# Times Series Analysis (IX) – Model Specification (II)


Jianyong Sun
School of Mathematics and Statistics
Xi'an Jiaotong University


16th Oct., 2017

## Recap

- Model specification – how to choose appropriate values for $p, d$, and $q$ for a given series
- ACF, sample ACF
  - The sampling properties of sample ACF
  - The standard error to the estimation of $\rho_k$ through $r_k$ can be obtained for both MA and AR models
- Partial ACF, sample PACF
  - It is developed to decide the order of an AR model since the autocorrelation of an AR does not have cut-off.
  - For an AR($p$), its PACF cuts off exactly at $k > p$.
  - For an MA($q$), its ACF cuts off exactly at $k > q$.
  - For an ARMA($p, q$), both ACF and PACF do not cut off.

## Recap

- Extended ACF (EACF)
  - For mixed ARMA model
  - The idea is to ‘filter out’ the autoregression to obtain a pure MA process, and enjoys the cutoff property of its ACF
  - The AR coefficients may be estimated by a sequence of regressions.

# Nonstationarity

- many series exhibit nonstationarity that can be explained by integrated ARMA models.
- The nonstationarity will frequently be apparent in the time series plot of the series.
- The definition of the sample autocorrelation function implicitly **assumes** stationarity; for example, we use lagged products of deviations from the overall mean, and the denominator assumes a constant variance over time. Thus it is not at all clear what the sample ACF is estimating for a nonstationary process
- Nevertheless, for nonstationary series, the sample ACF typically fails to die out rapidly as the lags increase. This is due to the tendency for nonstationary series to drift slowly, either up or down, with apparent “trends".

## Nonstationarity

Consider the oil price time series:

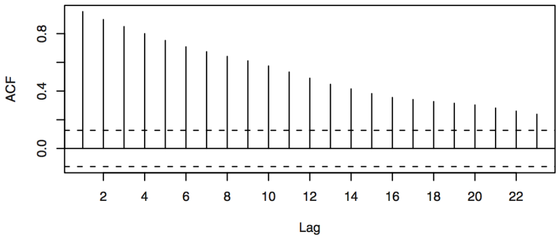



**Figure:** Sample ACF for the Difference of the Log Oil Price Time Series.

- The sample ACF for the logarithms of these data is displayed.
- All values shown are “significantly far from zero", and the only pattern is perhaps a linear decrease with increasing lag.
- The sample PACF (not shown) is also indeterminate.

# Nonstationarity

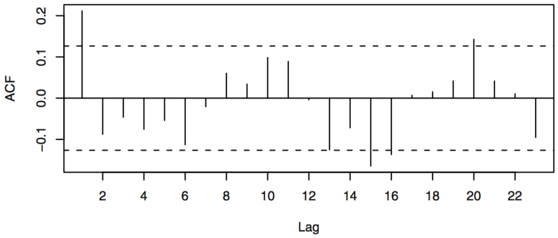



**Figure:** Sample ACF for the Difference of the Log Oil Price Time Series.

- The sample ACF computed on the first differences of the logs of the oil price series is shown.
- Now the pattern emerges much more clearly — after differencing, a moving average model of order 1 seems appropriate.
- The model for the original oil price series would then be a nonstationary IMA(1,1) model. (The “significant" ACF at lags 15, 16, and 20 are ignored for now.)

# Nonstationarity

- If the first difference of a series and its sample ACF do not appear to support a stationary ARMA model, then we take another difference and again compute the sample ACF and PACF to look for characteristics of a stationary ARMA process.
- Usually one or at most two differences, perhaps combined with a logarithm or other transformation, will accomplish this reduction to stationarity.
- Additional properties of the sample ACF computed on nonstationary data are given in Wichern (1973), Roy (1977), and Hasza (1980). See also Box, Jenkins, and Reinsel (1994, p. 218).

# Overdifferencing

- we know that the difference of any stationary time series is also stationary.
- However, overdifferencing introduces unnecessary correlations into a series and will complicate the modeling process.

