Homework 6

Due: Never, To Be Used as Preparation for Exam 2

Testing Nonstationarity

For $n \in \{10, 100, 500\}$ simulate 100 time series according to each of the following models, all of which assume $w_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_w^2\right)$:

(i) $x_t = 0.5x_{t-1} + 0.25x_{t-2} + 0.125x_{t-3} + w_t$;

(ii) $x_t = -10(t/n - 1/2)^2 + w_t$;

(iii) $x_t = 10(t/n - 1/2)^2 w_t$;

(iv) $x_t = x_{t-1} + w_t$;

(v) $x_t = 0.5x_{t-1} + 0.25x_{t-2} + 0.25x_{t-3} + w_t$;

(vi) $x_t = -x_{t-1} + w_t$.

In total, you will simulate $3 \times 6 \times 100$ time series. For each time series, use the `ndiffs` function from the `forecast` library to perform an Augmented Dickey-Fuller an a Phillips-Perron test of the null hypothesis that the undifferenced time series is non-stationary rejects the null. Record whether or not the null is rejected.

(a) Plot the rejection rates as a function of n for each model.

(b) The probability that a test rejects the null when the alternative is true is the **power** of a test. Based on your figure in (a), describe how the power of the tests depend on n.

(c) Based on your plot in (a), explain in one sentence how well the tests can detect a non-stationary $\mathbf{AR}(3)$ process if ϕ_1 is not greater than 1 and n is large.

(d) Based on your plot in (a), explain in one sentence how well the tests can detect variance nonstationary when n is large.

(e) Based on your plot in (a), explain in one sentence how well the tests can detect a non-stationary random walk when n is large.

(f) Based on your plot in (a), explain in one sentence how well the tests can detect a nonlinear mean function when n is large.

(g) Based on your plot in (a), explain in one sentence how the tests perform when x_t is a non-stationary $\mathbf{AR}(1)$ time series with an autoregressive polynomial root that is not equal to 1 and n is large.

(h) Based on everything you've observed here, is it sufficient to just apply non-stationarity tests to determine whether or not an observed time series is stationary, or is it aso important to look at the data? Answer in at most one sentence.

An ARIMA Example

On Homework 4, we forecasted future values of the `chicken` chicken data by fitting a linear time trend and applying an $\mathbf{ARMA}(p, q)$ model for the residuals. This homework will ask you to return to that. Specifically, you should have had a figure like the following:

```r
library(astsa)

data(chicken)

r <- lm(chicken~time(chicken))$res
```

```
fit <- arima(r, order = c(2, 0, 3))
plot(chicken, type = "l", xlim = c(min(time(chicken)), 2020),
     ylim = c(60, 130), xlab = "Time", ylab = "Chicken Prices")
lm <- lm(chicken~time(chicken))
abline(a = lm$coef[1], b = lm$coef[2], col = "purple")
n.ahead <-12*3 + 6
pred <- predict(fit, n.ahead = n.ahead, se.fit = TRUE)
t.new <- max(time(chicken)) + 1:length(pred$pred)/12
lines(t.new, lm$coef[1] + lm$coef[2]*t.new + pred$pred, col = "blue")
lines(t.new, lm$coef[1] + lm$coef[2]*t.new + pred$pred + qnorm(0.975)*pred$se,
      col = "blue", lty = 2)
lines(t.new, lm$coef[1] + lm$coef[2]*t.new + pred$pred - qnorm(0.975)*pred$se,
      col = "blue", lty = 2)
```
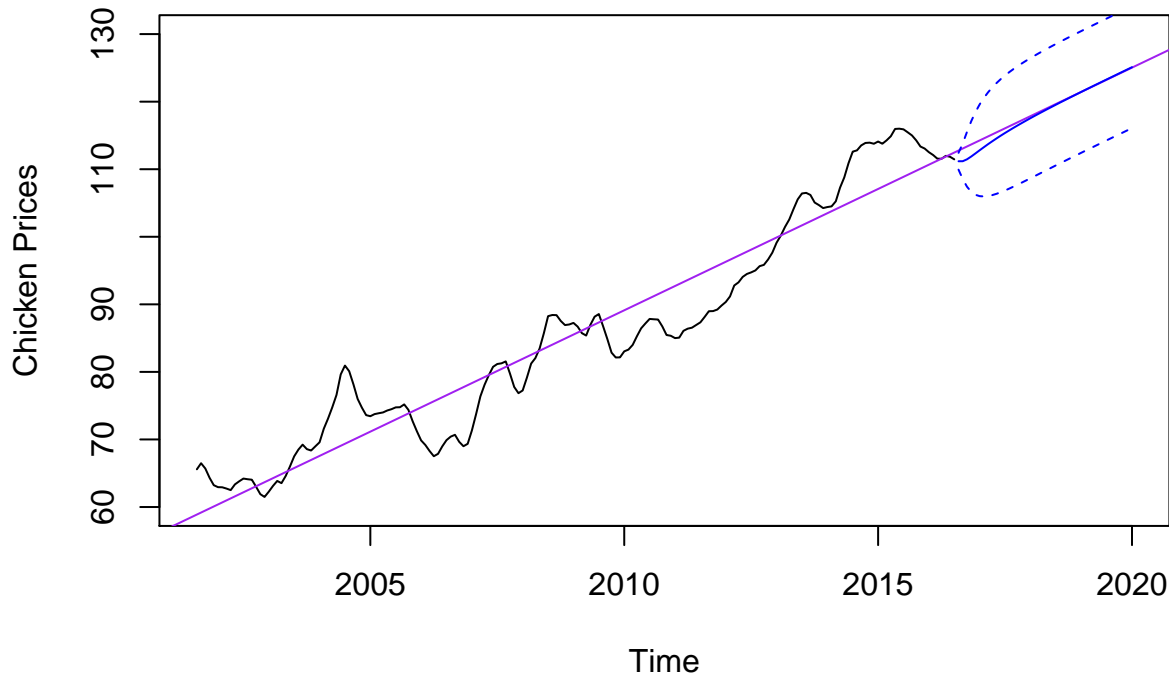

