Time Series Analysis

ECE 6563, Electrical and Computer Engineering
University of New Brunswick

Recap

In the last section, we looked at using our models to predict/forecast

- Model fitting, Information Criteria
  MLE, AIC, BIC, HQIC
- Forecasting
  Naïve, Drift, Exponential, ARIMA, Prediction Intervals
- Evaluation
  Fitting, Hold-out, Rolling Window, MAPE, RMSE, R^2
- Anomaly Detection
  Static (statistical, classification) vs Residuals

In this section, we'll

- Revisit Regression
- Multivariate Time Series

Regression

When we first looked at regression, used it fit a trend line from for a time series

- We used a linear (or non-linear) weighting of the independent variable, t, to predict the value of x_t
- This allowed us to determine a trend by looking at the slope
- We could also use it as a simple model for prediction

Parametric Trend Estimation

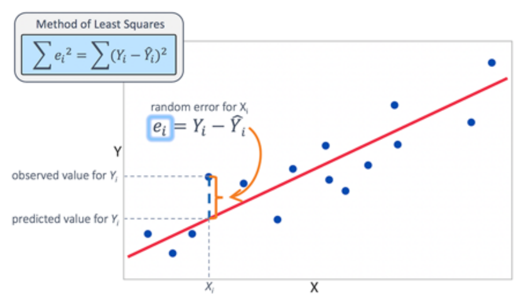
One approach to estimating a trend line is to assume that it can be modeled using a finite set of parameters

- Fitting a linear or quadratic curve using regression
- Use least squares regression (think “find the line of best fit”)
- Use time (t) as the independent variable.
- Linear Example:

$$X_t = \beta_0 + \beta_1 t$$

Intercept

Slope

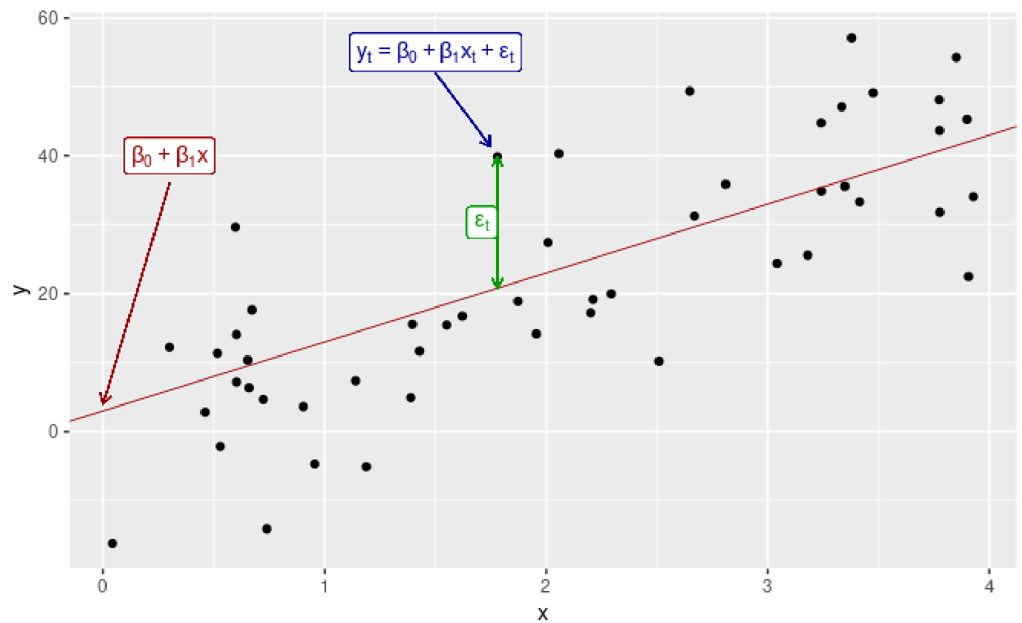


Regression

When we first looked at regression, used it fit a trend line from for a time series

- We used a linear (or non-linear) weighting of the independent variable, t, to predict the value of x_t
- This allowed us to determine a trend by looking at the slope
- We could also use it as a simple model for prediction

- We then removed the trend, to allow us to focus on modeling the residuals, ε_t



Multi-variate Time Series

But what if we wanted to include more information in our models

- Two or more *predictor* variables – features that describe the problem
- In our univariate time series case, predictor was time, *t*

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t$$

- But we also created a model that combined trend and seasonality that incorporated different *functions* of time

$$x_t = (B_0 + B_1 t + B_2 t^2) + \left[\sum_{j=1}^{k} \alpha_j \cos\left(2\pi f_j t\right) + \gamma_j \sin\left(2\pi f_j t\right)\right] + \varepsilon_t$$

More generally, we can write a multivariate regression model as

$$x_t = \beta_0 + \beta_1 f_{1t} + \beta_2 f_{2t} \ldots \beta_k f_{kt} + \varepsilon_t$$

where: x_t is the variable being forecast, f_{Nt} are the *k* predictor variables
β_N are the coefficients that determine the *marginal* effects for each predictor (i.e. the effect of that one, considering all other effects)

Multi-variate Time Series

These 'other' predictors can be either exogenous or endogenous variables

Exogenous:

- A variable that affects our model, but isn't affected by the other variables in the model
- For example, weather, pest levels, and the availability of seed are all exogenous to crop production
- Take as "given" or fixed, not affected or explained by the model

