Winning Times in Marathon

- Sports records are often interesting because they reflect the evolution of human development.

- This data file contains the winning times, in seconds, for the Boston Marathon from 1897-2015.
Winning Times in Marathon

- Pro equals 1 since 1986 to reflect the payment of prize money
- Women have participated in this race since 1966, so the winning times for women are included from 1966-2015

Variables

m_sec (winning time for men in seconds, so two hours = 60*60*2 = 7200 seconds)

Note that the winning time in 1918 is not available because of WW I

w_sec (winning time for women in seconds, since 1966)

time (a time trend = -63 in 1897 and =55 in 2015)
Variables

**pro** (dummy variable = 1 since 1986, and 0 otherwise)

**Weather_bad** if commentary on history web site describes conditions as unfavorable (hot, headwind, or very cold and rainy, for example)

**Weather_good** if commentary on history web site describes conditions as favorable (e.g., strong tailwind)

**Entrants** number of runners in the race

---

Exponential Trend Model for Male Winning Times

I include an interaction term between time and PRO to reflect the fact that winning times quit falling after prize money attracted pro runners

Before prize money, winning time decreased by about 0.13% per year

After prize money, winning time is constant (sum of time and pro*time coefficients)
Exponential Trend Model for Male Winning Times

Prize money lowered the winning time by 6.4%

Bad weather is associated with winning times that are 1.6% higher

A one percent increase in the number of runners decreases the winning time by 1.7%

Residual Autocorrelations

AR(3)?

Partial ACF cuts off after lag 3
Exponential Trend Model for Male Winning Times, AR(3)

AR terms improve the model

Winning time declines by about .13% per year before prize money was offered

Winning times for pros are about 7.5% lower, bad weather slows races by about 1.6% and good weather speeds them up by about 2.3%

### Exponential Trend Model for Male Winning Times, AR(3)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>9.117910</td>
<td>0.054370</td>
<td>167.7026</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME</td>
<td>-0.001334</td>
<td>0.000717</td>
<td>-1.809090</td>
<td>0.0656</td>
</tr>
<tr>
<td>PRO</td>
<td>-0.074859</td>
<td>0.021760</td>
<td>-3.449136</td>
<td>0.0008</td>
</tr>
<tr>
<td>PRO*TIME</td>
<td>0.002040</td>
<td>0.000756</td>
<td>2.696284</td>
<td>0.0062</td>
</tr>
<tr>
<td>WEATHER_BAD</td>
<td>0.015599</td>
<td>0.008133</td>
<td>1.917999</td>
<td>0.0579</td>
</tr>
<tr>
<td>WEATHER_GOOD</td>
<td>-0.023067</td>
<td>0.010461</td>
<td>-2.205030</td>
<td>0.0297</td>
</tr>
<tr>
<td>LOG(ENTRANTS)</td>
<td>-0.012606</td>
<td>0.008146</td>
<td>-1.547609</td>
<td>0.1248</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.370461</td>
<td>0.131684</td>
<td>2.813261</td>
<td>0.0069</td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0.093263</td>
<td>0.131404</td>
<td>-0.709742</td>
<td>0.4756</td>
</tr>
<tr>
<td>AR(3)</td>
<td>0.292338</td>
<td>0.150545</td>
<td>1.941863</td>
<td>0.0549</td>
</tr>
</tbody>
</table>

R-squared: 0.839981
Adjusted R-squared: 0.825862
S.E. of regression: 0.032496
S.D. dependent var: 0.077878
Akaike info criterion: -3.930265
Schwarz criterion: -3.687483
Hannan-Quinn criterion: -3.831726
Durbin-Watson stat: 2.030343

Q-stat for 12 lags (9 df) has p-value of .583

Note that the residual variance seems to be getting smaller in the “modern” era
Forecast Number of Entrants

Note that to forecast winning time it is necessary to forecast how many runners will enter in future years.

Use time trend and dummy variables for Centennial race and the year after the bombing to account for unusual number of entrants.

Since prize money has been paid in 1986, this model works pretty well.

Number of entrants was much higher in 1996 (the centennial race) and about 6.6% higher last year after the bombing in the prior year.
Forecast Number of Entrants

Substitute forecasts of entrants into entrants series so it can be used to forecast winning times through 2050

Forecasts of Winning Times

Note that forecasted winning time is 2:08:08 in 2050
I include an interaction term between time and PRO to reflect the fact that winning times quit falling after prize money attracted pro runners.

Before prize money, winning time decreased by about 2.1% per year.

After prize money, winning time is constant (sum of time and pro*time coefficients).

Prize money lowered the winning time by 51%.

AR(2) helps clean up residual autocorrelations.

---

**Exponential Trend Model for Female Winning Times, AR(2)**

Dependent Variable: LOG(W_SEC)
Method: ARMA Conditional Least Squares (Marquardt - EViews legacy)
Sample (adjusted): 1968 2015
Included observations: 48 after adjustments
Convergence achieved after 6 iterations
White heteroskedastic-consistent standard errors & covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>9.623970</td>
<td>0.062532</td>
<td>153.9057</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME</td>
<td>-0.021290</td>
<td>0.001732</td>
<td>-12.29306</td>
<td>0.0000</td>
</tr>
<tr>
<td>PRO</td>
<td>-0.519555</td>
<td>0.024725</td>
<td>-20.64904</td>
<td>0.0000</td>
</tr>
<tr>
<td>PRO*TIME</td>
<td>0.002122</td>
<td>0.001219</td>
<td>18.10527</td>
<td>0.0000</td>
</tr>
<tr>
<td>WEATHER_BAD</td>
<td>0.002197</td>
<td>0.0006504</td>
<td>0.332662</td>
<td>0.7411</td>
</tr>
<tr>
<td>WEATHER_GOOD</td>
<td>-0.003728</td>
<td>0.016137</td>
<td>-0.231044</td>
<td>0.8165</td>
</tr>
<tr>
<td>LOG(ENTRANTS)</td>
<td>-0.007629</td>
<td>0.010989</td>
<td>-0.732455</td>
<td>0.4682</td>
</tr>
<tr>
<td>AR(2)</td>
<td>-0.440019</td>
<td>0.171336</td>
<td>-2.568159</td>
<td>0.0141</td>
</tr>
</tbody>
</table>

R-squared: 0.957857
Adjusted R-squared: 0.950481
S.E. of regression: 0.022776
Akaikes info criterion: -4.575301
Schwarz criterion: -4.263424
Hannan-Quinn criterion: -4.457446
Durbin-Watson stat: 1.775379

---

**Actual & Residual Plot**

Note that the trend disappears after 1986.
The log of the men’s winning time has a coefficient of about .53 (t-stat of 2.44)

Higher winning men’s times are associated with higher women’s times (elasticity of 0.53)

Could reflect different racing conditions, beyond weather comments

Forecasts for 2016-2050

Note that forecasted winning time is 2:28:36 in 2050
<table>
<thead>
<tr>
<th>Links</th>
</tr>
</thead>
</table>

**Eviews worksheet**
http://schwert.ssb.rochester.edu/a425/a425_boston.wf1

**Excel worksheet (zipped)**
http://schwert.ssb.rochester.edu/a425/a425_boston.zip

**APS 425 Home Page**
http://schwert.ssb.rochester.edu/a425/a425main.htm