

# Just how predictable are the Oscars?

Iain Pardoe

September 5, 2005

Each year, hundreds of millions of people worldwide watch the television broadcast of the Academy Awards ceremony, at which the Academy of Motion Picture Arts and Sciences (AMPAS) honors film-making from the previous year. Almost 6000 members of AMPAS vote for the nominees and final winners of Academy Awards, more commonly known as Oscars, in a wide range of categories for directing, acting, writing, editing, etc. Oscars have been presented for outstanding achievement in film every year since 1928, and are generally recognized to be the premier awards of their kind since AMPAS voting members are themselves the foremost workers in the motion picture industry. In a comparison with other movie awards and movie guide ratings, Simonton (2004) finds substantial validity for the Oscars, and notes that “Those who take an Oscar home can have a strong likelihood of having exhibited superlative cinematic creativity or achievement.”

As well as honoring film-makers, Oscars can boost the box-office performance of nominated and winning films. It has even been shown that winning a Best Actor or Best Actress Oscar is associated with a gain in life expectancy, perhaps four extra years of life (Redelmeier and Singh, 2001). However, while studies into the factors that impact a movie’s economic success show that awards can boost revenues, there is little overall association between budget and box office variables and the most important movie awards, such as the Oscars.

This article does not consider the economic and aesthetic aspects of movies in relation to the Oscars, but rather focuses purely on the goal of predicting the winners of the four major awards—picture, director, actor in a leading role, actress in a leading role—from those nominated each year. Although many in the media (as well as movie-loving members of the public) make their own annual predictions, it appears that very few researchers have conducted a formal statistical analysis

for this purpose.

In terms of data, since the goal is to predict the eventual winner from a list of nominees, any information on the nominees that is available before the announcement of the winner is potentially useful, including other Oscar category nominations, previous nominations and wins, and other (earlier) movie awards. I use a discrete choice model to provide annual predictions, and then assess predictive accuracy using one-year-ahead, out-of-sample errors. The modeling approach used allows prediction of the four major Oscars from 1938 to 2004 (earlier years had yet to accumulate sufficient information to provide satisfactory predictions). The final results reveal interesting insights into just how predictable the four major Oscars are, which factors play an important role in the predictions, and also how these have changed over time. It is also possible to identify past winners with an exceptionally low estimated probability of winning, and past nominees with a very high estimated probability of winning who did not actually win.

## **Data**

All data have been obtained from two reliable Internet sources: “The Fennec Awards Database” ([awards.fennec.org](http://awards.fennec.org)) and “The Internet Movie Database” ([us.imdb.com](http://us.imdb.com)). Tables 1 to 5 contain a description of the explanatory variables used to predict the four major Oscar winners from 1938 to 2004, with data ranges for the predicted years’ awards indicated with square brackets “[ ]” (each variable was included only for the years in which it provided some predictive power). Table 6 contain a description of additional variables that were considered but ultimately not used.

## **Estimation**

The goal is to predict the four major Oscar winners for each year from 1938 to 2004 using any information on the nominees that is available before the announcement of the winner. This can be framed as a series of discrete choice problems with one winner selected in each category each year from a discrete set of nominees (usually five, although up until 1936 the number of director and acting nominees varied between 3 and 8, while up until 1944 the number of picture nominees varied between 5 and 10).

In this particular discrete-choice application, the explanatory variables described earlier take

different values for different response (nominee) choices. McFadden (1974) proposed a discrete-choice model for just such a case where explanatory variables are characteristics of the choices. This model also permits the choice set to vary across choice experiments, which in this case are each of the four categories (picture, director, actor, actress) in each of the years (1938–2004).

