5. Autoreg & Distributed Lags

author: ECONOMIC FORECASTING date: summer 2017 autosize: true

Model Autocorrelation

Generalized differences and thoughtful model specification are attempts at removing autocorrelation.

What if autocorrelation is "pure"? - We can try to model autocorrelation - Autoregression

Assume:

For an autoregressive process of order 1: $\epsilon_t = \rho\epsilon_{t-1} + u_t$ where $|\rho| < 1 \implies$ the time series is stationary. When the autoregression is of order greater than $1, |\rho_1 + \rho_2 + \cdots + \rho_p| < 1$ and $\rho_j \to 0$ as j gets large.

Definition

Autoregression:

A time series, $\{x_t\}$, is an **autoregressive** process of order p denoted AR(p), if

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + w_t$$

where $\{w_t\} \sim WN$

In backshift notation…

$$\theta_p(\mathbf{B})x_t = (1 - \alpha_1\mathbf{B} - \alpha_2\mathbf{B}^2 - \cdots - \alpha_p\mathbf{B}^p)x_t = w_t$$

Identifying Stationarity

Characteristic Equation:

$$\theta_p(\mathbf{B}) = 0$$

For stationarity, the roots of the characteristic equation must all be strictly greater than 1 in absolute value.

For example:

$$x_t = x_{t-1} + w_t$$
$$(1 - \mathbf{B})x_t = w_t$$

$\mathbf{B} = 1 \not> 1 \implies$ nonstationarity

Example

$$
\begin{aligned}
x_t &= -\frac{3}{2}x_{t-1} + x_{t-2} + w_t \\
x_t + \frac{3}{2}x_{t-1} - x_{t-2} &= w_t \\
\left(1 + \frac{3}{2}\mathbf{B} - \mathbf{B}^2\right)x_t &= w_t \\
-1\left(\mathbf{B}^2 - \frac{3}{2}\mathbf{B} - 1\right)x_t &= w_t \\
-1(\mathbf{B} - 1/2)(\mathbf{B} + 2)x_t &= w_t
\end{aligned}
$$

Thus, the roots are 1/2 and 2, implying $\{x_t\}$ is nonstationary

Another Example

$$
\begin{aligned}
x_t &= -\frac{3}{4}x_{t-1} - \frac{1}{8}x_{t-2} + w_t \\
\frac{1}{8}x_{t-2} - \frac{3}{4}x_{t-1} + x_t &= w_t \\
\left(\frac{1}{8}\mathbf{B}^2 - \frac{3}{4}\mathbf{B} + 1\right)x_t &= w_t \\
\frac{1}{8}(\mathbf{B}^2 - 6\mathbf{B} + 8)x_t &= w_t \\
\frac{1}{8}(\mathbf{B} - 4)(\mathbf{B} - 2)x_t &= w_t
\end{aligned}
$$

Thus, the roots are 4 and 2, implying $\{x_t\}$ is stationary

Properties of AR(1)

Recall from the random walk model that

$$
x_t = \alpha x_{t-1} + w_t \Rightarrow x_t = \alpha(\alpha x_{t-2} + w_{t-1}) + w_t = \sum_{i=0}^{\infty} \alpha^i w_{t-i}
$$

So the expected value can be derived

$$
\begin{aligned}
\mathbf{E}(x_t) &= \mathbf{E}\left(\sum_{i=0}^{\infty} \alpha^i w_{t-i}\right) \\
&= \sum_{i=0}^{\infty} \alpha^i \mathbf{E}(w_{t-i}) \\
&= 0
\end{aligned}
$$

Properties of AR(1)

$$
\begin{aligned}
\gamma_k(x_t, x_{t+k}) &= \mathbf{C}\left(\sum_{i=0}^{\infty} \alpha^i w_{t-i}, \sum_{i=0}^{\infty} \alpha^j w_{t+k-j}\right) \\
&= \sum_{j=k+i}^{\infty} \alpha^i \alpha^j \mathbf{C}(w_{t-i}, w_{t+k-j}) \\
&= \alpha^k \sigma_w^2 \sum_{i=0}^{\infty} \alpha^{2i} \\
&= \frac{\alpha^i \sigma_w^2}{(1-\alpha^{2i})}
\end{aligned}
$$

So, what is γ_k if $\alpha = 1$ (i.e. random walk)?

