Optimal band-pass filtering and the reliablity of current analysis

Simon van Norden
HEC (Montréal)
CIRANO
CRDE

Abstract: This paper shows how existing band-pass filtering techniques and their extensions may be applied to the common problem of estimating current trends or cycles. These techniques give estimates which are “optimal” given the available data, so their standard errors represent a lower bound on what can be achieved with other univariate techniques. Applications to the problems of estimating current trend productivity growth, core inflation and output gaps are considered. These illustrate the different factors which determine how accurately the underlying trend is measured and the degree to which estimated cycles tend to lead or lag the true cycle.

1.0 Introduction

A common problem in macroeconomics is that of measuring the business cycle, or more generally, of separating long-run trends from short-term movements. A technique which does so may be thought of as a *filter,* one which is applied to raw economic data prior to analysis. The best known is the Hodrick-Prescott (HP) filter, which has become the benchmark against which all other filters in macroeconomics are compared.[1]

The HP filter is unabashedly arbitrary; it was proposed and adopted largely on the basis that it gave results which “looked reasonable.” Its use has since been rationalized as an approximate band-pass filter. A band-pass filter is one which isolates movements in a series between a specified upper and lower frequency or duration; movements outside this desired frequency band are eliminated. As commonly used with quarterly economic data, the HP filter eliminates or greatly reduces most long-run movements in the series while preserving those at roughly business cycle frequencies.[2] The result is a detrended series which looks like a business cycle and has served as an agnostic basis for much economic analysis.[3]

Given its preeminent role and its arbitrary nature, the HP filter has been the focus of much emulation and innovation in recent years. Baxter and King (1999) argue in favour of replacing the HP filter with a more exact band-pass filter, arguing that better results come from using a better approximation. Gomez (2001) and Pollock (2000) propose the use of other ad hoc filters which are commonly used in engineering and which also approximate band-pass filters.[4] Pedersen (1998) and Kaiser and Maravall (2001) propose extensions or modifications to the HP filter to improve its performance. However, most of this literature has ignored the application to current analysis.

The distinguishing feature of current analysis is that it interprets the most recently available information; the cycle (or trend) of interest is that at the end of the data sample. The dominant focus of the above literature is historical analysis, in which we care mostly about the cycle (or trend) somewhere in the middle of the sample. This distinction is sometimes critical. Most of the analysis in the existing literature is restricted to symmetric filters; to isolate the cycle or trend at time *t*, such filters use an equal number of observations from before and after *t*.[5] This precludes their use at the end of sample. Other filters are justified

1. Its current popularity stems from its use in the seminal working paper Hodrick and Prescott (1977), finally published as Hodrick and Prescott (1997), although the technique dates from the 1920s.
2. In practice, business cycles are defined to have durations between 6 and 32 quarters. This definition gained popularity after Baxter and King (1999) cited Burns and Mitchell (1947) as characterizing business cycles in this way. It is useful to remember that these numbers are not written in stone; Stock and Watson (1998) quote Burns and Mitchell (1947, p. 3) as stating that business cycles vary in duration “... from more than one year to ten or twelve years.”
3. The claim that HP-filtered output “looks like” a business cycle is a popular misstatement. Business cycle measurement and analysis, since its infancy, has used the HP filter (or even simpler moving-average filters which produce similar results.) It is therefore probably more accurate to say that the output of these filters have *defined* what we think of as business cycles. See Morley, Nelson and Zivot (1999).
4. Both examine Butterworth filters, of which they note the HP filter is simply a special case.

on the basis of their mid-sample properties, but may behave quite differently at the end of sample.[6]

This paper considers the filtering problem from the perspective of current analysis. Rather than use ad hoc approximations to band-pass filters, it shows how to construct one-sided band-pass filters which are optimal in a minimum mean-squared-error sense. Unlike the filters mentioned in the above literature, the optimal filter will vary with the properties of the data series to be filtered. While such filters are little known in macroeconomics, they are not new; the annex of this paper reviews the contributions of Koopmans (1974) and Christiano and Fitzgerald (1999). The body of this paper gives an overview of these optimal filters and applies them to three problems of widespread interest; estimating the current output gap, the current trend growth rate of productivity and the current trend rate of inflation.

Related to the filtering literature discussed above is another one which examines the reliability of filtered estimates of trends and cycles.[7] Since the current analysis filter discussed in this paper minimizes a MSE criterion, we can relate its reliability to those of other filters examined in this literature. It also establishes an upper bound on the accuracy that any such filters can hope to achieve. Because this bound depends in a complicated way on the properties of the data analysed, we investigate the properties of the optimal filter for the three common problems in current analysis.

The next section of the paper provides a non-technical overview of band-pass filtering and the optimal one-sided band-pass filter. A derivation and a more detailed discussion of this filter may be found in the technical appendix. Section 3.0 applies the filter to three common macroeconomic problems of measuring trend and cycles and discusses the results. Section 4.0 summarizes the conclusions and suggests avenues for further research.

5. Baxter and King (1999), for example, suggest reserving about five years of data from each end of the sample to provide the necessary leads and lags.

6. The HP filter is a case in point. Comparisons to band-pass filters are based on its symmetric MA representation, which is a limit the filter approaches in the middle of a large sample. Its representation at the end of sample is quite different; see St. Amant and van Norden (1997).

7. Examples of this include Setterfield et al. (1992), Staiger, Stock and Watson (1997), Orphanides and van Norden (1999, 2002) and Cayen and van Norden (2002).

