

## Robust estimation of the mAR index of high grossing films at the US box office, 1935 to 2005

Nick Redfern

*Leeds Trinity University*

*Abstract:* The modified autoregressive (mAR) index has been proposed as a description of 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 index underestimates both the level of shot clustering in a film 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 identifies differences between genres overlooked when using the classical index.

*Key words:* autocorrelation, film editing, modified autoregressive index, robust methods

### 1. Introduction

Cutting, De Long, and Nothelfer(2010) proposed the modified autoregressive (mAR) index as a statistic of film style measuring the degree to which shots of similar duration cluster together in a motion picture. They calculate the mAR index as the intercept of the negative exponential function  $1/[1 + lag]^\beta$  fitted to the partial autocorrelation function out to lag-20 with a critical value based on the average number of shots in a motion picture from a sample. Applying this method to 150 high grossing films at the US box office released from 1935 to 2005 they identified a tendency for shots to become increasingly more correlated in length with their neighbours over time and also noted variations in the degree of shot clustering between genres.

Though the mAR index can be a useful description of film style there is good reason to doubt the validity of these conclusions. The mAR values reported by

Cutting, De Long, and Nothelfer are derived from the classical, moment-based estimator of the auto-covariance function, which is well known to be non-resistant to the presence of outliers (Ma and Genton, 2000; Marrona, Martina, and Yohai, 2006). Typically, the distribution of shot lengths in a motion picture is positively skewed and contains a number of shots of atypically long duration that adversely affect the moments this function is calculated from (i.e. the mean and variance). Consequently, estimates of the mAR index determined in this way will not accurately describe the style of a film and lead to incorrect conclusions about the nature of film style. Because these long takes are ‘true’ outliers representing the decisions of filmmakers about the arrangement of stylistic elements (staging, cinematography, editing, etc) we are interested in using robust statistical methods that perform reliably in the presence of outliers and departures from the assumptions that underpin statistical methods (Marrona, Martina and Yohai, 2006). This paper calculates robust estimates of the mAR index using a rank-based autocorrelation function (rmAR) and compares these values to the index based on classical autocorrelation function (cmAR).

## 2. Classical and rank-based autocorrelation

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

$$\gamma(h, \mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \bar{X})(X_{i+h} - \bar{X}), \quad (2.1)$$

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 semi-definite. The auto-covariance function when  $h = 0$  is equal to the variance and standardising (1) by this value gives the autocorrelation function

$$\rho(h, \mathbf{x}) = \frac{\gamma(h, \mathbf{x})}{\gamma(0, \mathbf{x})}. \quad (2.2)$$

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 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.

The above functions are not resistant to the influence of outlying data points. 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 a single outlier. Therefore, the above functions, being based on these statistics, are similarly affected by the presence of outliers. The presence of outliers in the upper tail of a shot length 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 tend to overestimate positive autocorrelation and underestimate negative autocorrelation. 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 as they appear quadratically in the denominator in (2) (see Marrona, Martina and Yohai, 2006, pp. 250-252). Consequently, the presence of outliers leads to underestimation of the strength of autocorrelation between observations in a time series. The lack of robustness of the classical auto-covariance 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 (Ma and Genton, 2000). Furthermore, 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).

Rank-based methods provide an obvious alternative to the classical functions (Ferguson, Genest and Hallin, 2000) and have been explored since Wolfowitz (1943). Although some information is lost when ranking data, rank-autocorrelation functions have a number of attractive properties: they are distribution-free while also being as powerful as classical methods (and in many cases more powerful); they are robust being relatively resistant to the influence of outliers and nonlinear distortions; and they are conceptually simple (Hallin and Puri, 1992). A rank-based approach to identifying serial dependency and periodicities in time series by Ahdesmaki, Lahdesmaki, Pearson, Huttunen and Yli-Harja (2005) calculates the autocorrelation function of a time series as

$$\hat{\rho}_s = \frac{1}{n} \frac{12}{(n-h)^2 - 1} \sum_{i=1}^{n-h} \left( R_x(i) - \frac{n-h+1}{2} \right) \left( R'_x(i) - \frac{n-h+1}{2} \right), \quad (2.3)$$

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. This function is biased and is directly comparable to the biased autocorrelation function based on (1), though it is not guaranteed to be positive semi-definite.

### 3. Methods

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 a reduced sample size of 134 films.

