Appendix 2
Filters and frequency domain statistics

This appendix presents and discusses more formally the filters we apply in order to extract cyclical components of the raw data series, and discusses in some detail how we estimate the frequency domain statistics.

A2.1 The transfer function and phase

In this section we will define and analyze filters in terms of their effect on cyclical behavior. A univariate time invariant linear filter transforms the time series x (input) to the new series y (output) by forming a weighted moving average using the weights v :

(A2.1) $$y_t = \sum_{s=-\infty}^{\infty} v_s \, x_{t-s}.$$

The *transfer function* for the filter is the Fourier transform of the weights

(A2.2) $$B(\omega) = \sum_{s=-\infty}^{\infty} v_s \exp(-i\omega s),$$

where $\exp(-i\omega s) = cos(\omega s) - i^* sin(\omega s)$ is the complex exponential function (this is like replacing x_{t-s} with $\exp(-i\omega s)$ in (A2.1)). The frequency $\omega \in [-\pi, \pi]$, but spectra are symmetric around zero.

The reason why the transfer function is interesting is that the spectrum for y, $S^y(\omega)$, is given by

(A2.3) $$S^y(\omega) = |B(\omega)|^2 \, S^x(\omega),$$

where $S^x(\omega)$ is the spectrum for x and $|\ |$ denotes the modulus of the complex function $B(\omega)$, that is, if $B(\omega) = q + ip$ then $|B(\omega)|^2 = q^2 + p^2$.

The transfer function can equivalently be expressed as

(A2.4) $$B(\omega) = |B(\omega)| \exp[-iP^{xy}(\omega)],$$

where $P^{xy}(\omega)$ is the *phase* and the magnitude $|B(\omega)|$ is often called the *gain*. The phase is defined by

(A2.5) $$P^{xy}(\omega) = arg B(\omega) = tan^{-1}\{Im[B(\omega)]/Re[B(\omega)]\},$$

where Im and Re are operators that pick out the imaginary and real part of the complex number, respectively. One way of interpreting the phase is to note that

(A2.6) $$\tau(\omega) = P^{xy}(\omega)/\omega$$

is a measure of the phase shift at frequency ω in terms of time (which is measured in years in the present paper). The convention is that when $\tau(\omega) < 0$, then on average y (output) is lagged relative to x (input).

A2.2 Definitions of the filters

This subsection provides formal definitions for the three filters that we use in the paper. The computational procedures are described in a subsequent section.

The *Band–pass* filter (denoted BA) operates in the frequency domain. We use a filter with the transfer function

(A2.7) $$B^{BA}(\omega) = \begin{cases} 1 \text{ if } 2\pi/3 \geq |\omega| \text{ and } |\omega| \geq 2\pi/8 \\ 0 \text{ otherwise} \end{cases},$$

which means that cycles with a period between 3 and 8 years passes through the filter untouched, but all other cycles are completely removed.

The *Whittaker–Henderson* filter type A, introduced to economists by Hodrick and Prescott (1980) and described in some detail by Danthine and Girardin (1989), is given by the solution to the following optimization problem:

Split the series x into a cyclical component y and a trend g (where, of course, $x_t = y_t + g_t$) so as to minimize

(A2.8) $$\sum_{t=1}^{n} y_t^2 + \lambda \sum_{t=3}^{n} [(g_t - g_{t-1}) + (g_{t-1} - g_{t-2})]^2$$

Hence, the trend's tracking of the x series (giving small y^2) is traded off against the smoothness of the trend (giving small $[(g_t-g_{t-1})+(g_{t-1}-g_{t-2})]^2$). We have chosen $\lambda=10$,

which implies that the trend tracks the original series fairly closely, in order to get an average cycle length of approximately 5 years for the cyclical component (y) of the GDP series.

The *first difference* filter is simply $y_t = x_t - x_{t-1}$, which corresponds approximately to growth rates for all series that are expressed in logarithms.

A2.3 Computational procedures for the filters

This section describes the actual computational procedures that we apply. All calculations are made in the GAUSS programming language. Let us assume that we have n observations of the time series x (input) $x_0, x_1, \ldots, x_{n-1}$.