# Overdifferencing

suppose our observed series, $\{Y_t\}$, is in fact a random walk so that one difference would lead to a very simple white noise model

$$\nabla Y_t = Y_t - Y_{t-1} = e_t$$

However, if we difference once more (that is, overdifference) we have

$$\nabla^2 Y_t = e_t - e_{t-1}$$

which is an MA(1) model but with $\theta = 1$.

# Overdifferencing

- If we take two differences in this situation we unnecessarily have to estimate the unknown value of $\theta$. Specifying an IMA(2,1) model would not be appropriate here.
- The random walk model, which can be thought of as IMA(1,1) with $\theta = 0$, is the correct model.
- Overdifferencing also creates a noninvertible model, which create serious problems when we attempt to estimate their parameters.

# Overdifferencing

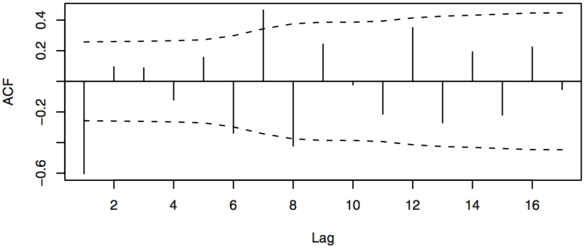



**Figure:** Sample ACF of Overdifferenced Random Walk

- Based on this plot, we would likely specify at least an IMA(2,1) model for the original series and then estimate the unnecessary MA parameter.
- We also have a significant sample ACF value at lag 7 to think about and deal with.

# Overdifferencing

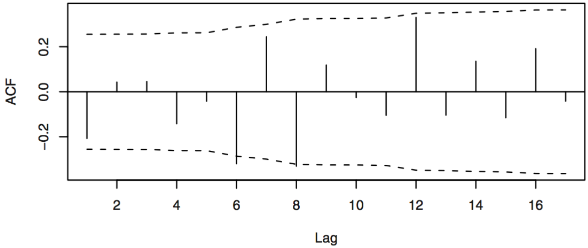



**Figure:** Sample ACF of Correctly Differenced Random Walk

To avoid overdifferencing, we recommend looking carefully at each difference in succession and keeping the principle of parsimony always in mind — models should be simple, but not too simple $\rightarrow$ Occam's Razor Principle (**Entities should not be multiplied unnecessarily**)

# The Dickey-Fuller Unit Root Test

- While the approximate linear decay of the sample ACF is often taken as a symptom that the underlying time series is nonstationary and requires differencing,
- It is also useful to quantify the evidence of nonstationarity in the data-generating mechanism.
- The use of the first difference can be too severe a modification in the sense that the nonstationary model might represent an overdifferencing of the original process.
- This can be done via hypothesis testing.

## The Dickey-Fuller Unit Root Test

Consider a casual AR(1) model,

$$Y_t = \phi Y_{t-1} + e_t$$

Through differencing, we have:

$$\nabla Y_t = \phi \nabla Y_{t-1} + e_t - e_{t-1}$$

introduces extraneous correlation and invertibility problems.

That is, while $Y_t$ is a casual AR(1) process, working with the differenced process $Y_t$ will be problematic because it is a noninvertible ARMA(1,1).

# The Dickey-Fuller Unit Root Test

A unit root test provides a way to test whether previous equation is a random walk (the null case) as opposed to a causal process (the alternative). That is, it provides a procedure for testing

$$H_0 : \phi = 1 \qquad \textit{versus} \qquad H_1 : |\phi| < 1.$$

An obvious test statistic would be to choose $(\hat{\phi} - 1)$, appropriately normalized, in the hope to develop an asymptotically normal test statistic, where $\hat{\phi}$ is one of the optimal estimators.

To examine the behaviour of $(\hat{\phi} - 1)$ under the null hypothesis that $\phi = 1$, or more precisely that the model is a random walk, $Y_t = \sum_{j=1}^{t} e_t$, or $Y_t = Y_{t-1} + e_t$ with $Y_0 = 0$, consider the least square estimator of $\phi$.

