3.10.2 Gas furnace series

We next explain how a simple adjustment to the methods described can improve the estimates of the dependence between two series when there is an appreciable time lag in the relationship between them. The example we use for illustration is the series of input and output measurements on a gas furnace that have been previously modeled by spectral analysis in Jenkins and Watts (1968), and by lagged correlation methods in Box and Jenkins (1970). They are shown in Figure 3.18 but with the Input series negated so that the relationship between the series is positive. It is easy to see a strong similarity in the patterns of the two series, and the slight lag of variations in the Output, behind those in the Input. Figure 3.19(a) shows the lagged response between the series, estimated using a bandwidth of 0.05. The response peaks at a lag of 5 sampling intervals, which is also the lag at which the cross-correlation between the series has its maximum. Figure 3.19(b) shows the corresponding estimated squared coherency, confirming a strongly significant relationship. These estimates can, however, be improved if we first align the series by lagging the Input by 5 sampling intervals. The lagged response estimated after this alignment is shown in Figure 3.19(c) after adjusting the lag to restore the correct alignment.

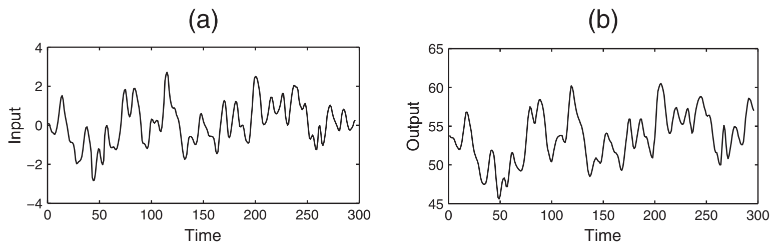

Figure 3.18 *Gas furnace time series: (a) Input and (b) Output.*

Notice that the peak in the response becomes more pronounced and the significance limits are centered on the chosen alignment lag. However, the more dramatic revision is to the estimated squared coherency shown in Figure 3.19(d). This is now much closer to 1 at frequencies below 0.05. The reason for the improvement can be seen from the plots in Figure 3.20 of the estimated gain and phase for this example. The gain reduces by a relatively small factor over this low frequency range, but the phase increases rapidly. The coefficients in the relationship between the harmonic components, as presented in the harmonic regression equation (3.32), are therefore rapidly varying. This regression cannot explain such a high proportion of the variability using a fixed pair of coefficients across the bandwidth, leading to a reduction in the squared coherency.

An appropriate value for the alignment lag is the slope of the estimated phase close to the origin, which in this example, by visual inspection of Figure

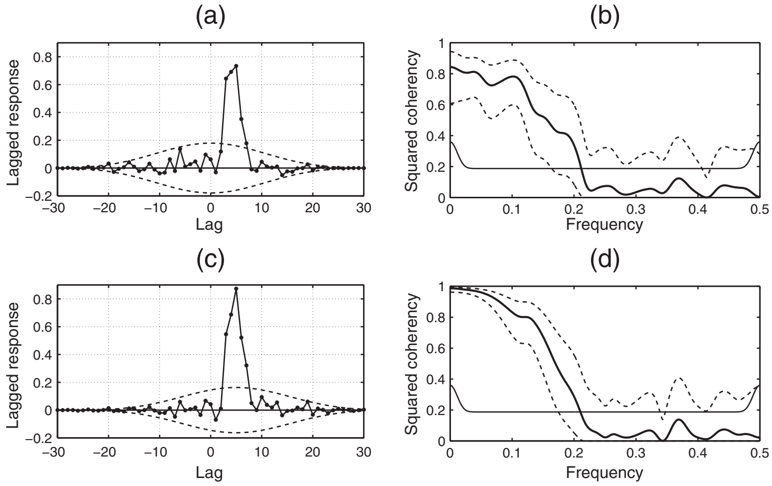

Figure 3.19 *Spectral estimates for the Gas furnace series: (a) the lagged response from Input to Output and (b) the squared coherency between the series, both estimated without alignment. (c) The lagged response from Input to Output and (d) the squared coherency between the series, both estimated with alignment.*

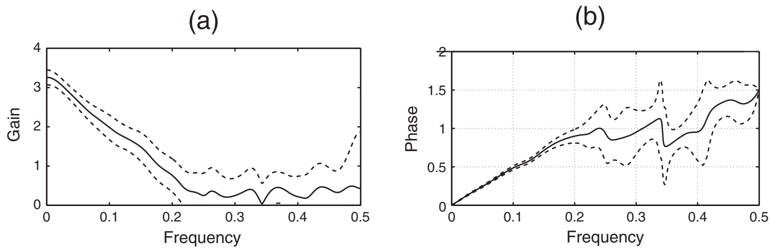

Figure 3.20 *Spectral estimates for the Gas furnace series: (a) the gain of the response from Input to Output and (b) the phase.*

3.20(b), is seen to be 5. Aligning the series by a lag ℓ subtracts the linear term ℓf from the phase, where f is the frequency. The coefficients in the harmonic regression are then more nearly uniform at low frequency, which improves the estimation of squared coherency, gain and phase. The phase estimated after alignment of the series is then corrected by adding on the term ℓf, restoring an improved estimate of the phase for the series prior to alignment. The overall result can also be achieved by centering the lag window w_k on the cross-covariance between the series at lag ℓ instead of lag zero.

