

## *Univariate Filtering*

### **1. Introduction**

Most of the tools that we use in time series analysis are designed for the treatment of stationary processes. However, most economic data is non-stationary and often presents seasonal behavior. For example, U.S. GDP exhibits an upward trend over time that is consistent with a growing economy, seasonal behavior characterized by slow winters and summers and strong springs and falls, and a cyclical pattern of expansions and recessions. Most research in macroeconomics dedicated to the explanation of business cycles therefore relies on pre-filtering methods from which the trend and seasonality are removed only to isolate the cyclical components. Other disciplines in and outside economics often have similar pre-filtering requirements. However, it is important to highlight that the motivation for pre-filtering the data does not really come from statistical necessity imposed by estimation, testing or forecasting of time series models (although admittedly, stationary models are more tractable). Rather, it reflects the preference of applied researchers to focus on certain features of the data generating mechanism that occur with specific regularity.

Except for trending behavior, features such as seasonality, business cyclicity and noise reflect patterns that tend to repeat themselves with a certain frequency (in the case of quarterly data, a seasonal pattern reappears every fourth quarter, in terms of business cycles, expansions exhibit cycles of varying length that tend to last anywhere between two to eight years, etc.). It is therefore common to investigate issues of filtering in the frequency domain since this type of analysis more naturally lends itself to the decomposition of a time series into sums of “periodic patterns.” However, I will not present the discussion of filtering in terms of spectral analysis (the more general term used to refer to analysis in the frequency domain) here because it would require that I spend some time presenting the essential tools of this type of analysis (for a nice introduction see chapter 6 in Hamilton, 1994). However, note that most reviews of

filtering methods approach the subject from the frequency domain perspective and it is thus to your advantage to become acquainted with this set of tools.

Summarizing, filtering of data is done to:

- Remove trends
- Remove seasonality
- Remove other cyclical phenomena

You should be aware that as I present the different methods to filter trends, and seasonal effects, they are not mutually exclusive and are often used in combination.

## 2. Common Detrending Methods

### 2.1 Deterministic Time Trends

Perhaps the simplest de-trending method consists in specifying the trend as a polynomial in time,

$$\begin{aligned}y_t &= y_t^T + y_t^c \\y_t^T &= \sum_{i=0}^K \alpha_i t^i \\y_t^c &= u_t\end{aligned}\tag{1}$$

where  $y_t^i$  for  $i = T, c$  denotes the trend component and the cyclical component respectively (I will use this nomenclature rather loosely throughout),  $K$  is a finite integer and  $u$  is a stationary process. By far, the most common choice is to set  $K = 1$  so that

$$\begin{aligned}y_t &= \alpha_0 + \alpha_1 t + u_t \\ \psi(L)u_t &= \phi(L)\varepsilon_t\end{aligned}\tag{2}$$

where  $\varepsilon_t$  is white noise,  $\psi(L)$  is a stationary AR polynomial, and  $\phi(L)$  is a finite order moving average polynomial.

Remarks:

- Deterministic detrending implies that the “trend” and the “cycle” are uncorrelated with each other.
- It is easy to compute: one can use all the data points to consistently estimate the trend via conventional techniques.
- Because its simplicity, it is the least likely to distort the short-run dynamics of the cyclical component for low orders of the deterministic polynomial.
- From a forecasting point of view, it has the rather strong implication that the long-run forecast error variance converges to a fixed value. In practice, we would expect the forecast error variance of a point in the distant future to grow as the forecast horizon increases.

## 2.2 Stochastic Trends

Perhaps the most common filter in macroeconomics is the first-difference operator which in more general terms is

$$\psi(L)(1-L)^d y_t = \phi(L)\varepsilon_t; \quad \Delta_d y_t = y_t^T \quad (3)$$

Thus the first difference corresponds  $d = 1$ . This is by far the most common value of  $d$  although for a few U.S. macro series and many European macro series it is common to use  $d = 2$ . Financial data often exhibit long-memory properties (i.e. autocorrelations that decay at a hyperbolic rate rather than at an exponential rate) for which a value of  $d$  between zero and one half is often specified (there is a whole literature dedicated to investigating non-integer values of  $d$  and long-memory processes). The first-difference operator is particularly intuitive since it transforms the log of any series into its growth rate.