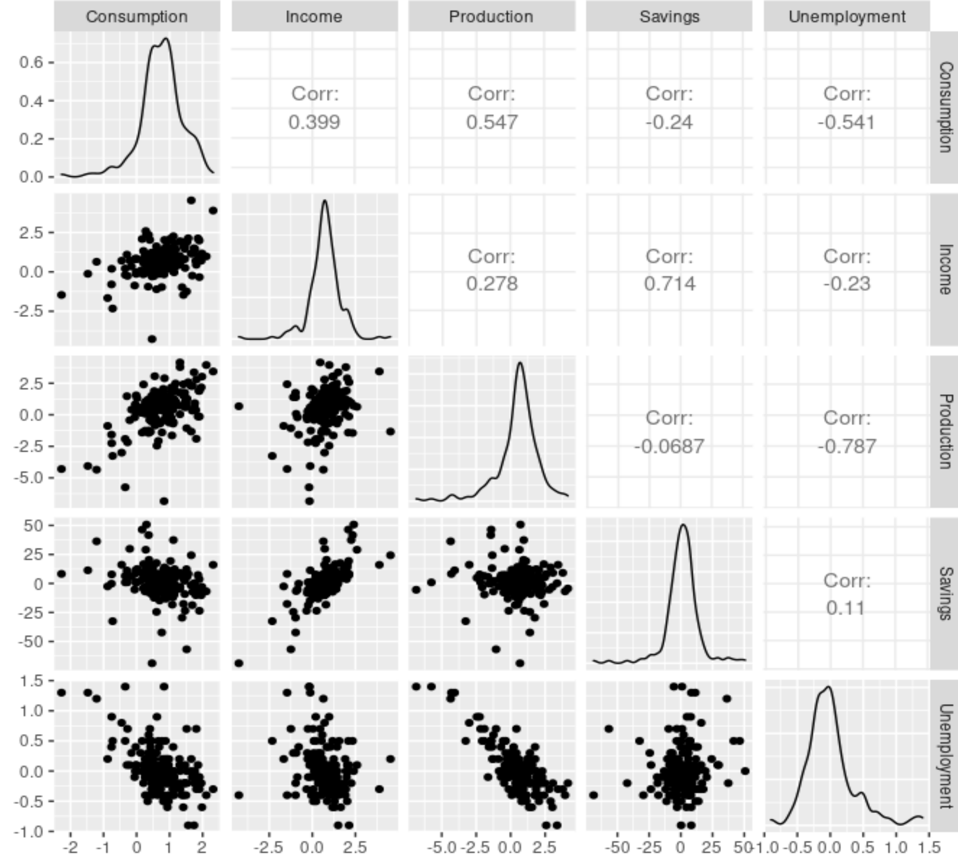
Endogenous:

- A variable that is affected by the other variables in the model
- Usually, the output of the model
- For example, crop production is affected by sunlight, so it is endogenous
- We want to be careful to avoid having endogenous variables as predictors in our models, because of interactions with other variables

Multi-variate Time Series

Scatterplots with predictors (exogenous variables)

- Plot the variable to be forecast (here consumption) against potential predictors



- Positive relationship with income and industrial production
- Negative relationship with savings and unemployment.
- The strength of the relationships are shown by the correlation coefficients
- Others show the relationships between the predictors.

https://otexts.com/fpp2/regression-intro.html

Multi-variate Time Series

Generally, when we use linear regression, we assume that the remaining error:

- Is zero mean (otherwise the model is biased)
- Is unrelated to the predictor values (otherwise, or fit is off)
- It is not autocorrelated
- Useful to assume it is normally distributed with constant variance

$$x_t = \beta_0 + \beta_1 f_{1t} + \beta_2 f_{2t} \ldots \beta_k f_{kt} + \varepsilon_t$$

- This allows us to fit our model by finding the coefficients that give us the Least Squared Errors:

$$\sum_{t=1}^{T} {\varepsilon_t}^2 = \sum_{t=1}^{T} (x_t - \beta_0 + \beta_1 f_{1t} + \beta_2 f_{2t} \ldots \beta_k f_{kt})^2$$

- When applied to time series, however, this “not autocorrelated” assumption is likely violated

ARIMAX Models

A logical extension of multivariate regression for time series, is then to combine it with ARIMA models

- We can simply extend our ARIMA (or SARIMA) models by adding the regression terms (or extend our regression model with ARIMA terms)
- For example, if we add one covariate term:

$$x_t = \beta_0 + \beta_1 f_{1t} + \phi_1 x_{t-1} + \cdots \phi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots \theta_q \varepsilon_{t-q} + \varepsilon_t$$

ARIMAX Models

A logical extension of multivariate regression for time series, is then to combine it with ARIMA models

- We can simply extend our ARIMA (or SARIMA) models by adding the regression terms (or extend our regression model with ARIMA terms)
- For example, if we add one covariate term:

$$x_t = \underbrace{\beta_0 + \beta_1 f_{1t}}_{\text{Reg. terms}} + \underbrace{\phi_1 x_{t-1} + \cdots \phi_p x_{t-p}}_{\text{AR terms}} + \underbrace{\theta_1 \varepsilon_{t-1} + \cdots \theta_q \varepsilon_{t-q}}_{\text{MA terms}} + \underbrace{\varepsilon_t}_{\text{IID}}$$

- We can easily extend this to include more f_{Nt} terms to include additional exogenous/predictor variables
- A drawback of this approach is that the β_1 coefficient term loses some of its interpretability (as compared to regular regression)

    β_N is no longer only the effect of f_{Nt} on x_t because there are lagged (x_{t-k}) terms on the right side now too.

ARIMAX Models

For this reason, instead of directly using:

$$x_t = \beta_0 + \beta_1 f_{1t} + \phi_1 x_{t-1} + \cdots \phi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots \theta_q \varepsilon_{t-q} + \varepsilon_t$$

ARIMAX models are often written as

$$x_t = \beta_0 + \beta_1 f_{1t} + \cdots \beta_N f_{Nt} + n_t$$

where the n_t term is denoted separately so that the recursions operate only on it, and not the exogenous variable terms.

$$n_t = \phi_1 n_{t-1} + \cdots \phi_p n_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots \theta_q \varepsilon_{t-q} + \varepsilon_t$$

In this way, n_t is considered to be an error series which may follow an ARIMA model.

These forms of model, which include both the regression and ARIMA models are sometimes referred to as dynamic regression models.

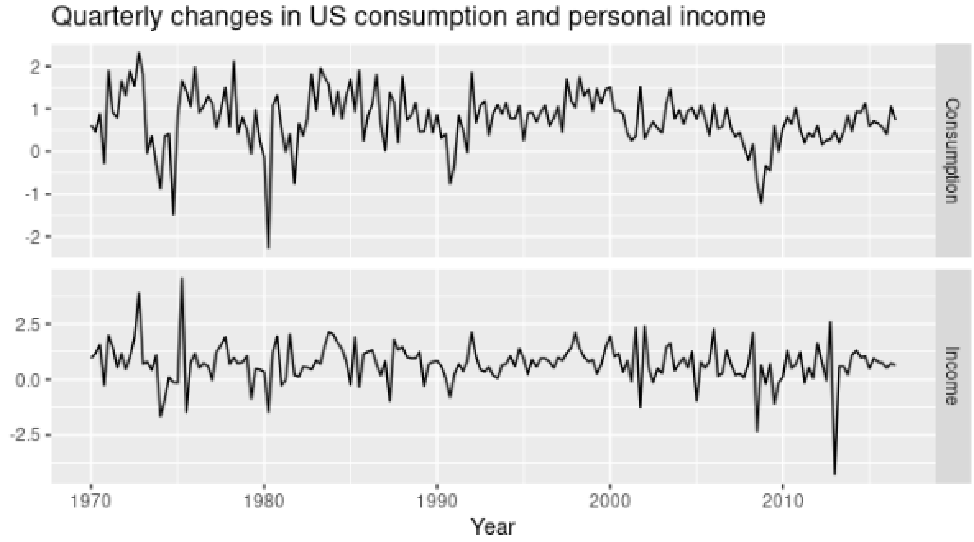
ARIMAX Models

Although you can usually solve for the coefficients in regression using LSE, you cannot in ARIMAX because of the inclusion of the MA terms.