3.10.3 Weekly moth trappings

We now consider issues illustrated by the investigation of how the numbers of Lepidoptera (moths) trapped each week are dependent upon two climate variables, temperature and rainfall. Figure 3.21(a) shows the Moth counts for the 345 weeks beginning in the first week of 1993 at Drayton in the UK. These are the daily average numbers of moths trapped. Figure 3.22 shows two climate variables for the same times and place. The Temperature series is the average daily temperature, and the Rainfall series is the average daily rainfall. The main practical point in this example is the importance of correcting for large fixed cycles arising from seasonal patterns in both explanatory and response series. A further consideration, which arises more generally in regression modeling, is the appropriate use of data transformations.

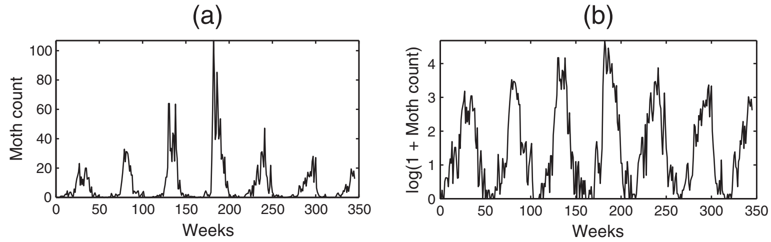

Figure 3.21 *Daily average numbers of moths trapped weekly at Drayton from the first week of 1993: (a) original counts, (b) the counts transformed by adding one then taking logarithms.*

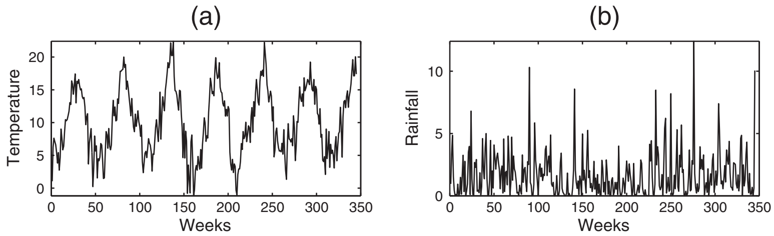

Figure 3.22 *Climate variables recorded weekly at Drayton from the first week of 1993: (a) a daily air temperature in degrees Celsius, (b) average daily rainfall in mm.*

Both the Moth count and the Temperature series show a strong annual cycle with peaks in summer and troughs in winter. Inevitably, this leads to a strong correlation between these two series, but this does *not* necessarily characterize the dependence in which we are interested. A similar strong correlation would be found if we used temperatures from almost any mid-latitude northern hemisphere site. Our interest is in discovering how the local temperature and rainfall affect the variations in moth trappings from week to week, and for this reason we will *correct* both the Moth count and Temperature series for a fixed annual cycle. However, before applying this correction to the Moth count, we note the marked asymmetry of this series, with much higher variability and a sharper peak in the summer than in the winter. The standard statistical approach to this issue is to consider a transformation of the series. Figure 3.21(b) shows the Moth count after applying a logarithmic

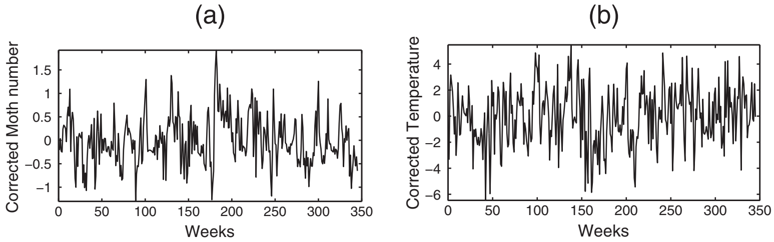

Figure 3.23 *Series corrected for annual cycles: (a) weekly Moth number, (b) weekly Temperature.*

transformation which results in a much more symmetric annual pattern of variation. We will call this transformed series the Moth number.