Remarks:

- Because the trend is now stochastic, the variance of the forecast error grows with the forecast horizon (note that:  $\Delta y_t = \varepsilon_t \Rightarrow y_t = \sum_{i=1}^t \varepsilon_{t-i} + y_0$ ).
- It is a natural transformation of the log of many macro time series (because it then becomes a growth rate).
- Assumes that the trend and the cycle are orthogonal to each other.
- Testing the value of  $d$  in practice is based on tests whose distribution has to be calculated with non-standard asymptotic theory.
- Many business cycle analysts perceive this expression of the trend as delivering a cycle that is “too noisy.”

### 2.3 Beveridge-Nelson Decomposition

In this decomposition, the secular and cyclical components are perfectly correlated instead. In particular, if

$$\begin{aligned} \Delta y_t &= u_t \\ u_t &= \phi(L)\varepsilon_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j} \end{aligned} \tag{4}$$

with  $\phi(L)$  a stationary and possibly infinite polynomial, then one can always decompose  $y$  as

$$\begin{aligned} y_t &= y_t^T + y_t^C = \underbrace{\phi(1)(\varepsilon_1 + \dots + \varepsilon_t)}_{Trend} + \underbrace{(\eta_t - \eta_0)}_{Cycle} \\ \eta_t &= \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j} \quad \alpha_j = -(\phi_{j+1} + \phi_{j+2} + \dots) \end{aligned} \tag{5}$$

Remarks:

- The B-N decomposition is an equivalent expression of the infinite moving average representation of any process. It will become very useful when we investigate the asymptotics of unit roots.
- Unlike the previous filters, the trend and the cycle are perfectly correlated.

### 3. Seasonal Adjustment

Although in macro-time series the data is often available “seasonally adjusted” (you will hear the words X11 or more recently X12, which are methods designed and used by the U.S. Census Bureau and are as close as it gets to a black box) it is important to review some of the most common seasonal filters. Even the most sophisticated seasonal filters will invariable affect the cyclical properties of the data, a fact few researchers, particularly in the business cycle literature, seem to consider practically relevant. However, there is little research supporting this statement, which is largely an artifact of convenience.

#### 3.1 Deterministic Seasonal Adjustment

Let  $s$  index the “season” to which observation  $t$  belongs. For example,  $s = 1, 2, 3, 4$  for quarterly data while  $s = 1, 2, \dots, 12$  for monthly. Further, let  $d_t^s = 1$  if  $t$  belongs to the  $s^{\text{th}}$  season, 0 otherwise. Then, the seasonal component of  $y_t$  denoted  $y_t^s$  is

$$y_t^s = \sum_{i=0}^K \sum_{s=1}^S \alpha_i^s d_t^s t^i \quad (6)$$

where as before,  $K$  is an integer denoting the order of the polynomial in  $t$  and usually ranges from 0 to 1 although in general higher values are possible. Expression (6) is therefore a direct extension of expression (1). A common simplification is to assume that the trend is common to all the seasons and that the only difference resides in the average levels so that,

$$y_t^s = \sum_{s=1}^S \alpha_0^s d_t^s + \alpha_1 t \quad (7)$$

which implies the restriction  $\alpha_1^1 = \dots = \alpha_1^s = \dots = \alpha_1^S$ .