- Requires solution using MLE

We can use LSE if the q order is 0, but you must minimize the ε_t term (not the full n_t, as we just discussed). Otherwise

- Some of information would be missing – the β_N terms wouldn't be optimal
- Information criteria guides would be incorrect
- Statistical tests tend to be overconfident/overfit

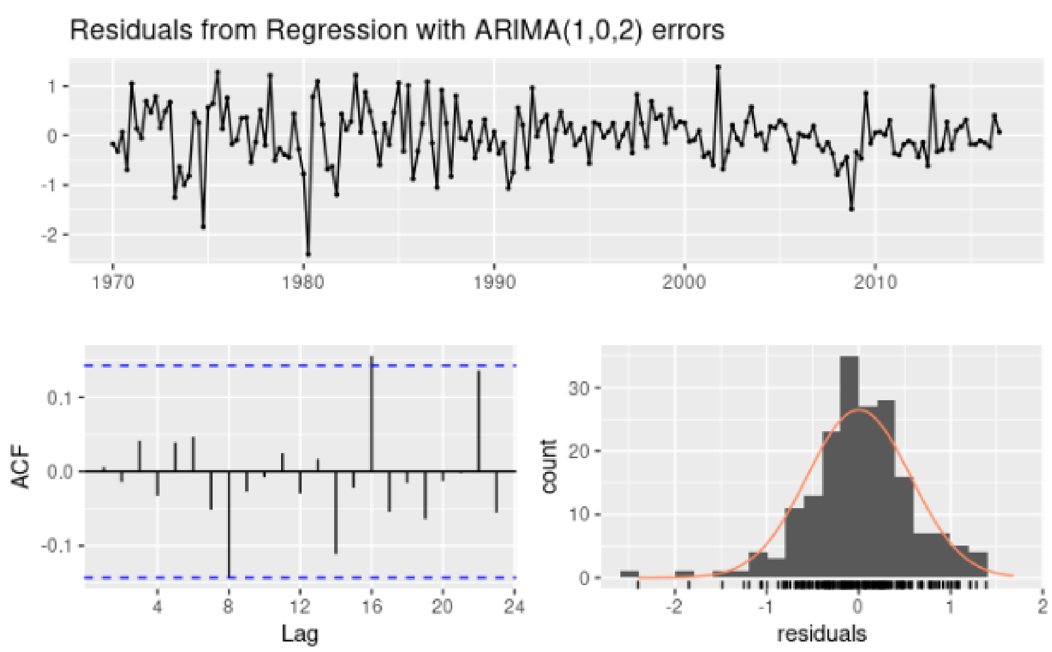


https://otexts.com/fpp2/regarima.html

ARIMAX Models

Just like ARIMA, we need all variables to first be stationary

- To maintain relationships between variables, it is common to difference all variables by the same order/amount
- We can then model the full system, and evaluate the residuals of the ε_t




```
In [ ]: model_ar_1_i_1_ma_1_Xspx = ARIMA(df.market_value, exog = df.spx, order=(1,1,1))
```

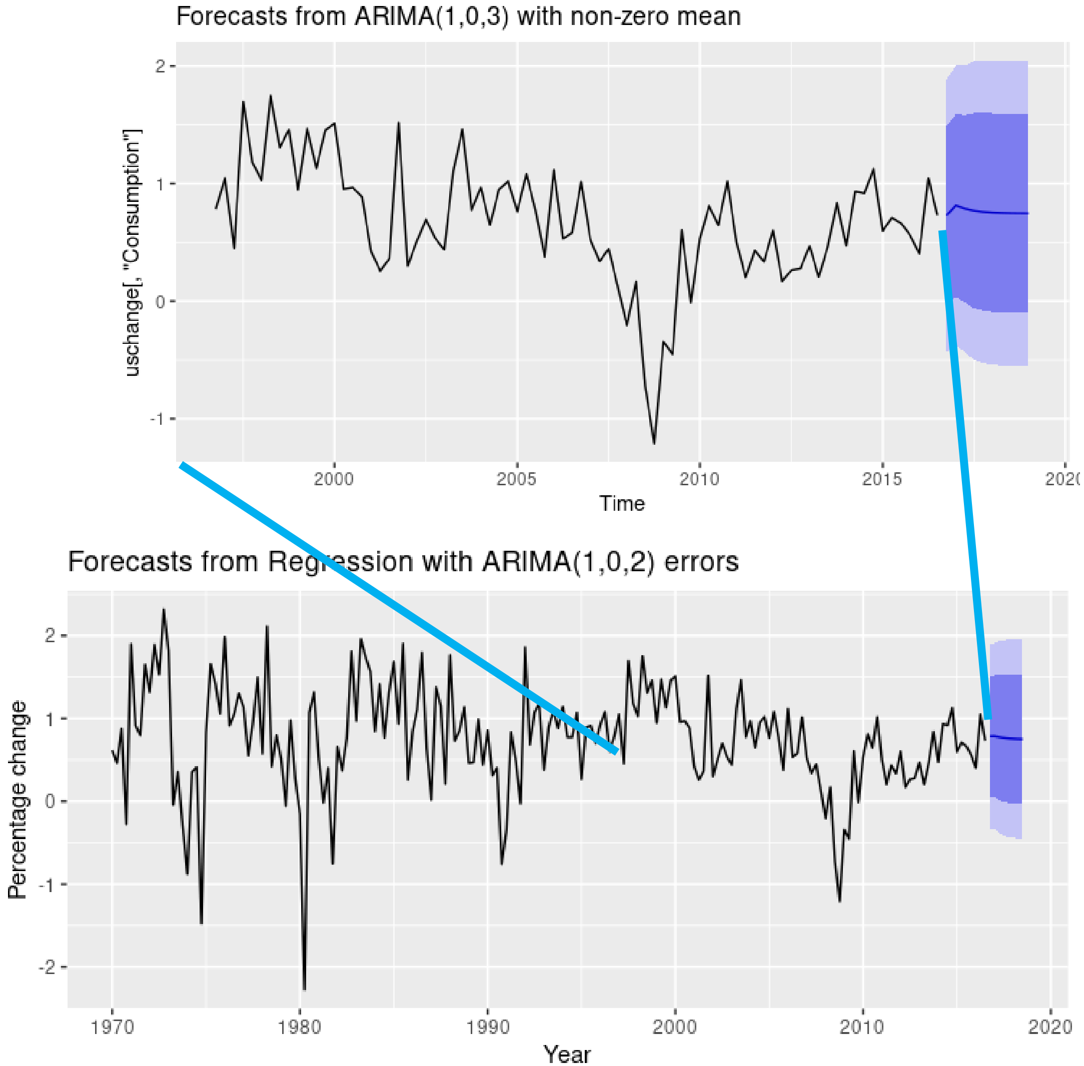
ARIMAX Models

ARIMA model alone:

- Prediction of consumption based on consumption alone

ARIMAX model:

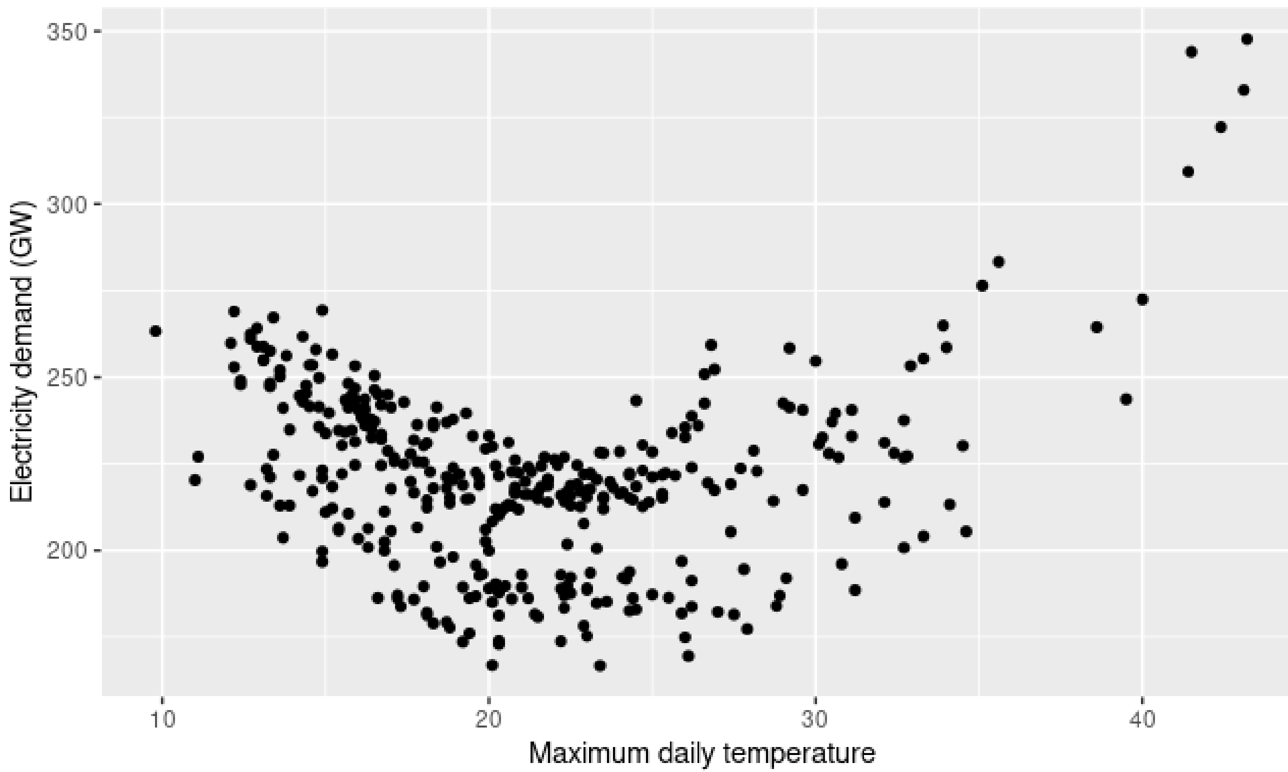
- includes income as an additional predictor
- Smaller prediction intervals



ARIMAX Models

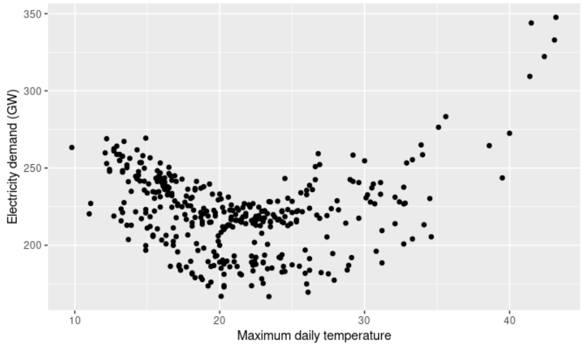
Example: Electricity Demand

- The temperature can be used to predict electricity demand
- Too hot, we use A/C; too cold, we use heat



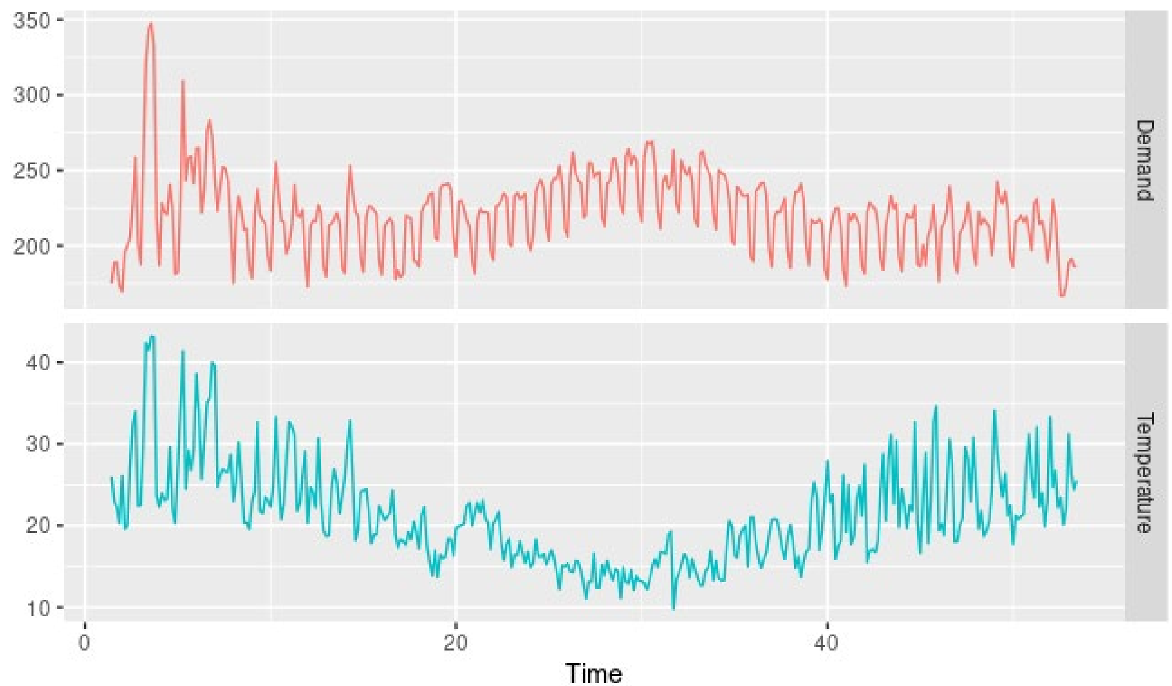
https://otexts.com/fpp2/forecasting.html

