Periodic Autoregressive Models for New Zealand Hydro Catchment Inflows: An Evaluation of Their Ability to Forecast the Risk of Persistent Low Inflows

David Harte & Peter Thomson
Statistics Research Associates
Wellington
New Zealand

email: david@statsresearch.co.nz

Abstract

We wish to model weekly inflows to a number of NZ lakes used for hydro electricity generation. We are particularly interested in the extent to which we can generate forward realisations over time scales of up to 2–3 years. Of particular interest are extremely low inflows that can occur during the summer months, typically, about once in 5–10 years. A primary objective is to estimate persistently low inflows, which could cause considerable risk to stable electricity generation.

Three models were fitted to the data. The base model is a periodic autoregression model (PAR). The two other models are semi-parametric variations on this model. The standard PAR model has strictly periodic stochastic properties that do not account for dynamically changing seasonal patterns. The two variants attempt to incorporate dynamic seasonality and longer term variability.

The first variant involves block bootstrapping the innovations from the fit of the PAR model to the historical series. This includes building into the simulated sequence, in a non-parametric way, structure that occurs in the historical series that cannot be accounted for by the PAR model.

The second variant involves extracting evolving trend and seasonal components using conventional smoothing windows. With these subtracted, a standard PAR model is fitted to the residuals. A simulated sequence is generated by simulating a pure PAR process, then adding back the evolving seasonal and trend components by block bootstrapping them from those estimated from the historical data.

We determine which model fits “best” by comparing the characteristics of the simulated sequences from the three models to those of the observed time series. As well as the usual methods of evaluating model goodness of fit, we have derived

others that are more special to the problem concerned. For example, if we simulate a long sequence of data from each of the three models, what is the probability distribution of having a run length (weeks) of inflows below a given threshold, given that the run starts in a given week? How does this distribution compare to that using the historical data?

Residual structure after fitting all three models indicates that the inflow series probably contain episodic or abrupt changes in level. These could be caused by changes between seasons, where the times of these changes can be somewhat random from year to year. A possible modification to the above models would be to include a Markov switching component, where the change points between the Markov states (seasons) is hidden.

This study was undertaken as a benchmarking exercise for the New Zealand Electricity Commission. In the light of our findings, other hopefully more appropriate risk forecasting models are proposed.

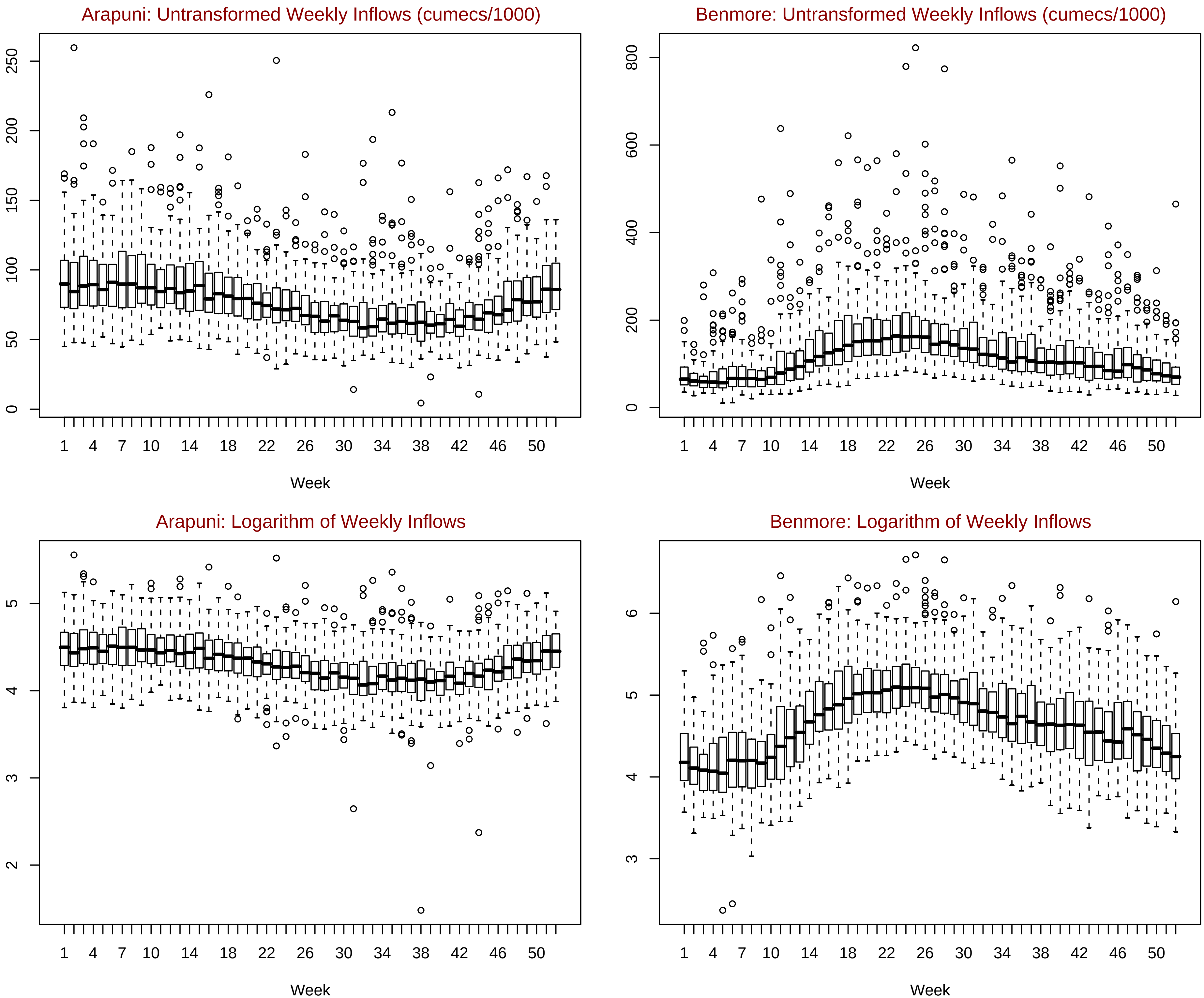
Introduction

NZ lakes used for hydro electricity generation

Problem: Persistently low inflows in some years (extremes) during summer – insufficient generation potential for winter

Weekly time series $\{X_t\}$ of inflows, from beginning 1931 to end 2004

9 lakes, different parts of country, different seasonality



Arapuni: Untransformed Weekly Inflows (cumecs/1000)
Benmore: Untransformed Weekly Inflows (cumecs/1000)
Arapuni: Logarithm of Weekly Inflows
Benmore: Logarithm of Weekly Inflows
Week
0
50
100
150
200
250
200
400
600
800
2
3
4
5
6
1
4
7
10
14
18
22
26
30
34
38
42
46
50

Data Transformation

Fit log-normal distribution

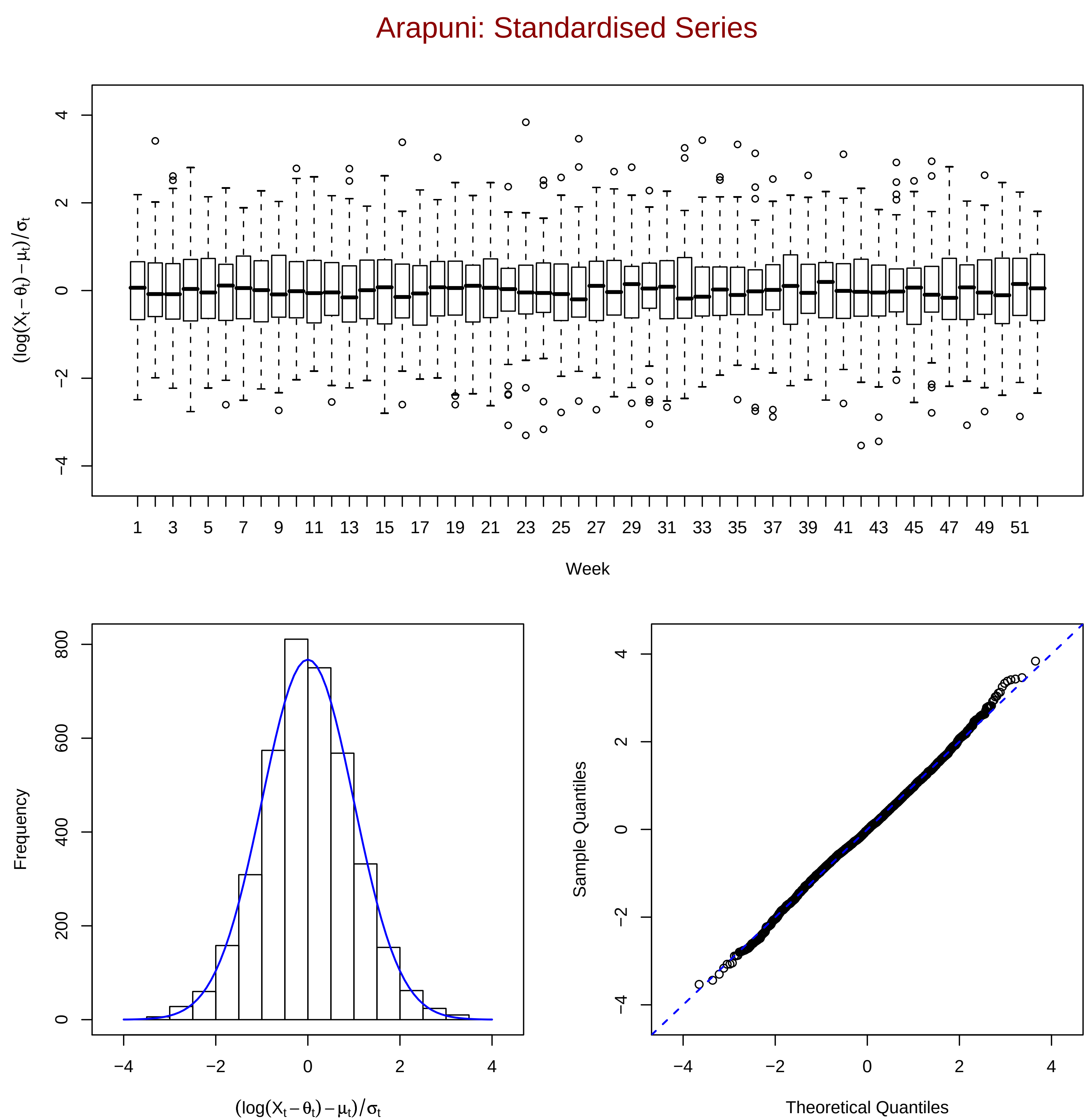
$$Z_t = \frac{\log(X_t - \theta_t) - \mu_t}{\sigma_t}$$

where

$$\begin{aligned}
\theta_t &= \theta_{t+52} \\
\mu_t &= \mu_{t+52} \\
\sigma_t &= \sigma_{t+52}
\end{aligned}$$

θ_t does not form part of the seasonal dynamics

Estimate θ_t using concentrated likelihood, 13 week moving window



Arapuni: Standardised Series
$(\log(X_t-\theta_t)-\mu_t)/\sigma_t$
Week
Frequency
$(\log(X_t-\theta_t)-\mu_t)/\sigma_t$
Sample Quantiles
Theoretical Quantiles

