APS 425
Fall 2015

Boston Marathon Data

Instructor: G. William Schwert
585-275-2470
schwert@schwert.ssb.rochester.edu

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 wining 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)

Dependent Variable: LOG(M_SEC)
Method: Least Squares
Sample: 1897 2015
Included observations: 118
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.145135	0.036044	253.7240	0.0000
TIME	-0.001337	0.000376	-3.555457	0.0006
PRO	-0.061154	0.011450	-5.341118	0.0000
PRO*TIME	0.002160	0.000316	6.841592	0.0000
WEATHER_BAD	0.015984	0.008562	1.866993	0.0645
WEATHER_GOOD	-0.021495	0.016522	-1.300995	0.1960
LOG(ENTRANTS)	-0.017247	0.005426	-3.178267	0.0019

R-squared	0.791867	Mean dependent var	9.047257
Adjusted R-squared	0.780617	S.D. dependent var	0.082195
S.E. of regression	0.038499	Akaike info criterion	-3.618894
Sum squared resid	0.164519	Schwarz criterion	-3.454531
Log likelihood	220.5148	Hannan-Quinn criter.	-3.552158
F-statistic	70.38568	Durbin-Watson stat	1.170851
Prob(F-statistic)	0.000000	Wald F-statistic	105.3442
Prob(Wald F-statistic)	0.000000		

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%

Dependent Variable: LOG(M_SEC)
Method: Least Squares
Sample: 1897 2015
Included observations: 118
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.145135	0.036044	253.7240	0.0000
TIME	-0.001337	0.000376	-3.555457	0.0006
PRO	-0.061154	0.011450	-5.341118	0.0000
PRO*TIME	0.002160	0.000316	6.841592	0.0000
WEATHER_BAD	0.015984	0.008562	1.866993	0.0645
WEATHER_GOOD	-0.021495	0.016522	-1.300995	0.1960
LOG(ENTRANTS)	-0.017247	0.005426	-3.178267	0.0019

R-squared	0.791867	Mean dependent var	9.047257
Adjusted R-squared	0.780617	S.D. dependent var	0.082195
S.E. of regression	0.038499	Akaike info criterion	-3.618894
Sum squared resid	0.164519	Schwarz criterion	-3.454531
Log likelihood	220.5148	Hannan-Quinn criter.	-3.552158
F-statistic	70.38568	Durbin-Watson stat	1.170851
Prob(F-statistic)	0.000000	Wald F-statistic	105.3442
Prob(Wald F-statistic)	0.000000		

Residual Autocorrelations

AR(3)?

Partial ACF cuts off after lag 3

Correlogram of Residuals

Sample: 1897 2015
Included observations: 118

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.392	0.392	18.641	0.000
		2	0.212	0.068	24.113	0.000
		3	0.331	0.269	37.585	0.000
		4	0.252	0.049	45.483	0.000
		5	0.118	-0.038	47.217	0.000
		6	0.138	0.028	49.613	0.000
		7	0.185	0.075	53.982	0.000
		8	-0.018	-0.170	54.022	0.000
		9	0.087	0.131	55.007	0.000
		10	0.039	-0.115	55.206	0.000
		11	0.010	0.056	55.220	0.000
		12	-0.011	-0.056	55.237	0.000

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%

Dependent Variable: LOG(M_SEC)
Method: ARMA Conditional Least Squares (Marquardt - EViews legacy)
Sample (adjusted): 1900 2015
Included observations: 112 after adjustments
Convergence achieved after 6 iterations
White heteroskedasticity-consistent standard errors & covariance

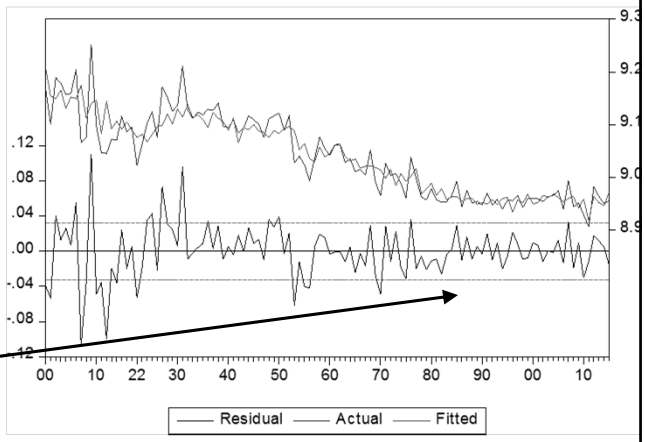
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.117910	0.054370	167.7026	0.0000
TIME	-0.001334	0.000717	-1.860900	0.0656
PRO	-0.074859	0.021760	-3.440136	0.0008
PRO*TIME	0.002040	0.000756	2.696284	0.0082
WEATHER_BAD	0.015599	0.008133	1.917899	0.0579
WEATHER_GOOD	-0.023067	0.010461	-2.205030	0.0297
LOG(ENTRANTS)	-0.012606	0.008146	-1.547609	0.1248
AR(1)	0.370461	0.131684	2.813261	0.0059
AR(2)	-0.093263	0.131404	-0.709742	0.4795
AR(3)	0.292338	0.150545	1.941863	0.0549