ARIMAX Models



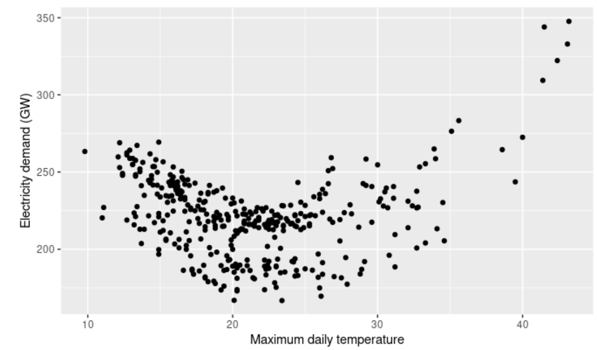
We can fit a model that includes

- a quadratic regression for the temperature influence
- an ARIMA model for the remaining errors, with 7-day seasonality reflecting the work week



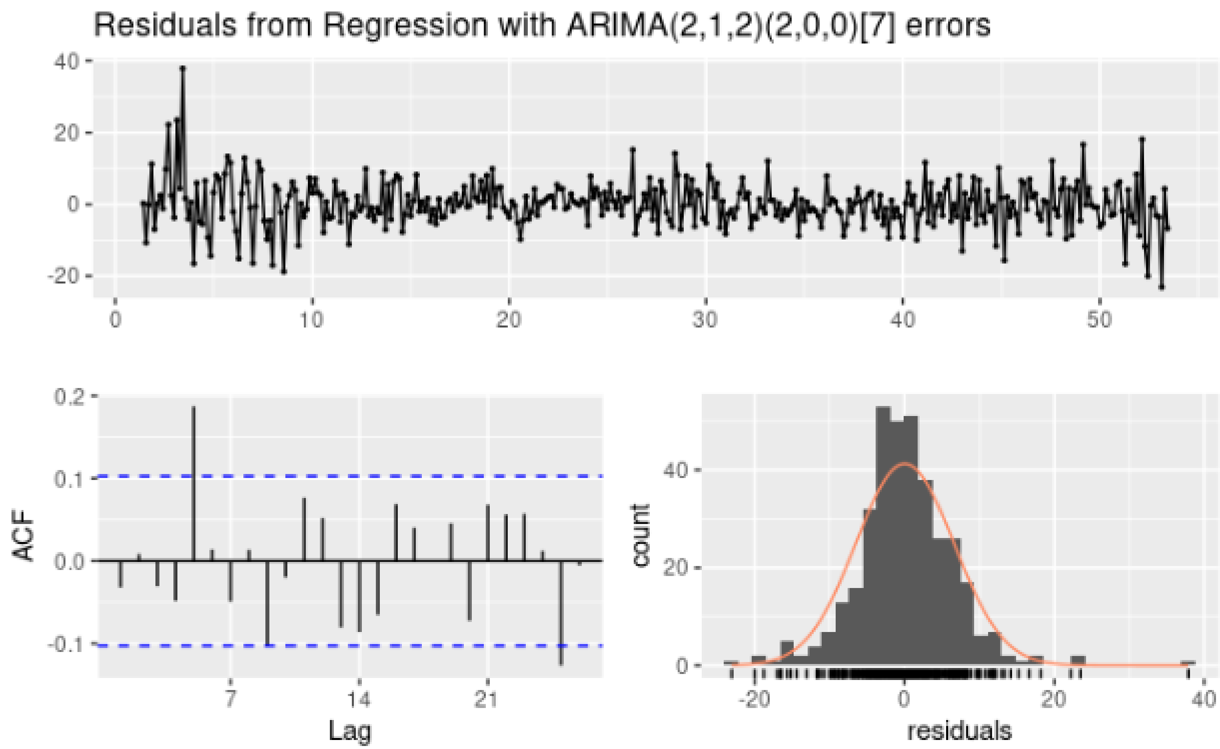
https://otexts.com/fpp2/forecasting.html

ARIMAX Models



We can fit a model that includes

- a quadratic regression for the temperature influence
- an ARIMA model for the remaining errors, with 7-day seasonality reflecting the work week



https://otexts.com/fpp2/forecasti

Vector Autoregression

Vector Autoregression is another extension of ARIMA-style models

- Unlike ARIMAX, VAR is used when two or more time series influence each other
- The main difference is that in ARIMAX models – the predictors influence the dependent variable, but not the other way around
- In VAR-based approaches (VAR, VARMA), the variables are allowed to be interrelated

VAR treats each variable as a function of the lagged values of all of the variables in the system

- In AR, a signal is modeled as a linear combination of past value of itself.
- In VAR, a signal is modeled as a linear combination of past value of itself and of the other variables.
- Because we are solving for multiple times series that influence each other, they are modeled as a system of equations

  (i.e. One equation per signal/time series)

Vector Autoregression

Consider an original AR(1) model:

$$x_t = \beta_0 + \phi_1 x_{t-1} + \varepsilon_t$$

A corresponding a VAR(1) model with 2 covariates is as follows:

$$x_{1,t} = \beta_1 + \phi_{11,1} x_{1,t-1} + \phi_{12,1} x_{2,t-1} + \varepsilon_{1,t}$$
$$x_{2,t} = \beta_2 + \phi_{21,1} x_{1,t-1} + \phi_{22,1} x_{2,t-1} + \varepsilon_{2,t}$$

where $x_{1,t-1}$ and $x_{2,t-1}$ are the 1st lag of $x_{1,t}$ and $x_{2,t}$

$\phi_{AB,L}$ is the effect of the L^{th} lag of variable *A* on variable *B*

For a VAR(2) model with 2 covariates

$$x_{1,t} = \beta_1 + \phi_{11,1} x_{1,t-1} + \phi_{12,1} x_{2,t-1} + \phi_{11,2} x_{1,t-2} + \phi_{12,2} x_{2,t-2} + \varepsilon_{1,t}$$
$$x_{2,t} = \beta_2 + \phi_{21,1} x_{1,t-1} + \phi_{22,1} x_{2,t-1} + \phi_{21,2} x_{1,t-2} + \phi_{22,1} x_{2,t-2} + \varepsilon_{2,t}$$

