Seasonality, filters, linear regularity

Christopher A. Sims
Princeton University
sims@princeton.edu

December 4, 2017

Filtering

- If $Y_t = a * X_t$, then $\tilde{Y}(\omega) = \tilde{a}(\omega)\tilde{X}(\omega)$ and $S_Y(\omega) = |\tilde{a}(\omega)|^2 S_X(\omega)$.

- So to remove a peak in the spectral density of X, we choose an a such that $|\tilde{a}|$ is small over the band of frequencies in which the peak is large, replace X with $a * X$.

- Seasonal adjustment: doing this with bands around all the seasonal frequencies $2\pi j/S$, where S is the number of observations per year and j is an integer. (In continuous time, S is just the length of the year in our time unit.)

- This leaves a lot of room for various methods: width of band, how close to zero with $|\tilde{a}|$.

Optimal seasonal adjustment?

1. Since the non-seasonal variation that interests us presumably has a spectral density that is smooth across seasonal bands, why not adjust so the adjusted series has smooth spectral density across these bands?

2. Or, why not have a model: $X_t = X_t^N + X_t^S$, $X^N \perp X^S$, X_t^N non-seasonal, X_t^S entirely seasonal?

3. Or, why not just “wipe out” seasonal variation, setting $|\tilde{a}| = 0$ in seasonal bands?

There is no optimum

- All these approaches require adapting adjustment method to the series being considered.
- First two require both bandwidth and degree of damping to adapt.
- Last requires bandwidth to adapt.
- Second implies that adjusted series should have *dips* at the seasonal frequencies.

Dips?

$$\hat{X}_t^N = a * X_t\,, \quad a \text{ minimizes } E\big[(\hat{X}_t^N - X_t^N)^2)\big]\,.$$

- If a were of fixed finite length, we could calculate the least squares fit from knowledge of S^N and S^S, the spectral densities of the seasonal and non-seasonal components, because these would allow us to compute the autocovariance functions and cross-covariance functions of the components and of X itself.

- But it is easier to understand what is going on by translating to the frequency domain and thinking of making separate linear projections, frequency by frequency, of $\tilde{X}^N$ on $\tilde{X}$.

Frequency-by-frequency projection

- $\tilde{X}(\omega)$ is independent across frequencies.
- Project $\tilde{X}^N$ on $\tilde{X}$ at each frequency to get

$$\tilde{a}(\omega) = \frac{\mathrm{Cov}(\tilde{X}(\omega, \tilde{Y}(\omega))}{\mathrm{Var}(\tilde{X}(\omega))} = \frac{S_{X^N X}}{S_X}\,.$$

- This formula can't be interpreted literally because Fourier transforms of processes do not exist as random variables at individual frequencies.

A more careful statement

- If we want to project the random variable $\int c(\omega)\tilde{Y}(\omega)d\omega$ on the $\tilde{X}$ process by finding a function b that minimizes the variance of

$$\int c\tilde{Y}d\omega - \int b\tilde{X}d\omega\,,$$

  the optimal choice is $b = cS_{YX}/S_X$

- With $Y = X^N$, $X = X^N + X^S$, this gives us the formula from the previous slide.

- With c the Fourier transform of a time-domain function that truncates X to a finite time span, This tells us that the optimal time domain deseasonalization filter will be approximately the inverse FT of S_{X^N}/S_X.

Qualitative implications

- Since S_{X^N} is smooth across seasonal frequencies and S_{X^S} is near zero except for peaks in seasonal bands, $\tilde{a}$ will have dips at seasonal frequencies, with the size of the dips dependent on the ratio of the height of the peaks compared to the height of S_X at the neighboring non-seasonal frequencies.

- The minimum variance estimate of X based on a noisy observation of it will have lower variance than X itself, with the degree of damping of variance dependent on the ratio of noise to non-noise variance in the observation.

- In the seasonal bands, $\tilde{X}$ is a very noisy observation on $\tilde{X}^N$, so variance there of the best predictor will be much lower than the variance of $\tilde{X}^N$ itself.

