

# Applying logisitic regression framework to predict Academy Awards

Christopher Lee

The Academy Awards are the most important event in the Hollywood film industry. Winning an Academy Award is not only prestigious but financially lucrative for a film's box office revenue, a nominee's future earnings and the large speculative market that has developed around the annual award show. This study will attempt to model the outcome of the Academy Awards with data from a period of 1970 to 2013 and identify significant covariates for Academy Award success for the four acting categories, the directing award and the best picture award. The paper will conclude with the model's predictions for the 2014 Academy Award winners and review the predictions against the resulting winners.

## Introduction

The Hollywood film industry has well-defined seasonal activity. The summer season is dedicated towards large grossing Blockbuster films, while the late fall (December and November) is the release target for films looking to be competitive during the awards season. The award season begins January and concludes with the most important event of the industry: the Academy Awards. The Academy Awards stand apart from the other award shows as the final and definitive award show of the year. For reasons of prestige, and perceived increases in box-office revenue, campaigning for Oscars is a huge and expensive endeavor. There is also intense public intrigue and well-endowed prediction markets, as the Academy Awards is a widely viewed televised event and one of the largest non-sporting events for speculative betting.

This paper discusses the challenge of identifying covariates and variables which have an effect on the odds of winning an Academy Award. There have been several papers which have addressed this topic. Most notably, a study with similar motivations and methods appeared before the 2008 Oscar Ceremonies (Pardoe & Simonton, 2008). Other papers attempt other modelling techniques for Academy Award prediction (Silver, 2013), (Bernard, 2005), (Deepak Pathaka, 2013), (Ghomi, Shirzadi, & Movassaghi, 2013) (Krauss, Nann, Simon, Fischbach, & Gloor, 2008) which range from analyzing social networks to polling data and even Twitter activity.

This paper makes no claim to have found or imply a set of characteristics that produce the 'best film' nor does this study present models outlining voting behaviour among the members of the Academy of Motion Picture Arts and Sciences. The purpose of this paper is solely to identify covariates which change the odds of winning a select group of awards at the Oscar ceremonies. In doing so, we set aside any claims on 'the best film of the year' or having identified voting preferences from AMPAS members. This study merely address what factors can contribute to Oscar success,

whether or not the Oscars themselves are fair meritocracies or biased elections.

## Data

To address this question, this paper uses data that has been web-scraped from several sources of publically accessible web domains, including [imdb.com](http://imdb.com), [boxofficemojo.com](http://boxofficemojo.com) and [metacritic.com](http://metacritic.com). This study has created a dataset with several quantitative and qualitative variables. The data records box office revenue and production budgets, nominee information (age, past Oscar record, ethnicity, etc.), film details (genre, rating, topic, etc.) and Oscar results from the years 1970 to 2013. In addition this study gathers data from the most relevant preceding film awards, that is, the Golden Globe Awards, the British Academy of Film and Television Arts Awards (BAFTA), the Screen Actors Guild Awards (SAG), the Directors Guild Awards (DGA), the Writer's Guild Awards (WGA), the American Cinema Editors Award (Eddie) and the Critics Choice Awards. An effort will be made to preserve the limited sample size of this data wherever possible. Many variables are truncated as they were only recently created/recorded. Given a reasonable level of statistical power, it would not be appropriate to introduce these variables into the models for many years.

## Descriptives

Several key traits among nominees are identified as having potentially strong effects on the odds of winning an Oscar. This study has collected data on the MPAA rating, release date, genre and subject matter of the nominee pool. Many of these classifications are interesting, in that they present an image of the general tastes and preferences of the Academy. It is unclear whether the Academy favors films that have late fall release dates or if the most worthy films are released during those dates. Regardless, the months of November and December claim a disproportionate share of Oscar nominees and winners, in terms of the films' month of release. Equally