To correct the Moth number and the Temperature series for the annual cycle, we regress each upon the set of functions $\cos(2\pi kt/p)$ and $\sin(2\pi kt/p)$ and subtract the fitted values. The corrected series are shown in Figure 3.23. The period p used to define the regressors is $365.24/7 = 52.18$ weeks in the year. The regressor for $k = 1$ defines the fundamental annual cycle but the regressor with $k = 2$, for the second harmonic of this cycle, was also included for each series. Figure 3.24 shows how this decision was made for the Moth number; the same applied to the Temperature series. Figure 3.24(a) is the sample spectrum of the series with the prominent peak indicating the fundamental seasonal cycle. Figure 3.24(b) is the sample spectrum after the series has been corrected for this cycle. Again there is a peak that is prominent, though much less so than before, and now at the second seasonal harmonic frequency of 0.0383. One might also imagine peaks at the third and fourth harmonic frequencies (0.0575, 0.0767) of the annual cycle. These are, however, comparable to the general level of the neighboring sample spectra and will not, individually, contribute substantially to the variance of the series or the correlation between them. Correcting for the first fundamental and second harmonic of the seasonal cycle is sufficient to ensure that these features do not dominate in our modeling of the dependence between the series.

These peaks in the spectrum corresponding to the annual cycles are known as discrete spectral components. Failure to correct for these can distort the analysis, as we have just pointed out and will illustrate below for the coherency estimates. Note also that if a series has a non-zero mean that is *not* corrected, this will be evident as a discrete spectral component at frequency zero and can similarly lead to distortion of the analysis.

Figure 3.25 shows the estimated lagged response in the dependence of the corrected Moth number on the corrected Temperature series and the Rainfall series. Note that there is no lag dependence in the first of these. The Temperature has only a simultaneous positive effect on the Moth number. In contrast, the Rainfall has an extended lagged negative effect both in the immediately following weeks and also after a lag of about 7 weeks, associated with the typical summer breeding cycle of the moths.

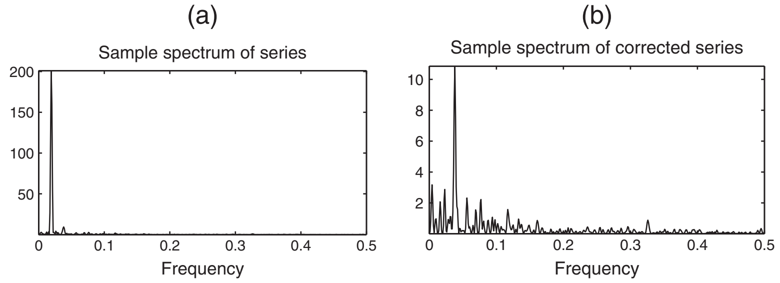

Figure 3.24 *Sample spectrum of Moth number: (a) before correction, (b) after correction by the fundamental annual cycle.*

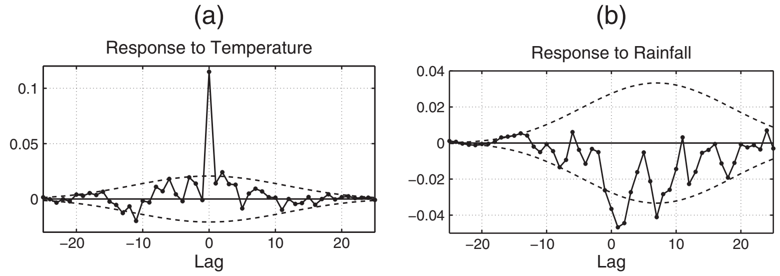

Figure 3.25 *Estimated responses of corrected Moth number to (a) corrected Temperature and (b) Rainfall.*

We do not show the estimated responses using the Moth number and the Temperature series *without* correcting for the seasonal cycle. They do look very similar to those shown in Figure 3.25 for the corrected series. To understand this, Figure 3.26(a) shows a scatter plot of the harmonic contrasts, both cosine and sine, of the Moth number against those of the Temperature, *before* correction. The plot shows only the harmonic frequencies, and we have restricted the series length to 313, corresponding to six annual cycles. The contrasts corresponding to the fundamental frequencies of these cycles are then at exactly the sixth harmonic frequency. The two outliers in this plot correspond to these cycles, all other points clustering close to the origin. Figure 3.26(b) shows the plot when the outliers are removed. In fact, the slope of the regression is quite close for both plots, indicating that the effects of temperature on the larger scale of the annual cycle are very similar to those on the small scale of week to week variations. The wisdom of correcting for annual cycles is that, were the effects *not* similar on these two scales, the estimation of the week to week response would be distorted by including the dominating effect of the large magnitude annual cycle.

