J. R. Statist. Soc. A (2008)
171, *Part* 2, *pp.* 375–394

Applying discrete choice models to predict Academy Award winners

Iain Pardoe

University of Oregon, Eugene, USA

and Dean K. Simonton

University of California at Davis, USA

[Received September 2005. Revised June 2007]

Summary. Every year since 1928, the Academy of Motion Picture Arts and Sciences has recognized outstanding achievement in film with their prestigious Academy Award, or Oscar. Before the winners in various categories are announced, there is intense media and public interest in predicting who will come away from the awards ceremony with an Oscar statuette. There are no end of theories about which nominees are most likely to win, yet despite this there continue to be major surprises when the winners are announced. The paper frames the question of predicting the four major awards—picture, director, actor in a leading role and actress in a leading role—as a discrete choice problem. It is then possible to predict the winners in these four categories with a reasonable degree of success. The analysis also reveals which past results might be considered truly surprising—nominees with low estimated probability of winning who have overcome nominees who were strongly favoured to win.

Keywords: Bayesian; Conditional logit; Films; Forecasting; Mixed logit; Motion pictures; Movies; Multinomial logit

1. Introduction

Hundreds of millions of television viewers world wide watch the annual Academy Awards ceremony to see the Academy of Motion Picture Arts and Sciences (AMPAS) honour outstanding achievement in film from the previous year. Since 1928, AMPAS members have voted for the nominees and final winners of Academy Awards, which are more commonly known as Oscars, in a wide range of categories for directing, acting, writing, editing, etc. The Oscars are generally recognized to be the premier awards of their kind since the almost 6000 AMPAS members are among the foremost workers in the motion picture industry.

Besides honouring cinematic accomplishments in the most glamorous manner, Oscars have direct practical repercussions. For instance, winning a Best Actor or Best Actress Oscar can increase the income that recipients can later command and the quality of screenplays that are sent their way. In addition, Oscar awards and nominations can boost the box office performance of films by millions of dollars (see, for example Deuchert *et al.* (2005), Dodds and Holbrook (1988) and Nelson *et al.* (2001)). Furthermore, although many factors are associated with a film's gross earnings (Collins *et al.*, 2002; Simonoff and Sparrow, 2000; Terry *et al.*, 2005a, b), the financial consequence of Oscars operates independently of other significant predictors

Address for correspondence: Iain Pardoe, Department of Decision Sciences, Lundquist College of Business, University of Oregon, Eugene, OR 97403-1208, USA.
E-mail: ipardoe@lcbmail.uoregon.edu

 0964–1998/08/171375

(Litman, 1983; Sochay, 1994). As an example, although a film's production costs are positively correlated with gross earnings (Prag and Casavant, 1994; Simonton, 2005a), there is little, if any, association between budget and the most important movie awards, such as the directing, acting and screenplay Oscars (Simonton, 2005a, b).

Yet are these noteworthy repercussions of Oscar recognition actually justified? Many critics believe that the Academy Awards are almost completely contaminated by local Hollywood politics and provincial tastes (e.g. Peary (1993)). If so, then the individual and financial rewards are bestowed without merit and thus may be totally unfair. This issue has been addressed in two major ways.

First, investigators can simply determine whether Oscar nominations and awards are positively and significantly related to alternative ways of assessing cinematic achievement. For example, research shows that Oscar recognition is strongly associated with other awards, such as the Golden Globes which are bestowed by the Hollywood Foreign Press Association, a group of international journalists based in Southern California (Simonton, 2004a, b), and with critical acclaim, as gauged by the ratings that films receive in various movie guides (Simonton, 2002, 2004a). On the basis of these and other statistical relationships, Simonton (2004a), page 171, observed that 'Those who take an Oscar home can have a strong likelihood of having exhibited superlative cinematic creativity or achievement'. Indeed, among all major awards the Oscars appear to be the strongest indicators of merit (Ginsburgh, 2003; Simonton, 2004a).

Second, researchers can adopt a more ambitious predictive modelling strategy. Rather than simply considering correlations between pairs of variables, models for predicting Oscar outcomes can provide more detailed information about the magnitude of correspondence between the variables, and the number and extent of prediction errors. For instance, when Bennett and Bennett (1998) attempted to predict the winners of the best acting Oscars from 1936 to 1996, they achieved a successful prediction rate of 47% for Best Actor and 39% for Best Actress. Given the number of nominees in each of these two categories (generally 5), these results amply exceed baseline expectation. Although many in the media (as well as movie-loving members of the public) make their own annual predictions, this study, although somewhat limited by arbitrary variable definitions and *ad hoc* estimation techniques, appears to be the only previous analysis of this type in the literature.

The current investigation aims at developing this second strategy well beyond previous efforts. We focus on predicting the winners of the four major awards—picture, director, actor in a leading role and actress in a leading role—from those who were nominated each year. By developing statistically rigorous models for this task and identifying the prediction errors, it is possible to reveal specific cases in which the evaluative system that is implicit in the Oscar process may have been unjust, arbitrary or capricious. Beyond this, such error identification could lead to the eventual development of predictive models that introduce additional variables that isolate the source of the error. Some of these variables may be indicative of merit, but other variables could represent factors that are extraneous to actual cinematic achievement.

In contrast with the limited previous research on predictive modelling of Oscar outcomes, there is a wealth of research on motion picture industry economics (see De Vany (2004, 2006) and Walls (2005) for surveys). As pointed out by a referee, Caves (2000) referred to the 'nobody knows principle' (which was originally coined by screenwriter William Goldman), whereby much information on previous movie success fails to help us to predict reliably how successful the next movie will be. This extreme *ex ante* uncertainty dominates the distribution of profits, which tend to be stable Pareto distributed with a heavy upper tail (De Vany, 2006). As a result, Schulze (2005), page 157, concluded that

> 'As *ex ante* uncertainty is a dominant feature of the movie industry and informational cascades determine the dynamics of a movie and thus its financial success, it seems necessary to analyze how this information transmission can be influenced once the movie has been released. The most visible quality signals are nominations for the Academy Awards, and the awards themselves.'

The research that is reported in the current paper would appear to complement this economics-based literature, although its focus on predicting movie award outcomes is very different from predicting financial performance outcomes. In addition, although extreme uncertainty shapes movie industry economics, this work shows that a far greater degree of predictability appears to pervade some major movie awards.

The outline of the paper is as follows. Section 2 describes the data that are used—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. Section 3 motivates the discrete choice models that are used to provide annual predictions and discusses the modelling process. The modelling approach that is used allows 1-year-ahead, out-of-sample prediction of the four major Oscars from 1938 to 2006 (earlier years had yet to accumulate sufficient information to provide satisfactory predictions). Presentation of the final results in Section 4 includes 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 revealing 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. Finally, Section 5 contains a discussion, including ideas for how this work could be further developed in the future.

This paper extends earlier work in Pardoe (2005). In particular, this paper compares Bayesian estimation with maximum likelihood estimation for this application and also considers whether a more complex *mixed logit* model provides additional benefits over the standard *multinomial logit* model. This paper also expands on details of variable selection and model assessment, draws more extensively on economics and econometrics literature, includes two more years of data and contains additional discussion and many more references.

2. Data

All data were obtained from a reliable Internet source, namely 'The Internet movie database' (`us.imdb.com`). Table 1 outlines the explanatory variables that are used to predict the four major Oscar winners from 1938 to 2006 and also provides data ranges for the predicted years' awards (each variable was included only for the years in which it provided some predictive power—see also Section 3.3). Additional details on the variables follow (see also Pardoe (2005)).