We calculated the classical and rank autocorrelation functions for the linearly detrended shot lengths of films in the sample to  $h = (n + 1)/2$  if  $n$  is odd and  $h = n/2$  if  $n$  is even, where  $n$  is the number of shots in a film. The rank function in (3) was calculated using unpackaged **R** functions by Bernhard Spangl,<sup>1</sup> and verified as a valid positive definite sequence in each case. The partial autocorrelations for each measure were calculated recursively using the Durbin-Levinson algorithm, and the mAR indices determined by fitting the negative exponential function  $1/[1 + h]^\beta$  to the partial autocorrelation functions for lags 0 to 20 using non-linear least squares ( $df = 20$ ). The value of an index is the intercept between the fitted function and a critical value of  $2/\sqrt{N} = 0.0611$ , where  $N = 1070$  and is the median number of shots in a film in the sample. The methods used here differ from that originally used by Cutting, De Long and Nothelfer and so our estimates of the cmAR differ from theirs.

The full set of results is in the supplementary material attached to this article. We were unable to determine the cmAR 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. It is unclear how Cutting, De Long and Nothelfer obtained mAR values for these films.

To describe trends in film style over time we fit a locally weighted (LOESS) regression smoother to the descriptive statistics. LOESS is a nonparametric method for graphically depicting the relationship between independent and dependent variables in a scatter-plot by fitting a low-order polynomial to only those observations in the neighbourhood of a point on the x-axis ( $x_i$ ) rather than fitting the trendline globally. Observations within this window are inversely weighted according to their distance from the evaluation point so that points closest to  $x_i$  have more influence on the placement of the LOESS curve than more distant observations. The degree of smoothing is controlled by the span, which specifies

---

<sup>1</sup>These functions can be accessed at: <http://lists.r-forge.rproject.org/pipermail/robust-ts-commits/2009-March/000000.html>.

the proportion of the data included in the window. In this study the span was determined separately for each time series using a generalized cross-validation procedure, and a bootstrapped 95% confidence interval gives the precision of the LOESS curve. See Jacoby (2000) for an overview of LOESS regression.

#### 4. Results

Comparing the cmAR and rmAR values calculated for each film we see that they are very different. Specifically, the classical, moment-based method underestimates both the degree to which shots of similar duration are clustered together within a film and differences in film style between films in the sample. The cmAR indices are less than the rank mAR index for 96% of films in the reduced 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. We also note the dispersion of the rank index is greater for films in the sample indicating more variation in editing style than that suggested by the cmAR index: the range and standard deviation for the rmAR index are 7.37 and 1.40, respectively, and the corresponding statistics for the cmAR index are 4.01 and 0.79. Figure 1 presents the times series plots of the classical and rank mAR indices. 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 2.a 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.

Cutting, De Long, and Nothelfer state that the trend in the cmAR index is not an artefact of decreases in the mean shot length, but because neither statistic is resistant to outliers this claim is dubious. From Figure 2.a we see that measures of location and scale for shot length distributions are strongly related, and so we combined the median and interquartile range (IQR) of the shot lengths of each film using principal components analysis to produce a new dummy variable that retains most of the information of the original variables (see Abdi and Williams (2010) for an overview of PCA). As an alternative descriptive statistic this dummy variable can be thought of as a size measure with films with a low score having a low median and low IQR a stronger tendency to more rapid editing while high-scoring films with a high median and high IQR are edited more slowly. Plotting this score against year of release (Figure 2.b) 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 2.b for the 1940s and 1950s indicates much greater stylistic variation between films from those decades; while the trendline after the 1950s shows that high grossing films have converged to a single, rapidly-edited style. These trends correspond to the trends in shot clustering in Figure 1.b, indicating that changes in the rapidity of cutting and in the degree of shot clustering are a part of the same overall transformation of film style. Again, this is a relationship overlooked when using non-robust methods.

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. The beanplots in Figure 3 present the classical and rank mAR indices sorted by the above genre categories, and the differences in the level and dispersion of the indices are stark. To compare the distribution of indices for each genre we performed a Kruskal-Wallis ANOVA test, and pairwise post-hoc Dunn tests assuming 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. All reported test statistics are corrected for ties. The omnibus test for the cmAR index shows a statistically significant difference between genres ( $\chi^2(4) = 28.50, p = < 0.01$ ), with significant pairwise differences between the action genre and comedy films ( $Z = 4.09$ ), and drama films ( $Z = 4.49$ ). For the rmAR index we also see a statistically significant difference ( $\chi^2(4) = 38.61, p = < 0.01$ ), with pairwise differences between action films and the animation ( $Z = 3.22$ ), comedy ( $Z = 4.59$ ), and drama ( $Z = 5.22$ ) genres and between adventure films and comedy ( $Z = 3.15$ ) and drama ( $Z = 3.60$ ) films. (The difference between adventure films and animated films was not quite significant [ $Z = 2.46$ ]). Using the classical mAR index would therefore lead us to miss key differences between the style of films in particular genres.