Decomposition of Shifted-Log Series

$$\text{shifted-log series (black line)} = \log(X_t - \theta_t)$$

$$\text{strictly periodic weekly series (red line)} = \mu_t$$

Trend T_t estimated with moving 3 yr window

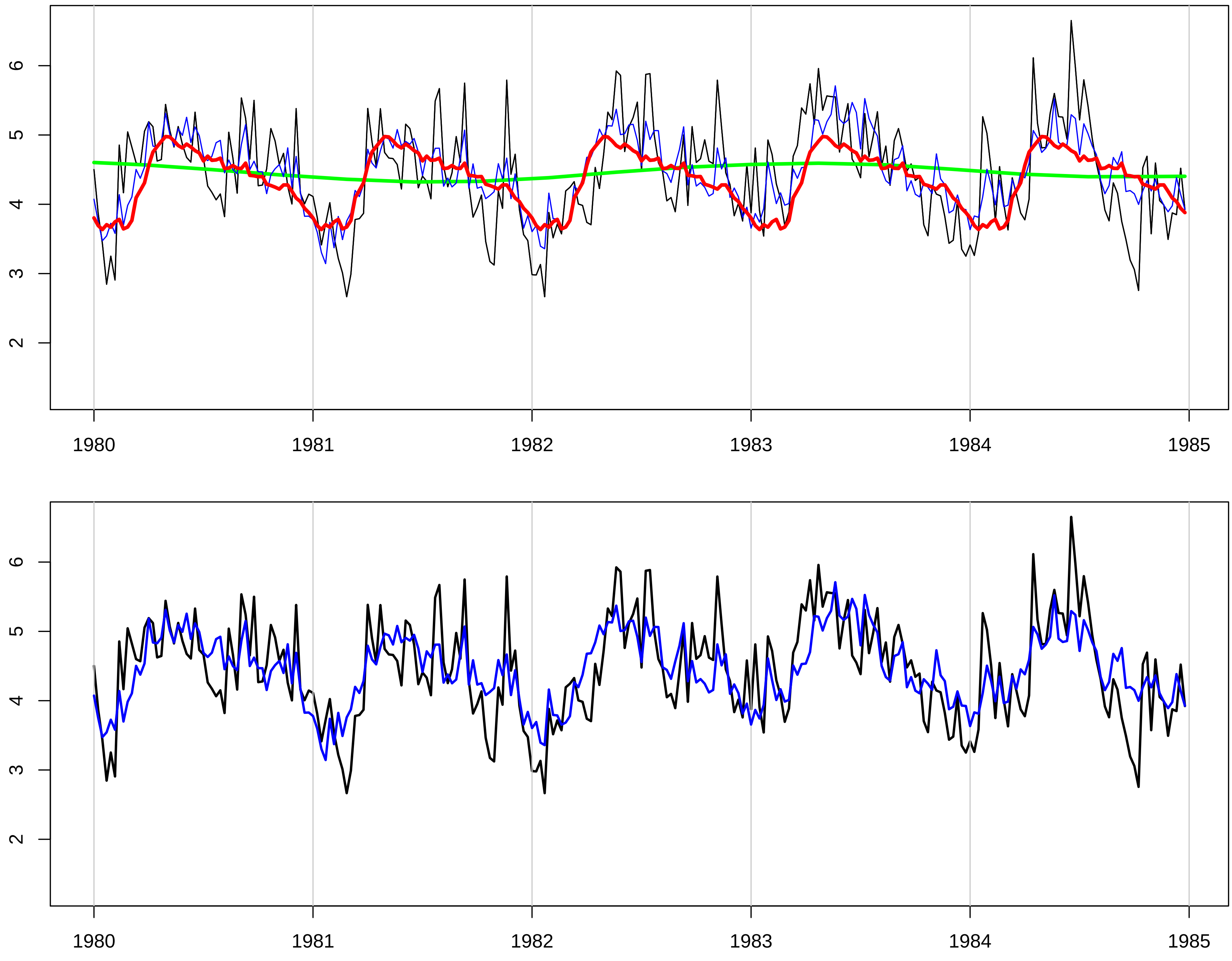
Seasonal S_t estimated with a 7 yr window

$$\log(X_t - \theta_t) - \mu_t = T_t + S_t + R_t$$

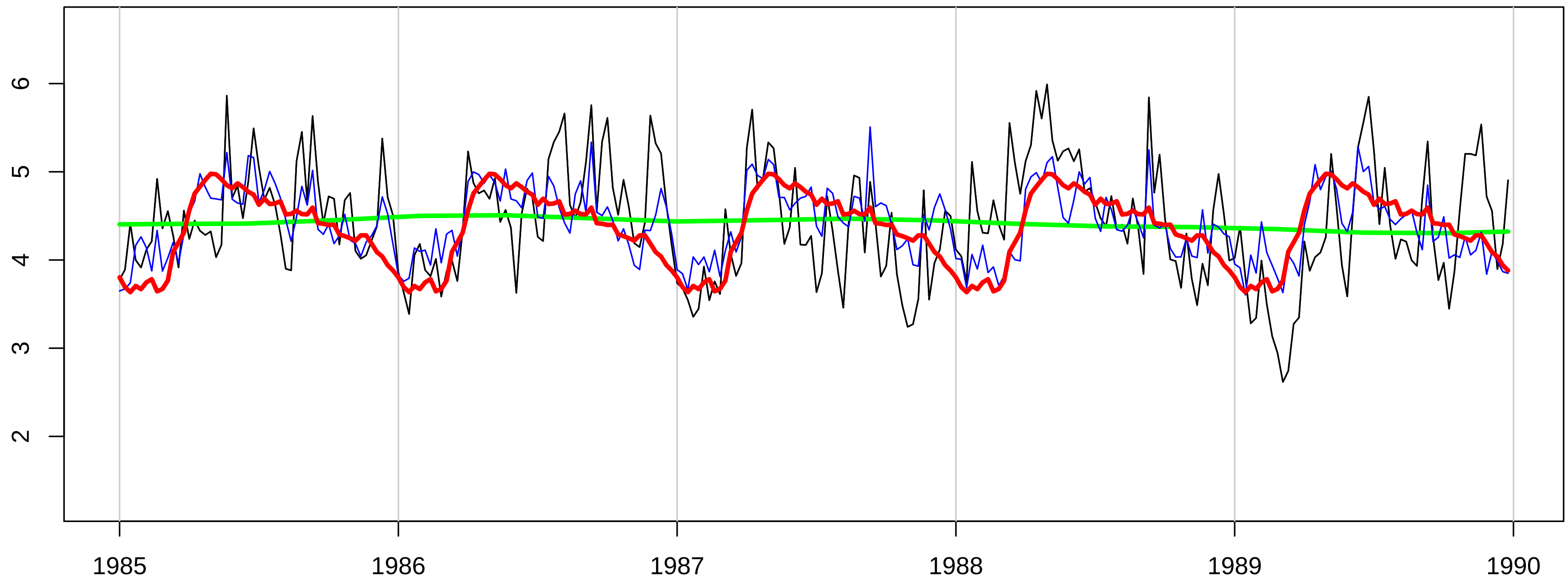
$$\text{green line} = \overline{\mu_t} + T_t$$

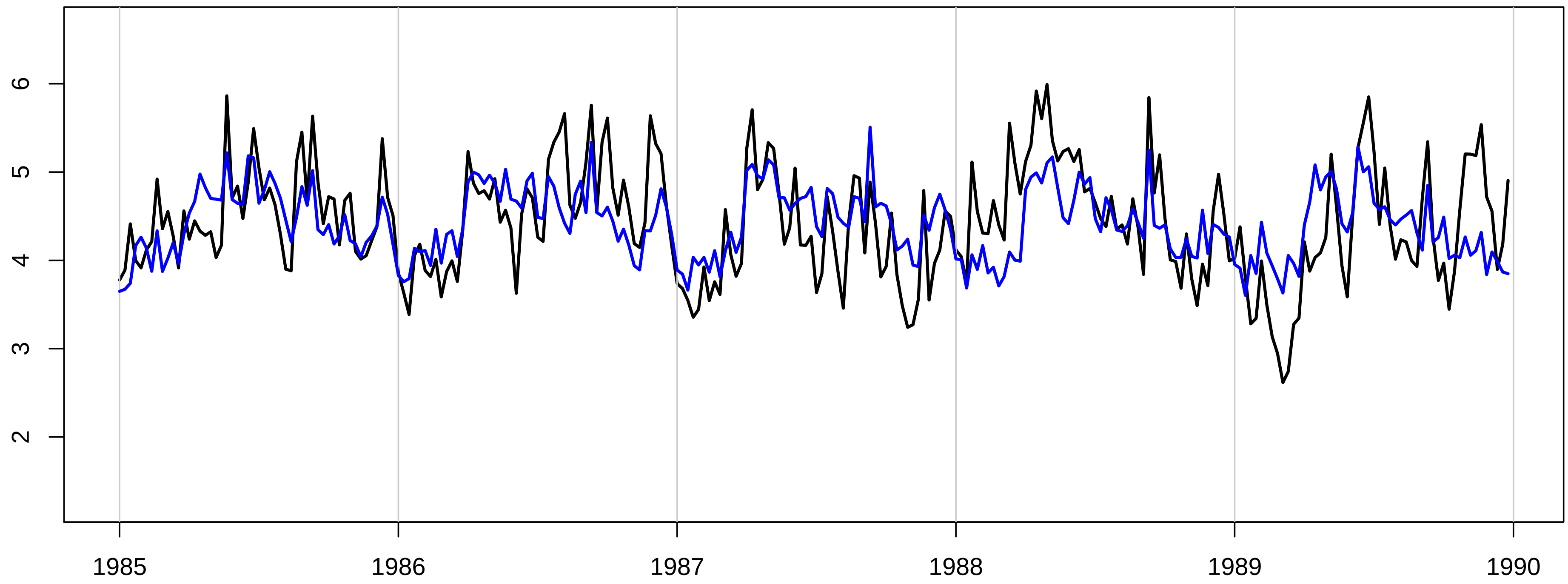
$$\text{blue line} = \mu_t + T_t + S_t$$