For experiment  $i$  and response choice  $j$ , let  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$  denote the values of  $p$  explanatory variables, and let  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{ip})$ . Conditional on the choice set  $C_i$  for experiment  $i$ , the model for the probability of selecting choice  $j$  is

$$\Pr(Y = j | \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_{ij})}{\sum_{h \in C_i} \exp(\boldsymbol{\beta}^T \mathbf{x}_{ih})},$$

where  $Y$  is the categorical response variable representing the winning nominee. For each pair of choices  $a$  and  $b$ , this model has the logit form

$$\log[\Pr(Y = a | \mathbf{x}_i) / \Pr(Y = b | \mathbf{x}_i)] = \boldsymbol{\beta}^T (\mathbf{x}_{ia} - \mathbf{x}_{ib}).$$

Conditional on the choice being  $a$  or  $b$ , a variable's effect depends on the difference in the variable's values for those choices. If the values are the same, then the variable has no effect on the choice between  $a$  and  $b$ . Thus McFadden originally referred to this model as a conditional logit model, although it is now more commonly called a multinomial logit model because the underlying likelihood is a multinomial distribution.

Multinomial logit models can be fit with a variety of statistical software packages. For reasons of flexibility, convenience, and familiarity, Bugs is used here for model estimation, with R used to process data and results. All data available before the announcement of the 1938 Oscars is used to fit a model which can predict the winners for that year. Then, the actual outcome of the 1938 Oscars is appended to the previous dataset, and used to fit a new model which can predict the winners of the 1939 Oscars. The process repeats, adding new variables as they become available, up to the most recent Oscars in 2004. To assess the predictive accuracy of the analysis, one-year-ahead, out-of-sample errors were used. For example, the four major Oscars winners for 1938 were predicted from a model fit to data from 1928–1937. Then, the winners for 1939 were predicted from a model fit to data from 1928–1938, and so on.

## Results

Using the modeling approach just described, 186 of the 268 Best Picture, Director, Actor, and Actress Oscar winners from 1938–2004 were correctly identified, corresponding to an overall prediction accuracy of 69%. With more data available in the later years, prediction accuracy has improved over time. For example, the overall prediction accuracy for the last 30 years (1975–2004) is 97 correct predictions out of 120, or 81%. Figure 1 summarizes overall results across the four categories. Overall, the Best Director Oscar has been the most predictable, then the Oscars for Best Picture, Best Actor, and Best Actress, respectively. Each of the categories have become more predictable over time, particularly Best Actress, which was very hard to predict up until the early 1970s.

The roles of the explanatory variables in helping to predict Oscar winners change over time; Figure 2 illustrates. The importance of receiving a Best Director nomination (for Best Picture nominees) of a Best Picture nomination (for Best Director, Actor, or Actress nominees) has tended to increase over time (except perhaps for actors), as shown by the trends in the lines labeled “P”. Previous nominations appear to have remained approximately equally important for Best Director nominees, but were more important for Best Actor nominees in the past than they have been more recently (lines labeled “N”). Previous wins seemed to hurt Best Actor nominees less in the 1960s and 1970s than in the 1940s and more recently, while previous wins have tended to become less important for Best Actress nominees over time (lines labeled “W”).

The Golden Globes have remained useful predictors of future Oscar success since their inception. The changing fortunes of dramas (lines labeled “D”) and musicals and comedies (lines labeled “M”) can be traced in Figure 2, with musicals and comedies appearing to hold an advantage over dramas in the 1960s with respect to Best Picture wins, but with acting wins tending to favor dramas, particularly for males. Guild awards have clearly enabled quite accurate prediction of Best Director winners, and to a lesser extent Best Picture winners (lines labeled “G”). Since they have had a much shorter history, it is not clear whether SAG awards will be just as helpful in predicting acting wins, although early indications would suggest so.

The effect of the total number of Oscar nominations (lines labeled “T”) on prediction of the Best Picture and Director Oscars remains reasonably steady. Since the number of nominations a

movie can receive has ranged in the past between 1 and 14, this variable is more influential than it appears to be in the graphs (which show the effects of the number of nominations increasing by *one*). The effects of the “front runner” variables—which cut across all four categories—are not shown in Figure 2 (they appeared to be less important than the other variables, having estimates with smaller magnitudes and larger standard errors).

The analysis also reveals which past nominees have really upset the odds (winners with low estimated probability of winning), and which appear to have been truly robbed (losers with high estimated probability of winning). Table 7 provides details of the three “most surprising” outcomes in each category (based on the model results). A complete listing of the results is available at the web-site <http://lcb1.uoregon.edu/ipardoe/oscars/>—this site will be updated each year so predictions for the 2005 Oscars should appear soon.

