimates of the final gross are always within 5% of the true value, and that with two exceptions the averaged data provides a more accurate estimate.

TABLE 2 *Estimated parameter values using (6.7). Estimates from raw data are in Roman; estimates from averaged data are in italics*

Title	$\tilde{\alpha}$		$\tilde{\kappa}$		r	
<i>Yogi Bear</i>	0.2692	<i>0.2763</i>	6.3850	<i>6.3851</i>	0.0362	<i>0.0326</i>
<i>The Karate Kid</i>	0.7756	<i>0.7649</i>	6.3846	<i>6.3846</i>	0.0575	<i>0.0459</i>
<i>Toy Story 3</i>	0.7224	<i>0.7225</i>	6.3847	<i>6.3846</i>	0.0330	<i>0.0297</i>
<i>Alice in Wonderland</i>	0.8860	<i>0.8794</i>	6.3851	<i>6.3846</i>	0.0371	<i>0.0322</i>
<i>Iron Man 2</i>	1.1914	<i>1.1907</i>	6.4673	<i>6.3847</i>	0.0373	<i>0.0339</i>
<i>Twilight Saga: Eclipse</i>	1.6477	<i>1.6472</i>	6.3846	<i>6.3847</i>	0.0431	<i>0.0382</i>
<i>Harry Potter 7</i>	1.5203	<i>1.5202</i>	148.8178	<i>6.3846</i>	0.0018	<i>0.0374</i>
<i>Inception</i>	0.5280	<i>0.5279</i>	6.3846	<i>6.3846</i>	0.0321	<i>0.0293</i>
<i>Shrek Forever After</i>	0.6033	<i>0.6043</i>	6.3850	<i>6.3851</i>	0.0628	<i>0.0529</i>
<i>How To Train Your Dragon</i>	0.3782	<i>0.3788</i>	6.3861	<i>6.3848</i>	0.0295	<i>0.0267</i>

TABLE 3 *Estimated parameter values Using (6.7). Estimates from raw data are in Roman; estimates from averaged data are in italics*

Title	β		$\bar{G}(\tilde{t}_{\max})$	
<i>Yogi Bear</i>	0.9059	<i>0.9095</i>	0.9912	<i>0.9933</i>
<i>The Karate Kid</i>	0.9528	<i>0.8547</i>	0.9659	<i>0.9625</i>
<i>Toy Story 3</i>	0.8547	<i>0.8547</i>	0.9536	<i>0.9550</i>
<i>Alice in Wonderland</i>	0.9681	<i>0.8547</i>	0.9687	<i>0.9668</i>
<i>Iron Man 2</i>	0.9590	<i>0.9664</i>	0.9643	<i>0.9649</i>
<i>Twilight Saga: Eclipse</i>	1.0000	<i>1.0000</i>	0.9688	<i>0.9691</i>
<i>Harry Potter 7</i>	1.0000	<i>0.9989</i>	0.9628	<i>0.9629</i>
<i>Inception</i>	1.0000	<i>0.9958</i>	0.9832	<i>0.9841</i>
<i>Shrek Forever After</i>	0.9232	<i>0.9169</i>	0.9861	<i>0.9873</i>
<i>How To Train Your Dragon</i>	0.9274	<i>0.9241</i>	0.9851	<i>0.9863</i>

6.3 One-parameter fit

Recall from our model that $\tilde{\kappa}$, r and β should be the same for each movie. The averaged data estimates for $\tilde{\kappa}$ italicized in Table 2 are very nearly the same, and the values of r vary in a narrow range.

We use this information to calculate a single set $\{\tilde{\kappa}, r, \beta\}$ of movie-independent parameters by averaging the values in Tables 2 and 3, yielding

$$\tilde{\kappa} = 6.3847, \quad r = 0.0359, \quad \beta = 0.9276. \tag{6.8}$$

We then fit only $\tilde{\alpha}$ using the parameters in (6.8) for *all* the films; the results are shown in Table 4.