2.0 An Optimal Band-Pass Filter for Current Analysis

2.1 Filters: A Primer[8]

A filter may be thought of as an algorithm for processing a time series in order to get a more meaningful statistic; e.g. the process of averaging measured rainfall at a given location to get "average" rainfall. We can describe this mathematically as

$$S = f(\vec{y}) \qquad \text{(EQ 1)}$$

where S is our statistic, $\vec{y}$ is our time series and $f(.)$ is our filter. While such processing could be complex, attention often focuses on a particularly simple and tractable case; the *linear*, *time-invariant* filter. Such filters can be described as

$$S_t = \vec{\beta} \cdot \vec{y} = \sum_{i=-\infty}^{\infty} \beta_i \cdot y_{t-i} \qquad \text{(EQ 2)}$$

The distinguishing features of such filters are that the weight β_i we put on a particular observation does not depend on t, and that the operation is linear in $\vec{y}$.[9] If we think of y_t as a random error term, the Wold decomposition theorem tells us that this class of filters is related to the class of ARMA process. In this situation, since the properties of $\vec{y}$ are held fixed, the properties of *S* are determined by $\vec{\beta}$.

We are often particularly interested in the dynamics of *S,* which may be uniquely characterized via frequency or spectral analysis. The idea is to decompose all the movements in *S* into cycles of varying frequency and amplitude. Such a unique decomposition exists if *S* is stationary (and if not, we assume that we can difference it until it is.) Furthermore, since cycles of different frequencies are uncorrelated in the long run, the variance of *S* will simply be the sum of its variances over all frequencies. The relative importance of these different frequencies in the overall variance tells us something about the dynamic behaviour of the series. For example, an i.i.d. error will display a constant variance at all frequencies, while a random walk will have much more variance at low frequencies (long cycles) than at high (short cycles.) The function decomposing the total variance by frequency is commonly called the *spectrum* or *spectral density* and is typically shown graphed from 0 (lowest frequencies, infinitely-long cycles) to π (highest observable frequencies, cycles of 2 periods.)

The spectrum of *S* depends on the properties of both $\vec{\beta}$ and $\vec{y}$. To understand the effects of $\vec{\beta}$, we can divide the spectrum of *S* by that of $\vec{y}$ to define the *squared gain* or *transfer*

8. This section provides an intuitive introduction to the filtering terms used in the rest of this paper. It may be skipped without loss of continuity by those familiar with spectral analysis.
9. Several well-known filters do not belong in this class. Examples include the HP filter (where the weights vary with *t*) and the Hamilton filter for the probability of being in a particular regime (which is a non-linear function of $\vec{y}$.)

function of $\vec{\beta}$. Frequencies at which the squared gain is greater than 1 are accentuated in *S*, while those at which the squared gain is (close to) zero are (nearly) removed from *S*. The aim of *band-pass* filtering is to choose β to match a particular kind of squared gain function; one which has a gain of 1 over a particular frequency range $(l,u) \mid 0 \le l < u \le \pi$ and zero elsewhere. The case where $l = 0$ is called a low-pass filter while $u = \pi$ is called a high-pass filter.

If a filter has the property $\beta_j = \beta_{-j}$ $\forall j$ it is called a *symmetric* filter, whereas if $\beta_j = 0$ $\forall j > 0$ then the filter is said to be *one-sided*.[10] Symmetric filters have the property that *S* will tend neither to lead nor to lag movements in the corresponding components of *y*; the same is not generally true for non-symmetric filters.[11] This effect of non-symmetric filters is called *phase shift* and in general will vary from one frequency to another.

2.2 Business Cycle Filters

Band-pass filtering is an approach to the measurement of trends or cycles in macroeconomics whose appeal rests on two key assumptions;

1. We can agree on some threshold duration such that we wish to interpret movements of longer duration as trends and shorter duration as cycles.
2. Aside from this, we wish to remain fairly agnostic about the economic or stochastic processes generating the data.

In this case we may detrend the data using a low-pass filter (one which passes all frequencies *below* the threshold, i.e. all durations *above* the threshold) to isolate the trend, or equivalently, use a high-pass filter to isolate the cycle. Two of the three applications we study below use low-pass filters to isolate such trends. The case of business cycles is more complex since we wish to exclude both the trend component and a seasonal/short-lived component. We therefore need a band-pass filter to block both the very long and very short duration movements.

The ideal band-pass filter would have a gain of zero outside the (*l,u*) interval and a gain of one inside. Deviations from the former condition allow leakage from undesired frequencies, while deviations from the later distort the "true" cycle present in the data. The unique filter with such properties exists and is given by the formula

$$\begin{aligned} B_j &= \frac{\sin ju - \sin jl}{\pi j} \text{ for } |j| \ge 1 \\ &= \frac{u-l}{\pi} \text{ for } j = 0 \end{aligned} \quad \text{(EQ 3)}$$

One problem with this ideal filter is that we require the sum in (EQ 2) to go from $-\infty$ to ∞. Truncating this sum at some finite values, say -*N* and *N*, results in a approximate filter

10.The only one-sided symmetric filter is the trivial filter which just multiplies y_t by a constant; we'll typically ignore that special case and talk as if these two classes are mutually exclusive.

11.For example, consider the difference between a centre-weighted and a one-sided moving average.

in which desired rectangular shape of its squared gain is contaminated by sinusoidal imperfections. (See Figure 1.) Baxter and King (1999) suggest that using values of N as small as 20 in quarterly data gives reasonable results for US business cycles if we adopt the Burns and Mitchell cut-offs of 6 and 32 quarters.

