

PREDICTING ACADEMY AWARD WINNERS USING DISCRETE CHOICE MODELING

Iain Pardoe

Lundquist College of Business
University of Oregon

August 10, 2005

`ipardoe@lcbmail.uoregon.edu`

`http://lcb1.uoregon.edu/ipardoe/`

Presentation Outline

1. Introduction
2. Data
3. Estimation
4. Results

1. Introduction

- Academy of Motion Picture Arts and Sciences (AMPAS)
- Since 1928, recognized outstanding achievement in film
- Academy Awards, or Oscars, in various categories, e.g. Best Picture, Director, Lead Actor, Lead Actress
- Nominees selected by AMPAS members early in the year
- Winners (voted by AMPAS members) announced at ceremony in late Feb
- Intense media/public interest in predicting winners

2. Data

- Best Picture
 - Total Oscar nominations
 - Best Director Oscar nomination
 - Best Picture (drama) Golden Globe
 - Best Picture (musical or comedy) Golden Globe
 - Guild award (PGA or DGA)
- Best Director
 - Total Oscar nominations
 - Best Picture Oscar nomination
 - Previous Best Director Oscar nominations
 - Guild award (DGA)

- Best Actor in a Leading Role
 - Best Picture Oscar nomination
 - Previous Best Actor Oscar nominations
 - Previous Best Actor Oscar wins
 - Best Actor (drama) Golden Globe
 - Best Actor (musical or comedy) Golden Globe
 - Guild award (SAG)
- Best Actress in a Leading Role
 - Best Picture Oscar nomination
 - Previous Best Actress Oscar wins
 - Best Actress (drama) Golden Globe
 - Best Actress (musical or comedy) Golden Globe
 - Guild award (SAG)

- All four categories
 - Indicator for the first “front-running movie”
 - Indicator for the second “front-running movie”
 - Indicator for the third “front-running movie”
- Excluded variables
 - Previous Best Director Oscar wins
 - Previous Best Actress Oscar nominations
 - Total Oscar nominations for predicting acting winners
 - Age of Actor/Actress nominees
 - Supporting actor nominations/wins, genre, MPAA rating, length, release date, critic ratings, other awards

