STAT 5170: Applied Time Series

Course notes for part A of learning unit 5

Section 5A.1: Model building and predictive assessment.

In this learning unit, our aim is to apply the tools we have discussed thus far to *explore time-series data* and *build time-series models*. One approach toward these aims relies heavily on the use of regression methods to identify aspects of time series data that are non-stationary, or that are otherwise systematic. For a time series, (x_t), a basic decomposition is

$$x_t \quad = \quad \mu_t + \epsilon_t,$$

where μ_t is a non-random value that changes over time, and (ϵ_t) is a zero-mean, stationary time series. As a rudimentary model, it describes (x_t) as exhibiting stationary behavior around a *trend*, and defines the trend as $E(x_t) = \mu_t$. The stationary time series (ϵ_t) collects the *residual deviations* from the trend. One challenge of working with data the exhibit a combination of trend and stationarity is that strong trends in μ_t can sometimes obscure any dependencies within the time series of residuals, (ϵ_t). In what follows, we will discuss diagnostic tools that can help to unravel such characteristics.

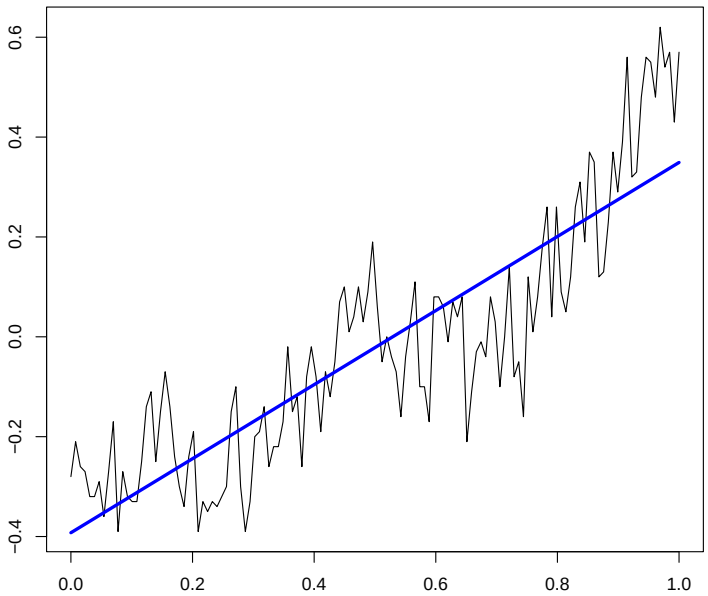
Later, we will discuss a refinement of the basic decomposition described above into

$$x_t \quad = \quad \mu_t + s_t + \epsilon_t,$$

where the component s_t is added to capture the possibility of periodic behavior, which may reflect seasonal patterns. However, before discussing periodic behavior, let us consider the following exploration of time series data, guided by the basic time-series decomposition into trend and stationarity residual deviations from the trend.

Example: Exploring average global temperatures

The graph below displays a time series of global temperature values from 1880-2009, shifted from a baseline temperature.