The band–pass filter is implemented in four steps. First, in order to reduce the wrap–around effect (see below) the x series is prefiltered with a Whittaker–Henderson filter with $\lambda=2500$ and padded with zeros to four times its original length. If needed, it is further padded with zeros to get a number of elements T equal to a power of 2. Second, a Fast (finite) Fourier Transform (FFT) is applied, to form the random spectral measure

(A2.9) $$Z(\omega_j) = (1/T)\sum_{t=0}^{T-1} x_t \exp(-it\omega_j), \quad \omega_j = 2\pi j/T \text{ where } j=0,1,2,\ldots,T-1.$$

The frequency runs from 0 to $2\pi(T-1)/T$, but since the spectral measure is periodic with a period of 2π, the values for $\pi \leq \omega < 2\pi$ equals the values for $-\pi \leq \omega < 0$. Third, the spectral measure is multiplied with the transfer function $B^{BA}(\omega)$ from (A2.7), with the definition changed in accordance with the frequency domain of the FFT

(A2.10) $$B^{BA}(\omega) = \begin{cases} 1 \text{ if } 2\pi/8 \leq \omega \leq 2\pi/3 \text{ or } 2\pi(1-1/3) \leq \omega \leq 2\pi(1-1/8) \\ 0 \text{ otherwise} \end{cases}.$$

Fourth, the inverse FFT is calculated as

(A2.11) $$y_t = \sum_{j=0}^{T-1} Z(\omega_j) B^{BA}(\omega_j) \exp(it\omega_j), \quad \text{where } \omega_j = 2\pi j/T,$$

and the n first observations are picked out. The result of these operations constitutes our filtered series.

The Whittaker–Henderson type A filter operates in the time domain and is simple to use. Danthine and Girardin (1989) shows that the filter can be computed as

$$(\text{A2.11a}) \qquad y_t = [\mathrm{I} - (\mathrm{I}+\lambda \boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}]\, x_t,$$

where I is the $n\times n$ identity matrix and $\boldsymbol{\Gamma}$ the $n\times n$ toeplitz matrix

$$(\text{A2.11b}) \qquad \boldsymbol{\Gamma} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{bmatrix}.$$

The first–difference filter is simply

$$(\text{A2.12}) \qquad y_t = x_t - x_{t-1}, \qquad t \in [1, n-1],$$

which of course means that one observation at the beginning of the sample is lost.

A2.4 Characteristics of the filters

This section discusses the characteristics of the filters in terms of their transfer and phase functions. Special attention is paid to the effect of the finite sample length.

Figure A2.1a–c shows in the upper panel the the gain functions $(|B(\omega)|)$ of the three filters, and in the lower panel the phase functions. The x–axis is expressed in frequencies (ω) which run from zero to π. For convenience, cycle lengths of 16, 8, 3, and 2 years are marked with vertical dotted lines. Note that the cycle length in years and the frequencies (ω) are related according to

$$(\text{A2.13}) \qquad \text{years/cycle} \doteq 2\pi/\omega.$$

In Figure A2.1a the gain for the ideal (asymptotic) *band pass* filter (A2.7) is plotted as the thick solid curve. This filter obviously keeps all cycles between 3 and 8 years but shuts out all others cycles. Unfortunately, this filter is not realizable for a finite sample. The reason is easy to see if the band pass filter is transformed to the time domain by applying the inverse Fourier Transform. It can be to shown that the theoretical (asymptotic) weights in the time domain v_s in (A2.1) for the band pass filter (A2.7) are

$$v_s = \begin{cases} [\,\sin(s2\pi/3) - \sin(s2\pi/8)\,]/(s\pi) & s=\pm1,\pm2,\dots \\ [\,\sin(2\pi/3) - \sin(2\pi/8)\,]/\pi & s=0 \end{cases}.$$

These weights decline fairly slowly as $|s|$ increases, which is at the root of the problem. With a finite sample, some of the v_s are disregarded implying that the cycles to some extent are blurred with each other, a phenomenon usually called leakage. How does the the gain function actually look like for our sample length? In order to understand that, we have to deal with the question about how the FFT that we use for calculations treats a time series. As a matter of fact, it treats the series as periodic and assumes the the last observation is identical to the observation preceding the first observation. This "wrap–around" can seriously distort the time series. Therefore, we try to eliminate the wrap–around by padding the time series with a lot of zeros. All this means that the finite sample problem is indeed present.

In Figure A2.1a we have plotted, along with the ideal gain function discussed above, the actual gain function for observation t=64 (thin solid line) and t=1 (dotted line) in a sample of length 128. Obviously, for t=64, the finite sample problem is insignificant, but for t=1 (and also for t=128 which is identical to t=1) the leakage is considerable. It can be shown that for t=5, the leakage is dramatically smaller than for t=1. Furthermore, for the observations at the beginning (end) of the sample, only weights v_s for $s<0$ ($s>0$) has any effect. This will introduce a phase shift for these observations. Even if the ideal band–pass filter, and the actual band pass–filter for t=64 has no phase, as illustrated in the lower panel of Figure A2.1a, this is not universally true. For t=1, the phase shift is negative for longer cycles, around zero for the pass band, and positive for shorter cycles (the pattern for t=128 is exactly the opposite). Again, for t=5, the phase shift is dramatically smaller than for t=1.