R-squared	0.839981	Mean dependent var	9.041494
Adjusted R-squared	0.825862	S.D. dependent var	0.077878
S.E. of regression	0.032498	Akaike info criterion	-3.930206
Sum squared resid	0.107727	Schwarz criterion	-3.687483
Log likelihood	230.0915	Hannan-Quinn criter.	-3.831726
F-statistic	59.49159	Durbin-Watson stat	2.030343
Prob(F-statistic)	0.000000		

Actual & Residual Plot

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

Dependent Variable: LOG(ENTRANTS)
Method: Least Squares
Sample (adjusted): 1898 2050
Included observations: 152 after adjustments

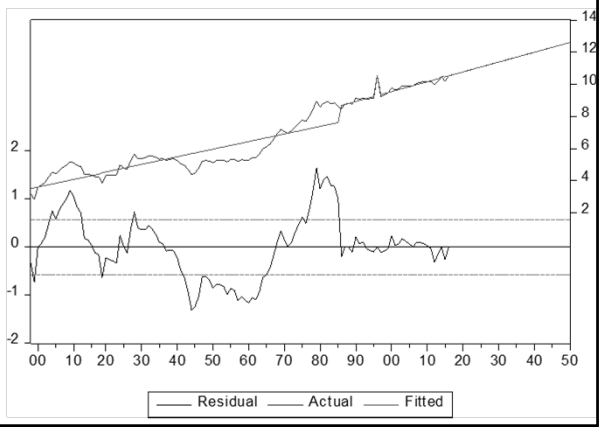
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.442485	0.075208	85.66246	0.0000
TIME	0.047261	0.002404	19.65800	0.0000
PRO	0.659802	0.245731	2.685058	0.0081
PRO*TIME	0.014058	0.004506	3.119573	0.0022
DUM_CENT	1.254019	0.580887	2.158799	0.0325
DUM_BOMB(-1)	0.070915	0.574818	0.123370	0.9020

R-squared	0.961243	Mean dependent var	7.761100
Adjusted R-squared	0.959916	S.D. dependent var	2.847342
S.E. of regression	0.570065	Akaike info criterion	1.752541
Sum squared resid	47.44623	Schwarz criterion	1.871905
Log likelihood	-127.1931	Hannan-Quinn criter.	1.801031
F-statistic	724.2203	Durbin-Watson stat	0.139386
Prob(F-statistic)	0.000000		

Forecast Number of Entrants

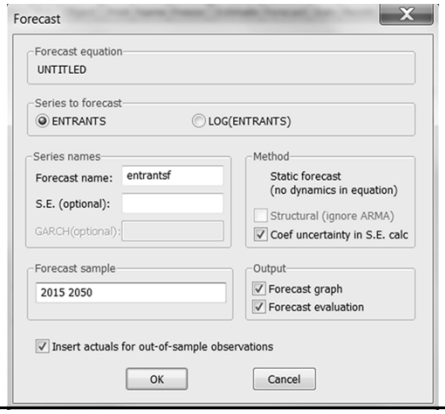
Since prize money has been paid in 1986, this model works pretty well

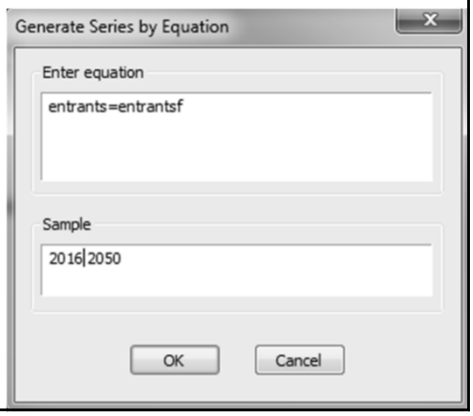
Number of entrants was much higher in 1996 (the centennial race) and about 6.6% higher last year after the bombing in the prior year