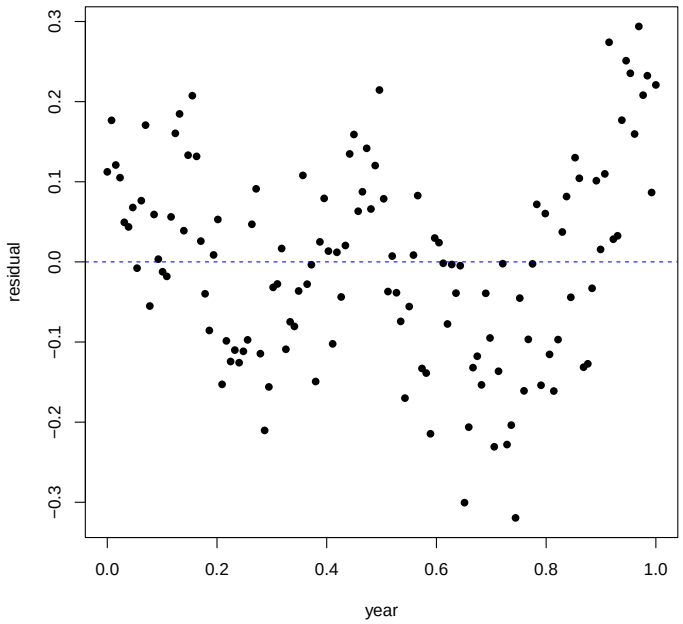


Added to the plot is a *fitted* regression line calculated under the model

$$x_t = \beta_1 + \beta_2 t + \epsilon_t,$$

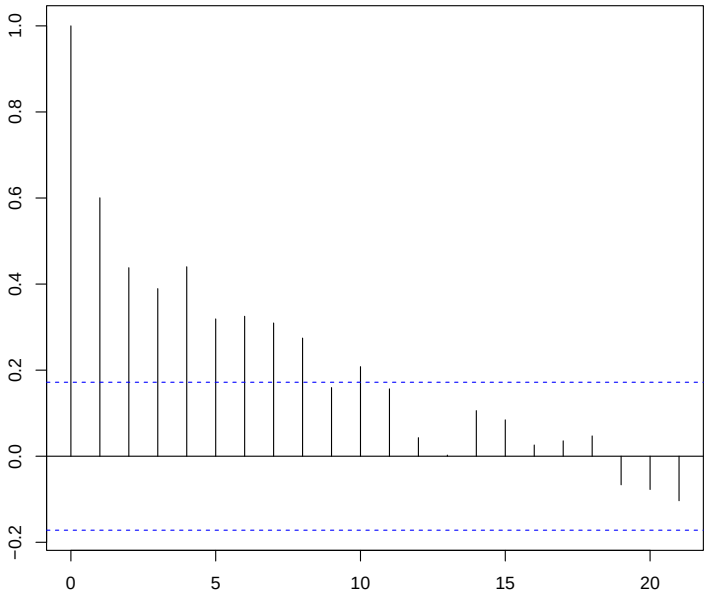
where β_1 and β_2 are regression parameters and (ϵ_t) is Gaussian white noise with variance parameter σ_ϵ^2. Recall from a previous example that this type of model, where the time-index is a regressor variable, would be used to de-trend a time series. Under this regression model, the coefficient β_2, describes an expected increase in temperature, x_t, per one unit increase in time. Here, the mean function is $\mu_t = \beta_1 + \beta_2 t = \boldsymbol{z}_t^T \boldsymbol{\beta}$, wherein the terms written in vector-matrix notation are $\boldsymbol{z}_t = [1 \mid t]^T$ and $\boldsymbol{\beta} = [\beta_1^T \mid \beta_2^T]^T$. The values that define the fitted regression line in the above plot are $\hat{\mu}_t = \boldsymbol{z}_t^T \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is obtained by least-squares analysis. Recall that such $\hat{\mu}_t$ correspond to the posterior means of μ_t in the Bayesian setup under a non-informative prior distribution.

The time series is possibly de-trended in the sense that the regression model provides a route to recovering the residual deviations by subtracting μ_t from each x_t; that is, each $\epsilon_t = x_t - \mu_t$. Adapting this operation to the least-squares fitted regression line yields a time series of fitted residual deviations, $(\hat{\epsilon}_t)$, defined by $\hat{\epsilon}_t = x_t - \hat{\mu}_t$. These are plotted below for the global temperature data.



One approach to diagnostic analysis would be to incorporate model-checking into a Bayesian analysis that is implemented by numerical simulation; this option is discussed further in the next portion of the learning unit. An alternative, simplistic approach is to check for trends and autocorrelations among the fitted residual deviations, $(\hat{\epsilon}_t)$.

If the time series of residual deviations, (ϵ_t), is to resemble Gaussian white noise, there should be no trends observed in the time-series plot of $(\hat{\epsilon}_t)$, and no patterns in variability over time. A quick visual check is sometimes sufficient to identify glaring problems. Examining the sample autocorrelation of $(\hat{\epsilon}_t)$ may suggest the presence of of autocorrelations. For the global temperature data, a plot of this function is



Note the indications of autocorrelation in this plot.

If a diagnostic analysis illuminates worrisome patterns in the residual deviations, a next step would be to build a different model that offers a better description of patterns exhibited in the data. Reflecting the basic decomposition, $x_t = \mu_t + \epsilon_t$, two potential directions to pursue are to modify the trend component μ_t, or adopt a more sophisticated time-series model for the stationary component (ϵ_t), or do both.

A potentially helpful modification to the trend component would be to add a quadratic term for time, which yields

$$x_t = \beta_1 + \beta_2 t + \beta_3 t^2 + w_t.$$

As for the stationary component, a potentially helpful modification would be to adopt an autoregressive $AR(1)$ model for (ϵ_t). An exploration of these modification is discussed in a later example. □

As referenced in the example, above, the model-checking approach discussed in our earlier discussion of Bayesian inference can sometimes assist in the task of model building. An additional set of diagnostics are also available for choosing between models based on their *performance in making predictions.* The rationale that connects these diagnostics to prediction is not obvious from their mathematical form, and is best discussed in a specialized course; however, this should not prevent us from using them for building a time series model in an applied context.