Spectral density matrices

If X is an m-dimensional vector valued process, we can FT each of its components to obtain $\tilde{X}(\omega)$, where as before this has to be interpreted as $dZ(\omega)$, that is the differential of a complex-valued Gaussian process Z in the frequency domain that has independent increments. We can FT the autocovariance function R_X to obtain S_X, but now both $R_X(t)$ and $S_X(\omega)$ are $m \times m$ matrices. As in one dimension,

$$\mathrm{Var}\big(Z(\omega_2) - Z(\omega_1)\big) = \int_{\omega_1}^{\omega_2} S_X(\omega)\, d\omega\,.$$

Note that for a vector valued process it is no longer true that $R_X(t) = R_X(-t)$. But if we interpret the $'$ operator as, in the time domain, both transposing a matrix and reversing the sign of the time argument, we do have $R_X(t) = R'_X(t)$. In other words, $R_X(t)$ is the transpose of $R_X(-t)$. This means that $S_X(\omega) = S'_X(\omega)$, if in the frequency domain we maintain the convention that $'$ both transposes and takes the complex conjugate. A complex matrix A for which $A' = A$, where $'$ both conjugates and transposes, is called **Hermitian**.

The off-diagonal elements of S_X are just the FT's of the cross-covariance functions of the corresponding components of the X vector. The $S_X(\omega)$ matrices are always positive semidefinite, i.e. $c'S_X(\omega)c$ real and non-negative for all real or complex vectors c.

Relations among adjusted series

Suppose $Y = b * X + \varepsilon$, with X_t independent of ε_s for all t, s and $b_s = 0$ for $s < 0$. If b has just finitely many non-zero coefficients, we can estimate this consistently by OLS. But suppose instead we have to estimate it using seasonally adjusted series, $Y^* = a_Y * Y$ and $X^* = a_X * X$, where a_Y, a_X are seasonal adjustment filters.

Then $\tilde{b} = S_{YX}/S_X$, and this is the FT of what we will recover in large samples by OLS estimates using Y and X. If instead we use Y^* and X^*, we get

$$\tilde{b}^* = \frac{S_{Y^*X^*}}{S_{X^*}} = \frac{\tilde{a}_Y \bar{\tilde{a}}_X S_{YX}}{|\tilde{a}_X|^2} = \frac{\tilde{a}_Y}{\tilde{a}_X}\tilde{b}\,.$$

Note that if X has been seasonally adjusted, $\tilde{a}_X$ will be very small in seasonal bands, so if Y has not been seasonally adjusted, or has been seasonally adjusted "less aggressively" (meaning a_Y does not go so close to zero in the seasonal bands), $\tilde{b}^*$ will have peaks in absolute value at seasonal frequencies. In the reverse case, there will be seasonal dips in $\left|\tilde{b}^*\right|$.

The impossibility of $\tilde{a}(\omega) = 0$ over non-zero length intervals

Any discrete-time process X that can be represented as a one-sided (possibly infinite-order) MA (with, of course, square-summable weights) $X = a * \varepsilon$ is linearly regular, even if a has roots inside the unit circle. This must be true, because it is easy to see that, because of square-summability of a, the forecast error variance in predicting X_t from $\{\varepsilon_s, s < t - v\}$ must converge to $\mathrm{Var}(X_t)$ as $v \to \infty$, and we know that forecasts based on past ε's must be at least as good as those based on past X's. (The forecasts are the same if the representation is fundamental.)

So, given any one-sided filter b, we know that $\left|\tilde{b}\right|^2 = S_X$ for some linearly regular X and therefore that $\left|\tilde{b}\right|$ cannot vanish over any interval of non-zero length. So a filter that “wipes out” variation in seasonal bands cannot be one-sided.

In fact a filter a with $\tilde{a}(\omega) = 1$ outside seasonal bands and $\tilde{a}(\omega) = 0$ inside a band around each seasonal is necessarily two-sided and symmetric. While actual seasonal filters used in practice are not closely approximated by this limiting case, they are generally two-sided and symmetric. A crude example, sometimes used for a quick check: $a_0 = 1$, $a_s = -1/(2n)$ for $s \neq 0$ and $s = jS$ for integer j, and $|j| \leq n$, and $a_s = 0$ otherwise. Here S is the length of the season (e.g. 12 for monthly data). This makes the adjusted series the deviation between the current level and the average of corresponding months between n years ago and n years from now.

Obviously the adjustment filters cannot be two-sided at the start and at the end of the series, so special adjustments are made there.

Pathologies of autoregressive models with seasonally adjusted data

We have observed that seasonal adjustment is likely to create dips in the spectral density of the adjusted series at the seasonal frequencies. Where seasonality is strong, optimal adjustment according to the $X^N + X^S$ model would make these dips deep. Recall that the log of the one-step-ahead forecast error variance of a linearly regular process is $\exp\left(1/(2\pi)\right)\int \log S_X \, d\omega$. The integral in this expression can become much smaller if S_X is made to dip close to zero, even over a small interval, because $\log S_X$ approaches $-\infty$ as $S_X \to 0$. So seasonal adjustment can in principle create major distortions in one-step-ahead forecast error variance, which can be an important issue in rational expectations models. This should not be surprising. Seasonal adjustment applies a filter that makes the current value of the series depend on future as well as past values of the data. By exploiting the fact that the adjusted data contain information about the future, it can become possible to greatly reduce forecast error.