## Remarks

Discrete choice modeling of past data on Oscar nominees in the four major categories—Best Picture, Director, Actor, and Actress—enables prediction of the winners in these categories with a reasonable degree of success. If recent trends persist, it should be possible to predict future winners with a prediction success rate of approximately 77% for Picture, 93% for Director, 77% for Actor, and 77% for Actress. The analysis could also be extended to other Oscar categories, such as the supporting acting and screen-writing awards.

A limitation of the model is that it can give very extreme predictions which cannot (of course) account for unmeasured factors. I recall the surprise of Denzel Washington winning over Russell Crowe in the 2001 Oscar race for Best Actor, but the surprise wasn’t nearly as extreme as implied by the model predictions. Another example is “The Aviator” failing to win Best Picture for 2004 (after winning a Golden Globe and the PGA award). Again, the surprise of “Million Dollar Baby” winning instead was not as extreme as implied by the model predictions—but the model was unable to make use of the late surge that “Million Dollar Baby” made (in unquantifiable “Hollywood buzz” terms) as the Oscar Ceremony drew near.

Further exploration of the results reveals additional insights into the predictability—or lack thereof—of winning an Oscar. For example, there has been much media speculation about leg-

endary individuals who have never won an Oscar, such as Alfred Hitchcock and Martin Scorsese, each with five directing nominations, Peter O'Toole with seven lead actor nominations, Richard Burton with six lead actor nominations, and Deborah Kerr with six lead actress nominations. Of these, the unluckiest was probably O'Toole who came closest to winning in 1968 (with an 89% modeled probability of winning) and 1964 (83% probability). Kerr came close in 1956 (72% probability), as did Burton in 1977 (62% probability), while Hitchcock's nearest miss was for Rebecca in 1940 (42% probability). Hitchcock, Kerr, and O'Toole were subsequently awarded honorary Oscars.

## **Additional Reading**

- Dodds, J. C. and M. B. Holbrook (1988). What's an Oscar worth? An empirical estimation of the effects of nominations and awards on movie distribution and revenues. *Current Research in Film: Audiences, Economics and Law* 4, 72–87.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.
- Nelson, R. A., M. R. Donihue, D. M. Waldman, and C. Wheaton (2001). What's an Oscar worth? *Economic Inquiry* 39, 1–16.
- Redelmeier, D. A. and S. M. Singh (2001). Survival in Academy Award-winning actors and actresses. *Annals of Internal Medicine* 134, 955–962.
- Simonoff, J. S. and I. R. Sparrow (2000). Predicting movie grosses: Winners and losers, blockbusters and sleepers. *Chance* 13(3), 15–24.
- Simonton, D. K. (2004b). Film awards as indicators of cinematic creativity and achievement: A quantitative comparison of the Oscars and six alternatives. *Creativity Research Journal* 16, 163–172.

Table 1: Explanatory variables for Best Picture

1. Total number of Oscar nominations [1938–2004]. Nominees for Best Picture are often also represented by multiple nominees in other categories, and the chances of winning are generally thought to increase the higher the total number of nominations. For example, the median number of nominations for winners of the Best Picture Oscar since its inception (1928–2004) is nine, whereas the median number of nominations for losing Best Picture nominees is six.
2. Indicator for Best Director Oscar nomination [1938–2004]. Only three movies have won the Best Picture Oscar without also receiving a Best Director nomination (*Wings* in 1928, *Grand Hotel* in 1932, and *Driving Miss Daisy* in 1989).
3. Indicator for winning a Golden Globe for Best Picture or for Best Picture (Drama) [1946–2004]. The Hollywood Foreign Press Association (a group of Southern California-based international journalists) has awarded its Golden Globes every year since 1944 to honor achievements in film during the previous calendar year. Since Oscars are presented some time after Golden Globes (up to two months later), winning a Golden Globe often precedes winning an Oscar. For example, of the 62 Best Picture Oscar winners from 1943 to 2004, 34 had won a Golden Globe for Best Picture (Drama) a few weeks earlier.
4. Indicator for winning a Golden Globe for Best Picture (Musical or Comedy) [1956–2004]. The Golden Globe award for Best Picture was separated into two distinct categories in 1951: Drama and Musical or Comedy. Of the 54 Best Picture Oscar winners from 1951 to 2004, ten had won a Golden Globe for Best Picture (Musical or Comedy) a few weeks earlier.
5. Indicator for winning a Directors Guild of America (DGA) award (between 1951 and 1988) or a Producers Guild of America (PGA) award (since 1989) [1951–2004]. DGA has been awarding its honors for best Motion Picture Director since 1949 (with all but two early awards made before the announcement of the Best Picture Oscar). Since 1989, PGA has been awarding its honors to the year's most distinguished producing effort (with all but the first awarded before the announcement of the Best Picture Oscar). Of the 40 Best Picture Oscar winners from 1949 to 1988, 31 had already won a DGA award (and two would subsequently win one). Of the 16 Best Picture Oscar winners from 1989 to 2004, ten had already won a PGA award (and one would subsequently win one).