The first derives from a statistic called *Akiake's information criterion* (AIC), and is understood in terms of the *sum-of-squared-errors statistic* of a regression analysis, together with the model's *number of parameters.* It is described below initially for the case in which the time-series of residual deviations (ϵ_t) is Gaussian white noise, after which a more general version is presented. We start by considering the use of AIC as a diagnostic tool specifically for the purpose of identifying a suitable form for the trend component of the decomposition, treating the time series of residual deviations (ϵ_t) as white noise.

For the context at hand, an underlying setup that is sometimes helpful to keep in mind is one in which there is a pool of q regressor variables available, which, at time t, are

$$z_{t1}, \ldots, z_{tq}$$

Any of these are suitable for inclusion into the regression model, but it is suspected that some are not contributing a great deal of information about the regression relationship. Accordingly, rather than working with a regression model with all q regressor variables included, a smaller model, including just r regressor variables, is defined. A question that arises is whether the model with just r regressor variables includes enough variables, or perhaps includes too many (*i.e.*, some are not contributing a great deal of information); or, it could be that some other model with a distinct set of regressor variables is best. To emphasize, for any model that is to be considered, r is its number of regressor variables, including the intercept term.

Under a multiple linear regression model with r regressor variables,

$$x_t = \beta_1 z_{t1} + \cdots + \beta_r z_{tr} + \epsilon_t,$$

and residual deviations (ϵ_t) that are Gaussian white noise, the sum-of-squared-errors statistic is

$$SSE = \sum_{t=1}^{n} (x_t - \boldsymbol{z}_t^T \hat{\boldsymbol{\beta}})^2,$$

where $\hat{\boldsymbol{\beta}}$ is the vector of fitted regression coefficients obtained by least-squares analysis. This is one of the statistics that arises in classical analysis of variance (ANOVA). The ordinary AIC statistic has the formula

$$AIC = n \log(2\pi SSE/n) + n + 2p.$$

where $p = r + 1$, which counts the total number of parameters in the model. Recall there are r regression coefficients, $\beta_1, \ldots, \beta_r$, plus the white-noise variance parameter, σ_ϵ^2. A smaller value of AIC is interpreted as an indication of better predictive performance.

A general version of AIC, which is applicable to extensions of the standard linear regression model, is defined from its data-generating density. The statistic is

$$AIC \quad = \quad -2 \log p(\boldsymbol{x} | \hat{\boldsymbol{\theta}}^{MLE}) + 2p,$$

where $\hat{\boldsymbol{\theta}}^{MLE}$ is the maximum likelihood estimate of the multi-dimensional parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$.

For example, a *multiple linear regression model with autocorrelated residual errors* is such that $\boldsymbol{x} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, wherein $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_t]$ is defined from a stationary time series (ϵ_t) with respective autocovariance and autocorrelation functions $\gamma_\epsilon(h)$ and $\rho_\epsilon(h)$. The regression model may also be specified according to $\boldsymbol{x} | \boldsymbol{\beta}, \sigma_\epsilon^2 \sim N(\boldsymbol{Z}\boldsymbol{\beta}, \sigma_\epsilon^2 \boldsymbol{R})$, having written $\sigma_\epsilon^2 = \gamma_\epsilon(0)$,

and defined $\boldsymbol{R}$ as the $n \times n$ matrix with (s,t) entry $\rho_\epsilon(|s-t|)$. The data-generating density under this model is

$$p(\boldsymbol{x}|\boldsymbol{\beta},\sigma_\epsilon^2) \;=\; (2\pi\sigma_\epsilon^2)^{-n/2}|\boldsymbol{R}|^{-1/2}\exp\left\{-\frac{1}{2\sigma_\epsilon^2}(\boldsymbol{x}-\boldsymbol{Z}\boldsymbol{\beta})^T\boldsymbol{R}^{-1}(\boldsymbol{x}-\boldsymbol{Z}\boldsymbol{\beta})\right\},$$

wherein $|\cdot|$ denotes the absolute determinant of a matrix. Maximum likelihood estimators of $\boldsymbol{\beta}$ and σ_ϵ^2 are

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{x} \quad \text{and} \quad \hat{\sigma}_\epsilon^2 = SSE/n,$$

respectively, where

$$SSE = (\boldsymbol{x}-\boldsymbol{Z}\hat{\boldsymbol{\beta}})^T\boldsymbol{R}^{-1}(\boldsymbol{x}-\boldsymbol{Z}\hat{\boldsymbol{\beta}})$$

is the revised sum-of-squared-errors statistic. Substituting these statistics into the general AIC formula yields

$$AIC \;=\; n\log(2\pi SSE/n) + \log|\boldsymbol{R}| + n + 2p,$$

as can be seen from

$$\begin{aligned} -2\log p(\boldsymbol{x}|\hat{\boldsymbol{\theta}}^{MLE}) &= n\log(2\pi\hat{\sigma}_\epsilon^2) + \log|\boldsymbol{R}| + (\boldsymbol{x}-\boldsymbol{Z}\hat{\boldsymbol{\beta}})^T\boldsymbol{R}^{-1}(\boldsymbol{x}-\boldsymbol{Z}\hat{\boldsymbol{\beta}})/\hat{\sigma}_\epsilon^2 \\ &= n\log(2\pi SSE/n) + \log|\boldsymbol{R}| + n. \end{aligned}$$