The problem with this approach is that it cannot be used for current analysis. Using N=20 implies that our most recent estimates of the business cycle would be 20 quarters prior to the last quarter for which we had data. One way around this would to use the Baxter-King formula at the end of the sample, simply omitting (i.e. replacing by zero) the missing observations which are not yet known. As shown in Figure 1, this gives poor results, even for large values of N; the resulting filters have a gain which is far from 1, vary considerably over the frequency band of interest, and leak much more of the frequencies outside the desired band. Stock and Watson (1998) use a different *ad hoc* solution. They fit the available time series to a simple AR model, then use forecasts from the fitted model in place of the required future observations. Unfortunately, they do not provide a justification for this procedure, nor do they examine how closely it approximates the ideal filter.

Another criticism that has been made of this approach is that even if the filter we use has a gain function which is close to that of the ideal filter, this need not imply that the series it produces will be a good approximation of the ideally filtered series.[12] The problem is that many economic series display a "Typical Granger Spectral Shape"; the density in their spectrum is highly concentrated at the lowest frequencies. This in turn means that for band-pass or high-pass filtering (e.g. for measuring business cycles), we care much more about how well we approximate the ideal filter at low frequencies than at high frequencies.[13]

2.3 Optimal Current-Analysis Filters

To adapt the band-pass filtering approach to a current analysis context, we would like to have some optimal filter $\{\hat{B}_j\}$ which minimizes

$$E\left[\sum_{j=0}^{T-1} \hat{B}_j \cdot y_{T-j} - \sum_{j=-\infty}^{\infty} B_j \cdot y_{T-j}\right]^2 \qquad \text{(EQ 4)}$$

In other words, given T observations on the series we wish to filter $\{y_t\}$, the optimal filter will give us the minimum mean squared error (MSE) estimate of what the ideal filter would give us with data from ∞ to $-\infty$. As shown in the Annex to this paper, this problem has a unique solution under fairly standard conditions.[14] In the case where $\{y_t\}$ is stationary, we find that

12.For example, see Guay and St. Amant (1997)

13.Pedersen (1998) re-examines HP filters from this perspective and suggests alternatives to the traditional value of 1600 for its smoothing parameter.

14.The annex gives solutions for the case of ARIMA(p,d,q) processes ($-1 < d < 3$) and surveys related contributions in the literature.

$$\vec{\hat{\beta}} = \vec{\Sigma}_y^{-1} \cdot \vec{B} \cdot \vec{\sigma}_y \qquad \text{(EQ 5)}$$

where

$\vec{\hat{\beta}} = [\hat{B}_0, \ldots, \hat{B}_{T-1}]'$, a length *T* column vector

$\bar{\Sigma}_y = [\vec{\sigma}_{y,0}, \vec{\sigma}_{y,1}, \vec{\sigma}_{y,2}, \ldots, \vec{\sigma}_{y,T-1}]'$, a *T x T* matrix.

$\vec{\sigma}_{y,j} = [\sigma_y(-j), \sigma_y(-j+1), \sigma_y(-j+2), \ldots, \sigma_y(T-1-j)]'$, a length *T* column vector.

$\vec{\sigma}_y = [\sigma_y(-q), \sigma_y(-q+1), \sigma_y(-q+2), \ldots, \sigma_y(q)]'$, a *2q+1* column vector.

$\sigma_y(q) = cov(y_t, y_{t+q})$

$\vec{B} = [\vec{B}^0, \vec{B}^1, \vec{B}^2, \ldots, \vec{B}^{T-1}]'$, a *T x 2q+1* matrix

$\vec{B}^j = [B_{j-q}, B_{j-q+1}, B_{j-q+2}, \ldots, B_{j+q}]'$, a *2q+1* column vector

B_j is the weight of the ideal filter, given by (EQ 3).

In general, these optimal weights depend on (a) the number of observations *T* we have, (b) the dynamics of our series y_t, as measured by its autocovariances, and (c) the ideal weights B_j from (EQ 3). Furthermore, they have an intuitive interpretation as the solution to a regression problem; one where we regress the doubly-infinite set $\{B_j \cdot y_{T-j}\}$ on our *T* observations $\vec{y}$. The resulting coefficients are our optimal weights, and therefore our minimization problem (EQ 4) simply seeks to minimize the variance of the regression residuals.[15]

We can better understand this formula and the above intuition if we consider two special cases. First, suppose that y_t is i.i.d. This means that $\sigma_y(l) = 0$ for all $l \neq 0$, so we may set *q*=0. This makes $\vec{\sigma}_y$ a scalar, equal to the variance of *y*, and Σ_y is simply the identity matrix times this variance. This means that (EQ 5) further simplifies to $\vec{\hat{\beta}} = [B_0, \ldots, B_{T-1}]$. Put another way, in this case the optimal solution is simply a truncated version of ideal weights, precisely the same solution which was seen to give very poor results in Figure 1.

Now suppose that *y* follows a stationary MA(*Q*) process. Since its autocovariances will be zero for leads and lags greater than *Q*, this again effectively determines *q* in (EQ 5). However, suppose that instead of using the optimal weights, we use the Stock and Watson approach of padding our *T* observations with *Q* forecasts/backcasts from the MA model at each end of the sample, then using the Baxter-King approximate filter with *N*=*Q*. The esti-

15.This interpretation is developed further in the Annex, particularly in Section 2.1.

mate from this two-step ad hoc procedure will be identical to the estimate from our optimal filter. This is because the optimal weights given in (EQ 5) reflect both the weights used to form forecasts/backcasts at the ends of the available sample, as well as the weights which the Baxter-King filter would place on them. Put another way, (EQ 5) implies that the Stock-Watson two-step procedure will give optimal estimates at the end of sample provided that (1) we use the "right" forecasting model to pad our data, (2) we pad our sample until our forecasts have converged to zero.