Remarks:

- It is also common to combine deterministic seasonal adjustment with stochastic detrending so that, for example,

$$\psi(L)(1-L)^d y_t = \sum_{s=1}^S \alpha_0^s d_t^s + \phi(L)\varepsilon_t \quad (8)$$

However, note that there may still be seasonality in the cycles themselves (embedded in the specific forms of the polynomials  $\psi(L)$  and  $\phi(L)$ ). For example, suppose that  $y_t = \log GDP_t$  and that  $d = 1$ , then expression (8) (as well as expression (7)) states the assumption that output growth at time  $t$  depends on previous values of output growth irrespective of what the season is. However, the average rate of growth of output is allowed to vary with the season. This may or may not be a sensitive strategy depending on whether this assumption is a reasonable one.

- It is simple to implement and in its simpler form (i.e., when the seasonals only affect the constant term) it has the least impact on the cyclical behavior of the original series.

### 3.2 Stochastic Seasonal Trends

The same way we have generalized deterministic detrending to account for seasonal effects one could consider stochastic seasonal detrending so that

$$\psi(L)(1-L^S)^d y_t = \phi(L)\varepsilon_t \quad (9)$$

for  $S$  the length of the seasonal periodicity (i.e.  $S = 4$  for quarterly data,  $12$  for monthly, etc.).

Remarks:

- In principle,  $y_t$  may contain both a stochastic trend and seasonal stochastic trends and it is common to see transformations such as

$$\psi(L)(1 - L^S)(1 - L)y_t = \phi(L)\varepsilon_t \quad (10)$$

However, it is important to note that, for example, for quarterly data  $(1 - L^4) = (1 - L)(1 + L)(1 + iL)(1 - iL)$ , which already contains a unit root in the term  $(1 - L)$  and thus expression (10) imposes two unit roots, a feature that may not be desirable or warranted by the data.

- Note that the seasonal stochastic detrending in expression (9) still remains silent with respect to seasonal short-run dynamics embedded in the polynomials  $\psi(L)$  and  $\phi(L)$ .

### 3.3 Short-run Seasonality

As we discussed in previous sections, the goal of filtering is to remove features of the data that the researcher does not wish to explain. The hope is that the filter will remove these unwanted features while leaving the properties of the remaining features of interest reasonably unaffected. In the context of seasonal data, an alternative strategy would consist in explicitly modeling the seasonal component. The traditional approach of this type of treatment is based on the Box-Jenkins methodology and consists in modeling the data with multiplicative seasonal lagged polynomials, so that

$$\psi_S(L^S)\psi(L)y_t = \phi_S(L^S)\phi(L)\varepsilon_t \quad (11)$$

Note that the seasonal polynomials are expressed in terms of seasonal lags so that, for example,  $\psi_S(L^4) = 1 + \psi_{S1}L^4 + \psi_{S2}L^8 + \dots + \psi_{Sp}L^{4p}$  for quarterly data.

Expression (11) implies that the seasonal dynamics are common to all the seasons so that the dependency of the spring of this year on the spring of last year is the same as the dependency between the summer of this year and the summer of last year, and so on. A more parameter-intensive method of modeling seasonality directly, which does not impose this restriction, is to use *periodic models*. The basic strategy of this approach consists in specifying lagged polynomials specific to each season. However, because of their high dimensionality, they are of limited practicality beyond quarterly data.

#### 4. Moving Average Smoothers

Generally speaking, there is a large class of filters that can be expressed in terms of the following moving average form,

$$y_t^T = \sum_{i=-N}^M w_i y_{t-i} \quad (12)$$

for  $N, M$  two integers and the  $w_i$  a set of weights. The simplest example is a centered moving average

$$y_t^T = \frac{1}{2K+1} \sum_{i=-K}^K y_{t-i} \quad (13)$$

for  $K = 2$  if the data were quarterly,  $K = 6$  if monthly, etc. In fact, many of the filters designed to remove certain frequency components in the frequency domain can be expressed in the time domain as simple moving average smoothers. You may come across with the nomenclature “Low-Pass filter” which refers to, essentially, the type of detrending methods described in the earlier sections, while a “high-pass filter” removes the low frequency components instead. Business cycle researchers often talk of “band-pass” filters which are designed to remove frequencies ranging between a lower bound

and an upper bound of the frequency spectrum. I discuss a couple of band-pass filters below but only in terms of their moving average representation.