When calculating the AIC statistic, the number of parameters, p, in the general formula is to include the number of parameters that define the autocorrelated residual-error time series. For example, if (ϵ_t) is specified as autoregressive $AR(1)$ time series, then the total number of parameters is $p = r + 1 + 1$, which accounts for the r regressor variables, the autoregressive parameter, ϕ_1, and the white-noise variance parameter, σ_w^2.

A procedure that is sometimes helpful for choosing among possible models is to assess each model in terms of its predictive performance, and select the model that appears best. For this purpose, AIC offers one criterion for predictive performance; it would be calculated on each model that is under consideration, and the preferred model would be the one that produces the smallest value of the diagnostic.

Example: Exploring average global temperatures (continued)

The following table lists the values of AIC for two sets of de-trending models for the global temperatures data, the first such that the residual deviations are specified as Gaussian white noise, and the second such that the residuals are an $AR(1)$ time series with the autoregressive parameter estimated by moment-matching. Within each set, the de-trending model involves a mean function that is specified as either a linear, quadratic, or cubic polynomial.

	AIC	
μ_t	white noise	$AR(1)$
linear	-406.361	-471.859
quadratic	-437.110	-474.817
cubic	-446.119	-475.140

Observe that the most complicated model, that with autocorrelated residual deviations and a cubic polynomial mean function, achieves the lowest value of the AIC diagnostic. □

There is a great variety of measures of predictive performance beyond AIC, none of which are universally agreed-upon as most preferred. Two additional statistics are given below that arise within the Bayesian setup. When choosing among models, some analysts may prefer to focus on AIC, while a different group of analysts may prefer to focus on a different statistic. A widely recommended guideline is to examine several statistical criteria for predictive performance, at least to assess their degree of consistency in the preferred models they indicate, and perhaps to identify a small handful of similar models that appear to exhibit good predictive performance in multiple diagnostic criteria.

To describe the Bayesian predictive-performance statistics, let us suppose the analysis is to be implemented by numerical simulation, and that $\boldsymbol{\theta}^{(0)}, \ldots, \boldsymbol{\theta}^{(m-1)}$ are the sample values produced in that simulation to represent a posterior distribution with density $p(\boldsymbol{\theta}|\boldsymbol{x})$. Let us further suppose that the simulated values produced in the first $b-1$ iterations are discarded, as would be the case when the simulation is an MCMC algorithm whose burn-in phase ends at iteration $b-1$.

The *deviance information criterion* (DIC) has the formula

$$DIC \quad = \quad -2 \log p(\boldsymbol{x}|\hat{\boldsymbol{\theta}}^{BAY}) + 2p_{DIC},$$

where $\hat{\boldsymbol{\theta}}^{BAY}$ and p_{DIC} are computed as follows. If the parameter is multi-dimensional, so $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$, then the i'th entry, $\hat{\theta}_i^{BAY}$, of $\hat{\boldsymbol{\theta}}^{BAY}$ is the posterior mean,

$$\hat{\theta}_i^{BAY} \approx E[\theta_i|\boldsymbol{x}] \approx \frac{1}{m-b} \sum_{k=b}^{m-1} \theta_i^{(k)},$$

where $\theta_i^{(k)}$ is the i'th entry of $\boldsymbol{\theta}^{(k)}$. The quantity p_{DIC} is called the *effective number of parameters*, and has the computational formula

$$p_{DIC} \quad \approx \quad 2 \left(\log p(\boldsymbol{x}|\hat{\boldsymbol{\theta}}^{BAY}) - \frac{1}{m-b} \sum_{k=b}^{m-1} \log p(\boldsymbol{x}|\boldsymbol{\theta}^{(k)}) \right)$$

The connection between p_{DIC} and a model's number of parameters is not obvious, and is made in the deeper mathematical motivation for this criterion, which is beyond our scope. It is, in any case, a feature of the DIC statistic that it automatically incorporates a concept of "number of parameters."