## 5. 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 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 due to the presence of outliers in shot length data.

Because the power spectral density of a time series is the Fourier transform of its autocorrelation function the problem of outliers in shot length data will be transmitted to the spectral analysis of time series. This raises questions about Cutting, De Long, and Nothelfer's claim that the editing of films in the sample shows an increasing tendency to be well fitted by a  $1/f$  noise pattern over time. This should not lead us to reject the idea that a  $1/f$  noise pattern is a characteristic of film editing. On the contrary, given that the rank mAR index indicates the correlation between shots has been underestimated it is likely the role of  $1/f$  noise in entraining viewers' attention in the cinema has also been underestimated due to the incorrect identification of white noise in the composite power spectra of each film. Future research on the relationship between film style and attention will therefore need to employ robust methods of spectral analysis (see, for example, Spangl, 2008) to determine if this is the case.

### Acknowledgements

I would like to thank Mike Baxter for a timely observation on an early draft of this article.

### References

- Abdi, H. and Williams, L. J. (2010). Principal components analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**, 433-459.
- Lahdesmaki, H., Pearson, R., Huttunen, H., and Yli-Harja, O. (2005). FRobust detection of periodic time series measured from biological systems. *BMC Bioinformatics* **6**, 117.
- Hallin, M., and Puri M. L. (1992). *Statistical Factor Analysis and Related Methods: Theory and Applications*. Wiley, New York.
- Chatfield, C. (2004). *The Analysis of Time Series: An Introduction*, 6th edition. Chapman & Hall/CRC, Boca Raton, FL.

- Cutting, J. E., De Long, J. E, and Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film. *Psychological Science* **21**, 432-439.
- Ferguson, T. S., Genest, C., & Hallin, M. (2000). Kendall's tau for serial dependence. *Canadian Journal of Statistics* **28**, 587-604.
- Jacoby, W. G. (2000). Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies* **19**, 577-613.
- Ma, Y. and Genton, M. G. (2000). Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis* **21**, 663-684.
- Maronna, R., Martin, D., and Yohai, V. (2006). *Robust Statistics: Theory and Method*. John Wiley and Sons, Chichester.
- Spangl, B. (2008). *On Robust Spectral Density Estimation*. Ph.D. Thesis, Vienna Technical University.
- Wald, A., and Wolfowitz, J. (1943). An exact test of randomness in the non-parametric case based on serial correlation, *Annals of Mathematical Statistics* **14**, 378-388.



## Supplementary material

*Key:* Med = median shot length (in seconds), IQR = interquartile range (in seconds), PC1 = first principle component score, cmAR = classical modified autoregressive index, rmAR = rank modified autoregressive index