Table 1  
Mean and proportions of descriptives

	Nominees	Winners
<b>Genre</b>		
drama	0.909	0.898
comedy	0.213	0.200
biopic	0.220	0.238
rom	0.234	0.238
action	0.094	0.087
<b>Film Traits</b>		
Runtime (min.)	125.217	130.626
TIFF premiere	0.193	0.170
Nov./Dec. release	0.503	0.547
Domestic.Gross <sup>a</sup>	59.581	88.640
Budget <sup>a</sup>	28.245	26.494
adapted	0.592	0.574
R.rated	0.576	0.574
Hot topic <sup>b</sup>	0.196	0.249
Dark topic <sup>c</sup>	0.220	0.272
<b>Nominee Traits</b>		
age	44.915	45.582
white	0.909	0.900
black	0.050	0.055
american	0.715	0.732
british	0.149	0.127

<sup>a</sup>Measured in million USD <sup>b</sup>Contains the following keywords: "racism, slavery, gay rights, abortion, poverty, aids, hiv or homophobia" <sup>c</sup>Contains the following keywords: "terrorism, rape, gang violence, sex trade, human trafficking, torture, heroin, cocaine, serial killer or mass murderer"

as telling, is the proportion of genres and ratings where we see a very strong results for dramas and R-rated films.

We examine the proportions and means of these characteristics and any differences between the sample of nominees, and sample of winners. This study seeks to predict the winner selection rather than the nominee selection therefore many of these traits are of less use to us than the differential between nominees and winners shown in Table 1.

Many of the more jarring descriptives such as ninety percent of the nominee pool being dramas are less impactful because the respective winner pool has a similarly steep mean. We see very little variation between the means of nominees vs. winners. This makes it difficult to distinguish whether or not these traits affect the odds of winning an Oscar or merely affecting the odds of being included in the sample (being nominated). It seems that, in terms of film/nominee traits, the sample of nominees and winners are mostly indistinguishable by the criterion we have considered thus far. Their discrimination for the winner selection may be carried out by means which we do not have the ability to capture (i.e.

personal taste, artistic merit, significance, quality ect.)

### Co-movement of the Award Season

The award season for the film industry, culminating in the Oscar ceremony, is closely followed and valued as important events which affect nominees' odds at the Academy Awards. Preceding film award shows such as the Golden Globes, BAFTA awards and recently the Critic's Choice awards all confer equivalent awards to the Oscars while the professional guild award shows (WGA, SAG, DGA, PGA etc.) specialize in awards inherent to their respective crafts. Even so, the co-movement between all these events and the Oscars are incredible, considering the voting bodies of each event differ in composition, size and selection procedures. Academy Award winners are voted upon by members of the Academy for Motion Picture Arts and Sciences (AMPAS) of an electorate of approximately 6,000 voters. AMPAS uses a complex preferential system and the electoral equivalent of a *single transferable vote* for the nomination process and a plurality vote to select the winner - except for Best Picture, which continues a preferential voting system (Gehrlein & Kher, 2004). The Golden Globes are voted on by members of the Hollywood Foreign Press Association that normally has fewer than 100 active members eligible to vote - none of which are members of the voting body of AMPAS. The BAFTAs are voted on by 6,000 with plurality vote with a sizeable cross-over between voting bodies, and professional guild awards are decided by guild members with plurality vote.

The Golden Globes are the furthest removed from the Oscars but it is perceived as the most relevant event preceding the Academy Awards. In Table 2 we examine the historical performance of all the major award shows for the 6 races to see how often precursor award-winners wins the respective Oscar prize. In all categories, the Golden Globe select the Oscar winner roughly half the time while the Directors Guild Awards boast an eighty-two and eighty-six percent success rate for predicting the Oscar directing and Picture races respectively.