Vector Autoregression

VAR is sometimes more popular than VARMA because it doesn't include MA terms

- Approximates any them using extra AR lags
- Can be a less compact than directly using MA terms (like the ARIMAX model)
- But enables direct coefficient estimation using least squares approach

Example:

- 8 Quarterly Values related to Inflation

```
1. rgnp  : Real GNP.
2. pgnp  : Potential real GNP.
3. ulc   : Unit labor cost.
4. gdfco : Fixed weight deflator for personal consumption expenditure excluding food and ene
5. gdf   : Fixed weight GNP deflator.
6. gdfim : Fixed weight import deflator.
7. gdfcf : Fixed weight deflator for food in personal consumption expenditure.
8. gdfce : Fixed weight deflator for energy in personal consumption expenditure.
```

Vector Autoregression

Example:

- 8 Quarterly Values related to Inflation
- Very similar trends across all of them – so are they 'interrelated'?

https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/

Granger Causality

Granger Causality is used to determine if one time series will be useful to forecast another

- Tests whether including past information from one variable will improve the prediction of another
- The Null hypothesis is that a lagged time series does not “Granger-cause” another series

  Note: this is not the same as “caused”
- That is, the past values provide no predictive value for the target series
- If the P-Values significant (< 0.05, for example) then you reject the null hypothesis and determine that it IS indeed useful.

Granger Causality

Granger Causality is used to determine if one time series will be useful to forecast another

- Tests whether including past information from one variable will improve the prediction of another
- The Null hypothesis is that a lagged time series does not “Granger-cause” another series
- That is, the past values provide no predictive value for the target series
- If the P-Values significant (< 0.05, for example) then you reject the null hypothesis and determine that it IS indeed useful.

Granger Causality

Granger Causality is used to determine if one time series will be useful to forecast another

- Tests whether including past information from one variable will improve the prediction of another
- The Null hypothesis is that a lagged time series does not “Granger-cause” another series
- That is, the past values provide no predictive value for the target series
- If the P-Values significant (< 0.05, for example) then you reject the null hypothesis and determine that it IS indeed useful.

$$x_{1,t} = \sum_{j=1}^{p} \beta_{11,j} x_{1,t-j} + \sum_{j=1}^{p} \beta_{12,j} x_{2,t-j} + \varepsilon_{1,t}$$

$$x_{2,t} = \sum_{j=1}^{p} \beta_{21,j} x_{1,t-j} + \sum_{j=1}^{p} \beta_{22,j} x_{2,t-j} + \varepsilon_{2,t}$$

If $x_{2,t-j}$ contains no info about $x_{1,t}$, then $\beta_{12,j} \approx 0$

- We say “Granger-cause”, because it may not be truly causal, but the variables at least provide predictive information

Causality vs Correlation

Side note: it's worth considering that causality ≠ correlation

Example: predict the number of drownings at a beach resort each month based on the number of ice-cream cones sold over that period.

- When it is hot outside, there are more people at the beach, more people eating ice-cream, and more people likely to drown
- So, ice cream and drowning are correlated, but they are both caused by a third variable (say, temperature)
- Here, if temperature wasn't built into the model, but has causation over a predictor (which then affects the outcome), it is considered a confounder

You can build a predictive model based on correlated variables, or even backwards causation

- e.g. predict rain based on the number of cyclists on the road
- but, it is usually best to include causation in the model if you can

Granger Causality

Example:

- 8 Quarterly Values related to Inflation
- Very similar trends across all of them – so are they ‘interrelated’?
- Results of a Granger-Causality Test:

	rgnp_x	pgnp_x	ulc_x	gdfco_x	gdf_x	gdfim_x	gdfcf_x	gdfce_x
rgnp_y	1.0000	0.0003	0.0001	0.0212	0.0014	0.0620	0.0001	0.0071
pgnp_y	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ulc_y	0.0000	0.0000	1.0000	0.0002	0.0000	0.0000	0.0000	0.0041
gdfco_y	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
gdf_y	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
gdfim_y	0.0011	0.0067	0.0014	0.0083	0.0011	1.0000	0.0004	0.0000
gdfcf_y	0.0000	0.0000	0.0008	0.0008	0.0000	0.0038	1.0000	0.0009
gdfce_y	0.0025	0.0485	0.0000	0.0002	0.0000	0.0000	0.0000	1.0000

https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/

Granger Causality

Example:

- 8 Quarterly Values related to Inflation
- Very similar trends across all of them – so are they ‘interrelated’?
- Results of a Granger-Causality Test:

	rgnp_x	pgnp_x	ulc_x	gdfco_x	gdf_x	gdfim_x	gdfcf_x	gdfce_x
rgnp_y	1.0000	0.0003	0.0001	0.0212	0.0014	0.0620	0.0001	0.0071
pgnp_y	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ulc_y	0.0000	0.0000	1.0000	0.0002	0.0000	0.0000	0.0000	0.0041
gdfco_y	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
gdf_y	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
gdfim_y	0.0011	0.0067	0.0014	0.0083	0.0011	1.0000	0.0004	0.0000
gdfcf_y	0.0000	0.0000	0.0008	0.0008	0.0000	0.0038	1.0000	0.0009
gdfce_y	0.0025	0.0485	0.0000	0.0002	0.0000	0.0000	0.0000	1.0000

Almost all variables are Granger-Causing each other!

https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/

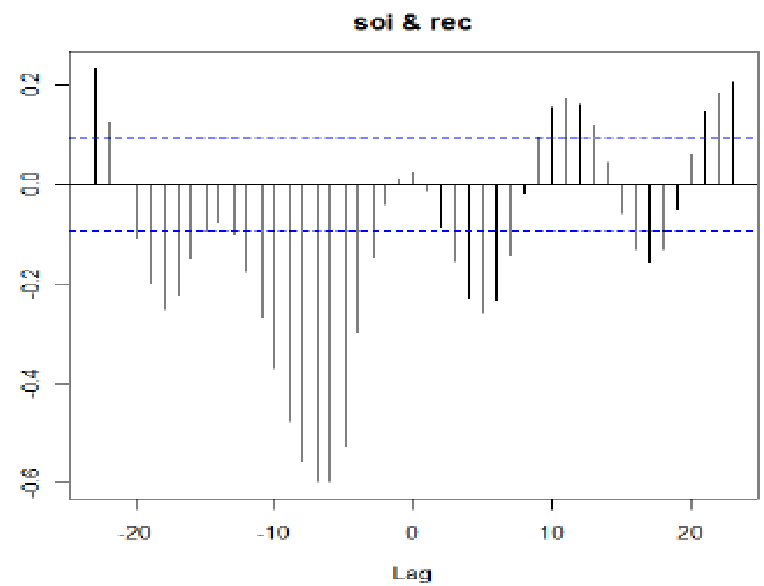
Cross Correlation Function

Just as we used the ACF and PACF to help determine the AR and MA orders for the univariate case, the Cross Correlation Function (CCF) can be used to help determine relationships between variables.