Table 2: Explanatory variables for Best Director

1. Total number of Oscar nominations [1939–2004]. As for Best Picture, nominees for Best Director are often for movies which are also represented by multiple nominees in other categories. The median number of nominations for movies with Best Director winners since 1928 is nine, whereas the median number of nominations for movies with losing Best Director nominees is six.
2. Indicator for Best Picture Oscar nomination [1944–2004]. Only two directors have won a Best Director Oscar for a movie that did not receive a Best Picture nomination (Lewis Milestone who won a Best Director (Comedy) Oscar for *Two Arabian Nights* in 1928, and Frank Lloyd who won a Best Director Oscar for *The Divine Lady* in 1929).
3. Natural logarithm of the number of previous Best Director Oscar nominations [1938–2004]. A director's chance of winning an Oscar tends to increase the more times they have been nominated in previous years. For example, 18 percent of Best Director Oscar nominees with no previous directing nominations have won the Oscar, whereas 24 percent of Best Director Oscar nominees with one or more previous directing nominations have won. This variable has been log-transformed because it is highly skewed.
4. Indicator for winning a Golden Globe for Best Director (between 1945 and 1950) or a Directors Guild of America award (from 1951) [1945–2004]. Of the 62 Best Director Oscar winners from 1943 to 2004, 33 had already won a Golden Globe for Best Director. Of the 56 Best Director Oscar winners from 1949 to 2004, 49 had already won a DGA award (and one would subsequently win one). Separate indicators were not included for both the Golden Globe Best Director and DGA awards from 1949 on because of collinearity between the two awards.

Table 3: Explanatory variables for Best Actor in a Leading Role

1. Indicator for Best Picture Oscar nomination [1939–2004]. Only 12 actors have won the Best Actor Oscar for a movie that did not receive a Best Picture nomination (most recently, Denzel Washington for *Training Day* in 2001).
2. Natural logarithm of the number of previous Best Actor in a Leading Role Oscar nominations [1938–2004]. 19 percent of Best Actor Oscar nominees with no previous lead actor nominations have won the Oscar, whereas 23 percent of Best Actor Oscar nominees with one or more previous lead actor nominations have won. This variable has been log-transformed because it is highly skewed.
3. Natural logarithm of the number of previous Best Actor in a Leading Role Oscar wins [1939–2004]. An actor’s chance of winning an Oscar tends to *decrease* the more times they have won in previous years. For example, 23 percent of Best Actor Oscar nominees with no previous lead actor wins have won the Oscar, whereas ten percent of Best Actor Oscar nominees with one or more previous lead actor wins have won. This variable has been log-transformed because it is highly skewed.
4. Indicator for winning a Golden Globe for Best Actor in a Leading Role (Drama) [1944–2004]. Of the 62 Best Actor Oscar winners from 1943 to 2004, 39 had won a Golden Globe for Best Actor (Drama) a few weeks earlier.
5. Indicator for winning a Golden Globe for Best Actor in a Leading Role (Musical or Comedy) [1965–2004]. The Golden Globe award for Best Actor in a Leading Role was separated into two distinct categories in 1950: Drama and Musical or Comedy. Of the 55 Best Picture Oscar winners from 1950 to 2004, six had won a Golden Globe for Actor (Musical or Comedy) a few weeks earlier.
6. Indicator for winning a Screen Actor’s Guild (SAG) award [1995–2004]. Since 1994 SAG has awarded five statuettes, known as “The Actor,” for achievements in film (always before the Oscar Ceremony), including Male Actor in a Leading Role and Female Actor in a Leading Role. Of the 11 Best Actor Oscar winners since 1994, seven had already won a SAG award.