Note that $\mu_Y = 0$, the least square estimator can be written as

$$\hat{\phi} = \frac{\frac{1}{n}\sum_{t=1}^{n} Y_t Y_{t-1}}{\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}^2} = 1 + \frac{\frac{1}{n}\sum_{t=1}^{n} e_t Y_{t-1}}{\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}^2}$$

where we have written $Y_t = Y_{t-1} + e_t$ in the numerator.

## The Dickey-Fuller Unit Root Test

Recall that $Y_t = 0$, and in the least square setting, we are regressing $Y_t$ on $Y_{t-1}$ for $t = 1, 2, \cdots, n$. Hence under $H_0$, we have that

$$\hat{\phi} - 1 = \frac{\frac{1}{n\sigma_e^2} \sum_{t=1}^n e_t Y_{t-1}}{\frac{1}{n\sigma_e^2} \sum_{t=1}^n Y_{t-1}^2}$$

Note that first squaring both sides of $Y_t = Y_{t-1} + e_t$, we obtain $Y_t^2 = Y_{t-1}^2 + 2Y_{t-1}e_t + e_t^2$ so that

$$Y_{t-1}e_t = \frac{1}{2}(Y_t^2 - Y_{t-1}^2 - e_t^2)$$

Hence

$$\frac{1}{n\sigma_e^2} \sum_{t=1}^{n} e_t Y_{t-1} = \frac{1}{2} \left( \frac{Y_n^2}{n\sigma_e^2} - \frac{\sum_{t=1}^n e_t^2}{n\sigma_e^2} \right)$$

Because $Y_n = \sum_{t=1}^{n} e_t$, we have that $Y_n \sim \mathcal{N}(0, n\sigma_e^2)$, so that $\frac{1}{n\sigma_e^2} Y_n^2 \sim \chi_1^2$ (i.e. the chi-squared distribution with one degree of freedom.

Because $e_t$ is white Gaussian noise, $\frac{1}{n} \sum e_t^2 \to_p \sigma_e^2$ or $\frac{1}{n\sigma_e^2} \sum e_t^2 \to_p 1$. Consequently, as $n \to \infty$,

$$\frac{1}{n\sigma_e^2} \sum_{t=1}^{n} e_t Y_{t-1} \to \frac{1}{2}(\chi_1^2 - 1)$$

Consider the denominator $\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}^2$. According to the functional central limit theorem, we know that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_{t-1}}{\sigma_e\sqrt{n}}\right)^2 \to_d \int_0^1 W^2(t)dt$$

where $W(t)$ is standard Brownian motion which statisfies

(i) $W(0) = 0$

(ii) $\{W(t_2) - W(t_1), \cdots, W(t_n) - W(t_{n-1})\}$ are independent

(iii) $W(t + \Delta(t)) - W(t) \sim \mathcal{N}(0, \Delta t)$ for $\Delta t > 0$

# The Dickey-Fuller Unit Root Test

We thus obtain

$$n(\hat{\phi}-1)=\frac{\frac{1}{n\sigma_e^2}\sum_{t=1}^{n}e_tY_{t-1}}{\frac{1}{n^2\sigma_e^2}\sum_{t=1}^{n}Y_{t-1}^2}\to\frac{\frac{1}{2}(\chi_1^2-1)}{\int_0^1 W^2(t)dt}$$

The test statistic $n(\hat{\phi}-1)$ is known as the root or Dickey-Fuller (DF) statistic.

Because the distribution of the test statistic does not have a closed form, quantiles of the distribution must be computed by numerical approximation or by simulation

# The Dickey-Fuller Unit Root Test

Toward a more general model, we note that the DF test was established by noting that if $Y_t = \phi Y_{t-1} + X_t$, then $\nabla Y_t = (\phi - 1)Y_{t-1} + e_t = \gamma Y_{t-1} + e_t$. One could test $H_0 : \gamma = 0$ by regressing $\nabla Y_t$ on $Y_{t-1}$.

A Wald statistic and its limiting distribution can be derived (previous derivation is based on Brownian motion is due to Philips (1987).

The test can then be extended to accommodate AR($k$) model as follows. Consider the model $Y_t = \sum_{j=1}^{p-1} \phi_j Y_{t-j} + e_t$.