Another feature of (EQ 5) is that it lets us solve for the minimum value of (EQ 4). This is useful, since it tells us how well our best end-of-sample estimates can approximate the ideal estimates. The general solution is given by

$$E\left[\sum_{j=-\infty}^{\infty} B_j \cdot y_{T-j}\right]^2 - E\left[\sum_{j=0}^{T-1} \hat{B}_j \cdot y_{T-j}\right]^2 \tag{EQ 6}$$

In the above case where y is i.i.d., this reduces to

$$\left(\sum_{j=-\infty}^{\infty} B_j^2 - \sum_{j=0}^{T-1} B_j^2\right) \cdot \sigma_y^2 = \left(\sum_{j=-\infty}^{-1} B_j^2 + \sum_{j=T}^{\infty} B_j^2\right) \cdot \sigma_y^2 \tag{EQ 7}$$

3.0 Applications

We now examine the performance of the optimal filter by applying it to three problems of common interest; estimating the current output gap, the current trend growth rate of productivity and the current trend rate of inflation. The first of these differs from the other two in that we analyse the raw data in levels, not growth rates, and that we seek to isolate the intermediate frequencies rather than the low frequencies. Because the optimal filter is a function of the dynamic properties of the series analysed, our results can be expected to differ across applications.

3.1 Data

The output series (Y_t) is the natural logarithm of Euro-zone real GDP for the period 1991Q1 to 2003Q2. The inflation series (π_t) is the monthly difference of the natural logarithm of the Euro-zone Harmonized Index of Consumer Prices (seasonally adjusted) covering the period January 1995 to August 2003. The productivity series (Q_t) is difference of the natural logarithm of quarterly data on Real GDP per person employed from 1991Q1 to 2003Q2. Note that deterministic components were removed from all three series prior to analysis. In the case of Q_t the series were demeaned; for Y_t and π_t a deterministic linear trend was also removed.

For each series, two different estimates of the autocovariance function $\sigma_y(q)$ were then constructed. The first fit a low-order ARMA model to the data (Table 1), then used the estimated parameters of the ARMA model to calculate the implied covariances.[16] The second used a nonparametric kernel estimate.[17] In both cases, each data sample of *N* observations was used to calculate *N-1* autocovariances. The two approaches gave sometimes similar estimates, as shown in the top left panel of Figure 2 through Figure 4, with the nonparametric kernel tending to capture somewhat more complex and persistent dynamics. The second panel (top right) in each of these figures shows the corresponding spectrum for each series.

The dynamics of the three series look quite different. Quarterly productivity growth shows little persistence and is close to white noise; its ARMA spectrum is nearly flat and the kernel-estimated spectrum shows no clear tendancy to rise or fall as the frequency increases. Output shows slowly decaying autocovariances, consistent with estimated autoregressive roots of nearly 0.95. Its spectrum displays the typical Granger shape, with density powerfully concentrated in the low frequencies. Monthly inflation falls between these two extremes, with the dynamics of its ARMA approximation showing less persistence that those of the nonparametric kernel estimate.[18] The result is two very different-looking spectra, with the ARMA-based estimate looking quite flat, but the kernel-based estimate

16.The *ARMABIC3()* procedure from the COINT module for GAUSS by Ouliaris and Phillips (1995) was used for estimation and model selection. This uses the BIC criterion for model selection and a 2 or 3 stage Hannan-Rissanen iterative estimation procedure.

17.The results presented here use the Quadratic-Spectral kernel (without the data-dependent band-width selection.) Limited experimentation suggested that the results were not sensitive to this choice.

showing even more concentration in the low frequencies than the spectrum for output growth.

3.2 Filtering Productivity Growth

Trends in productivity growth are the subject of considerable analysis, with interest in current trends having intensified in recent years. While it is widely acknowledged that labour productivity is procyclical, most analysis of productivity growth does little to explicitly separate its trend and cyclical components beyond examining averages of growth rates over several years. It would therefore be useful to construct optimal estimates of current trend productivity, as well as to know how reliable such estimates may be. This is precisely what the results from Section 2.0 now enable us to do.

Since we are trying to remove business cycle influences from productivity growth, we adopt the Burns-Mitchell-Baxter-King characterization of these cycles as having durations of up to 8 years in length. Our ideal filter for quarterly data is therefore a symmetric low-pass filter which blocks all frequencies above $\pi / 16$. This together with the results presented in Section 3.1 are all we need to construct the optimal filter. Its properties are described in Figure 2 and Table 2.

The third panel in Figure 2 (middle row, left column) compares the weights of the optimal and ideal filters. In the case of the ARMA model, the two are almost identical. This is because the ARMA model implies quarterly productivity growth is almost serially uncorrelated; this is almost the simple case discussed in Section 2.3. The weights for the kernel model are similar, but die away more slowly and are somewhat more volatile, reflecting the somewhat greater persistence in the series which the kernel detects.

Panels 4 and 6 in Figure 2 (right column, middle and bottom rows) compare the spectral properties of the three filters and the two optimally-filtered series. The gain of the optimal kernel filter shows a pattern similar to that shown in Figure 1 for other truncated ideal filters; the biggest difference being the absence of a lower bound on the filter. [19] The optimal ARMA filter has a somewhat similar shape, but tends to have a lower squared gain at most of the frequencies shown. Note that the kernel filter has a particularly high gain at the start of the stop band (about 0.8 near $\omega = 0.2$); this is roughly double the corresponding squared gain for the ARMA filter. This difference is to be expected given the peaks and

18. ARMA model selection for the CPI data was problematic. It is doubtful that the MA(12) adequately captures the persistence of inflation, since this would imply that inflation shocks completely die out in 12 months.

