

Analysis of Time Series

Chapter 6: Extending the ARMA model: Seasonality, integration and trend

Edward L. Ionides

Outline

- 1 Seasonality
 - The SARMA model
- 2 Differencing and integration
 - The ARIMA model
 - The SARIMA model
- 3 Trend estimation: regression with ARMA errors

Seasonal autoregressive moving average (SARMA) models

- A general $\text{SARMA}(p, q) \times (P, Q)_{12}$ model for monthly data is

$$[\text{S1}] \quad \phi(B)\Phi(B^{12})(Y_n - \mu) = \psi(B)\Psi(B^{12})\epsilon_n,$$

where $\{\epsilon_n\}$ is a white noise process and

$$\begin{aligned}\mu &= \mathbb{E}[Y_n] \\ \phi(x) &= 1 - \phi_1 x - \cdots - \phi_p x^p, \\ \psi(x) &= 1 + \psi_1 x + \cdots + \psi_q x^q, \\ \Phi(x) &= 1 - \Phi_1 x - \cdots - \Phi_P x^P, \\ \Psi(x) &= 1 + \Psi_1 x + \cdots + \Psi_Q x^Q.\end{aligned}$$

- SARMA is a special case of ARMA, where the AR and MA polynomials are factored into a **monthly** polynomial in B and an **annual polynomial** (also called **seasonal polynomial**) in B^{12} .
- Everything we learned about ARMA models (including assessing causality, invertibility and reducibility) also applies to SARMA.

Choosing the period for a SARMA model

- For the $\text{SARMA}(p, q) \times (P, Q)_{12}$ model, 12 is called the **period**.
- One could write a SARMA model for some period other than 12.
- A $\text{SARMA}(p, q) \times (P, Q)_4$ model could be appropriate for quarterly data.
- In principle, a $\text{SARMA}(p, q) \times (P, Q)_{52}$ model could be appropriate for weekly data, though in practice ARMA and SARMA may not work so well for higher frequency data.
- The seasonal period should be appropriate for the system being modeled. It is usually inappropriate to fit a $\text{SARMA}(p, q) \times (P, Q)_9$ model just because you notice a high sample autocorrelation at lag 9.

Consider the following two models:

$$[S2] \quad Y_n = 0.5Y_{n-1} + 0.25Y_{n-12} + \epsilon_n,$$

$$[S3] \quad Y_n = 0.5Y_{n-1} + 0.25Y_{n-12} - 0.125Y_{n-13} + \epsilon_n,$$

Question 6.1. Which of [S2] and/or [S3] is a SARMA model?

Notice the factorization $1 - \frac{1}{2}x - \frac{1}{4}x^{12} + \frac{1}{8}x^{13} = (1 - \frac{1}{2}x)(1 - \frac{1}{4}x^{12})$ (*)

From (*), [S3] is

$$(1 - 0.5B)(1 - 0.25B^{12}) Y_n = \epsilon_n$$

so, [S3] is a SARMA model.

$$\text{SARMA}(1, 0, 0) \times (1, 0, 0)_{12}$$

[S2] lacks the $B \cdot B^{12}$ term so is not SARMA. $(1, 0, 0) \times (1, 0, 0)_{12}$

[S2] could be SARMA $(12, 0, 0) \times (0, 0, 0)_{12}$
i.e. ARMA $(12, 0, 0)$ with coefficient constraint.

Question 6.2. Why do we assume a multiplicative structure in the SARMA model, [S1]? What theoretical and practical advantages (or disadvantages) arise from requiring that an ARMA model for seasonal behavior has polynomials that can be factored as a product of a monthly polynomial and an annual polynomial?

The multiplicative form is convenient for data analysis: it separates annual and monthly effects, with simple and interpretable mts to the ARMA polynomials. There is no scientific reason to prefer [S3] to [S2].

