INCREASING RETURNS TO INFORMATION AND PARETO'S LAW: NEW AUSTRALIAN BOX OFFICE REVENUE EVIDENCE

Jordi McKenzie*

May 7, 2007

Abstract

Recent empirical evidence has suggested that box office revenue distributions exhibit 'increasing returns to information'. This conclusion has been proposed as an explanation of the observed autocorrelation in box office revenue growth rates in the context of the famous Pareto law model that, among other things, has been used to model firm size. Presenting a similar theoretical model to that of Ijiri and Simon (1974), this paper tests the findings of De Vany and Walls (1996), Walls (1997) and Hand (2001) on box office revenue data using a large new data set of Australian box office revenues. Support is found for a violation of the log-linear Pareto model suggesting autocorrelation in revenue growth rates to be a feature of the 'weekly box office revenue' and 'cumulative box office revenue' distributions. This paper extends previous research by re-examining the empirical findings when the cumulative revenue distribution is disaggregated in terms of opening week number of screens and also by considering the model's suitability for describing the distribution of 'weekly screen average revenues'. It is observed that the autocorrelation is still evident leading to the conclusion that 'increasing returns to information' is a strong and general feature of demand for motion pictures.

Keywords: Box Office Revenue, Pareto's law, Increasing Returns to Information
Classification Numbers: C16, L00, Z11

* *Correspondence*: Jordi McKenzie, Discipline of Economics, Merewether Building H04,University of Sydney, NSW, Australia, 2006. Email: J.Mckenzie@econ.usyd.edu.au

1 Introduction

Recent empirical evidence on box-office revenue data has suggested that the word of mouth opinion shared between consumers may produce, what has become termed, 'increasing returns to information'.[1] The fact that the top ranking movies seem to earn the majority share of total revenue in any given week suggests a convex relation between ordinal ranking and revenue that was informally termed Murphy's law after Art Murphy, a *Variety* film magazine journalist, noted the trend in the revenue of films. Such an observation is similar to a 'winner take all' outcome and can be likened to a highly unequal income distribution that would be represented by a highly convex Lorenz curve. Formally, this result was first really noted by De Vany and Walls (1996) using data on the United States when they likened the empirical distribution to the well known Pareto distribution but with the presence of autocorrelated growth similar to the model of Ijiri and Simon (1974). They subsequently explained this finding of autocorrelation in the growth rates of box office revenue as support for the importance of information transmission among consumers.[2] Their model of information transmission was based on a Bayesian updating process that exhibits Bose-Einstein dynamics.[3] These dynamics are important as they a imply box office revenue distribution with remarkable uncertainty in outcomes – a critical observation that has implications for conduct, structure and performance of the industry.

[1] De Vany and Walls (1996).
[2] The process can also be viewed as an information cascade where individuals place relatively more weight on the information provided by a film's recent viewers. See, for example, Banerjee (1992), Bikhchandani, Hirshleifer and Welch (1992), Chamley and Gale (1994), Smith and Sorenson (2000) , De Vany and Lee (2001) and Sgroi (2002).
[3] The Bose-Einstein name originates in physics and relates to the Bose-Einstein condensate, a phase of matter formed by bosons cooled to temperatures very near to absolute zero, and is of interest for the statistical distribution of particles over energy states.

Since the initial contribution of De Vany and Walls (1996) – Walls (1997) and Hand (2001) have also supported these findings using data from Hong Kong and the United Kingdom respectively.[4] The initial motivations for this research are firstly, to provide an initial contribution to the literature utilising a new and unique Australian data set that is significantly larger than those used in the previous studies, and secondly, to test whether there is supportive evidence for 'increasing returns to information' being a feature of the Australian film industry. This paper also considers an extension to the previous research by considering films as different types in relation to the number of screens on which they are shown. Failure to account for films screening on differing number of screens imposes an obvious bias on ranking films performance at either a weekly or cumulative level. For example, to consider the world wide box office hit and the fourth Harry Potter film, *Goblet of Fire*, which opened on 552 screens, and the Australian film starring Jack Thompson, *Under the Light House*, which only opened on one screen, in the same sample creates an undesirable bias in either a weekly or cumulative sample. To better understand the role of information transmission, and to properly assess its importance on shaping the distribution, such a bias needs to be removed. This paper does this in two respects. Firstly, by disaggregating the cumulative revenue sample into four groups defined by the opening week number of screens, and secondly, by considering the ranked 'weekly *screen average* revenue' distribution, rather than the 'weekly revenue' distribution per se. This paper is organised as follows. Section 2 provides a brief background to the Pareto law, Section 3 describes a theoretical model that explains the role of autocorrelation in growth rates of box office revenue, Section 4 describes the

[4] Maddison (2004) has also found support for 'increasing returns to information' in Broadway theatre productions.

empirical method used to test the model, Section 5 describes the data and empirical results, and Section 6 provides a concluding discussion.

2 The Pareto law and Size Variance

The empirical methodology utilised in this paper involves a simple test of a scaling phenomena commonly known as *Pareto's* law. Pareto's law (1897) was originally utilised in economics as a description of income distributions but the scaling phenomena has also been observed in many other areas in natural and social sciences states the probability of occurrence of an event starts high and tapers off. [5] Thus, a few occur very often while many others occur rarely. Pareto's law has also been proposed as a model of firm size – as gauged by assets, sales revenues, number of employees, etc[6] – and the formal methodology of this research bears much resemblance to this particular interpretation. The Pareto law can be expressed by the following relationship

$$SR^{\beta}=A \tag{1}$$

where A and β are constants, S denotes size, and R denotes rank. The largest firm, for example, is rank 1 (and size A). Ijiri and Simon (1971) use the constant β as an indicator of the degree of concentration in a paper examining the effect of mergers and acquisitions on business firm concentration, because it indicates the frequencies of large firms relative to smaller firms. The greater the β, the greater the relative size of a large firm (small rank) compared with a smaller firm (large rank).

[5] Worthy of mention in relation to this much-documented phenomena include Benford (1938), Yule (1944), Zipf (1949), Simon (1955) and Mandelbrot (1963).
[6] Steindl (1965) found it applies to the relationship between firm size and rank for a number of industries.