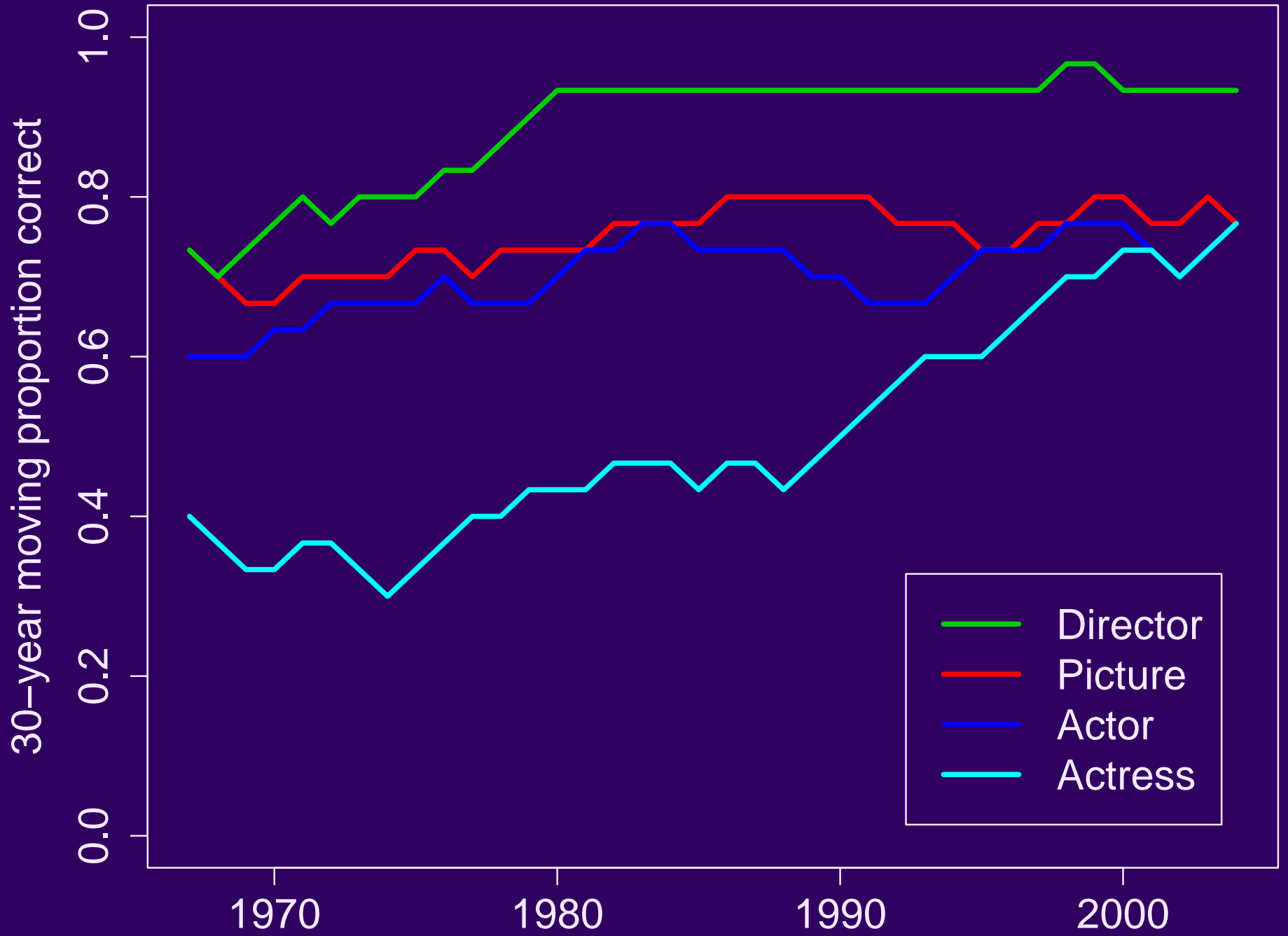
3. Estimation

- McFadden's conditional logit model
(more commonly called multinomial logit now)
- $\Pr(Y = j | \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_{ij})}{\sum_{h \in C_i} \exp(\boldsymbol{\beta}^T \mathbf{x}_{ih})}$
- IIA seems reasonable for this application
- Data up to 1937 used to fit model to predict 1938 winners
Append 1938 results, update model to predict 1939 winners
and so on . . . up to 2004 winners
- Use R and Bugs (flexibility, convenience, familiarity)

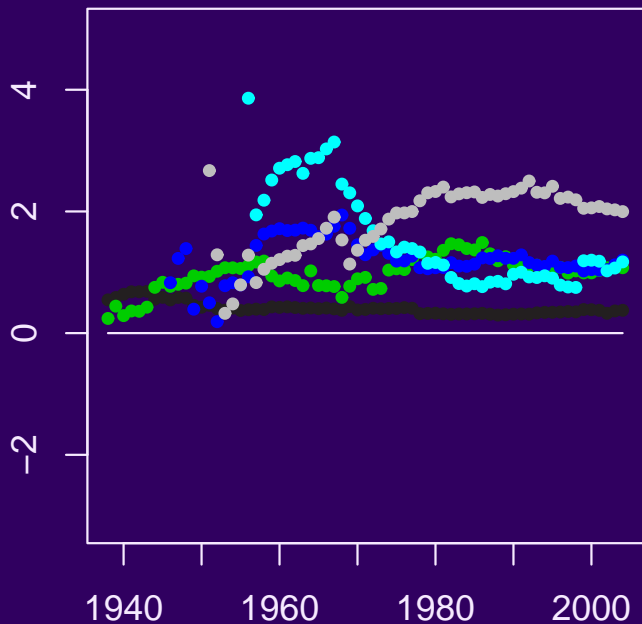
- Priors
 - noninformative (mean zero, variance ten)
 - informative using previous year's model posteriors
- Time series
 - use weights to down-weight older data
 - moving window uses only previous N years of data
- Variable selection
 - availability (e.g., Golden Globes 1944 on, SAG 1995 on)
 - sufficient predictive power, expected coefficient sign
- Predictive accuracy
 - minimize one-year-ahead out-of-sample errors