To conclude this example, Figure 3.27 shows the partial coherency between the Moth numbers and the Temperature series, both before and after correcting the two series for the annual cycle. The very high coherency between the uncorrected series at low frequencies is typical of that resulting from a strong

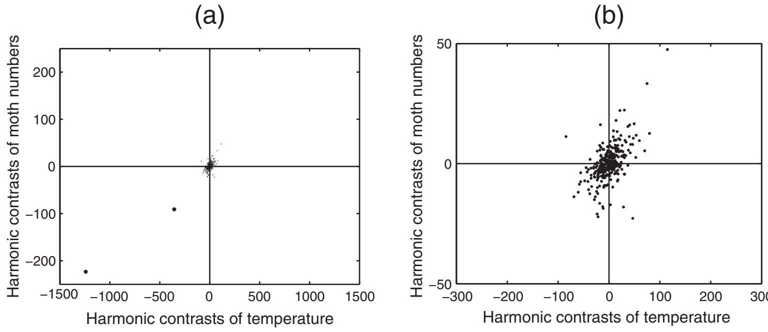

Figure 3.26 *Scatter plot of harmonic contrasts of the Moth numbers against those of the Temperature series: (a) showing all points and (b) scaled to exclude outliers.*

deterministic cycle of the same period in both series. In this example there is also strong coherency at higher frequencies to give assurance of significant dependence between the series. It is, however, always good statistical practice to correct for strong seasonal cycles before investigating the dependence between series.

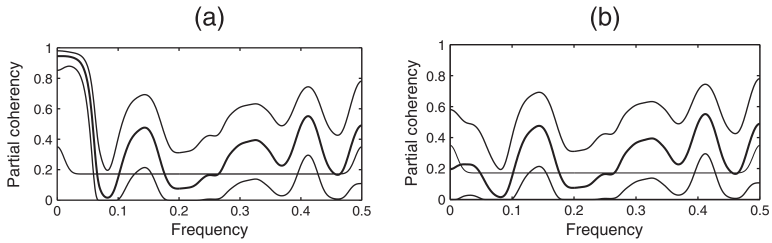

Figure 3.27 *Partial coherency between the Moth number and the Temperature series: (a) before correcting for the annual cycle and (b) after correction.*

3.10.4 Natural gas sendout series

Our next example also involves correction for seasonal effects, but also shows how an apparently small non-causal response at a negative lag could be explained by contamination of the explanatory variable by noise. Figure 3.28(a) shows the natural Gas sendout, as it is known, in a southeast region of England from 1st September 1977 to 31st August 1978. This should be the amount of gas consumed except that some is lost in transmission. Figure 3.28(b) shows a Meteorological indicator which is designed to help explain and possibly predict the Gas sendout. The main component of the indicator is temperature, but wind chill is also considered. The relationship between the series is clearly negative and quite strong, as the plot of one against the other shows in Figure 3.28(e).

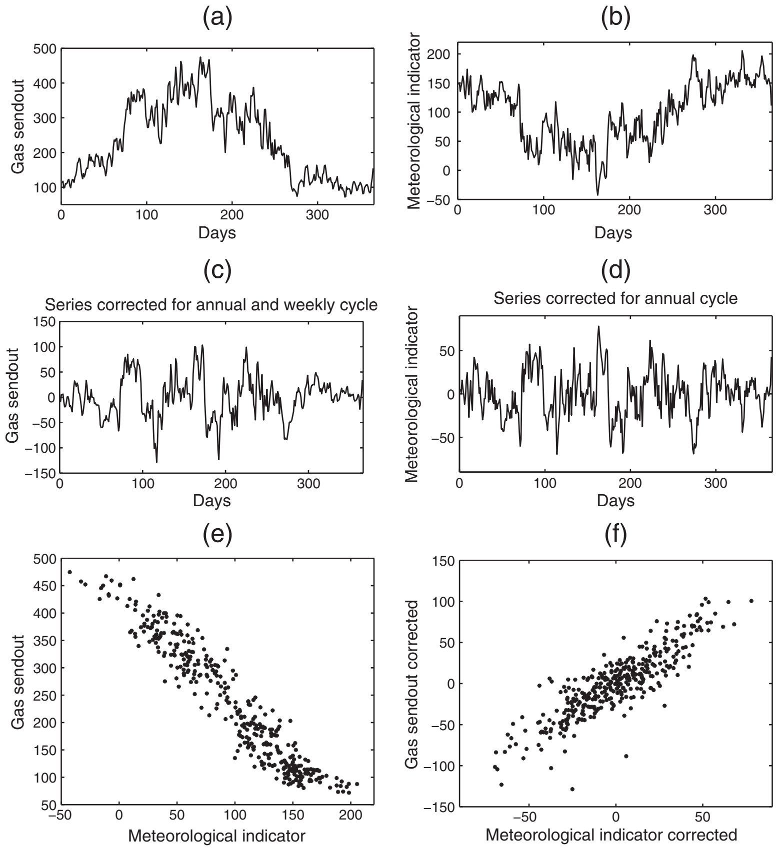

Figure 3.28 *(a) Daily Gas sendout over the period of one year. (b) A Meteorological indicator designed as a predictor of sendout. (c) The Gas sendout corrected for annual and weekly cycles. (d) The Meteorological indicator corrected for an annual cycle and reversed in sign. (e) The scatter plot of Gas sendout against the Meteorological indicator. (f) The scatter plot of the corrected Gas sendout against the corrected Meteorological indicator.*