<b>Title</b>	<b>Year</b>	<b>Genre</b>	<b>Med</b>	<b>IQR</b>	<b>PC1</b>	<b>cmAR</b>	<b>rmAR</b>
<i>A Night at the Opera</i>	1935	Comedy	4.0	6.2	-0.21	N/A	4.87
<i>A Tale of Two Cities</i>	1935	Drama	5.2	7.0	0.31	2.49	4.94
<i>Anna Karenina</i>	1935	Drama	6.5	8.0	0.89	0.73	2.61
<i>Captain Blood</i>	1935	Action	4.4	4.8	-0.31	2.70	3.89
<i>Les Miserables</i>	1935	Drama	5.0	7.4	0.31	2.29	4.34
<i>Mutiny on the Bounty</i>	1935	Drama	4.5	7.6	0.18	2.51	4.83
<i>The 39 Steps</i>	1935	Drama	4.1	6.5	-0.13	2.67	3.31
<i>The Informer</i>	1935	Drama	7.4	9.0	1.34	1.62	2.41
<i>Top Hat</i>	1935	Comedy	5.4	7.3	0.42	3.09	3.93
<i>Westward Ho</i>	1935	Drama	4.0	5.7	-0.29	2.43	4.55
<i>Fantasia</i>	1940	Animation	6.5	8.8	1.02	3.33	4.45
<i>Foreign Correspondent</i>	1940	Drama	4.3	5.9	-0.16	1.92	3.92
<i>Grapes of Wrath</i>	1940	Drama	6.8	9.0	1.15	1.48	2.45
<i>Pinocchio</i>	1940	Animation	4.0	4.4	-0.51	2.36	2.45
<i>Rebecca</i>	1940	Drama	4.8	7.5	0.26	1.27	2.06
<i>Santa Fe Trail</i>	1940	Action	4.3	6.3	-0.10	4.36	5.92
<i>The Great Dictator</i>	1940	Comedy	9.2	14.3	2.80	N/A	1.07
<i>The Letter</i>	1940	Drama	7.1	13.0	1.91	1.17	2.35
<i>Thief of Bagdad</i>	1940	Adventure	3.6	4.0	-0.70	2.07	4.19
<i>Bell's of St. Mary's</i>	1945	Drama	6.0	9.0	0.89	1.62	2.47
<i>Blood On The Sun</i>	1945	Action	9.1	12.9	2.53	2.00	2.85
<i>Brief Encounter</i>	1945	Drama	9.2	12.9	2.57	1.65	3.24
<i>Detour</i>	1945	Drama	12.3	16.2	4.11	N/A	0.58
<i>In Pursuit to Algiers</i>	1945	Adventure	5.8	13.3	1.54	0.91	1.75
<i>Leave Her to Heaven</i>	1945	Drama	6.0	9.1	0.91	1.93	2.00
<i>Lost Weekend</i>	1945	Drama	8.2	11.6	2.03	1.67	2.75
<i>Spellbound</i>	1945	Drama	5.1	9.4	0.67	1.75	3.46
<i>All About Eve</i>	1950	Drama	4.9	8.4	0.44	2.35	1.83
<i>Annie Get Your Gun</i>	1950	Comedy	5.0	13.2	1.26	4.43	7.57
<i>Born Yesterday</i>	1950	Comedy	6.9	16.5	2.42	1.93	2.49
<i>Cheaper By The Dozen</i>	1950	Comedy	7.3	14.4	2.20	1.60	2.51
<i>Cinderella</i>	1950	Animation	3.1	2.8	-1.06	3.18	3.18
<i>Harvey</i>	1950	Comedy	13.2	26.7	6.13	0.87	1.23
<i>The Asphalt Jungle</i>	1950	Drama	6.6	8.1	0.94	1.91	1.84
<i>The Flame and the Arrow</i>	1950	Action	4.0	5.1	-0.39	3.50	4.90
<i>Battle Cry</i>	1955	Drama	6.0	7.6	0.66	1.72	2.75

<b>Title</b>	<b>Year</b>	<b>Genre</b>	<b>Med</b>	<b>IQR</b>	<b>PC1</b>	<b>cmAR</b>	<b>rmAR</b>
<i>East Of Eden</i>	1955	Drama	5.1	7.8	0.41	1.91	2.83
<i>Lady And The Tramp</i>	1955	Animation	3.8	3.6	-0.70	2.99	2.84
<i>Mr Roberts</i>	1955	Comedy	6.5	8.7	1.01	1.41	2.00
<i>Night Of The Hunter</i>	1955	Drama	4.9	6.9	0.19	1.67	2.90
<i>Rebel Without A Cause</i>	1955	Drama	4.8	6.0	0.01	3.28	4.54
<i>Seven Year Itch</i>	1955	Comedy	11.3	26.1	5.42	1.27	1.26
<i>The Ladykillers</i>	1955	Comedy	5.1	7.1	0.29	1.36	2.00
<i>The Trouble With Harry</i>	1955	Comedy	5.1	7.2	0.31	1.03	1.25
<i>Butterfield 8</i>	1960	Drama	6.1	10.7	1.21	2.23	3.11
<i>Exodus</i>	1960	Action	13.8	20.3	5.27	2.93	4.42
<i>Inherit the Wind</i>	1960	Drama	7.5	10.9	1.69	1.18	1.57
<i>The Magnificent Seven</i>	1960	Adventure	4.5	5.7	-0.13	2.16	4.45
<i>Ocean's 11</i>	1960	Comedy	7.5	9.4	1.44	2.70	1.94
<i>Peeping Tom</i>	1960	Drama	5.2	8.2	0.50	1.40	2.74
<i>Spartacus</i>	1960	Action	4.9	6.8	0.18	2.28	3.68
<i>Swiss Family Robinson</i>	1960	Adventure	3.1	3.7	-0.91	3.92	5.66
<i>The Apartment</i>	1960	Comedy	8.3	15.7	2.74	2.62	3.58
<i>The Time Machine</i>	1960	Adventure	3.4	5.0	-0.60	2.10	3.78
<i>Dr Zhivago</i>	1965	Drama	5.9	7.2	0.57	1.66	2.45
<i>Flight Of The Phoenix</i>	1965	Adventure	3.3	3.5	-0.88	3.04	5.26
<i>Those Magnificent Men in Their Flying Machines</i>	1965	Adventure	3.9	4.4	-0.54	2.03	4.65
<i>Help</i>	1965	Comedy	2.6	2.8	-1.22	2.02	2.96
<i>Shenandoah</i>	1965	Drama	4.0	5.2	-0.38	2.51	2.75
<i>Sound Of Music</i>	1965	Drama	4.2	5.1	-0.33	2.64	4.73
<i>That Darn Cat</i>	1965	Comedy	3.4	4.0	-0.77	0.81	2.32
<i>The Great Race</i>	1965	Action	4.0	6.1	-0.23	2.73	5.29
<i>Thunderball</i>	1965	Action	2.5	2.7	-1.27	3.22	5.40
<i>What's New Pussycat</i>	1965	Comedy	4.2	6.4	-0.11	2.11	4.20
<i>Airport</i>	1970	Drama	4.7	6.2	0.01	1.79	2.50
<i>Aristocats</i>	1970	Animation	3.3	2.7	-1.01	2.57	3.15
<i>Beneath The Planet Of The Apes</i>	1970	Action	2.8	3.7	-1.01	3.62	5.79
<i>Catch 22</i>	1970	Comedy	5.6	10.8	1.06	1.71	4.46
<i>Five Easy Pieces</i>	1970	Drama	3.6	4.9	-0.55	3.14	3.70
<i>Kelly's Heroes</i>	1970	Action	4.1	5.5	-0.29	2.98	4.57
<i>Patton</i>	1970	Drama	5.0	7.7	0.36	2.73	4.90
<i>Tora! Tora! Tora!</i>	1970	Action	4.5	5.9	-0.10	2.38	4.54
<i>Barry Lyndon</i>	1975	Drama	9.8	11.4	2.51	2.86	3.73
<i>Three Days Of The Condor</i>	1975	Drama	3.4	4.9	-0.62	2.80	4.09
<i>Dog Day Afternoon</i>	1975	Drama	3.1	3.4	-0.96	3.22	4.67