Subtract $Y_{t-1}$ from the model, we obtain

$$Y_t - Y_{t-1} = \gamma Y_{t-1} + \sum_{j=1}^{k-1} \psi_j \nabla Y_{t-j} + e_t$$

where $\gamma = \sum_{j=1}^{k} \phi_j - 1$ and $\psi_j = -\sum_{j=i}^{k} \phi_j$ for $j = 2, \cdots, p$.

For a quick check, when $p = 2$, note that
$Y_t = (\phi_1 + \phi_2) Y_{t-1} - \phi_2 (Y_{t-1} - Y_{t-2}) + e_t$.

# The Dickey-Fuller Unit Root Test

To test the hypothesis that the process has a unit root at 1 (i.e., the AR polynomial $\phi(z) = 0$ when $z = 1$), we can test $\mathbf{H}_0 : \gamma = 0$ by estimating $\gamma$ in the regression of $\nabla Y_t$ on $\nabla Y_{t-1}, \nabla Y_{t-2}, \nabla Y_{t-p+1}$, and forming a Wald test based on

$$t_\gamma = \hat{\gamma}/se(\gamma).$$

where *se* is the standard error of the regression. This leads to the so-called **augmented Dickey-Fuller test** (ADF).

# The Dickey-Fuller Unit Root Test

- The augmented Dickey-Fuller (ADF) test statistic is the $t$-statistic of the estimated coefficient of $\lambda$ from the method of least squares regression.
- However, the ADF test statistic is not approximately $t$-distributed under the null hypothesis; instead, it has a certain non-standard large-sample distribution under the null hypothesis of a unit root.
- Fortunately, percentage points of this limit (null) distribution have been tabulated; see Fuller (1996)

## The Dickey-Fuller Unit Root Test

- In practice, even after first differencing, the process may not be a finite-order AR process, but it may be closely approximated by some AR process with the AR order increasing with the sample size.
- Said and Dickey (1984) showed that with the AR order increasing with the sample size, the ADF test has the same large-sample null distribution as the case where the first difference of the time series is a finite-order AR process.
- Often, the approximating AR order can be first estimated based on some information criteria (for example, AIC or BIC) before carrying out the ADF test.

## The Dickey-Fuller Unit Root Test

- In some cases, the process may be trend nonstationary in the sense that it has a deterministic trend (for example, some linear trend) but otherwise is stationary.
- A unit-root test may be conducted with the aim of discerning difference stationarity from trend stationarity.
- This can be done by carrying out the ADF test with the detrended data.
- Equivalently, this can be implemented by regressing the first difference on the covariates defining the trend, the lag 1 of the original data, and the past lags of the first difference of the original data.
- The t-statistic based on the coefficient estimate of the lag 1 of the original data furnishes the ADF test statistic, which has another nonstandard large-sample null distribution

## The Dickey-Fuller Unit Root Test

Illustrate how the ADF test do with the simulated random walk:

- First, we consider testing the null hypothesis of unit root versus the alternative hypothesis that the time series is stationary with unknown mean.
- Hence, the regression defined by $Y_t - Y_{t-1} = \gamma Y_{t-1} + \sum_{j=1}^{k-1} \psi_j \nabla Y_{t-j} + e_t$ is augmented with an intercept to allow for the possibly nonzero mean under the alternative hypothesis.
  - Determine $p$, according to the AIC with the first difference of the data, $k$ is found to be 8, in which case the ADF test statistic becomes -0.601, with the $p$-value being greater than 0.1.
  - On the other hand, setting $k = 0$ (the true order) leads to the ADF statistic -1.738, with $p$-value still greater than 0.1.
  - Thus there is strong evidence supporting the unit-root hypothesis.

