Robust estimation of the modified autoregressive index for high grossing films at the US box office, 1935 to 2005

Nick Redfern

Abstract

The modified autoregressive (mAR) index describes the clustering of shots of similar duration in a motion picture. In this paper we derive robust estimates of the mAR index for high grossing films at the US box office using a rank-based autocorrelation function resistant to the influence of outliers and compare this to estimates obtained using the classical, moment-based autocorrelation function. The results show that (1) The classical mAR function underestimates both the level of shot clustering and the variation in style among the films in the sample.; (2) there is a decline in shot clustering from 1935 to the 1950s followed by an increase from the 1960s to the 1980s and a levelling off thereafter rather than the monotonic trend indicated by the classical index, and this is mirrored in the trend of the median shot lengths and interquartile range; and (3) the rank mAR index indentifies differences between genres missed by the classical index.

Keywords: robust statistics, autocorrelation, modified autoregressive index, film style, editing, Hollywood

1. Introduction

Cutting, De Long, and Nothelfer (2010) proposed the modified autoregressive (mAR) index as a measure of the degree to which shots of similar length cluster together in a motion picture. Applying this method to 150 high grossing films at the US box office released from 1935 to 2005 they identified a tendency for the shots in a motion picture to cluster over time and for action films to exhibit a greater degree of shot clustering than other genres. They derived the mAR index of a film from the partial autocorrelation function of its shot lengths, where the function used was the classical, moment-based estimator. However, this function is well known to be non-resistant to the presence of outliers (Chan 1995), resulting in mAR index values that do not accurately describe the clustering of shots and lead to flawed interpretations of film style.

In this paper we calculate robust estimates of the mAR index using a rank-based autocorrelation function and compare this to the index based on classical autocorrelation functions. The next section describes the two time series functions, and in section three we compare these functions by looking at the shot length data used by Cutting, De Long, and Nothelfer in terms of the large scale structure described by the autocorrelation functions, the mAR index, the relationship between the mAR index and other statistics of film style, and variation of the mAR index between genres.

2. Classical and rank autocorrelation functions

The autocovariance function of a dataset describes the statistical dependence between the values taken by a stochastic process at two points in time. The classical, moment-based autocovariance function for a weakly stationary time series $\mathbf{x} = (X_1, \dots, X_n)^{\mathrm{T}}$ is defined as

$$\gamma(h, \mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \bar{X})(X_{i+h} - \bar{X}) \,, \tag{1}$$

where $\bar{X}$ is the mean and h is a lag operator specifying the distance between the observations X_i and X_{i+h}. The denominator in (1) is the total sample size (n), and so this function is biased and positive semidefinite. The autocovariance function when $h = 0$ is equal to the variance and standardising (1) by this value gives the autocorrelation function

$$\rho(h, \mathbf{x}) = \frac{\gamma(h, \mathbf{x})}{\gamma(0, \mathbf{x})} . \tag{2}$$

The autocorrelation function ranges from –1 to 1, with negative autocorrelation at lag h reflecting a tendency of observations to lie on opposite sides of the mean and positive autocorrelation a tendency for observations tend to lie on the same side of the mean. The partial autocorrelation function ($\alpha(h, \mathbf{x})$) is the correlation between X_i and X_{i+h} with the linear dependence of the intervening lags removed, and can be calculated recursively using the Durbin-Levinson algorithm:

$$\alpha(h, \mathbf{x}) = \frac{\rho_h - \sum_{j=1}^{k-1} \rho_{h-j} \phi_{k-1,j}}{1 - \sum_{j=1}^{k-1} \rho_j \phi_{k-1,j}} \,, \tag{3}$$

where $\phi_{a,b} = \phi_{a-1,b} - \alpha_a \phi_{a-1,a-b}$ and $\rho_h$is the autocorrelation function at lag h.

These functions are not resistant to outliers. The mean and the variance of a dataset have finite sample breakdown points of $1/N$ and unbounded influence functions, and can be arbitrarily bad estimates of location and dispersion in the presence of even single outlier. As the above time series functions are based on the mean and variance of a dataset they are similarly affected by the presence of outliers. The presence of outliers in the upper tail of a distribution inflates the mean so that the majority of observations will tend to lie on the same side of the mean irrespective of the underlying structure of the time series. Consequently, the autocorrelation function will overestimate positive correlation and underestimate negative correlation. The presence of outliers inflates the variance introducing a bias of $\rho(h, \mathbf{x})$ toward zero that becomes stronger as the magnitude of the outlier increases (see Maronna, Martin, & Yohai 2006: 250-252). Consequently, the presence of outliers leads to underestimation of the level of autocorrelation between observations in a time series. At the same time, if a time series contains more than one outlier we may find spuriously large autocorrelation coefficients when h is equal to the distance between outliers (Chatfield 2004: 27). The lack of robustness of the classical autocovariance and its derived functions mean that the information it carries about the structure of a time series can be destroyed by just a single outlier.

As an alternative to the classical function, a rank-based approach by Ahdesmäki et al. (2005) calculates the autocorrelation function of a time series as

$$\hat{\rho}_S(h,\mathbf{x}) = \frac{1}{n}\frac{12}{(n-h)^2-1}\sum_{i=1}^{n-h}\left(R_x(i) - \frac{n-h+1}{2}\right)\left(R'_x(i) - \frac{n-h+1}{2}\right), \qquad (4)$$

where $R_x(i)$ are the ranks of x_i in $S = \{x_t, t = 1, \dots n-h\}$ and $R'_x(i)$ are the ranks of x_{i+h} in $S' = \{x_{t+h}, t = 1, \dots n-h\}$. As a moving-window extension of Spearman's rank correlation statistic, $\hat{\rho}_S$ measures the monotonicity of the relationship between two observations and does not assume linearity. Because $\hat{\rho}_S$ is based on the ranks of the data it is resistant to the influence of outliers (see Boudt, Cornelisson, & Croux 2012). This function is biased and is directly comparable to the biased autocorrelation function based on (1), though it is not guaranteed to be positive semidefinite.

3. Results

The data set used in this study comprises the same data used by Cutting, De Long, and Nothelfer accessed via the Cinemetrics database (http://www.cinemetrics.lv/index.php). However, we were unable to use all 150 films from the original study because the minimum shot length for nine films was given as 0.0 seconds and was less than 0.0s for seven films, presumably due to rounding or data entry errors. These films were excluded from the study to give an actual sample size of 134 films. The full set of results is in the supplementary material attached to this article.