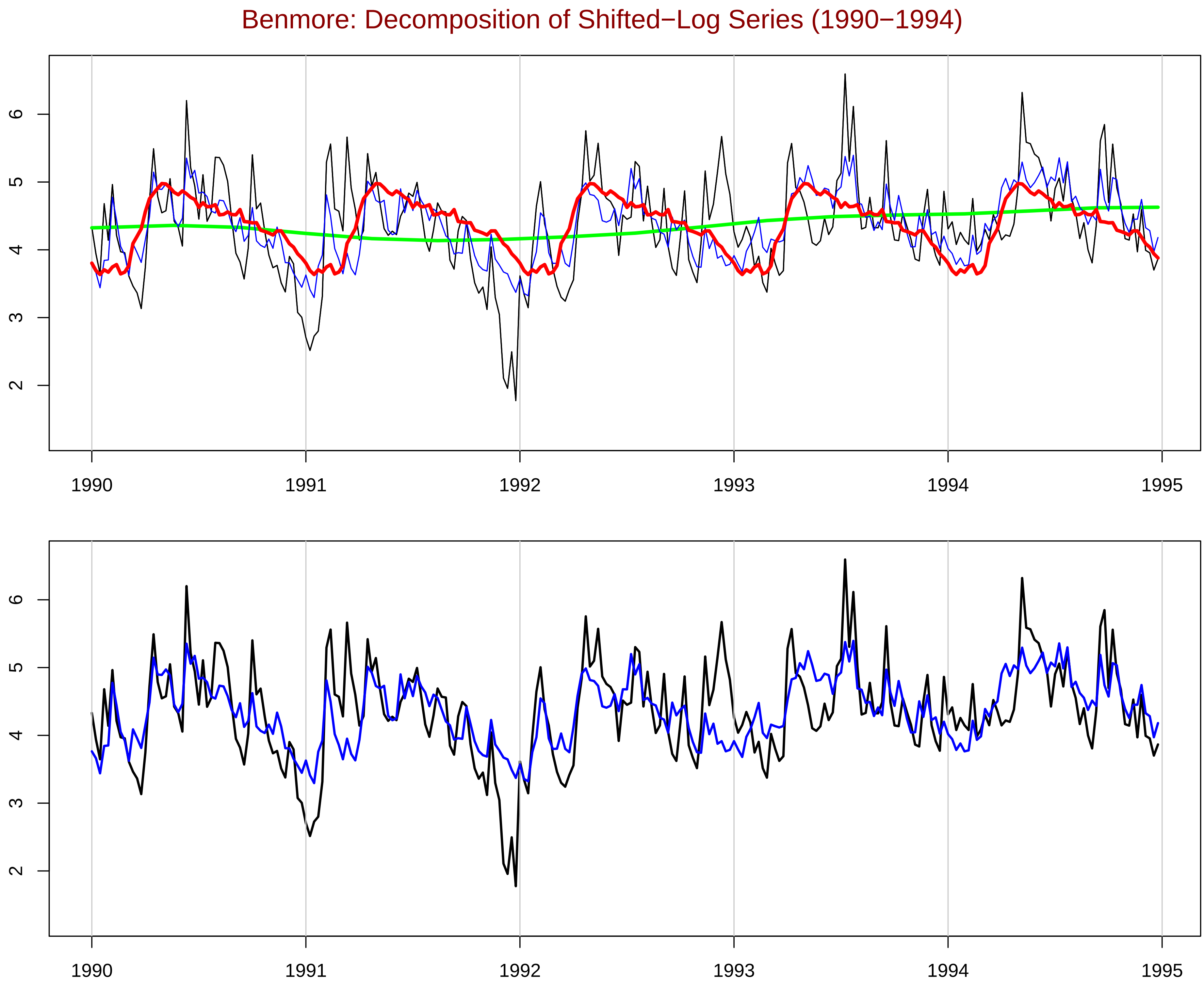
Benmore: Decomposition of Shifted−Log Series (1980−1984)



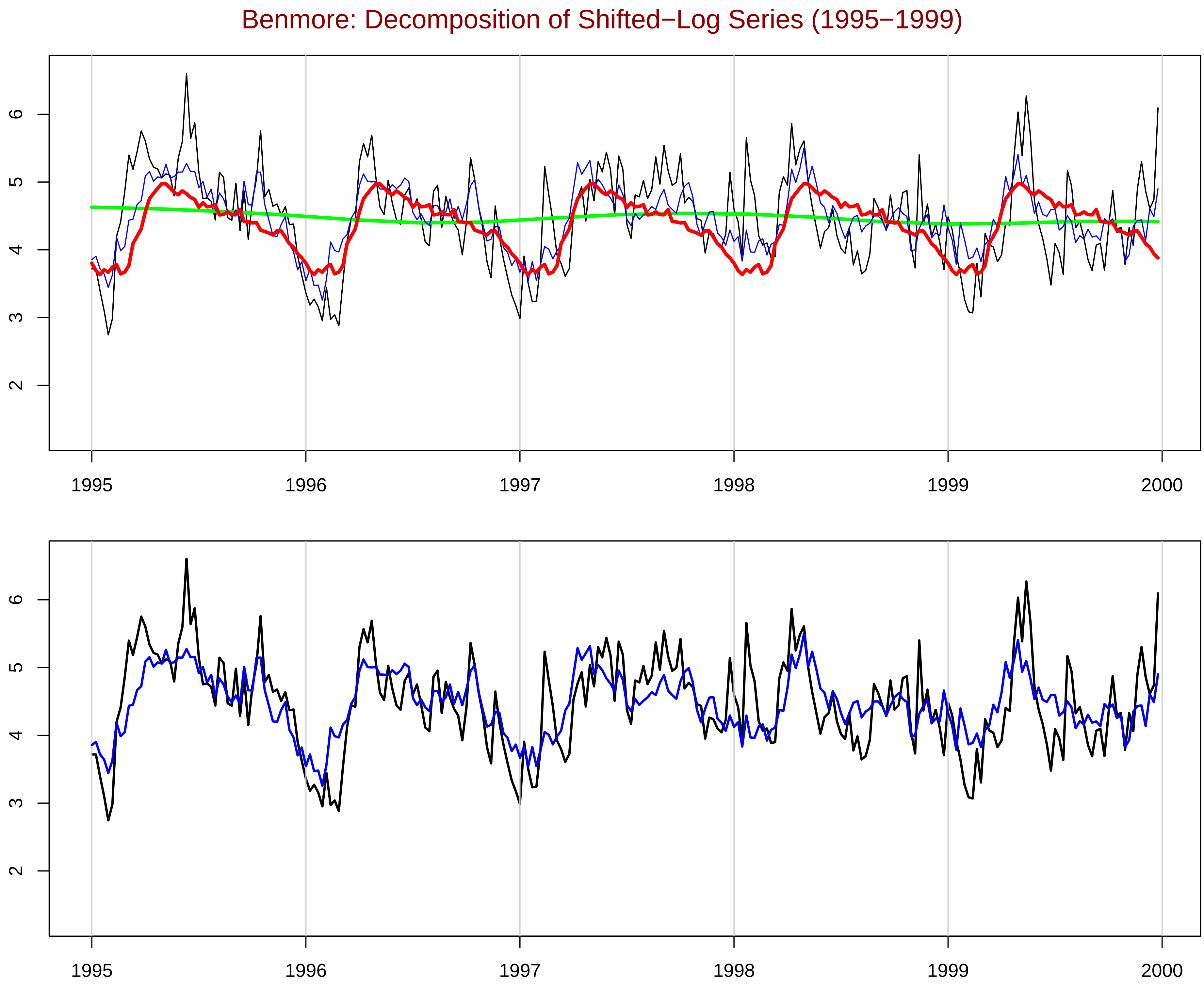
Benmore: Decomposition of Shifted−Log Series (1985−1989)







Benmore: Decomposition of Shifted−Log Series (1990−1994)
2
3
4
5
6
1990
1991
1992
1993
1994
1995
2
3
4
5
6
1990
1991
1992
1993
1994
1995



Benmore: Decomposition of Shifted−Log Series (1995−1999)
2
3
4
5
6
1995
1996
1997
1998
1999
2000
2
3
4
5
6
1995
1996
1997
1998
1999
2000

Parametric Linear PAR(1) Model

$\{Y_t\}$ is a 1st order *periodic autoregressive process* if Y_t has mean μ_t and variance σ_t^2, and satisfies

$$(Y_t - \mu_t) = \phi_t(Y_{t-1} - \mu_{t-1}) + \epsilon_t$$

ϵ_t has zero mean and variance $\sigma_t^2 - \phi_t^2\sigma_{t-1}^2$

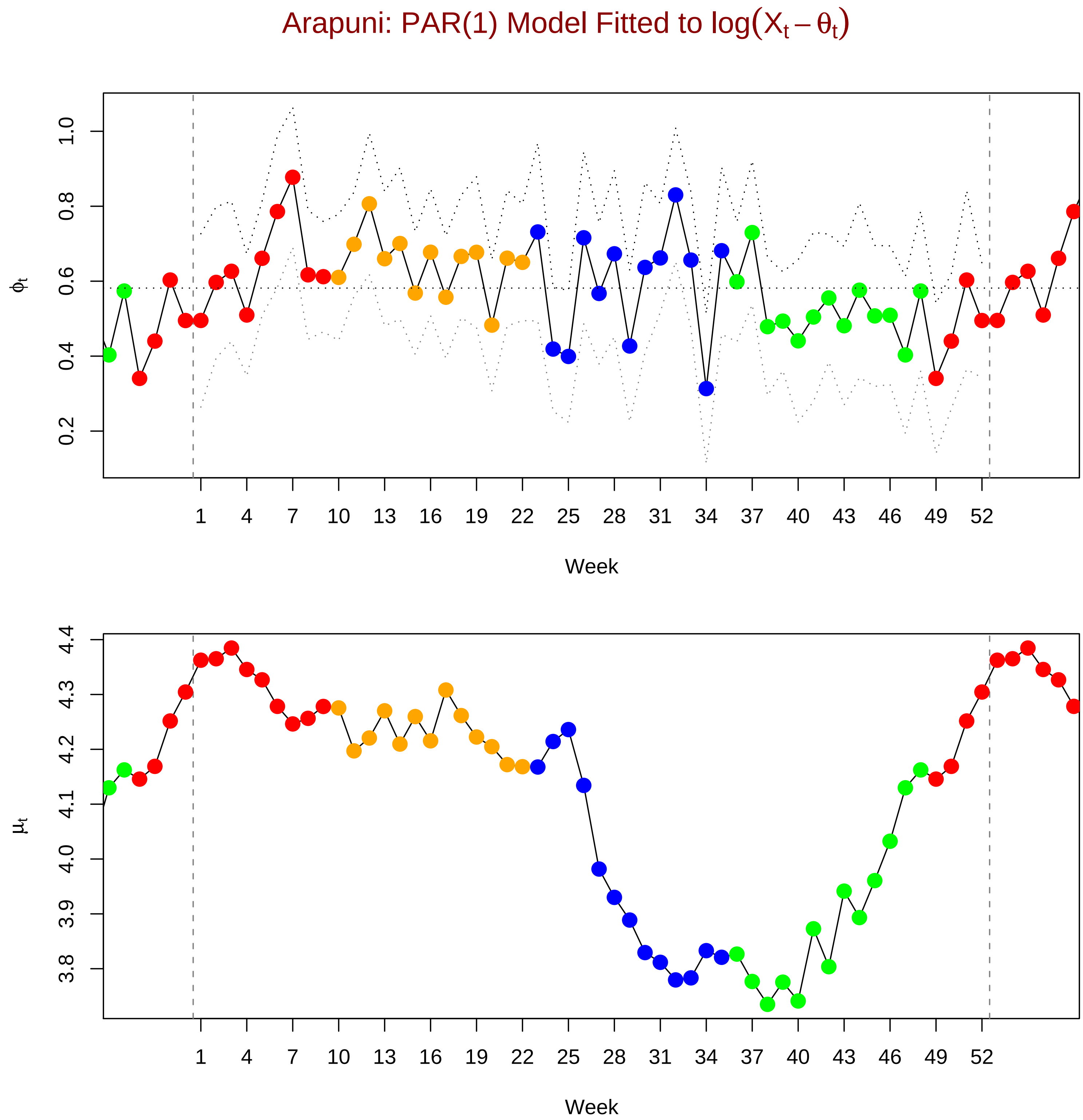
$$\begin{aligned} \phi_t &= \phi_{t+52} \\ \mu_t &= \mu_{t+52} \\ \sigma_t &= \sigma_{t+52} \\ \theta_t &= \theta_{t+52} \end{aligned}$$

Shifted-log series:

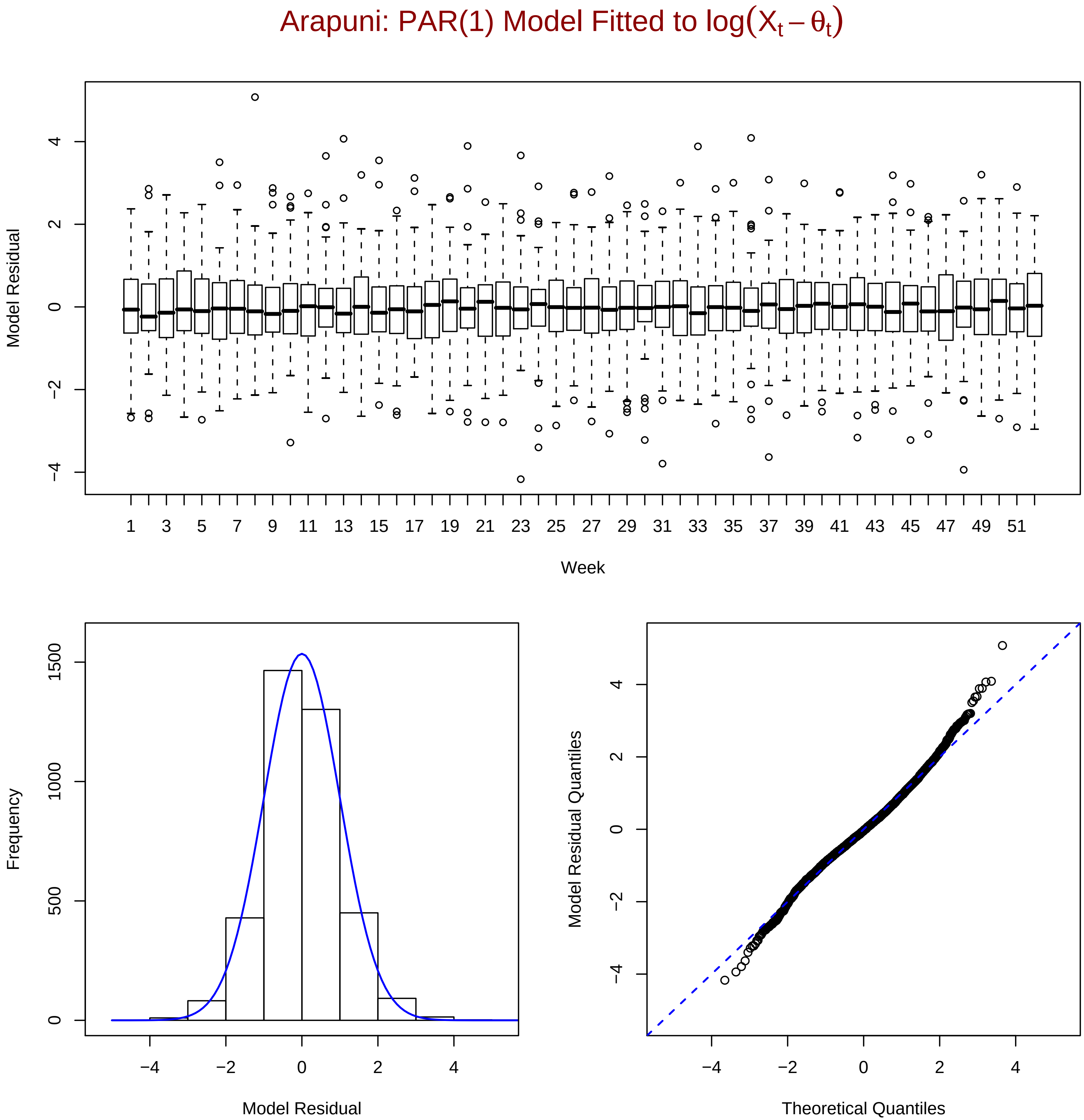
$$Y_t = \log(X_t - \theta_t)$$

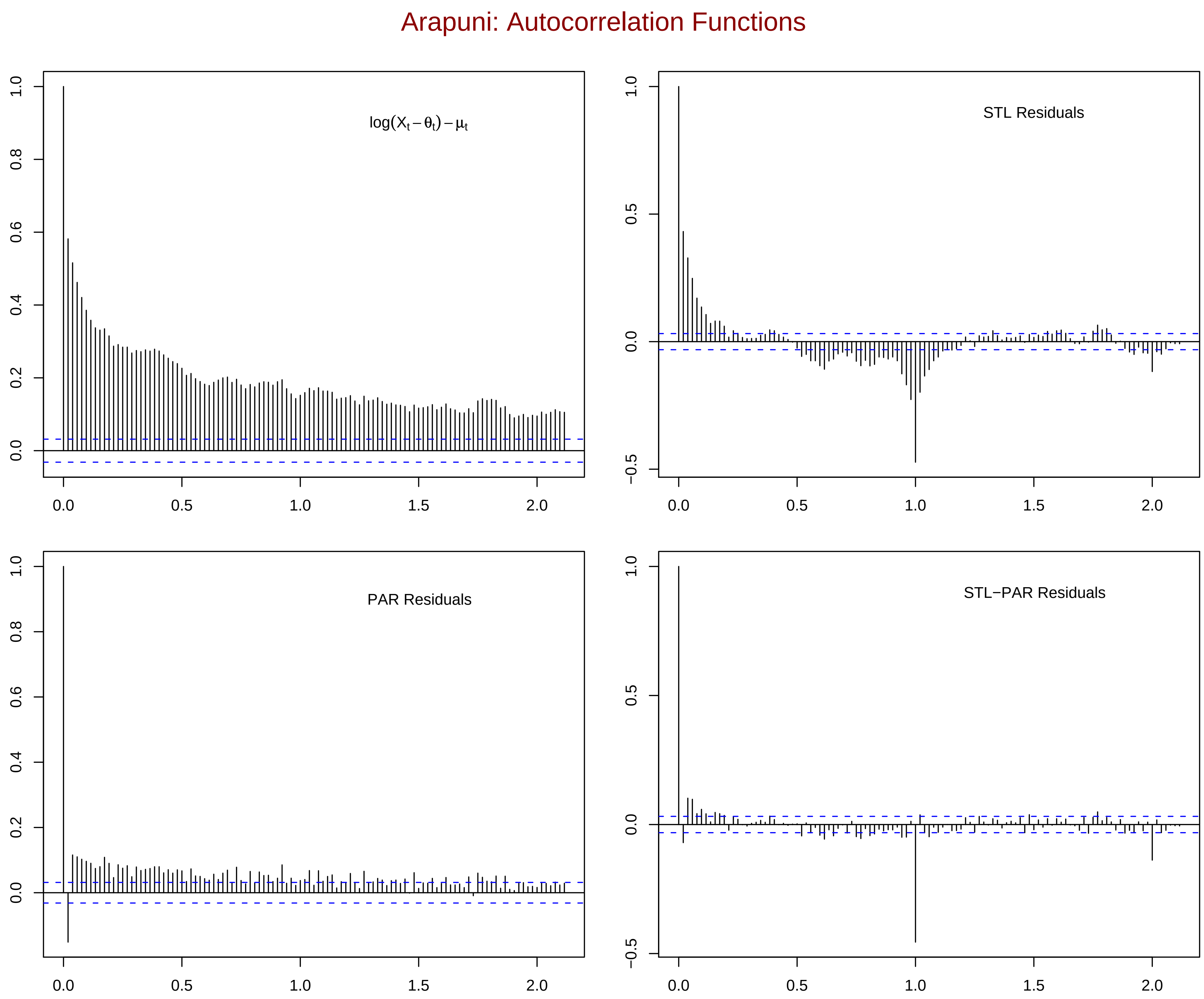
STL residual series:

$$R_t = \log(X_t - \theta_t) - \mu_t - T_t - S_t$$



Arapuni: PAR(1) Model Fitted to $\log(X_t - \theta_t)$
ϕ_t
1.0
0.8
0.6
0.4
0.2
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52
Week
μ_t
4.4
4.3
4.2
4.1
4.0
3.9
3.8
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52
Week


Arapuni: PAR(1) Model Fitted to $\log(X_t - \theta_t)$
Model Residual
Week
Frequency
Model Residual
Model Residual Quantiles
Theoretical Quantiles



Arapuni: Autocorrelation Functions
$\log(X_t - \theta_t) - \mu_t$
STL Residuals
PAR Residuals
STL−PAR Residuals
0.0
0.2
0.4
0.6
0.8
1.0
−0.5
0.5
1.5
2.0

Semiparametric Model: Srinivas & Srinivasan

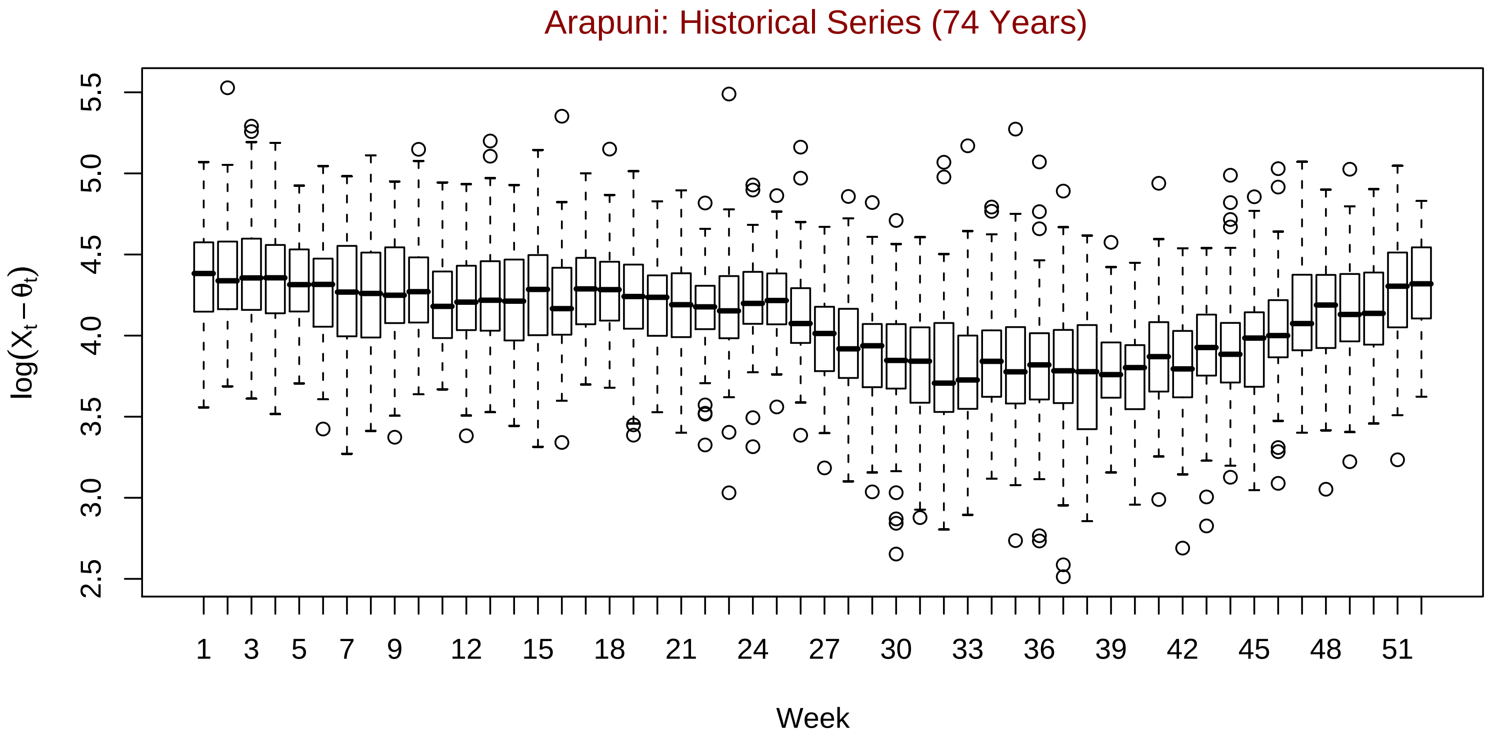
Fit PAR(1) to $Y_t = \log(X_t - \theta_t)$