Further still, in the event of consensus, an Oscar winner is all but determined. In history, the Golden Globes, BAFTAs, Critic's Choice and SAG awards have agreed upon seven actors, three actresses, four supporting actors and five supporting actresses. Only two such times (Russell Crowe in *A Beautiful Mind* and Jessica Chastain in *The Help*) has the consensus choice failed to win the Oscar. Meanwhile, the ACE Eddies and DGA have agreed on nineteen directors (all of which went on to win the Oscar for Best Directing) and twenty films where all but *Saving Private Ryan* won the Best Picture race at the Oscars. Daniel Day-Lewis and Steven Spielberg remain the only two individuals who have twice been the consensus choice.

Table 2  
Accuracy of precursor awards for future Oscar winners

Oscar Won	Globes	BAFTA	SAG	Critics'
<b>Leading</b>				
Actor	0.61	0.41	0.79	0.67
Actress	0.55	0.41	0.68	0.44
<b>Supporting</b>				
Actor	0.66	0.23	0.58	0.44
Actress	0.48	0.34	0.63	0.50
Director	0.52	0.30		0.78
Picture	0.51	0.38		0.58
	DGA	Eddie	WGA	PGA
Director	0.86	0.48	0.48	0.63
Picture	0.82	0.51	0.53	0.71

### Methodology

This paper will focus on modelling and prediction for the six most important award races of the Oscars: Best Actor in a Leading Role, Best Actress in a Leading Role, Best Actor in a Supporting Role, Best Actress in a Supporting Role, Best Directing and Best Picture. To do this, this we employ a logistic regression framework to identify key predictors of Academy Award success. The logistic regression framework has been found appropriate for this problem, as the dependent variable is a binary vector reporting a value of 1 if a film has won an Academy Award in one of six categories, and 0 otherwise. On the issue of dependence between observations, we follow the precedence of previous literature on Academy Award prediction. Given an award category, every year there must necessarily exist one observation which reports a 1 which forces all other observations which report 0, because the awards are competitive and exclusive. By removing any temporal variables, we remove the yearly dimension and avoid observational dependence. We treat the remainder of the analysis as though there are 220 nominees all in a stationary time period, and we are classifying 44 of them as winners simultaneously.

### Estimation

Preliminary models first include all relevant awards, given the high differences in the proportions of Oscar nominees and Oscar winners. Attempts are made to classify winners according to film/nominee traits, however they are not nearly as strong. In variable selection, this study heeds a series of commonly held expert/pundit views on the Academy Awards. These views are as follows: (1) Serious Oscar contenders for Best Picture have secured nominations in Editing and most importantly, Directing. (2) Guild awards are the strongest predictors of Oscar winners in guild-specific

racers. (3) Biographical feature films are highly favored by the Academy. (4) November/December release dates are indicative of highly competitive films. (5) The four acting races are largely similar to one another.

We first fit saturated model including all variables except those which have very high missing value rates (SAG, Critic Choice, budget). We must resign ourselves to the fact that for reasons of inadequate sample size, we cannot hope to include SAG and maintain an appropriate level statistical power despite exploratory analysis and a priori reasoning showing the SAG to be a crucial predictor of Oscar success.

### Model Selection

We consider several criteria for model selection including the receiver operator characteristic (*ROC*) curve and its *AUC*, which measures the models's true positives (sensitivity) against its false positives (1 - specificity). We also consider information theoretic criteria such as the Bayesian information criterion (*BIC*), Aikake information criterion (*AIC*) and the second-order *AIC* (*AIC<sub>c</sub>*) which asymptotically converges to *AIC* but is proven to correct the systematic overfitting of *AIC* (Kass & Raftery, 1995) in small samples (Burnham & Anderson, 2004). These criteria will be used to perform backwards stepwise regression which minimizes the information criteria.