A similar statistic is the *Watanabe-Akaike information criterion* (WAIC), which has the computational formula

$$WAIC \quad \approx \quad -2 \sum_{t=1}^{n} \log \left(\frac{1}{m-b} \sum_{k=b}^{m-1} p(x_t|\boldsymbol{\theta}^{(k)}) \right) + 2p_{WAIC},$$

where its effective number of parameters is

$$p_{WAIC} \approx 2\sum_{t=1}^{n}\left(\log\left(\frac{1}{m-b}\sum_{k=b}^{m-1}p(x_t|\boldsymbol{\theta}^{(k)})\right)-\frac{1}{m-b}\sum_{k=b}^{m-1}\log p(x_t|\boldsymbol{\theta}^{(k)})\right)$$

Observe that the WAIC statistic is calculated from the marginal data-generating densities $p(x_t|\boldsymbol{\theta})$, and is thus said to assess "pointwise" predictive performance.

Observe that the form AIC and each of the Bayesian predictive criteria is a sum of two terms, which we can write generically as $IC = -2\log L^* + 2p^*$, where L^* and p^* indicate "likelihood" and "number of parameters," respectively. The likelihood term may be thought of as measuring the "fit" of the model to the data; mathematically, we know that a complex enough model—one that involves many parameters—can fit the data so well that it essentially just connects the data-points to one another without reflecting intuition about uncertainty. A rough interpretation of a predictive criterion's mathematical form is that it accounts for the property that a model that incorporates a larger number of parameters would produce a larger value of "likelihood," which is desirable, but it could also include a subset of parameters that are spurious, in the sense that they do not contribute to predictive performance. The "number of parameters" term is thus interpreted as a penalty that would accrue with the admission of spurious parameters in the model. Operationally, the form of such a criterion implies that a model with a smaller value of the diagnostic would be preferred.

The following example demonstrates the expanded list of criteria.

Example: Exploring average global temperatures (continued)

The following table extends the previous table to include models that account for autocorrelated residual deviations, and the statistics DIC and $WAIC$.

	AIC		DIC		WAIC	
μ_t	white noise	$AR(1)$	white noise	$AR(1)$	white noise	$AR(1)$
linear	-406.361	-471.859	-171.4	-229.6	-171.6	-169.4
quadratic	-437.110	-474.817	-204.1	-240.4	-204.3	-203.3
cubic	-446.119	-475.140	-215.1	-244.1	-215.4	-214.4

All diagnostics indicate that the model with a cubic term and autocorrelated errors yields the best predictive performance. □

Section 5A.2: Seasonality.

Having now discussed predictive criteria that would be used in model building, let us return to the topic of model building itself. Recall the refined decomposition of the time series, (x_t), given by

$$x_t = \mu_t + s_t + \epsilon_t,$$

where μ_t is trend, (ϵ_t) is a stationary time series of residual deviations, and s_t is a seasonal term that captures periodic behavior. Our present focus is on the seasonal term.

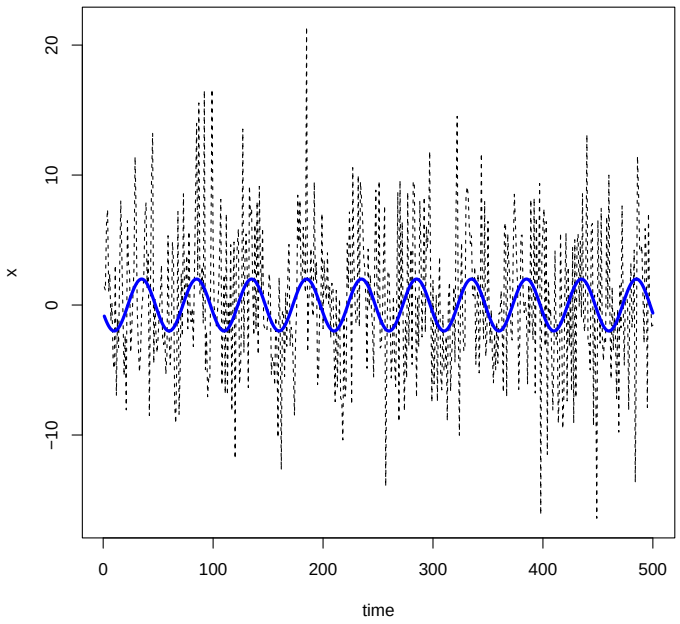
A starting point to modeling seasonality is to consider the simple periodic function with the form

$$s_t \quad = \quad A\cos(2\pi\omega t + \phi).$$

Related terminology describes the quantity A as the amplitude of s_t, ω as its frequency, $1/\omega$ as its period, and ϕ as its phase.

Example: Simulated time series with a periodic signal

The plot below displays a simulated time series that is strongly influenced by a periodic signal. The time series is measured at $n = 500$ time points, and the signal has a period of 50 units.



