

Modeling and Analysis of Time Series Data
Chapter 6: Extending the ARMA model:
Seasonality, integration and trend
STATS 531, Winter 2026

Edward Ionides

Seasonal autoregressive moving average (SARMA) models

- ▶ A general SARMA(p, q) \times (P, Q)₁₂ model for monthly data is

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

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

$$\begin{aligned}\mu &= \text{E}[Y_n] \\ \phi(x) &= 1 - \phi_1x - \dots - \phi_px^p, \\ \theta(x) &= 1 + \theta_1x + \dots + \theta_qx^q, \\ \Phi(x) &= 1 - \Phi_1x - \dots - \Phi_Px^P, \\ \Theta(x) &= 1 + \Theta_1x + \dots + \Theta_Qx^Q.\end{aligned}$$

- ▶ SARMA is a special case of ARMA, where the AR and MA polynomials are factored into a **monthly** polynomial in B and an **annual (or 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 SARMA(p, q) \times (P, Q)₁₂ model, 12 is called the **period**.
- ▶ One could write a SARMA model for a period other than 12.
- ▶ A SARMA(p, q) \times (P, Q)₄ model could be appropriate for quarterly data.
- ▶ In principle, a 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 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. Which of [S2] and/or [S3] is a SARMA model?

$$[S3]: \quad (1 - 0.5B)(1 - 0.25B^{12})Y_n = \epsilon_n$$

comes from the
multiplicative
assumption of SARMA.

[S2]: cannot be written in product form.

Question. 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?

This is a matter of convenience, not science.

Mathematically, the separation is elegant. Does it typically hold for data?

The SARMA assumption greatly reduces the number of models to look at.

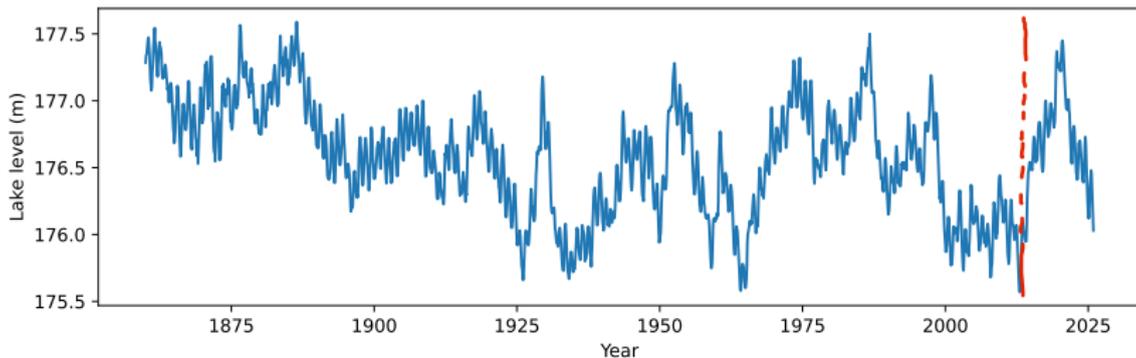
Fitting a SARMA model

We fit a monthly version of the Lake Huron depth data.

```
dat = pd.read_csv("huron_level.csv", comment='#')
dat.columns = dat.columns.str.strip()
print(dat.iloc[:3, :6])
```

	Year	Jan	Feb	Mar	Apr	May
0	1860	177.285	177.339	177.349	177.388	177.425
1	1861	177.077	177.105	177.224	177.254	177.382
2	1862	177.227	177.181	177.272	177.321	177.397

```
# Reshape data to monthly time series
huron_level = dat.iloc[:, 1:13].values.flatten()
years = np.repeat(dat['Year'].values, 12)
months = np.tile(np.arange(12), len(dat))
time = years + months/12
```



Based on our previous analysis, we try fitting AR(1) for the annual polynomial. We try ARMA(1,1) for the monthly part, giving

$$(1 - \Phi_1 B^{12})(1 - \phi_1 B)(Y_n - \mu) = (1 + \theta_1 B)\epsilon_n. \quad (1)$$