Which criteria should be used and when has been discussed in model selection literature. Very generally, the *BIC* will select the model with the greatest posterior probability (the *true* model in the set) while the *AIC* will find the model that minimizes the expected Kullback-Leibler Information loss (the best approximation to a full reality model). We have strong a priori reason to believe that the *true* model, even if it exists, is not captured in our model set (exclusion of SAG is one of many reasons). But others (Burnham & Anderson, 2004) argue the *BIC*'s controversial assumption that *there exists a true model from which data was generated* is exaggerated and only holds asymptotically at that. Burnham & Anderson go on to state that the correct criteria should not be determined by philosophical dichotomies between frequentist and Bayesian perspectives but by measurable performance. To this end we will select *BIC*, *AIC* and *AIC<sub>c</sub>* and choose the criteria whose subsequent models demonstrate the lowest classification error. The more conservative and parsimonious *BIC* models are shown in Table 3 while the larger *AIC<sub>c</sub>* models are in the Appendix.

### Results

The regression analysis follows conventional wisdom as it relates to the relationship between film award shows. Indeed, there appears to be considerable co-movement between the Golden Globes Awards, the BAFTA awards and the Oscars. An important feature identified by the models is the spurious relationship between winning incidence for Best Picture and

Table 3  
Backwards stepwise BIC selection

	Dependent variable:					
	Best Actor	Best Actress	Best Supporting Actor	Won Oscar Best Supporting Actress	Best Director	Best Picture
globes	10.489*** (4.689,24.523)	8.223*** (3.757,18.485)	19.130*** (8.455,46.615)	4.294*** (1.989,9.356)		
bafta	4.411*** (1.781,11.064)	4.992*** (2.105,11.947)		5.239*** (2.160,12.874)		7.462*** (1.698,36.338)
script.nom	3.882** (1.461,11.969)					
comedy			3.536*** (1.361,9.343)			
edit.nom			3.003** (1.296,7.279)	2.867*** (1.340,6.144)		9.056** (1.517,79.095)
eddie					7.058** (1.457,34.582)	6.584*** (1.732,27.979)
dga					321.080*** (83.470,1,780.902)	117.394*** (32.390,605.544)
wga						6.181*** (1.758,26.416)
Constant	0.030*** (0.009,0.077)	0.092*** (0.052,0.152)	0.040*** (0.016,0.087)	0.082*** (0.043,0.142)	0.023*** (0.007,0.055)	0.001*** (0.0001,0.011)
Observations	220	221	221	221	221	239
BIC	176.733	188.641	177.244	202.455	84.479	106.654
Area under ROC	0.856	0.794	0.838	0.768	0.944	0.974

Note: Values are in terms of odds ratios and 95% exponentiated confidence intervals. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Oscar nominations for Best Directing, Editing and Screenplay. As only four nominees in history have won Best Picture without first being nominated for Best Directing (Argo being the most recent) and only three nominees have won without a Best Editing nomination (last in 1981), popular belief is that director and editing nominations are reliable signals of a solid Best Picture contender. Indeed, preliminary models of the Best Picture race agree but upon the introduction of the DGA and Eddie variables (which award for director and editing respectively), the previous directing and editing nomination variables have large changes in coefficient values, or else lose predictive power. We can safely identify the DGA and Eddies as confounding factors for the widely held belief of direct causal links between Best Picture and nominations for Best Directing and Best Editing. By supplementing the director nomination and editing nomination with their respective confounders, we find substantial increases in all measures of goodness-of-fit.

Another surprising fact is that among the four acting races, the BIC models show homogeneity between Lead Actor and Lead Actress but the Supporting races differ greatly in effect size and composition. The differences between Lead and Supporting races have often been blurred in the media but the analysis shows the races may be far from similar. This may follow from a belief that Supporting awards are less prestigious than Lead awards, though distinguishing between Lead and Supporting roles is often vague and unfounded. For ex-

ample, Al Pacino was nominated for a Supporting Role in *The Godfather* despite having more screentime than co-star and eventual Lead Actor winner Marlon Brando. This contrasts with Beatrice Straight, who secured a Supporting Actress Oscar in *The Network* with a total of five minutes and forty seconds of screentime - four percent of the runtime.