The natural logarithmic transformation of (1) yields the following equation

$$\ln S = \ln A - \beta \ln R. \tag{2}$$

So on a log-log scale, β measures the slope of the straight line. This relationship was initially used by Steindl (1965) to describe the firm size relation, however it soon was noted that many actual observed firm size distributions displayed curvature that violated the implicit predicted size of (2).[7] This departure from a linear relation may be known as 'size variance'. The downward concavity shows an upward variance for middle ranked firms relative to small or large sized firms. The existence of the concavity can be directly seen and quantified with the introduction of a quadratic variable to equation (2) giving the following

$$\ln S = \ln A - \beta \ln R + \gamma(\ln R)^2 \tag{3}$$

where the coefficient γ depends on the curvature. The log-log distribution is convex downwards if γ is positive and concave downwards if γ is negative.

Ijiri and Simon (1967) provided initial empirical support for the existence of concavity in the firm size distribution and noted that "autocorrelation" between recent growth and future growth may be the source of this concavity, a conjecture that was formalised in their 1974 article. They noted that, of the firms sampled in 1954-58 with above the average levels of growth had higher levels of growth carried onto the

[7] For example, see Ijiri and Simon (1967, 1971, 1974) and Vining (1976).

next period sampled 1958-62 for those firms. Such observations led some to question the validity of the Pareto curve as an appropriate description of the firm size distribution, however its use was subsequently defended by Ijiri and Simon (1974) who demonstrated how the introduction of autocorrelation in growth rates may explain observed divergences from predictions. They argued for the retention of Pareto law primarily on the grounds of its simplicity and moreover economically plausible assumptions regarding the size independence of percentage growth rates (Gilbrat's 1939 law of proportionate effect) and constancy of entry rate.[8] In defending the Pareto distribution as an appropriate model of the firm size relationship, Ijiri and Simon (1974) describe a model that may lead to a distribution that approximates the Pareto distribution but with curvature as observed in the log-log plot. Although not formally presented in their research, this model of autocorrelation forms the basis for conclusions and inference formed by De Vany and Walls (1996), Walls (1997), and Hand (2001) regarding box office revenue and an adapted version of Ijiri and Simon's (1974) model is now detailed.

3 A Model of Autocorrelated Growth in Film Revenue

Define the set of films in the population at time t as N_t. At time t each film has size (revenue) and weight defined $S_{i,t}$ and $W_{i,t}$ respectively for films $i = 1 \ldots N_t$. At some point in time ($t = 0$) there are several films in the population N_0, each having a unit size and weight ($S_{i,0} = 1$ and $W_{i,0} = 1$) for all films $i \in N_0$.[9] The weighting of each film determines its potential for revenue growth relative to that of the other films in the population. At each subsequent time period t = 1, 2, 3 etc, one unit of sales goes

[8] Ijiri and Simon (1974, p. 317).
[9] Placing several films in the population is the initialisation for the model, however, Ijiri and Simon (1964) show that after large t, the resulting distribution is insensitive to the number of firms (films) chosen.

to a film selected at random. Define $s_{i,t}$ as a binary variable that takes the value 1 if a unit of sale goes to film *i* at time *t* and 0 otherwise. Therefore, it follows that the size of film *i* at time *t* may be written

$$S_{i,t} = \sum_{k=0}^{t} s_{i,k} \ . \tag{4}$$

At time t =1, the number of films in the population, N_1, may become greater than, or remain equal to N_0, depending upon whether a new film has been selected or not. Similarly, at t = 2, N_2 may become greater than or equal to N_1, or may still retain the value N_0 if no unit sales have been added to new films in the first two rounds.

There is constant probability

$$\Pr(s_{j,t} = 1 \mid j \notin N_{t-1}) = \alpha \tag{5}$$

that the unit of sales at time *t* goes to a new film *j* which was previously not in the population of films in N_{t-1}. If this happens, the new film *j* is added to the population with size and weight equal to 1. If the sales unit at time *t* does not go to a new film, it goes to an old film. By logical argument from (5), the probability that the sale goes to one of the old films is (1-α). If it does go to an old film, the probability of a *given* old film receiving the sales unit is proportional to its weight from the previous period – that is,

$$\Pr(s_{i,t} = 1 \mid i \in N_{t-1}) = \frac{W_{i,t-1}}{\sum_{i \in N_{t-1}} W_{i,t-1}} \qquad \forall i \in N_{t-1} \tag{6}$$

implying that the probability of sale to an old film is equal to the film's weight divided by the sum of the weights in the population at time $t - 1$ for the population of films at N_{t-1}. The size (and weight) of the old film selected is subsequently increased by one unit in time t.

After a new film is created or an old film is selected among films in the population, the weight of each old film in the population is discounted by multiplying its current weight by δ. Then, if a new film is selected, its weight is set equal to one. If an old film is selected, its weight (after discounting) is increased by 1.[10]

This implies that the total weight of film i at time t can be written as

$$W_{i,t} = \sum_{k=0}^{t} \delta^{t-k} s_{i,k} \ . \tag{7}$$

Following Ijiri and Simon's (1974) model of autocorrelation in firm size growth, the model of autocorrelation in film revenue is introduced by specifying the variable

[10] Ijiri and Simon (1964) show, by simulation, that the firm size distribution exhibits downward concavity if $\delta < 1$, is a straight line if $\delta = 1$, and has downward convexity if $\delta > 1$. The curvature is greater as $|\delta - 1|$ is greater. The parameter δ has a simple economic interpretation and describes the relative significance of recent sales, as compared with earlier sales, in generating future sales. Both recent and prior sales contribute to future sales but the latter contributes less than the former if $\delta < 1$ and vice-versa. This is shown analytically in their (1974) paper for the case where $0 < \delta < 1$ (distribution concave downward).

$$W_t = \sum_{i \in N_t} W_{i,t} \tag{8}$$

as the sum of all weights of all films in the population after time t. It follows that,

$$W_t = \delta W_{t\text{-}1} + 1, \tag{9}$$

implying that the sum of all previous weights are discounted by δ and a weighting of 1 unit is added to the population from the new unit of sales. By substitution, this must therefore mean that for large t

$$W_t = 1 + \delta + \delta^2 + \ldots\ldots + N_0\, \delta^{t\text{-}1} \approx \frac{1}{1-\delta} \qquad (0< \delta <1). \tag{10}$$

For a film, j, that was created at time t, the size of the film, $S_{j,t}$, and its weight, $W_{i,t}$, are both equal to 1. It follows that the probability that this film will experience a sale at time t+1 is the probability that the sale goes to an old film (1-α) times the probability that this particular film experiences the sale. Using (6) and (10), this can be defined

$$\Pr(s_{j,t+1} = 1 \mid S_{j,t} = 1) = (1-\alpha)\frac{W_{j,t}}{\sum_{i \in N_t} W_{i,t}} = (1-\alpha)\frac{1}{W_t} = (1\text{-}\alpha)(1\text{-}\delta) \equiv \theta\,. \tag{11}$$