<b>Title</b>	<b>Year</b>	<b>Genre</b>	<b>Med</b>	<b>IQR</b>	<b>PC1</b>	<b>cmAR</b>	<b>rmAR</b>
<i>Jaws</i>	1975	Adventure	3.6	4.7	-0.59	1.76	5.67
<i>The Man Who Would Be King</i>	1975	Action	4.9	5.8	0.01	2.07	2.81
<i>Monty Python And The Holy Grail</i>	1975	Comedy	2.6	3.5	-1.11	2.36	3.29
<i>One Flew Over the Cuckoo's Nest</i>	1975	Drama	3.6	4.1	-0.69	2.58	3.59
<i>Return Of The Pink Panther</i>	1975	Comedy	3.7	6.9	-0.19	1.71	3.56
<i>The Rocky Horror Picture Show</i>	1975	Comedy	3.3	4.4	-0.73	2.39	5.12
<i>Shampoo</i>	1975	Drama	6.1	8.6	0.86	1.79	3.40
<i>Airplane</i>	1980	Comedy	4.3	5.8	-0.18	1.84	3.27
<i>Coal Miner's Daughter</i>	1980	Drama	5.3	9.2	0.70	2.34	3.75
<i>The Empire Strikes Back</i>	1980	Action	2.9	3.1	-1.08	3.45	5.45
<i>Nine To Five</i>	1980	Comedy	3.9	4.5	-0.52	2.07	3.26
<i>Ordinary People</i>	1980	Drama	3.5	4.5	-0.65	2.25	4.27
<i>Popeye</i>	1980	Comedy	3.3	3.5	-0.88	3.46	5.47
<i>Stir Crazy</i>	1980	Comedy	4.4	5.4	-0.21	1.20	2.41
<i>Superman 2</i>	1980	Action	2.5	2.9	-1.24	3.96	6.17
<i>The Blue Lagoon</i>	1980	Adventure	3.1	3.5	-0.95	2.54	4.06
<i>Urban Cowboy</i>	1980	Drama	3.5	4.0	-0.73	2.70	4.86
<i>Back To The Future</i>	1985	Action	2.7	3.9	-1.01	3.51	7.95
<i>Cocoon</i>	1985	Adventure	3.9	4.6	-0.51	2.99	5.18
<i>Out Of Africa</i>	1985	Drama	3.5	3.9	-0.75	2.94	5.55
<i>Police Academy 2</i>	1985	Comedy	3.0	3.8	-0.93	2.39	4.57
<i>Rambo II</i>	1985	Action	2.0	2.1	-1.53	2.96	6.90
<i>Spies Like Us</i>	1985	Comedy	2.5	3.0	-1.22	1.97	3.58
<i>The Color Purple</i>	1985	Drama	4.7	6.0	-0.02	3.01	3.87
<i>Witness</i>	1985	Drama	4.2	4.9	-0.36	2.81	4.41
<i>Dick Tracy</i>	1990	Action	2.8	3.2	-1.09	3.65	5.17
<i>Die Hard 2</i>	1990	Action	2.1	2.3	-1.46	2.69	4.52
<i>Ghost</i>	1990	Comedy	3.4	3.9	-0.78	3.26	5.34
<i>Goodfellas</i>	1990	Drama	4.2	5.5	-0.26	2.08	3.05
<i>Home Alone</i>	1990	Comedy	3.1	3.4	-0.96	2.93	4.34
<i>Hunt For Red October</i>	1990	Action	4.7	5.7	-0.07	2.18	3.82
<i>Pretty Woman</i>	1990	Comedy	3.8	4.5	-0.56	1.80	2.83
<i>Teenage Mutant Ninja Turtles</i>	1990	Action	2.8	3.2	-1.09	2.38	4.16
<i>Total Recall</i>	1990	Action	2.4	2.8	-1.29	3.41	6.78
<i>Ace Ventura 2</i>	1995	Comedy	2.7	3.4	-1.09	1.92	3.79
<i>Apollo 13</i>	1995	Adventure	3.5	3.7	-0.78	2.61	3.76
<i>Batman Forever</i>	1995	Action	2.4	2.8	-1.29	3.72	6.40
<i>Casper</i>	1995	Comedy	4.1	5.8	-0.24	3.20	4.26
<i>Goldeneye</i>	1995	Action	2.3	2.7	-1.33	3.60	5.95
<i>Jumanji</i>	1995	Adventure	2.5	2.6	-1.29	3.57	5.29
<i>Pocohontas</i>	1995	Animation	2.8	2.7	-1.17	2.51	3.51