We calculated the classical and rank autocorrelation and partial autocorrelation functions for the linearly detrended shot length of films in the sample. The rank-based function in (4) was calculated using **R** functions by Spangl (2009). To ensure positive semi-definiteness we arranged the rank autocorrelation function as a toeplitz matrix and found the nearest positive definite correlation matrix using the `nearPD` function in the **R** Matrix package, which implements Higham's (2002) algorithm. The partial autocorrelations were determined recursively using the Durbin-Levinson algorithm. We then determined the order of the autocorrelation and partial autocorrelation functions by considering values at lag h so long as previous values were greater than the positive bound $2/\sqrt{N}$, where N is the number of shots in a film, thereby assuming a null hypothesis of white noise. The methodology used here diverges slightly from that used by Cutting, De Long, and Nothelfer, and so there are some discrepancies between their values for the classical autocorrelation function and the estimates used here.

The autocorrelation function describes the large scale structure of the time series of each film. In some cases the functions indicate a similar extent of significant autocorrelation (the order of the autocorrelation) while under-estimating the degree of autocorrelation, but in others the functions describe very different editing patterns. From the autocorrelation data, we see *A Night at the Opera* (1935) is an extreme case of how the functions can differ, with the classical function leading us to assume a white noise mode of uncorrelated shot lengths while the rank function shows that shots in this film are correlated up to lag-30 (see supplementary material). To illustrate these differences Figure 1 plots the classical and rank autocorrelation functions for two films. The autocorrelation functions for *Santa Fe Trail* have similar large scale structure but the classical function underestimates the degree of correlation between shots. The functions for *Jaws* show very different patterns. The classical function for this film indicates short-range correlation between shots with no linear relationship between distant shots, whereas the rank autocorrelation function shows much stronger correlation between shots at short lags and between shot lengths at much longer lags.

This has implications for the claim that editing of films in the sample show an increasing tendency over time to a $1/f$ noise pattern. The power spectral density of a time series is the Fourier transform of its autocorrelation function, and so the problem of outliers in the data will be transmitted to spectral analysis. In the case of *Santa Fe Trail* the power spectral density of the robust autocorrelation function will not be dramatically different from that of the classical function given the similarities between their large scale structures. However, the 'waves' in the rank function for *Jaws* show long range correlation between shots rather than the white noise indicated by the classical function and so we should expect to see some key differences between the estimates of the spectral density based on these two functions.

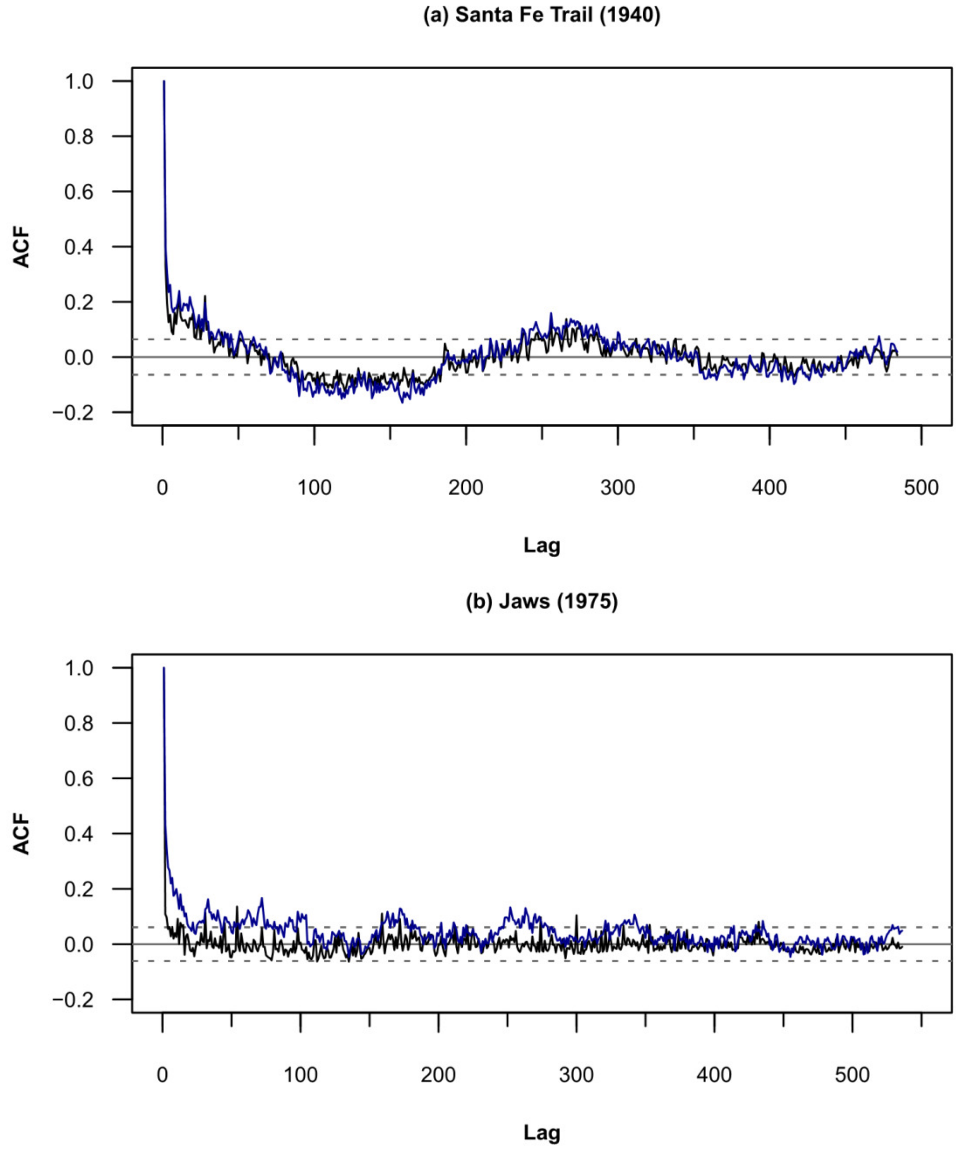

Figure 1 Classical (black) and rank (blue) autocorrelation functions of two films for lags 0 to $(N/2) - 1$. The critical values are set at $\pm 2/\sqrt{N}$.

