

APS 425 Winter 2008

Boston Marathon Data

Instructor: G. William Schwert

585-275-2470

schwert@schwert.simon.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 1897-2007.

Winning Times in Marathon

Women have participated in this race since 1966, so the winning times for women are included from 1966-2001.

- Finally, as this race has become more famous and the prize for winning has become lucrative, world-class runners have begun to compete and win, so I also include variable indicating whether the winning racers were from the U.S., or not.

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 = -53 in 1897 and =57 in 2007)

Variables

usm (dummy variable = 1 when the male winner is from the US, and 0 otherwise)

Note that I have set this = 0 from 2008-2050 on the assumption that foreign runners will continue to win in the future)

usw (dummy variable = 1 when the female winner is from the US, and 0 otherwise)

Note that I have set this = 0 from 2008-2050 on the assumption that foreign runners will continue to win in the future)

Exponential Trend Model for Male Winning Times

This is an exponential trend model with the winning time declining by about .21% per year

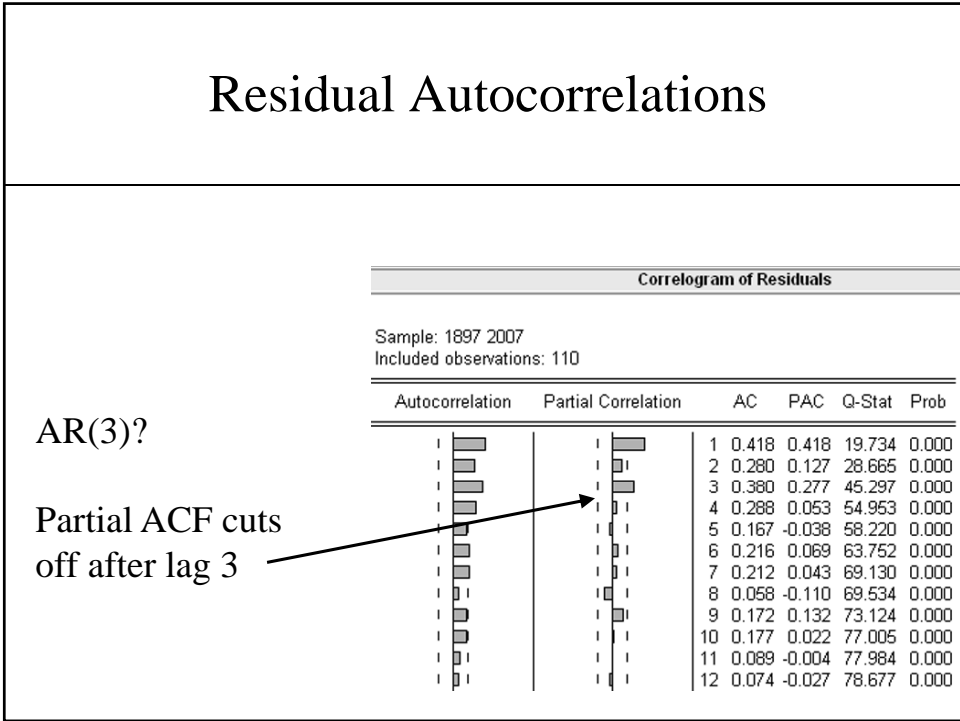
Winning times for US men are about .04% higher, but this difference is not significant

Dependent Variable: LOG(M_SEC)
 Method: Least Squares
 Sample (adjusted): 1897 2007
 Included observations: 110 after adjustments
 White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.059217	0.004986	1816.988	0.0000
TIME	-0.002137	0.000142	-15.09109	0.0000
USM	0.000422	0.009488	0.044465	0.9646

R-squared	0.734916	Mean dependent var	9.054447
Adjusted R-squared	0.729961	S.D. dependent var	0.080308
S.E. of regression	0.041732	Akaike info criterion	-3.488183
Sum squared resid	0.186351	Schwarz criterion	-3.414533
Log likelihood	194.8500	F-statistic	148.3225
Durbin-Watson stat	1.106478	Prob(F-statistic)	0.000000

Residual Autocorrelations



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

AR terms improve the model

Winning time declines by about .18% per year

Winning times for US men are about 1.6% higher, and significant

Dependent Variable: LOG(M_SEC)
 Method: Least Squares
 Sample (adjusted): 1900 2007
 Included observations: 104 after adjustments
 Convergence achieved after 7 iterations
 White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.050806	0.011853	763.5710	0.0000
TIME	-0.001799	0.000370	-4.863226	0.0000
USM	0.015752	0.007508	2.098054	0.0385
AR(1)	0.409585	0.128514	3.187075	0.0019
AR(2)	-0.069022	0.131863	-0.523438	0.6019
AR(3)	0.305691	0.141866	2.154780	0.0336