Why seasonal adjustment is not quite as bad as it looks

Consider the case of our regression model $Y = a * X + \varepsilon$. Suppose that instead of estimating a freely, we set up a parametric model in which a finite parameter vector θ determines a as $a(\theta)$, and we choose our parameterization so that $\tilde{a}(\theta)$ cannot have sharp peaks or dips at seasonal frequencies, no matter what the value of θ. Least squares estimates of θ will then choose $\hat{\theta}$ to minimize (using $\hat{a}$ to represent $a(\hat{\theta})$)

$$\mathrm{Var}(Y_t - \hat{a} * X_t) = \mathrm{Var}(\varepsilon_t) + \mathrm{Var}\big((\hat{a} - a) * X_t\big) = \mathrm{Var}(\varepsilon_t) + \int_{-\pi}^{\pi} \left|\tilde{\hat{a}} - \tilde{a}\right|^2 S_X(\omega)\, d\omega$$

In other words, the estimate of θ will be chosen to minimize a weighted average of squared errors in the frequency domain, with the weights given by the spectral density of X.

If we have chosen our parameterization well, we may hope that, using the true non-seasonal components of Y and X we would find a and $a(\hat{\theta})$ extremely close. If we instead must rely on seasonally adjusted data, the results may still be extremely accurate, because S_X will dip at the seasonal frequencies. Even though the parameterization constrains $\tilde{a}(\theta)$ to be smooth across the seasonal bands and thus probably not to match the a that best fits the adjusted data, the weight on errors at these frequencies will be small, so the estimated $\tilde{a}(\theta)$ is likely to match the true $\tilde{a}$ for these data well at non-seasonal frequencies, less well at seasonals, and thus on the whole to match fairly well what would be obtained with the unobservable non-seasonal component.

If we use the unadjusted data, but maintain our parameterization, the results are likely to be very bad. The unadjusted data have peaks at the seasonal frequencies, so the approximation error is weighted especially strongly there. The estimated $\tilde{a}(\theta)$ therefore is likely to match the least squares $\tilde{a}$ well at the seasonal frequencies, poorly at non-seasonal frequencies. It will therefore probably be very different from what would have been obtained with the unobservable non-seasonal data.

One can make a similar, slightly more subtle, argument concerning autoregressive models and forecast error variances, but we will not go through it explicitly here.

The conclusion is that if you know how to construct a parameterization that is likely to fit well with unobservable non-seasonal data, and if the parameterization makes sharp rises or falls in $|\tilde{a}|$ at seasonal frequencies impossible, then the approximation error involved in using seasonally adjusted data is likely to be small — and smaller the greater the reduction induced by the seasonal adjustment in variance in the seasonal bands.

There are several important "if"'s in this optimistic conclusion, however. It may not be easy to construct a parameterization with the required qualities. A simple low order MA operator $a(L)$ must be smooth across any narrow band of frequencies. But we will shortly be considering models of the ARMA form, $a(L)/b(L)$. Even if a and b are both low order, such models can have arbitrarily sharp peaks at arbitrary frequencies. If they are low order, they can have only a small number of such peaks, but seasonal effects can sometimes be concentrated at one or two frequencies.

ARMA models for seasonal data

$$Q(L^S)A(L)y_t = P(L^S)B(L)\varepsilon_t \,,$$

with $Q(1) \doteq 0$, $P(1) \doteq 0$, $P(Z)/Q(Z) \xrightarrow[Z\to 1]{} 1$.

- If A, B both low-order, $\left|A(e^{-i\omega})\right| = \left|\tilde{A}(\omega)\right|$ bounded away from zero except possibly near $\omega = 0$, then $\left|\tilde{B}\right| \,/\, \left|\tilde{A}\right|$ will be smooth, without sharp peaks, while our conditions on P and Q imply that $\left|\tilde{P}/\tilde{Q}\right|$ can have sharp peaks near seasonal frequencies.

- Often people use simple versions of this model, e.g. with no $P(L^S)$, $Q(L^S) = 1 - L^S$ or $Q(L^S) = (1 - L^S)^2$.

- Here, as in the frequency domain, it is important to distinguish between *eliminating* seasonality as a nuisance, and *modeling* seasonality. The $1 - L^S$ factor will correctly extrapolate any extremely persistent seasonal. It will not generally capture the nature of time variation in the seasonal.