To determine the mAR indices we fit the negative exponential function $1/[1+h]^{\beta}$ to the classical and robust partial autocorrelation functions for lags 0 to 20 using nonlinear least squares ($df = 20$) with the value of mAR determined to be the intercept between the fitted function and a critical value of 0.0611 based on the median number of shots in a film in the sample (N = 1070). We were unable to determine the classical mAR index for three films (*A Night at the Opera* [1935], *The Great Dictator* [1940], *Detour* [1945]) because the lag-1 autocorrelation was negative resulting in a singular gradient when fitting by non-linear least squares, and so these films are excluded from discussion of the classical mAR index but the rank index of each film is included.

The classical mAR index is less than the rank mAR index for 96 per cent of films in the sample, with a median difference between the two indices of -1.42 (95% CI: -1.62, -1.21). The largest difference is for *Charlie's Angels* (2000), which has a classical mAR index of 1.82 but a rank index of 6.82. Figure 2 shows the partial autocorrelation and fitted negative exponential functions for *Santa Fe Trail* and *Jaws*, and clearly illustrates the large differences between the two measures of film style.

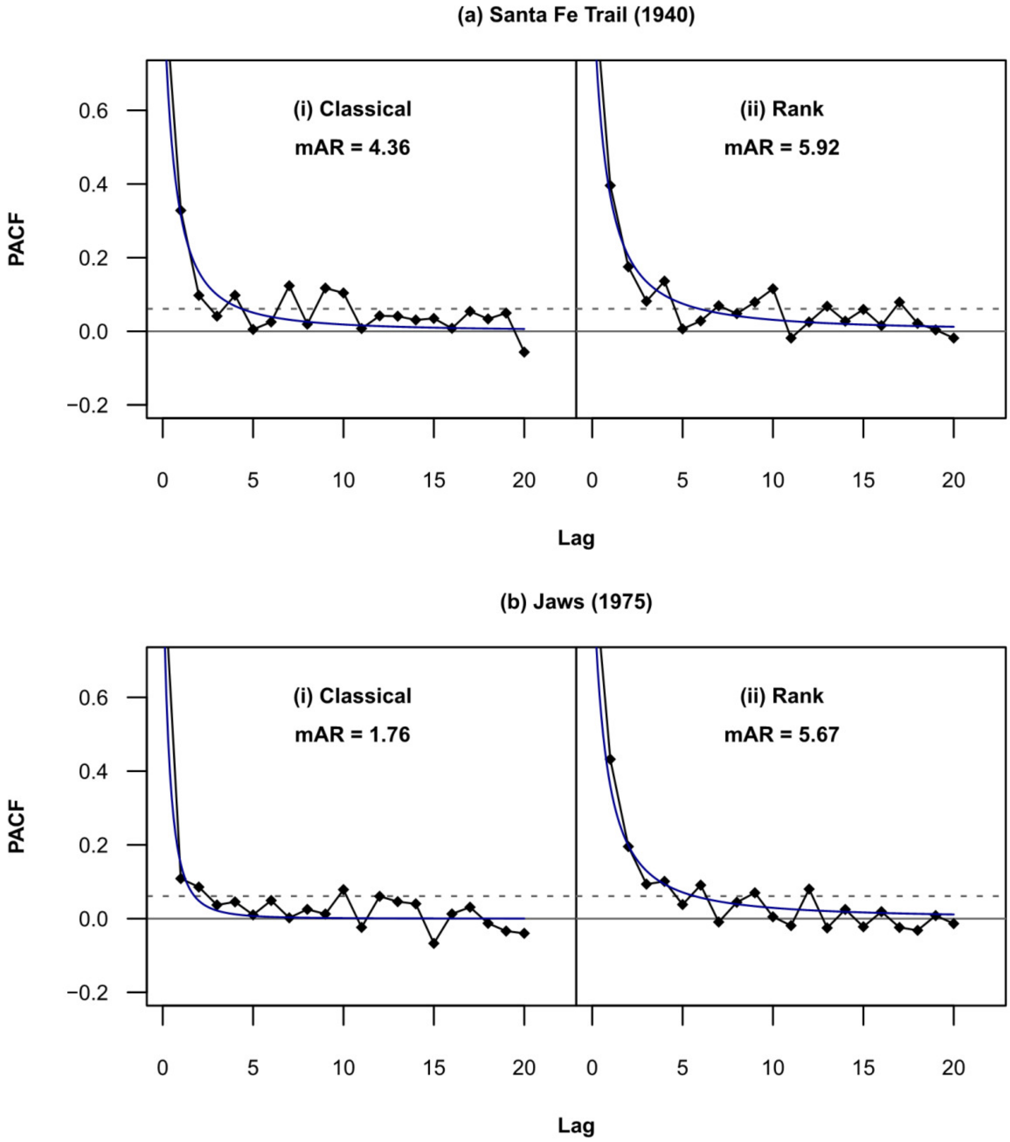

Figure 2 Partial autocorrelation and fitted negative exponential functions for two films. The critical value is at 0.0611. The ordinate has been truncated at 0.6.

The rank mAR index of a film is strongly correlated with its year of release: *r* = 0.40 (95% CI: 0.25, 0.53), *t* (129) = 5.01, *p* = <0.01), This is greater than the correlation based on our estimates of the classical mAR index (*r* = 0.32 (95% CI: 0.16, 0.47), *t* (129) = 3.88, *p* = <0.01); and is comparable to those reported by Cutting, De Long, and Nothelfer (*r* = 0.43). These similarities should not lead us to conclude the classical and rank mAR indices show the same pattern from over time. Figure 3 presents the times series plots of the classical and rank mAR indices with fitted LOESS trendline and bootstrapped 95% confidence interval. The classical mAR index shows a gradual trend to increased shot clustering, and is consistent with the monotonic trend reported by Cutting, De Long, and Nothelfer in Figure 2a of their article. The trendline for the rank mAR index shows a very different pattern of changes in film style with a decline in the clustering of shots from 1935 to the 1950s followed by an increase from the 1960s to the 1980s with a levelling off after 1985. Basing our analyses of changes in film style overtime on the classical mAR index would thus lead us to incorrectly describe changes in films style over time and to underestimate the size of those changes.