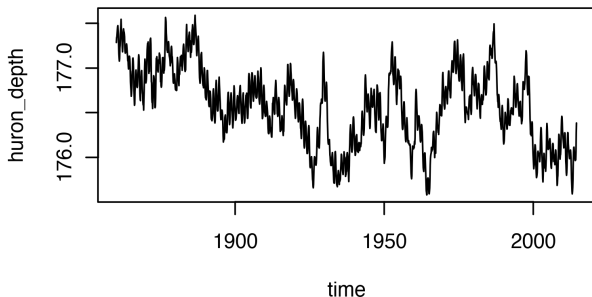
Fitting a SARMA model

We fit the full, monthly, version of the Lake Huron depth data described earlier.

```
head(dat,3)
```

Date	Average	year	month
1860-01-01	177.285	1860	1
1860-02-01	177.339	1860	2
1860-03-01	177.349	1860	3

```
huron_depth <- dat$Average  
time <- dat$year + dat$month/12  
# Note: this treats December 2011 as time 2012.0  
plot(huron_depth~time,type="l")
```



Based on our previous analysis, we try fitting AR(1) for the annual polynomial. Let's try ARMA(1,1) for the monthly part. In other words, we consider the model

$$(1 - \Phi_1 B^{12})(1 - \phi_1 B)Y_n = (1 + \psi_1 B)\epsilon_n. \quad (1)$$

```
huron_sarma11x10 <- arima(huron_depth,  
  order=c(1,0,1),  
  seasonal=list(order=c(1,0,0),period=12)  
)
```



```
huron_sarma11x10
```

Call:

```
arima(x = huron_depth, order = c(1, 0, 1), seasonal = list(order = c(1, 0, 0),  
  period = 12))
```

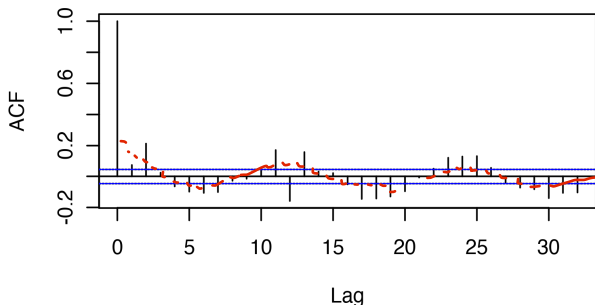
Coefficients:

	ar1	ma1	sar1	intercept
	0.9641	0.3782	0.5104	176.5714
s.e.	0.0063	0.0203	0.0218	0.0909

sigma² estimated as 0.002592: log likelihood = 2884.36, aic = -5758.72

Residual analysis

- Residual analysis is similar to non-seasonal ARMA models.
- We look for residual correlations at lags corresponding to multiples of the period (here, 12, 24, 36, ...) for misspecified annual dependence.



Question 6.3. What do you conclude from this residual analysis? What would you do next? *The oscillatory component of the residuals suggests AR(2) for the residuals, suggesting $(3, 0, 1) \times (1, 0, 0)_{12}$. This heuristic suggests a model to fit, but is not a substitute for AIC or LRT.*

ARMA models for differenced data

- Applying a difference operation to the data can make it look more stationary and therefore more appropriate for ARMA modeling.
- This can be viewed as a **transformation to stationarity**
- We can transform the data $y_{1:N}^*$ to $z_{2:N}$

$$z_n = \Delta y_n^* = y_n^* - y_{n-1}^*. \quad (2)$$

- Then, an ARMA(p,q) model $Z_{2:N}$ for the differenced data $z_{2:N}$ is called an **integrated autoregressive moving average** model for $y_{1:N}^*$ and is written as ARIMA(p,1,q).
- Formally, the ARIMA(p,d,q) model with intercept μ for $Y_{1:N}$ is

$$[S4] \quad \phi(B)((1-B)^d Y_n - \mu) = \psi(B) \epsilon_n,$$

where $\{\epsilon_n\}$ is a white noise process; $\phi(x)$ and $\psi(x)$ are ARMA polynomials.

- It is unusual to fit an ARIMA model with $d > 1$.

- We see that an ARIMA(p,1,q) model is almost a special case of an ARMA(p+1,q) model with a **unit root** to the AR(p+1) polynomial.

Question 6.4. Why “almost” not “exactly” in the previous statement?

Note the differences in the treatment of μ .

For ARIMA, μ is the mean of the differenced process.

