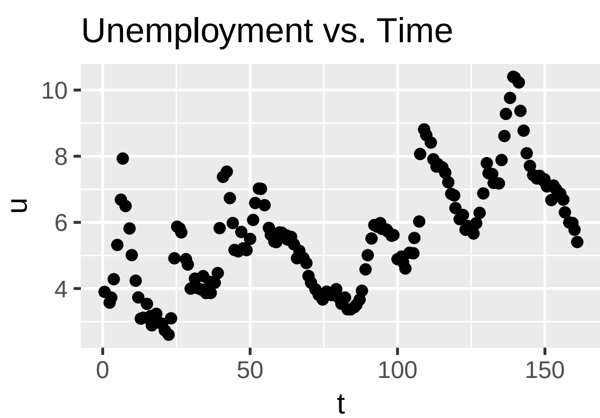
ARIMAX Modeling - US Economy

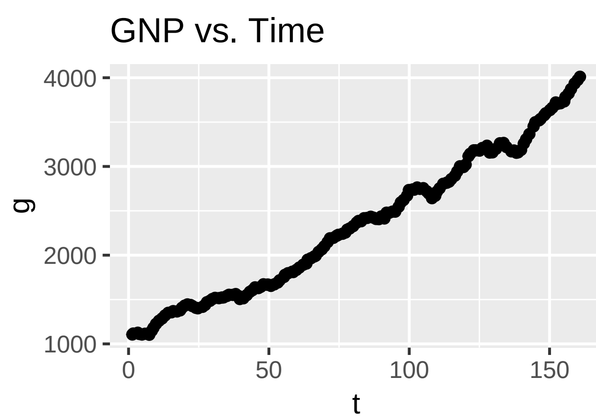
Andira Putri

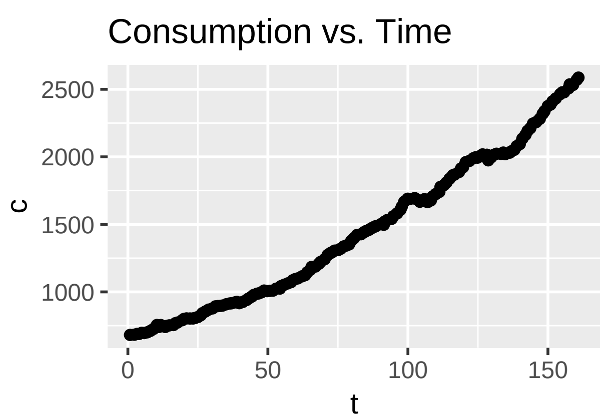
This exercise is taken from **Time Series Analysis and Its Applications: With R Examples** by Shumway and Stoffer. We will be using the `econ5` data set from the `astsa` library. `econ5` is a five quarterly economic series containing the following numeric variables: quarterly U.S. unemployment, GNP, consumption, government investment, and private investment. There are 161 observtions spanning from 1948-III to 1988-II.

Consider the data set `econ5`. The seasonal component has been removed from the data. Concentrating on unemployment (U_t), GNP (G_t), and consumption (C_t), fit a vector ARMA model to the data after first logging each series, and then removing the linear trend. That is, fit a vector ARMA model to $x_t = (x_{1t}, x_{2t}, x_{3t})^t$ where, for example, $x_{1t} = log(U_t) - \hat{\beta}_0 - \hat{\beta}_1 t$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimates for the regression of $log(U_t)$ on time t. Run a complete set of diagnostics on the residuals.

Curious Plotting







From the plots, GNP and consumption would be interesting variables to use as regressors. We can use time, GNP, and consumption to predict unemployment in the United States.