Nominees for Best Picture and Best Director 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 and Best Director Oscars since their inception (1928–2006) is 9, whereas the median number of nominations for losing nominees is 6.

Nominees tend to fare better if they are nominated for movies receiving Best Picture and/or Best Director Oscar nominations.

(a) 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). In 1928, the Best Unique and Artistic Picture winner, *Sunrise*, also did not receive a Best Director nomination.

Table 1. Explanatory variables that are used to predict the four major Oscar winners from 1938 to 2006 and data ranges

Variable	*Picture*	*Director*	*Lead actor*	*Lead actress*
Total Oscar nominations	1938–2006	1939–2006	—	—
Director Oscar nomination	1938–2006	—	—	—
Picture Oscar nomination	—	1944–2006	1939–2006	1939–2006
Golden Globe drama	1946–2006	1945–1950	1944–2006	1944–2006
Golden Globe musical or comedy	1956–2006	—	1965–2006	1952–2006†
Guild award	1951–2006‡	1951–2006§	1995–2006	1996–2006
Previous Oscar nominations§§	—	1938–2006	1938–2006	—
Previous Oscar wins§§	—	—	1939–2006	1938–2006
1st front-running movie	1938–2006	1938–2006	1938–2006	1938–2006
2nd front-running movie	1959–2006	1959–2006	1959–2006	1959–2006
3rd front-running movie	1959–2006	1959–2006	1959–2006	1959–2006

†Variable dropped between 1961 and 1972 because the standard error greatly exceeded the estimate.
‡Directors Guild of America for 1951–1988; Producers Guild of America for 1989–2006.
§Separate indicators were not included for both the Golden Globe Best Director and Directors Guild of America awards from 1951 onwards because of collinearity between the two awards.
§§Transformed to natural logarithms.

(b) Only two directors have won a Best Director Oscar for a movie that was not nominated for Best Picture (Lewis Milestone for *Two Arabian Nights* in 1928 and Frank Lloyd for *The Divine Lady* in 1929).
(c) Only 13 actors have won the Best Actor Oscar for a movie that was not nominated for Best Picture (most recently, Forest Whitaker for *The Last King of Scotland* in 2006).
(d) Only 26 actresses have won the Best Actress in a Leading Role Oscar for a movie that was not nominated for Best Picture (most recently, Reese Witherspoon for *Walk the Line* in 2005).

The Hollywood Foreign Press Association has awarded its Golden Globes every year since 1944 to honour achievements in film during the previous calendar year. Since Oscars are presented some time after Golden Globes (up to 2 months later), winning a Golden Globe often forecasts winning an Oscar.

(a) Of the 64 Best Picture Oscar winners from 1943 to 2006, 34 had previously won the Golden Globe for Best Picture (Drama).
(b) The Golden Globe award for Best Picture was separated into two distinct categories in 1951: Drama and Musical or Comedy. Of the 56 Best Picture Oscar winners from 1951 to 2006, 10 had previously won the Golden Globe for Best Picture (Musical or Comedy).
(c) Of the 64 Best Director Oscar winners from 1943 to 2006, 35 had already won the Golden Globe for Best Director.
(d) Of the 129 Best Actor Oscar winners from 1943 to 2006, 41 males had previously won the Golden Globe for Best Actor (Drama) and 32 females had previously won the Golden Globe for Best Actress (Drama).
(e) Of the 115 Best Actor Oscar winners from 1950 to 2006, six males had previously won the Golden Globe for Best Actor (Musical or Comedy) and 12 females had previously won the Golden Globe for Best Actress (Musical or Comedy).

The Directors Guild of America (DGA) has been awarding its honours for Best Motion Picture Director since 1949 (with all except two early awards made before the announcement of the Best Picture Oscar). Since 1989, the Producers Guild of America has been awarding its honours to the year's most distinguished producing effort (with all except the first awarded before the announcement of the Best Picture Oscar). Since 1994, the Screen Actor's Guild has awarded five statuettes, which are 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.

(a) Of the 40 Best Picture Oscar winners from 1949 to 1988, 31 had already won a DGA award (and two would subsequently win one).
(b) Of the 18 Best Picture Oscar winners from 1989 to 2006, 10 had already won a Producers Guild of America award (and one would subsequently win one).
(c) Of the 58 Best Director Oscar winners from 1949 to 2006, 51 had already won a DGA award (and one would subsequently win one).
(d) Of the 26 Best Actor Oscar winners since 1994, nine males had already won a Screen Actor's Guild award and 10 females had already won one.

Nominees for Director and Lead Actor seem to have an *increased* chance of winning the more times they have been *nominated* in previous years, whereas nominees for Lead Actor and Lead Actress seem to have a *decreased* chance of winning the more times they have *won* in previous years. These variables were log-transformed because they are highly skewed.

(a) 17% of Best Director Oscar nominees with no previous directing nominations have won the Oscar, whereas 24% of Best Director Oscar nominees with one or more previous directing nominations have won.
(b) 20% of Best Actor Oscar nominees with no previous lead actor nominations have won the Oscar, whereas 22% of Best Actor Oscar nominees with one or more previous lead actor nominations have won.
(c) 23% of Best Actor Oscar nominees with no previous lead actor wins have won the Oscar, whereas 10% of Best Actor Oscar nominees with one or more previous lead actor wins have won.
(d) 24% of Best Actress Oscar nominees with no previous lead actress wins have won the Oscar, whereas 12% of Best Actress Oscar nominees with one or more previous lead actress wins have won.

The indicator variable for the first 'front-running movie' allows for the possibility that a nominee's chance of winning an Oscar could be linked to the fortunes of other nominees for the same movie. Each year often a handful of movies are 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 on the basis of previous Best Picture Oscar winners (e.g. the Best Director category usually ranks highly since Best Picture winners nearly always also have a Best Director nomination). Then, a 'nomination score' was calculated for each nominated movie on the basis of these rankings (e.g. 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 1 for all nominees who are associated with this movie. Indicator variables for the second and third front-running movies were derived similarly.

Although a variable for previous Best Director Oscar nominations was included, adding a variable for the number of previous Best Director Oscar *wins* tended to worsen rather than to improve predictions. Conversely, although a variable for previous Best Actress Oscar wins was included, adding a variable for the number of previous Best Actress Oscar *nominations* tended to worsen predictions. Also, although 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 (Markson and Taylor, 1993; Gilberg and Hines, 2000). For example, the median age of Best Actress Oscar winners between 1928 and 2006 was 33 years, whereas that for Best Actor was 42 years. However, within-gender age differences 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 winners was 29 years *versus* that of losing nominees of 33 years. Comparable figures for the second third (1954–1979) are 34 years *versus* 34 years, and for the final third (1980–2006) are 35 years *versus* 37 years. 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 years *versus* 38 years, for the second third are 43 years *versus* 39 years and for the final third are 43 years *versus* 45 years. 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 in the models failed to improve predictions of winners.

Other variables that were investigated but that did not improve results include supporting actor Oscar nominations and wins, genre of the nominated movie (drama, comedy, etc.), Motion Picture Association of America rating ('PG', 'R', etc.), running time (i.e. the length of the movie), 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). Some of these, although perhaps correlated to some extent with Oscar wins, failed to improve on variables that were already included—for example, of all the pre-Oscar awards, the Golden Globes and Guild awards are the most predictive of future Oscar wins. Other variables were excluded partly because of difficulties in obtaining reliable measurements over time—for example, it is difficult to find a long time series of consistent movie critic ratings that would have been available before a particular year's Oscar results.