Note that the difference between the two estimates for $\tilde{\alpha}$ is very small (within 0.2). The starkest difference is for *The Karate Kid*, the results for which are shown in Fig. 11. Note that the change in $\tilde{\alpha}$ leads to only a slight improvement in our results overall (most noticeable for moderate t), but that the estimate for $\bar{G}(\tilde{t}_{\max})$ is actually slightly worse. In fact, using a one-parameter fit actually *improved* our estimate of $\bar{G}(\tilde{t}_{\max})$ in 6 of the 10 cases. In the four that were worse, both estimates were within 1% of each other.

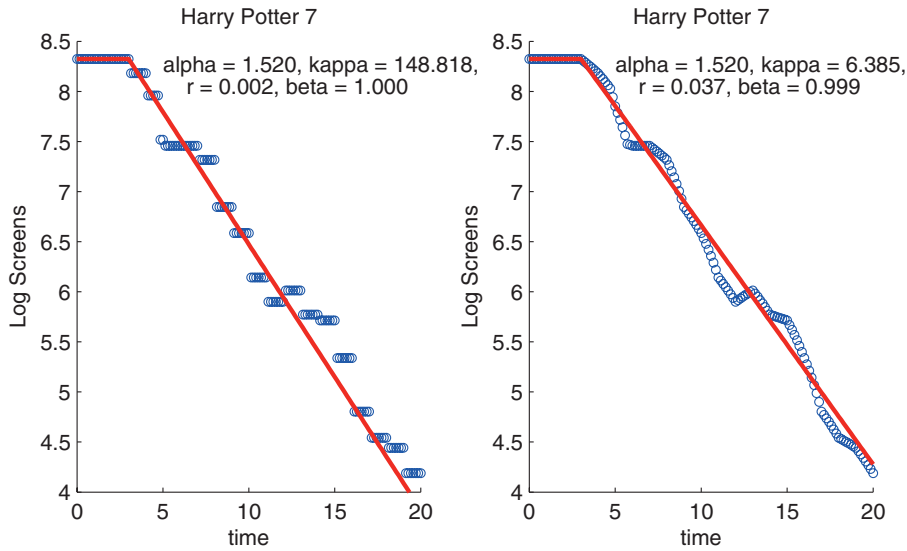


FIG. 10. Screens data for *Harry Potter 7* using (6.7). On left: raw data (circles) and fit (curve). On right: averaged data (circles) and fit (curve). Note the jump in screens around $t = 12$.

TABLE 4 Estimated values of $\tilde{\alpha}$ and $\bar{G}(\tilde{t}_{\max})$ using (6.7)

Title	Four-parameter estimate		One-parameter estimate	
	$\tilde{\alpha}$	$\bar{G}(\tilde{t}_{\max})$	$\tilde{\alpha}$	$\bar{G}(\tilde{t}_{\max})$
<i>Yogi Bear</i>	0.2763	0.9933	0.2845	0.9965
<i>The Karate Kid</i>	0.7649	0.9625	0.7828	0.9590
<i>Toy Story 3</i>	0.7225	0.9550	0.7234	0.9652
<i>Alice in Wonderland</i>	0.8794	0.9668	0.8809	0.9718
<i>Iron Man 2</i>	1.1907	0.9649	1.1882	0.9661
<i>Twilight Saga: Eclipse</i>	1.6472	0.9691	1.6478	0.9687
<i>Harry Potter 7</i>	1.5202	0.9629	1.5203	0.9627
<i>Inception</i>	0.5279	0.9841	0.5124	0.9892
<i>Shrek Forever After</i>	0.6043	0.9873	0.6174	0.9797
<i>How To Train Your Dragon</i>	0.3788	0.9863	0.3657	1.0013

6.4 Other ways to estimate

There are other ways to estimate $d\bar{G}/d\tilde{t}(0)$ besides using the first day's value, as in (6.7). For instance, one could use the gross after 1 week:

$$\frac{d\bar{G}}{d\tilde{t}} = \frac{\bar{\mu}(\bar{S})}{\bar{\mu}(1)} \bar{G}(\text{day } 7) e^{-\tilde{\alpha}\tilde{t}}. \tag{6.9}$$

One could also use multiple data points for \bar{G} with appropriate weighting, but the consideration of such schemes is beyond the scope of this manuscript.