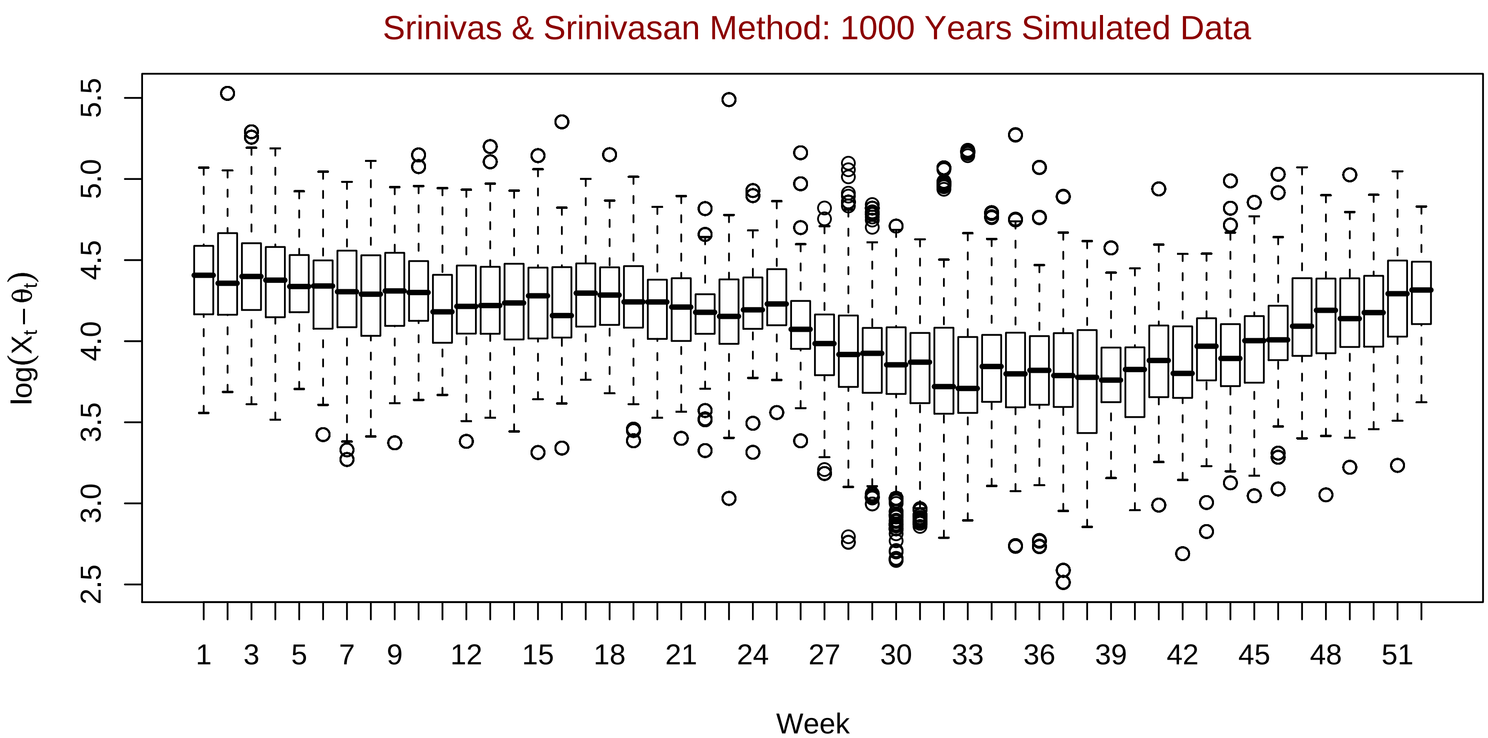
Extract innovations: $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$

Simulation: divide into 2 year overlapping blocks:

$$\begin{gathered}\epsilon_1, \cdots, \epsilon_{104} \\ \epsilon_{53}, \cdots, \epsilon_{156} \\ \epsilon_{105}, \cdots, \epsilon_{208} \\ \epsilon_{157}, \cdots, \epsilon_{260} \\ \vdots \\ \epsilon_{n-103}, \cdots, \epsilon_n\end{gathered}$$

To simulate m years, sample with replacement $m/2$ blocks and post-blacken


Arapuni: Historical Series (74 Years)
log(Xt − θt)
2.5
3.0
3.5
4.0
4.5
5.0
5.5
1 3 5 7 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51
Week


Srinivas & Srinivasan Method: 1000 Years Simulated Data
log(Xt − θt)
2.5
3.0
3.5
4.0
4.5
5.0
5.5
1 3 5 7 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51
Week

The historical data can be thought of as containing components:

$$\{\text{PAR(1)}\} + \{\text{evolving } T_t \text{ and } S_t\} + \{\text{other}\}$$

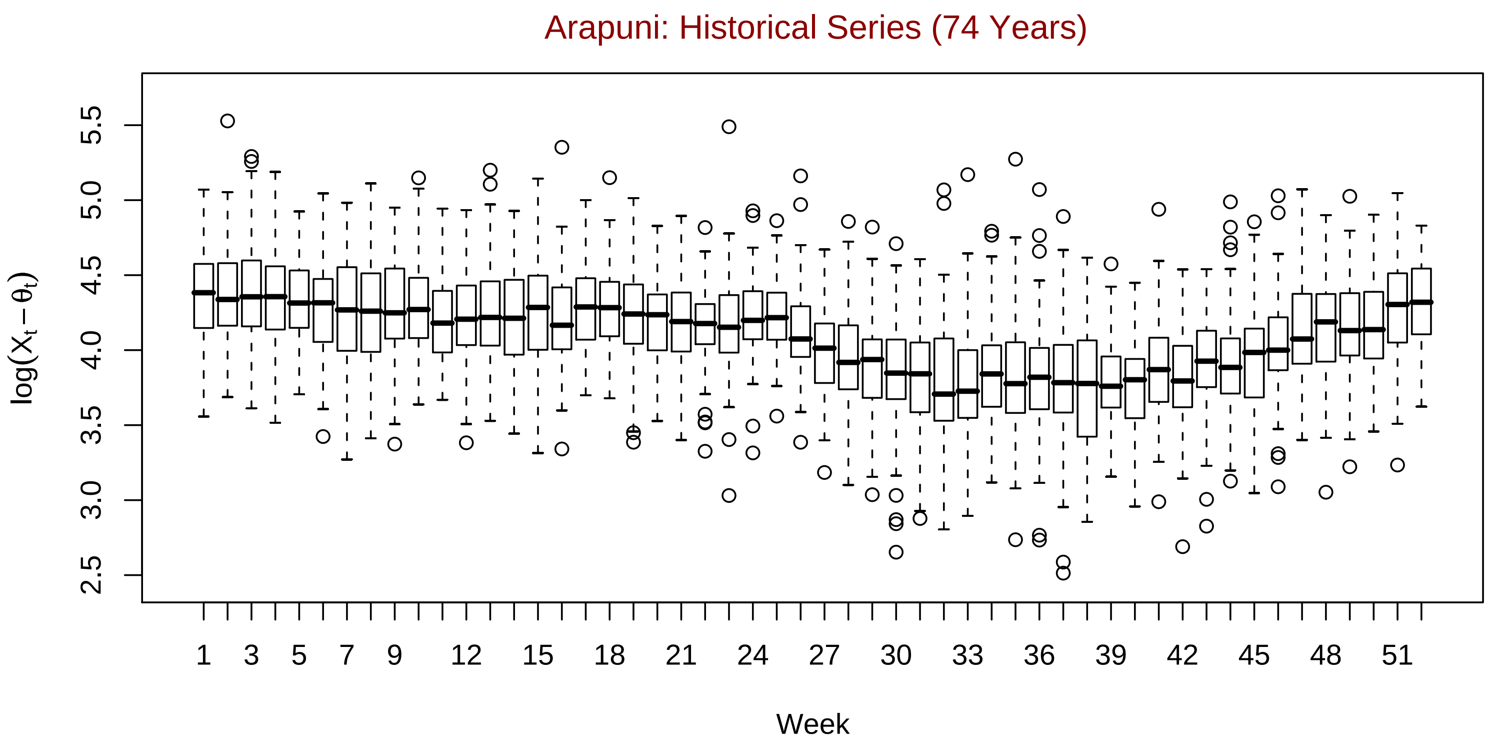
Semiparametric Modified Model

Fit PAR(1) to: $\quad R_t = \log(X_t - \theta_t) - \mu_t - T_t - S_t$

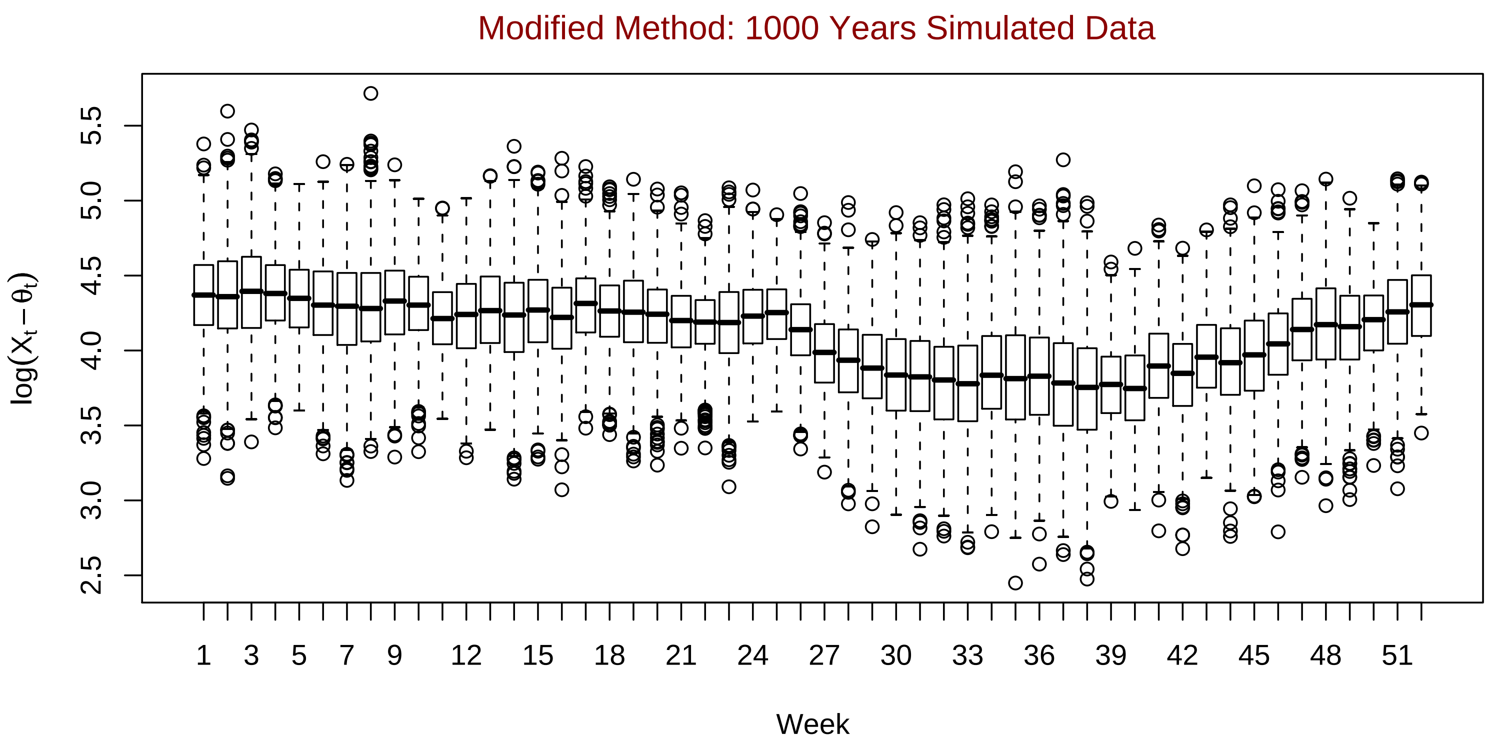
Simulate R_t by generating white noise and post-blackening