Table 4: Explanatory variables for Best Actress in a Leading Role

1. Indicator for Best Picture Oscar nomination [1939–2004]. Only 25 actresses have won the Best Actress in a Leading Role Oscar for a movie that did not receive a Best Picture nomination (most recently, Charlize Theron for *Monster* in 2003).
2. Natural logarithm of the number of previous Best Actress in a Leading Role Oscar wins [1938–2004]. 24 percent of Best Actress Oscar nominees with no previous lead actress wins have won the Oscar, whereas 13 percent of Best Actress Oscar nominees with one or more previous lead actress wins have won. This variable has been log-transformed because it is highly skewed.
3. Indicator for winning a Golden Globe for Best Actress in a Leading Role (Drama) [1944–2004]. Of the 62 Best Actress Oscar winners from 1943 to 2004, 31 had won a Golden Globe for Best Actress (Drama) a few weeks earlier.
4. Indicator for winning a Golden Globe for Best Actress in a Leading Role (Musical or Comedy) [1952–2004]. Of the 55 Best Actress Oscar winners from 1950 to 2004, 11 had won a Golden Globe for Best Actress (Musical or Comedy) a few weeks earlier.
5. Indicator for winning a Screen Actor’s Guild award [1996–2004]. Of the 11 Best Actress Oscar winners since 1994, eight had already won a SAG award.

Table 5: Explanatory variables included for all four categories

1. Indicator for the first “front-running movie” [1938–2004]. This variable allows for the possibility that the chance of a nominee winning an Oscar could be linked to the fortunes of other nominees for the same movie. Each year there are often a handful of movies considered to be the Oscar front-runners—movies with multiple nominations in the more high-profile categories (including picture, director, and acting). To identify these front-runners, the Oscar categories were ranked each year based on previous Best Picture Oscar winners (for example, the Best Director category usually ranks highly since Best Picture Oscar winners nearly always also have a Best Director nomination). Then, a “nomination score” can be calculated for each movie nominated for one of the four major Oscars based on these rankings (for example, movies with many nominations in the top-ranked categories will have higher nomination scores than movies with few nominations). The indicator variable then identifies the top front-runner as the movie with the highest nomination score, and takes the value one for all nominees associated with this movie.

2 and 3. Indicators for the second and third front-running movies [1959–2004]. These variables identify the movie with the second and third highest nomination scores, and take the value one for all nominees associated with these movies.

Table 6: Excluded variables

While a variable for previous Best Director Oscar nominations is included, including the number of previous Best Director Oscar *wins* tended to worsen rather than improve predictions. Conversely, while a variable for previous Best Actress Oscar wins is included, the number of previous Best Actress Oscar *nominations* tended to worsen predictions. Also, while a variable for the total number of nominations improves predictions of the Best Picture and Best Director Oscar winners, such a variable worsens predictions of the acting Oscar winners.

It is well documented that female winners of acting Oscars tend to be younger than male winners. For example, the median age of Best Actress Oscar winners between 1928 and 2004 was 33, whereas that for Best Actor was 42. However, the age differences within gender between Oscar winning and losing nominees are less dramatic. In the first third of the Oscars' history (1928–1953), the median age of Best Actress Oscar winners was 29 versus that of losing nominees of 33. Comparable figures for the second third (1954–1979) are 34 versus 34, and for the final third (1980–2004) are 35 versus 37. In other words, actress nominee ages have increased over time, with winning nominees tending to be slightly younger than losing nominees (less so during the middle period). For Best Actor nominees, comparable figures for the first third are 41 versus 38, for the second third are 43 versus 39, and for the final third are 43 versus 45. Thus, actor nominee ages have also increased over time, with winning nominees tending to be slightly older than losing nominees initially, but tending to be slightly younger more recently. Age effects of this nature on the chance of winning an acting Oscar can be picked up by adding *age* and *age-squared* variables (i.e., quadratic terms) to the models for Best Actor and Best Actress. Nevertheless, incorporating quadratic terms for age into the models failed to improve predictions of winners.

