Marathon Record Progression

Raw Data

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 ... enter page number 27 > >>

Rank	Mark	Competitor	DOB	Nat	Pos	Venue	Date	Results Score
1	2:01:25	Kelvin KIPTUM	**02 DEC 1999**	KEN	1	London (GBR)	23 APR 2023	1307
2	2:03:47	Bashir ABDI	**10 FEB 1989**	BEL	1	Rotterdam (NED)	16 APR 2023	1266
3	2:03:50	Timothy KIPLAGAT		KEN	2	Rotterdam (NED)	16 APR 2023	1265
4	2:04:09	Bernard Kiprop KOECH	**31 JAN 1988**	KEN	1	Hamburg (GER)	23 APR 2023	1259
5	2:04:23	Geoffrey KAMWOROR	**22 NOV 1992**	KEN	2	London (GBR)	23 APR 2023	1255
6	2:04:33	Joshua BELET	**10 FEB 1998**	KEN	2	Hamburg (GER)	23 APR 2023	1252
7	2:04:59	Gadisa SHUMIE	**15 SEP 1992**	ETH	1	Sevilla (ESP)	19 FEB 2023	1245
7	2:04:59	Tamirat TOLA	**11 AUG 1991**	ETH	3	London (GBR)	23 APR 2023	1245
9	2:05:06	Marius KIMUTAI	**10 DEC 1992**	BRN	1	Barcelona (ESP)	19 MAR 2023	1243
10	2:05:08	Samwel Nyamai MAILU	**07 FEB 1993**	KEN	1	Wien (AUT)	23 APR 2023	1242

Data Retrieved from World Athletics Website

General Background Information

- Course > = 42.195 km (26.219 mi).
- WR: Eliud Kipchoge, 2:01:09 = 7269s

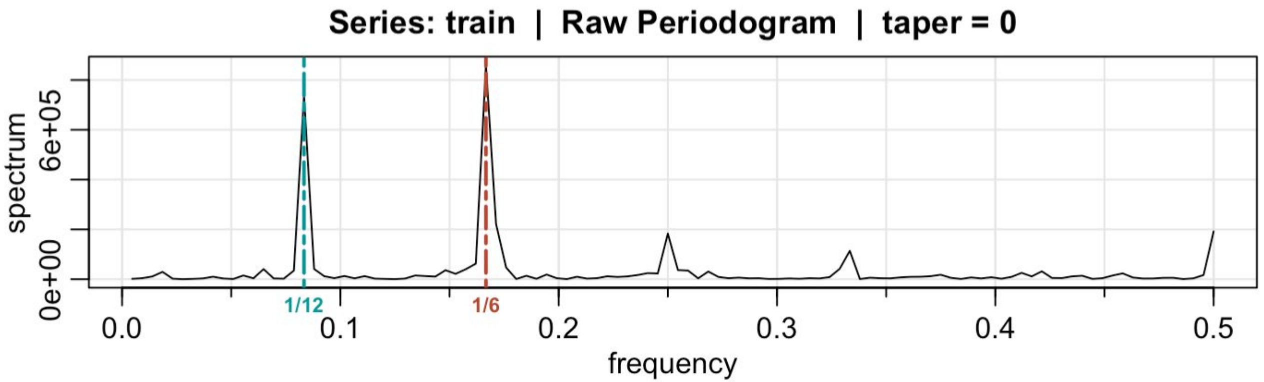
- We can see that due to hot weather, August 2007 and July 2021 has no official race.
- And due to covid, April to September 2020 has no official race