Add in the $T_t + S_t$ component by block bootstrapping (2 yr blocks) from:

$$\begin{gathered} T_1 + S_1, \cdots, T_{104} + S_{104} \\ T_{53} + S_{53}, \cdots, T_{156} + S_{156} \\ T_{105} + S_{105}, \cdots, T_{208} + S_{208} \\ T_{157} + S_{157}, \cdots, T_{260} + S_{260} \\ \vdots \\ T_{n-103} + S_{n-103}, \cdots, T_n + S_n \end{gathered}$$



Arapuni: Historical Series (74 Years)
log(Xt−θt)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
1 3 5 7 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51
Week



Modified Method: 1000 Years Simulated Data
log(Xt−θt)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
1 3 5 7 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51
Week

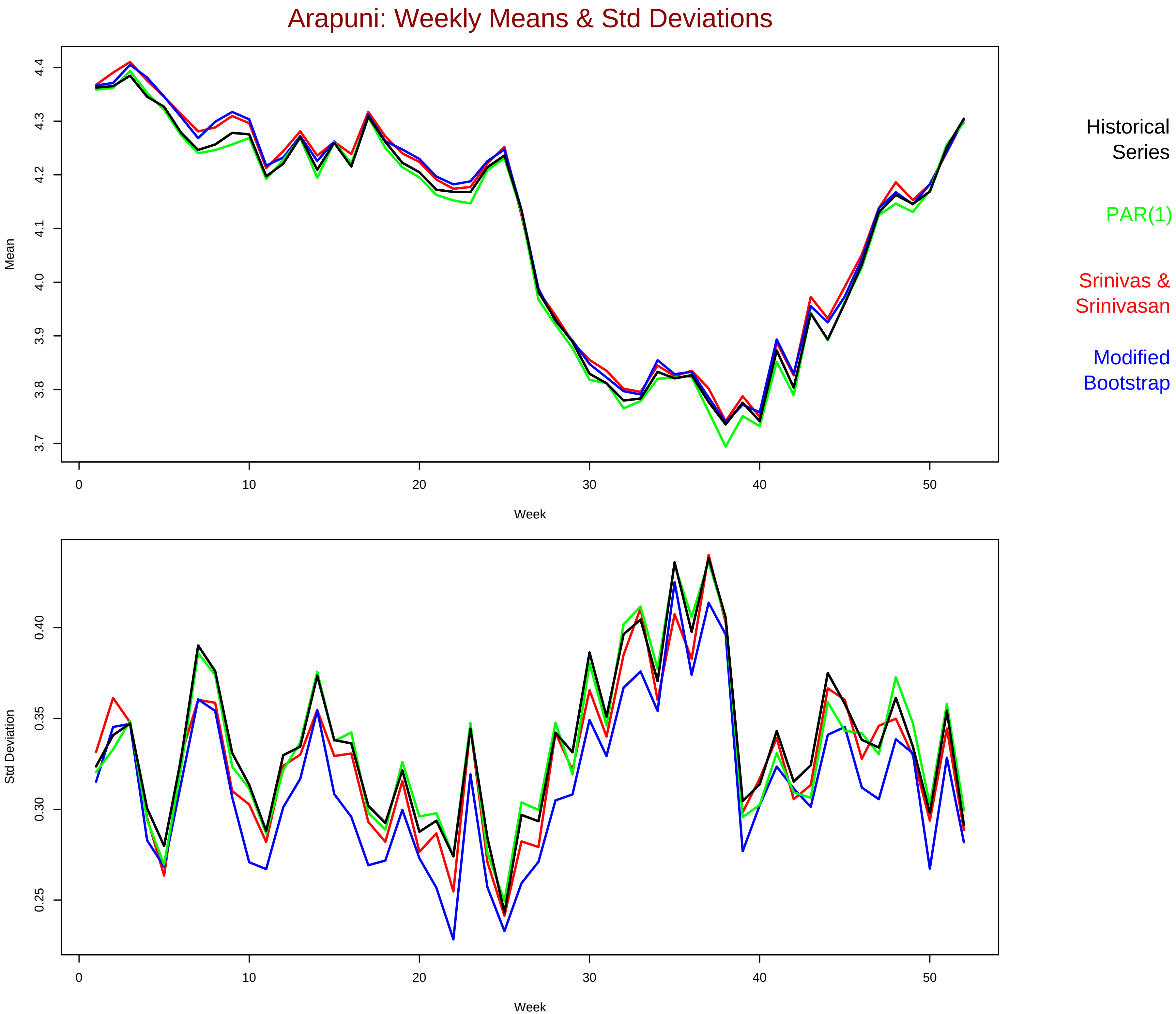
Evaluation of Forecasting Performance

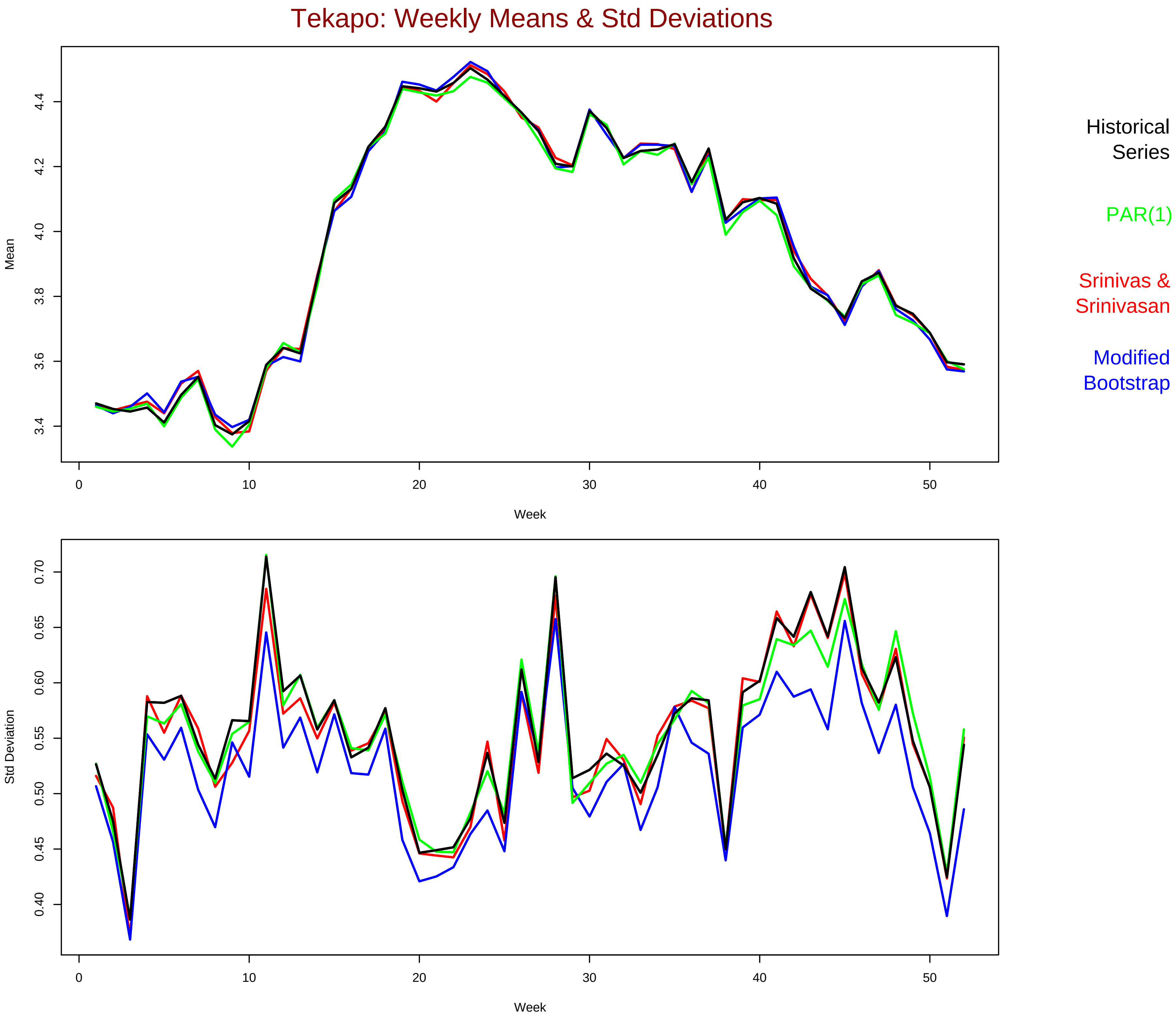
The historical data can be thought of as containing components:

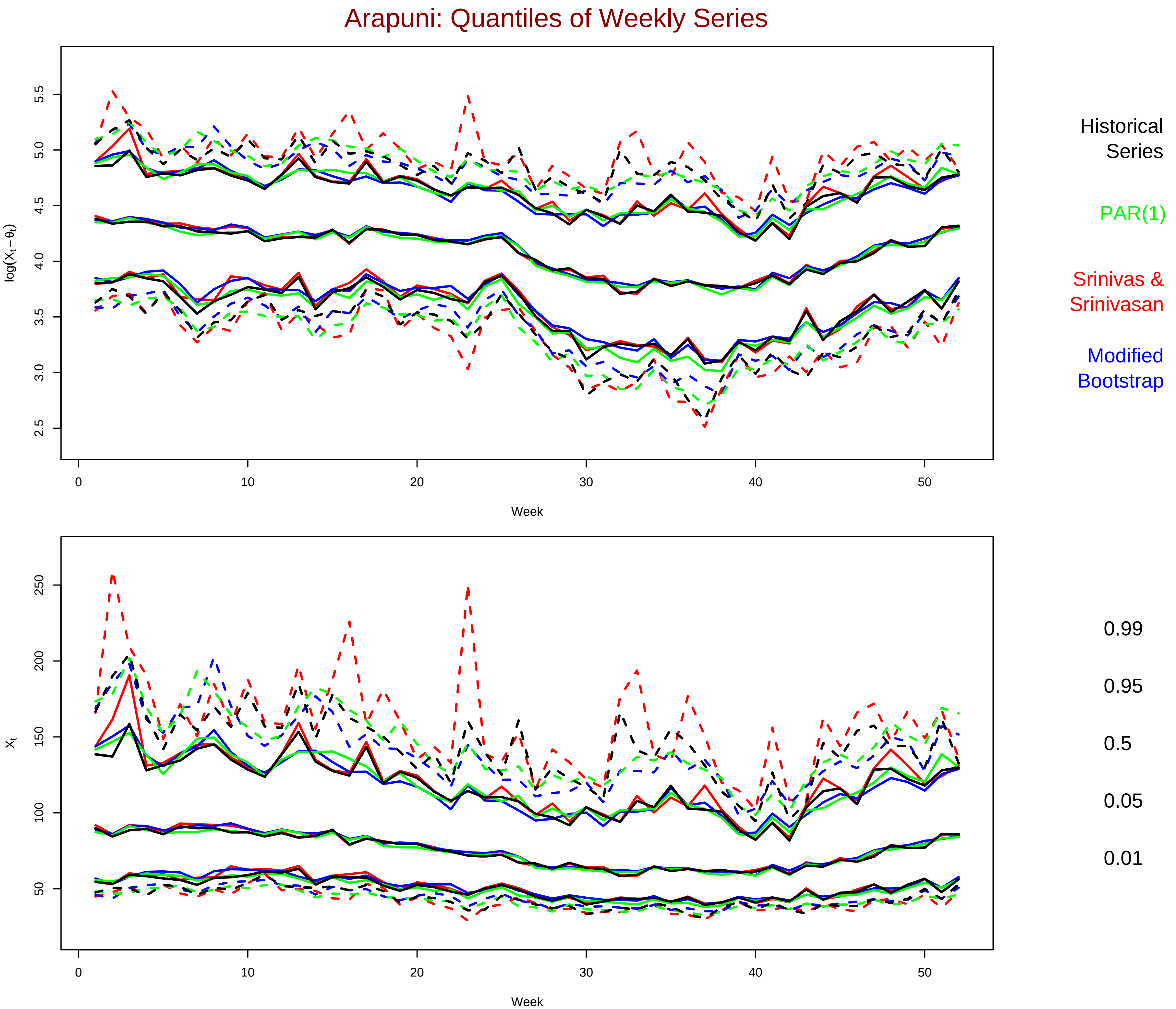
$$\{\text{PAR(1)}\} + \{\text{evolving } T_t \text{ and } S_t\} + \{\text{other}\}$$