Defining the probability (1-α)(1-δ) from (11) as θ, implies that the probability film j doesn't get the sale at time t+1 is (1-θ). If the film didn't get the sale at t+1, the weighting is discounted by δ, implying that the probability of getting a sale at t+2, is $\theta\delta$ and the probability of not getting the sale is (1-$\theta\delta$). It follows that the probabilities of getting a sale at t+3 is $\theta\delta^2$ and not getting a sale is 1- $\theta\delta^2$. The probability that film j will never get a sale after time t is defined by the following

$$\Pr(s_{j,\forall T>t} = 0 \mid S_{j,t} = 1) = (1-\theta)(1-\theta\delta)(1-\theta\delta^2)(1-\theta\delta^3)\ldots\ldots \quad (12)$$

and implies that the film will always remain at size 1. From (11) it can be easily verified that for δ close to 1, θ is close to zero and so is $\theta\delta^t$ for low t. If δ is not close to 1, $\theta\delta^t$ becomes close to zero for high t.[11] Taking the logarithmic transformation of (12) yields the following equation

$$\ln\{\Pr(s_{j,\forall T>t} = 0 \mid S_{j,t} = 1)\} = \ln(1-\theta) + \ln(1-\theta\delta) + \ln(1-\theta\delta^2) + \ln(1-\theta\delta^3)\ldots\ldots \quad (13)$$

Assuming θ is small, the approximation $\ln(1+x) \approx x$ may be used to obtain

$$\ln\{\Pr(s_{j,\forall T>t} = 0 \mid S_{j,t} = 1)\} \approx -\theta(1+\delta+\delta^2+\delta^3+\ldots) = -\frac{\theta}{1-\delta} = -\frac{(1-\alpha)(1-\delta)}{1-\delta} \quad (14)$$

implying the following

$$\Pr(s_{j,\forall T>t} = 0 \mid S_{j,t} = 1) \approx e^{-(1-\alpha)}. \quad (15)$$

[11] Ijiri and Simon (1964) show in their simulation that a slight deviation of δ = 0.95 was sufficient to create considerable curvature for the distribution.

The implication of this result is that the asymptotic proportion of films having a size of 1 depends only on α (the probability of sale going to a new film) and is independent of δ.

Now assume a film j has weight $W_{j,t}*$ and size $S_{j,t}*$ at time t. Consideration of equation (8) reveals that the probability of no future sale for this film at t+1, can simply be defined as $\theta W_{j,t}*$, i.e the numerator of 1 from (11) is replaced by $W_{j,t}*$. Therefore, for the film that initially has weight $W_{j,t}*$, equation (15) can easily be shown to generalise to the expression

$$\Pr(s_{j,\forall T>t} = 0 \mid S_{j,t} = S_{j,t}*) = e^{-(1-\alpha)W^*_{j,t}} \quad . \tag{16}$$

Recalling equation (4), it can be seen that the derivative $dW_{j,t}*/d\delta > 0$ for any given history of a film, meaning that as the discount rate (δ<1) approaches unitary, the film specific weighting also increases. But, the probability a film doesn't receive any more sales is given by equation (16), implying $d\Pr(.)/dW_{j,t}* < 0$, and therefore it must follow that $d\Pr(.)/d\,\delta < 0$ as well.[12]

The implication of this result is that the probability of having further sales (1-Pr(.)) for old films becomes smaller as δ becomes smaller. Growth becomes more difficult the higher the discount factor. By consideration of equation (7) and (16) it is also apparent that as the size of the film, $S_{i,t}$, becomes larger, the effect of reducing δ on $W_{i,t}$ becomes more severe owing to the extra terms on the right hand side of the

[12] Ijiri and Simon (1974, p.321).

equation. This implies that it is not only the discount factor that affects the probability of future growth, but also the size of the film.

4 The Pareto Law and Box Office Revenue

Similar to initial observations of Ijiri and Simon (1974) regarding firm size data, De Vany and Walls (1996) noted deviations from the ‘theoretical’ size predicted by the straight line for box office revenues in the U.S. market. Their analysis begins by considering the natural logarithmic transformation and rearrangement of equation (1) and defining of weekly revenue as a measure of relative size

$$\ln(WKREV_{i,t}) = \alpha + \beta \ln(WKRNK_{i,t}) + \varepsilon_{i,t} \qquad (17)$$

where α is $\ln A$ and $WKREV_{i,t}$ and $WKRNK_{i,t}$ denote weekly box office revenues (i.e. size) and ordinal rank of film *i* in week *t*. The coefficient β is actually negative as this is implicit in the logarithmic transformation of equation (1). Logarithms are to the natural base *e* and $\varepsilon_{i,t}$ is the disturbance term with mean zero and finite variance.

Formally, De Vany and Walls (1996) showed the departure from the linear rank revenue relationship suggested in equation (17) with the inclusion of a quadratic variable, $\ln(WKRNK)^2$, that was found to have a significant coefficient. This finding has subsequently been explored and confirmed by Walls (1997) and Hand (2001) for Hong Kong and U.K. data respectively. The existence of non-linearity in the log-log plot of rank and size (revenue) signals the significance of the curvature of actual rank-revenue relation that has implications for the sharing of information between filmgoers and the autocorrelation of film growth rates theoretically presented in the previous section. To formally test for a departure from a linear Pareto rank revenue relation equation (17) is modified and the following equation is estimated

$$\ln(WKREV_{i,t}) = \alpha + \beta \ln(WKRNK_{i,t}) + \gamma(\ln(WKRNK_{i,t}))^2 + \varphi(DUWK_t) + \varepsilon_{i,t}\,. \qquad (18)$$

where $DUWK_t$ is a week specific dummy variable to remove seasonal effects. Finding that $\gamma \neq 0$ is evidence that the Pareto law does not hold in the linear sense and there may be 'autocorrelated growth of revenue' in the context proposed by Ijiri and Simon (1974) and found by De Vany and Walls (1996), Walls (1997) and Hand (2001) for motion picture revenues. The implication of such a finding is that film revenues may grow in a manner that is related to their *relative* performance. Finding of $\gamma > 0$ (convex downward) or $\gamma < 0$ (concave downward) may signify negative or positive autocorrelated growth respectively. The previous cited studies of box office data have rejected a linear Pareto rank revenue relation and found statistically negative estimates of γ. These results have been interpreted as evidence of 'increasing returns to information' based on the model of Ijiri and Simon (1974) outlined above. That is to say, a film that has enjoyed recent growth is more likely to grow faster than a film whose growth occurred further in the past.