===

$x_t = \frac{3}{4}x_{t-1} - \frac{1}{8}x_{t-2} + w_t$

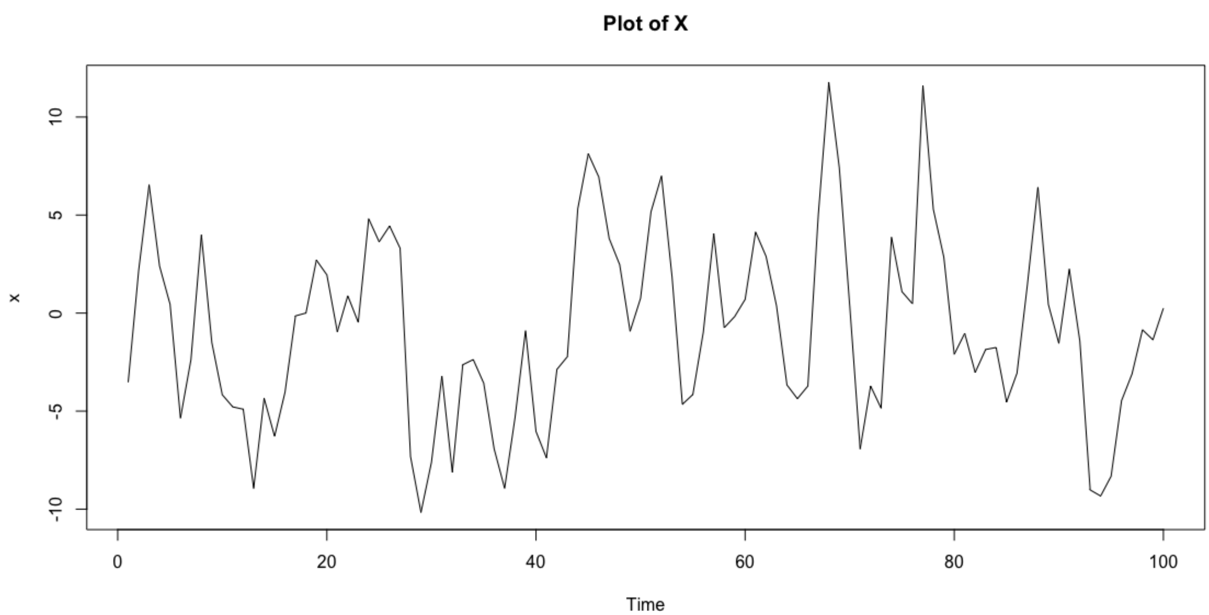

Figure 1:

===

$x_t = \frac{3}{4}x_{t-1} - \frac{1}{8}x_{t-2} + w_t$

Partial ACF

The partial autocorrelation function (ϕ_{kk}) is a measure of the correlation that remains at lag k after the correlation from all the previous periods have been removed.

Properties of the White Noise process:

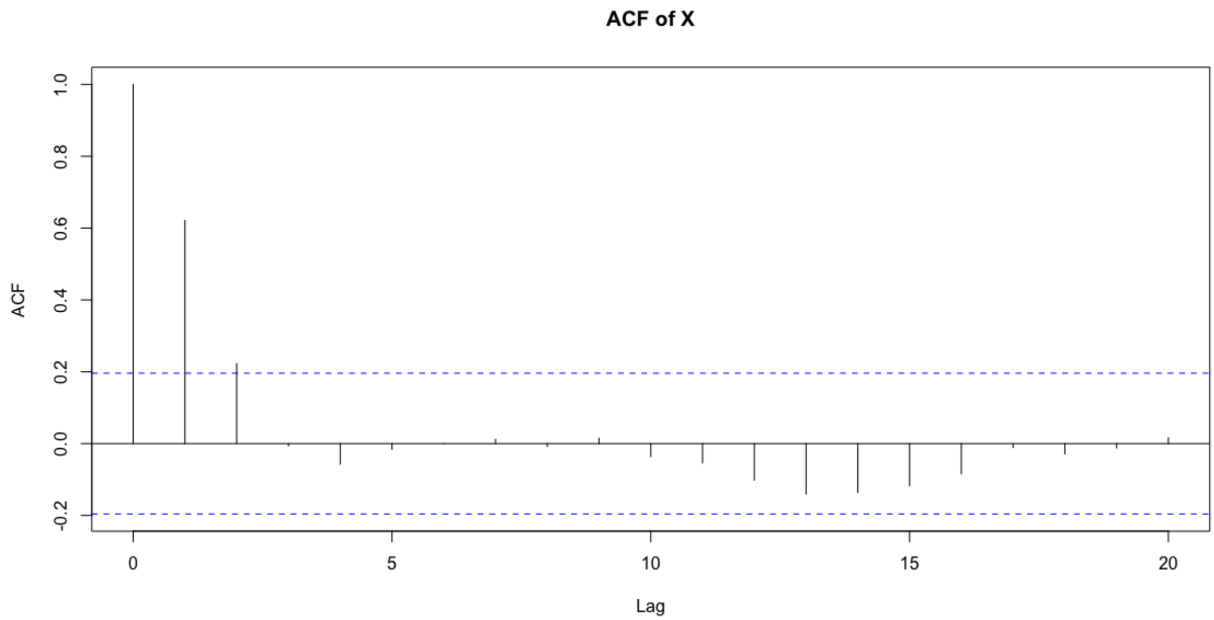

Figure 2:

$$\phi_{kk} = 0 \;\forall k$$

Properties of the AR(p) process:

$$\phi_{kk} = 0 \;\; k > p$$

In R, `pacf()`

Partial ACF

$x_t = \frac{3}{4}x_{t-1} - \frac{1}{8}x_{t-2} + w_t$
Only 1 lag is significant even though it is an AR(2) process

Forecasting AR

Suppose that we have an AR(1) process:

$$x_t = \alpha x_{t-1} + w_t$$

One way to forecast would just be to estimate α such that our forecast would be

$$\hat{x}_{t+1} = \hat{\alpha}\, x_t$$

$\hat{\alpha}$ is chosen by minimizing the sum of squared residuals. However, as evidenced by the ACFs and PACFs above, choosing the correct AR(p) process if NOT trivial.

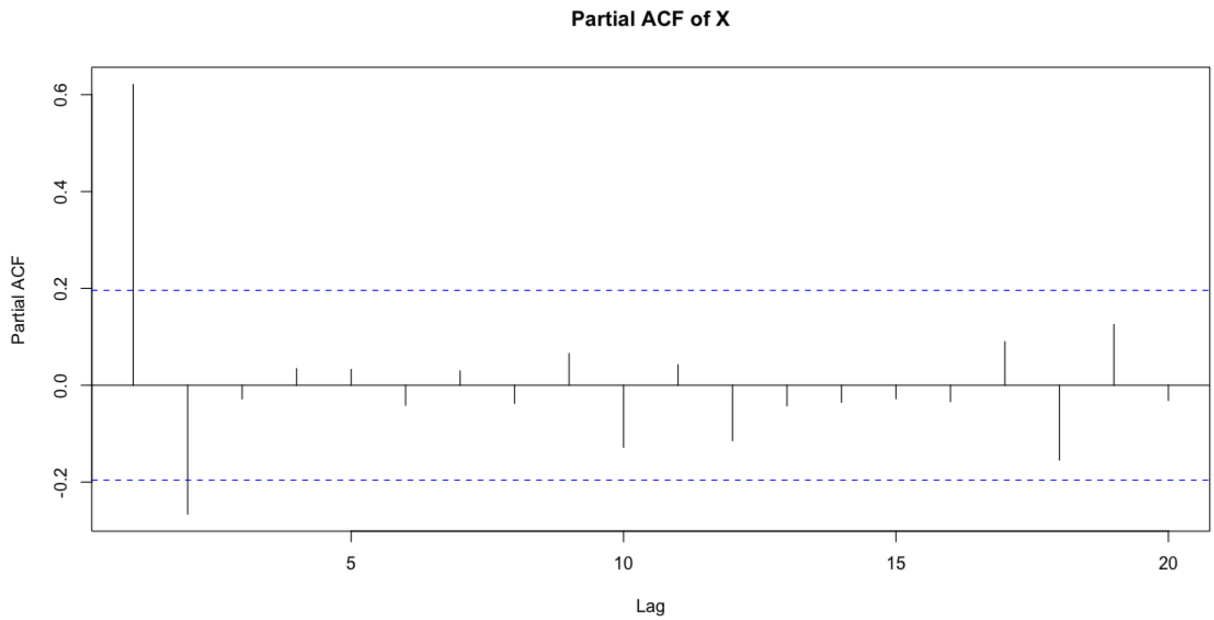

Figure 3:

AIC

Akaike's Information Criterion (AIC)

$$AIC = 2k - 2\ln(\mathcal{L})$$

- $\mathcal{L}$ is the log-likelihood
- Punishes inclusion of independent variables more heavily than $\bar{R}^2$
- The smaller the AIC, the better the fit

The `ar()` function in R automatically creates this.

SIC

Schwartz-Bayesian Information Criterion (SIC)

$$SIC = k\ln(n) - 2\ln(\mathcal{L})$$

- Punishes the inclusion of independent variables even *more* heavily than the AIC
- Again, the smaller the SIC, ther better the fit.

Corrected AIC

The AIC can be prone to overfitting. Thus, we have a correction for the AIC known as the **Corrected AIC**:

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

- Punishes the inclusion of independent variables more heavily than the AIC
- The smaller the AIC_c , the better the fit
- Easily calculable in R

AR Wrap-Up

- Visual inspection of the dsata is crucial
  - Plots, ACFs, PACFs
- Choose model based on minimized AIC or AIC_c
- Difficulty of indentifying the model displays the art of forecasting

Distributed Lag Models

Lagged independent variables appear in many regressions - Past influences the future - Especially prevelant in time series data

Distributed Lag Models - Economic changes can be distributed over a number of time periods - A series of lagged explanatory variables accounts for the time adjustment process

Specifying the Model

$$
\begin{aligned}
y_t &= \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots \\
&= \alpha + \sum_{s=0}^{\infty} \beta_s x_{t-s} + \epsilon_t
\end{aligned}
$$

If we use OLS, we run into two problems:

1. Degrees of Freedom problem
   - Remember, N must be greater than k
2. Multicollinearity

To deal with these problems, we can specify some conditions about the structure of the lag.

Geometric Lag

Assumes that the lagged independent are all positive and decline geometrically

$$
\begin{aligned}
y_t &= \alpha + \beta(x_t + w x_{t-1} + w^2 x_{t-2} + \cdots) + \epsilon \\
&= \alpha + \beta \sum_{s=0}^{\infty} w^s x_{t-s} + \epsilon_t \ , \ 0 < w < 1
\end{aligned}
$$

Using the Model

$$
y_t = \alpha + \beta(x_t + w x_{t-1} + \cdots) + \epsilon \tag{1}
$$

$$y_{t-1} = \alpha + \beta(x_{t-1} + wx_{t-2} + \cdots) + \epsilon_{t-1} \tag{2}$$

Multiplying (2) by w and subtracting this relation from (1), we get:

$$y_t - wy_{t-1} = \alpha(1-w) + \beta x_t + u_t \tag{3}$$

where $u_t = \epsilon_t - w\epsilon_{t-1}$. Rewriting (3), we obtain:

$$y_t = \alpha(1-w) + wy_{t-1} + \beta x_t + u_t \tag{4}$$

Equation (4) is an **Autoregressive Distributed Lag** (ARDL) model

==

The long-run effect of x_t is

$$\frac{\beta}{(1-w)}$$

Be careful in using OLS on this model: - There is a lagged dependent variable - Error vector must be transformed - Must test for serial correlation - Breusch-Godfrey LM Test

==

Clearly, the geometric lag places a severe restriction on how the lags are distributed

Polynomial Distibuted Lag

Add structure to the geometric lag model:

$$y_t = \alpha(1-w) + wy_{t-1} + \beta x_t + \gamma x_{t-1} + \delta x_{t-2} + u_t$$

This allows $\beta, \gamma,$ and δ to all have differential effects

Use AIC or AIC_c to help specify the model