APS 425
Winter 2010

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 1897-2009.

Winning Times in Marathon

Women have participated in this race since 1966, so the winning times for women are included from 1966-2009.

- 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 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 = -53 in 1897 and =59 in 2009)

Variables

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

Note that I have set this = 0 from 2010-2050 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 2010-2050 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 2009
Included observations: 112 after adjustments
White Heteroskedasticity-Consistent Standard Errors & Covariance

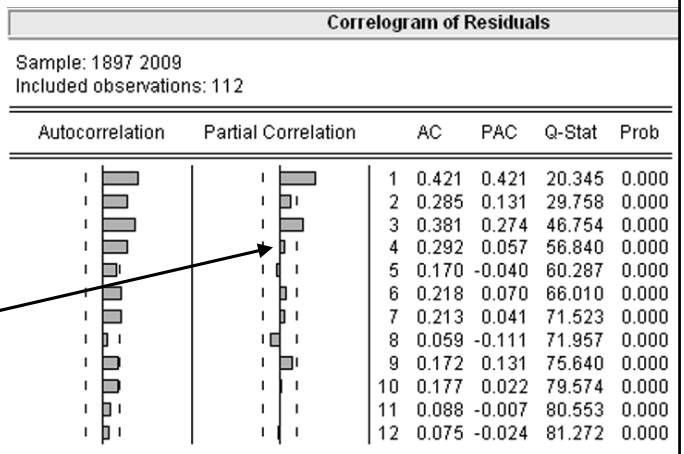
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.059414	0.004989	1815.899	0.0000
TIME	-0.002124	0.000137	-15.46872	0.0000
USM	0.000444	0.009487	0.046846	0.9627

R-squared	0.742468	Mean dependent var	9.052549
Adjusted R-squared	0.737743	S.D. dependent var	0.080829
S.E. of regression	0.041393	Akaike info criterion	-3.504984
Sum squared resid	0.186760	Schwarz criterion	-3.432167
Log likelihood	199.2791	Hannan-Quinn criter.	-3.475440
F-statistic	157.1245	Durbin-Watson stat	1.116137
Prob(F-statistic)	0.000000		

Residual Autocorrelations

AR(3)?

Partial ACF cuts off after lag 3



Correlogram of Residuals

Sample: 1897 2009
Included observations: 112

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.421	0.421	20.345	0.000
		2	0.285	0.131	29.758	0.000
		3	0.381	0.274	46.754	0.000
		4	0.292	0.057	56.840	0.000
		5	0.170	-0.040	60.287	0.000
		6	0.218	0.070	66.010	0.000
		7	0.213	0.041	71.523	0.000
		8	0.059	-0.111	71.957	0.000
		9	0.172	0.131	75.640	0.000
		10	0.177	0.022	79.574	0.000
		11	0.088	-0.007	80.553	0.000
		12	0.075	-0.024	81.272	0.000

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 2009
Included observations: 106 after adjustments
Convergence achieved after 7 iterations
White Heteroskedasticity-Consistent Standard Errors & Covariance

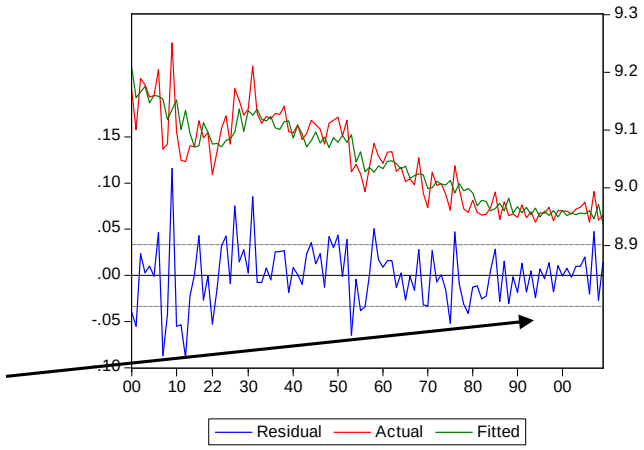
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.050752	0.011698	773.7111	0.0000
TIME	-0.001824	0.000347	-5.258689	0.0000
USM	0.015520	0.007511	2.066382	0.0414
AR(1)	0.396680	0.126573	3.133991	0.0023
AR(2)	-0.054293	0.129558	-0.419064	0.6761
AR(3)	0.297394	0.141447	2.102509	0.0380

R-squared	0.818120	Mean dependent var	9.046760
Adjusted R-squared	0.809026	S.D. dependent var	0.076518
S.E. of regression	0.033439	Akaike info criterion	-3.903251
Sum squared resid	0.111816	Schwarz criterion	-3.752490
Log likelihood	212.8723	Hannan-Quinn criter.	-3.842147
F-statistic	89.96288	Durbin-Watson stat	2.021301
Prob(F-statistic)	0.000000		