De Vany and Walls (1996) also consider the distribution of cumulative box office revenues for their U.S. sample by considering the following regression that has much similarity to equation (18)

$$\ln(CUMREV_i) = \alpha + \beta \ln(CUMRNK_i) + \gamma(\ln(CUMRNK_i))^2 + \varepsilon_{i,} \qquad (19)$$

where $CUMREV_i$ and $CUMRNK_i$ refer to the 'cumulative revenue' and 'cumulative ranking' of film *i* respectively. The obvious difference between (18) and (19) being the loss of the time subscript as films are compared in respect to their total cumulative revenue and overall cumulative ranking in the sample – not their weekly rankings – and the loss of the weekly dummy variables. In their sample of 300 films De Vany and Walls also find a departure from the linear in logs specification, supporting the existence of autocorrelated growth rates in the cumulative sample.

A potential shortcoming of the previous studies cited and the regression equations (18) and (19) relates to the fact that films of differing release types are being compared with each other. In the cumulative model (19), this would mean that wide release films, such as *Harry Potter – Goblet of Fire* that opened on above 550 screens, are being compared with small independent titles that may only enjoy a single screen release.[13] Surely comparing such films in a single sample has the potential to bias results in a direction that would lead to a rejection of the simple linear model in favour of the quadratic specification. In this respect, the data set is disaggregated to consider films in terms of their opening number of screens. Four sub-samples are partitioned that include films opening on less than 50 screens, films opening on 50-149 screens, films opening on 150-249 screens, and those opening on 250+ screens. In the weekly model (18), a similar potential bias may also exist as implicitly films on differing numbers of screens are being compared. By considering 'weekly screen average' rather than solely 'weekly revenue' this avoids the potential bias and allows further insight to the level and degree of information sharing and the possible existence of autocorrelation in growth rates. Such assessment is unique to the

[13] In the current sample of 2,429 films there are 209 films that only opened on one screen.

present study and only permissible with the current Motion Picture Distributors Association of Australia (MPDAA) Australian data set. This is because most previous research that has originated in the U.S. and U.K has used Nielsen EDI or *Variety* figures that only include *theatre* (location) numbers and not *screen* (engagement) numbers. It is therefore impossible to get an exact screen count, and hence screen average, when it is very common for multiplex theatres to screen more than one print per theatre location for a wide release picture.[14]

To test for the existence of autocorrelation in the weekly screen average distribution a regression of the following form is considered

$$\ln(WKSCRNAVREV_{i,t}) = \alpha + \beta \ln(WKSCRNAVRNK_{i,t}) + \gamma(\ln(WKSCRNAVRNK_{i,t}))^2 + \varphi(DUWK_t) + \varepsilon_{i,t} \qquad (20)$$

where $WKSCRNAVREV_{i,t}$ and $WKSCRNAVRNK_{i,t}$ are 'weekly screen average revenue' and 'weekly screen average rank' respectively for film i in week t. Two regressions of this form are estimated – one for all films in the sample (16,472 observations) and one for films that opened on 50 or more screens and are classified as 'wide' release by the MPDAA (8,828 observations). This was done to avoid the upward bias on screen averages that is observed of films released on low number of screens and, in particular, to avoid the high screen averages associated with Large Screen Format (LSF) and IMAX films.[15]

[14] For example, the recent Australian block buster *Happy Feet* reported 3,804 theatres for its opening week in the US but its screen/print count was estimated to be in excess of 5,600.
[15] For example, the IMAX film *Everest* had the top opening week screen average of A$134k compared with the highest 'wide release' opening week screen average of A$46k attained by *Lord of the Rings – Two Towers* which would have ranked 23 in the full sample in terms of opening week screen average.

5 Data and Empirical Results

The MPDAA data utilised in this section includes the cumulative revenue and cumulative ranks of all films released at the Australian box office between January 1, 1997 and December 31, 2005 covering 2,429 distinct titles. The data set also includes weekly revenue, weekly screens and weekly rankings of films released between the April 17, 1997 and December 31, 2005 covering 16,472 observations.[16] Table 1 reports the top five highest grossing films of the sample where it can be seen that *Titanic* achieved the highest ranking in terms of cumulative revenue taking in excess of $57.6m at the Australian box office or 0.8% of total box office of the sample. The extent to which the ranked cumulative distribution exhibits convexity is apparent in Table 2, where it is observed that the top 10% of films take over 55% of total revenue and the top 20% take over 80% - a striking observation in support of the famous 80/20 principle.

[INSERT TABLE 1&2 NEAR HERE]

Summary statistics for 'cumulative revenue', 'weekly revenues' and 'weekly ranks' are provided in Table 3 where the suggestion of a heavily right skewed distribution with large variance is apparent for the 'cumulative revenue' and disaggregated 'cumulative revenue' samples and also at each of the 'weekly revenue' levels. The extremity of the skew is evident from the calculated mean consistently being significantly greater than the median. For example, the mean of the 'cumulative revenue' distribution is over $2.8m but the median is less than $0.6m. In terms of ranking it can be observed that, on average, each film debuted at a rank of 14.3 in its first week and this average steadily declined over each week of the run and by the tenth week the average film was at rank 27.4.

---

[16] The original form of the data necessitated censoring the first 15 weeks of 1997 for the weekly analysis.

[INSERT TABLE 3 NEAR HERE]

Table 4 reports similar information in relation to 'weekly screen averages' and 'weekly screen average ranks' for the full sample and for films defined as 'wide' release (opening on 50 or more screens).[17] It can similarly be observed that the mean does exceed the median in all weeks for both the full sample (All Films) and the sample of films that never played on more than 50 screens (Wide Release Films). It is apparent, however, that the deviation between the mean and median is by nowhere near as significant as it was at the weekly levels. The summary statistics in terms of rankings tell a similar story as before and it can be observed that, on average, films debut in the chart at rank 10.8 (4.32 for wide release films) – in terms of 'weekly screen average revenue – and fell away week by week until at week 10 they were at position 27.1 (15 for wide release films).

[INSERT TABLE 4 NEAR HERE]

The noted features of an extremely skewed distribution arise primarily because of the 'extreme events' in the sample, or the blockbuster type movies that drag the mean away from the median. This feature suggests a convex downward relation between films' ranks and their revenues for the cumulative revenue sample, the weekly revenue sample, and the weekly screen average distributions that can be seen in Figure 1. The logarithmic transformation of these variables allows a visual test of Pareto's law and it can be seen in Figure 2 that the transformation leads to an outcome suggestive of a non-linear relation between ln(*Revenue*) and ln(*Rank*) for each definition. These observations are now formally tested.