#### 4.1 The Hodrick-Prescott Filter

The H-P filter has become a popular choice among business cycle analysts. The original presentation of the filter in a 1980 working paper did not see its publication debut until 1997. The filter is obtained by solving the minimization problem

$$\min_{\{y_t^T\}_{t=1}^T} \sum_{t=1}^T (y_t - y_t^T)^2 + \lambda [(y_{t+1}^T - y_t^T) - (y_t^T - y_{t-1}^T)]^2 \quad (14)$$

where  $\lambda$  is an arbitrary constant that penalizes the variability in the smoother so that when  $\lambda = 0$  the smooth component is the data itself and no smoothing takes place. Conversely, as  $\lambda$  grows large the smooth component is a linear trend. For quarterly data, Hodrick and Prescott recommended  $\lambda = 1600$ , although an “optimal” choice is given by the ratio of the variance of the trend to the variance of the cycle. Nelson and Plosser (1982) suggested instead that  $\lambda$  should be in the range  $[1/6, 1]$ .

King and Rebelo (1999) show that the filter can be written as a moving average smoother of the form

$$y_t^T = \frac{\theta_1 \theta_2}{\lambda} \left[ \sum_{j=0}^{\infty} (A_1 \theta_1^j + A_2 \theta_2^j) y_{t-j} + \sum_{j=0}^{\infty} (A_1 \theta_1^j + A_2 \theta_2^j) y_{t+j} \right] \quad (15)$$

where  $\theta_1$  and  $\theta_2$  are complex conjugates whose values depend on  $\lambda$  and  $A_1$  and  $A_2$  are functions of  $\theta_1$  and  $\theta_2$ . Obviously the filter requires infinite observations but Baxter and King (1999) show how to appropriately truncate expression (15) in practice.

Remarks:

- One of the virtues and downfalls of the H-P filter is its flexibility: McCallum (2000) showed that the trend produced by the H-P filter applied to GDP data in the early turn of the century suggest the 1930's depression was only a mild recession. This result stems from the inability to calculate an approximately "optimal"  $\lambda$  for each series via estimation, for example.
- The filter depends on the choice of  $\lambda$  which makes the resulting cyclical component and its statistical properties highly sensitive to this choice.

#### 4.2 The Baxter-King Filter

The B-K filter is a band-pass filter designed to isolate business cycle fluctuations with a period of length ranging between 6 to 32 quarters (the typical ranges for U.S. expansions) although other characterizations are possible. The resulting filter is a centered moving average with symmetric weights

$$y_t^T = \sum_{i=-K}^K w_i y_{t-i} \quad (16)$$

with  $w_0 = 0.2777$ ;  $w_1 = w_{-1} = 0.2204$ ;  $w_2 = w_{-2} = 0.0838$ ;  $w_3 = w_{-3} = -0.0521$ ;  $w_4 = w_{-4} = -0.1184$ ;  $w_5 = w_{-5} = -0.1012$ ;  $w_6 = w_{-6} = -0.0422$ ;  $w_7 = w_{-7} = 0.0016$ ;  $w_8 = w_{-8} = 0.0015$ ;  $w_9 = w_{-9} = -0.0279$ ;  $w_{10} = w_{-10} = -0.0501$ ;  $w_{11} = w_{-11} = -0.0423$ ;  $w_{12} = w_{-12} = -0.0119$ .

Remarks:

- Notice that, like all moving average smoothers,  $K$  observations at the beginning and at the end of the sample are lost in the computation of the filter.
- Like many moving average smoothers, the B-K filter has been criticized on the grounds that it induces spurious dynamics in the cyclical component.