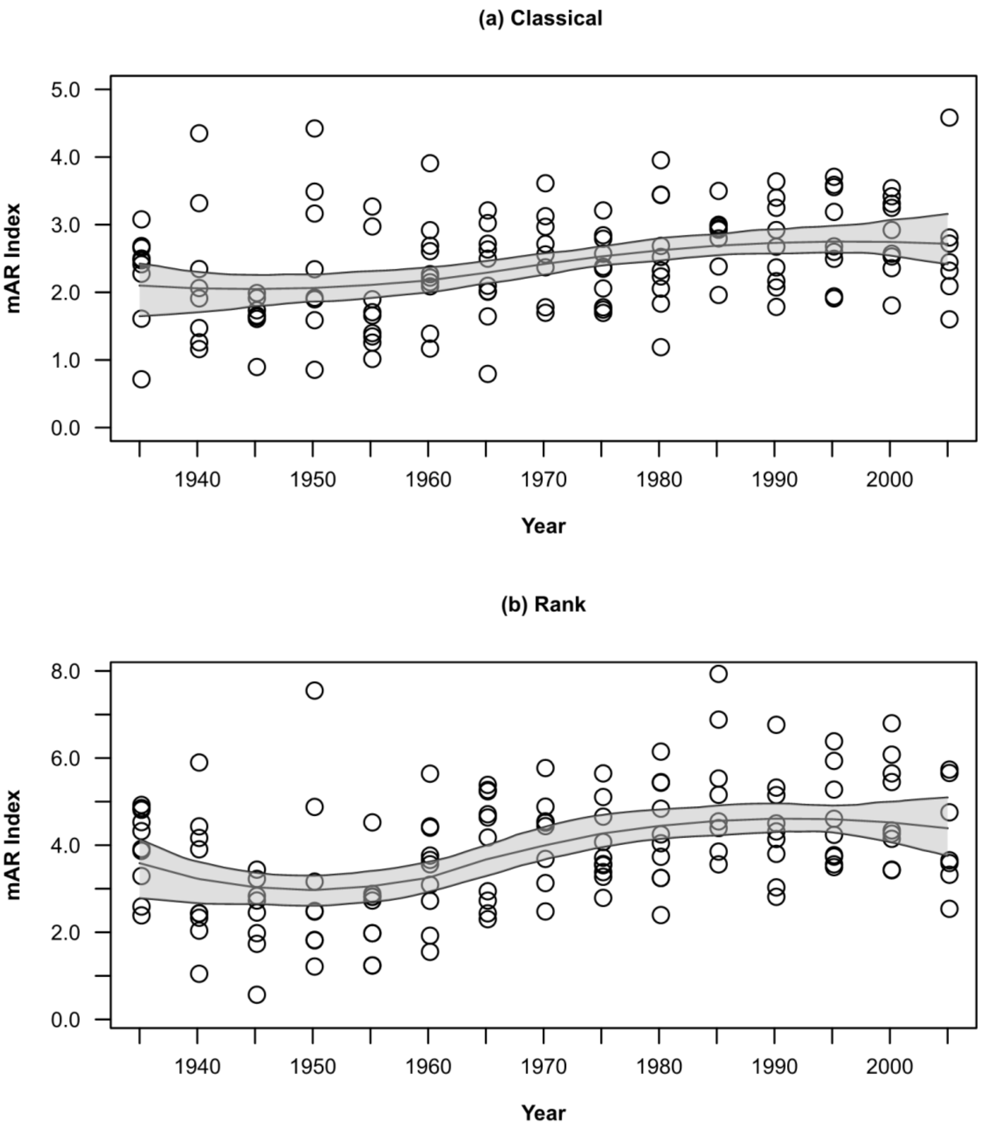

Figure 3 Time series of classical and rank mAR indices with fitted LOESS trendlines and bootstrapped 95% confidence intervals

Cutting, De Long, and Nothelfer state that the increase in the mAR index overtime is not an artefact of changes in the average shot length. Though the shot length distribution of a motion picture is positively skewed the average shot length they refer to is the mean. The failure to link changes in the mAR index to changes in the mean shot length may be due to the mistaken impression gained from the classical mAR index of a monotonically increasing trend that did not correspond to the time series of other statistics suggesting two separate trends in film style or to the greater variation in the mean shot length resulting from its own lack of robustness. They did not discuss changes in the mAR in relation to the dispersion of shot lengths in a motion picture.

Measures of location and scale for shot length distributions are highly correlated, and so we combined the median shot length and interquartile range of each film into a single dummy variable using principal components analysis with films with a low median and low IQR having a low score and a stronger tendency to more rapid editing.[1] A score of zero represents the average of the dummy variable. Plotting this score against year of release (Figure 4) we see the same trend in film style evident in the rank mAR index, with above average scores tending to come in the early decades of the sample while later films tend have lower-than-average scores. There is a slowing down of film editing from 1935 to the 1950s as the median and dispersion of shot lengths increases, followed by a decrease on both measures from the mid-1960s to 2005. The differences between a group of films and the one immediately preceding it have become smaller over time as editing has stabilised into a single Hollywood style. The greater variation in the scores in Figure 4 for the 1940s and 1950s indicates much greater stylistic variation between films from those decades, while the shows that films have converged to a single, rapidly-edited style.

These results show there is no evidence to support the widely held opinion that the trend to rapid cutting in Hollywood cinema was initiated with the advent of MTV in the early 1980s (see, for example, Dixon 2001; King 2002: 247; Rombes 2005: 13). Shot lengths have been steadily decreasing for over fifty years, with the largest shift to rapid editing occurring in the early 1960s, and far from increasing with the introduction of MTV, this process has slowed after 1985. The time has come to rid the study of film of this factoid once and for all as demonstrably untrue (see Calavita 2007).

---

[1] We did not repeat the autoregressive analysis on the log-transformed shot length data of these films because the lognormal distribution is not an appropriate parametric model for this data (Redfern, in press).