- Helpful for identifying useful lags of the predictor variables
- Correlation between our signal of interest, x_t , and a predictor signal, say, y_{t+h}
- Note that when h is –ve, it precedes x_t, when it is +ve, it lags it
- Because we want to use y_{t+h} to predict future values of x_t, we usually look at the -ve values

Relationship between
- Southern Oscillation Index (soi)
  - a measure of weather

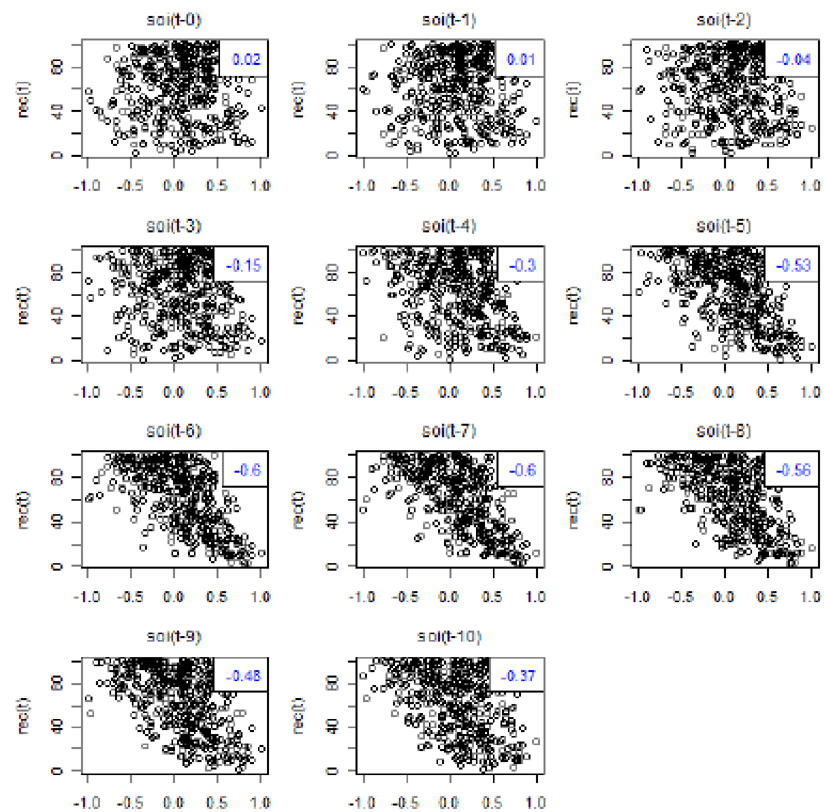
and
- Recruitment (rec)
  - a measure of fish population



https://online.stat.psu.edu/stat510/lesson/8/8.2

Scatter/Lag Plots

Similarly, we can re-use our lag plots, but this time look at lagged versions of the predictors vs. our forecasting signal



Finally, we can determine the appropriate lags using an iterative Information Criteria approach

- Pick the lags that minimize AIC

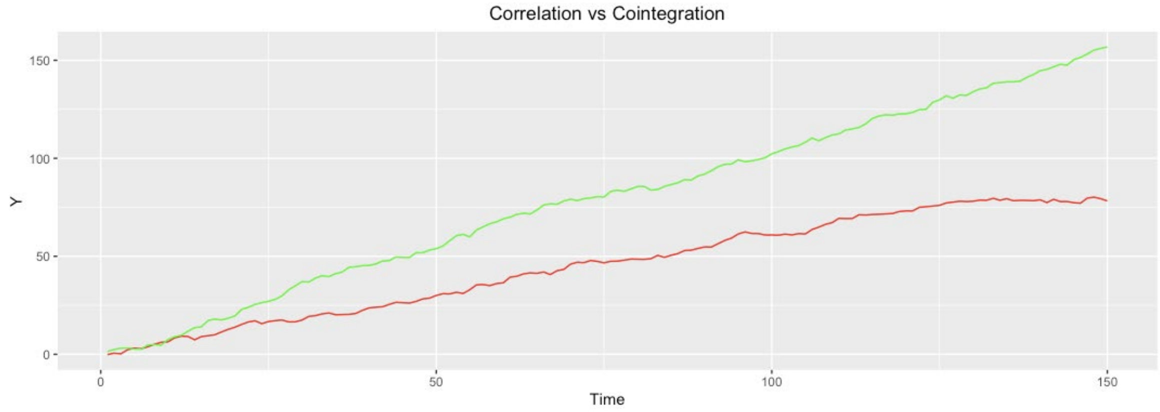
https://online.stat.psu.edu/stat510/lesson/8/8.2

Cointegration

- A related, but separate concept that is important in VAR
- A set of times series are considered cointegrated if there exists a linear combination of them that has an order of integration (*d*) less than that of the individual series
- Two main tests: Engle–Granger test, and the Johanssen test
- What does this mean? They have a long-term statistically significant relationship… sounds like correlation?

Cointegration

- A related, but separate concept that is important in VAR
- A set of times series are considered cointegrated if there exists a linear combination of them that has an order of integration (*d*) less than that of the individual series
- Two main tests: Engle–Granger test, and the Johanssen test
- What does this mean? They have a long-term statistically significant relationship… sounds like correlation? But it isn't
- Cointegration is considered a stronger/better measure of their relationship