4. Results

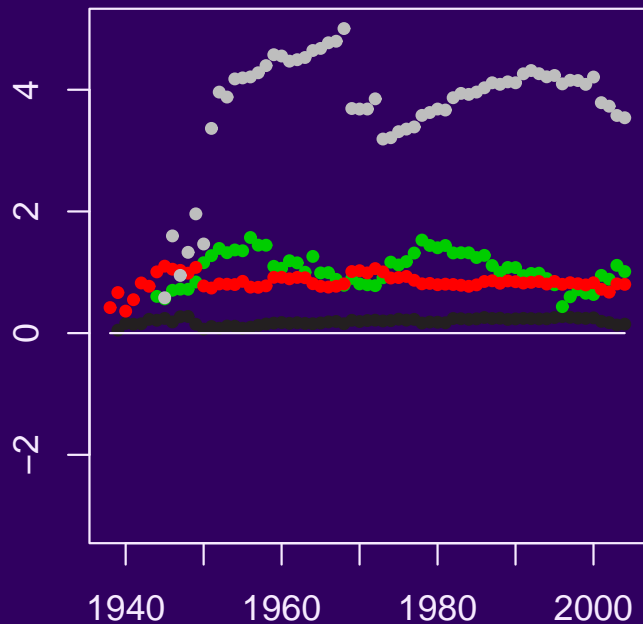
- Correctly identified 186 of the 268 Best Picture, Director, Actor, and Actress Oscar winners from 1938–2004
- Corresponds to an overall prediction accuracy of 69%
- Prediction accuracy has improved over time (see plot on next slide)
- Overall prediction accuracy for last 30 years (1975–2004) is 97/120, or 81%
- Role of explanatory variables in helping to predict Oscar winners changes over time (see plot on next slide but one)



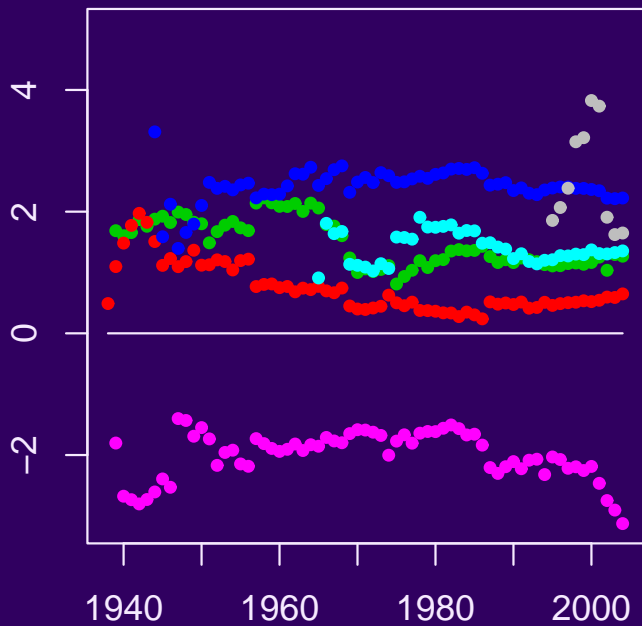
Picture



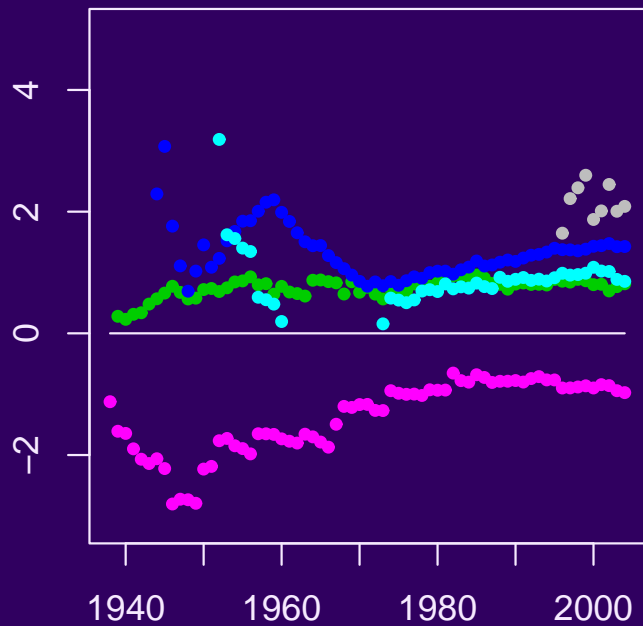
Director



Actor



Actress



- Noms
- Dir/Pic
- Prev nom
- Prev win
- GG dra
- GG m/c
- Guild

Winners in four major categories by year (dark fill indicates incorrect prediction)