[INSERT FIGURES 1&2 NEAR HERE]

[17] Films were re-ranked according to their 'weekly screen average revenue' for both the full sample and 'wide release' sub-sample.

Equations (17) and (18) are estimated using OLS where $WKREV_{i,t}$ is taken as each film's weekly revenue in 1997 dollars[18] and $WKRNK_{i,t}$ is the weekly rank as reported by the MPDAA. These regression equations were estimated with and without the inclusion of 52 week dummy variables for the total sample comprising 16,472 observations. Weekly dummy variables were included to remove any seasonal effects and are estimated relative to the week number 53 dummy variable that was used twice in the sample to avoid losing days due to there being 365 days in a year.[19] Equation (19) was also estimated for cumulative revenue ($CUMREV_i$) and cumulative rank ($CUMRNK_i$) of the 2,429 distinct films and once again final box office was deflated to 1997 dollars. The results, reported in Table 5, show remarkable self-similarity between the regressions that utilised individual films' 'weekly revenues' and the regression that utilised 'cumulative revenue'.

[INSERT TABLE 5 NEAR HERE]

The results for both sets of regressions suggest that the rank-revenue relationship for films tends to depart from the linear Pareto distribution in a direction that may suggest autocorrelated growth as defined by Ijiri and Simon (1974) and has been noted by De Vany and Walls (1996), Walls (1997) and Hand (2001) for box office data. That is to say, the $(\ln(Rank))^2$ coefficient is observed to be significant and negative. In both specifications of the weekly model, inclusion of dummy variables only marginally changed results. The results are very similar to those obtained in the previous studies cited, as can be seen in Table 6, where the results of the regression utilising weekly revenues, weekly rank and weekly dummy variables are compared. The implications from these results suggest that the Australian box office revenue distribution is right

[18] Using annual Reserve Bank of Australia (RBA) inflation data the revenue dollars were deflated to January 1997 prices. Inflation increments were defined bi-annually.

[19] This allowed week 52 of the sample to always include Boxing Day (December 26) – the biggest single day of the year at the Australian box office.

skewed with a mean that is dominated by the blockbuster type movies. The top films each week (and overall) earn a disproportionate large share of total week (and overall) revenue and this leads to a downwardly convex relation between rank and revenue. The rejection of a linear in logs relation is rejected in favour of a quadratic specification that suggests a violation of the simple Pareto law.

[INSERT TABLE 6 NEAR HERE]

The models discussed so far, however, may suffer an implicit bias in their construction when one considers that films are marketed and released with different target audiences and, moreover, different (perhaps non-comparable) revenue earning potential and ability. Failure to treat films as different types therefore may cause a bias in favour of rejection of the simple Pareto law. An extension of this research to that of the studies cited previously is to attempt to address this by disaggregating the cumulative revenue data set in terms of release size (i.e. Opening week number of screens) and to also consider the Pareto model for the 'weekly screen average revenue' distribution, rather than just in relation to 'weekly revenue'. Again using deflated cumulative revenue data, equation (19) is estimated for films that opened on less than 50, 50-149, 150-249 and 250+ screens when films were re-ranked according to their performance within the relevant sub sample.[20] Table 7 provides the results of the disaggregated 'cumulative revenue' regressions and it can be clearly seen that in each release window considered there is still a notable violation of the linear Pareto model as the parameter γ is consistently estimated significantly negative. It is also of interest to note that the estimated constant and coefficient to the quadratic variable display some fluctuation. By examining the predicted values of these regressions (Figure 3) the reason for this becomes evident as the curvature imposed by the

[20] Not all films reported opening week number of screens, therefore 137 films were lost in the disaggregated analysis.

quadratic specification leads to some variation in the constant term – particularly for the 'Opening Screens <50' regression. Nonetheless, over the majority of the data the models appear to give a reasonable fit that is confirmed with the adjusted R-squared statistic.

[INSERT TABLE 7 NEAR HERE]

[INSERT FIGURE 3NEAR HERE]

The results of the 'weekly screen average revenue' model of equation (20) are presented in Table 8. Films were once again re-ranked according to their relative performances in terms of 'week screen average revenue' for both the full sample (16,472 observations) and the 'wide release' sub-sample (8,828 observations).[21] Estimations were carried out on the (1997 prices) deflated weekly revenue figure divided by the corresponding week's number of cinema screens. Such a variable obviously removes the bias of the 'weekly revenue' variable that could represent a film playing on 1 or 500 screens. Once again, the γ parameter was estimated significantly negative in both the full sample (All Films) and reduced sample (Wide Release Films) suggesting there is still sufficient curvature in the log Pareto model to warrant the inclusion of the quadratic term.[22] This implies that the theoretical model of autocorrelation is still highly valid when films revenues are considered in this way and provides stronger empirical support for the 'increasing returns to information' doctrine that has been posited for the industry.

[INSERT TABLE 8 NEAR HERE]

[21] Recall the 'wide release' sub sample was portioned to avoid the cluster of films that occupy low screen counts but have excessively high screen averages (e.g. IMAX films).

[22] Inclusion of the 52 week specific dummy variables only marginally changed the results and, again, these are all relative to the week 53 dummy variable used twice in the nine year sample to maintain consistency between week number and dates.

6 Conclusion

This paper has considered the 'increasing returns to information' model using a large new Australian data set. Reconsidering the model of Ijiri and Simon (1974) a theoretical model was presented to show how autocorrelation in the growth of revenue may impact on the simple (linear in logs) Pareto model. This research has also considered whether disaggregation of the 'cumulative revenue' sample by opening week release size had any bearing on the observed empirical results. It didn't. A similar question was also asked in relation to the 'weekly revenue' model by instead considering the Pareto law's appropriateness for the 'weekly *screen average* revenue' distribution. Again, it was impossible to retain the simple Pareto model and the model with the quadratic term was preferred. The best explanation for such a finding is that the growth of box office revenue may be autocorrelated in the sense suggested by De Vany and Walls (1996). This implies that a film that has experienced recent growth in revenue is more likely to experience growth than a film that experienced growth further back. This finding lends support to the idea of 'increasing returns to information', where consumers share information on films they like and this is leveraged into admissions and ensuing revenue. The evidence presented using this large new Australian data set suggests that 'increasing returns to information' is a general characteristic of the film industry.

Acknowledgements

I acknowledge support from the Motion Picture Distributors Association of Australia (MPDAA) in this research.

References

Banjeree, A. (1992) A Simple Model of Herd Behaviour, *Quarterly Journal of Economics*, **107**, 797-817.