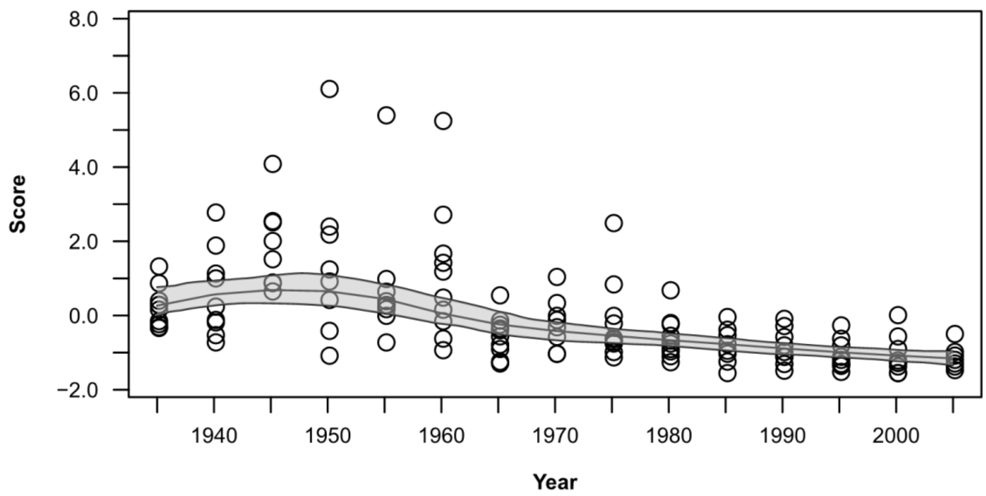

Figure 4 First principal component score against year of release for high grossing films at the US box office, 1935 to 2005 (*N* = 134). The dummy variable is a linear combination of the median shot length and interquartile range, and accounts for 96.5% of the variance of the original variables.

Finally, Cutting, De Long, and Nothelfer assigned the films in their sample to one of five genres (action, adventure, animation, comedy, drama) and compared the distribution of the mAR index of each genre. They found that action films tend to have a higher mAR index than films in other genres along with smaller differences between the other genres, though they did not correct for multiple comparisons. For each method of calculating the mAR index, we performed a Kruskal-Wallis ANOVA test, and pairwise post-hoc Dunn tests. Post-hoc testing was carried out based on an experiment-wise error rate of 0.10 and 10 tests giving a two-tailed Sidak-corrected *p*-value of 0.0105 and a critical Z-value of 2.56. A pairwise Z-statistic greater than or equal to the critical value indicates a statistically significant difference between the distributions of the mAR index for those genres. All statistics are corrected for ties.

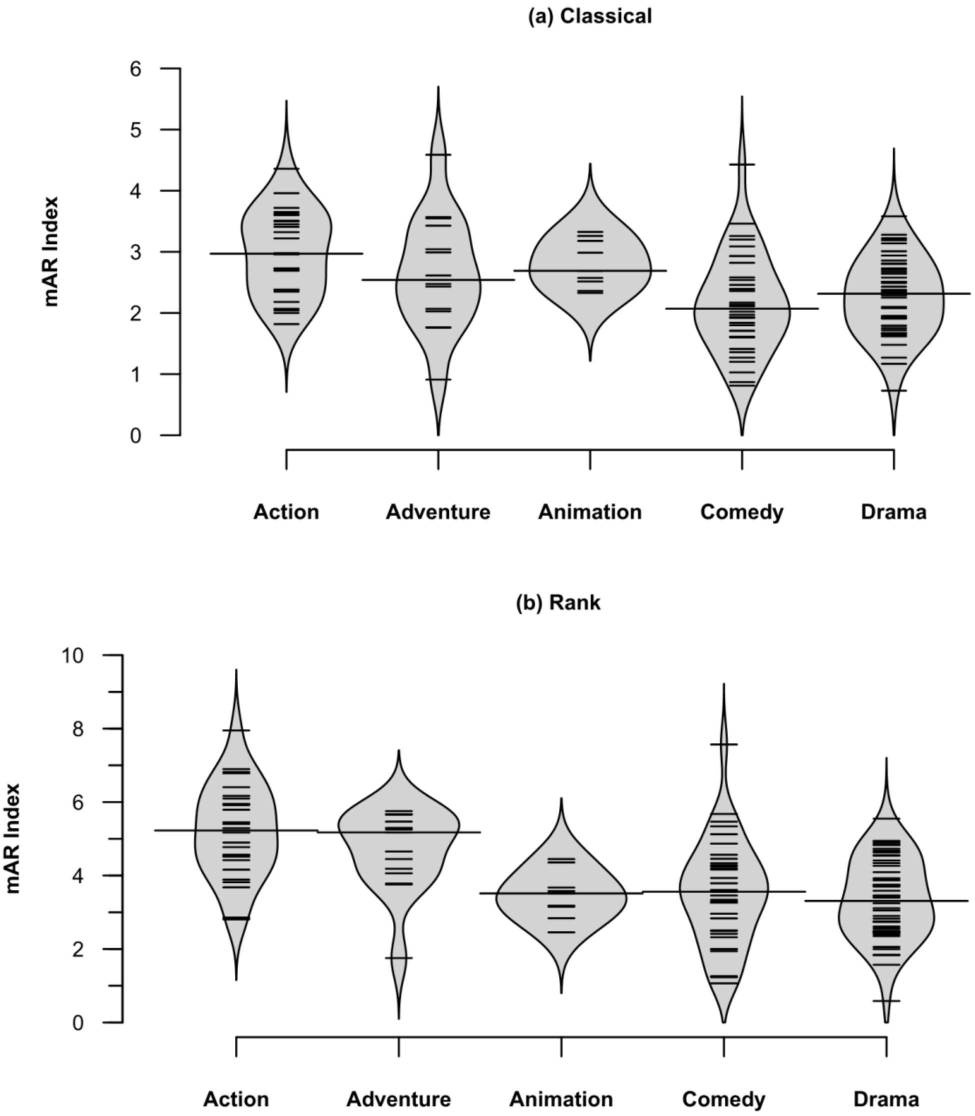

Figure 5 Beanplots of classical and rank mAR indices sorted by genre. The average beanlines are set at the sample median.

The beanplots in Figure 5 present the classical and rank mAR indices sorted by the above genre categories._The results for the classical mAR index show that there is a statistically significant difference (KW ANOVA: χ^2 (4) = 28.50, p = <0.01), and post-hoc testing shows this to between action films and comedy films (Z = 4.09) and action films and drama films (Z = 4.49). There were no other pairwise significant differences. Turning to the rank mAR values, there is again a statistically significant difference (KW ANOVA: χ^2 (4) = 38.61, p = <0.01) with significant pairwise differences between action films and animated films (Z = 3.22), comedy films (Z = 4.59), and drama films (Z = 5.22), and between adventure films and the comedy (Z = 3.15) and drama (Z = 3.60) genres. Using the classical mAR index would therefore lead us to miss key differences between genres. Looking at the shape of the distributions in Figure 5, the rank mAR index shows much greater variation in the editing style among films in the sample.