Consider the regular periodic time series at the center of the plot, displayed as a thick line. This time series is calculated according to the formula $s_t = 2\cos(2\pi t/50 + 3/5\pi)$. Its amplitude, $A = 2$ is the maximum height its graph reaches above or below zero. Its period, $1/\omega = 50$ is the number of time units the series has moved through when its cyclic pattern begins to repeat. Its frequency, $\omega = 1/50$, is the proportion of the length of that cyclic pattern that the time series moves through per time unit. Its phase, $\phi = 3/5\pi$, defines the starting and ending times of the pattern, shifted along the horizontal axis. □

Additional insight into this form can be gained by applying the trigonometric identity

$$\cos(\alpha + \beta) \quad = \quad \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta),$$

from which it is deduced that

$$\begin{aligned} s_t &= A\cos(2\pi\omega t + \phi) \\ &= A\cos(\phi)\cos(2\pi\omega t) - A\sin(\phi)\sin(2\pi\omega t) \\ &= \beta_1\cos(2\pi\omega t) + \beta_2\sin(2\pi\omega t), \end{aligned}$$

where

$$\beta_1 = A\cos(\phi) \quad \text{and} \quad \beta_2 = -A\sin(\phi).$$

The interesting perspective that follows from such derivation is that s_t has been written as a linear function of the variables $z_1 = \cos(2\pi\omega t)$ and $z_2 = \sin(2\pi\omega t)$, with coefficients β_1 and β_2. In other words, it is possible to understand the function $A\cos(2\pi\omega t + \phi)$, which is *non-linear* in A and ϕ, by transforming it to linear function, a form with which we are already familiar from our study of linear regression models.

Taking this further, additional properties come to light when the study of periodicities is set up as a regression problem. Suppose the time series (x_t) is measured at time points $t = 1, \ldots, n$, and $\omega_j = j/n$, for some integer $j = 0, \ldots, (n-1)/2$. The regressor variables are

$$z_{t1} = \cos(2\pi\omega_j t) \quad \text{and} \quad z_{t2} = \sin(2\pi\omega_j t).$$

As it turns out, formulas for the associated least-squares fitted regression coefficients are surprisingly easy to write down:

$$\hat{\beta}_1 = \frac{2}{n}\sum_{t=1}^{n} x_t \cos\left(2\pi t\omega_j\right) \quad \text{and} \quad \hat{\beta}_2 = \frac{2}{n}\sum_{t=1}^{n} x_t \sin\left(2\pi t\omega_j\right).$$

It is also the case that if include additional pairs of regressor variables, each defined at a different value of the integer $j = 0, \ldots, (n-1)/2$, the least-squares fitted regression coefficients remain easy to write down. That is, suppose r distinct values $j_1, \ldots, j_r = 0, \ldots, (n-1)/2$ define r frequencies, $\omega_{j_k} = j_k/n$, and $2r$ pairs of response variables

$$z_{t1}(\omega_{j_k}) = \cos(2\pi\omega_{j_k} t) \quad \text{and} \quad z_{t2} = \sin(2\pi\omega_{j_k} t),$$

for $k = 1, \ldots, r$. The formulas for all of the associated least-squares fitted regression coefficients exactly parallel the formulas displayed above:

$$\hat{\beta}_1(\omega_{j_k}) = \frac{2}{n}\sum_{t=1}^{n} x_t \cos\left(2\pi t\omega_{j_k}\right) \quad \text{and} \quad \hat{\beta}_2(\omega_{j_k}) = \frac{2}{n}\sum_{t=1}^{n} x_t \sin\left(2\pi t\omega_{j_k}\right),$$

for $k = 1, \ldots, r$.

In addition, it is suggested from this type of calculation that the sum of the two squared fitted regression coefficients defined within a pair corresponding to a particular frequency ω_{j_k} provides a sense of the *overall strength* of the periodicity exhibited within (x_t) at frequency ω_{j_k}. The relevant formula is expressed below as the definition of a statistical tool for explore periodicities:

Definition: The *(scaled) periodogram* of the time series (x_t) is the set of values

$$P(\omega_j) \quad = \quad \hat{\beta}_1^2(\omega_j) + \hat{\beta}_2^2(\omega_j)$$

where evaluated across all $\omega_j = j/n$ such that $j = 0, \ldots, (n-1)/2$.

A continuation of the previous example demonstrates these ideas.