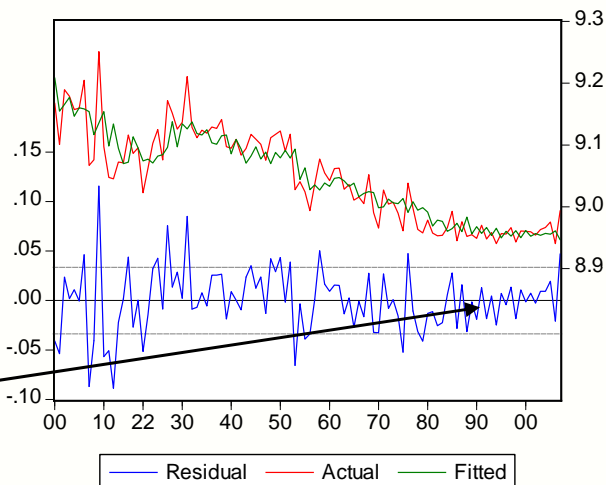
R-squared	0.813668	Mean dependent var	9.048655
Adjusted R-squared	0.804161	S.D. dependent var	0.076001
S.E. of regression	0.033633	Akaike info criterion	-3.890629
Sum squared resid	0.110859	Schwarz criterion	-3.738068
Log likelihood	208.3127	F-statistic	85.58841
Durbin-Watson stat	1.996809	Prob(F-statistic)	0.000000

Inverted AR Roots	.80	-.20+.59i	-.20-.59i
-------------------	-----	-----------	-----------

Actual & Residual Plot

Q-stat for 12 lags (9 df) has p-value of .279

Note that the residual variance seems to be getting smaller in the “modern” era



Exponential Trend Model for Female Winning Times, AR(1)

AR term improves the model

Winning time declines by about .22% per year (insignificant)

Winning times for US women are about .08% higher, not significant

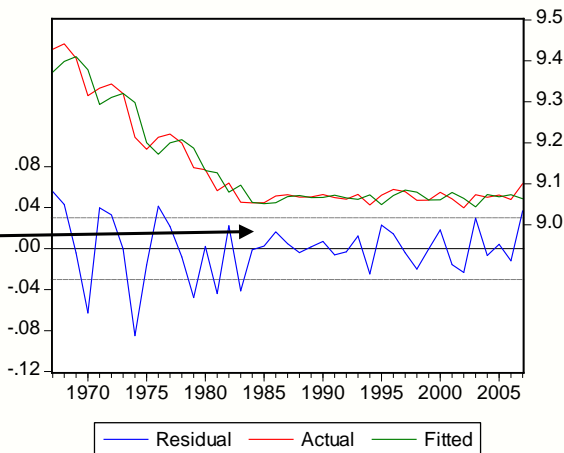
Dependent Variable: LOG(W_SEC)
 Method: Least Squares
 Sample (adjusted): 1967 2007
 Included observations: 41 after adjustments
 Convergence achieved after 5 iterations
 White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.932555	0.623496	14.32656	0.0000
USW	-0.000789	0.010371	-0.076068	0.9398
TIME	0.002211	0.010111	0.218693	0.8281
AR(1)	0.929716	0.074975	12.40033	0.0000

R-squared	0.938848	Mean dependent var	9.140603
Adjusted R-squared	0.933889	S.D. dependent var	0.117282
S.E. of regression	0.030156	Akaike info criterion	-4.072430
Sum squared resid	0.033646	Schwarz criterion	-3.905252
Log likelihood	87.48481	F-statistic	189.3490
Durbin-Watson stat	1.910012	Prob(F-statistic)	0.000000

Actual & Residual Plot

Note that the residual variance seems to be getting smaller since 1985



Female Winning Times Related to Male Winning Times, AR(1)

The log of the men's winning time has a coefficient of about .5 (t-stat of almost 4)

Higher winning men's times are associated with higher women's times (elasticity of 1/2)

Could reflect different racing conditions

Dependent Variable: LOG(W_SEC)
 Method: Least Squares
 Sample (adjusted): 1967 2007
 Included observations: 41 after adjustments
 Convergence achieved after 7 iterations
 White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.450348	1.296194	3.433396	0.0015
USW	-0.011431	0.009050	-1.263104	0.2147
TIME	0.004316	0.007982	0.164929	0.8699
LOG(M_SEC)	0.504714	0.128574	3.925472	0.0004
AR(1)	0.931159	0.070735	13.16399	0.0000

R-squared	0.951852	Mean dependent var	9.140603
Adjusted R-squared	0.946502	S.D. dependent var	0.117282
S.E. of regression	0.027127	Akaike info criterion	-4.262730
Sum squared resid	0.026491	Schwarz criterion	-4.053758
Log likelihood	92.38597	F-statistic	177.9225
Durbin-Watson stat	1.830500	Prob(F-statistic)	0.000000

Links

Eviews worksheet

http://schwert.simon.rochester.edu/a425/a425_boston.wf1

APS 425 Home Page

<http://schwert.simon.rochester.edu/a425/a425main.htm>