| Year | Category | Actual Winner | Prob | Predicted Winner | Prob |
|------|----------|------------------------|------|----------------------------|------|
| 2004 | Picture | Million Dollar Baby | 0.02 | The Aviator | 0.97 |
| | Director | Clint Eastwood | 0.95 | Clint Eastwood | 0.95 |
| | Actor | Jamie Foxx | 0.56 | Jamie Foxx | 0.56 |
| | Actress | Hilary Swank | 0.95 | Hilary Swank | 0.95 |
| 2003 | Picture | The Return of the King | 0.94 | The Return of the King | 0.94 |
| | Director | Peter Jackson | 0.92 | Peter Jackson | 0.92 |
| | Actor | Sean Penn | 0.85 | Sean Penn | 0.85 |
| | Actress | Charlize Theron | 0.88 | Charlize Theron | 0.88 |
| 2002 | Picture | Chicago | 0.95 | Chicago | 0.95 |
| | Director | Roman Polanski | 0.02 | Rob Marshall | 0.92 |
| | Actor | Adrien Brody | 0.32 | Daniel Day-Lewis | 0.49 |
| | Actress | Nicole Kidman | 0.08 | Renee Zellweger | 0.90 |
| 2001 | Picture | A Beautiful Mind | 0.25 | The Fellowship of the Ring | 0.38 |
| | Director | Ron Howard | 0.92 | Ron Howard | 0.92 |
| | Actor | Denzel Washington | 0.00 | Russell Crowe | 0.99 |
| | Actress | Halle Berry | 0.36 | Halle Berry | 0.36 |
| 2000 | Picture | Gladiator | 0.97 | Gladiator | 0.97 |
| | Director | Steven Soderbergh | 0.01 | Ang Lee | 0.95 |
| | Actor | Russell Crowe | 0.62 | Russell Crowe | 0.62 |
| | Actress | Julia Roberts | 0.93 | Julia Roberts | 0.93 |

Best Picture winners: ordered from most to least surprising

| Year | Actual Winner | Prob | Predicted Winner | Prob | Ratio |
|------|-----------------------------|------|----------------------------|------|-------|
| 1948 | Hamlet | 0.01 | Johnny Belinda | 0.97 | 0.01 |
| 1981 | Chariots of Fire | 0.01 | Reds | 0.87 | 0.01 |
| 2004 | Million Dollar Baby | 0.02 | The Aviator | 0.97 | 0.02 |
| 1967 | In the Heat of the Night | 0.02 | The Graduate | 0.94 | 0.02 |
| 1951 | An American in Paris | 0.02 | A Place in the Sun | 0.77 | 0.03 |
| 1943 | Casablanca | 0.03 | The Song Of Bernadette | 0.89 | 0.03 |
| 1956 | Around the World in 80 Days | 0.03 | The King and I | 0.82 | 0.04 |
| 1998 | Shakespeare in Love | 0.08 | Saving Private Ryan | 0.91 | 0.09 |
| 1992 | Unforgiven | 0.09 | The Crying Game | 0.74 | 0.12 |
| 1968 | Oliver! | 0.12 | The Lion in Winter | 0.78 | 0.15 |
| 1952 | The Greatest Show on Earth | 0.08 | The Quiet Man | 0.52 | 0.15 |
| 1977 | Annie Hall | 0.11 | The Turning Point | 0.61 | 0.18 |
| 1941 | How Green Was My Valley | 0.22 | Sergeant York | 0.44 | 0.50 |
| 1969 | Midnight Cowboy | 0.31 | Anne of the Thousand Days | 0.53 | 0.58 |
| 1945 | The Lost Weekend | 0.30 | The Bells Of St. Mary's | 0.46 | 0.65 |
| 2001 | A Beautiful Mind | 0.25 | The Fellowship of the Ring | 0.38 | 0.66 |
| 1995 | Braveheart | 0.30 | Apollo 13 | 0.41 | 0.73 |
| 1973 | The Sting | 0.44 | The Exorcist | 0.50 | 0.88 |
| 1938 | You Can't Take It With You | 0.32 | You Can't Take It With You | 0.32 | 1.00 |
| 1944 | Going My Way | 0.38 | Going My Way | 0.38 | 1.00 |
| 1985 | Out of Africa | 0.46 | Out of Africa | 0.46 | 1.00 |
| 1949 | All the King's Men | 0.49 | All the King's Men | 0.49 | 1.00 |
| 1950 | All About Eve | 0.50 | All About Eve | 0.50 | 1.00 |
| 1991 | The Silence of the Lambs | 0.56 | The Silence of the Lambs | 0.56 | 1.00 |