Both series are corrected for an annual cycle of period 365.24. The Gas sendout series is also corrected for a weekly pattern of period 7 by regression on the complete set of cosine and sine cycles of frequencies 1/7, 2/7 and 3/7. Figures 3.28(c) and (d) show the corrected series with the sign of the Meteorological indicator reversed so that it becomes a direct predictor of Gas sendout, displaying a positive relationship. Figure 3.28(f) shows the graph of the corrected series against each other, with the appearance of a standard bivariate normal scatter. The small number of lower outliers arise from the Christmas period. A further correction could be made for these, but their presence will not greatly affect the analysis. Figures 3.29(a) and (b) show the estimated gain and lagged response function between these corrected series. The dominant feature of the lagged response is the coefficient of 0.915 at lag zero, but note that the gain at frequency zero is 90% greater at 1.75. There is a low level of response at positive lags which, though not apparently significant beyond lag 3, is cumulatively substantial. A unit step in the Meteorological indicator would imply an immediate increase in Gas sendout of 0.915, rising to a level of 1.75 over the next few days as lower temperatures gradually permeate structures. The regression slope of Figure 3.28(f) is 1.3, a compromise between these two values which would over-estimate the immediate impact and under-estimate the longer term.

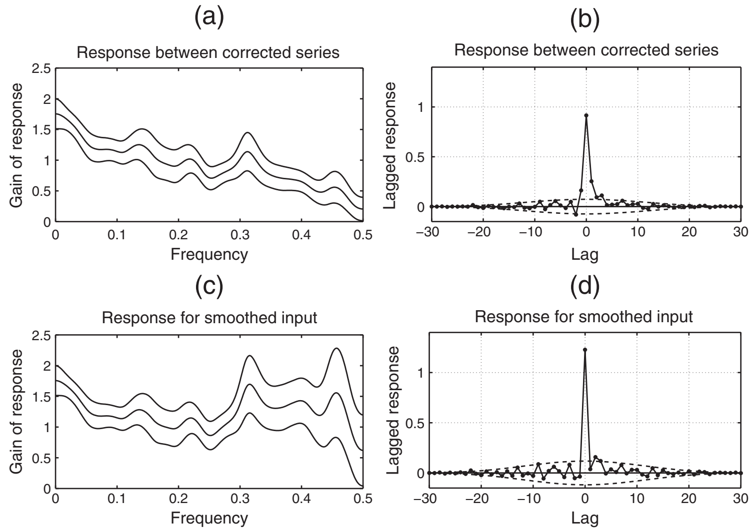

Figure 3.29 *(a) The gain of the frequency response and (b) the lagged response of the corrected Gas sendout to the corrected Meteorological indicator. (c) The gain of the frequency response and (d) the lagged response of the corrected Gas sendout to the smoothed corrected Meteorological indicator.*

However, we also note a small but quite significant response coefficient at a lag of -1. Both non-causal dependence and feedback are implausible explanations of this, but it could be explained if the Meteorological indicator were considered to be affected by noise in the form of measurement error. Figures 3.29(c) and (d) show the gain and response function re-estimated after applying a small amount of smoothing to the Meteorological indicator. The smoothing was a centered symmetric moving average of total span 5 lags, with weights of $(1-a)$ at lag 0, $4a/6$ at lags ± 1 and $-a/6$ at lags ± 2: the choice of $a = 0.25$ was made following a small amount of experimentation. This approximates a cubic spline smoother and is designed to preserve a cubic through these points whilst reducing the variance of added white noise. Note now that in Figure 3.29(d) there is no longer any significant non-causal response at negative lags and that the response at lag 0 has increased to 1.23.

The gain at zero frequency is unchanged at 1.75, because the smoother leaves the spectrum unchanged at this frequency. There is still therefore a 43% increase beyond the initial response in the eventual response to a step change of the Meteorological indicator.

Consider a response estimated from a stationary input series contaminated by white noise that is uncorrelated with both input and output. The noise-free input can be optimally estimated by a projection operator which reduces to the application of a smoothing filter. The estimated response to the contaminated input is given in the frequency domain as the product of the transfer function of the smoothing filter with the transfer function of the output response from the noise-free input series. In the time domain the lagged responses of the two are convoluted: the response to the contaminated input is given by applying the smoothing weights to the response to the uncontaminated input. This appears to be a plausible explanation of what is seen in this example.