- ▶ As discussed earlier, we analyze data only up to 2014, shown by a red line on the plot.

```

huron_level_sub = huron_level[time < 2014.99]
time_sub = time[time < 2014.99]

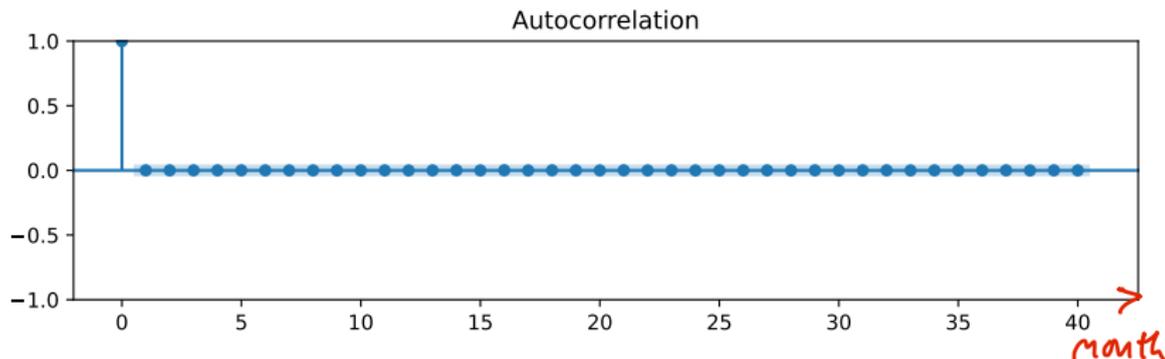
# Fit SARMA(1,1)x(1,0)_12 model
huron_sarma11x10 = SARIMAX(huron_level_sub,
    order=(1, 0, 1),
    seasonal_order=(1, 0, 0, 12)).fit()

```

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	1.0000	4.83e-06	2.07e+05	0.000	1.000	1.000
ma.L1	0.7868	1.21e-06	6.5e+05	0.000	0.787	0.787
ar.S.L12	1.0000	3.66e-06	2.73e+05	0.000	1.000	1.000
sigma2	0.0050	3.56e-07	1.41e+04	0.000	0.005	0.005

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. What do you conclude from this residual analysis? What would you do next?

These look good! Too good to be true?

YES! SEE UPDATED SLIDES

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 $z_{2:N}$

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

Definition: 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

$$[\text{S4}] \quad \phi(B)[(1 - B)^d Y_n - \mu] = \theta(B) \epsilon_n,$$

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

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

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

$$[54]: \phi(B) \left[(1-B)^d Y_n - \mu \right] = \theta(B) \epsilon_n$$

similar ARMA(p+1,q), with a unit root.

$$\phi(B) (1-B)^d (Y_n - \mu) = \theta(B) \epsilon_n$$

there's a small difference in treatment of μ .

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. *1st difference removes a linear trend.
2nd difference removes a quadratic trend.*
- ▶ 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” [1] 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 SARIMA(p, d, q) \times (P, D, Q) model

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

$$[S5] \quad \phi(B)\Phi(B^{12})[(1-B)^d(1-B^{12})^D Y_n - \mu] = \theta(B)\Theta(B^{12})\epsilon_n,$$

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

- ▶ The 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).

Trend estimation: regression with ARMA errors

Definition: 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.
- ▶ A **signal plus ARMA(p,q) noise model** occurs when $\{\eta_n\}$ is a stationary, causal, invertible, mean zero ARMA(p,q) process.

- ▶ As well as the $p + q + 1$ parameters in the ARMA(p,q) noise model, there will usually be unknown parameters in the mean function.
- ▶ When the mean (also called trend) has a linear specification,

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

the model is **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,

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

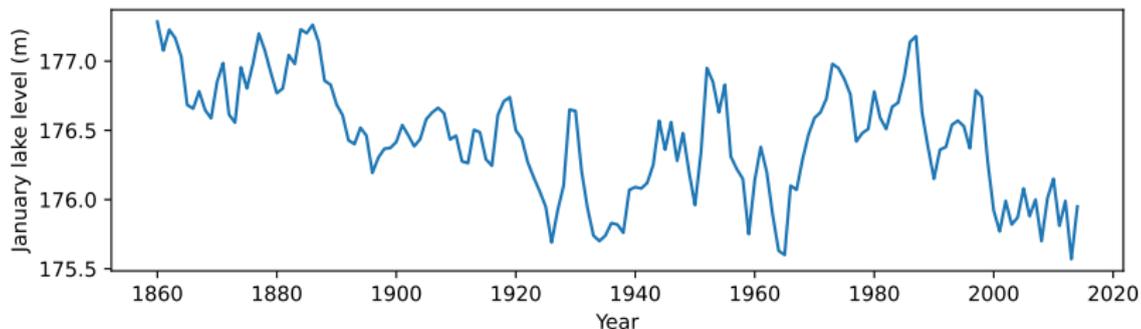
- ▶ From this equation, $Y - Z\beta$ is ARMA so likelihood evaluation and numerical maximization can build on ARMA methods.