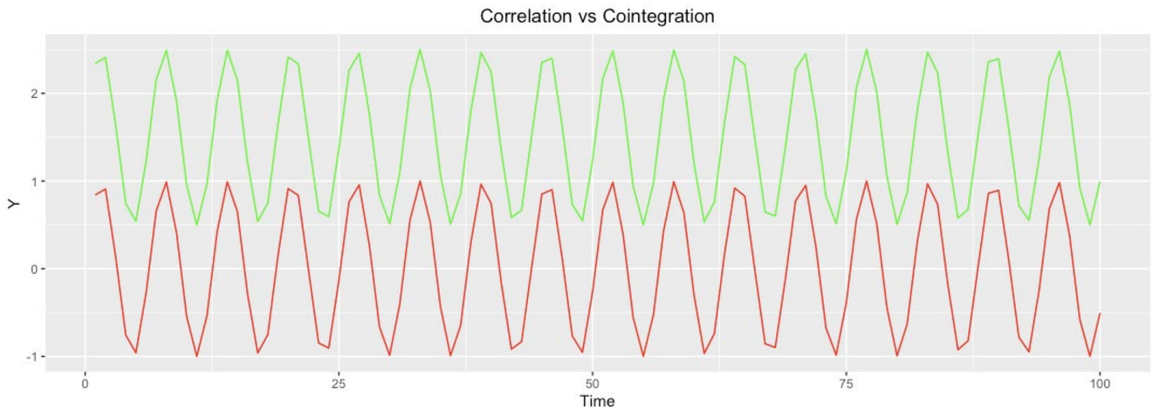


Correlated, but not cointegrated

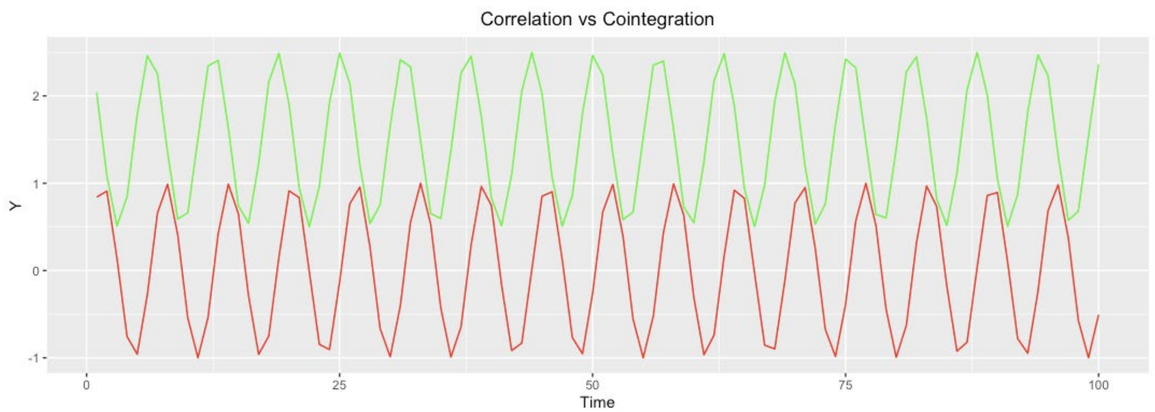
https://medium.com/ro-data-team-blog/measuring-correlation-ii-cointegration-for-time-series-analysis-f0f5e6f65f5

Cointegration

Correlated & Cointegrated



Not Correlated , but Cointegrated (due to added phase lag)



If two signals are cointegrated, then at least one likely Granger-causes the other

- Not necessarily the other way around

https://medium.com/ro-data-team-blog/measuring-correlation-ii-cointegration-for-time-series-analysis-f0f5e6f65f5

Vector Error Correction

For VAR, we've assumed that the individual signals are stationary

- In cases of cointegration, even if the individual signals are non-stationary, the relationship between them may be
- More generally, if the linear combination between of variables has an order of integration (d) less than that of the series themselves, they are cointegrated

When cointegration is identified between variables in a VAR model, it is possible to instead use Vector Error Correction (VEC)

- The advantage of VEC over VAR (when this condition exists) is that the coefficient estimates can be more efficient
- In Python, and in many texts, this is called VECM

So many models!

There are many, many different ARIMA-style and ARIMA-inspired models

- Although they may seem like a lot, they are mostly small tweaks
- ARMA models combined AR and MA models (*p*,*q*)
- ARIMA added the differencing capabilities within the model definition
- SARIMA extended ARIMA by adding a seasonal treatment on the period of *m*
- ARIMAX added the ability to include a regression component based on exogenous variables (external factors)
- VARMA is our first real multivariate approach, and added the ability to include interrelated endogenous variables
- If cointegration is determined, VEC can be used instead of VAR
- SARIMAX, VARX, VARMAX, etc…

https://www.statsmodels.org/stable/examples/index.html#time-series-analysis

https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/

Helpful Resources

machinelearningmastery.com

Machine Learning Mastery
Making Developers Awesome at Machine Learning

Click to Take the FREE Time Series Crash-Course

Search...

Get Started | Blog | Topics | EBooks | FAQ | About | Contact

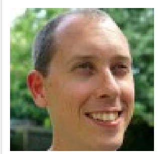
11 Classical Time Series Forecasting Methods in Python (Cheat Sheet)

by **Jason Brownlee** on August 6, 2018 in **Time Series**

Tweet | Share | Share

Last Updated on December 10, 2020

Machine learning methods can be used for classification and forecasting on time series problems.



Welcome!
I'm *Jason Brownlee* PhD
and I **help developers** get results
with **machine learning**.
Read more

Never miss a tutorial:

https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/

Helpful Resources

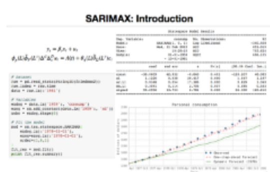
Statsmodels.org

statsmodels v0.12.2

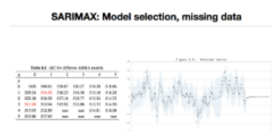
- Installing statsmodels
- Getting started
- User Guide
- Examples
  - Linear Regression Models
    - Ordinary Least Squares
    - Generalized Least Squares
    - Quantile regression
    - Recursive least squares
    - Example 2: Quantity theory of money
    - Example 3: Linear restrictions and formulas
    - Rolling Regression
    - Regression diagnostics
    - Weighted Least Squares
    - Linear Mixed Effects Models
    - Comparing R lmer to statsmodels MixedLM
    - Variance Component Analysis
  - Plotting
    - Regression Plots
    - Plot Interaction of Categorical Factors
    - Box Plots

State space models

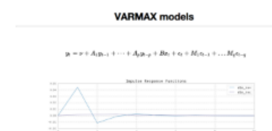
SARIMAX: Introduction



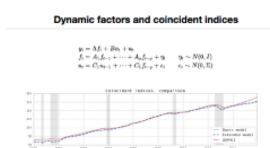
SARIMAX: Model selection, missing data



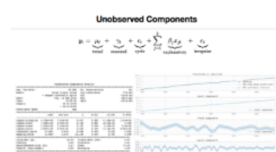
VARMAX: introduction



Dynamic Factor Models: Application



Unobserved Components: Application



https://www.statsmodels.org/stable/examples/index.html#time-series-analysis

Q&A