3. Estimation

Our goal is to predict the winners of the four major Oscar categories for each year from 1938 to 2006 by using nominee information that is available before announcement of the winners. 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 5, although until 1936 the number of director and acting nominees varied between 3 and 8, and until 1944 the number of picture nominees varied between 5 and 10.

In this particular application, the explanatory variables in Table 1 take different values for different response (nominee) alternatives. McFadden (1974) proposed a discrete choice model for just such a case where explanatory variables are characteristics of the choice alternatives. 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 and actress) in each of the years (1938–2006).

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

$$\Pr(Y = j|\mathbf{x}_i) = \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ij}) \bigg/ \sum_{h \in C_i} \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{ih}), \tag{1}$$

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

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

Conditional on the choice being a or b, a variable's effect depends on the difference in the variable's values for those alternatives. 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. In contrast with this model in which the explanatory variables are characteristics of the choice alternatives, a similar model in which the explanatory variables are specific to the choice situation (and constant across alternatives) is the multinomial logit model of Nerlove and Press (1973). However, Maddala (1983) showed that the two models are equivalent and the distinction between them is somewhat artificial (see also Greene (2003), page 720). Both types of explanatory variable can be handled together within the same framework with appropriate use of interactions (although the resulting model is sometimes called a mixed logit (ML) model, we reserve this terminology for the hierarchical model that is discussed below). Within such a framework, the model is often just called the multi-nomial logit (MNL) model (e.g. Hausman and McFadden (1984)), and we follow that lead here (see also the reference guide for the NLOGIT software of Greene and Hensher (2002)).

The MNL model exhibits a property that is known as the *independence of irrelevant alternatives* (IIA) (Luce, 1959). For example, in a choice set containing two alternatives a and b, the addition of a third alternative can have no effect on the ratio $\Pr(Y = a|\mathbf{x}_i)/\Pr(Y = b|\mathbf{x}_i)$. In other words, the new alternative gains share proportionately from the choice shares of the existing alternatives in the set. There are contexts in which this property fails to describe observed behaviour. For example, suppose that two soft drink beverages are available in a choice set: one cola flavoured and the other lemon flavoured. The introduction of an alternative cola-flavoured soft drink (with a different name but otherwise indistinguishable from the existing cola) would most probably take most of its market share from the other cola rather than equally from both existing drinks. However, in the Oscars application, it seems reasonable to assume IIA, since nominees are unlikely to be considered close substitutes for one another. Exceptions to this might be nominated movies of the same genre that are closer substitutes than those from different genres or the relatively rare occasion when an individual receives multiple nominations in a category in the same year. To date, this latter phenomenon has happened only three times for Best Director (Clarence Brown in 1930, Michael Curtiz in 1938 and Steven Soderbergh in 2000); the Oscar rules prevent this from happening in the lead acting categories. IIA is also supported by the manner in which the winner is selected (using plurality voting) as the nominee who receives the most votes from all active and lifetime members of the AMPAS (see Gehrlein and Hemant (2004)). To evaluate the IIA assumption empirically, we applied the IIA test of Hausman and McFadden (1984) for the MNL model—results appear in Section 3.4.

An extension of the MNL model that places a probability distribution on some or all of the parameters $\boldsymbol{\beta}$ in equation (1) is the ML model, which is also known as the random-parameters, or random-coefficients or random-effects logit, or, from a Bayesian perspective, hierarchical MNL (see Hensher and Greene (2003), Revelt and Train (1998), Rossi *et al.* (2005) and Train

(2003)). (We use the ML terminology in this paper since it appears to be the most common in current discrete choice literature.) The ML model generalizes equation (1) so that the parameters $\boldsymbol{\beta}_i$ are specific to choice experiment i and the (unconditional) probability of selecting alternative j averages over a mixing distribution $f(\boldsymbol{\beta})$:

$$\Pr(Y=j|\mathbf{x}_i)=\int \frac{\exp(\boldsymbol{\beta}_i^{\mathrm{T}}\mathbf{x}_{ij})}{\sum\limits_{h\in C_i}\exp(\boldsymbol{\beta}_i^{\mathrm{T}}\mathbf{x}_{ih})}\, f(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta}. \tag{2}$$

Two important features of the ML model in this context are that it does not require the IIA assumption and it can approximate any random utility choice model to any degree of accuracy through appropriate specification of f (McFadden and Train, 2000). However, Hensher and Greene (2003), page 133, cautioned that ML models require 'extremely high quality data *if* the analyst wishes to take advantage of the extended capabilities of such models'. ML models are often most successfully applied in situations where the choice experiments i are represented by individuals making the choices, and either demographic data are available on the individuals, or the individuals make repeated choices over a sequence of similar choice experiments (i.e. panel data) or both. These additional data can facilitate estimation of the ML model by providing additional information on the distribution of the individual parameters $f(\boldsymbol{\beta})$. In the application that is considered in this paper, the 'individuals making the choices' are the Oscar competitions for each category within each year, which have no repeated measurements (each Oscar competition is essentially unique) and which have no obvious associated data equivalent to demographics. Thus, for this particular application, it is not clear that an ML model will necessarily outperform the MNL model (which is just a special case of the ML model with a degenerate distribution for f). Nevertheless, we compare the performance of both MNL and ML models in Section 3.4 later.

MNL and ML models can be fitted with a variety of statistical software packages. We experimented with two estimation methods: classical maximum likelihood by using NLOGIT (Greene and Hensher, 2002) and Bayesian estimation by using WinBUGS (Spiegelhalter *et al.*, 2003). NLOGIT is one of the main software packages for MNL estimation by using maximum likelihood (Hensher and Greene, 2003), whereas WinBUGS uses Bayesian estimation techniques that are based on Markov chain Monte Carlo simulation. Both packages are relatively easy to use, although they do require some limited programming. Other software packages for discrete choice modelling include the software language GAUSS (`www.aptech.com`) and the R package `bayesm` (Rossi *et al.*, 2005), although currently `bayesm` requires the number of alternatives in each choice set to be constant (which they are not for this application).

The `R2WinBUGS` functions of Sturtz *et al.* (2005) provide an interface between WinBUGS and R (R Development Core Team, 2005) that facilitates processing of data and results (`R2WinBUGS` is available as an R package at the Comprehensive R Archive Network (`http://cran.r-project.org`). This is particularly useful in this application since 69 models are fitted (one for each of the years 1938–2006). All data that were 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 data set 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 2006.

The following sections compare the two estimation methods using NLOGIT and WinBUGS for this application and also provide details on variable selection and model assessment. In the same spirit as Train and Sonnier (2005), we employ methods and interpret our results from both Bayesian and classical perspectives (see also Box (1980) and Rubin (1984)).

3.1. Maximum likelihood estimation by using NLOGIT

NLOGIT obtains maximum likelihood estimates for the MNL model by using Newton's method. Greene and Hensher (2002) have provided details of how to carry out the specification test of Hausman and McFadden (1984) to test the IIA assumption—results are in Section 3.4. We also used the programming capabilities of NLOGIT to carry out the Lagrange multiplier test that was proposed in McFadden and Train (2000) to determine whether mixing (of the MNL model) is needed (i.e. do the data suggest that extension to an ML model might be beneficial?). Again, results are in Section 3.4, which also contains some results for ML models; NLOGIT uses simulated maximum likelihood estimation to fit ML models (which it calls random-parameters logit models), with various distributions available for f (including normal, log-normal, uniform and triangular).