Inference for the linear regression model with ARMA errors

- ▶ Maximum likelihood estimation of $\theta = (\phi_{1:p}, \theta_{1:q}, \sigma^2, \beta)$ is a nonlinear optimization problem.
- ▶ Fortunately, Python's SARIMAX can do it for us using the `exog` argument.
- ▶ 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 level?

We return to annual data, say the January level, 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.

```

from io import StringIO
fit0 = SARIMAX(huron_jan, order=(1, 0, 0),
               trend='c').fit()
params_csv = fit0.summary().tables[1].as_csv()
pd.read_csv(StringIO(params_csv))

```

		coef	std err	z	P> z	[0.025	0.975]
0	intercept	26.2506	6.887	3.812	0.0	12.753	39.748
1	ar.L1	0.8512	0.039	21.814	0.0	0.775	0.928
2	sigma2	0.0440	0.005	8.439	0.0	0.034	0.054

AIC: -37.06

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

```
fit1 = SARIMAX(huron_jan, order=(1, 0, 0),
               exog=year_jan).fit()
```

SE($\hat{\beta}$) from empirical Fisher Info

		coef	std err	z	P> z	[0.025	0.975]
0	x1	-0.0055	0.243	-0.023	0.982	-0.482	0.471
1	ar.L1	1.0000	0.001	780.962	0.000	0.997	1.003
2	sigma2	0.0466	0.005	8.689	0.000	0.036	0.057

NEED trend='c', but even then the optimization is inconsistent.

AIC: -15.71

SEE UPDATED SLIDES

Setting up a formal hypothesis test

- ▶ To talk formally about these results, we must set 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

$$(1 - \phi_1 B)(Y_n - \underbrace{\mu - \beta t_n}_{\text{linear trend}}) = \epsilon_n, \quad (5)$$

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. How do we test $H^{(0)}$ against $H^{(1)}$? Construct two different tests using the Python output above. Which test do you think is more accurate, and why?

1. Coefficient test: test statistic $\frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{-0.0055}{0.243} = -0.023$
P-value ≈ 1 , but note we don't trust this analysis,

2. Likelihood ratio test.

$$2\Delta \log \text{lik} = \Delta \text{AIC} + 2\Delta \text{d.f.} = -37.06 - (-15.71) + 2 = -19.35$$

compared with χ^2 . This is negative \Rightarrow mathematically impossible \Rightarrow there's an error.

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

We might expect LRT to be more accurate than F.T. due to greater robustness to asymptotics.

Here, the LRT makes clear the error in the nesting.

SEE UPDATED SLIDES

Testing for trend vs unit root: The ADF test

Definition: A time series model has a **unit root** if its first difference is stationary.

- ▶ For linear time series models, this corresponds to a $(1 - L)$ factor in the AR specification, and hence an AR polynomial root of 1.
- ▶ A unit root corresponds to a “stochastic trend” (formally an oxymoron) described by a random walk.

The **augmented Dickey-Fuller (ADF)** test has this null hypothesis. *[there is a unit root].*

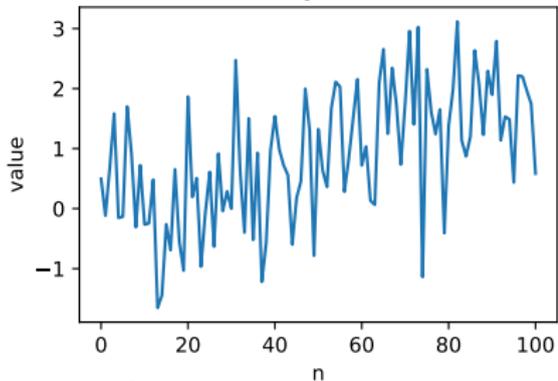
- ▶ The ADF test is sometimes called a test for “stationarity”. However, non-stationary models may remain non-stationary after any amount of differencing. e.g., any model with non-polynomial trend.
- ▶ Assessing evidence for a trend is better than applying an ADF test for a unit root if a model with trend is statistically (or

Question. Which of y_1 , y_2 , y_3 , y_4 do you think will be considered non-stationary by a unit root test?