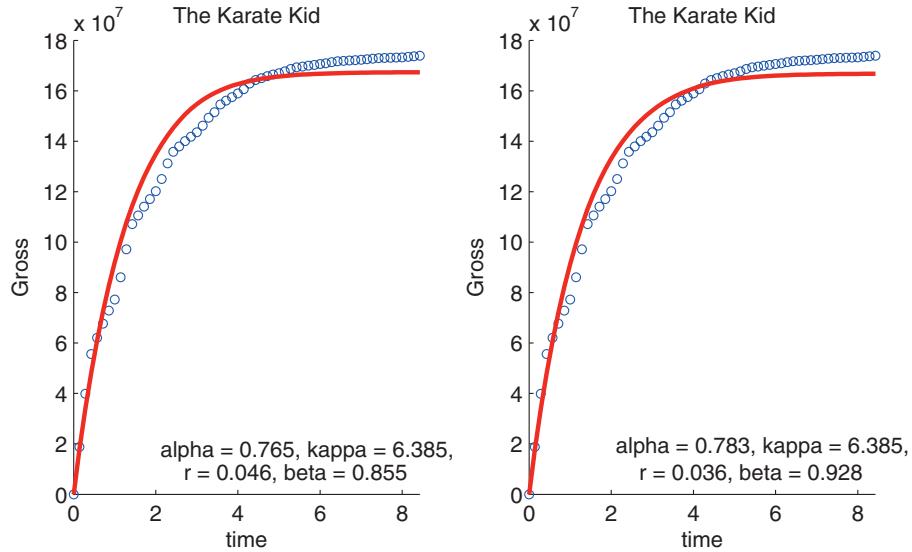


FIG. 11. Grosses data for *The Karate Kid* using (6.7). On left: fit from four-parameter estimation. On right: fit from one-parameter estimation.

TABLE 5 Estimated values of $\tilde{G}(\tilde{t}_{\max})$

Title	Estimates using (6.9)		
	Four-parameter	One-parameter	Exponential fit
<i>Yogi Bear</i>	1.0071	1.0145	0.9969
<i>The Karate Kid</i>	1.0199	1.0135	0.9960
<i>Toy Story 3</i>	0.9994	1.0133	0.9872
<i>Alice in Wonderland</i>	1.0178	1.0237	0.9911
<i>Iron Man 2</i>	1.0027	1.0064	0.9865
<i>Twilight Saga: Eclipse</i>	1.0006	1.0013	0.9880
<i>Harry Potter 7</i>	0.9905	0.9922	0.9806
<i>Inception</i>	0.9991	1.0116	0.9947
<i>Shrek Forever After</i>	1.0100	1.0024	0.9964
<i>How To Train Your Dragon</i>	1.0077	1.0336	1.0011

If we use (6.9) instead of (6.7), some calculation shows that the averages of the parameters are given by

$$\tilde{\kappa} = 7.4713, \quad r = 0.0392, \quad \beta = 0.8923. \tag{6.10}$$

Note the similarity of r and β to the values given in (6.8). The differing value of $\tilde{\kappa}$ is the result of an outlying value of $\tilde{\kappa}$ estimated using (6.9) for *Shrek Forever After*; if that outlier were removed, the values in (6.10) would be much closer to those in (6.8). However, since *a priori* one may not know which movies are the outliers, we did not remove it.

We show the final estimates for $\tilde{G}(\tilde{t}_{\max})$ using (6.9) in Table 5. As one might expect theoretically, using a one-parameter fit made our estimates worse in all but three cases. In the worst case, the estimate

was more than 2.5% worse. (However, note that the estimates for $\bar{G}(\tilde{t}_{\max})$ were almost all better than the one using the daily fit.)