3.2. Bayesian estimation by using WinBUGS

WinBUGS uses Markov chain Monte Carlo techniques, specifically Gibbs sampling (see Casella and George (1992)) and the Metropolis–Hastings algorithm (see Chib and Greenberg (1995)), to estimate statistical models. In contrast with NLOGIT, which has the MNL and ML models 'built in', the user must program WinBUGS to estimate these models. However, the programming is relatively straightforward (for example, the MNL model that is used here is a mere 10 lines of code), and it is easy to extend the models in non-standard ways (for example, the variations for prior distributions and data weighting that are discussed below are relatively easy to carry out in WinBUGS, whereas they are not possible with NLOGIT). WinBUGS code for the MNL and ML models that are considered in this paper is available at `http://lcb1.uoregon.edu/ipardoe/research.htm`.

Another potential difficulty of using Bayesian estimation here is specification of the prior distributions for the model parameters. For the MNL model (1), standard non-informative normal priors for $\boldsymbol{\beta}$ (independent and centred at zero, with variance 10) produced stable results with reasonable predictive accuracy. The results in Section 4 are based on the last halves of three chains of 4000 simulations each (the first half of each chain—which was considered burn-in—was discarded). Since the 0.975-quantiles of the corrected scale reduction factor (Brooks and Gelman (1998), page 438) were each 1.1 or less, and trace plots showed good mixing of the three chains, convergence to stationary posterior distributions (all unimodal) seems likely.

It is also possible to use more informative prior distributions for the MNL models in this application. In particular, since 69 models are fitted, one after another, it is possible to use normal approximations of the posterior distributions of $\boldsymbol{\beta}$ for the model fit to predict year t as the prior distributions for the model fit to predict year $t+1$, and so on. Using such priors produced equivalent, but not better, results to those by using the non-informative priors that were discussed above.

The time series nature of the iterative estimation process also permits some modelling flexibility. The process as described uses *all* previous data for predicting any particular year's Oscars. However, it is possible that more accurate models might be estimated if older data were downweighted in some way relatively to more recent data. One approach to doing this might be to *weight* the data and to adjust the estimation process to take account of the weights in fitting the model. Experiments with weighting schemes of this nature failed to improve predictive accuracy, however. An alternative method for downweighting older data is to use a *moving window* approach whereby each model is fitted using just the previous N years of data. Setting N too low (say 30) for this application produced less stable parameter estimates with correspondingly worse predictions. Setting N too high (say using all previous data) might have produced parameter estimates that remain overly affected by very early Oscar voting patterns. However, systematic experimentation with the moving window length N ultimately suggested using all previous data.

Programming the ML model (2) in WinBUGS requires just a few extra lines of code to specify the distribution for f and the 'hyperprior' distribution for the parameters of the f-distribution. Any reasonable distribution can be used for f, although it is typically normal or log-normal. Section 3.4 compares results for MNL and ML models fitted with both NLOGIT and WinBUGS.

3.3. Variable selection

As indicated in Section 2, the explanatory variables enter the models at various points between 1938 and 2006. The main restriction on when a variable enters a model is the earliest date at which the variable is available. For example, since the first Golden Globes were for 1943 movies, the earliest that Golden Globe variables can be used is in the prediction of 1944 Oscars. However, variables were also omitted for years in which they provided little predictive power and counterintuitive parameter estimates. For example, although a high number of Oscar nominations generally improves the chance that a nominee will win a Best Picture or Best Director Oscar, this association only became established for the directing Oscar from 1938 onwards.

The general modelling strategy follows the insight of Box (1979), page 202, that 'all models are wrong, but some are useful'. We do not claim that our final selected model uniquely represents the actual voting dynamics of AMPAS members during the Oscar season. Yet we do hope that our modelling endeavours and results shed some light on interesting questions around the predictability of the four major Oscar categories, and whether this can tell us anything useful about the intent of the Oscars to recognize outstanding achievement in film. The approach that is used to determine which variables are included and excluded for each model for each year is based on standard regression modelling methodology such as that found in Weisberg (2005) or Dielman (2005). For example, variables were initially selected that represent all the phenomena that are used by Oscar prognosticators in the media (for which data are available). Given the relatively sparse nature of the response data—one choice from a set of 5 (usually) for each category over a limited time series—and the collinearity between some variables (e.g. Golden Globe for Best Director and DGA awards), many variable effects are poorly estimated (with relatively large standard errors). Such variables were then dropped from the model, although they could re-enter a model later in the time series if their parameter estimates became sufficiently large relative to their standard errors.

Although our approach was quite flexible in this regard (for example, we did not use a rigid criterion such as 5% significance), we believe that our general modelling strategy compares reasonably favourably for this application with a more formal econometric methodology such as that of Spanos (1986). For example, Spanos noted the six most important criteria for model selection as theory consistency, goodness of fit, predictive ability, robustness (including lack of collinearity), encompassing other empirical studies or models and parsimony. Our approach would appear to fare well with regard to these criteria. One remaining question is more problematic: are there any omitted variables that could be biasing the results, as might be revealed through non-random systematic errors? There are undoubtedly factors at play in Oscar voting dynamics that we fail to capture, such as studio advertising, temporal fads or trends, but there is little evidence to suggest that their absence from the modelling process produces systematic predictable errors. Residual analysis for logit models is notoriously challenging, but there is little to suggest a serious problem from examining the residuals for the models that are considered here (there might be scope for extending the graphical model checking ideas in Pardoe and

Cook (2002) to consider this more extensively, but such an extension lies beyond the scope of this paper).

3.4. Model assessment

Train (2003) recommended the *likelihood ratio index* to measure how well a discrete choice model fits the data:

$$\rho = 1 - \frac{\mathrm{LL}(\hat{\beta})}{\mathrm{LL}(\mathbf{0})},$$

where $\mathrm{LL}(\hat{\beta})$ is the value of the log-likelihood function at the estimated parameters and $\mathrm{LL}(\mathbf{0})$ is its value when all the parameters are set equal to 0. This statistic is somewhat analogous to R^2 in linear regression although it does not have the same 'explained variation' interpretation. To illustrate, the value of ρ for the MNL model fitted by using NLOGIT to all the data up to 2006 is 0.47. The analogous quantity for the same model fitted by using WinBUGS is also 0.47 (as noted in Section 3, to ease comparison of our maximum likelihood and Bayesian approaches we follow Train and Sonnier (2005) by evaluating Bayesian results by using both Bayesian and classical procedures).

The two sets of MNL model results (from NLOGIT and WinBUGS) are in fact very similar (which is not too surprising given the relatively uninformative priors that are used for the Bayesian approach). This begs the question, why use the more complicated Bayesian approach, when the simpler likelihood approach using NLOGIT offers similar results? There are two reasons why we prefer the Bayesian approach for this application.