(a) First - on Homework 4 we took a somewhat ad-hoc two-step approach by first regressing out the time trend and then adding it back later to obtain forecasts, ignoring any estimation error. Using the `arima` function, find the AIC minimizing values for $0 \leq p \leq 3$ and $0 \leq q \leq 3$ for the model:

$$\phi(B)(x_t - a + bt) = \theta(B) w_t,$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$. Do you get the same AIC minimizing values of p and q?

(b) Add predicted values and prediction standard errors from the model in (a) to your plot. In one sentence, comment on why they do or do not match the predicted values and prediction standard errors we obtained in Homework 4.

(c) Instead of modeling x_t as stationary about a linear trend, we might use an **ARIMA**(p, d, q) model. Use the `ndiffs` function from the `forecast` package to determine how many times you should difference the data. Include a linear trend and use a level-0.05 Augmented Dickey Fuller test. What does the test indicate you should choose for d?

(d) Use the `binned.mv.plot` function to make a plot of the residuals from the linear model fit, $x_t - a - \hat{b}t$, using bins of size 30. Do you think that the deviations from a the fitted time trend look stationary?

```r
binned.mv.plot <- function(x, bin.size = 30) {

  bins <- floor(0:(length(x) - 1)/bin.size)
  small.bins <- which(table(bins) == min(table(bins)) & !table(bins) == max(table(bins)))
  bins[bins == small.bins - 1] <- NA

  bin.means <- aggregate(as.numeric(x), list("bins" = bins), mean)$x
  bin.vars <- aggregate(as.numeric(x), list("bins" = bins), var)$x

  par(mfrow = c(1, 2))
  plot(bin.means, type = "b", xlab = "Bin", ylab = "Mean", ylim = range(x))
  plot(bin.vars, type = "b", xlab = "Bin", ylab = "Variance", ylim = c(0, max(x^2)))
}
```

(e) Using the `arima` function, find the AIC minimizing values for $0 \le p \le 3$ and $0 \le q \le 3$ for the model:

$$\phi(B)(\nabla x_t - c) = \theta(B) w_t,$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$. To force the inclusion of the intercept d, you will need to use the `xreg` argument. Add predicted values and prediction standard errors from the AIC minimizing model to the plot from (b). Because you used the `xreg` argument to get the ARIMA fit with an intercept, you will need to use the `newxreg` argument of the `predict` function applied to the ARIMA object.

(f) Use the `binned.mv.plot` function to make a plots of the mean and variance of ∇x_t and $\nabla^2 x_t$, using bins of size 30. This will yield two sets of two plots each. Do the first or second differences look more stationary?

(g) Using the `arima` function, find the AIC minimizing values for $0 \le p \le 3$ and $0 \le q \le 3$ for the model:

$$\phi(B)(\nabla^2 x_t - k) = \theta(B) w_t,$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$. To force the inclusion of the intercept k, you will need to use the `xreg` argument. Add predicted values and prediction standard errors from the AIC minimizing model to the plot from (b). Because you used the `xreg` argument to get the ARIMA fit with an intercept, you will need to use the `newxreg` argument of the `predict` function applied to the ARIMA object.

(h) In one sentence, comment on why the predicted values and prediction standard errors change the way they do as we increase d.

(i) Given everything you saw, explain in one sentence which predicted values and prediction standard errors you would use for forecasting for this data.