4. Conclusion

This paper compared estimates of the degree of shot clustering using the classical, moment-based autocorrelation function and a rank-autocorrelation function resistant to the influence of outliers. The results show that the classical mAR index is not robust to outliers and underestimates both the level of shot clustering and the variation in style among the films in the sample. We also found that this index gives a misleading impression of changes in film style over time and that the trends identified by the rank mAR index are consistent with trends in other statistics describing the editing style of these films. Finally, the classical mAR index failed to identify key differences between films in the sample when sorted by genre. These results show that the mAR index can be a useful statistics of film style but that it is necessary to use robust methods die to the presence of outliers in shot length data.

References

Ahdesmäki M, Lähdesmäki H, Pearson R, Huttunen H, and Yli-Harja O 2005 Robust detection of periodic time series measured from biological systems, *BMC Bioinformatics* 6: 117.

Boudt K, Cornelisson J, and Croux C 2012 The Gaussian rank correlation estimator: robustness properties, *Statistics and Computing* 22 (2): 471-483.

Calavita M 2007 'MTV aesthetics' at the movies: interrogating a film criticism fallacy, *Journal of Film and Video* 59 (3): 15-31.

Chan WS 1995 Understanding the effects of time series outliers on sample autocorrelation, (1): 179-186.

Chatfield C 2004 *The Analysis of Time Series: An Introduction*. Boca Raton, FL: Chapman & Hall/CRC.

Cutting JE, De Long JE, and Nothelfer CE 2010 Attention and the evolution of Hollywood film, *Psychological Science* 21 (3): 432-439.

Dixon WW 2001 Twenty-five reasons why it's all over, in J Lewis (ed.) *The End of Cinema as We Know It: American Film in the Nineties*. New York: NYU Press: 356-66.

Higham NJ 2002 Computing the nearest correlation matrix – a problem from finance, *IMA Journal of Numerical Analysis* 22 (3): 329-343.

King G 2002 *New Hollywood Cinema: An Introduction*. London: IB Tauris.

Maronna R, Martin D, and Yohai V 2006 *Robust Statistics: Theory and Method*. Chichester: John Wiley & Sons.

Spangl B 2009 [Robust-ts-commits] r9 - pkg/R (04 Mar 2009), http://lists.r-forge.rproject.org/pipermail/robust-ts-commits/2009-March/000000.html, accessed 23 March 2012.

Redfern N *in press* The lognormal distribution is not an appropriate parametric model for shot length distributions of Hollywood films, *Literary and Linguistic Computing*.

Rombes N 2005 *New Punk Cinema*. Edinburgh: Edinburgh University Press.

Supplementary Material

Key: Med = median shot length (in seconds), IQR = interquartile range (in seconds), PC1 = first principle component score, cACF = order of classical autocorrelation function, cAR = classical Autoregressive index, cmAR = classical modified Autoregressive index, rACF = order of rank autocorrelation function, rAR = rank Autoregressive index, rmAR = rank modified Autoregressive index

Title	Year	Genre	Med	IQR	PC1	cACF	cAR	cmAR	rACF	rAR	rmAR
A Night at the Opera	1935	Comedy	4.0	6.2	-0.21	0	0	N/A	30	2	4.87
A Tale of Two Cities	1935	Drama	5.2	7.0	0.31	4	3	2.49	15	3	4.94
Anna Karenina	1935	Drama	6.5	8.0	0.89	0	0	0.73	5	3	2.61
Captain Blood	1935	Action	4.4	4.8	-0.31	5	4	2.70	17	5	3.89
Les Miserables	1935	Drama	5.0	7.4	0.31	4	1	2.29	20	4	4.34
Mutiny on the Bounty	1935	Drama	4.5	7.6	0.18	3	2	2.51	17	3	4.83
The 39 Steps	1935	Drama	4.1	6.5	-0.13	3	3	2.67	6	4	3.31
The Informer	1935	Drama	7.4	9.0	1.34	2	0	1.62	2	2	2.41
Top Hat	1935	Comedy	5.4	7.3	0.42	5	3	3.09	7	4	3.93
Westward Ho	1935	Drama	4.0	5.7	-0.29	4	3	2.43	18	3	4.55
Fantasia	1940	Animation	6.5	8.8	1.02	4	1	3.33	8	3	4.45
Foreign Correspondent	1940	Drama	4.3	5.9	-0.16	2	2	1.92	11	4	3.92
Grapes Of Wrath	1940	Drama	6.8	9.0	1.15	1	1	1.48	5	3	2.45
Pinocchio	1940	Animation	4.0	4.4	-0.51	4	3	2.36	4	2	2.45
Rebecca	1940	Drama	4.8	7.5	0.26	0	0	1.27	3	2	2.06
Santa Fe Trail	1940	Action	4.3	6.3	-0.10	29	2	4.36	32	4	5.92
The Great Dictator	1940	Comedy	9.2	14.3	2.80	0	0	N/A	0	0	1.07
The Letter	1940	Drama	7.1	13.0	1.91	0	0	1.17	2	2	2.35
Thief Of Bagdad	1940	Adventure	3.6	4.0	-0.70	2	2	2.07	15	5	4.19
Bells Of St Mary’s	1945	Drama	6.0	9.0	0.89	2	1	1.62	4	2	2.47
Blood On The Sun	1945	Action	9.1	12.9	2.53	2	1	2.00	2	2	2.85
Brief Encounter	1945	Drama	9.2	12.9	2.57	0	0	1.65	2	2	3.24