Other variables that were investigated but which did not improve results include supporting actor Oscar nominations and wins, nominated movie genre (e.g., drama, comedy, etc.), Motion Picture Association of America rating (e.g., PG, R, etc.), running time, release date, movie critic ratings, and other pre-Oscar awards (e.g., New York Film Critics Circle, Los Angeles Film Critics Association, National Society of Film Critics, and National Board of Review).

Table 7: Three outcomes in each of the major categories with the smallest estimated win probabilities for the actual winner relative to the predicted winner.

Year	Winner	Prob	Predicted	Prob
<b>Best Picture</b>				
1948	Hamlet	0.01	Johnny Belinda	0.97
1981	Chariots of Fire	0.01	Reds	0.87
2004	Million Dollar Baby	0.02	The Aviator	0.97
<b>Best Director</b>				
2000	Steven Soderbergh	0.01	Ang Lee	0.95
2002	Roman Polanski	0.02	Rob Marshall	0.92
1968	Carol Reed	0.03	Anthony Harvey	0.97
<b>Best Actor</b>				
2001	Denzel Washington	0.00	Russell Crowe	0.99
1968	Cliff Robertson	0.00	Peter O'Toole	0.89
1974	Art Carney	0.02	Jack Nicholson	0.86
<b>Best Actress</b>				
2002	Nicole Kidman	0.08	Renée Zellweger	0.90
1967	Katharine Hepburn	0.05	Faye Dunaway	0.43
1966	Elizabeth Taylor	0.08	Anouk Aimée	0.68

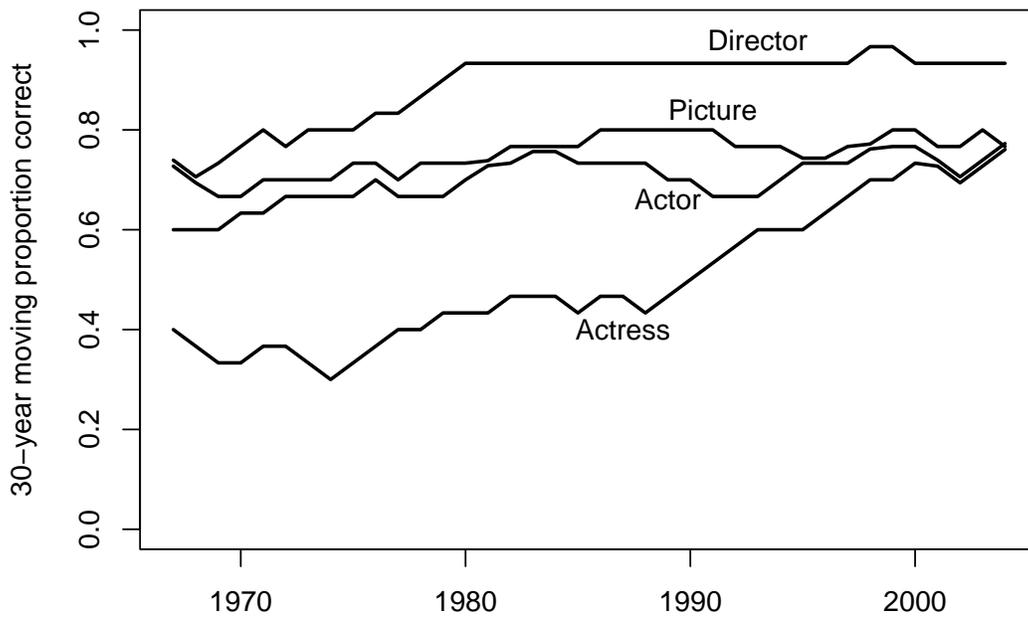


Figure 1: 30-year moving averages of the proportion of correct predictions in each of the four major Oscar categories. The moving average values are placed at the ends of the 30-year periods, e.g., at the far right of the graph the proportions of correct predictions over the period 1975–2004 are 93% for Best Director, 77% for Best Picture, 77% for Best Actor, and 77% for Best Actress.