Best Director winners: ordered from most to least surprising

| Year | Actual Winner | Prob | Predicted Winner | Prob | Ratio |
|------|----------------------|------|----------------------|------|-------|
| 2000 | Steven Soderbergh | 0.01 | Ang Lee | 0.95 | 0.01 |
| 2002 | Roman Polanski | 0.02 | Rob Marshall | 0.92 | 0.02 |
| 1968 | Carol Reed | 0.03 | Anthony Harvey | 0.97 | 0.03 |
| 1972 | Bob Fosse | 0.03 | Francis Ford Coppola | 0.96 | 0.03 |
| 1949 | Joseph L. Mankiewicz | 0.02 | Robert Rossen | 0.48 | 0.04 |
| 1950 | Joseph L. Mankiewicz | 0.10 | Billy Wilder | 0.78 | 0.13 |
| 1948 | John Huston | 0.20 | Jean Negulesco | 0.57 | 0.35 |
| 1946 | William Wyler | 0.20 | Frank Capra | 0.56 | 0.36 |
| 1940 | John Ford | 0.17 | Alfred Hitchcock | 0.42 | 0.40 |
| 1939 | Victor Fleming | 0.22 | Frank Capra | 0.35 | 0.63 |
| 1943 | Michael Curtiz | 0.30 | Henry King | 0.41 | 0.73 |
| 1941 | John Ford | 0.28 | William Wyler | 0.38 | 0.74 |
| 1938 | Frank Capra | 0.26 | Frank Capra | 0.26 | 1.00 |
| 1945 | Billy Wilder | 0.33 | Billy Wilder | 0.33 | 1.00 |
| 1944 | Leo McCarey | 0.39 | Leo McCarey | 0.39 | 1.00 |
| 1985 | Sydney Pollack | 0.49 | Sydney Pollack | 0.49 | 1.00 |
| 1977 | Woody Allen | 0.53 | Woody Allen | 0.53 | 1.00 |
| 1947 | Elia Kazan | 0.56 | Elia Kazan | 0.56 | 1.00 |
| 1995 | Mel Gibson | 0.59 | Mel Gibson | 0.59 | 1.00 |
| 1942 | William Wyler | 0.72 | William Wyler | 0.72 | 1.00 |
| 1976 | John G. Avildsen | 0.87 | John G. Avildsen | 0.87 | 1.00 |
| 1975 | Milos Forman | 0.87 | Milos Forman | 0.87 | 1.00 |
| 1986 | Oliver Stone | 0.90 | Oliver Stone | 0.90 | 1.00 |
| 1980 | Robert Redford | 0.90 | Robert Redford | 0.90 | 1.00 |