### Validation

For other studies, internal validation would be based off a classification rule, say  $\hat{p} > .5$ , to determine the necessary predicted response probability to be considered a predicted event. In our models, we must define a predicted event differently. Because only one award may (and must) be conferred per category each year, we define a predicted event to be the nominee with the highest probability in that specific year and category. In the very rare event that two competing nominees have equal predicted probabilities, they are both predicted as the winner. Previous literature (Steyerberg et al., 2001), (Austin & Tu, 2004) have suggested bootstrap validation as another method which may perform favorably against K-fold cross validation methods (i.e. 10-fold, jackknife ect.).

However, because our classification rule relies on all predicted probabilities in the same year, normal resampling with replacement and data-splitting procedures will not perform well. Random resampling and cross-validation may result in scenarios where nominees in the same year are partitioned

Table 4  
Confusion matrices for all 6 races  
Lead Actor

	Real	
	Lost	Won
Predicted		
Lost	158	15
Won	18	29

Internal Accuracy: .85  
Jackknife Estimate: .85  
Lead Actress

	Real	
	Lost	Won
Predicted		
Lost	157	20
Won	20	24

Internal Accuracy: .819  
Jackknife Estimate: .782

Supporting Actor

	Real	
	Lost	Won
Predicted		
Lost	158	15
Won	19	29

Internal Accuracy: .846  
Jackknife Estimate: .845  
Supporting Actress

	Real	
	Lost	Won
Predicted		
Lost	152	17
Won	25	27

Internal Accuracy: .81  
Jackknife Estimate: .81

Direction

	Real	
	Lost	Won
Predicted		
Lost	170	4
Won	7	40

Internal Accuracy: .955  
Jackknife Estimate: .931  
Picture

	Real	
	Lost	Won
Predicted		
Lost	188	7
Won	6	38

Internal Accuracy: .946  
Jackknife Estimate: .934

into separate folds. In this case our classification rule would perform disastrously. Instead we cross-validate by partitioning across years, rather than across observations. The accuracy of the models are cross-validated with the Jackknife (leave-one-out) procedure. Rather than  $n-1$  folds, we have  $r-1$  folds where  $r$  is the number of years in the datasets.

Limitations in stepwise regression have been noted for inclusion of noise variables and spurious relationships in the event of collinearity (Harrell, 2003). Given that we are including several film awards variables which are designed to quantify the same thing, collinearity would appear to be a diagnostic issue. However using variance inflation factors, we see there are no factors greater than 3 and strangely there is little multicollinearity between precursor award shows.

## 2014 Prediction and Results

For both internal validation and cross-validation, the  $AIC_c$  models perform better than the  $BIC$  models and the small-sample correction  $AIC_c$  targets the same models as  $AIC$ . Using the  $AIC_c$  models in this study, predicted probabilities for the 2014 Oscar nominees are calculated and compared with the implied probabilities taken from the modal fractional betting odds of major bookies, an hour before the 2014 ceremony. We select the candidate with the highest predicted probability as our official prediction for the Oscar winner. The results are shown in Table 5.

This study's predictions score 5/6 and are consistent with the speculative market's predictions for all categories except Best Picture. It is important to note that *Captain Phillips* has a much larger predicted probability than many would considered. This stems from its surprise win at the ACE Eddie awards and the Writer's Guild Awards. The WGA win is unique as neither frontrunner film for Best Picture was al-

lowed to contest - *12 Years a Slave* being ineligible for nomination due to a guild dispute and *Gravity* excluded for lack of merit. Furthermore, the 2014 Supporting Actress race was the closest. Lupita Nyong'o prevailed for *12 Years a Slave* over fellow nominee Jennifer Lawrence for *American Hustle*. This is attributed to our model showing a larger odds ratio for the BAFTA winner (Nyong'o) over the Golden Globes winner (Lawrence). This is the only acting race where the model

Table 5  
2014 Oscars: Predictions (Winner and Runner-up)