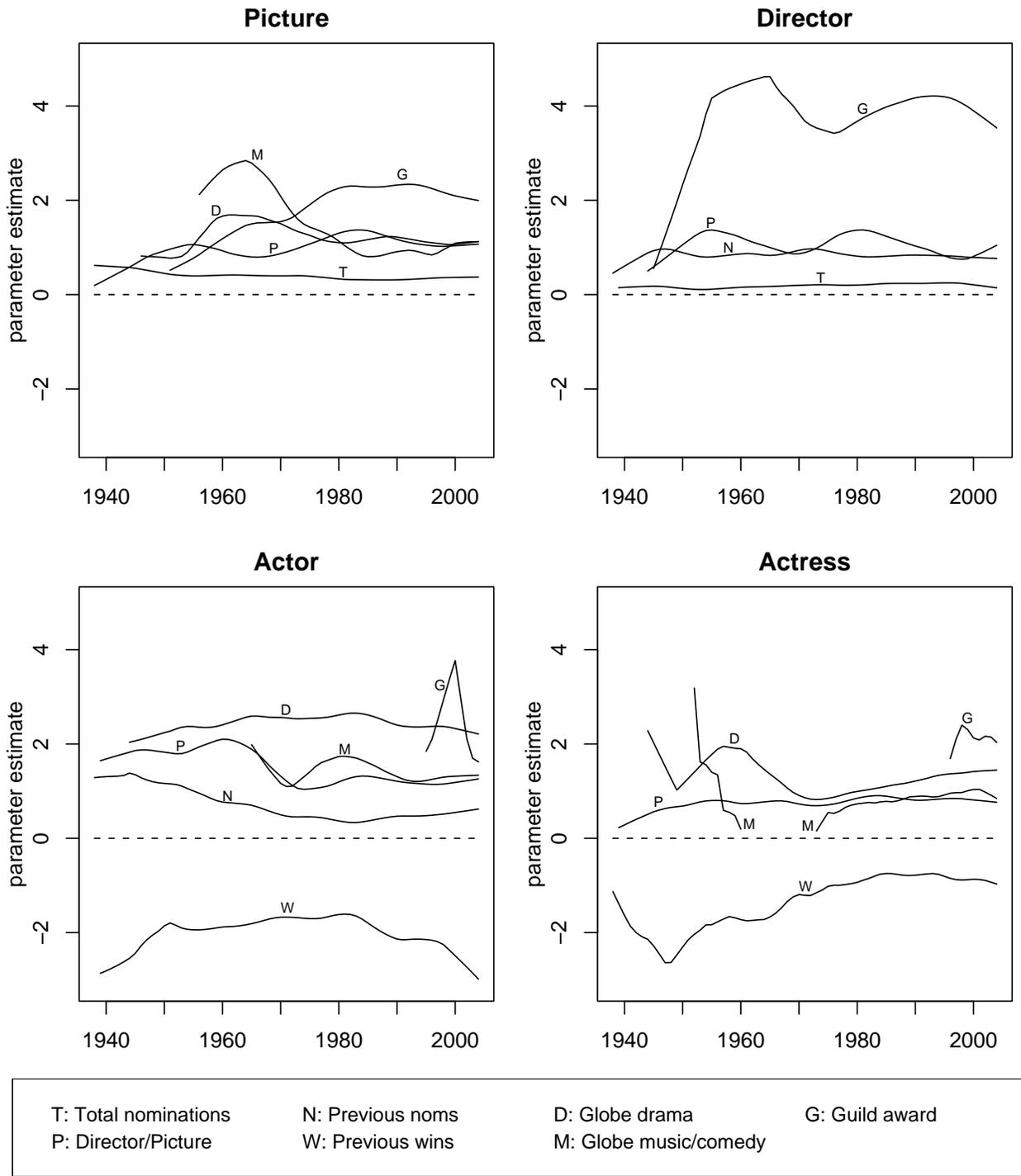


Figure 2: Smoothed parameter estimates for the explanatory variables for each of the four major Oscar categories. The explanatory variables are described in the text