Time: January 2001 - May 2023, with the above exceptions **imputed**

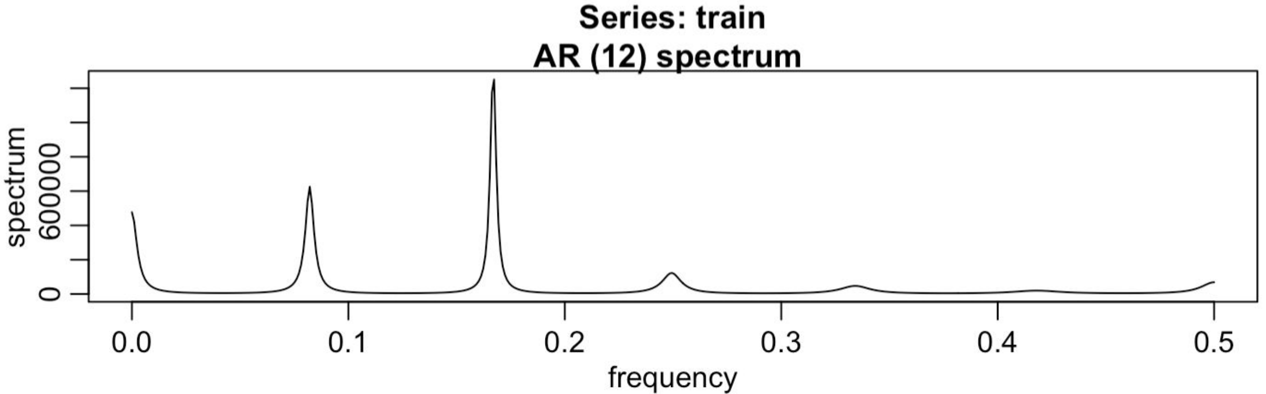
Reindex January 2001 as time 1, ..., and May 2023 as 266