<b>Title</b>	<b>Year</b>	<b>Genre</b>	<b>Med</b>	<b>IQR</b>	<b>PC1</b>	<b>cmAR</b>	<b>rmAR</b>
<i>Sense And Sensibility</i>	1995	Drama	3.8	4.2	-0.60	1.95	4.62
<i>Toy Story</i>	1995	Animation	2.1	2.1	-1.50	2.69	3.58
<i>Castaway</i>	2000	Adventure	4.5	6.7	0.03	3.43	5.67
<i>Charlie's Angels</i>	2000	Action	2.0	2.1	-1.53	1.82	6.82
<i>Dinosaur</i>	2000	Animation	2.8	2.2	-1.26	3.26	4.36
<i>Erin Brockovich</i>	2000	Drama	4.2	3.8	-0.54	2.37	3.44
<i>The Grinch Who Stole Christmas</i>	2000	Comedy	2.6	3.0	-1.19	2.58	4.17
<i>Scary Movie</i>	2000	Comedy	2.3	2.6	-1.35	2.54	3.46
<i>The Perfect Storm</i>	2000	Adventure	3.3	3.5	-0.88	3.55	5.47
<i>What Women Want</i>	2000	Comedy	2.5	2.8	-1.25	2.93	4.28
<i>X-men</i>	2000	Action	2.0	2.1	-1.53	3.32	6.10
<i>Hitch</i>	2005	Comedy	2.8	2.7	-1.17	2.46	3.61
<i>King Kong</i>	2005	Adventure	2.6	2.5	-1.27	4.59	5.75
<i>The Longest Yard</i>	2005	Comedy	2.3	2.0	-1.45	1.61	3.34
<i>Madagascar</i>	2005	Animation	3.0	3.6	-0.96	2.33	3.67
<i>Mr and Mrs Smith</i>	2005	Action	2.8	3.4	-1.06	2.73	4.77
<i>Walk The Line</i>	2005	Drama	4.3	4.0	-0.48	2.10	2.56
<i>The Wedding Crashers</i>	2005	Comedy	2.4	2.4	-1.35	2.82	5.68

Sixteen films were not included in the 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).

Received December 13, 2012; accepted August 19, 2013.