Similar problems arise in applying the Whittaker–Henderson filter. The upper panel of Figure A2.1b shows the gain of the Whittaker–Henderson filter for a sample with 128 observations, for observation 1 and 64. The value λ=10 has been used. Since the sample is finite, the actual filter (the v_s values) will differ between observations. The gain function for the first (t=1) and last (t=128) coincide, but differs from the gain for an observation in the middle of the sample (t=64). It is worth noting that while the former seems to keep more of the longer cycles than the latter, the opposite is true for shorter cycles. The lower panel of Figure A2.1b illustrates the obvious phase shifts at the beginning of the sample: at t=1 the phase is negative, that is, output lags input. This is natural, since at t=1 only contemporaneous and leaded values of the input series are available. For the seven year cycle, the lag is about one year. Analogously, the filter for t=128 gives a positive phase. But, in the middle of the sample the filter is symmetric in the time domain, which gives a zero phase. According to Figure A2.1b, the phase vanishes fairly quickly as one moves into the interior of the sample. The phase for t=5 is fairly unimportant.

Figure A2.1c shows the gain and the phase for the first–difference filter. From the upper panel it is clear that the filter keeps a great deal of the long cycles and actually magnifies the short cycles. The phase for the first–difference filter is positive. At seven year cycles the lead is about 1.25 year and at three year cycles it is about 0.25 year.

Figure A2.2 brings together the gain functions for the three different filters in order to highlight the differences. It is clear that the Whittaker–Henderson filter differs from the band pass filter by keeping more of the long cycles and all the short cycles. The first–difference filter is even more extreme in both these respects.

A2.5 Estimation of spectra and coherencies

This section summarizes the procedures for estimating spectra and coherencies. Most of the material is adapted from Koopmans (1974).

The estimation of spectra proceeds in three steps. First, the random spectral measure $Z(\omega_j)$ of the time series x is calculated by applying the FFT in (A2.9). Second, the *periodogram* is constructed by

(A2.14) $$I^x(\omega_j) = T/(2\pi)\ |Z(\omega_j)|^2.$$

Third, the weighted average over $\pm k$ frequencies of the periodogram is our estimate of the spectrum

(A2.15a) $$\hat{S}^x(\omega_j) = \sum_{i=-k}^{k} W_i I(\omega_{j-i}),$$

where W_i is a tent shaped sequence of weights which fulfill

(A2.15b) $$\sum_{i=-k}^{k} W_i = 1.$$

The *equivalent degrees of freedom* is given by

(A2.16a) $$r = 2 \Big/ \sum_{i=-k}^{k} W_i^2 ,$$

and a 100(1–a)% confidence interval is given by

(A2.16b) $$r\hat{S}^x(\omega_j)/b \ \leq\ S(\omega_j)\ \leq\ r\hat{S}^x(\omega_j)/a,$$

where a and b are given by $Prob(a \leq \chi_r^2 \leq b) = 1{-}a$.

The estimation of the coherencies of two series x and y goes in a similar way. First, the spectrum of each series is estimated. Second, using the random spectral measures of each series, $Z^x(\omega_j)$ and $Z^y(\omega_j)$, the multivariate periodogram for the two series is calculated as

(A2.17) $$I^{xy}(\omega_j) = T/(2\pi)\ Z^x(\omega_j)Z^y(\omega_j)^*,$$

where the star (*) denotes the complex conjugate. Third, weighting $I^{xy}(\omega_j)$ as in (A2.15), gives the estimated cross spectrum $\hat{S}^{xy}(\omega_j)$. Finally, the coherence is obtained by

(A2.18) $$\hat{C}^{xy}(\omega_j) = |\hat{S}^{xy}(\omega_j)| / (\hat{S}^x(\omega_j)\hat{S}^y(\omega_j))^{\frac{1}{2}}.$$

An approximate 100(1–a)% confidence interval is given by

$$\hat{C}^{xy}(\omega_j)_{min} = tanh\{\ arctanh(\hat{C}^{xy}(\omega_j)) - Crit\times(2(2k{-}1))^{-\frac{1}{2}} - (2(2k{-}1))^{-1}\ \}$$
$$\hat{C}^{xy}(\omega_j)_{max} = tanh\{\ arctanh(\hat{C}^{xy}(\omega_j)) + Crit\times(2(2k{-}1))^{-\frac{1}{2}} - (2(2k{-}1))^{-1}\ \}$$

where $Crit$ denotes the critical value for $a/2$ in the standard normal distribution.