For example, if $d=1$, we have $E[(1-B)Y_n] = \mu$

$$E[Y_{n+1}] = \mu + E[Y_n]$$

corresponding to a linear trend model.

$$E[Y_n] = A + \mu n,$$

Two reasons to fit an ARIMA(p,d,q) model with $d > 0$

1. You may really think that modeling the differences is a natural approach for your data. The S&P 500 stock market index analysis in Chapter 3 is an example of this, as long as you remember to first apply a logarithmic transform to the data.
2. Differencing often makes data look “more stationary” and perhaps it will then look stationary enough to justify applying the ARMA machinery.
 - We should be cautious about this second reason. It can lead to poor model specifications and hence poor forecasts or other conclusions.
 - The second reason was more compelling in the 1970s and 1980s. Limited computing power resulted in limited alternatives, so it was practical to force as many data analyses as possible into the ARMA framework and use method of moments estimators.

Practical advice on using ARIMA models

- ARIMA analysis is relatively simple to do. It has been a foundational component of time series analysis since the publication of the influential book “Time Series Analysis” (Box and Jenkins, 1970) which developed and popularized ARIMA modeling.
- A practical approach is:
 1. Do a competent ARIMA analysis.
 2. Identify potential limitations in this analysis and remedy them using more advanced methods.
 3. Assess whether you have in fact learned anything from (2) that goes beyond (1).

The $\text{SARIMA}(p, d, q) \times (P, D, Q)$ model

Combining integrated ARMA models with seasonality, we can write a general $\text{SARIMA}(p, d, q) \times (P, D, Q)_{12}$ model for nonstationary monthly data, given by

$$[S5] \quad \phi(B)\Phi(B^{12})((1-B)^d(1-B^{12})^DY_n - \mu) = \psi(B)\Psi(B^{12})\epsilon_n,$$

where $\{\epsilon_n\}$ is a white noise process, the intercept μ is the mean of the differenced process $\{(1-B)^d(1-B^{12})^DY_n\}$, and we have ARMA polynomials $\phi(x)$, $\Phi(x)$, $\psi(x)$, $\Psi(x)$ as in model [S1].

- The $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$ model has often been used for forecasting monthly time series in economics and business. It is sometimes called the **airline model** after a data analysis by Box and Jenkins (1970).

Modeling trend with ARMA noise

- A general **signal plus noise** model is

$$[S6] \quad Y_n = \mu_n + \eta_n,$$

where $\{\eta_n\}$ is a stationary, mean zero stochastic process, and μ_n is the mean function.

- If, in addition, $\{\eta_n\}$ is uncorrelated, then we have a **signal plus white noise** model. The usual linear trend regression model fitted by least squares in Chapter 2 corresponds to a signal plus white noise model.
- We can say **signal plus colored noise** if we wish to emphasize that we're not assuming white noise.
- Here, **signal** and **trend** are used interchangeably. In other words, we are assuming a deterministic signal.
- At this point, it is natural for us to consider a signal plus ARMA(p,q) noise model, where $\{\eta_n\}$ is a stationary, causal, invertible ARMA(p,q) process with mean zero.
- As well as the $p + q + 1$ parameters in the ARMA(p,q) model, there will usually be unknown parameters in the mean function.

Linear regression with ARMA errors

- When the mean function (also known as the trend) has a linear specification,

$$\mu_n = \sum_{k=1}^K Z_{n,k} \beta_k, \quad (3)$$

the signal plus ARMA noise model is known as **linear regression with ARMA errors**.

- Writing Y for a column vector of $Y_{1:N}$, μ for a column vector of $\mu_{1:N}$, η for a column vector of $\eta_{1:N}$, and Z for the $N \times K$ matrix with (n, k) entry $Z_{n,k}$, we have a general linear regression model with correlated ARMA errors,

$$Y = Z\beta + \eta. \quad (4)$$

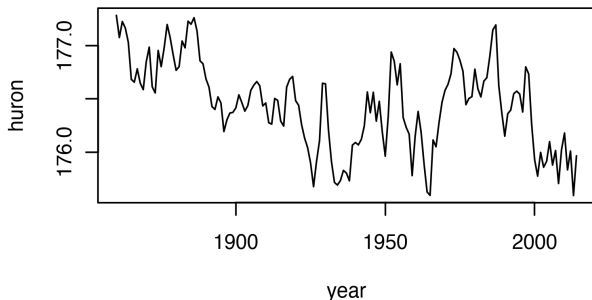
- From (4), $Y - Z\beta$ is ARMA so likelihood evaluation and numerical maximization can build on ARMA methods.