Initial Spectral Analysis

- Periodogram with narrow-band peak at around **1/12** and **1/6**

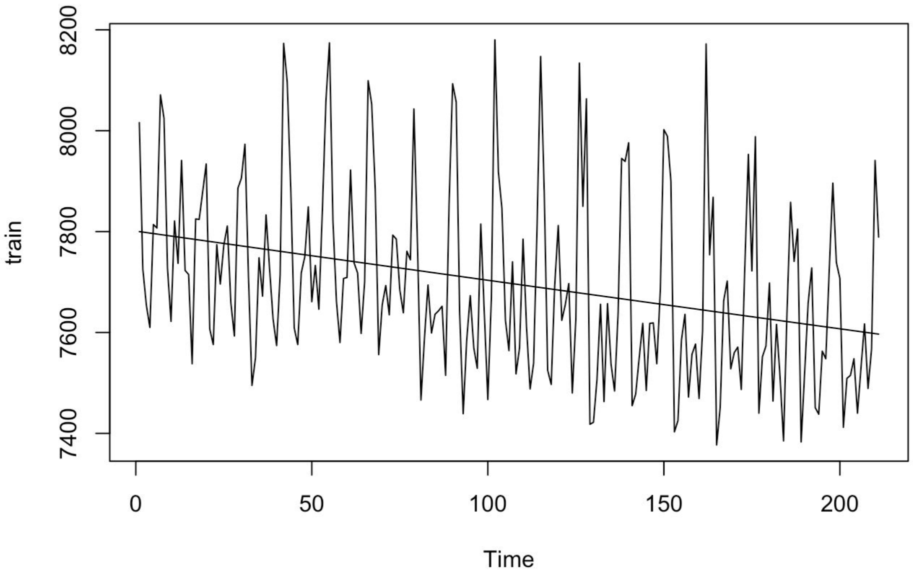


- Parametric AR(12) spectrum captures the peak and the harmonic

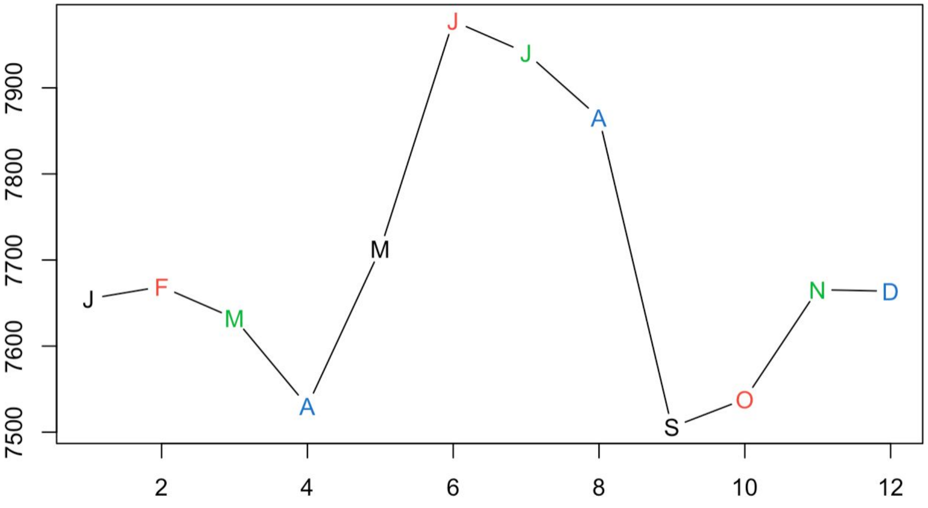


Trend & Cycle Removal

$$y = e^{8.962 - 0.0013 * t}$$

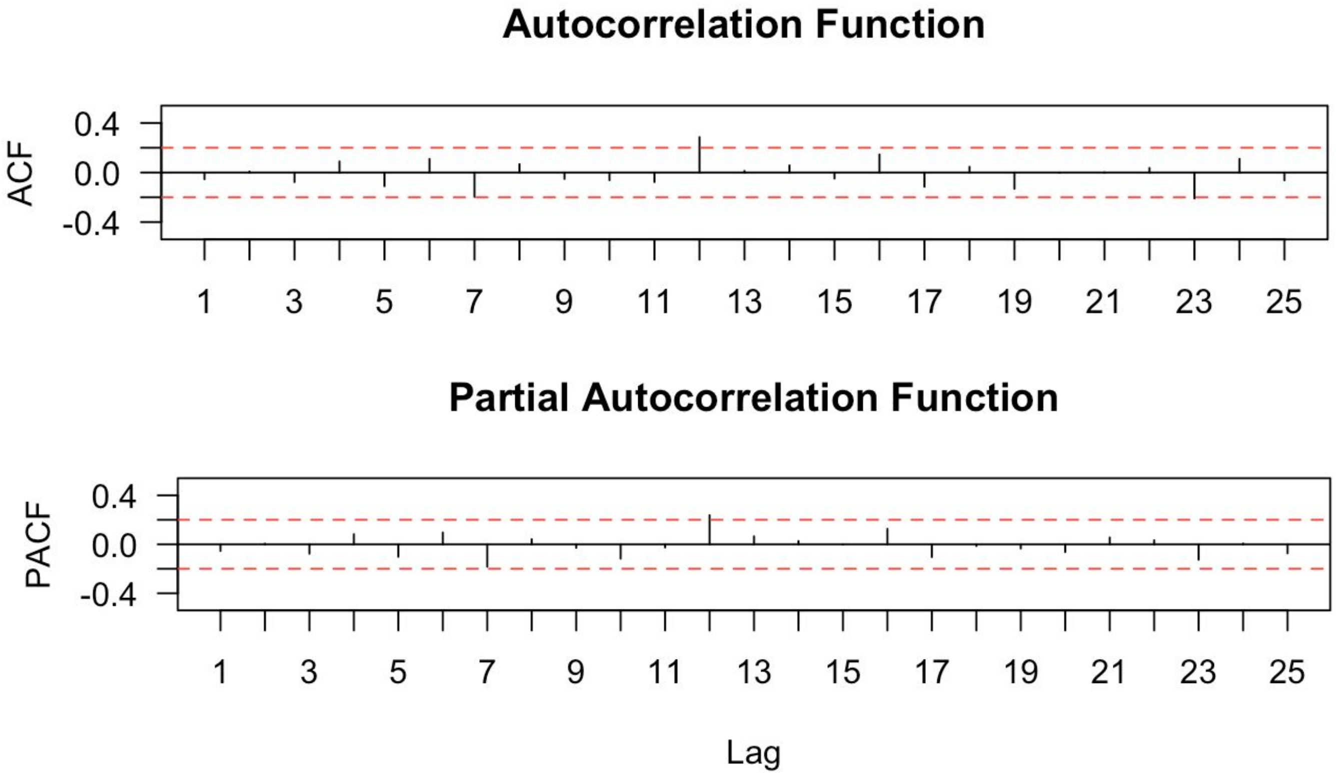


Seasonal Plot (Mean)



Residual Analysis

- Both ACF & PACF has shrinked after detrending
- Both ACF & PACF tails off.
  - Indication of ARMA model



Model Fitting

Grid Search w/ RMSE as metric → ARMA(7,6)

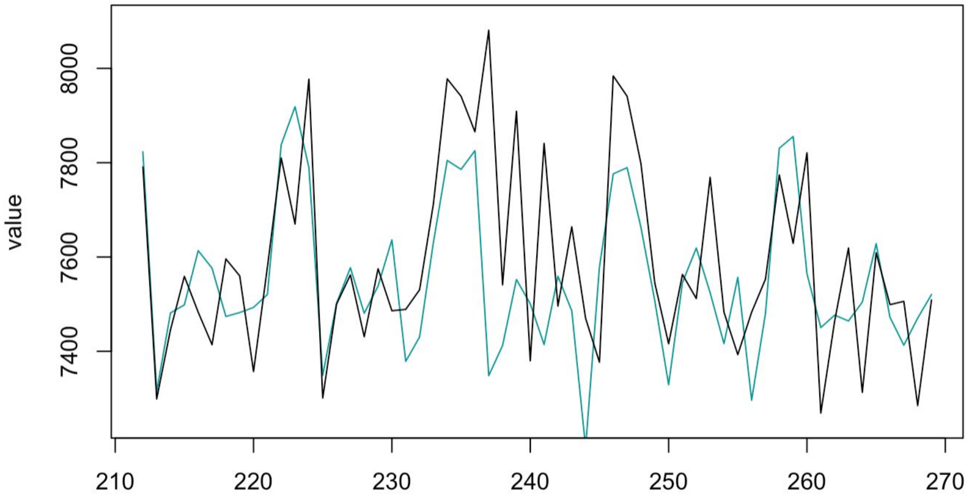
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	92.39901	95.68758	95.65219	94.36388	93.31652	91.05821	88.34669	87.61045	87.25847	86.04724
[2,]	90.64180	100.88067	87.42562	92.22620	86.91695	87.41688	101.47109	85.78565	85.88436	85.44019
[3,]	90.58195	115.88636	179.37185	107.17336	102.26921	87.35414	98.31313	89.43111	87.02495	91.38327
[4,]	90.37507	89.36075	93.04617	135.90803	106.04821	115.99287	89.21444	99.51944	96.33291	99.15321
[5,]	89.74203	86.79325	86.48071	86.79697	87.68128	87.71589	129.02656	125.04273	96.53153	98.33598
[6,]	89.93517	87.34736	86.92376	90.50506	87.44833	131.39315	104.47757	126.31991	93.88085	84.90459
[7,]	88.77410	94.66051	88.10991	147.18374	132.40931	146.17898	139.42975	89.71582	93.76902	137.54698
[8,]	85.53173	85.43159	86.67199	137.22468	162.73005	83.48060	83.08321	96.98145	100.19755	109.47116
[9,]	85.14055	85.23019	84.84785	86.45289	103.20927	85.04579	105.24499	91.52096	135.45307	87.24750
[10,]	84.78767	84.63329	120.04583	84.23023	134.66566	83.38553	85.75732	105.06556	105.19466	173.28177