Appendix 3
Statistical inference on filtered data

This appendix discusses some problems of statistical inference on filtered data.

A3.1 Bandpass filtered data

Statistical inference on band–pass filtered time–series turns out to be a little less straightforward than might be expected. This is due to the following theorem[22]:

A process is purely linearly non–deterministic iff:

$$\int_{-\pi}^{\pi} \log f(\omega)\mathrm{d}\omega > -\infty$$

where f is the spectral density of the process. Otherwise it is linearly deterministic, so that x_t can be predicted without error using a linear prediction function given the history of x up until t–1.

Since a pure band–pass filter is constructed to nullify the spectrum over a band of frequencies with an asymptotically positive measure, it follows that a filtered process becomes deterministic. Inference based on asymptotic properties of non–deterministic processes is thus invalid. Another important implication is that no band–pass filtered process can be Granger–caused by another process: since a deterministic process is perfectly linearly predictable from its own history, no other process can help predict it.

In the following, we discuss two ways of computing significance levels for band–pass filtered series. The first way appears more appealing but has problems for tests on sub–samples and also seems to have lower power. The second way is therefore used in this paper.

Method 1. First we note that zero correlation between the two filtered series y_f and x_f

22 See e.g. Whittle (1983) p. 26.

is equivalent to a zero regression coefficient β in the bivariate regression:

(A3.1) $$y_f = x_f\beta + \varepsilon .$$

Here, however, both y_f and x_f as well as ε are deterministic.

Let no subscript denote an unfiltered series and let the subscript r denote the remainder after filtering so that, for example, $y = y_f + y_r$. If we then run the regression:

(A3.2) $$y = x_f\beta + \varepsilon ,$$

we will have deterministic regressors, but y and thus the regression will be non–deterministic. Fortunately, regressions (A3.1) and (A3.2) will both yield the same estimate $\hat{\beta}$. This is because the space spanned by y_r is orthogonal to the space spanned by x_f Letting F represent the band–pass filter in the time domain and noting that F is symmetric and idempotent[23] we have that:

(A3.3)
$$\begin{aligned}\bar{\beta} &= (x_f'x_f)^{-1}x_f'y_f \\ &= (x_f'x_f)^{-1}x'F'F_y \\ &= (x_f'x_f)^{-1}x'F'y \\ &= (x_f'x_f)^{-1}x_f'y\end{aligned}$$

But since x_f and x_r are also orthogonal we may include x_f in the regression and run:

(A3.4) $$y = x_f\beta + x_r\gamma + v .$$

This will reduce the residual variance if there is some correlation between y and x_f and hence increase the power of the test. To test whether the correlation between filtered series is zero, we test if β in A3.4 is significantly different from zero. To account for non–spherical residuals we use a Newey–West estimator for the variance of $\hat{\beta}$.[24]

In our case we are interested in statistics for the sub–samples as well as for the

23 Following Engle (1974) the basic band–pass filter can be written $\mathbf{W^*AW}$ where $\mathbf{W}' = [W_0\ W_1 \ldots W_{T-1}]$, where $w_k = (1, e^{i\theta k}, e^{2i\theta k}, \ldots e^{(T-1)i\theta k}), \theta_k = 2\pi k/T$, $\mathbf{W^*}$ is the complex conjugate of the transpose of $\mathbf{W}$ and $\mathbf{A}$ is a diagonal matrix with a symmetric diagonal of ones and zeros such that a one represent included frequencies. Using this and the fact that $\mathbf{W'W} = \mathbf{WW'} = I$, it is easy to show that $\mathbf{F}$ is symmetric and idempotent.

24 We use four lags in the computation of var $(\hat{\beta})$ See Newey—West (1987).

whole period. In this case the method is less well suited. The problem arises because the orthogonality, which gives us an equal $\hat{\beta}$ in (A3.1) and (A3.2), does not hold if we filter the series over the full sample period but run the regression for a sub–sample. In practice, this will cause unreasonable results in small samples. For instance, correlation coefficient between two filtered series which is close to zero for a sub–sample may be spuriously significant. The remedy is obviously to use one filter for every sub–sample that we study. This, however, leaves us with too small samples.

Another problem is that in practice we don't use the basic filter described in Footnote 23. Among other things, we are padding the series with zeros to reduce some finite sample problems with band–pass filters.[25] This procedure will also produce non–orthogonality between x_f and x_r. We have therefore chosen another method to compute significance levels.