Srinivas & Srinivasan's *semiparametric* model contains all components

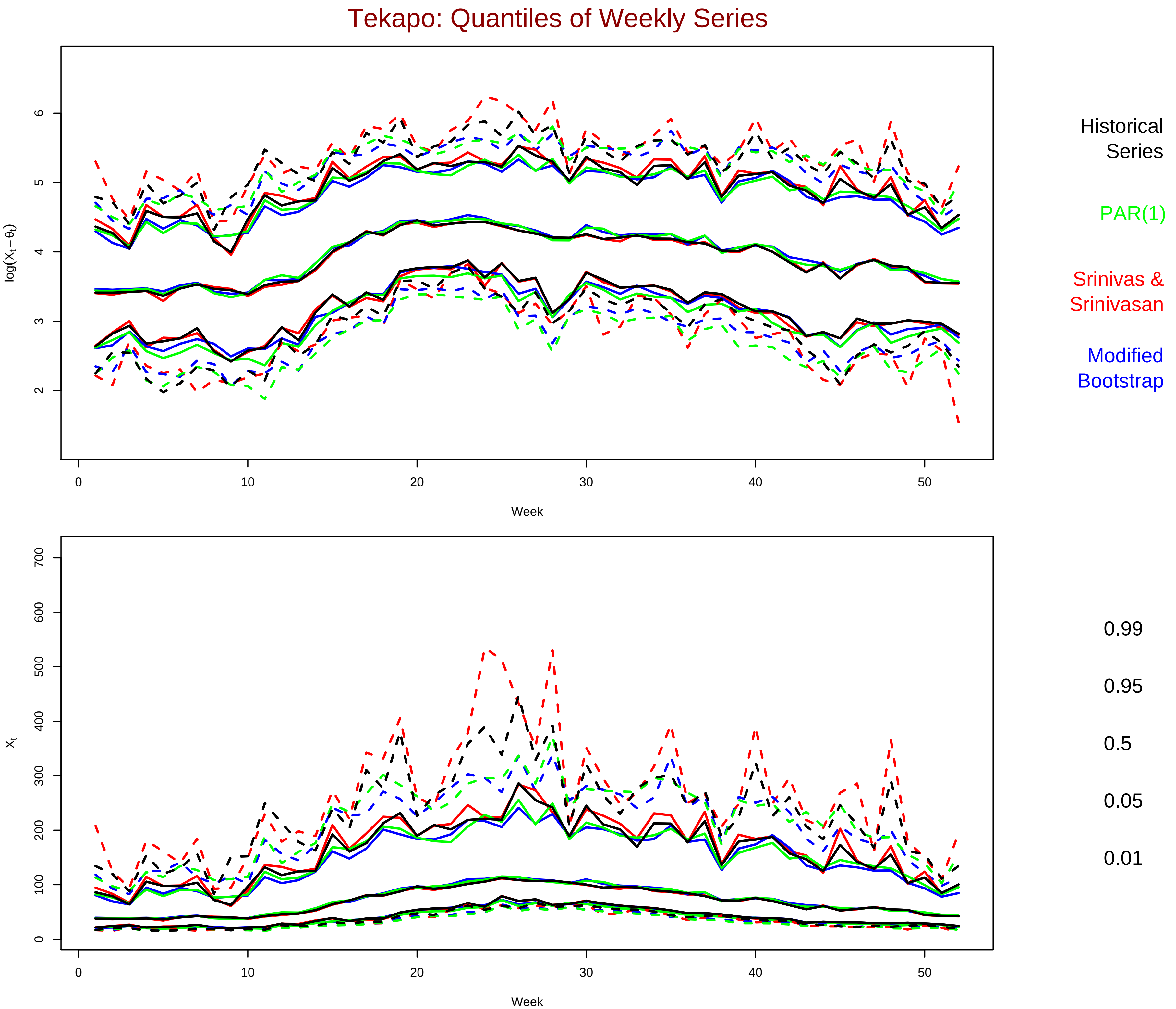
The modified *semiparametric* model contains components 1 and 2

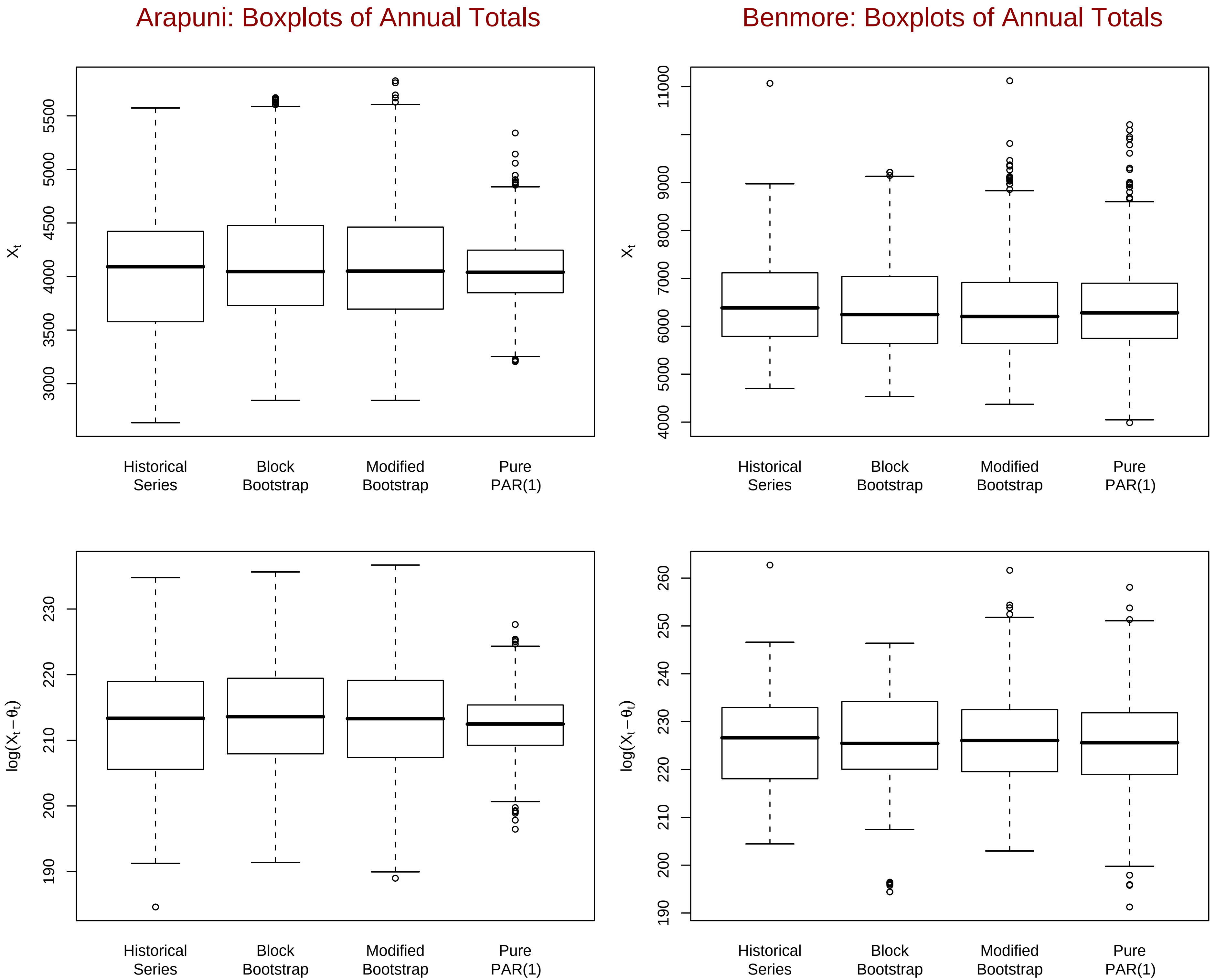

Arapuni: Weekly Means & Std Deviations
Mean
Week
Std Deviation
Week
Historical
Series
PAR(1)
Srinivas &
Srinivasan
Modified
Bootstrap


Tekapo: Weekly Means & Std Deviations
Mean
3.4
3.6
3.8
4.0
4.2
4.4
0
10
20
30
40
50
Week
Std Deviation
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0
10
20
30
40
50
Week
Historical Series
PAR(1)
Srinivas & Srinivasan
Modified Bootstrap



Arapuni: Quantiles of Weekly Series
$\log(X_t - \theta_t)$
5.5
5.0
4.5
4.0
3.5
3.0
2.5
0
10
20
30
40
50
Week
Historical Series
PAR(1)
Srinivas & Srinivasan
Modified Bootstrap
X_t
250
200
150
100
50
0
10
20
30
40
50
Week
0.99
0.95
0.5
0.05
0.01


Tekapo: Quantiles of Weekly Series
log(X_t − θ_t)
6
5
4
3
2
0
10
20
30
40
50
Week
Historical Series
PAR(1)
Srinivas & Srinivasan
Modified Bootstrap
X_t
700
600
500
400
300
200
100
0
0
10
20
30
40
50
Week
0.99
0.95
0.5
0.05
0.01



Arapuni: Boxplots of Annual Totals
Benmore: Boxplots of Annual Totals
X_t
$\log(X_t - \theta_t)$
3000
3500
4000
4500
5000
5500
4000
5000
6000
7000
8000
9000
11000
190
200
210
220
230
240
250
260
Historical Series
Block Bootstrap
Modified Bootstrap
Pure PAR(1)

Runs of Low Inflows

Define a “run” as a sequence of weekly values that are all less than some threshold value