(a) One extension which was discussed at the end of Section 3.2 (using informative priors for year $t+1$ based on results for year t) is only possible with the Bayesian approach (although this extension ultimately failed to improve overall results).
(b) Although results for models fitted to data up to more recent times are similar under both approaches, results for less recent times (when there were less data) diverged, sometimes considerably. In particular, NLOGIT was sometimes unable to make use of certain variables until a number of years after WinBUGS could make use of the same variable. To illustrate, DGA awards have been highly predictive of the Best Director Oscar since their inception (in 1949) and WinBUGS estimates for the corresponding parameter stabilized (with a relatively small standard error) from 1951 onwards. By contrast, NLOGIT estimates for the same parameter are highly unstable (with huge standard errors) until 1969. Consequently, WinBUGS can quite accurately predict Best Director Oscar winners over the period 1951–1969, but NLOGIT is far less accurate, relying instead on basing predictions on the Golden Globe Best Director winner (which is less clearly associated with the Best Director Oscar winner). WinBUGS can obtain stable estimates that NLOGIT cannot because the Bayesian approach can be thought of as 'shrinking' the maximum likelihood estimate towards the prior mean (0 in this case), particularly in early years when the data do not (yet) dominate the prior.

We next considered the IIA assumption of the MNL model by conducting the specification test of Hausman and McFadden (1984). For completeness, we again present results for both NLOGIT and WinBUGS. The test works by re-estimating the model with a smaller set of choices than in the complete data set; if IIA holds, results should be similar, and a statistic that is based on differences in the model parameter estimates and covariance matrices allows significance to be determined by using the χ^2-distribution. In this application, we can restrict the choice set by dropping one nominee in each year–category. Since the ordering of nominees within category contains no statistical information, we randomly selected one nominee to drop in each year–category (dropping the entire year–category if the nominee dropped happened to be the winner). To increase our confidence in the results, we conducted five such tests. For NLOGIT, p-values for the five tests for the MNL model fitted to all the data up to 2006 were 0.57, 1.00, 0.15, 1.00 and 0.99. The analogous quantities by using WinBUGS were 0.56, 1.00, 0.70, 1.00 and 0.15 (the values of 1.00 actually corresponded to negative test statistics; Greene and Hensher (2002) suggested that in such cases the right conclusion is probably that the test statistic should be 0). Both sets of results suggest that the IIA assumption seems reasonable for this application (the values of 0.15 raise a slight question, so we investigate further by considering ML models below).

Although these tests suggest that the MNL IIA assumption seems reasonable, the slight question that is raised by the 0.15-value motivated us to investigate ML models in addition. We applied the Lagrange multiplier test of McFadden and Train (2000) to determine whether mixing (of the MNL model) is suggested. The test works by first creating artificial variables that are based on squared deviations of explanatory variables from their weighted mean (the weights being the MNL fitted probabilities). An MNL model including the artificial variables can then be compared with the original MNL model by using a likelihood ratio test; if mixing seems unnecessary, the artificial variables should have estimates that are close to 0 and the two models should have a similar fit. Although the test is not designed to identify specifically *which* variables might benefit from mixing, McFadden and Train (2000) used t-statistics exceeding 1 to signal potential candidates for mixing in their example. For this application, the only variables that come close to benefiting from mixing are the first two 'front-running movie' indicators that cut across all four Oscar categories. For NLOGIT, including two artificial variables for these two variables for the MNL model fitted to all the data up to 2006 produces t-statistics that both (just) exceed 1, but the likelihood ratio test for their inclusion results in a p-value of 0.13. The analogous value by using WinBUGS is 0.41 (with one t-statistic close to 1 but the other close to 0). The results suggest that the data do not particularly support extension of the MNL model to an ML model.

A further check on whether the data support an ML model over the MNL model that is sometimes applied is a standard likelihood ratio test comparison between the two models. The most reasonable ML model based on extensive exploratory analysis has a multivariate normal mixing distribution for the first two front-running movie indicator variables (we need to allow for correlation since, in any particular year, a positive effect from being the first front-running movie is likely to be coupled with a negative effect from being the second front-running movie, and vice versa). For NLOGIT, the resulting p-value for models fitted to all the data up to 2006 is 0.48; the analogous value for WinBUGS is 0.34.

However, it is not clear that such a test is valid here since in the hierarchical ML model the number of model parameters is ambiguous. These test results assume that the additional number of parameters in the ML model is 3 (for two standard deviation parameters and one correlation parameter). But, from an alternative viewpoint, there is one additional parameter for each year–category for each front-running movie indicator variable (since each year–category has its own separate front-running movie impacts estimated). From a Bayesian perspective, the *effective number of parameters* lies somewhere in between and is calculated as twice the difference between the value of the (conditional) log-likelihood at the parameter posterior means and the posterior mean of the (conditional) log-likelihood—see Spiegelhalter *et al.* (2002). The *deviance information criterion* DIC is then minus twice the posterior mean of the (conditional) log-likelihood (i.e. the posterior mean of the deviance) plus the effective number of parameters. Given the difficulty in counting the number of parameters in Bayesian hierarchical models, Spiegelhalter *et al.* (2002) argued that the deviance information criterion is a more appropriate measure of goodness of fit for such models than Akaike's information criterion or the Bayesian information criterion. For this application, the value of the deviance information criterion for the WinBUGS MNL model fitted to all the data up to 2006 is 589, whereas for the comparable ML model it is 536. In contrast with the previous results, this actually suggests that the ML model does provide a better fit than the MNL model, even taking into account the additional complexity of the ML model.

Despite the DIC results, it is possible that, although the ML model may provide a superior fit *within sample*, the MNL model may prove more useful at predicting Oscar winners out of sample. Train (2003) cautioned against assessing goodness of fit for discrete choice models by using 'per cent correctly predicted'. For the usual applications for discrete choice models, the rationale is that

> 'in stating choice probabilities, the researcher is saying that if the choice situation were repeated numerous times (or faced by numerous people with the same attributes), each alternative would be chosen a certain proportion of the time'

(Train (2003), page 73). This is not quite the same as predicting the choice alternative with the highest fitted probability. However, in this application, the 'alternatives chosen a certain proportion of the time' correspond to the proportion of votes that are cast by AMPAS members for each nominee, and this is not really the focus of interest here (and, given the secret nature of the ballot, this is something that will always be unknowable). Since the focus of interest is rather the predictability of the single winner for each category in each year, it seems more reasonable to use per cent correctly predicted as a goodness-of-fit criterion in this context. Thus, to assess the predictive accuracy of the various modelling choices that have just been described, one-year-ahead, out-of-sample errors were used. For example, the four major Oscars winners for 1938 were predicted from a model fitted to data from 1928–1937. Then, the winners for 1939 were predicted from a model fitted to data from 1928–1938, and so on.

Using the modelling approach that was described in Section 3 for the MNL model in WinBUGS, 190 of the 276 Best Picture, Director, Actor and Actress Oscar winners from 1938 to 2006 were correctly identified, corresponding to an overall prediction accuracy of 69%. This compares with 185 (67%) for the MNL model in NLOGIT and 186 (67%) for the ML model in WinBUGS. The main reason for the reduced performance of the MNL model in NLOGIT appears to be related to the inability of the estimation method to use some variables (e.g. DGA wins) early in their history. The main reason for the reduced performance of the ML model in WinBUGS appears to be related to the nature of the data in the context of the ML model. In particular, in ML applications with repeated choices, it is possible to condition on an individual's previous choices to improve predictions of that individual's future (out-of-sample) choices (Revelt and Train, 2000). That is not possible here since each Oscar competition is essentially unique and there are no repeated choices in this sense.