19. To understand why a filter with a gain everywhere less than 1 may still be optimal, consider the effect of scaling all the filter weights by some constant k>1. This has the effect of scaling the squared gain everywhere by k^2. In the case of the ARMA filter, this will reduce the difference between its gain and that of the ideal filter at frequencies below the cutoff frequency (i.e. reducing compression), thereby improving the estimate. However, this benefit is counterbalanced by the effect of increasing the difference between the two filters at frequencies above the cutoff (i.e. increasing leakage.) The optimal scale is the one at which the marginal benefits at some frequencies of a change in scale are exactly equal to the marginal costs at all other frequencies.

valleys of the kernel-estimated spectrum in these ranges; there is a dip in the density around $\omega = 0.2$ and so leakage in this area is less important for the kernal filter than for the ARMA filter.

Multiplying the spectral density at a given frequency from Panel 2 by the squared gain at that frequency from Panel 4 gives us the spectrum of the filtered series shown in Panel 6. In contrast to the relatively flat spectra shown in Panel 2, our low-pass filters have succeed in powerfully concentrating the spectrum of the filtered series over the desired frequencies. At the same time, however, the imperfections of the optimal filters are clearly visible; the spectral density at the low frequencies is much less than that in the raw data, and the ARMA-filtered spectrum has a small artificial peak at the cutoff frequency, which implies a modest, but spurious, cycle in the filtered series.[20]

The overall performance of the filter across all frequencies is summarized by the statistics presented in Table 2. The variance of the raw series (line 1) is simply the area under the spectra shown in Panel 2 of Figure 2.[21] Similarly, the variance of the ideally filtered series (line 2) is the area under those spectra lying between frequencies 0 and $\pi / 16$. In this case, the ideally-filtered trend accounts for under 20% of the total variance of the observed series. The optimally-filtered estimate of that trend captures betwee half and three-quarters of the variance of the ideal trend. The difference between the ARMA and Kernel models becomes more apparent when this is expressed in terms of correlations and noise-signal ratios. Here we see that the Kernel model implies an N/S ratio less than half that of the ARMA model , as well as an 86% correlation with the ideal estimate compared to 74.5% for the ARMA model.

The fifth panel of Figure 2 (bottom row, left column) shows the degree of phase lag implied by the filters. The last row of Table 2 shows the mean phase lag, which simply uses the spectral density of the filtered series (the sixth panel in Figure 2) to produce a weighted average lag for the filtered series. We see that the phase lag is positive at all the frequencies shown, increasing as we move from the highest frequencies to those near the cutoff. However, these frequencies have little weight in the filtered series, whose density is concentrated in the pass-band below the 32Q cutoff. Here, the phase lag varies across the two estimates, with that for the kernel estimate dropping from a high of about 5 periods to a low near 1. The ARMA estimate gives a larger lag than the kernel estimate everywhere in the pass band, with the difference becoming larger as the frequency approaches zero. As a result, the mean phase lag is almost a full year (see Table 2) for the ARMA estimate but only just over half a year for the kernel estimate.

20. Of course, the current application is designed to produce a single point estimate rather than a series, so this point may be moot.

21. Strictly speaking, it is twice that area, since the full spectrum is symmetric about 0; the figure shows only half that range. This applies to the analysis of subsequent rows in this table as well.

3.3 Filtering Inflation

As the objective of price stability has moved to the forefront of monetary policy formulation throughout much of the world over the past decade and a half, more attention has been given to the question of how to measure and monitor progress towards this goal. Most monetary authorities propose a modified (or "core") measure of inflation which aims to capture persistent trends in inflation or inflationary pressures.[22] It would therefore be of interest to use band-pass filters to construct optimal measures of current trends in inflation, where these trends are again defined using a low-pass filter. The appropriate cut-off frequency to use for such a filter is debatable; for the example we study in this section a frequency of $\pi/24$ (corresponding to cycles lasting 48 months) is used in order to give seasonal influences and short-run nominal shocks ample time to dissipate. Results are presented in Table 3 and Figure 3.

The third panel (left column, middle row) in Figure 3 shows us that the optimal weights now put a much higher weight on the most recent observations than the ideal weights. Again, the ARMA-model weights converge much more quickly to the ideal weights than the kernel model, presumably reflecting the greater persistence in the kernel-estimated dynamics.

The fourth panel (right column, middle row) shows that the gain functions of the optimal filters are again different from that of the ideal filter in several respects. The gain function for the ARMA filter again resembles the shape we encountered in Figure 1, with a gain of less than 0.5 for most of the pass band, and a gain near 1 for only a narrow band near the cutoff frequency. Although the gain drops sharply beyond that point, it stays significantly above zero for the rest of the graphed frequency range. The gain of the kernel filter resembles that of the ARMA filter over the pass band, with the exception of a near-zero gain at frequency zero. Outside the pass-band, however, there are multiple peaks with gains close to or exceeding 50%.

The sixth panel (right column, bottom row) shows that despite the apparently irregular gain functions, both of the resulting filtered series capture most of the density of the raw series at the low frequencies and have a very sharp drop in density at the cutoff frequency, with very little density at the higher frequencies. This reflects the fact that both filters have sharp drops in gain at the cutoff frequency, and that the potentially large leakage they allow from much higher frequencies is relatively unimportant due to their lack of importance in the original spectrum. Similarly, we may note that the variable gain of the kernel filter within the pass-band does not appear to greatly distort the spectrum of inflation within this band. This is because peaks in the raw spectrum correspond to frequencies around which the gain is not far from 1, while areas of particularly low gain in the pass-band correspond to troughs in the raw spectrum.