As a final aside, in Table 5 we present the estimated value of $\bar{G}(\tilde{t}_{\max})$ using a simple two-parameter exponential fit:

$$\bar{G}(\tilde{t}) = G_1(1 - e^{-G_2\tilde{t}}).$$

Our one-parameter fit results are better for four of the films, while our four-parameter fit results are better for five of the films. This demonstrates that the model presented here is at least as good as a simple exponential fit, but our model has the advantage of providing information on the dynamics of the number of exhibition sites in addition to the final box-office gross.

7. Using the model

Our results from the previous section indicate that our model can be quite useful in estimating the final gross of films. In this section, we discuss how our model can be implemented as a practical matter to help in the estimates of grosses for movies that are still showing.

As discussed above, the parameters come in two flavours: movie-independent ($\tilde{\kappa}$, r , β) and movie-dependent ($\tilde{\alpha}$). From data freely available for previously released films, one can run simulations as in Section 6 to determine the movie-independent parameters. In order to increase the granularity of the estimates, it may be useful to isolate the films into genres (note all the films used in this paper were blockbusters). This would allow examination of the possibility that these parameters vary by genre (due to studio or exhibitor decisions or contracts, perhaps).

Once these parameters have been determined *a priori*, the next step is to estimate $\tilde{\alpha}$ (and hence the gross) for a particular film. This can be done repeatedly as new daily data arrive. However, given the close agreement to the final gross estimate by using either (6.7) or (6.9), we expect that estimates with even a small number of data points should be reasonably accurate. Such investigations are beyond the scope of this manuscript.

One final technical detail pertains to the normalization in (6.5). As $\tilde{G}(\tilde{t}_{\max})$ is not known for a movie currently being shown, the exact scaling in (6.5) will not work. However, recall that the purpose of (6.5) is simply to make errors in the fits of \tilde{S} and \tilde{G} roughly the same size. This can easily be done by normalizing the data by the median final gross of movies of the same genre.

8. Conclusions and further research

The filmmaking business is inherently a risky one (De Vany, 2004; Moul, 2005). The budget for film production is spent before any revenue begins to flow, and the product (a ticket) is one which the bulk of the customers will purchase only once, after even more money has been spent letting them know when the product will become available for consumption. In order to minimize the amount of promotional expenditures and maximize profits, distributors would like to predict a film's eventual final gross as early as possible during the film's exhibition lifetime. Due to the required contract period, an exhibitor or distributor will always have several weeks' worth of data to analyse for each film; however, a more desirable goal is to construct a model which can be used to estimate the gross as early as possible.

Rather than using a stochastic-based approach (Ishii *et al.*, 2012; Terry *et al.*, 2005; Walls, 2005a,b,c), we use a deterministic-based ordinary differential equation approach. Given that screen and gross data are available on a daily basis, we feel confident in our use of a continuous-time model as a reasonable approximation.

We have presented a geography-based, mathematical model for analysing film grosses. Any global area under study is divided into T smaller regions, and it is assumed that moviegoers within each region cannot travel to another region to see a film. Hence, if a movie is not screening within their region, they cannot view it. In this manuscript, we estimated T using a distance-based analysis, though one could just as easily use data on total population, population density or facts about the initial release data on films.

Although some of our simplifying assumptions may seem unrealistic, our results show that the results from the simplified model closely track real-life commerce, both in fitting the exhibition and gross curves, as well as predicting the final gross amount.

The geographic facet of the model is illustrated in (2.6). Equation (2.1) follows directly from the definition of the quantities \tilde{A} , \tilde{S} and \tilde{G} . Equation (2.3) models the effect of decreased attendance on the distributor's decision to change the number of sites. These equations contain the three movie-independent parameters to be fit: r , $\tilde{\kappa}$ and β . We used Matlab to fit these parameters, both separately for each film and for the entire data set as a whole. Using Matlab to fit the movie-independent parameters separately for each film produced only small changes. This lends credence to our contention that these are movie-independent, though our sample was small. This can continue to be verified with analyses of larger sets of movie releases, an area of further research.