Nominees	Bookies		Models
	Odds	Implied $p$	$\hat{p}$
<b>Best Leading Actor</b>			
Matthew McConaughey *	3/10	0.769	0.729
Chiwetel Ejiofor	20/1	0.048	0.404
<b>Best Leading Actress</b>			
Cate Blanchett *	1/25	0.962	0.674
Amy Adams	20/1	0.048	0.232
<b>Best Supporting Actor</b>			
Jared Leto *	1/12	0.923	0.836
Jonah Hill	66/1	0.015	0.239
<b>Best Supporting Actress</b>			
Lupita Nyong'o *	5/6	0.545	0.465
Jennifer Lawrence	7/5	0.417	0.383
<b>Best Directing</b>			
Alfonso Cuarón *	1/16	0.941	0.922
David O. Russell	66/1	0.015	0.032
<b>Best Picture</b>			
Gravity	7/2	0.222	0.470
Captain Phillips	250/1	0.004	0.170
12 Years a Slave *	1/3	0.750	0.156

Prediction, \* = winner

places a greater weight on the BAFTAs than on the Golden Globes. For Lead Actor, Lead Actress and Supporting Actor, wins at the Golden Globes increase their Oscar odds by factors of 14, 13 and 19 respectively.

### Discussion

Determining the best performances, direction and film of the year is hardly a scientific design. However with a rich set of data, there continue to be promising avenues to study the predictability of the Academy Awards. As seen in the analysis, previous film awards, particularly guild awards, have an unquestionably large effect on the odds of winning an Oscar in certain categories. There are many new award shows, such as the Spirits, the Critic's Choice Awards and SAG awards which we have excluded for the sake of statistical power. However in the coming years, we anticipate there will be sufficient data to re-model the Oscar ceremonies using new film award shows and these awards show strong signs of being promising indicators of the Oscar winning incidence rate.