These results slightly favour the MNL model over the ML model, and also Bayesian estimation using WinBUGS over maximum likelihood estimation using NLOGIT. This confirms most of the earlier results comparing the various modelling methods (the only exception being the deviance information criterion favouring the ML model). Since most of the model assessment methods favour the MNL model in WinBUGS, and noting the advice of Hensher and Greene (2003), page 171, that 'regardless of what is said about advanced discrete choice models, the MNL model should always be the starting point for empirical investigation', we selected this as our final modelling method.

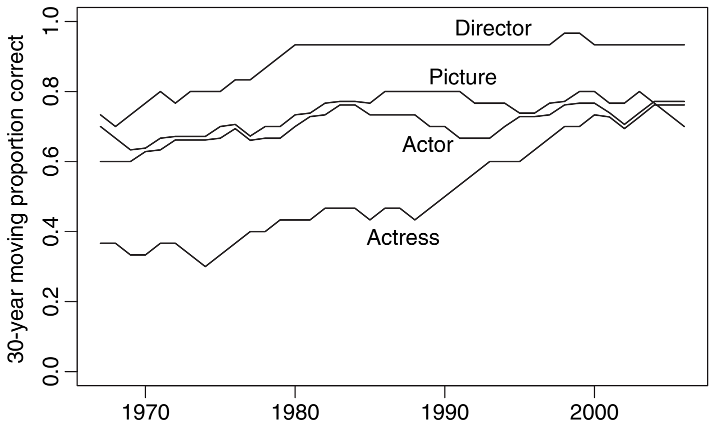

Fig. 1. 30-year moving averages of the proportion of correct predictions in each of the four major Oscar categories: the moving average values are placed at the ends of the 30-year periods (for example, at the far right-hand side of the graph the proportions of correct predictions over the period 1977–2006 are 93% for Best Director, 77% for Best Actor, 77% for Best Actress and 70% for Best Picture)

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 (1977–2006) is 95 correct predictions out of 120, or 79%. Fig. 1 summarizes overall results across the four categories. Overall, the Best Director Oscar has been the most predictable; then the Oscars for Best Picture (until recently), Best Actor and Best Actress. Each of the categories has tended to become more predictable over time, particularly Best Actress, which was very difficult to predict until the early 1970s (see Simonton (2004c) for some discussion of the contrast between movies with Best Actor and Best Actress nominations). The failure of the model to predict the Best Picture winner for the last three years has led to a recent reversal in the predictability trend for this category.

4. Results

From the modelling process that was described in Section 3, the roles of the explanatory variables in helping to predict Oscar winners can change over time; Fig. 2 illustrates with LOWESS smooths of posterior medians for the model parameters (we use medians here rather than means since some posterior distributions were slightly positively skewed). 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 curves labelled ‘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 (curves labelled ‘N’). Previous wins seemed to hurt Best Actor nominees less in the 1960s and 1970s than in the 1940s and more recently, whereas previous wins have tended to become less important for Best Actress nominees over time (curves labelled ‘W’).

The Golden Globes have remained useful predictors of future Oscar success since their inception. The changing fortunes of dramas (curves labelled ‘D’) and musicals and comedies (curves labelled ‘M’) can be traced in Fig. 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 favour dramas, particularly for males. Guild awards have clearly enabled quite accurate prediction of Best Director winners, and to a lesser extent Best Picture winners (curves labelled ‘G’). Since

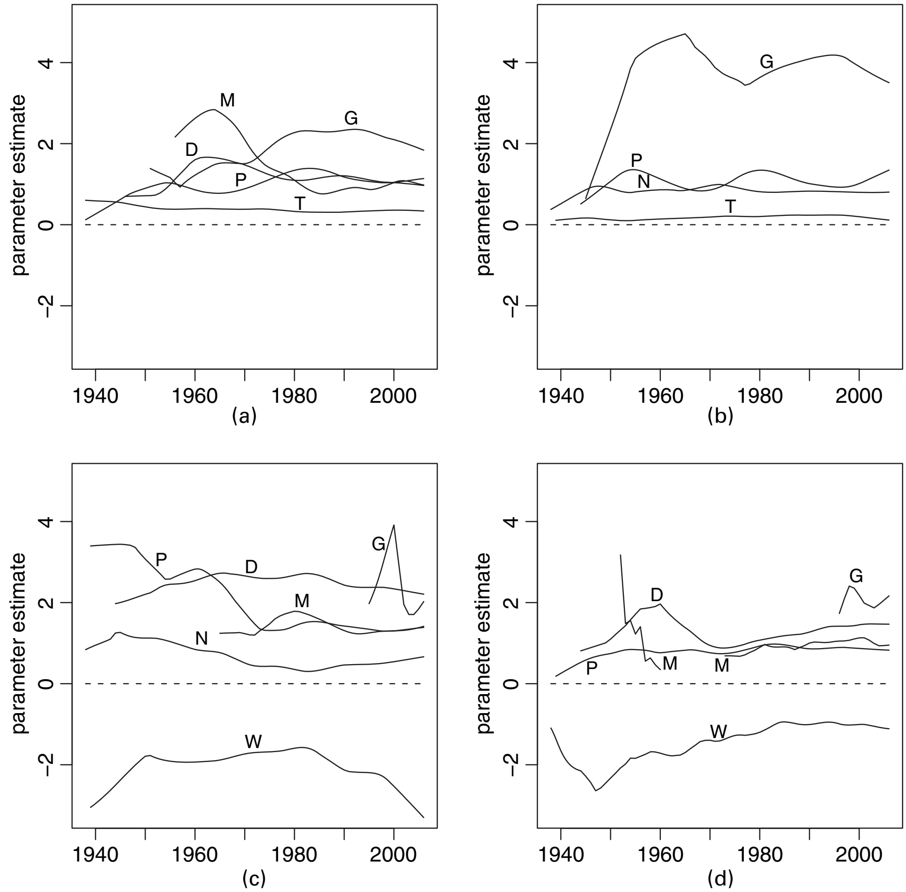

Fig. 2. Smoothed parameter estimates—posterior medians—for the explanatory variable for each of the four major Oscar categories (the explanatory variables are described in Section 2; T, total nominations; P, director or picture; N, previous nominations; W, previous wins; D, Globe drama; M, Globe musical or comedy; G, Guild award): (a) picture; (b) director; (c) actor; (d) actress

they have had a much shorter history, it is not clear whether Screen Actor's Guild awards will be just as helpful in predicting acting wins, although early indications suggest so.

The effect of the total number of Oscar nominations (curves labelled 'T') on prediction of the Best Picture and Best Director Oscars remains reasonably steady. Since the number of nominations that 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 *1*). The effects of the 'front-runner' variables—which cut across all four categories—are not shown in Fig. 2 (they appear less important than the other variables, with smaller magnitude estimates 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 2 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 `lcb1.uoregon.edu/ipardoe/oscars/` (which will be updated with future predictions in February of each year).