In our model, the only movie-dependent parameter is the decay rate $\tilde{\alpha}$ of demand. In our analysis, we found that the value of $\tilde{\alpha}$ did vary between films, but did not change significantly when we fixed the movie-dependent parameters. With the movie-independent parameters fixed, the estimate for $\tilde{\alpha}$ can be continually refined each day as new data trickles in. These refined estimates should produce more accurate estimates of the final gross. This is yet another project available for future research.

Acknowledgements

Any opinion, finding and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Funding

R.B. acknowledges that this work was based on work supported by the National Science Foundation, while working at the Foundation. Support for J.O.-G. for this project was provided by a fellowship from a Howard Hughes Medical Institute Undergraduate Science Education Grant.

REFERENCES

- BASUROY, S., CHATTERJEE, S. & RAVID, S. A. (2003) How critical are critical reviews? The box office effects of film critics, star power, and budgets. *J. Mark.*, **67**, 103–117.
- CHIOU, L. (2008) The timing of movie releases: evidence from the home video industry. *Int. J. Ind. Org.*, **26**, 1059–1073.
- CHISHOLM, D. (2011) The economics of motion pictures. *Handbook of Cultural Economics*, 2nd edn (R. Towse ed.). Northampton, MA: Edward Elgar.
- DE VANY, A. S. (2004) *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. London: Routledge.
- DE VANY, A. S. (2006) The movies. *Handbook of the Economics of Art and Culture*, vol. 1 (V. Ginsburgh & D. Throsby eds). Oxford: Elsevier, pp. 615–665.
- DE VANY, A. S. & ECKERT, R. (1991) Motion picture antitrust: the paramount cases revisited. *Res. Law Econ.*, **14**, 51–112.