### Appendix Tables References

- Austin, P. C., & Tu, J. V. (2004). Bootstrap methods for developing predictive models. *The American Statistician*, 58(2), pp. 131-137. Retrieved from <http://www.jstor.org/stable/27643521>
- BartoÅD, K. (2013). Mumin: Multi-model inference [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=MuMIn> (R package version 1.9.13)
- Bernard, A. (2005). An index of oscar-worthiness: predicting the academy award for best picture. *Unpublished paper*. Retrieved from [mba.tuck.dartmouth.edu/pages/faculty/andrew.bernard/oscars.pdf](http://mba.tuck.dartmouth.edu/pages/faculty/andrew.bernard/oscars.pdf)
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods Research*, 33(2), 261-304. Retrieved from <http://smr.sagepub.com/content/33/2/261.abstract> doi: 10.1177/0049124104268644
- Chongsuvivatwong, V. (2012). epicalc: Epidemiological calculator [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=epicalc> (R package version 2.15.1.0)
- Collins, A., & Hand, C. (2006). Vote clustering in tournaments: what can oscar tell us? *Creativity Research Journal*, 18(4), 427-434. doi: 10.1207/s15326934crj1804\_2
- Davison, D. V., A.C. Hinkley. (1997). *Bootstrap methods and their applications* (C. U. Press, Ed.). Cambridge.
- Deepak Pathaka, M. D., David Rothschild. (2013). A comparison of forecasting methods: fundamentals, polling, prediction markets, and experts. *Manuscript submitted for publication*.
- Gehrlein, W. V., & Kher, H. V. (2004, June). Decision rules for the academy awards versus those for elections. *Interfaces*, 34(3), 226-234. Retrieved from <http://dx.doi.org/10.1287/inte.1040.0072> doi: 10.1287/inte.1040.0072
- Ghomi, A., Shirzadi, E., & Movassaghi, A. (2013). Predicting the academy awards result by analyzing tweets. *Global Journal of Science, Engineering and Technology*, 8, 39-47.
- Harrell, F. E. (2003). *Regression modeling strategies with applications to linear models, logistic regression and survival analysis* (2nd ed.). New York: John Wiley Sons, Ltd. Retrieved from <http://dx.doi.org/10.1002/sim.1497> doi: 10.1002/sim.1497
- Hlavac, M. (2013). stargazer: Latex code and ascii text for well-formatted regression and summary statistics tables [Computer software manual]. Cambridge, USA. Retrieved from <http://CRAN.R-project.org/package=stargazer> (R package version 4.5.3)
- Hsieh, F. Y., Bloch, D. A., & Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, 17(14), 1623-1634. doi: 10.1002/(SICI)1097-0258(19980730)17:14<1623::AID-SIM871>3.0.CO;2-S
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773-795. Retrieved from <http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572>
- Krauss, J., Nann, S., Simon, D., Fischbach, K., & Gloor, P. (2008). Predicting movie success and academy awards through sentiment and social network analysis. *ECIS 2008 Proceedings*(Paper 116). Retrieved from <http://aisel.aisnet.org/ecis2008/116>
- Kuhn, M. (2014). caret: Classification and regression training [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=caret> (R package version 6.0-24)
- Litman, B. R. (1983). Predicting success of theatrical movies: an empirical study. *The Journal of Popular Culture*, 16(4), 159-175. doi: 10.1111/j.0022-3840.1983.1604\_159.x
- Pardoe, I., & Simonton, D. K. (2008). Applying discrete choice models to predict academy award winners. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2), 375-394. doi: 10.1111/j.1467-985X.2007.00518.x
- Silver, N. (2013). Oscar predictions, election-style. *fivethirtyeight.com*. Retrieved from <http://fivethirtyeight.blogs.nytimes.com/2013/02/22/oscar-predictions-election-style/>
- Simonton, D. (2004). Film awards as indicators of cinematic creativity and achievement: a quantitative comparison of the oscars and six alternatives. *Creativity Research Journal*, 16(2-3), 163-172. Retrieved from <http://escholarship.org/uc/item/6kp6d661>
- Steyerberg, E. W., Jr, F. E. H., Borsboom, G. J., Eijkemans, M., Vergouwe, Y., & Habbema, J. F. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, 54(8), 774 - 781. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0895435601003419> doi: [http://dx.doi.org/10.1016/S0895-4356\(01\)00341-9](http://dx.doi.org/10.1016/S0895-4356(01)00341-9)
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth ed.). New York: Springer. Retrieved from <http://www.stats.ox.ac.uk/pub/MASS4> (ISBN 0-387-95457-0)

Table A1  
*Final Models with backwards stepwise AIC<sub>c</sub> selection*

	<i>Dependent variable:</i>					
	Best Actor	Best Actress	Best Supporting Actor	Won Oscar Best Supporting Actress	Best Director	Best Picture
globes	14.379*** (5.960,37.704)	13.143*** (5.201,36.135)	18.995*** (8.180,48.009)	4.543*** (2.076,10.107)		
bafta	3.626*** (1.423,9.292)	4.582*** (1.749,12.352)		6.351*** (2.523,16.470)		9.152** (1.809,56.179)
script.nom	4.635*** (1.658,15.150)	2.025 (0.818,5.173)			11.623 (0.551,886.875)	
Hot	3.248** (1.159,9.236)	3.939*** (1.424,11.225)				5.293** (1.157,28.281)
fall	0.421** (0.173,0.981)					
past.win		0.272** (0.082,0.785)				
past.nom		1.407*** (1.110,1.805)				
drama		0.267 (0.057,1.517)				
rom		2.526** (1.024,6.473)				
Dark			2.507* (0.958,6.537)			
eddie					6.475** (1.284,32.763)	8.015*** (1.928,39.474)
comedy			3.999*** (1.510,10.814)	2.138* (0.927,4.998)	0.198 (0.018,1.530)	5.328** (1.083,29.526)
action			0.314 (0.065,1.261)			
edit.nom		2.397* (0.913,6.348)	3.410*** (1.399,8.796)	2.710** (1.193,6.203)		15.978** (2.238,178.153)
wga				2.766** (1.045,7.261)		6.696*** (1.759,32.714)
dga					355.224*** (82.578,2,641.651)	232.514*** (50.109,1,804.354)
Constant	0.030*** (0.008,0.091)	0.059*** (0.010,0.265)	0.031*** (0.011,0.073)	0.050*** (0.022,0.103)	0.003*** (0.00003,0.056)	0.0002*** (0.00001,0.004)
Observations	220	221	221	221	221	239
AIC <sub>c</sub>	158.694	168.916	163.268	186.131	74.361	82.544
AIC	158.300	167.868	162.875	185.739	74.082	81.918
Area under ROC	0.883	0.874	0.853	0.790	0.962	0.977