3.10.5 Extrusion process series

Our next example illustrates how understanding the properties of the estimates of transfer functions in the frequency domain can help to improve the estimates of lagged regression. We estimate the response of the times series of Extrusion pressure to the Heater current and Valve setting shown in Figure 1.9 for the plastic Film extrusion process. Figure 3.30(a) shows the partial coherency between Heater current and Extrusion pressure, and Figure 3.30(b) shows the gain of the frequency response from Heater current to Extrusion pressure. Figure 3.30(c) shows the estimated lagged response from Heater current to Extrusion pressure.

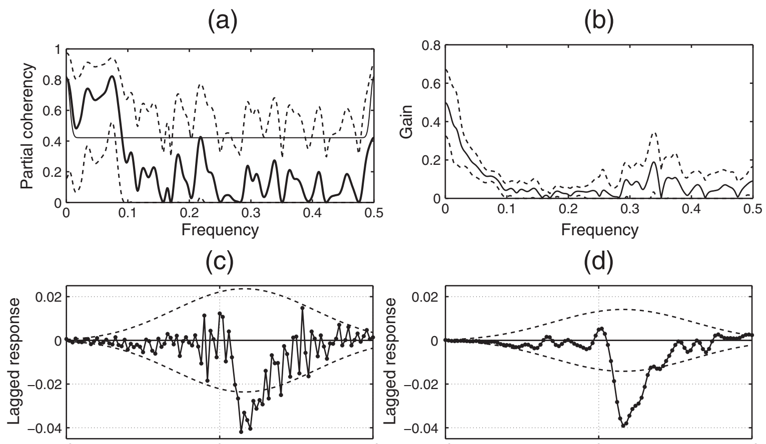

Figure 3.30 *(a) The partial coherency and (b) the gain of the frequency response from Heater current to Extrusion pressure. (c) The lagged response of the relationship and (d) the lagged response formed using a cutoff frequency of 0.2.*

Note that the partial coherency is only significant below frequency 0.1, and the gain estimates and standard errors fall away rapidly to frequency 0.2. However, above frequency 0.2, and in particular around frequency 0.35, the gain estimate is relatively large but poorly estimated with large standard errors, and not at all significant as assessed by the partial coherency. There is no significant (partial) dependence in this frequency range, but a relatively high variance of the harmonic contrasts in the response (Extrusion pressure) relative to those in the explanatory variable (Heater pressure) leads to these spuriously high estimates of gain. This effect increases the variability of *all* the lagged responses obtained by transformation of the frequency response, but it can be avoided by zeroing the estimated frequency response above a cutoff frequency, which we choose here as 0.2. The cutoff is tapered down to zero smoothly over an interval of twice the chosen bandwidth, centered on the cutoff frequency. The resulting lagged response is shown in Figure 3.30(d). It follows the same general shape as before this modification but is smoother in appearance and has much narrower significance bands. This cutoff procedure may be appropriate for similar process data where the sampling interval for recording the variables is chosen to avoid aliasing. The spectra and cross-spectra in the higher frequency range may then be low in magnitude and contain little information relating to the dependence between the series. The cutoff procedure then provides good estimates of lagged response without recourse to sub-sampling.

3.10.6 Sales series

The time series used in each of the previous examples can be considered to be jointly stationary processes. We now consider an example in which the weekly sales volume of a certain product, over a period of four years, is related to its price and a measure of spending on advertising promotions of the product. These series are shown in Figure 1.5, and none of them appears to be stationary. Both the Sales and the Price series have an obvious trend. The Promotion series varies considerably in magnitude with a twenty week period of zeros partway through. The Sales are also strongly affected by Christmas, and to investigate the impact of this season a further explanatory variable is introduced, an indicator variable taking the value of one in the Christmas week and zero otherwise. Brillinger (1981, Ch. 6) presents a careful treatment of the spectral analysis of series such as these. The methods we have been using can be extended to this context, with explanatory series that may be deterministic rather than stochastic and stationary, under two provisions:

1. The response series is stationary conditional upon the explanatory series; i.e., after removing the effects of the explanatory variable, the part remaining is stationary.
2. The harmonic contrasts of the explanatory variables are non-zero in any given bandwidth within the frequency range. These are required to explain the harmonic contrasts of the response, and a regression fails if the

regressors are zero. A clear exception to this provision is when an explanatory variable consists of one pure cycle (or the sum of a small number of cycles). The contrasts will be zero in any bandwidth that does not contain the frequency of this cycle. In this example the Christmas indicator can be represented as the sum of cycles at frequencies $k/52$ for integer k, but the chosen bandwidth of 0.1 contains several such frequencies so no problem arises.

Expressions for the large sample properties of the estimates of the transfer function are very slightly different for dependence upon deterministic series; see Brillinger (1981, p. 200, 201). The difference relates to the smoothing weights that should be applied to the sample spectrum of the dependent series. We will, however, use the same expressions as for dependence upon stationary series, because this makes very little difference in practical applications.