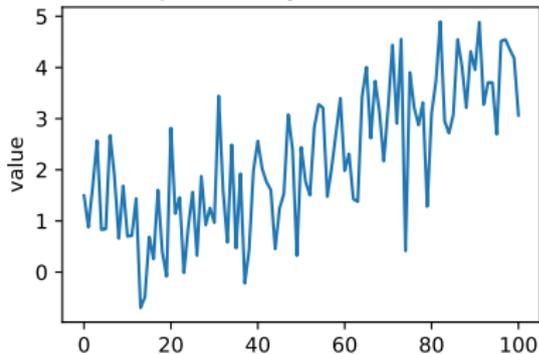
```
np.random.seed(42)
n = np.arange(101)
epsilon = np.random.normal(0, 1, len(n))
y1 = 2*n/100 + epsilon
y2 = np.exp(1.5*n/100) + epsilon
y3 = np.sin(2*np.pi*n/200)
y4 = n*(100-n)/2500
```

"ADF-stationary" if ADF null hypothesis is rejected.

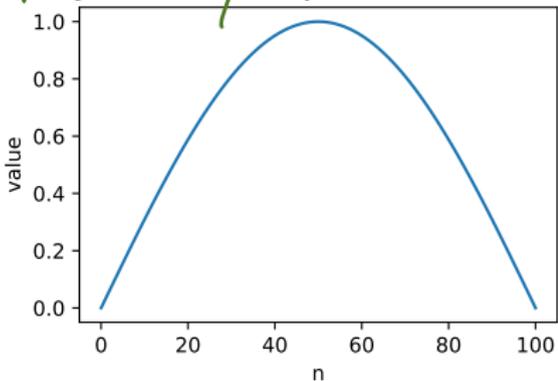
ADF-unit root y_1



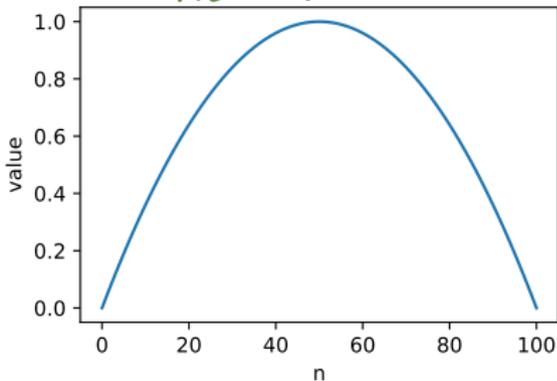
ADF-unit root y_2



ADF-stationary y_3



~~ADF-stationary~~ ADF-unit root y_4



Result of `adfuller()`

data: y1

Dickey-Fuller = -1.7834, Lag order = 4, p-value = 0.3887

Alternative hypothesis: stationary

data: y2

Dickey-Fuller = -0.8051, Lag order = 4, p-value = 0.8176

Alternative hypothesis: stationary

data: y3

Dickey-Fuller = -94217355042311.1562, Lag order = 11, p-value = 0.0000

Alternative hypothesis: stationary

data: y4

Dickey-Fuller = 0.0861, Lag order = 10, p-value = 0.9651

Alternative hypothesis: stationary

Note: these p-values are very different from
R: `adf.test()` reported in 531W25.

See updated slides

When to use the ADF test

- ▶ If you are interested in a null hypothesis of a random walk against an alternative of a stationary ARMA model, then this test is applicable.
- ▶ Generalizations of ADF claim to identify random walks on top of a nonlinear trend, but distinguishing these two options is hard and beyond our scope.
- ▶ Most midterm projects benefit from avoiding the ADF test and focusing on a direct investigation of the evidence for trend or other non-stationary behavior.

Further reading

- ▶ Section 3.9 of Shumway and Stoffer [2] discusses SARIMA modeling.
- ▶ Section 3.8 of Shumway and Stoffer [2] introduces regression with ARMA errors.

Acknowledgments

- ▶ Compiled on January 24, 2026 using Python.
- ▶ Licensed under the Creative Commons Attribution-NonCommercial license. Please share and remix non-commercially, mentioning its origin.
- ▶ We acknowledge previous versions of this course.

References I



G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. First. San Francisco: Holden-Day, 1970.



Robert H Shumway and David S Stoffer. *Time Series Analysis and its Applications: With R Examples*. 4th. Springer, 2017.