Benford, F. (1938) The Law of Anomalous Numbers, *Proceedings of the American Philosophical Society*, **78**, 551.

Bickhchandri, S., Hirshleifer, D., and Welch, I. (1992) A Theory of Fads, Fashion, Custom and Cultural Change as Information Cascades, *Journal of Political Economy*, **100**, 992-1026.

Chamley, C. and Gale, D. (1994) Information Revelation and Strategic Delay in a Model of Investment, *Econometrica*, **62(5)**, 1065-1085.

De Vany, A. (2004) *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*, Routledge, London.

De Vany, A. and Lee, C. (2001) Quality Signals in Information Cascades and the Dynamics of the Distribution of Motion Picture Box Office Revenues, *Journal of Economic Dynamics and Control*, **25(3-4)**, 593-614.

De Vany, A. and Walls, D. (1996) Bose-Einstein dynamics and adaptive contracting in the motion picture industry, *The Economic Journal*, **106**, 1493-1514.

Hand, C. (2001) Increasing Returns to Information: Further Evidence from the UK Film Market, *Applied Economics Letters*, **8**, 419-421.

Ijiri, Y. and Simon, H. (1964) Business Firm Growth and Size, *American Economic Review*, **54** (March), 77-89.

Ijiri, Y. and Simon, H. (1967) A Model of Business Firm Growth, *Econometrica*, **35** (April), 348-355.

Ijiri, Y. and Simon, H. (1971) Effects of Mergers and Acquisitions on Business Firm Concentration, *Journal of Political Economy*, **79** (March/April), 314-322.

Ijiri, Y. and Simon, H. (1974) Interpretations of Departures from the Pareto Curve Firm-Size Distributions, *Journal of Political Economy*, **82(2)**, 315-332.

Maddison, D. (2004) Increasing Returns to Information and the Survival of Broadway Theatre Productions, *Applied Economic Letters*, **11 (10)**, 639-644.

Mandlebrot, B. (1963) New Methods in Statistical Economics, *Journal of Political Economy*, **71**, 421-440.

MPDAA (Motion Picture Distributors Association of Australia) *Various Publications*, http://www.mpdaa.org.au

Pareto, V. (1897) *Cours d'economie politique*, Rouge, Paris.

Sgroi, D. (2002) Optimizing Information in the Herd: Guinea Pigs, Profit and Welfare, *Games and Economic Behaviour*, **39**, 137-166.

Smith, L. and Sorenson, P. (2000) Pathological Outcomes of Observational Learning, *Econometrica*, **68(2)**, 271-398.

Steindl, J. (1965) *Random Processes and the Growth of Firms: A Study of the Pareto Law*, Hafner, New York.

Vining Jr., D.R. (1976) Autocorrelated Growth Rates and the Pareto law: A Further Analysis, *Journal of Political Economy*, **84(2)**, 369-380.

Walls, D. (1997) Increasing Returns to Information: Evidence from the Hong Kong Movie Market, *Applied Economics Letters*, **4(5)**, 187-190.

Yule, G. (1944) *Statistical Study of Literary Vocabulary*, Cambridge University Press, Cambridge.

Zipf, G. (1949) *Human Behaviour and the Principle of Least Effort*, Hafner, New York.

Table 1
Top 5 Films – Cumulative Revenue and Share of Total Box Office

Top Five films	Gross Revenue ($A)	Share of Total Box Office (%)
Titanic	57,645,827	0.834
Shrek 2	50,388,327	0.610
Lord of the Rings: Return of the King	49,370,354	0.606
Lord of the Rings: Fellowship of the Ring	46,827,510	0.597
Lord of the Rings: Two Towers	45,538,891	0.572

Table 2
Ranked Distribution and Share of Total Box Office Revenue

Portion of Ranked Distribution	Gross Revenue ($A) (2,429 films)	Share of Box Office (%)
Top 10% (Films 1-243)	3,827,872,258	55.73
Top 20% (Films 1-486)	5,648,626,716	82.23
Bottom 20% (Films 1,943- 2,429)	14,226,282	0.021
Bottom 10% (Films 2,186-2,429)	3,013,966	0.004
Total	6,868,932,748	100

Table 3
Cumulative Revenue, Weekly Revenue and Weekly Rank Summary Statistics – All Films

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
Cumulative Revenue ($A)						
Full Sample	2,429	2,827,885	593,488	5,405,078	360	57,645,827
Opening Screens <50	1,199	484,938	153,186	998,687	360	12,286,683
Opening Screens 50-149	559	2,143,284	1,408,356	2,438,905	25,115	23,197,810
Opening Screens 150-249	367	6,061,141	4,757,130	4,971,728	502,561	57,645,827
Opening Screens 250+	177	15,939,976	12,893,809	9,340,261	4,032,114	50,388,327
Weekly Revenue ($A)						
Week 1	2,296	911,940	178,729	1,767,334	30	19,617,082
Week 2	1,995	716,773	197,791	1,198,175	98	12,464,352
Week 3	1,775	534,433	165,700	856,284	106	8,657,317
Week 4	1,602	375,483	113,411	618,575	37	6,079,753
Week 5	1,385	261,441	83,923	465,528	29	4,489,890
Week 6	1,211	182,517	58,652	347,347	50	3,848,849
Week 7	1,035	132,554	46,693	250,687	72	2,585,013
Week 8	867	103,062	36,557	202,967	129	2,428,437
Week 9	709	82,824	30,387	170,391	64	2,431,625
Week 10	576	72,697	25,525	154,905	39	2,089,127
Weekly Rank						
Week 1	2,296	14.31	12	11.58	1	55
Week 2	1,995	14.44	11	11.81	1	57
Week 3	1,775	15.84	12	12.24	1	55
Week 4	1,602	17.49	15	12.31	1	56
Week 5	1,385	19.25	17	12.12	1	57
Week 6	1,211	21.14	19	11.99	1	59
Week 7	1,035	22.75	21	11.90	1	58
Week 8	867	24.34	23	11.81	2	58
Week 9	709	25.55	25	11.53	2	55
Week 10	576	26.38	25	11.49	2	55

Table 4
Weekly Screen Average Revenue and Ranking Summary Statistics
– All Films and Wide Release Films