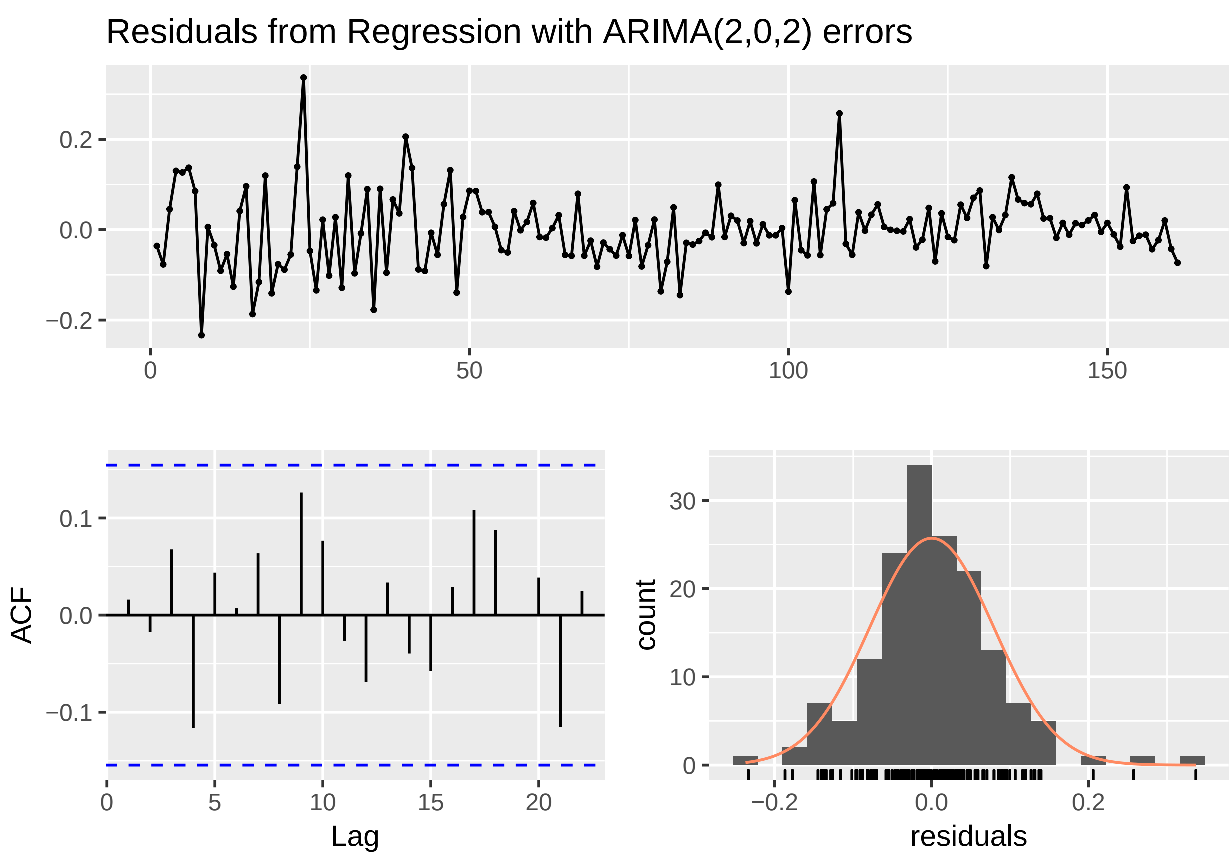
Model Fitting

```
# log transform
log.u=log(u)
log.g=log(g)
log.c=log(c)
```

```r
log.df=cbind(t,log.u,log.g,log.c)
# fit vector ARMA model
varma=auto.arima(log.u,xreg=log.g+log.c)
summary(varma)
```

```
## Series: log.u
## Regression with ARIMA(2,0,2) errors
##
## Coefficients:
##          ar1      ar2     ma1     ma2  intercept    xreg
##       1.3402  -0.4594  0.0105  0.1673    -0.5774  0.1519
## s.e.  0.1629   0.1505  0.1615  0.1218     1.2475  0.0838
##
## sigma^2 estimated as 0.006463:  log likelihood=179.12
## AIC=-344.24   AICc=-343.51   BIC=-322.67
##
## Training set error measures:
##                        ME       RMSE        MAE        MPE     MAPE
## Training set 0.0002232862 0.07887831 0.05894522 -0.2729806 3.777929
##                   MASE       ACF1
## Training set 0.8940719 0.01596416
```

Diagnostics




```
##
##  Ljung-Box test
```

```
##
## data:  Residuals from Regression with ARIMA(2,0,2) errors
## Q* = 9.3476, df = 4, p-value = 0.05297
##
## Model df: 6.   Total lags used: 10
```

Our residuals are normally-distributed for the most part, and there is no clear spike in the ACF diagram. The Ljung-Box test has a p-value of 0.05297 at the 5% level, so the data is independently distributed and there are no remarkable autocorrelations at any lag. This is a desirable result.

Forecasting

We randomly generate vectors for `log(g)` and `log(c)` to use as input values for our forecasting model. Both vectors are of length 8, so we aim to forecast the next 8 quarters' `log(u)` values. Note that for multiple regressors, we have to combine the vectors into a matrix for our `forecast()` function to work properly.