Inference for the linear regression model with ARMA errors

- Maximum likelihood estimation of $\theta = (\phi_{1:p}, \psi_{1:q}, \sigma^2, \beta)$ is a nonlinear optimization problem.
- Fortunately, `arima` in R can do it for us.
- As usual, we should look out for signs of numerical problems.
- Data analysis for a linear regression with ARMA errors model, using the framework of likelihood-based inference, is procedurally similar to fitting an ARMA model.
- This is a powerful technique, since the covariate matrix Z can include other time series. We can evaluate associations between different time series.
- With appropriate care (since **association is not causation**) we can draw inferences about mechanistic relationships between dynamic processes.

Evidence for systematic trend in Lake Huron depth?

Let's go back to annual data, say the January depth, to avoid seasonality.



- Visually, there seems some evidence for a decreasing trend, but there are also considerable fluctuations.
- Let's test for a trend, using a regression model with Gaussian AR(1) errors. We have previously found that this is a reasonable model for these data.
- First, for comparison, we fit a null model with no trend.

```
fit0 <- arima(huron, order=c(1,0,0))
```

Call:

```
arima(x = huron, order = c(1, 0, 0))
```

Coefficients:

	ar1	intercept
	0.8694	176.4588
s.e.	0.0407	0.1234

sigma^2 estimated as 0.04368: log likelihood = 22, aic = -38

- R drops trailing zeros which are useful for showing significant figures of accuracy. Here, this affects the likelihood. A solution is as follows:

```
as.character(22.00)
```

```
[1] "22"
```

```
library(broman)
```

```
myround(22.00)
```

```
[1] "22.0"
```

- We compare `fit0` with a linear trend model, coded as `fit1`.
- The covariate is included via the `xreg` argument.

```
fit1 <- arima(huron,order=c(1,0,0),xreg=year)
```

Call:

```
arima(x = huron, order = c(1, 0, 0), xreg = year)
```

Coefficients:

	ar1	intercept	year
	0.8240	186.0146	-0.0049
s.e.	0.0451	3.7417	0.0019

sigma² estimated as 0.0423: log likelihood = 24.62, aic = -41.25

Setting up a formal hypothesis test

- To talk formally about these results, we must down a model and some hypotheses.
- Writing the data as $y_{1:N}^*$, collected at years $t_{1:N}$, the model we have fitted is

$$(1 - \phi_1 B)(Y_n - \mu - \beta t_n) = \epsilon_n, \quad (5)$$

where $\{\epsilon_n\}$ is Gaussian white noise with variance σ^2 . Our null model is

$$H^{(0)} : \beta = 0, \quad (6)$$

and our alternative hypothesis is

$$H^{(1)} : \beta \neq 0. \quad (7)$$

Question 6.5. How do we test $H^{(0)}$ against $H^{(1)}$?

- Construct two different tests using the R output above.
- Which test do you think is more accurate, and why?

1. Z-test on the slope parameter, β .

test statistic: $\frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{-0.0049}{0.0019}$, so $|\frac{\hat{\beta}}{SE(\hat{\beta})}| > 1.96$, so

we reject H_0 at 5% level.

2. Likelihood ratio test (LRT).

$\Delta = \text{difference in log likelihood} = 24.62 - 22.00 = 2.62$

since $\Delta = 2.62 > 1.92$, we reject H_0 at 5% level

From Ch 5, we may trust the LRT more than the Fisher SE.

Question 6.6. How would you check whether your preferred test is indeed better? What other supplementary analysis could you do to strengthen your conclusions?

A simulation study (or bootstrap) can explore if either or both approximate tests (Fisher normal approximation and Wilks CRT) are accurate for this sample size.

Residual analysis and trying alternative models can reveal if there is room for improvement in the model.

Further reading

- Section 3.9 of Shumway and Stoffer (2017) discusses SARIMA modeling.
- Section 3.8 of Shumway and Stoffer (2017) introduces regression with ARMA errors.

License, acknowledgments, and links

- Licensed under the [Creative Commons Attribution-NonCommercial license](#). Please share and remix non-commercially, mentioning its origin.
- The materials builds on [previous courses](#).
- Compiled on February 12, 2021 using R version 4.0.3.



[Back to course homepage](#)

References

- Box GEP, Jenkins GM (1970). *Time Series Analysis: Forecasting and Control*. First edition. Holden-Day, San Francisco.
- Shumway RH, Stoffer DS (2017). *Time Series Analysis and its Applications: With R Examples*. Springer. URL <http://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf>.