Table 3 confirms the relatively good performance of optimal filter for both the ARMA and kernel-based models. Both give estimates which have correlations of over 88% with the

22. Unlike the techniques examined in this paper, many other approaches to measuring core inflation rely on the analysis of disaggregated price movements.

true trend rate of inflation, and their noise-signal ratios are both under 30%. This improved performance is consistent with the basic intuition we saw in Section 2.3, that persistence tends to improve the quality of current estimates of the trend. We can understand this in terms of the Stock-Watson two-step procedure; the more persistent a series is, the better we are able to forecast it and therefore the better we approximate the ideal filter. Another way to understand these results is to recall that as persistence increases, the spectrum of our data series becomes increasingly concentrated in the lowest frequencies. Since our goal is to design a low-pass filter, this in turn reduces the importance of leaking higher frequencies, allowing us to increase the average gain of the optimal filter and thereby better approximate the ideal filter.

Although the phase lag of these filters varies by frequency, the lags are smaller than in the productivity growth case and rarely exceed half a year (6 months.) The lag for both filters peaks at the cutoff frequency and for the kernel filter it then quickly falls to zero in both directions. As a result, its mean phase lag is a trivial 0.68 months. Phase lag for the ARMA filter falls off more gradually, however, resulting in a mean lag of just over one quarter.

3.4 Filtering GDP

Optimal filters for business cycles and their properties are described in Figure 4 and Table 4. Unlike the two previous filters we have considered, this is a band-pass rather than a low-pass filter, and it uses the values suggested by Baxter and King $(\pi/16, \pi/3)$ to define the frequencies of interest.

Panel 3 (left column, middle row) of Figure 4 shows that the optimal ARMA weights are indistinguishable from the ideal weights, while the kernel weights resemble the ideal weights in many respects. Panel 4 (right column, middle row) shows that the squared gain of the optimal ARMA filter again resembles the patterns we saw in Figure 1, but that for the kernel filter is different in some important respects. It has a non-trivial gain at zero frequency, a pronounced peak (with a gain > 1) in the middle of the pass-band and very low gain as it approaches the high-frequency cutoff. However, it closely resembles the squared gain of the ARMA filter near the low-frequency cut-off and above the high-frequency cut-off.

Panel 6 (right column, bottom row) shows that both filtered series have effectively blocked the high-frequency part of the original spectrum, but that both seem to pass significant amounts at the lowest frequencies. Results in Table 4, similar for both models, apparently reflect this low-frequency leakage. The correlation of optimal and ideal filter estimates is about 80%, and the noise-signal ratio ranges from 46% for the kernel model to 60% for the ARMA model.

Phase lag for these models is small for most frequencies with the exception of those below the lower cutoff, where it becomes positive and large (peaking with a lag of over two years) for the kernel model, and negative (i.e. a phase *lead*) and large (peaking at three years) for the ARMA model. However, the relatively low spectral density of the filtered series in this region greatly reduces their overall impact. As shown in Table 4, mean phase lag is negative for both series, with a lead of roughly two and a half quarters for the ARMA model and less than one quarter for the kernel model.

3.4.1 The reliability of current estimates of the output gap

Estimates of the reliability of output gaps are of particular interest for the design of optimal monetary policies. It is therefore of interest to compare the above results with other recent estimates in the literature.

Christiano and Fitzgerald (1999) solve roughly the same optimal filter problem solved in this paper.[23] However, on the basis of experiments with models in the IMA(1,q) class, they conclude that most economic time series can be nearly-optimally filtered if we assume that they are random walks and solve for the corresponding approximately optimal filter.[24] In that case, the filter weights become functions of only the cutoff frequencies and the sample size. They calculate that the correlation between their nearly-optimal filter and the ideally-filtered measure of the business cycle is roughly 0.65 and that the noise-signal ratio is 0.77.[25] This is a lower correlation than we find for either model, while their noise-signal ratio is higher.

Orphanides and van Norden (1999, 2002) and Cayen and van Norden (2002) study other filters which do not have optimal band-pass properties but are nonetheless used to measure business cycles. They compare the rolling estimates produced when such filters are applied at the end of sample to historical estimates produced after many subsequent years of data are available.[26] The size of this revision in estimated business cycles corresponds to the difference between our ideal and optimal filter estimates. Using US data, Orphanides and van Norden (1999) find correlation coefficients ranging from 0.63 to 0.96, depending on the model used.[27] Cayen and van Norden (2002) use Canadian data and finds correlation coefficients for the same models ranging from 0.70 to 0.84.[28] We can also reconstruct noise-signal ratios for the former paper based on the ratio of the reported standard deviation of the revisions to the standard deviation of the final measure of the output gap. These figures are reported in Table 5; the values range from 0.34 to 0.79.

---

23. See the Annex for a discussion.

24. The fact that they use low-order IMA models presumably guarantees that the optimal weights and their approximate weights will differ only for the last few observations, and even then not very much. It would be interesting to see whether the usefulness of the random walk approximation would be sustained if kernel or ARIMA models were instead used to derive optimal filters.

25. The correlation is taken from the end-point of the graph in the left column, middle row of their Figure 6, while the signal-noise ratio is given in the discussion on p. 21 in Christiano and Fitzgerald (1999). Note that their band-pass filter is set to pass all cycles with durations from 2 to 8 years versus the 6 to 32 quarters used here.

26. In the terminology of these papers, these are the Final and the QuasiFinal estimates. The Final - QuasiFinal revision is a better analogue to the estimation error considered in this paper since both ignore the role of uncertainty in the underlying data generating process.