The Sales, Price and Promotion series are shown in Figure 1.5. It is not wise to assume that the downward trend in the Price explains the upward trend in Sales, and in a simple regression context this assumption would be avoided by adding a linear trend as a further explanatory variable. An equivalent step would be to subtract from each series its trend. We have applied that correction before our cross-spectral analysis, the results of which are presented in Figure 3.31.

This example is typical of market response modeling, in that the effects are not strongly significant, even though the Christmas effect may appear strong. We note that:

1. The response to Price appears to be simultaneous with the price variation, with a lag zero response that is just significant at -0.6. However, the gain in Figure 3.31(d) shows that the magnitude of the response is close to 2 at higher frequencies, suggesting that the reaction to frequent price changes is more substantial. At low frequency the gain is much smaller and barely significant, suggesting a longer term insensitivity to price change, possibly reflecting brand loyalty. A simple regression of Sales on Price, Promotion, the Christmas indicator and a linear trend gives a coefficient of -0.7 for price, whilst a regression between the *differenced series*, which accentuates the effect of changes, gives a coefficient of -1.1 on price. Simple regression, omitting lags and neglecting error correlation, delivers a compromise coefficient. The gain gives the more comprehensive picture of the response.
2. The response to Promotions is again barely significant but remains positive up to lag 10. The corresponding low frequency gain supports a more strongly significant long term effect, which is valuable marketing knowledge.
3. The response to Christmas is clearly significant, with the pattern indicating a build up for a week or two before hand and a rebound in sales a week or two afterwards. This is not surprising, but the gain at low frequency shows considerable uncertainty as to whether Christmas had any net effect upon total sales over the season.

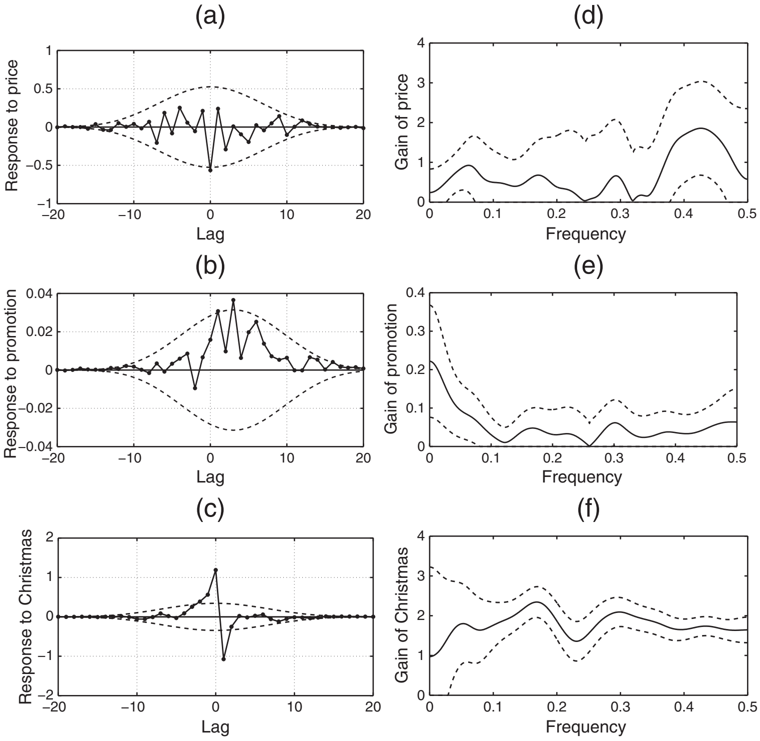

Figure 3.31 *The estimated responses of Sales to (a) Price, (b) Promotion and (c) Christmas, together with the frequency domain gain of (d) Price, (e) Promotion and (f) Christmas.*

There are many other factors, such as weather, that influence sales, resulting in much variability that is not accounted for by the chosen explanatory variables. We can derive the residual contrasts from the frequency domain regression and reconstitute the corresponding residual series which estimates the noise term n_t in the lagged relationship (3.23). This is the *discrepancy* between the Sales series and the part explained by the Price, Promotion and Christmas effect. Its estimated spectrum and autocorrelations can also be constructed. Figure 3.32(a) shows the Sales series corrected for linear trend together with this series of discrepancies, which confirms that there is much variability for which the model fails to account. The estimated autocorrelations of these discrepancies in Figure 3.32(b) decay over a period of two months. The model assumes time invariance for its coefficients, which may in fact vary from year to year. Other possible sources of variability of the discrepancy series may be suggested by its inspection. But even without further investigation, this analysis has provided a useful summary of the effects of the explanatory variables over the period of the data.