# The Dickey-Fuller Unit Root Test

- Second, recall that the simulated random walk appears to have a linear trend.
- Hence, linear trend plus stationary error forms another reasonable alternative to the null hypothesis of unit root (**difference nonstationarity**).
  - For this test, we include both an intercept term and the covariate time in the regression.
  - With $k = 8$, the ADF test statistic equals -2.289 with p-value greater than 0.1 $\rightarrow$ do not reject the null hypothesis
  - Setting $k = 0$, the ADF test statistic becomes -3.49 with $p$-value equal to 0.0501.
  - There is weak evidence that the process is linear-trend nonstationary; that is, the process equals linear time trend plus stationary error, contrary to the truth that the process is a random walk, being difference nonstationary!

This example shows that with a small sample size, it may be hard to differentiate between trend nonstationarity and difference nonstationarity.

## The Dickey-Fuller Unit Root Test

For a ADF test, there are three steps:

- Test for a unit root on

$$\nabla Y_t = \gamma Y_{t-1} + \sum_{j=1}^{p} \psi_j Y_{t-j} + e_t$$

- Test for a unit root with drift (non-zero mean)

$$\nabla Y_t = \alpha_0 + \gamma Y_{t-1} + \sum_{j=1}^{p} \psi_j Y_{t-j} + e_t$$

- Test for a unit root with drift and deterministic time trend:

$$\nabla Y_t = \alpha_0 + \alpha_1 t + \gamma Y_{t-1} + \sum_{j=1}^{p} \psi_j Y_{t-j} + e_t$$

# The Dickey-Fuller Unit Root Test

- Each version of the test has its own critical value which depends on the size of the sample
- In each case, the null hypothesis is that there is a unit root, $\gamma = 0$.
- The tests have low statistical power in that they often cannot distinguish between true unit-root processes ($\gamma = 0$) and near unit-root processes ( $\gamma$ is close to zero). This is called the "near observation equivalence" problem.

# Other Specification Methods

A number of other approaches to model specification have been proposed since Box and Jenkins's seminal work.

One of the most studied is **Akaike's (1973) Information Criterion** (AIC). This criterion says to select the model that minimizes

$$AIC = -2\log(\text{maximum likelihood}) + 2k$$

where $k = p + q + 1$ if the model contains an intercept or constant term and $k = p + q$ otherwise.

The addition of the term $2k$ serves as a "penalty function" to help ensure selection of parsimonious models and to avoid choosing models with too many parameters.

# Other Specification Methods

By adding another nonstochastic penalty term to the AIC, resulting in the corrected AIC, denoted by $\text{AIC}_c$ and defined by the formula

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

Here n is the (effective) sample size and again k is the total number of parameters as above excluding the noise variance.

Hurvich and Tsai (1989) suggest that for cases with $k/n$ greater than 10%, the $\text{AIC}_c$ outperforms many other model selection criteria, including both the AIC and BIC.

# Other Specification Methods

Another approach to determining the ARMA orders is to select a model that minimizes the Schwarz **Bayesian Information Criterion** (BIC) defined as

$$BIC = -2\log(\text{maximum likelihood}) + k\log(n)$$

If the true process follows an ARMA($p, q$) model, then it is known that the orders specified by minimizing the BIC are consistent; that is, they approach the true orders as the sample size increases.

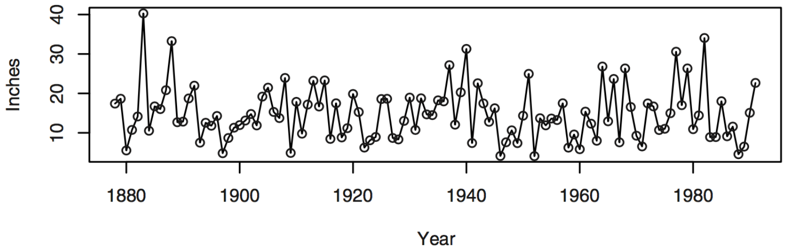
However, if the true process is not a finite-order ARMA process, then minimizing AIC among an increasingly large class of ARMA models enjoys the appealing property that it will lead to an optimal ARMA model that is closest to the true process among the class of models under study

## Other Specification Methods

- Regardless of whether we use the AIC or BIC, the methods require carrying out maximum likelihood estimation.
- However, maximum likelihood estimation for an ARMA model is prone to numerical problems due to multimodality of the likelihood function and the problem of overfitting when the AR and MA orders exceed the true orders.
- Hannan and Rissanen (1982)'s method
  - first fitting a high-order AR process with the order determined by minimizing the AIC.
  - then uses the residuals from the first step as proxies for the unobservable error terms.
  - an ARMA$(k, j)$ model can be approximately estimated by regressing the time series on its own lags 1 to k together with the lags 1 to $j$ of the residuals from the high order autoregression;
  - the BIC of this autoregressive model is an estimate of the BIC obtained with maximum likelihood estimation.