Grid Search w/ BIC as metric → ARMA(1,1)

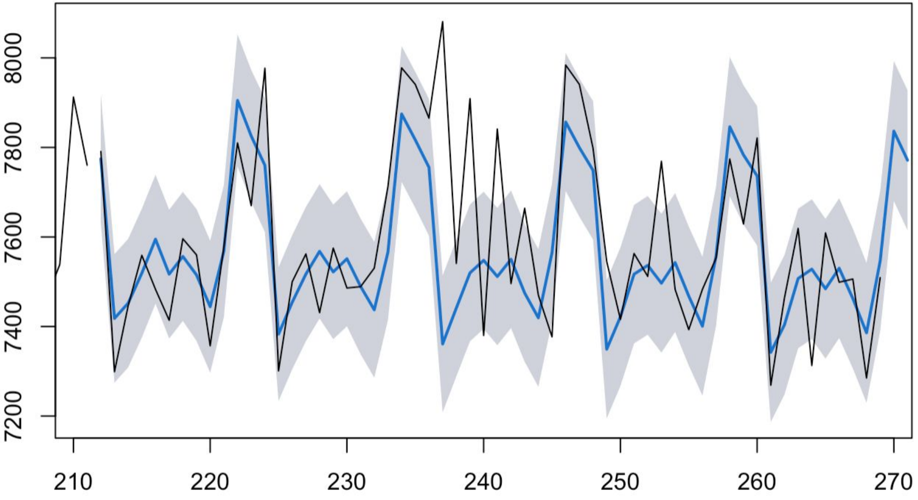
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	2519.517	2524.238	2529.590	2534.047	2538.581	2540.968	2544.472	2543.126	2548.449	2553.570
[2,]	2524.234	2516.400	2521.115	2523.900	2529.100	2534.445	2539.602	2548.456	2553.566	2558.316
[3,]	2529.580	2531.561	2531.782	2536.116	2541.095	2539.787	2544.961	2545.299	2550.430	2555.662
[4,]	2533.663	2524.230	2529.066	2533.153	2533.162	2538.183	2544.886	2550.413	2548.035	2560.268
[5,]	2537.504	2529.193	2534.407	2539.711	2541.870	2547.143	2542.597	2547.253	2558.058	2554.970
[6,]	2540.601	2534.320	2539.406	2541.744	2547.077	2545.167	2542.738	2546.628	2549.045	2557.589
[7,]	2543.827	2539.663	2544.758	2542.359	2541.381	2543.654	2545.392	2549.936	2554.396	2556.844
[8,]	2541.606	2546.578	2542.473	2543.822	2546.521	2547.779	2552.931	2547.675	2553.735	2555.191
[9,]	2546.589	2552.093	2557.281	2550.356	2550.785	2552.761	2557.423	2560.070	2555.718	2561.898
[10,]	2551.764	2556.576	2546.473	2557.151	2552.624	2557.805	2562.953	2556.908	2562.161	2565.597

Prediction

Forecast from ARIMA(7,0,6) with 1-step look ahead



Forecasts from ARIMA(7,0,6) with 60-period look ahead



RMSE: 177.4566

RMSE: 84.28

Conclusion

- ARIMA(7,0,6) seems to be a good fit of our training data

- However, for a different dataset (even using a fraction of our training) may lead to different model result.

- Future Work
  1. Deep dive into covariates (weather,)
  2. Dealing with missing data is key in this model
  3. Apply on other race events

Work Cited

- Data Source: Marathon - men - senior - outdoor - 2023
- Missing Value Handling Techniques: 4 Techniques to Handle Missing values in Time Series Data | by Satyam Kumar
- Forecast Package in R: forecast function - RDocumentation
- Shumway, Robert H., and David S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, 2017.