*Note:* Values are in terms of odds ratios and 95% exponentiated confidence intervals. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A2  
*Full 2014 Prediction and Speculation odds*

Nominee	Bookie spread		Model prediction
	Betting Odds	Implied Probability	Probability
<b>Lead Actor</b>			
Christian Bale	125/1	0.008	$0.054 \pm 0.02$
Bruce Dern	55/1	0.018	$0.12 \pm 0.04$
Leonardo DiCaprio	11/2	0.154	$0.054 \pm 0.02$
Chiwetel Ejiofor	20/1	0.048	$0.404 \pm 0.15$
Matthew McConaughey	3/10	0.769	$0.729 \pm 0.13$
<b>Lead Actress</b>			
Amy Adams	20/1	0.048	$0.232 \pm 0.11$
Cate Blanchett	1/25	0.962	$0.674 \pm 0.14$
Sandra Bullock	33/1	0.029	$0.014 \pm 0.01$
Judi Dench	40/1	0.024	$0.211 \pm 0.11$
Meryl Streep	100/1	0.01	$0.096 \pm 0.11$
<b>Supporting Actor</b>			
Barkhad Abdi	16/1	0.059	$0.211 \pm 0.08$
Bradley Cooper	100/1	0.01	$0.096 \pm 0.04$
Michael Fassbender	14/1	0.067	$0.211 \pm 0.08$
Jonah Hill	66/1	0.015	$0.239 \pm 0.11$
Jared Leto	1/12	0.923	$0.836 \pm 0.08$
<b>Supporting Actress</b>			
Sally Hawkins	50/1	0.02	$0.097 \pm 0.03$
Jennifer Lawrence	7/5	0.417	$0.383 \pm 0.10$
Lupita Nyong'o	5/6	0.545	$0.465 \pm 0.13$
Julia Roberts	66/1	0.015	$0.048 \pm 0.02$
June Squibb	50/1	0.02	$0.048 \pm 0.02$
<b>Directing</b>			
David O. Russell	66/1	0.015	$0.032 \pm 0.02$
Alfonso Cuaron	1/16	0.941	$0.922 \pm 0.06$
Alexander Payne	150/1	0.007	$0.032 \pm 0.02$
Steve McQueen	14/1	0.067	$0.032 \pm 0.02$
Martin Scorsese	100/1	0.01	$0.007 \pm 0.01$
<b>Picture</b>			
American Hustle	22/1	0.043	$0.004 \pm 0.00$
Captain Phillips	250/1	0.004	$0.17 \pm 0.11$
Dallas Buyers Club	40/1	0.024	$0.004 \pm 0.00$
Gravity	7/2	0.25	$0.47 \pm 0.17$
Her	250/1	0.004	$0.002 \pm 0.00$
Nebraska	250/1	0.004	$0 \pm 0.00$
Philomena	250/1	0.004	$0.001 \pm 0.00$
12 Years a Slave	1/3	0.75	$0.156 \pm 0.14$
The Wolf of Wall Street	66/1	0.015	$0.001 \pm 0.00$