# Specification of Some Actual Time Series

The Los Angeles Annual Rainfall Series




The rainfall amounts were not normally distributed.

# Specification of Some Actual Time Series

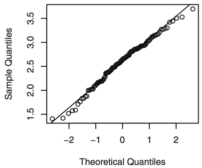



**Figure:** QQ Normal Plot of the Logarithms of LA Annual Rainfall.

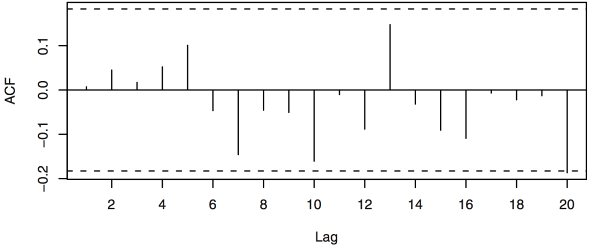



**Figure:** Sample ACF of the Logarithms of LA Annual Rainfall.

The Chemical Process Color Property Series

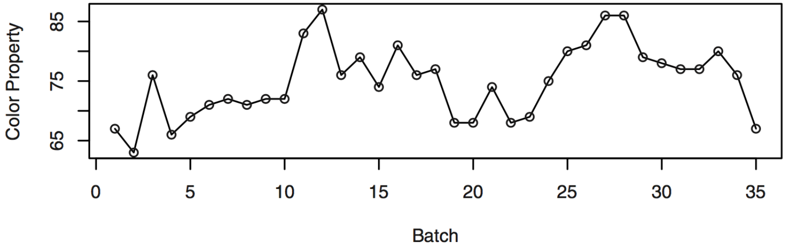



**Figure:** Time Series plot of Color Property from a Chemical Process.

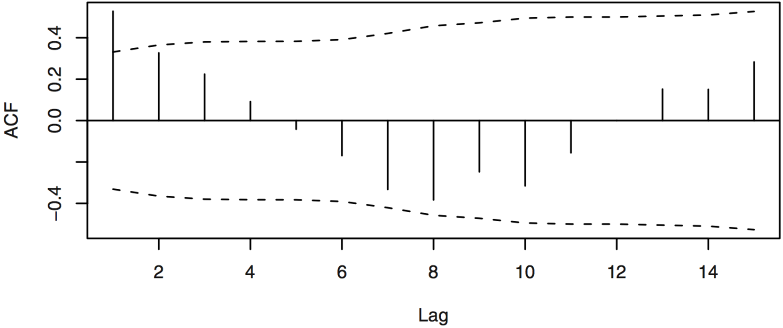



**Figure:** Sample ACF for the Color Property Series.

- The sample ACF plotted might suggest an MA(1) model, as only the lag 1 autocorrelation is significantly different from zero.
- However, the damped sine wave appearance encourages us to look further at the sample partial autocorrelation.