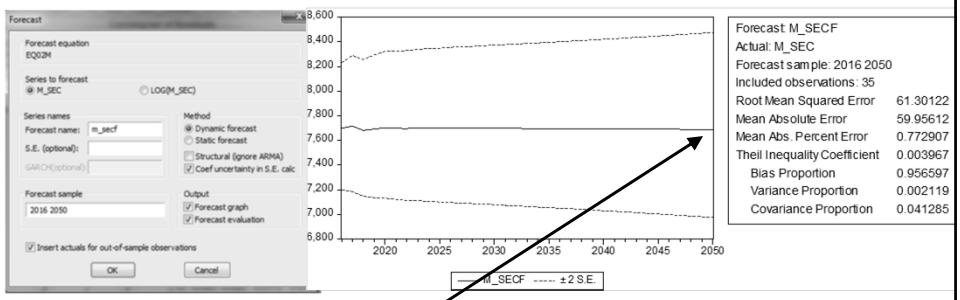
Forecast Number of Entrants

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

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

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

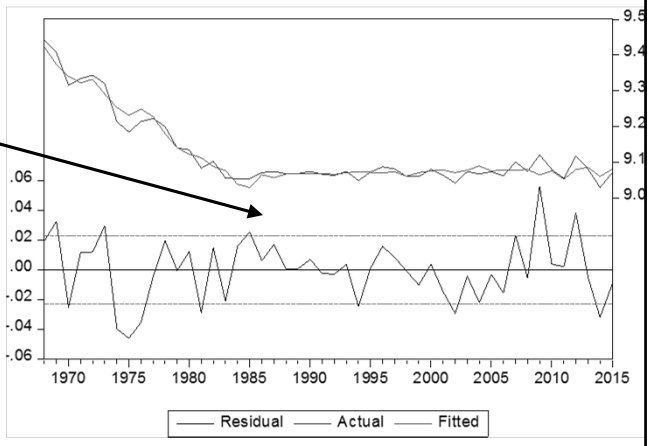
Dependent Variable: LOG(W_SEC)
Method: ARMA Conditional Least Squares (Marquardt - EViews legacy)
Sample (adjusted): 1968 2015
Included observations: 48 after adjustments
Convergence achieved after 6 iterations
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.623970	0.062532	153.9057	0.0000
TIME	-0.021290	0.001732	-12.29306	0.0000
PRO	-0.510555	0.024725	-20.64904	0.0000
PRO*TIME	0.022122	0.001219	18.15527	0.0000
WEATHER_BAD	0.002197	0.006604	0.332662	0.7411
WEATHER_GOOD	-0.003728	0.016137	-0.231044	0.8185
LOG(ENTRANTS)	-0.007829	0.010689	-0.732455	0.4682
AR(2)	-0.440019	0.171336	-2.568159	0.0141

R-squared	0.957857	Mean dependent var	9.124204
Adjusted R-squared	0.950481	S.D. dependent var	0.102346
S.E. of regression	0.022775	Akaike info criterion	-4.575301
Sum squared resid	0.020748	Schwarz criterion	-4.263434
Log likelihood	117.8072	Hannan-Quinn criter.	-4.457446
F-statistic	129.8769	Durbin-Watson stat	1.775379
Prob(F-statistic)	0.000000		

Actual & Residual Plot

Note that the trend disappears after 1986



Female Winning Times Related to Male Winning Times, AR(2)

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)

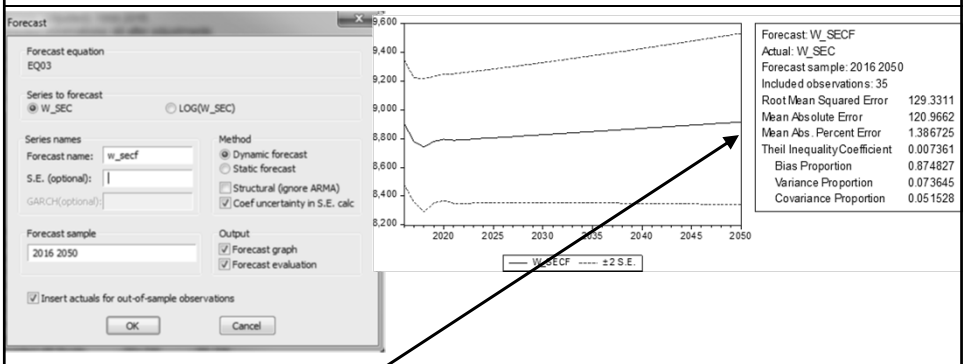
Could reflect different racing conditions, beyond weather comments

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 9 iterations
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.801485	1.952394	2.459281	0.0185
TIME	-0.021335	0.001468	-14.53543	0.0000
PRO	-0.489522	0.027394	-17.86977	0.0000
PRO*TIME	0.021489	0.001223	17.57302	0.0000
WEATHER_GOOD	0.014348	0.018238	0.786687	0.4362
WEATHER_BAD	-0.005144	0.006553	-0.784993	0.4372
LOG(ENTRANTS)	0.004554	0.008152	0.558644	0.5796
LOG(M_SEC)	0.526041	0.215528	2.440712	0.0193
AR(2)	-0.288408	0.204133	-1.412846	0.1656

R-squared	0.962915	Mean dependent var	9.124204
Adjusted R-squared	0.955308	S.D. dependent var	0.102346
S.E. of regression	0.021637	Akaike info criterion	-4.661503
Sum squared resid	0.018257	Schwarz criterion	-4.310653
Log likelihood	120.8761	Hannan-Quinn criter.	-4.528917
F-statistic	126.5800	Durbin-Watson stat	1.602227
Prob(F-statistic)	0.000000		

Forecasts for 2016-2050



Note that forecasted winning time is 2:28:36 in 2050

Links

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