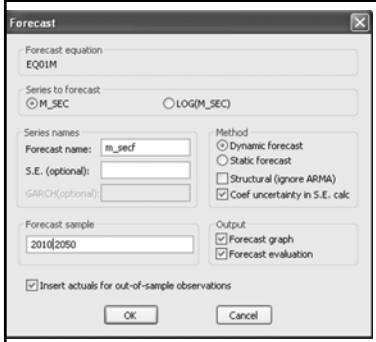
Actual & Residual Plot

Q-stat for 12 lags (9 df) has p-value of .254

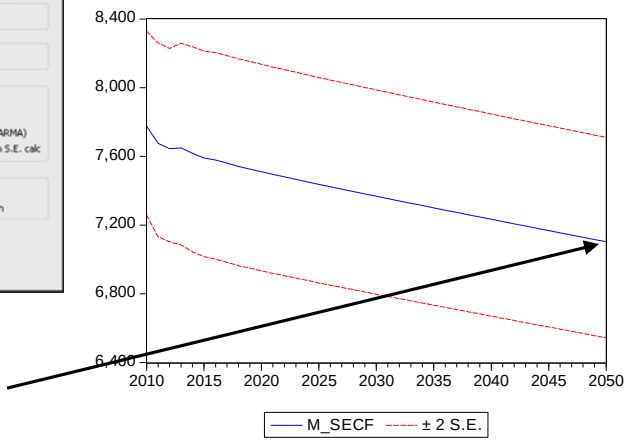
Note that the residual variance seems to be getting smaller in the "modern" era



Forecasts for 2010-2050



Note that forecasted winning time is 1:58:23 in 2050



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

AR term improves the model

Winning time increase by about .42% per year (insignificant)

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

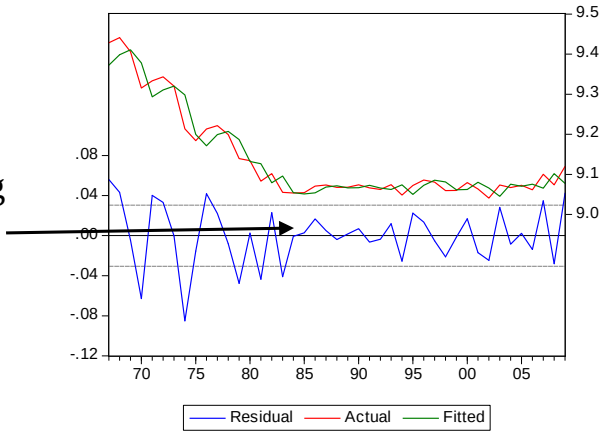
Dependent Variable: LOG(W_SEC)
Method: Least Squares
Sample (adjusted): 1967 2009
Included observations: 43 after adjustments
Convergence achieved after 5 iterations
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.824850	0.821447	10.74306	0.0000
USW	-0.000809	0.010311	-0.078475	0.9379
TIME	0.004158	0.012781	0.325326	0.7467
AR(1)	0.935746	0.073338	12.75935	0.0000

R-squared	0.934652	Mean dependent var	9.138574
Adjusted R-squared	0.929625	S.D. dependent var	0.114942
S.E. of regression	0.030492	Akaike info criterion	-4.054281
Sum squared resid	0.036261	Schwarz criterion	-3.890448
Log likelihood	91.16704	Hannan-Quinn criter.	-3.993865
F-statistic	185.9345	Durbin-Watson stat	2.028231
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 over 4)

Higher winning men's times are associated with higher women's times (elasticity of ½)

Could reflect different racing conditions

MEN

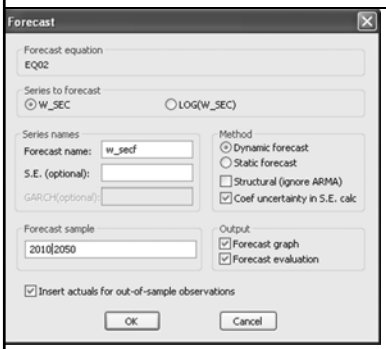
Modified: 1897 2050 // men=m_sec
Modified: 2010 2050 // men = m_secf

Dependent Variable: LOG(W_SEC)
Method: Least Squares
Sample (adjusted): 1967 2009
Included observations: 43 after adjustments
Convergence achieved after 8 iterations
White Heteroskedasticity-Consistent Standard Errors & Covariance

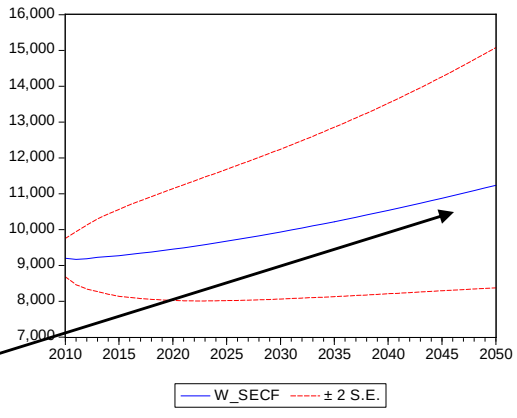
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.929441	1.966726	1.997961	0.0529
USW	-0.011758	0.008632	-1.362125	0.1812
TIME	0.008266	0.020834	0.396760	0.6938
LOG(MEN)	0.513815	0.117911	4.357641	0.0001
AR(1)	0.952886	0.069396	13.73116	0.0000

R-squared	0.949345	Mean dependent var	9.138574
Adjusted R-squared	0.944013	S.D. dependent var	0.114942
S.E. of regression	0.027197	Akaike info criterion	-4.262468
Sum squared resid	0.028108	Schwarz criterion	-4.057677
Log likelihood	96.64306	Hannan-Quinn criter.	-4.186948
F-statistic	178.0447	Durbin-Watson stat	1.835654
Prob(F-statistic)	0.000000		

Forecasts for 2010-2050



Note that forecasted winning time is 3:07:07 in 2050



Links

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

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