- DE VANY, A. S. & WALLS, W. D. (1996) Bose–Einstein dynamics and adaptive contracting in the motion picture industry. *Econ. J.*, **106**, 1493–1514.
- DE VANY, A. S. & WALLS, W. D. (2002) Does Hollywood make too many R-rated movies? Risk, stochastic dominance, and the illusion of expectation. *J. Bus.*, **75**, 425–451.
- DE VANY, A. S. & WALLS, W. D. (2004) Motion picture profit, the stable Paretian hypothesis, and the curse of the superstar. *J. Econ. Dyn. Control*, **28**, 1035–1057.
- DUAN, W., GU, B. & WHINSTON, A. B. (2008) Do online reviews matter? An empirical investigation of panel data. *Decis. Support Syst.*, **45**, 1007–1016.
- EDWARDS, D. A. & BUCKMIRE, R. (2001) A differential equation model of North American cinematic box-office dynamics. *IMA J. Man. Math.*, **12**, 41–74.
- ELBERSE, A. (2007) The power of stars: do star actors drive the success of movies? *J. Mark.*, **71**, 102–120.
- ELBERSE, A. & ANAND, B. (2007) The effectiveness of pre-release advertising for motion pictures: an empirical investigation using a simulated market. *Inf. Econ. Policy*, **19**, 319–343.
- ELIASHBERG, J., ELBERSE, A. & LEENDERS, M. A. A. M. (2006) The motion picture industry: critical issues in practice, current research, and new research directions. *Mark. Sci.*, **25**, 638–661.
- HAND, C. (2002) The distribution and predictability of cinema admissions. *J. Cult. Econ.*, **26**, 53–64.
- HANSSSEN, C. (2000) The block booking of films re-examined. *J. Law Econ.*, **43**, 395–426.
- HENNIG-THURAU, T., HOUSTON, M. B. & WALSH, G. J. (2007) Determinants of motion picture box office and profitability: an interrelationship approach. *Rev. Manag. Sci.*, **1**, 65–92.
- INTERNET MOVIE DATA BASE (2012) *Box Office Mojo*. Available at <http://www.boxofficemojo.com/>.
- ISHII, A., ARAKI, H., MATSUDA, N., UMEMURA, S., URUSHIDANI, T., YAMAGATA, N. & YOSHIDA, N. (2012) The ‘hit’ phenomenon: a mathematical model of human dynamics interactions as a stochastic process. *New J. Phys.*, **14**, 063018.
- LEE, K. J. & CHANG, W. (2009) Bayesian belief network for box-office performance: a case study on Korean movies. *Expert Syst. Appl.*, **36**, 280–291.
- LITMAN, B. R. (1983) Predicting success of theatrical movies: an empirical study. *J. Pop. Cult.*, **16**, 159–175.
- LIU, Y. (2006) Word of mouth for movies: its dynamics and impact on box office revenue. *J. Mark.*, **70**, 74–89.
- LOGAN, J. D. (2006) *Applied Mathematics*, 3rd edn. New York: Wiley.
- McKENZIE, J. (2012) The economics of movies: a literature survey. *J. Econ. Surv.*, **26**, 42–70.
- MOUL, C. (2005) *A Concise Handbook of Movie Industry Economics*. New York: Cambridge University Press.
- MOUL, C. (2007) Measuring word-of-mouth’s impact on theatrical movie admissions. *J. Econ. Manag. Strat.*, **16**, 859–892.
- NATIONAL ASSOCIATION OF THEATRE OWNERS (2012a) NATO statistics: *Average US ticket prices*. Available at <http://www.natonline.org/statisticstickets.htm>.
- NATIONAL ASSOCIATION OF THEATRE OWNERS (2012b) *NATO Statistics: US Cinema Sites*. Available at <http://www.natonline.org/statisticssites.htm>.
- NELSON, R., DONIHUE, M., WALDMAN, D. & WHEATON, C. (2001) What’s an Oscar worth? *Econ. Inq.*, **39**, 1–16.
- RAVID, S. A. (1999) Information, blockbusters, and stars: a study of the film industry. *J. Bus.*, **72**, 463–492.
- RAVID, S. A. & BASUROY, S. (2004) Managerial objectives, the R-rating puzzle, and the production of violent films. *J. Bus.*, **77**, S155–S192.
- RENHOFF, A. & WILBUR, K. (2011) The effectiveness of post-release movie advertising. *Int. J. Advert.*, **30**, 305–328.
- SAWHNEY, M. S. & ELIASHBERG, J. (1996) A parsimonious model for forecasting gross box-office revenues of motion pictures. *Mark. Sci.*, **15**, 113–131.
- SHARDA, R. & DELEN, D. (2006) Predicting box-office success of motion pictures with neural networks. *Expert Syst. Appl.*, **30**, 243–254.
- SOCHAY, S. (1994) Predicting the performance of motion pictures. *J. Media Econ.*, **7**, 1–20.
- TERRY, N., BUTLER, M. & DE’ARMOND, D. (2005) The determinants of domestic box office performance in the motion picture industry. *SW Econ. Rev.*, **32**, 137–148.
- WALLS, D. (2005a) Demand stochastics, supply adaptation, and the distribution of film earnings. *Appl. Econ. Lett.*, **12**, 619–623.

- WALLS, D. (2005b) Modeling movie success when ‘nobody knows anything’: conditional stable-distribution analysis of film returns. *J. Cult. Econ.*, **29**, 177–190.
- WALLS, D. (2005c) Modelling heavy tails and skewness in film returns. *Appl. Financ. Econ.*, **15**, 1181–1188.
- WALLS, W. D. (2008) Economics of motion pictures. *The New Palgrave Dictionary of Economics*, vol. 5, 2nd edn (L. Blume & S. Durlauf eds). London: Palgrave MacMillan.
- WOLFRAM|ALPHA (2012) Wolphram|Alpha. Available at www.wolframalpha.com.
- YEUNG, C. H., CIMINI, G. & JIN, C. H. (2011) Dynamics of movie competition and popularity spreading in recommender systems. *Phys. Rev. E*, **83**.