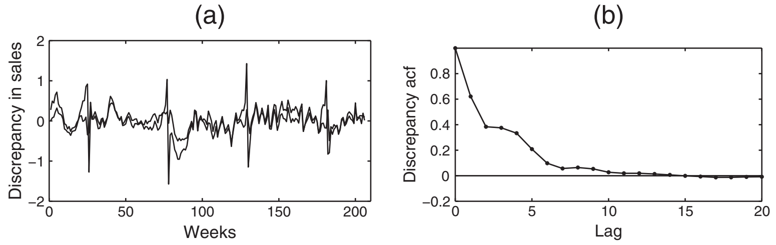

Figure 3.32 *(a) The discrepancy between the Sales series and the model (thick line) together with the mean and trend corrected Sales (thin line), and (b) the estimated autocorrelations of the discrepancy.*

An alternative to trend correction of the series is to use their (mean corrected) first differences in the analysis. The differencing must be applied to all the series to preserve the dynamic relationship between them. In large samples the estimated coherency, gain, phase and lagged responses are insensitive to this transformation; the differencing will, however, carry through to the discrepancy series. In this example, this alternative leads to very similar results. Whether trend correction or differencing is used, low frequency information is reduced, resulting in greater uncertainty in the estimation of the zero frequency gain that is of particular interest. This is unavoidable if we do not wish to risk distorting the estimated dependence of a trending series upon a trending explanatory variable, because of the omission of unknown trending factors which may also be part of the explanation.

3.10.7 Limitations of spectral analysis

Because spectral analysis is a form of non-parametric estimation, it has been thought necessary that the time series to which it is applied should be of a good length. However, as for other applications of non-parametric estimation, much depends on whether the spectrum has pronounced features which it is important to resolve, or if it is simple in form. In the latter case, spectral analysis might be usefully applied to series with length as low as 20, in which the dependence extends over a small number of lags.

Spectral analysis relies strongly on the assumption that the series are stationary. If applied to series in which the structure of lagged dependence is changing, it may fail to determine a significant relationship, or if it does, it will estimate some form of average dependence. This may be the case in the foregoing Sales example, where the impact of Promotions may change throughout the year. Similarly for the Moth count series, in the winter months when moth numbers are low, the effects of the weather may be quite different. Time varying spectral estimation may be carried out by applying the analysis to a moving window of the series. The graphic equalizer on much audio equipment shows the varying sound spectrum in different frequency bands. The topic of evolutionary spectra has been studied for many years (see, for example, Priestley (1965), Priestley (1981, Chap. 11) and Dahlhaus (1996)), but estimation of spectra is typically restricted to long, high-frequency sampled series. Wavelet analysis (Percival and Walden (2006)) is also used to analyze time-varying relationships. A recent practical approach to testing for second order stationarity and estimating local second order properties using wavelets is found in Nason (2013).

The description *non-stationary* is also applied to time series which are integrated stationary processes. Our approach to series which may be of this form has been to correct them for any regular trend but otherwise to apply the same spectral analysis as for stationary series. Evidence of integrated non-stationarity may be apparent in a spectral peak of very high power at frequency zero, though in a finite sample it is not possible to discriminate with certainty between an integrated process and a stationary process with red spectrum. Pre-whitening, possibly even in the form of differencing, may then improve the analysis. In practice, the estimates of lagged responses may not be greatly affected by this, as may be verified for the Sales series example.

In our examples we have also corrected series for fixed periodic annual and weekly cycles. This is not the same as applying a seasonal adjustment filter, which corrects further for slow variations in the seasonal pattern from year to year. Our correction removes harmonic components at the seasonal harmonic frequencies, corresponding to cycles with periods equal to the seasonal period or a simple fraction of that period. Seasonal adjustment also reduces the amplitude of harmonic components in the neighborhood of the seasonal harmonic frequencies. These may contain useful information about the relationship between the series. For example, if there is a trend to warmer summers, the moth trappings may increase correspondingly, but seasonal adjustment may remove this evidence. Many economic time series are seasonally adjusted, and the spectral estimation of their dependence may be adversely affected by this.

Spectral analysis is also directed at estimating linear lagged relationships. We determined a significant lagged response between variables of the Film extrusion process example, but this system almost certainly contains nonlinearities. These may not be large provided the variables do not depart far from their set point, and the linear relationships we identified may be sufficient to design successful control schemes.

Overall, spectral analysis is useful for demonstrating and quantifying linear dependence between time series. It is often of value for specifying the form of a linear parametric model for this dependence; this model may then be developed to allow for time variation and non-linearity if these features are found to be important. We refer again to Priestley (1981, Chap. 11) for models that may be considered in this context.

3.11 Harmonic contrasts in large samples

In the preceding sections we have presented the distributional properties of estimates of spectral quantities which are applicable in large samples. This