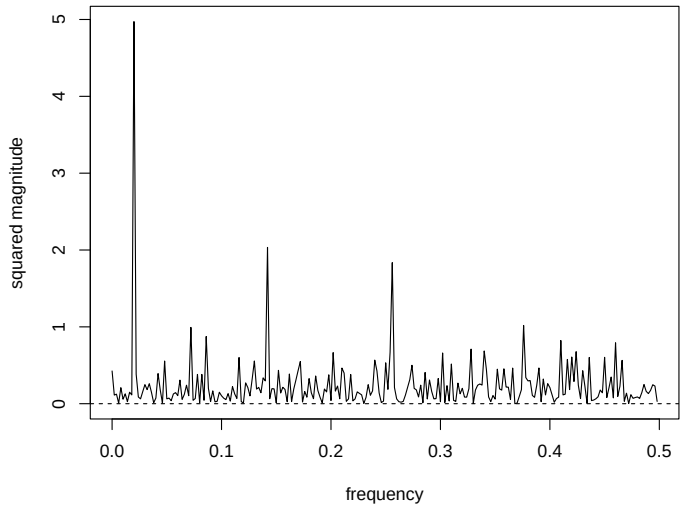
Example: Simulated time series with a periodic signal (continued)

An estimate of the underlying mean function is overlaid with the data and highlighted with a bold solid line. Its formula is

$$\hat{x}_t \;=\; \hat{\beta}_1 \cos(2\pi\omega t) + \hat{\beta}_2 \sin(2\pi\omega t) + y_t.$$

where $\omega = 1/50$, $\hat{\beta}_1 = -0.7083$, and $\hat{\beta}_2 = -2.5473$.

A plot of the scaled periodogram is shown below.



Observe the sharp peak at $\omega = 1/50 = 0.02$, which reflects the presence of a strong periodic signal at that frequency. □

The possible addition of periodic term in the decomposition

$$x_t \;=\; \mu_t + s_t + \epsilon_t,$$

adds a layer of complexity when applying regression techniques within time-series analysis. Some possible choices of regression functions are, for example, polynomials

$$\mu_t \;=\; \beta_0 + \beta_1 t + \cdots + \beta_r t^r,$$

trigonometric functions

$$s_t \;=\; \alpha_{11}\cos(2\pi\omega_{j_1}t) + \alpha_{12}\sin(2\pi\omega_{j_1}t) + \cdots + \alpha_{s1}\cos(2\pi\omega_{j_s}t) + \alpha_{s2}\sin(2\pi\omega_{j_s}t),$$

or a combination of both

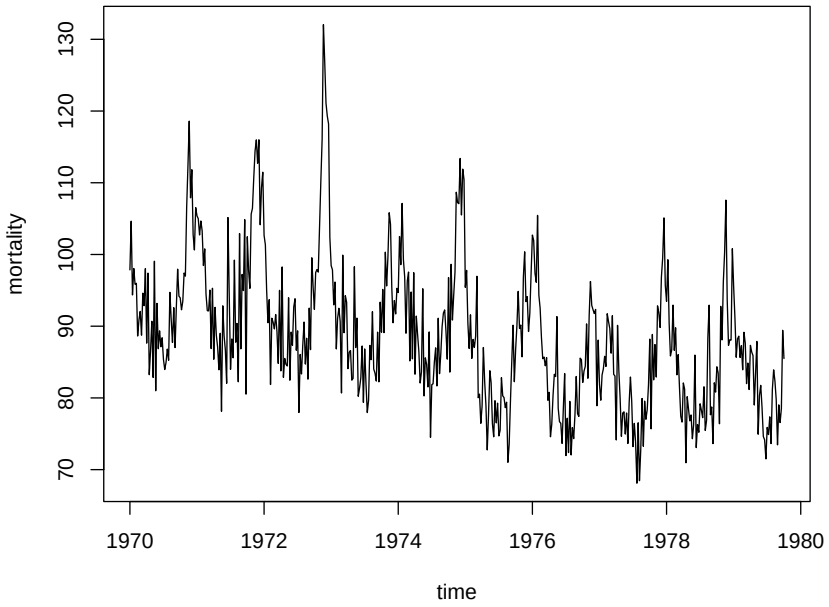
$$\begin{aligned}\mu_t + s_t \;=\;& \beta_0 + \beta_1 t + \cdots + \beta_r t^r \\ & + \alpha_{11}\cos(2\pi\omega_{j_1}t) + \alpha_{12}\sin(2\pi\omega_{j_1}t) \\ & + \cdots + \alpha_{s1}\cos(2\pi\omega_{j_s}t) + \alpha_{s2}\sin(2\pi\omega_{j_s}t).\end{aligned}$$

The possible presence of autocorrelated adds another dimension to the choices that might be considered during a model-building process. When selecting a suitable model from among the possibilities, the criteria for assessing predictive performance can play an important role.

The following example provides a simple illustration of a model-building process focusing on long-term trends.

Example: Cubic and periodic trends in cardiovascular mortality

The time series plotted below is of weekly measurements of cardiovascular mortality collected in Los Angeles over a ten year period.



Observe the variety of patterns in these data. Most striking are periodic oscillations, whose period is about one year, or about $1/\omega = 52$ weeks. However, there is also a slight decreasing trend, which may not be exactly linear.