Method 2. Our second method relies on the fact that it is only when the filter sets the spectral power for some frequencies (with strictly positive measure) *exactly* to zero that the series becomes deterministic. If we use a filter that reduce the spectral power of non–business cycle frequencies to some constant κ bounded away from zero, we may, by choosing κ small enough, get filtered series arbitrary close to the band–pass filtered series. In this case, our filtered process stays non–deterministic allowing the use of standard asymptotic results.

Instead of specifying κ to some arbitrary constant strictly bigger than zero, we have actually set it to exactly zero. This is purely for convenience since there exists some κ strictly bigger than zero for which all results would be close (e.g. within the numeric accuracy of GAUSS) to the reported results.

Significance levels for the reported correlations have thus been computed by running regression A3.1 and testing for β being different from zero using the consistent

25 See Appendix 2 for a description of the filter.

Newey–West estimator of the variance of $\hat{\beta}$. The Newey–West estimator is used since the residuals are not assumed to be spherical. In addition to the likely presence of non–spherical disturbances before filtering, the filtering procedure will introduce autocorrelation even if the original residual covariance matrix is diagonal.

We have compared the two methods of computing significance levels. The first method sometimes gives unreasonable results for sub–samples. We also find that the first method on average rejects zero correlation less frequently. For the full sample the first method rejects 165 of 840 contemporaneous correlations being zero on the 10% level. Reassuringly, all of these where also significantly different from zero according to the second method. But the second method yields another 240 significant correlations, totaling 405 (Table 4.2). Since the first method includes more noise than the second, it seems reasonable that its power is lower.

Given the discussion above, it is obvious that the reported significance levels must be interpreted with caution. This point is further strengthened by the fact that we don't know the small–sample behavior of the Newey–West covariance estimator in this case.

A3.2 Whittaker–Henderson filtered data

A Whittaker–Henderson filtered series includes spectral power at all frequencies. Therefore the problem of deterministic processes discussed above does not arise. Granger–causality tests between Whittaker–Henderson filtered series are still problematic, however, for the reason discussed in Section 5.2.

By filtering out virtually all of the lowest frequencies the Whittaker–Henderson filter typically introduces autocorrelation in the filtered series. All significance tests involving these series in the paper therefore rely on covariance matrices that have been recomputed with the Newey–West method.

Figure A2.1a: Band pass filter

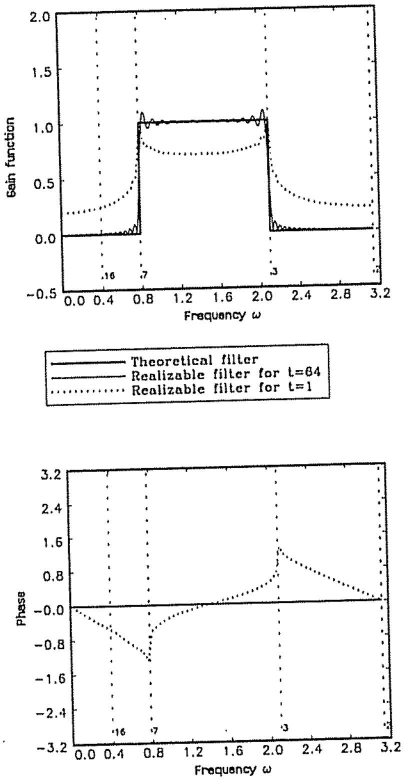

Figure A2.1b: Henderson-Whittaker λ=10

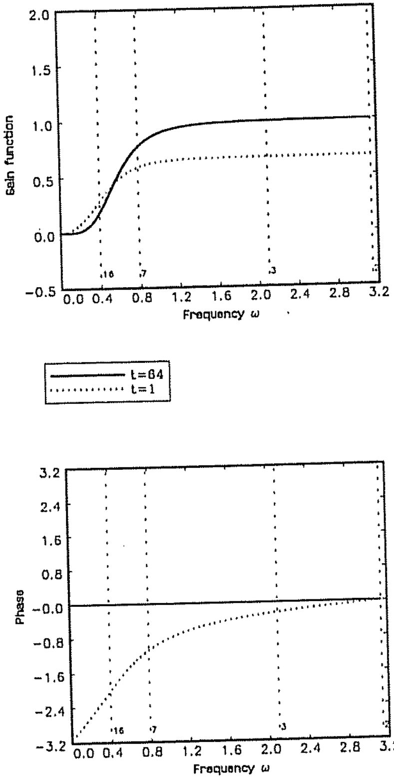

Figure A2.1c: First-difference