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
Weekly Screen Average Revenue ($A) - All Films						
Week 1	2,296	8,768	6,519	9,836	30	134,098
Week 2	1,995	7,190	5,350	9,660	90	152,331
Week 3	1,775	5,638	3,909	8,604	71	153,145
Week 4	1,602	4,775	2,900	10,327	37	191,884
Week 5	1,385	4,037	2,298	9,157	29	146,761
Week 6	1,211	3,640	1,980	8,670	25	135,970
Week 7	1,035	3,444	1,710	8,786	24	121,050
Week 8	867	3,444	1,617	9,244	44	149,997
Week 9	709	3,341	1,589	8,121	64	85,201
Week 10	576	3,439	1,596	8,205	39	101,630
Weekly Screen Average Rank - All Films						
Week 1	2,296	10.82	8	9.10	1	51
Week 2	1,995	13.48	10	10.62	1	56
Week 3	1,775	16.66	14	11.26	1	54
Week 4	1,602	19.59	18	11.21	1	57
Week 5	1,385	22.12	21	11.22	1	53
Week 6	1,211	24.14	23	11.63	1	59
Week 7	1,035	25.62	25	12.06	1	56
Week 8	867	26.85	27	12.37	1	56
Week 9	709	26.89	27	12.42	1	56
Week 10	576	27.08	27	12.48	1	58
Weekly Screen Average Revenue ($A) – Wide Release Films						
Week 1	1,109	9,276	7,824	6,449	289	46,267
Week 2	1,083	6,454	5,525	4,843	107	44,023
Week 3	1,021	4,607	3,767	3,604	106	37,393
Week 4	952	3,385	2,644	2,695	121	24,779
Week 5	841	2,631	2,063	2,122	103	15,175
Week 6	738	2,159	1,685	1,772	43	16,807
Week 7	629	1,885	1,475	1,528	24	12,764
Week 8	523	1,744	1,362	1,409	44	12,266
Week 9	406	1,688	1,287	1,281	64	10,132
Week 10	320	1,738	1,307	1,972	39	28,150
Weekly Screen Average Rank - Wide Release Films						
Week 1	1,109	4.32	3	3.85	1	21
Week 2	1,083	6.24	5	4.77	1	28
Week 3	1,021	8.43	7	5.50	1	27
Week 4	952	10.20	9	5.43	1	28
Week 5	841	11.79	12	5.37	1	28
Week 6	738	13.05	13	5.38	1	30
Week 7	629	14.07	14	5.48	1	28
Week 8	523	14.81	15	5.51	1	28
Week 9	406	14.93	15	5.52	3	29
Week 10	320	15.03	15	5.64	1	28

Table 5
Pareto Law Regression Results: Weekly and Cumulative Revenue

$\ln(WKREV_{i,t}) = \alpha + \beta \ln(WKRNK_{i,t}) + \gamma(\ln(WKRNK_{i,t}))^2 + \varphi(DUWK_t) + \varepsilon_{i,t}$ Regressions (1) – (4)

$\ln(CUMREV_i) = \alpha + \beta \ln(CUMRNK_i) + \gamma(\ln(CUMRNK_i))^2 + \varepsilon_i$ Regressions (5) – (6)

Parameter	(1)	(2)	(3)	(4)	(5)	(6)
α	16.674**	16.6025**	14.5663**	14.4427**	25.8368**	9.5835**
	(0.0218)	(0.1410)	(0.025727)	(0.1076)	(0.1634)	(0.281)
β	-2.124**	-2.116**	0.2921**	-0.2798**	-1.877**	4.0059**
	(0.0076)	(0.0074)	(0.0232)	(0.022)	(0.0238)	(0.096)
γ	-	-	-0.5437**	-0.5399**	-	-0.503**
			(0.0051)	(0.0048)		(0.0081)
Weekly dummies	No	Yes	No	Yes	-	-
Adj R^2	0.8258	0.8378	0.8977	0.9086	0.7195	0.8915
F - Statistic	78069.98	1606.73	72254.52	3032.31	6229.79	9977.75
(df_1, df_2)	(1, 16470)	(53, 16418)	(2, 16469)	(54, 16417)	(1, 2427)	(2, 2426)
Observations	16472	16472	16472	16472	2429	2429

Standard errors in parentheses.
* Significant at 10% level, ** significant at 5% level.

Table 6
Comparisons of Pareto Law Results

$\ln(WKREV_{i,t}) = \alpha + \beta \ln(WKRNK_{i,t}) + \gamma(\ln(WKRNK_{i,t}))^2 + \varphi(DUWK_t) + \varepsilon_{i,}$

Parameter	De Vany and Walls (1996)	Walls (1997)	Hand (2001)	Present Study
α	14.8405**	13.5601**	14.533**	14.4427**
	(0.0759)	(0.1096)	(0.0494)	(0.1076)
β	0.1859**	-0.0402	-0.4174**	-0.2798**
	(0.04033)	(0.0644)	(0.0636)	(0.022)
γ	-0.4033**	-0.443**	-0.339**	-0.5399**
	(0.0089)	(0.0257)	(0.0208)	(0.0048)
Weekly dummies	Yes	Yes	Yes	Yes
R^2	0.9397	0.8717	0.9134	0.9086
Observations	2000	710	780	16472

Standard errors in parentheses.
* Significant at 10% level, ** significant at 5% level.

Table 7
Pareto Law Results: Cumulative Revenue by Opening Screens

$\ln(CUMREV_i) = \alpha + \beta \ln(CUMRNK_i) + \gamma(\ln(CUMRNK_i))^2 + \varepsilon_i$

Parameter	Opening Screens < 50	Opening Screens 50 - 149	Opening Screens 150 - 249	Opening Screens 250+
α	12.2372**	14.8772**	16.6399**	17.3318**
	(0.2527)	(0.1824)	(0.1261)	(0.3981)
β	2.133**	1.1268**	0.3732**	0.2562**
	(0.0976)	(0.0818)	(0.062)	(0.033)
γ	-0.3544**	-0.2347**	-0.1274**	-0.1131**
	(0.0093)	(0.009)	(0.0074)	(0.0047)
Adj R^2	0.8957	0.8973	0.9039	0.9730
F - Statistic	5146.09	2438.86	1721.31	3174.67
(df_1, df_2)	(2,1196)	(2, 556)	(2, 364)	(2, 174)
Observations	1199	559	367	177

Standard errors in parentheses.
* Significant at 10% level, ** significant at 5% level.