Title	Year	Genre	Med	IQR	PC1	cACF	cAR	cmAR	rACF	rAR	rmAR
Detour	1945	Drama	12.3	16.2	4.11	0	0	N/A	0	0	0.58
In Pursuit To Algiers	1945	Adventure	5.8	13.3	1.54	0	0	0.91	1	1	1.75
Leave Her to Heaven	1945	Drama	6.0	9.1	0.91	2	1	1.93	2	2	2.00
Lost Weekend	1945	Drama	8.2	11.6	2.03	1	1	1.67	2	2	2.75
Spellbound	1945	Drama	5.1	9.4	0.67	2	2	1.75	12	2	3.46
All About Eve	1950	Drama	4.9	8.4	0.44	2	1	2.35	4	1	1.83
Annie Get Your Gun	1950	Comedy	5.0	13.2	1.26	9	3	4.43	23	2	7.57
Born Yesterday	1950	Comedy	6.9	16.5	2.42	1	1	1.93	2	2	2.49
Cheaper By The Dozen	1950	Comedy	7.3	14.4	2.20	2	2	1.60	2	2	2.51
Cinderella	1950	Animation	3.1	2.8	-1.06	10	5	3.18	9	5	3.18
Harvey	1950	Comedy	13.2	26.7	6.13	0	0	0.87	0	0	1.23
The Asphalt Jungle	1950	Drama	6.6	8.1	0.94	1	1	1.91	2	2	1.84
The Flame and the Arrow	1950	Action	4.0	5.1	-0.39	5	1	3.50	12	5	4.90
Battle Cry	1955	Drama	6.0	7.6	0.66	1	1	1.72	6	2	2.75
East Of Eden	1955	Drama	5.1	7.8	0.41	2	2	1.91	6	2	2.83
Lady And The Tramp	1955	Animation	3.8	3.6	-0.70	6	4	2.99	6	6	2.84
Mr Roberts	1955	Comedy	6.5	8.7	1.01	0	0	1.41	3	2	2.00
Night Of The Hunter	1955	Drama	4.9	6.9	0.19	2	2	1.67	8	2	2.90
Rebel Without A Cause	1955	Drama	4.8	6.0	0.01	8	3	3.28	14	3	4.54
Seven Year Itch	1955	Comedy	11.3	26.1	5.42	1	1	1.27	0	0	1.26
The Ladykillers	1955	Comedy	5.1	7.1	0.29	1	1	1.36	2	2	2.00
The Trouble With Harry	1955	Comedy	5.1	7.2	0.31	0	0	1.03	0	0	1.25
Butterfield 8	1960	Drama	6.1	10.7	1.21	3	3	2.23	3	2	3.11
Exodus	1960	Action	13.8	20.3	5.27	3	3	2.93	6	2	4.42
Inherit The Wind	1960	Drama	7.5	10.9	1.69	0	0	1.18	0	0	1.57
The Magnificent Seven	1960	Adventure	4.5	5.7	-0.13	5	2	2.16	19	3	4.45
Ocean’s 11	1960	Comedy	7.5	9.4	1.44	4	3	2.70	2	2	1.94

Title	Year	Genre	Med	IQR	PC1	cACF	cAR	cmAR	rACF	rAR	rmAR
Peeping Tom	1960	Drama	5.2	8.2	0.50	0	0	1.40	4	4	2.74
Spartacus	1960	Action	4.9	6.8	0.18	3	3	2.28	9	3	3.68
Swiss Family Robinson	1960	Adventure	3.1	3.7	-0.91	14	3	3.92	22	5	5.66
The Apartment	1960	Comedy	8.3	15.7	2.74	3	2	2.62	3	2	3.58
The Time Machine	1960	Adventure	3.4	5.0	-0.60	2	2	2.10	12	2	3.78
Dr Zhivago	1965	Drama	5.9	7.2	0.57	3	3	1.66	5	3	2.45
Flight Of The Phoenix	1965	Adventure	3.3	3.5	-0.88	16	5	3.04	41	5	5.26
Those Magnificent Men in Their Flying Machines	1965	Adventure	3.9	4.4	-0.54	2	1	2.03	17	2	4.65
Help	1965	Comedy	2.6	2.8	-1.22	2	2	2.02	6	3	2.96
Shenandoah	1965	Drama	4.0	5.2	-0.38	5	3	2.51	5	2	2.75
Sound Of Music	1965	Drama	4.2	5.1	-0.33	4	2	2.64	35	2	4.73
That Darn Cat	1965	Comedy	3.4	4.0	-0.77	0	0	0.81	4	3	2.32
The Great Race	1965	Action	4.0	6.1	-0.23	8	3	2.73	23	4	5.29
Thunderball	1965	Action	2.5	2.7	-1.27	6	7	3.22	23	4	5.40
What's New Pussycat	1965	Comedy	4.2	6.4	-0.11	1	1	2.11	12	3	4.20
Airport	1970	Drama	4.7	6.2	0.01	3	2	1.79	4	2	2.50
Aristocats	1970	Animation	3.3	2.7	-1.01	4	4	2.57	20	4	3.15
Beneath The Planet Of The Apes	1970	Action	2.8	3.7	-1.01	8	2	3.62	33	3	5.79
Catch 22	1970	Comedy	5.6	10.8	1.06	2	2	1.71	13	2	4.46
Five Easy Pieces	1970	Drama	3.6	4.9	-0.55	4	2	3.14	7	2	3.70
Kelly's Heroes	1970	Action	4.1	5.5	-0.29	4	3	2.98	20	2	4.57
Patton	1970	Drama	5.0	7.7	0.36	8	2	2.73	14	3	4.90
Tora! Tora! Tora!	1970	Action	4.5	5.9	-0.10	5	4	2.38	18	4	4.54
Barry Lyndon	1975	Drama	9.8	11.4	2.51	4	2	2.86	6	4	3.73
Three Days Of The Condor	1975	Drama	3.4	4.9	-0.62	3	2	2.80	10	4	4.09
One Flew Over the Cuckoo's Nest	1975	Drama	3.6	4.1	-0.69	7	2	2.58	14	2	3.59
Dog Day Afternoon	1975	Drama	3.1	3.4	-0.96	6	3	3.22	15	5	4.67