### 4.3 One-sided Moving Average Smoothers

Some of the filters we have introduced in sections 2 and 3 can be seen as special cases of expression (12) with  $N = 0$ , that is

$$y_t^T = \sum_{i=0}^M w_i y_{t-i} \quad (13)$$

The most common form of one-sided moving average filter is perhaps the exponential smoother. For this filter, the weights decline exponentially (hence the name) and it can be expressed as

$$y_t^T = \alpha \sum_{i=0}^{t-1} (1-\alpha)^i y_{t-i} \quad (14)$$

which can be recast as the simple recursion:  $y_t^T = \alpha y_t + (1-\alpha)y_{t-1}^T$ . The higher the value of  $\alpha$  the “smoother” the trend. Clearly, a first difference transformation,  $(1-L)$  is an example of a one-sided moving average but more generically, fractionally integrated processes with transformation  $(1-L)^d$  for  $-0.5 < d < 0.5$  can be expressed as follows:

$$(1-L)^d y_t = u_t \text{ for } |d| < 0.5 \text{ is equivalent to} \quad (15)$$

$$y_t = u_t + \sum_{i=1}^{\infty} \psi_i u_{t-i} \quad \text{with } \psi_i = \frac{(i+|d|-1)!}{i!(|d|-1)!}$$

### 5. Concluding Remarks

As I mentioned in the introduction, the wide variety of filters presented should not be viewed in isolation but rather as complements to one another. Additionally, I have not touched on other issues of specification, such as breaks in a deterministic time trend and

other ways to introduce more flexibility. These often become choices based on sophisticated test statistics and more frequently, the whim of the investigator. You should not think as there being a universal “optimal” filter which we have yet to find. Rather, each filter is designed to accomplish very specific goals. Depending on the goals particular to your research agenda, you should choose wisely what filter would be most appropriate.

You should also bear in mind I have not discussed any multivariate filtering methods. For instance, with regard to business cycles, it is plausible that many aggregate economic time series share common cyclical components. There are methods designed for taking advantage of features that are shared across time series for the purposes of detrending and filtering (cointegration would be an example but there are others).

## 6. References

Baxter, Marianne and Robert G. King (1999) “Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series,” *Review of Economics and Statistics*, 81(4): 575-593.

Burnside, Craig (1998) “Detrending and Business Cycle Facts: A Comment,” *Journal of Monetary Economics*, 41, 513-532.

Canova, Fabio (1998a) “Detrending and Business Cycle Facts,” *Journal of Monetary Economics*, 41, 475-512.

Canova, Fabio (1998b) “Detrending and Business Cycle Facts: A User’s Guide,” *Journal of Monetary Economics*, 41, 533-540.

Franses, Philip-Hans (1998) “Modeling Seasonality in Economic Time Series,” in Handbook of Applied Economics Statistics: Textbooks and Monographs, Aman Ullah and David E. A. Giles, editors. New York: Dekker, 553-577.

Gençay, Ramazan, Faruk Selçuk and Brandon Whitcher (2002) An Introduction to Wavelets and Other Filtering Methods in Finance and Economics, San Diego: Academic Press.

Hamilton, James D. (1994) Time Series Analysis, Princeton: Princeton University Press.

Miron, Jeffrey A. (1996) The Economics of Seasonal Cycles, Cambridge: MIT Press.

I recommend that you read the Baxter and King (1999) and Canova (1998a, b) and Burnside (1998) articles. The Baxter and King (1999) article contains a nice overview of desirable properties of filters and then goes on to derive the filter they propose.

Canova(1998a, b) and Burnside (1998) are in the same issue of the JME and also contain a nice discussion for practitioners. For an introduction to spectral analysis review

Hamilton's chapter 6. Gençay's book is a bit more terse but it has a nice discussion in Chapter 2. Franses (1998) gives you a nice flavor about how to model seasonality and Miron (1996) probably gives you more than you ever wanted to know about seasonality.