27. Orphanides and van Norden (1999), Table 1, p. 32. Results are correlations between Quasi-Final and Final estimates for the Watson, Clark and Harvey-Jaeger Models.

28. Cayen and van Norden (2002), Table 1, p. 33.

While it may appear counter-intuitive that non-optimal models appear sometimes to give better correlations or noise-signal ratios than the optimal measures developed in this paper, it should be remembered that the two are not strictly comparable since their definitions of trend and cycle differ. However, these results suggest that the optimal frequency-based techniques should not be expected to give markedly more accurate estimates than other sophisticated time-series methods.

4.0 Conclusions

The derivations contained in the annex show how to construct optimal band-pass filters for them ARIMA models commonly used with macroeconomic time series. Together with the above applications of such filters, this produces several interesting results.

First, it illustrates how the accuracy of such filters may vary considerably. When a series has little or no predictability, as is the case for the ARMA model of productivity growth, this limits our ability to measure the current long-term trend. We found therefore that current estimates of trend productivity growth based on the ARMA model have correlations of about 75% with comparable estimates constructed with the benefit of hindsight; put another way, their noise-signal ratio is about 80%. The ARMA model therefore implies that the measurement of productivity trends is the most difficult of the three problems considered in this paper. The kernel model suggests that productivity trend measurement is substantially easier, however. Reconciling these results is left to future work.

The results for inflation illustrate how increased predictability improves our measurement of current trends. Both of the models we examined imply that current inflation trends can be measured with considerable accuracy, giving about 90% correlations with the best ex-post measures and noise-signal ratios just above 20%.

Persistence acts as a double-edged sword, however, when we seek to measure current cycles, as shown by the results for business cycle measurement. On the one hand, it improves the amount of information available about the future of the series, thereby reducing the difference between current and future estimates of trend. On the other hand, by increasing the relative amount of noise to be filtered out, it increases the potential effects of leakage and therefore of measurement error. In the case of business cycle measurement, we see that optimal filters do not perform especially well; their correlations and noise-signal ratios are similar to those for productivity growth.

The results for business cycle measurement are somewhat surprising in light of previous work examining the performance of non-frequency-based models of trend and cycles. Comparison of these results seems to show that the latter will in some cases perform as well or better than the optimal methods analysed here. The comparisons are potentially misleading, however, since the definitions of trend and cycle are not comparable across models. The previous work also focused on models in which output was assumed to follow a stochastic trend rather than the deterministic trend assumed here. A reconciliation of these results should examine the extent to which the results for frequency-based filters are sensitive to the assumption of trend-stationarity.

More generally, the sensitivity of optimal filters to the assumed dynamics of the data series requires further evaluation. This would allow a closer scrutiny of Christiano and Fitzgerald's (1999) claim that the assumption of random walk dynamics is adequate for most macroeconomic time series, which is a potentially important simplification for applied work. It would also have implications for the accuracy with which business cycles may realistically be measured. The only source of error considered in this paper's analysis is the extent to which estimated cycles will be revised as new observations become available. As noted in Orphanides and van Norden (1999, 2002), other sources include estimation error in the autocovariance function, data revision, and model misspecification. The results presented above should therefore be viewed as lower bounds on the total measurement error in frequency-based estimation of current trends and cycles.

5.0 Annex: Derivation of Optimal Filter

This Annex presents a derivation of the optimal one-sided band-pass filter. It draws on Christiano and Fitzgerald (1999), who derive the optimal filtered estimated at each point in the sample for an IMA(1,q) process.The proof is given in two stages. In section 5.2 we derive the solution under the assumption that our process is a stationary MA(q) process with finite variance. Sections 5.3 and 5.4 discuss extensions to the ARMA and ARIMA cases. Section 6.0 then compares the contribution of Christiano and Fitzgerald (1999) to that of Koopmans (1974) and provides additional intuition about the nature of the solution.

After this Annex was prepared, Schleicher (2001) provided a more compact derivation of the optimal filter for the ARIMA case, which also extends the results to below to any arbitrary point in the sample instead of just the endpoint.

5.1 Notation and Basic Assumptions

We have a sample of T discrete observations on some stochastic series $\{y_t\}$. We assume that $\{y_t\}$ contains no deterministic components (constants or non-stochastic trends).[29] For the remainder of this section, we also assume that $\{y_t\}$ is covariance-stationary with a known MA(q) representation for finite q. We also require that the covariance matrix of $[y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-T+1}]$ exists and has a unique inverse. Finally, we will assume that the power spectrum of y also exists and is given by[30]

$$f_y(\omega) = \frac{1}{2\pi} \cdot \sum_{k=-\infty}^{\infty} \sigma_y(k) \cdot e^{-i\omega k} \qquad \text{(EQ 8)}$$

where $\sigma_y(k)$ is the covariance between y_t and y_{t-k}.

Ideally, we would want to choose $\{x_t, c_t\}$ to partition the spectrum of y_t, so that c_t contains all the fluctuations in y_t with frequencies between some lower limit l and some upper limit u, $0 < l < u \le \pi$, and x_t contains those fluctuations with frequencies below l or above u.[31] We can do so with an infinite-order time-invariant linear filter

$$B(L) \equiv \sum_{j=-\infty}^{\infty} B_j \cdot L^j \qquad \text{(EQ 9)}$$

29. Typically, one would regress the raw data series on a polynomial time trend and use the residuals as y_t, which may be stationary or stochastically integrated. As in all spectral analysis, we ignore the potentially important effects of any imperfections in this detrending.

30. Covariance stationarity plus finite variance is sufficient to guarantee the properties on the covariance matrix. The power spectrum conditions will be satisfied if the coefficients on the MA representation are absolutely summable; covariance stationarity guarantees that they are square summable.