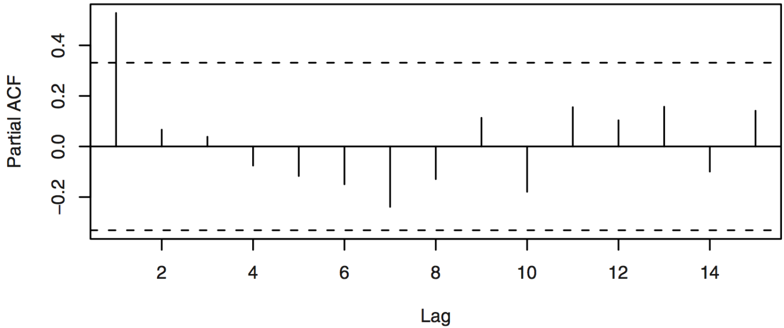
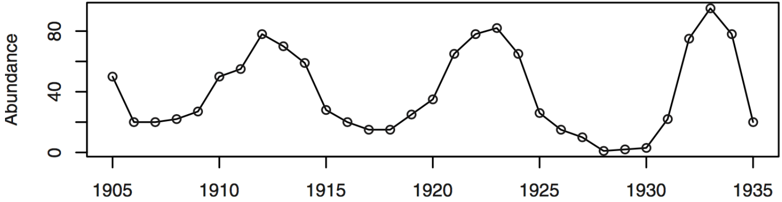



**Figure:** Sample PACF for the Color Property Series.

It is clear that an AR(1) model is worthy of first consideration. As always, our specified models are tentative and subject to modification during the model diagnostics stage of model building.

# Specification of Some Actual Time Series

The Annual Abundance of Canadian Hare Series




It has been suggested in the literature that a transformation might be used to produce a good model for these data

## Specification of Some Actual Time Series

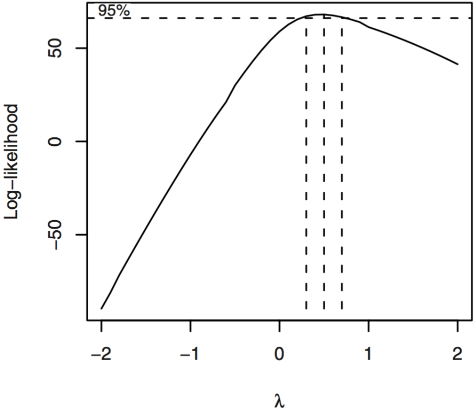



**Figure:** Box-cox power transformation results for hare abundance.

The plot shows the log-likelihood as a function of the power parameter $\lambda$. The maximum occurs at $\lambda = 0.4$, but a square root transformation with $\lambda = 0.5$ is well within the confidence interval

# Specification of Some Actual Time Series

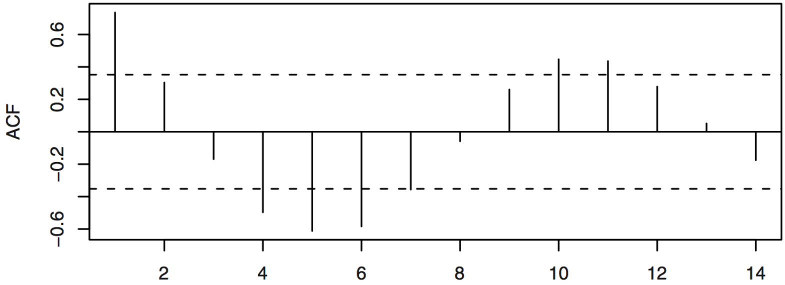



**Figure:** Sample ACF for Square Root of Hare Abundance.

The fairly strong lag 1 autocorrelation dominates but, again, there is a strong indication of damped oscillatory behavior.

## Specification of Some Actual Time Series

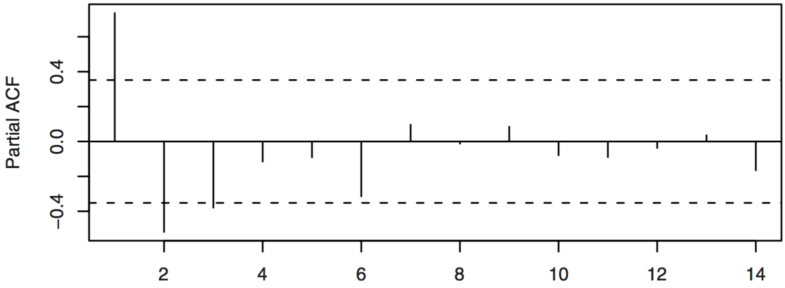



**Figure:** Sample partial ACF for Square Root of Hare Abundance.

It gives strong evidence to support an AR(2) or possibly an AR(3) model for these data.