Table 8
Pareto Law Regression Results: Weekly Screen Averages – All Films and Wide Release Films

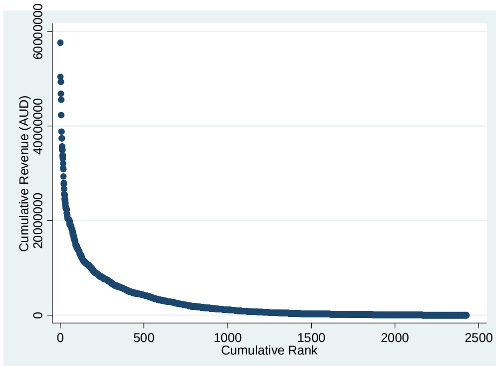
$\ln(WKSCRNAVREV_{i,t}) = \alpha + \beta \ln(WKSCNRNAVRNK_{i,t}) + \gamma(\ln(WKSCRNAVRNK_{i,t}))^2 + \varphi(DUWK_t) + \varepsilon_{i,t}$

	All Films		Wide Release Films	
Parameter	(1)	(2)	(1)	(2)
α	11.3195**	10.5569**	10.3998**	9.8437**
	(0.0724)	(0.0654)	(0.0861)	(0.0743)
β	-1.075**	-0.2246**	-1.061**	-0.1131**
	(0.0038)	(0.0133)	(0.0054)	(0.0173)
γ	-	-0.1921**	-	-0.2759**
		(0.0029)		(0.0049)
Weekly dummies	Yes	Yes	Yes	Yes
Adj R^2	0.8384	0.8724	0.8232	0.8707
F - Statistic	1613.30	2086.68	776.56	1102.05
(df_1, df_2)	(53, 16418)	(54, 16417)	(53, 8774)	(54, 8773)
Observations	16472	16472	8828	8828

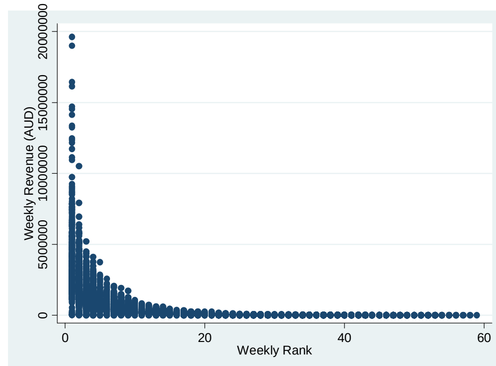
Standard errors in parentheses.
* Significant at 10% level, ** significant at 5% level.

Figure 1
Pareto Rank Distribution: Revenue vs. Rank

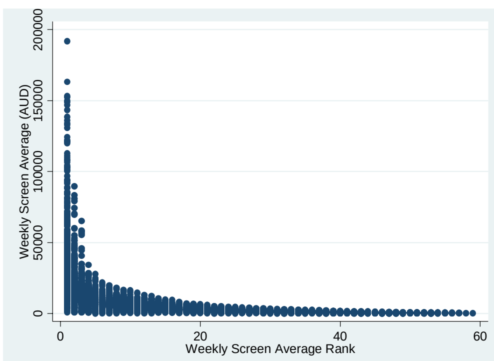
Cumulative Revenue -All Films



Weekly Revenue -All Films



Weekly Screen Average -All Films



Weekly Screen Average -Wide Release Films

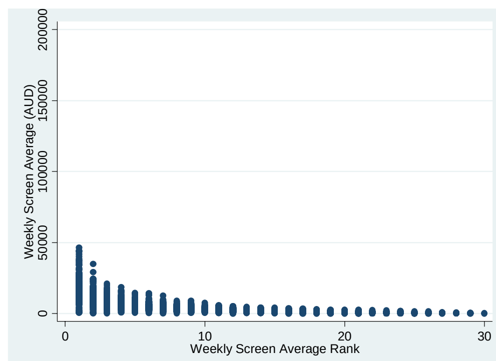


Figure 2
Pareto Rank Distribution: Log(Revenue) vs. Log(Rank)

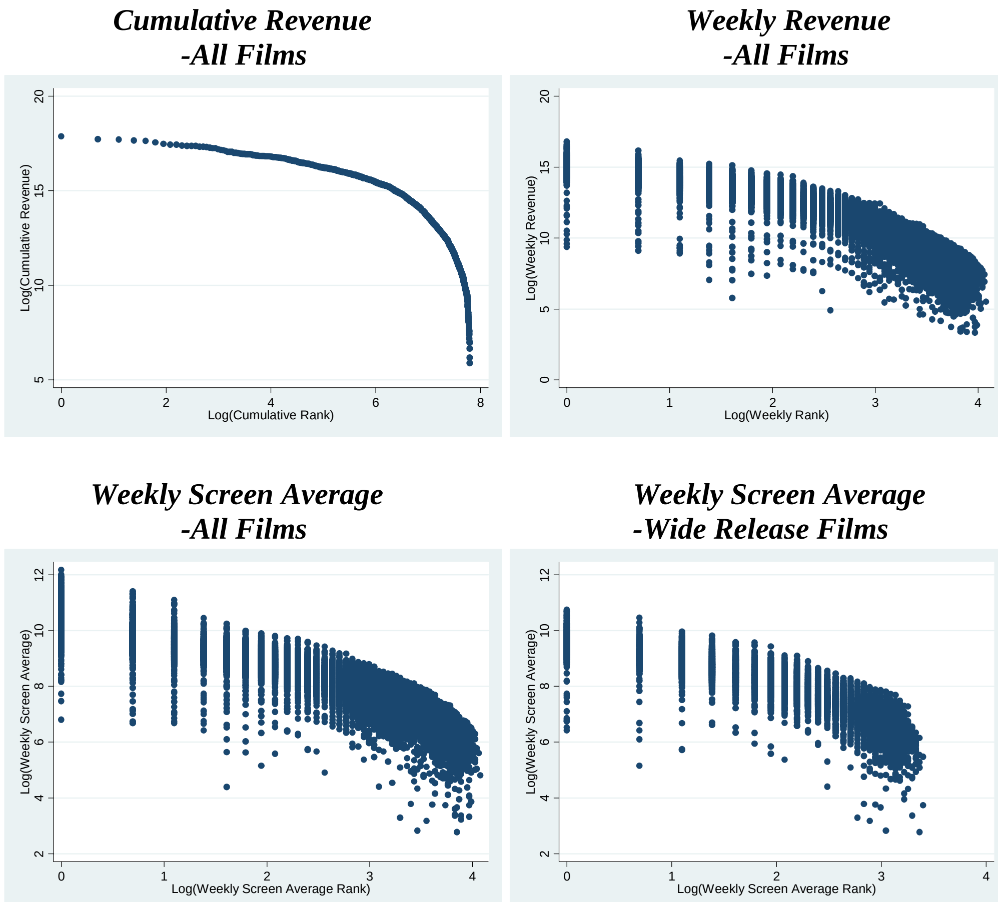


Figure 3
Log(Cumulative Revenue) vs. Log(Cumulative Rank)