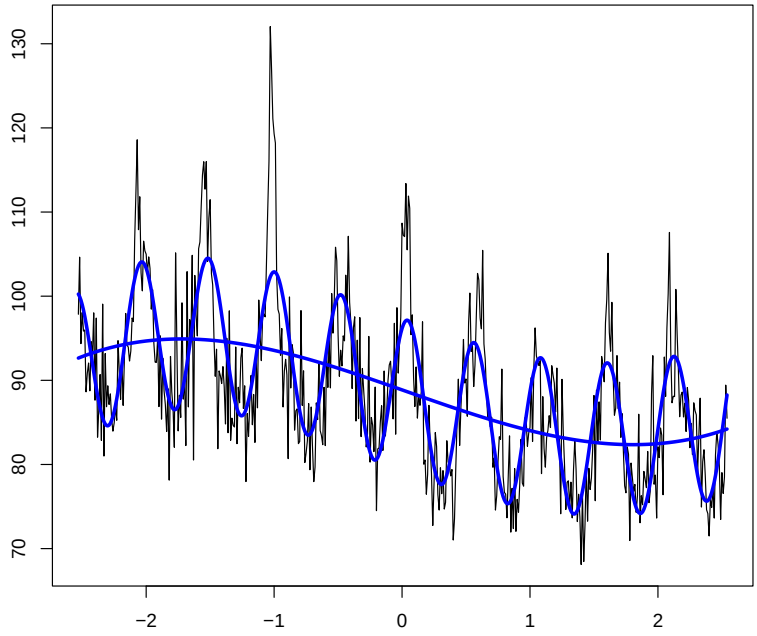
The patterns in these data are familiar enough to attempt to construct a model for the mean function, μ_t using familiar parametric functions. The periodic oscillations are modeled using a single pair of trigonometric functions (sine and cosine) at frequency $\omega = 1/52$, and the decreasing trend is modeled using a cubic polynomial. Putting these pieces together, the model for the mean function is

$$\mu_t \quad = \quad \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \alpha_1 \cos(2\pi\omega t) + \alpha_2 \sin(2\pi\omega t).$$

A graph of this function, fitted to the cardiovascular mortality series by least-squares regression, is shown below, superimposed on the data. Also shown in the plot is a graph of the fitted mean function without the periodic component. The form is

$$\mu_t \quad = \quad \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3,$$

and its effect in the plot is to highlight the long-term trend in the time-series.



The following table lists values of AIC, DIC, and WAIC that would be used to compare degrees of a polynomial fit for the long-term trend as well as to assess the impact of including a seasonal component.

	no seasonality			with seasonality		
μ_t	AIC	DIC	WAIC	AIC	DIC	WAIC
linear	2732.3	3661.9	3662.5	2381.3	3306.9	3307.8
quadratic	2734.2	3661.8	3662.3	2382.6	3306.2	3307.1
cubic	2723.1	3648.7	3649.1	2345.1	3266.8	3267.6
quartic	2722.7	3646.4	3646.8	2346.0	3265.7	3266.4

Here one sees at each degree of polynomial that adding the periodic component drastically decreases AIC. Across the varying degrees, one sees that as the polynomial increases in complexity, AIC decreases from linear to quadratic and from quadratic to cubic, but then increases from cubic to quartic. The minimum AIC is thus exhibited at the form of mean function originally presented, that for a cubic trend plus a periodic component. Both the DIC and WAIC statistic suggest that predictive performance would be improved by using the more complex model, that with a cubic trend and periodicity. □

As an additional note regarding the seasonality term, s_t, the regression interpretation encourages a perspective by which the coefficients $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{s1}, \alpha_{s2}$ in

$$s_t = \alpha_{11}\cos(2\pi\omega_{j_1}t) + \alpha_{12}\sin(2\pi\omega_{j_1}t) + \cdots + \alpha_{s1}\cos(2\pi\omega_{j_s}t) + \alpha_{s2}\sin(2\pi\omega_{j_s}t),$$

are constant parameters that are to be estimated. From this point of view, the term s_t would have the interpretation as a component that is supplemental to μ_t in describing long-term trends within the time series.

A separate perspective, which we will carefully explore in the next learning unit, is that the coefficients α_{11}, α_{12}, …, α_{s1}, α_{s2} are random variables. From this point of view, the term s_t

defines a distinct time series (s_t). Furthermore, it can be shown that, under straightforward conditions, the time series (s_t) is *stationary*. This suggests that the term s_t would have the interpretation as a component that is supplemental to (ϵ_t) in describing stationary residual deviations from long-term trends. This second interpretation is consistent with the definition $\mu_t = E[x_t]$, which implies that $E[s_t] = E[\epsilon_t] = 0$.