The Oil Price Series

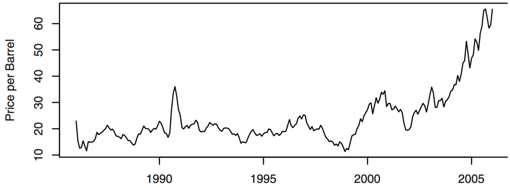



**Figure:** The Oil Price Series.

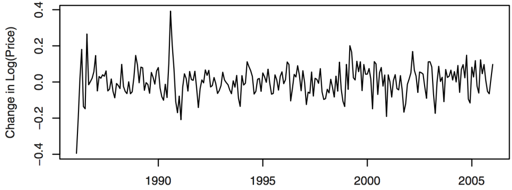



**Figure:** The First Difference of log Oil Price Series.

## Specification of Some Actual Time Series

- It is argued graphically that the difference of the logarithms could be considered stationary.
- Software implementation of the Augmented Dickey-Fuller unit-root test applied to the logs of the original prices leads to a test statistic of -1.1119 and a $p$-value of 0.9189.
- With stationarity as the alternative hypothesis, this provides strong evidence of nonstationarity and the appropriateness of taking a difference of the logs.
- For this test, the software chose a value of $k = 6$ based on large-sample theory.

## Specification of Some Actual Time Series

| AR / MA | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | x | o | o | o | o | o | o | o | o | o | o | o | o | o |
| 1 | x | x | o | o | o | o | o | o | o | o | x | o | o | o |
| 2 | o | x | o | o | o | o | o | o | o | o | o | o | o | o |
| 3 | o | x | o | o | o | o | o | o | o | o | o | o | o | o |
| 4 | o | x | x | o | o | o | o | o | o | o | o | o | o | o |
| 5 | o | x | o | x | o | o | o | o | o | o | o | o | o | o |
| 6 | o | x | o | x | o | o | o | o | o | o | o | o | o | o |
| 7 | x | x | o | x | o | o | o | o | o | o | o | o | o | o |

**Figure:** Extended ACF for Difference of Logarithms of Oil Price Series.

This figure shows the summary EACF table for the differences of the logarithms of the oil price data.

This table suggests an ARMA model with $p = 0$ and $q = 1$.

## Specification of Some Actual Time Series

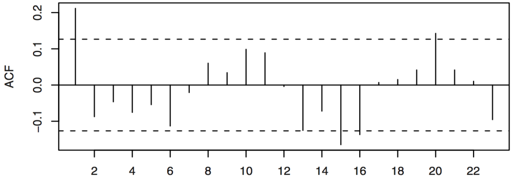



**Figure:** Sample ACF of Difference of Logged Oil Prices.

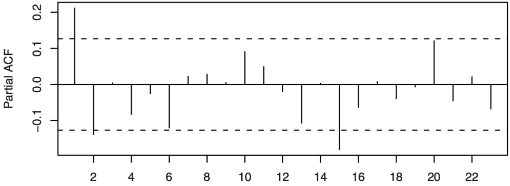



**Figure:** Sample PACF of Difference of Logged Oil Prices.

## Specification of Some Actual Time Series

- The sample ACF figure suggests that we specify an MA(1) model for the difference of the log oil prices,
- The sample PACF says to consider an AR(2) model (ignoring some significant spikes at lags 15, 16, and 20).
- We will want to look at all of these models further when we estimate parameters and perform diagnostic tests
- We will see later that to obtain a suitable model for the oil price series, the outliers in the series will need to be dealt with.

# Summary

- we considered the problem of specifying reasonable but simple models for observed times series.
- we investigated tools for choosing the orders ($p$, $d$, and $q$) for ARIMA($p, d, q$) models.
- Three tools, the sample autocorrelation function, the sample partial autocorrelation function, and the sample extended autocorrelation function, were introduced and studied to help with this difficult task.
- The Dickey-Fuller unit-root test was also introduced to help distinguish between stationary and nonstation-ary series.
- These ideas were all illustrated with both simulated and actual time series.

Questions?