Arapuni: Threshold Value $= 65$

Historical Series

	0	2	4	6	8	10	12	14	16	18	20	22	24	26
28	0.44	0.68	0.75	0.85	0.86	0.88	0.92	0.93	0.95	0.96	0.99	1	1	1
32	0.36	0.63	0.70	0.75	0.81	0.85	0.89	0.92	0.97	0.99	1	1	1	1
36	0.41	0.56	0.71	0.78	0.88	0.90	0.97	0.99	1	1	1	1	1	1
40	0.38	0.63	0.81	0.89	0.97	0.99	1	1	1	1	1	1	1	1
44	0.48	0.81	0.95	0.97	0.99	1	1	1	1	1	1	1	1	1
48	0.71	0.95	0.96	0.99	1	1	1	1	1	1	1	1	1	1
52	0.85	0.95	0.99	1	1	1	1	1	1	1	1	1	1	1
4	0.92	1	1	1	1	1	1	1	1	1	1	1	1	1
8	0.86	0.95	0.96	0.96	0.96	0.97	0.99	0.99	0.99	0.99	0.99	0.99	1	1
12	0.89	0.95	0.95	0.97	0.99	0.99	0.99	0.99	0.99	0.99	1	1	1	1
16	0.78	0.95	0.97	0.97	0.97	0.99	0.99	0.99	1	1	1	1	1	1
20	0.74	0.88	0.92	0.95	0.96	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
24	0.68	0.89	0.92	0.95	0.96	0.97	0.97	0.97	0.99	0.99	0.99	0.99	0.99	1

Pure PAR(1) Model

	0	2	4	6	8	10	12	14	16	18	20	22	24	26
28	0.46	0.73	0.80	0.90	0.94	0.96	0.97	0.99	0.99	1	1	1	1	1
32	0.41	0.71	0.83	0.90	0.93	0.96	0.99	0.99	1	1	1	1	1	1
36	0.46	0.68	0.79	0.89	0.96	0.98	0.99	1	1	1	1	1	1	1
40	0.31	0.65	0.87	0.96	0.99	1	1	1	1	1	1	1	1	1
44	0.50	0.82	0.94	0.99	1	1	1	1	1	1	1	1	1	1
48	0.68	0.93	0.99	1	1	1	1	1	1	1	1	1	1	1
52	0.85	0.98	1	1	1	1	1	1	1	1	1	1	1	1
4	0.90	0.98	0.99	1	1	1	1	1	1	1	1	1	1	1
8	0.87	0.98	0.99	1	1	1	1	1	1	1	1	1	1	1
12	0.89	0.97	0.99	1	1	1	1	1	1	1	1	1	1	1
16	0.82	0.95	0.98	0.99	0.99	1	1	1	1	1	1	1	1	1
20	0.75	0.88	0.95	0.98	0.98	0.99	0.99	1	1	1	1	1	1	1
24	0.65	0.87	0.92	0.95	0.97	0.98	0.99	0.99	1	1	1	1	1	1

Srinivas and Srinivasan Model

	0	2	4	6	8	10	12	14	16	18	20	22	24	26
28	0.44	0.72	0.80	0.88	0.90	0.91	0.94	0.95	0.96	0.97	0.99	1	1	1
32	0.37	0.64	0.72	0.77	0.82	0.86	0.91	0.94	0.98	0.99	1	1	1	1
36	0.41	0.56	0.71	0.77	0.89	0.92	0.98	0.99	1	1	1	1	1	1
40	0.40	0.63	0.83	0.91	0.98	0.99	1	1	1	1	1	1	1	1
44	0.51	0.84	0.95	0.98	0.99	1	1	1	1	1	1	1	1	1
48	0.73	0.94	0.95	0.98	1	1	1	1	1	1	1	1	1	1
52	0.85	0.94	0.99	1	1	1	1	1	1	1	1	1	1	1
4	0.93	1	1	1	1	1	1	1	1	1	1	1	1	1
8	0.90	0.96	0.98	0.98	0.98	0.99	1	1	1	1	1	1	1	1
12	0.91	0.96	0.96	0.99	1	1	1	1	1	1	1	1	1	1
16	0.81	0.97	0.99	0.99	0.99	1	1	1	1	1	1	1	1	1
20	0.77	0.90	0.93	0.95	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
24	0.70	0.89	0.93	0.97	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1

Modified Model

	0	2	4	6	8	10	12	14	16	18	20	22	24	26
28	0.47	0.76	0.84	0.90	0.93	0.95	0.96	0.97	0.98	0.99	0.99	1	1	1
32	0.45	0.73	0.84	0.89	0.91	0.94	0.96	0.98	0.99	1	1	1	1	1
36	0.47	0.69	0.77	0.86	0.92	0.96	0.98	0.99	1	1	1	1	1	1
40	0.34	0.68	0.86	0.93	0.97	0.99	0.99	1	1	1	1	1	1	1
44	0.53	0.83	0.95	0.98	0.99	1	1	1	1	1	1	1	1	1
48	0.74	0.94	0.98	1	1	1	1	1	1	1	1	1	1	1
52	0.86	0.98	0.99	1	1	1	1	1	1	1	1	1	1	1
4	0.92	0.97	0.99	1	1	1	1	1	1	1	1	1	1	1
8	0.92	0.99	1	1	1	1	1	1	1	1	1	1	1	1
12	0.91	0.98	0.99	1	1	1	1	1	1	1	1	1	1	1
16	0.85	0.96	0.99	0.99	0.99	1	1	1	1	1	1	1	1	1
20	0.81	0.91	0.96	0.98	1	1	1	1	1	1	1	1	1	1
24	0.70	0.89	0.94	0.97	0.97	0.98	0.99	0.99	0.99	1	1	1	1	1

Te Anau: Threshold Value = 260

Historical Series

	0	2	4	6	8	10	12	14	16	18	20	22	24	26
28	0.41	0.70	0.85	0.88	0.90	0.93	0.95	0.96	0.97	0.99	0.99	0.99	0.99	0.99
32	0.37	0.63	0.78	0.88	0.92	0.93	0.96	0.99	0.99	0.99	0.99	0.99	0.99	0.99
36	0.40	0.73	0.85	0.90	0.95	0.97	0.97	0.97	0.97	0.99	0.99	0.99	0.99	1
40	0.56	0.79	0.90	0.95	0.95	0.96	0.97	0.99	0.99	0.99	0.99	1	1	1
44	0.41	0.71	0.84	0.88	0.92	0.97	0.97	0.97	0.97	1	1	1	1	1
48	0.41	0.73	0.82	0.93	0.93	0.97	0.97	1	1	1	1	1	1	1
52	0.30	0.68	0.77	0.88	0.93	0.97	0.99	1	1	1	1	1	1	1
4	0.23	0.53	0.75	0.85	0.95	0.99	0.99	0.99	1	1	1	1	1	1
8	0.29	0.56	0.82	0.96	0.99	0.99	1	1	1	1	1	1	1	1
12	0.51	0.79	0.93	0.95	0.99	1	1	1	1	1	1	1	1	1
16	0.59	0.84	0.96	0.99	0.99	1	1	1	1	1	1	1	1	1
20	0.62	0.84	0.95	0.96	0.97	0.99	0.99	0.99	0.99	1	1	1	1	1
24	0.44	0.82	0.93	0.95	0.97	0.97	0.97	0.99	0.99	1	1	1	1	1

Pure PAR(1) Model

	0	2	4	6	8	10	12	14	16	18	20	22	24	26
28	0.43	0.66	0.80	0.88	0.92	0.95	0.98	0.99	0.99	1	1	1	1	1
32	0.33	0.58	0.73	0.86	0.94	0.97	0.98	0.99	1	1	1	1	1	1
36	0.33	0.67	0.87	0.93	0.96	0.99	0.99	0.99	1	1	1	1	1	1
40	0.49	0.74	0.86	0.96	0.98	0.99	0.99	1	1	1	1	1	1	1
44	0.37	0.73	0.86	0.93	0.96	0.97	0.98	0.99	0.99	1	1	1	1	1
48	0.41	0.66	0.79	0.87	0.91	0.95	0.97	0.98	0.99	1	1	1	1	1
52	0.34	0.60	0.73	0.83	0.90	0.94	0.98	0.99	1	1	1	1	1	1
4	0.24	0.55	0.73	0.84	0.93	0.97	1	1	1	1	1	1	1	1
8	0.25	0.53	0.79	0.91	0.98	0.99	1	1	1	1	1	1	1	1
12	0.54	0.81	0.96	0.98	0.99	1	1	1	1	1	1	1	1	1
16	0.63	0.87	0.94	0.97	0.99	0.99	1	1	1	1	1	1	1	1
20	0.57	0.81	0.92	0.96	0.99	0.99	1	1	1	1	1	1	1	1
24	0.46	0.78	0.91	0.94	0.97	0.98	0.99	0.99	1	1	1	1	1	1

Srinivas and Srinivasan Model

	0	2	4	6	8	10	12	14	16	18	20	22	24	26
28	0.42	0.71	0.84	0.88	0.90	0.93	0.94	0.96	0.98	0.99	0.99	0.99	0.99	0.99
32	0.35	0.60	0.76	0.87	0.91	0.93	0.96	0.99	0.99	0.99	0.99	0.99	0.99	0.99
36	0.40	0.73	0.86	0.91	0.95	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	1
40	0.54	0.79	0.90	0.95	0.95	0.96	0.98	0.99	0.99	0.99	0.99	1	1	1
44	0.41	0.71	0.85	0.89	0.94	0.98	0.98	0.98	0.98	1	1	1	1	1
48	0.41	0.74	0.84	0.95	0.95	0.98	0.98	1	1	1	1	1	1	1
52	0.29	0.68	0.77	0.89	0.95	0.97	0.99	1	1	1	1	1	1	1
4	0.23	0.55	0.76	0.84	0.95	0.99	0.99	0.99	1	1	1	1	1	1
8	0.30	0.53	0.83	0.96	0.99	0.99	1	1	1	1	1	1	1	1
12	0.51	0.79	0.93	0.94	0.98	1	1	1	1	1	1	1	1	1
16	0.57	0.83	0.95	0.98	0.98	1	1	1	1	1	1	1	1	1
20	0.59	0.82	0.93	0.95	0.98	0.99	0.99	0.99	0.99	1	1	1	1	1
24	0.44	0.80	0.93	0.95	0.97	0.97	0.97	0.98	0.98	1	1	1	1	1

Modified Model

	0	2	4	6	8	10	12	14	16	18	20	22	24	26
28	0.43	0.69	0.81	0.87	0.91	0.95	0.98	0.99	0.99	0.99	1	1	1	1
32	0.36	0.59	0.73	0.87	0.94	0.97	0.98	0.99	1	1	1	1	1	1
36	0.28	0.65	0.84	0.92	0.95	0.98	0.99	0.99	1	1	1	1	1	1
40	0.47	0.76	0.87	0.94	0.97	0.98	0.99	0.99	1	1	1	1	1	1
44	0.36	0.72	0.89	0.93	0.96	0.98	0.98	0.99	1	1	1	1	1	1
48	0.38	0.63	0.78	0.86	0.90	0.95	0.97	0.98	0.99	0.99	1	1	1	1
52	0.29	0.59	0.71	0.84	0.90	0.94	0.97	0.98	0.99	1	1	1	1	1
4	0.25	0.58	0.74	0.85	0.94	0.96	0.99	1	1	1	1	1	1	1
8	0.25	0.54	0.79	0.90	0.97	0.99	0.99	1	1	1	1	1	1	1
12	0.48	0.81	0.95	0.98	0.99	1	1	1	1	1	1	1	1	1
16	0.60	0.85	0.94	0.97	0.98	0.99	1	1	1	1	1	1	1	1
20	0.59	0.82	0.92	0.96	0.99	0.99	1	1	1	1	1	1	1	1
24	0.45	0.80	0.91	0.95	0.97	0.98	0.99	0.99	1	1	1	1	1	1