Title	Year	Genre	Med	IQR	PC1	cACF	cAR	cmAR	rACF	rAR	rmAR
Jaws	1975	Adventure	3.6	4.7	-0.59	2	2	1.76	19	4	5.67
The Man Who Would Be King	1975	Action	4.9	5.8	0.01	2	2	2.07	2	2	2.81
Monty Python And The Holy Grail	1975	Comedy	2.6	3.5	-1.11	2	2	2.36	4	2	3.29
Return Of The Pink Panther	1975	Comedy	3.7	6.9	-0.19	2	2	1.71	4	2	3.56
The Rocky Horror Picture Show	1975	Comedy	3.3	4.4	-0.73	3	2	2.39	17	2	5.12
Shampoo	1975	Drama	6.1	8.6	0.86	2	1	1.79	5	4	3.40
Airplane	1980	Comedy	4.3	5.8	-0.18	3	2	1.84	3	2	3.27
Coal Miner’s Daughter	1980	Drama	5.3	9.2	0.70	4	2	2.34	5	2	3.75
The Empire Strikes Back	1980	Action	2.9	3.1	-1.08	6	2	3.45	61	5	5.45
Nine To Five	1980	Comedy	3.9	4.5	-0.52	2	2	2.07	9	4	3.26
Ordinary People	1980	Drama	3.5	4.5	-0.65	2	2	2.25	6	2	4.27
Popeye	1980	Comedy	3.3	3.5	-0.88	7	5	3.46	38	4	5.47
Stir Crazy	1980	Comedy	4.4	5.4	-0.21	1	1	1.20	2	2	2.41
Superman 2	1980	Action	2.5	2.9	-1.24	18	4	3.96	26	4	6.17
The Blue Lagoon	1980	Adventure	3.1	3.5	-0.95	5	4	2.54	6	2	4.06
Urban Cowboy	1980	Drama	3.5	4.0	-0.73	5	1	2.70	32	5	4.86
Back To The Future	1985	Action	2.7	3.9	-1.01	18	3	3.51	64	6	7.95
Cocoon	1985	Adventure	3.9	4.6	-0.51	5	2	2.99	23	4	5.18
Out Of Africa	1985	Drama	3.5	3.9	-0.75	11	3	2.94	25	6	5.55
Police Academy 2	1985	Comedy	3.0	3.8	-0.93	4	3	2.39	12	4	4.57
Rambo II	1985	Action	2.0	2.1	-1.53	13	3	2.96	64	5	6.90
Spies Like Us	1985	Comedy	2.5	3.0	-1.22	4	3	1.97	16	5	3.58
The Color Purple	1985	Drama	4.7	6.0	-0.02	4	3	3.01	6	2	3.87
Witness	1985	Drama	4.2	4.9	-0.36	3	1	2.81	15	3	4.41
Dick Tracy	1990	Action	2.8	3.2	-1.09	9	2	3.65	17	3	5.17
Die Hard 2	1990	Action	2.1	2.3	-1.46	7	2	2.69	59	5	4.52

Title	Year	Genre	Med	IQR	PC1	cACF	cAR	cmAR	rACF	rAR	rmAR
Ghost	1990	Comedy	3.4	3.9	-0.78	2	1	3.26	39	5	5.34
Goodfellas	1990	Drama	4.2	5.5	-0.26	2	1	2.08	6	2	3.05
Home Alone	1990	Comedy	3.1	3.4	-0.96	5	2	2.93	13	4	4.34
Hunt For Red October	1990	Action	4.7	5.7	-0.07	4	2	2.18	12	4	3.82
Pretty Woman	1990	Comedy	3.8	4.5	-0.56	2	2	1.80	4	2	2.83
Teenage Mutant Ninja Turtles	1990	Action	2.8	3.2	-1.09	3	2	2.38	14	2	4.16
Total Recall	1990	Action	2.4	2.8	-1.29	14	3	3.41	23	5	6.78
Ace Ventura 2	1995	Comedy	2.7	3.4	-1.09	2	2	1.92	11	4	3.79
Apollo 13	1995	Adventure	3.5	3.7	-0.78	5	2	2.61	10	4	3.76
Batman Forever	1995	Action	2.4	2.8	-1.29	22	7	3.72	31	8	6.40
Casper	1995	Comedy	4.1	5.8	-0.24	3	1	3.20	12	3	4.26
Goldeneye	1995	Action	2.3	2.7	-1.33	10	3	3.60	43	5	5.95
Jumanji	1995	Adventure	2.5	2.6	-1.29	15	4	3.57	25	5	5.29
Pocohontas	1995	Animation	2.8	2.7	-1.17	6	3	2.51	7	4	3.51
Sense And Sensibility	1995	Drama	3.8	4.2	-0.60	4	4	1.95	9	4	4.62
Toy Story	1995	Animation	2.1	2.1	-1.50	3	3	2.69	13	4	3.58
Castaway	2000	Adventure	4.5	6.7	0.03	7	4	3.43	15	5	5.67
Charlie’s Angels	2000	Action	2.0	2.1	-1.53	7	2	1.82	64	6	6.82
Dinosaur	2000	Animation	2.8	2.2	-1.26	8	2	3.26	12	4	4.36
Erin Brockovich	2000	Drama	4.2	3.8	-0.54	3	2	2.37	8	3	3.44
The Grinch Who Stole Christmas	2000	Comedy	2.6	3.0	-1.19	7	3	2.58	9	5	4.17
Scary Movie	2000	Comedy	2.3	2.6	-1.35	4	2	2.54	9	2	3.46
The Perfect Storm	2000	Adventure	3.3	3.5	-0.88	12	4	3.55	41	5	5.47
What Women Want	2000	Comedy	2.5	2.8	-1.25	10	3	2.93	24	3	4.28
X -Men	2000	Action	2.0	2.1	-1.53	14	4	3.32	68	7	6.10
Hitch	2005	Comedy	2.8	2.7	-1.17	4	2	2.46	9	3	3.61
King Kong	2005	Adventure	2.6	2.5	-1.27	33	5	4.59	48	6	5.75

Title	Year	Genre	Med	IQR	PC1	cACF	cAR	cmAR	rACF	rAR	rmAR
The Longest Yard	2005	Comedy	2.3	2.0	-1.45	5	5	1.61	22	5	3.34
Madagascar	2005	Animation	3.0	3.6	-0.96	2	2	2.33	8	2	3.67
Mr and Mrs Smith	2005	Action	2.8	3.4	-1.06	8	4	2.73	17	4	4.77
Walk The Line	2005	Drama	4.3	4.0	-0.48	3	2	2.10	2	2	2.56
The Wedding Crashers	2005	Comedy	2.4	2.4	-1.35	8	2	2.82	57	6	5.68

Sixteen films were not included in the this study: *Philadelphia Story* (1940), *Anchors Aweigh* (1945), *Mildred Pierce* (1945), *King Solomon's Mines* (1950), *Sunset Boulevard* (1950), *To Catch a Thief* (1955), *Little Big Man* (1970), *Mash* (1970), *Jewel of the Nile* (1985), *Rocky IV* (1985), *Dances with Wolves* (1990), *The Usual Suspects* (1995), *Mission Impossible 2* (2000), *Chicken Little* (2005), *Harry Potter and the Goblet of Fire* (2005), and *Star Wars: Episode III – The Revenge of the Sith* (2005).