Nick Redfern  
 Department of Journalism, Media, and Business,  
 Leeds Trinity University, Leeds, LS18 5HD, United Kingdom  
 n.redfern@leedstrinity.ac.uk

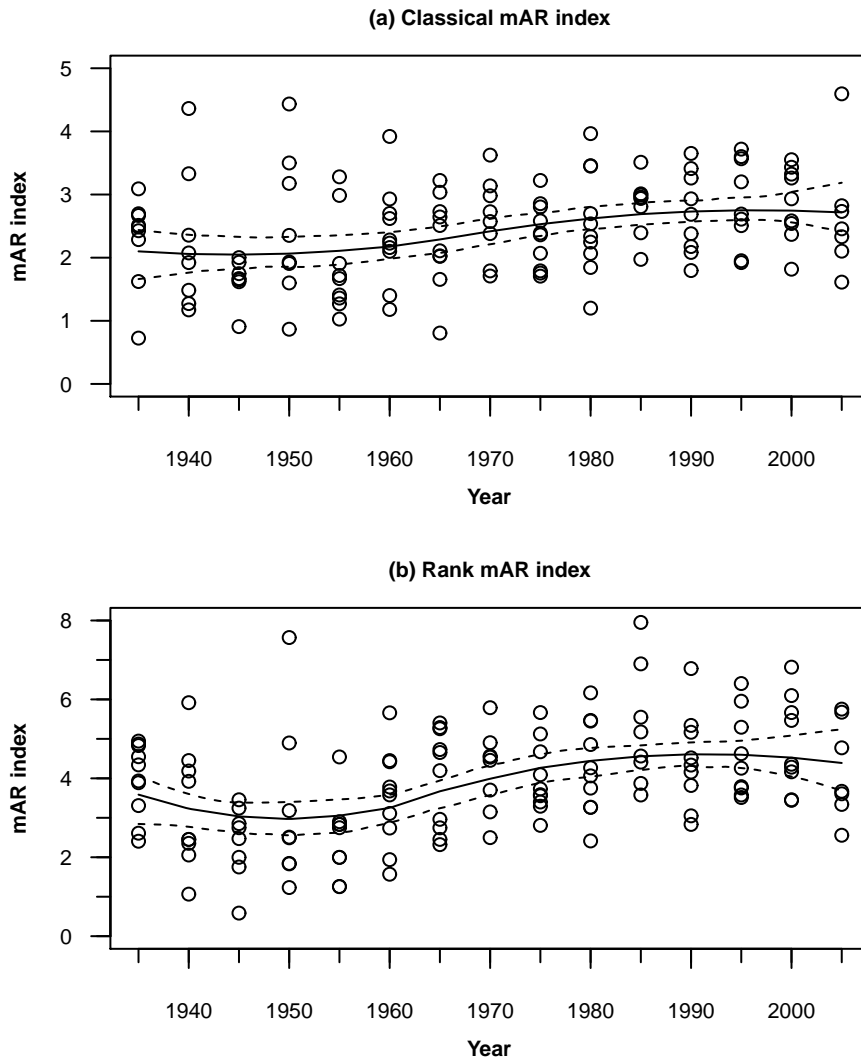


Figure 1: Scatterplot of (a) classical and (b) rank mAR indices against year of release with fitted LOESS trendlines and bootstrapped 95% confidence intervals.