Table 2. 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*	*Probability*	*Predicted winner*	*Probability*
Best Picture				
1948	*Hamlet*	0.01	*Johnny Belinda*	0.97
2004	*Million Dollar Baby*	0.01	*The Aviator*	0.97
1981	*Chariots of Fire*	0.01	*Reds*	0.88
Best Director				
2000	Steven Soderbergh	0.01	Ang Lee	0.95
1968	Carol Reed	0.02	Anthony Harvey	0.97
1972	Bob Fosse	0.03	Francis Ford Coppola	0.96
Best Actor				
2001	Denzel Washington	0.00	Russell Crowe	0.99
1968	Cliff Robertson	0.00	Peter O'Toole	0.88
1974	Art Carney	0.02	Jack Nicholson	0.87
Best Actress				
2002	Nicole Kidman	0.07	Renée Zellweger	0.90
1985	Geraldine Page	0.07	Whoopi Goldberg	0.70
1950	Judy Holliday	0.09	Gloria Swanson	0.76

5. Discussion

Discrete choice modelling 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 surprisingly high degree of success. If recent trends persist, it should be possible to predict future winners with a prediction success rate of approximately 70% for picture, 93% for director, 77% for actor and 77% for actress. Interestingly, although predictive accuracy has been largely increasing (Fig. 1), the parameter estimates for predictors often exhibit non-monotonic trends (Fig. 2). Hence, the results indicate that the dynamics are more complicated than would be expected if it were a simple matter of the Academy voters slowly converging on the assessments underlying the Golden Globes and Guild awards. The findings also imply a greater complexity in the process underlying the distribution of the awards across films than suggested by cumulative advantage effects that are produced by opinion diffusion models (see, for example, Collins and Hand (2006)).

It is also noteworthy that the historical changes in these parameters do not display the same trajectories for the four categories. Most strikingly, the Best Actor and Best Actress categories not only have different predictors, and different parameter estimates for the predictors that they share, but also the changes in the parameter estimates do not always follow the same trend (e.g. the effect of previous wins). Such differences may eventually provide an explanation for gender differences in the degree that acting awards or nominations (whether lead or supporting) are coupled with a Best Picture award or nomination (Simonton, 2004c).

As pointed out by a referee, predictive modelling of Oscar outcomes is an example of a 'beauty contest' (in the parlance of information economics and game theory). The object is not to choose who *should* win, but to predict the behaviour of Academy voters (i.e. to choose who *will* win). Media prognosticators often make the same distinction and offer their readers or viewers two such lists of Oscar predictions. Our focus on who will win informs our modelling strategy (e.g. select those variables that have historically been most likely to provide accurate predictions), and also helps to shed some light on when certain Oscar categories seem to go awry (e.g. mismatches between who will win and who should win). We discuss this latter point in more detail below (see also Table 2).

Nevertheless, parameter estimates are generally consistent with commonly held views about the Oscars. Heavily nominated films with complementary Best Picture and Best Director nominations tend to do well, and nominees with previous Golden Globe or Guild awards are also favoured. To the extent that the corresponding predictors reflect some measure of merit, the Oscar process would seem to be meeting its goal of recognizing outstanding achievement in film. Further, the results support the notion that previous Oscar nominations benefit directors and actors, whereas previous Oscar wins reduce the chance that an actor or actress will win again. However, the actual merit of an acting performance should probably not be contingent on how many times an individual was nominated in the past or on how many times he or she has won the award before. This would seem to indicate that the directing and lead acting Oscars may be as reflective of achievement over a lifetime as much as achievement for a particular film. Finally, actors and actresses who are nominated for performances in heavily nominated films (particularly those with a Best Picture nomination) do indeed fare better than their peers in less highly acclaimed films. Again, this observation might seem to be at odds with a purely objective assessment of the quality of an acting performance.

The results also help us to assess those occasions when the Oscars fail to be awarded in a manner that is consistent with true merit. For example, there has been much media speculation about legendary individuals who have never won an Oscar, such as Alfred Hitchcock with five directing nominations, Peter O'Toole with eight lead actor nominations, Richard Burton with six lead actor nominations and Deborah Kerr with six lead actress nominations. According to the predictions, the unluckiest was probably O'Toole who came closest to winning in 1968 (with an 89% modelled probability of winning) and 1964 (with 83% probability). Kerr came close in 1956 (with 72% probability), as did Burton in 1977 (with 62% probability), and Hitchcock's nearest miss was for *Rebecca* in 1940 (with 42% probability). In some instances the Academy decided to make corrections for various inadvertent false negative results. In particular, Hitchcock, Kerr and O'Toole were eventually awarded honorary Oscars as compensation.

To be sure, sometimes the errors of prediction may be too great, making the Oscars look worse than they really deserve. Although many were surprised when Denzel Washington won over Russell Crowe in the 2001 Oscar race for Best Actor, the surprise may not be nearly as extreme as implied by the model predictions in Table 2. Another example is the failure of *Brokeback Mountain* to win Best Picture for 2005 after winning both Golden Globe and Producers Guild of America awards. Nevertheless, the surprise of *Crash* winning instead was not as great as that implied by the model predictions of 0.03 probability for *Crash versus* 0.90 probability for *Brokeback Mountain*. Clearly, the model was unable to make use of the late surge that *Crash* made and the apparent backlash against *Brokeback Mountain* (in unquantifiable 'Hollywood buzz' terms) as the Oscar ceremony drew near. In so far as such intangibles confound the Oscar outcomes, complete predictive precision is probably unattainable. Nonetheless, it is also conceivable that the addition of new predictors can reduce the number and magnitude of observed errors. In addition, it would be instructive to extend the analyses that were described in this paper to other Oscar categories, such as the supporting acting and screenwriting awards.

However, future research on this question should not just focus on empirical prediction. In the long run, theoretical explanation would be no less important. For instance, it would be important to identify any causal process that underlies the various awards. Previous research has suggested two alternative kinds of process (Simonton, 1991). On the one hand, the various award measures for a given category could be the consequence of a single underlying factor or latent variable that indicates the actual relative merit of the pictures, directors, actors and actresses. The Oscar awards and nominations would then be a function of this latent factor plus some random-shock or error term (sympathy votes, current events, etc.). On the other hand, the diverse awards may reflect a more dynamic process—which is a reasonable hypothesis given that the award ceremonies are not all on the same day. Because the Oscars are the last awards that are bestowed in a given year, the Academy voters may benefit in a substantial way from previous awards and nominations, or even by rumours of awards and nominations. According to this model, the merit of a picture, director, actor or actress is not a stable attribute behind all awards, but rather a transient assessment that evolves over time. Although prior empirical investigations have supported the former, single-factor, model over the latter, auto-regressive, model, these two theoretical accounts have not yet been tested for movie awards.

Improving both prediction and explanation may prove more valuable than just satisfying our intellectual curiosity of a highly visible event. Conceivably, predictive and explanatory models can indicate the specific ways that the Oscars can go wrong. This knowledge might lead to recommendations about how to improve the Academy's selection and voting process. Even if the Academy were unwilling or unable to introduce the necessary improvements, these prediction methods can provide a useful antidote to the dramatic announcements of Oscar night. In particular, the discrete choice models can help filmmakers and moviegoers alike to assess how much faith they can place in the identity of the 'best' picture, director, actor or actress. To what extent does a given award reflect the merits of the actual cinematic performance rather than represent less relevant processes operating among Academy voters? Given the tremendous consequences of taking an Oscar home, these results would have far more than academic interest.

References

Bennett, K. L. and Bennett, J. M. (1998) And the winner is. . . : a statistical analysis of the best actor and actress Academy Awards. *Statistics*, **23**, 10–17.

Box, G. E. P. (1979) Robustness in the strategy of scientific model building. In *Robustness in Statistics* (eds R. Launer and G. Wilkinson). New York: Academic Press.

Box, G. E. P. (1980) Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *J. R. Statist. Soc.* A, **143**, 383–430.

Brooks, S. P. and Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *J. Computnl Graph. Statist.*, **7**, 434–455.