Best Actor winners: ordered from most to least surprising

| Year | Actual Winner | Prob | Predicted Winner | Prob | Ratio |
|------|-------------------|------|------------------|------|-------|
| 2001 | Denzel Washington | 0.00 | Russell Crowe | 0.99 | 0.00 |
| 1968 | Cliff Robertson | 0.00 | Peter O'Toole | 0.89 | 0.00 |
| 1974 | Art Carney | 0.02 | Jack Nicholson | 0.86 | 0.02 |
| 1946 | Fredric March | 0.03 | Gregory Peck | 0.83 | 0.04 |
| 1944 | Bing Crosby | 0.03 | Alexander Knox | 0.78 | 0.04 |
| 1964 | Rex Harrison | 0.06 | Peter O'Toole | 0.83 | 0.07 |
| 1986 | Paul Newman | 0.08 | Bob Hoskins | 0.74 | 0.11 |
| 1956 | Yul Brynner | 0.07 | Kirk Douglas | 0.55 | 0.13 |
| 1989 | Daniel Day-Lewis | 0.08 | Tom Cruise | 0.61 | 0.13 |
| 1973 | Jack Lemmon | 0.08 | Al Pacino | 0.58 | 0.14 |
| 1991 | Anthony Hopkins | 0.10 | Nick Nolte | 0.64 | 0.16 |
| 1943 | Paul Lukas | 0.09 | Walter Pidgeon | 0.32 | 0.28 |
| 1965 | Lee Marvin | 0.17 | Oskar Werner | 0.53 | 0.32 |
| 1985 | William Hurt | 0.13 | Jack Nicholson | 0.40 | 0.33 |
| 1977 | Richard Dreyfuss | 0.26 | Richard Burton | 0.62 | 0.42 |
| 1942 | James Cagney | 0.23 | Ronald Colman | 0.52 | 0.44 |
| 1953 | William Holden | 0.17 | Montgomery Clift | 0.38 | 0.45 |
| 1940 | James Stewart | 0.27 | Laurence Olivier | 0.51 | 0.53 |
| 1938 | Spencer Tracy | 0.21 | Leslie Howard | 0.37 | 0.57 |
| 2002 | Adrien Brody | 0.32 | Daniel Day-Lewis | 0.49 | 0.65 |
| 1950 | Jose Ferrer | 0.32 | William Holden | 0.38 | 0.84 |
| 1951 | Humphrey Bogart | 0.25 | Fredric March | 0.29 | 0.86 |
| 1959 | Charlton Heston | 0.27 | Charlton Heston | 0.27 | 1.00 |
| 1969 | John Wayne | 0.33 | John Wayne | 0.33 | 1.00 |

Best Actress winners: ordered from most to least surprising

| Year | Actual Winner | Prob | Predicted Winner | Prob | Ratio |
|------|---------------------|------|------------------|------|-------|
| 2002 | Nicole Kidman | 0.08 | Renee Zellweger | 0.90 | 0.09 |
| 1967 | Katharine Hepburn | 0.05 | Faye Dunaway | 0.43 | 0.12 |
| 1966 | Elizabeth Taylor | 0.08 | Anouk Aimee | 0.68 | 0.12 |
| 1985 | Geraldine Page | 0.08 | Whoopi Goldberg | 0.65 | 0.12 |
| 1999 | Hilary Swank | 0.10 | Annette Bening | 0.79 | 0.13 |
| 1950 | Judy Holliday | 0.10 | Gloria Swanson | 0.73 | 0.14 |
| 1952 | Shirley Booth | 0.12 | Susan Hayward | 0.83 | 0.14 |
| 1961 | Sophia Loren | 0.09 | Geraldine Page | 0.60 | 0.15 |
| 1973 | Glenda Jackson | 0.09 | Ellen Burstyn | 0.58 | 0.16 |
| 1956 | Ingrid Bergman | 0.12 | Deborah Kerr | 0.72 | 0.17 |
| 1946 | Olivia de Havilland | 0.08 | Rosalind Russell | 0.48 | 0.17 |
| 1969 | Maggie Smith | 0.09 | Genevieve Bujold | 0.53 | 0.17 |
| 1974 | Ellen Burstyn | 0.07 | Gena Rowlands | 0.35 | 0.20 |
| 1945 | Joan Crawford | 0.16 | Ingrid Bergman | 0.73 | 0.22 |
| 1959 | Simone Signoret | 0.16 | Elizabeth Taylor | 0.68 | 0.24 |
| 1970 | Glenda Jackson | 0.15 | Ali MacGraw | 0.61 | 0.25 |
| 1981 | Katharine Hepburn | 0.08 | Meryl Streep | 0.31 | 0.26 |
| 1988 | Jodie Foster | 0.16 | Melanie Griffith | 0.53 | 0.30 |
| 1960 | Elizabeth Taylor | 0.09 | Shirley MacLaine | 0.29 | 0.31 |
| 1947 | Loretta Young | 0.13 | Rosalind Russell | 0.41 | 0.32 |
| 1951 | Vivien Leigh | 0.12 | Shelley Winters | 0.35 | 0.34 |
| 1962 | Anne Bancroft | 0.22 | Geraldine Page | 0.61 | 0.36 |
| 1965 | Julie Christie | 0.23 | Samantha Eggar | 0.60 | 0.38 |
| 1963 | Patricia Neal | 0.20 | Leslie Caron | 0.48 | 0.42 |