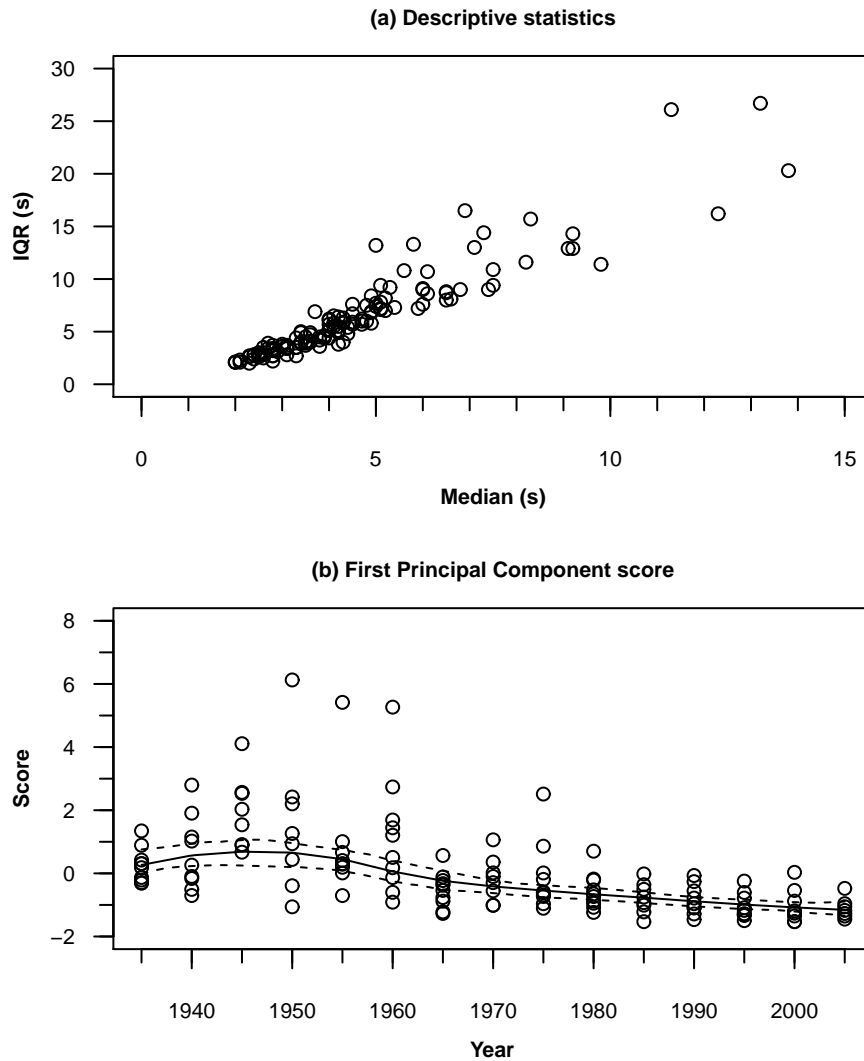


Figure 2: (a) Scatterplot of median shot lengths and interquartile ranges for high grossing films at the US box office, 1935 to 2005 ( $N = 134$ ). (b) Scatterplot of first principal component score against year of release with fitted LOESS trendline and bootstrapped 95% confidence interval. The dummy variable is a combination of these descriptive statistics, and accounts for 96.5% of the variance of the original variables. PC scores are calculated from the correlation matrix, and the correlation of both the median and IQR with the first principal component is 0.98. The loadings for the median and the IQR on the first principal component is 0.7071 for both variables.

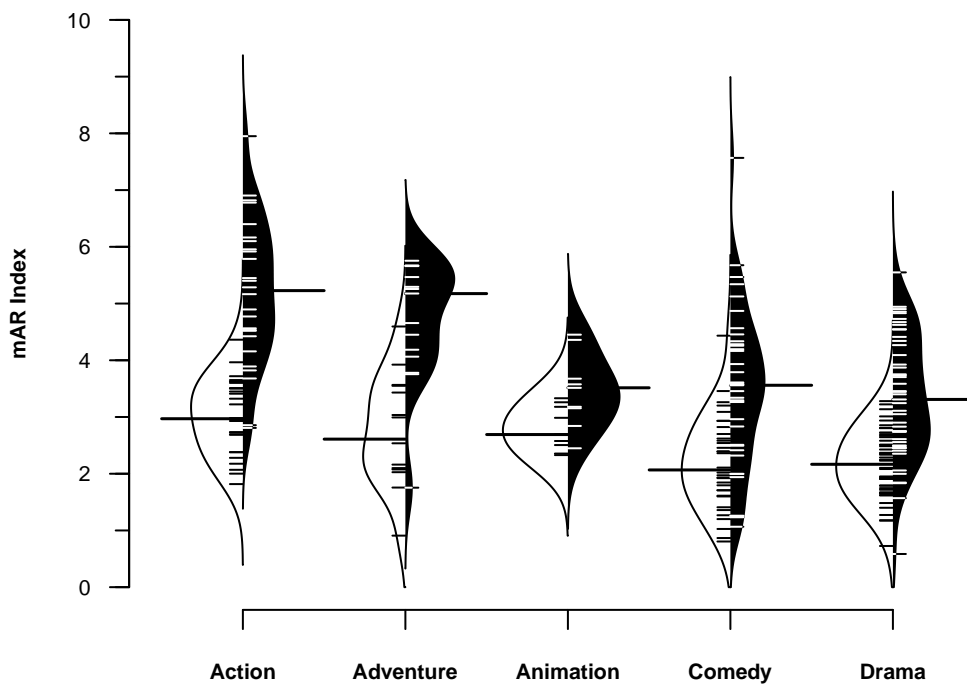


Figure 3: Classical (left) and rank (right) mAR indices of high grossing films at the US box office, 1935 to 2005, sorted by genre. The average beanlines are set to the sample medians.