Casella, G. and George, E. I. (1992) Explaining the Gibbs sampler. *Am. Statistn*, **46**, 167–174.

Caves, R. (2000) *Creative Industries*. Cambridge: Harvard University Press.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm. *Am. Statistn*, **49**, 327–335.

Collins, A. and Hand, C. (2006) Vote clustering in tournaments: what can Oscar tell us? *Creatvty Res. J.*, **18**, 427–434.

Collins, A., Hand, C. and Snell, M. C. (2002) What makes a blockbuster?: economic analysis of film success in the United Kingdom. *Mangrl Decisn Econ.*, **23**, 343–354.

Deuchert, E., Adjamah, K. and Pauly, F. (2005) For Oscar glory or Oscar money. *J. Cultrl Econ.*, **29**, 159–176.

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

De Vany, A. S. (2006) The movies. In *Handbook on the Economics of Art and Culture* (eds V. Ginsburgh and D. Throsby). Amsterdam: North-Holland.

Dielman, T. (2005) *Applied Regression Analysis for Business and Economics*, 4th edn. Pacific Grove: Duxbury.

Dodds, J. C. and Holbrook, M. B. (1988) Whats an Oscar worth?: an empirical estimation of the effects of nominations and awards on movie distribution and revenues. *Curr. Res. Film Aud. Econ. Law*, **4**, 72–87.

Gehrlein, W. V. and Hemant, V. K. (2004) Decision rules for the Academy Awards versus those for elections. *Interfaces*, **34**, 226–234.

Gilberg, M. and Hines, T. (2000) Male entertainment award winners are older than female winners. *Psychol. Rep.*, **86**, 175–178.

Ginsburgh, V. (2003) Awards, success and aesthetic quality in the arts. *J. Econ. Perspect.*, **17**, 99–111.

Greene, W. (2003) *Econometric Analysis*, 5th edn. Englewood Cliffs: Prentice Hall.

Greene, W. and Hensher, D. (2002) *NLOGIT Version 3.0 Reference Guide*. Plainview: Econometric Software.

Hausman, J. and McFadden, D. (1984) Specification tests for the multinomial logit model. *Econometrica*, **52**, 1219–1240.

Hensher, D. A. and Greene, W. H. (2003) The mixed logit model: the state of practice. *Transportation*, **30**, 133–176.

Litman, B. R. (1983) Predicting success of theatrical movies: an empirical study. *J. Poplr Cult.*, **16**, 159–175.

Luce, R. D. (1959) *Individual Choice Behavior*. New York: Wiley.

Maddala, G. S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Markson, E. W. and Taylor, C. A. (1993) Real versus reel world: older women and the Academy Awards. *Women Ther.*, **14**, 157–172.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (ed. P. Zarembka), pp. 105–142. New York: Academic Press.

McFadden, D. and Train, K. (2000) Mixed MNL models for discrete response. *J. Appl. Econ.*, **15**, 447–470.

Nelson, R. A., Donihue, M. R., Waldman, D. M. and Wheaton, C. (2001) What's an Oscar worth? *Econ. Inq.*, **39**, 1–16.

Nerlove, M. and Press, J. (1973) Univariate and multivariate log-linear and logistic models. *Report R-1306-EDA/NIH*. RAND Corporation, Santa Monica.

Pardoe, I. (2005) Just how predictable are the Oscars? *Chance*, **18**, 32–39.

Pardoe, I. and Cook, R. D. (2002) A graphical method for assessing the fit of a logistic regression model. *Am. Statistn*, **56**, 263–272.

Peary, D. (1993) *Alternate Oscars: One Critic's Defiant Choices for Best Picture, Actor, and Actress from 1927 to the Present*. New York: Delta.

Prag, J. and Casavant, J. (1994) An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. *J. Cult. Econ.*, **18**, 217–235.

R Development Core Team (2005) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Revelt, D. and Train, K. (1998) Mixed logit with repeated choices: households' choices of appliance efficiency level. *Rev. Econ. Statist.*, **80**, 647–657.

Revelt, D. and Train, K. (2000) Specific taste parameters and mixed logit. *Technical Report E00–274*. Department of Economics, University of California, Berkeley.

Rossi, P. E., Allenby, G. M. and McCulloch, R. E. (2005) *Bayesian Statistics and Marketing*. Hoboken: Wiley.

Rubin, D. B. (1984) Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, **12**, 1151–1172.

Schulze, G. (2005) Nobody knows anything—or do we?: introduction to the special issue on the movie industry. *J. Cult. Econ.*, **29**, 157–158.

Simonoff, J. S. and Sparrow, I. R. (2000) Predicting movie grosses: winners and losers, blockbusters and sleepers. *Chance*, **13**, no. 3, 15–24.

Simonton, D. K. (1991) Latent-variable models of posthumous reputation: a quest for Galton's G. *J. Personlty Socl Psychol.*, **60**, 607–619.

Simonton, D. K. (2002) Collaborative aesthetics in the feature film: cinematic components predicting the differential impact of 2,323 Oscar-nominated movies. *Empir. Stud. Arts*, **20**, 115–125.

Simonton, D. K. (2004a) Film awards as indicators of cinematic creativity and achievement: a quantitative comparison of the Oscars and six alternatives. *Creatvty Res. J.*, **16**, 163–172.

Simonton, D. K. (2004b) Group artistic creativity: creative clusters and cinematic success in 1,327 feature films. *J. Appl. Socl Psychol.*, **34**, 1494–1520.

Simonton, D. K. (2004c) The "best actress" paradox: outstanding feature films versus exceptional performances by women. *Sex Rol.*, **50**, 781–794.

Simonton, D. K. (2005a) Cinematic creativity and production budgets: does money make the movie? *J. Creatv. Behav.*, **39**, 1–15.

Simonton, D. K. (2005b) Film as art versus film as business: differential correlates of screenplay characteristics. *Empir. Stud. Arts*, **23**, 93–117.

Sochay, S. (1994) Predicting the performances of motion pictures. *J. Media Econ.*, **7**, 1–20.

Spanos, A. (1986) *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.

Spiegelhalter, D. J., Thomas, A., Best, N. G. and Lunn, D. (2003) *WinBUGS Version 1.4 User Manual*. Cambridge: Medical Research Council Biostatistics Unit.

Sturtz, S., Ligges, U. and Gelman, A. (2005) R2WinBUGS: a package for running WinBUGS from R. *J. Statist. Softwr.*, **12**. (Available from `http://www.jstatsoft.org/`.)

Terry, N., Butler, M. and De'Armond, D. (2005a) The determinants of domestic box office performance in the motion picture industry. *Sthwestrn Econ. Rev.*, **32**, 137–148.
Terry, N., Butler, M. and De'Armond, D. (2005b) The determinants of worldwide box office performance in the motion picture industry. *Sthwest Rev. Int. Bus. Res.*, **16**, 195–204.
Train, K. (2003) *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.
Train, K. and Sonnier, G. (2005) Mixed logit with bounded distributions of correlated partworths. In *Applications of Simulation Methods in Environmental Resource Economics* (eds A. Alberini and R. Scarpa), ch. 7, pp. 117–134. Dordrecht: Springer.
Walls, W. D. (2005) Modeling movie success when "nobody knows anything": conditional stable-distribution analysis of film returns. *J. Cult. Econ.*, **29**, 177–190.
Weisberg, S. (2005) *Applied Linear Regression*, 3rd edn. Hoboken: Wiley.