31. For example, isolating the fluctuations between 6 and 32 quarters in length with quarterly data would correspond to $l = \pi/16$ and $u = \pi/3$.

such that $c_t = B(L) \cdot y_t$ and $x_t = y_t - c_t = y_t \cdot (1 - B(L))$ where L is the lag operator. The coefficients of this ideal filter are given by [32]

$$B_j = \frac{\sin ju - \sin jl}{\pi j} \text{ for } j \neq 1 \qquad \text{(EQ 10)}$$
$$= \frac{u-l}{\pi} \text{ otherwise}$$

Our problem is to choose a set of filter weights $\{\hat{B}_j\}$ to solve

$$\min_{\{\hat{B}_j\}} \int_{-\pi}^{\pi} \left|B(e^{-i\omega}) - \hat{B}(e^{-i\omega})\right|^2 \cdot f_y(\omega) d\omega \qquad \text{(EQ 11)}$$

where

$$\hat{B}(z) = \sum_{j=0}^{T-1} \hat{B}_j \cdot z^j \qquad \text{(EQ 12)}$$

for complex z on the unit circle.

5.2 Solution for the MA(q) case

5.2.1 First Order Conditions

Rewrite (EQ 11) as

$$\min_{\{\hat{B}_j\}} \int_{-\pi}^{\pi} \delta(\omega) \cdot \delta(-\omega) \cdot f_y(\omega) d\omega \qquad \text{where } \delta(\omega) = B(e^{-i\omega}) - \hat{B}(e^{-i\omega}) \qquad \text{(EQ 13)}$$

Differentiating (EQ 13) with respect to each element of $\{\hat{B}_j\}$, we obtain T first order conditions

$$0 = \int_{-\pi}^{\pi} \left(\delta(\omega) \cdot \frac{\partial}{\partial \hat{B}_j}\delta(-\omega) + \delta(-\omega) \cdot \frac{\partial}{\partial \hat{B}_j}\delta(\omega)\right) \cdot f_y(\omega) d\omega \qquad \forall j = 0, \ldots, T-1 \quad \text{(EQ 14)}$$

In fact

32.Christiano and Fitzgerald (1999) further impose the restriction that $\sum_j B_j = 0$, which is reasonable except in the special case of low-pass filters.

$$\frac{\partial}{\partial \hat{B}_j}\delta(-\omega) = -e^{i\omega j} \text{ and } \frac{\partial}{\partial \hat{B}_j}\delta(\omega) = -e^{-i\omega j} \qquad \text{(EQ 15)}$$

so (EQ 14) becomes

$$0 = \int_{-\pi}^{\pi} (\delta(\omega)\cdot -e^{i\omega j} + \delta(-\omega)\cdot -e^{-i\omega j})\cdot f_y(\omega)d\omega$$

$$\therefore \int_{-\pi}^{\pi} [B(e^{-i\omega})\cdot e^{i\omega j} + B(e^{i\omega})\cdot e^{-i\omega j}]\cdot f_y(\omega)d\omega = \qquad \text{(EQ 16)}$$

$$\int_{-\pi}^{\pi} [\hat{B}(e^{-i\omega})\cdot e^{i\omega j} + \hat{B}(e^{i\omega})\cdot e^{-i\omega j}]\cdot f_y(\omega)d\omega \qquad \forall j = 0, \ldots, T-1$$

Noting that $f_{\Delta y}(\omega)$ is symmetric about 0, we obtain

$$\int_{0}^{\pi} [B(e^{-i\omega})\cdot e^{i\omega j} + B(e^{i\omega})\cdot e^{-i\omega j}]\cdot f_y(\omega)d\omega = \qquad \text{(EQ 17)}$$

$$\int_{-\pi}^{0} [B(e^{-i\omega})\cdot e^{i\omega j} + B(e^{i\omega})\cdot e^{-i\omega j}]\cdot f_y(\omega)d\omega$$

and

$$\int_{0}^{\pi} [\hat{B}(e^{-i\omega})\cdot e^{i\omega j} + \hat{B}(e^{i\omega})\cdot e^{-i\omega j}]\cdot f_y(\omega)d\omega = \qquad \text{(EQ 18)}$$

$$\int_{-\pi}^{0} [\hat{B}(e^{-i\omega})\cdot e^{i\omega j} + \hat{B}(e^{i\omega})\cdot e^{-i\omega j}]\cdot f_y(\omega)d\omega$$

(EQ 16) implies

$$\int_{0}^{\pi} [B(e^{-i\omega})\cdot e^{i\omega j} + B(e^{i\omega})\cdot e^{-i\omega j}]\cdot f_y(\omega)d\omega = \int_{0}^{\pi} [\hat{B}(e^{-i\omega})\cdot e^{i\omega j} + \hat{B}(e^{i\omega})\cdot e^{-i\omega j}]\cdot f_y(\omega)d\omega$$

$$\therefore \int_{-\pi}^{\pi} B(e^{-i\omega})\cdot e^{i\omega j}\cdot f_y(\omega)d\omega = \int_{-\pi}^{\pi} \hat{B}(e^{-i\omega})\cdot e^{i\omega j}\cdot f_y(\omega)d\omega \qquad \forall j = 0, \ldots, T-1 \qquad \text{(EQ 19)}$$

Using the fact that $\int_{-\pi}^{\pi} e^{i\omega j}d\omega = 2\pi$ when $j = 0$ and $\int_{-\pi}^{\pi} e^{i\omega j}d\omega = 0$ when $j = 1, 2, 3, \ldots$, we can see that (EQ 19) simply picks out one term at a time from the convolutions of the filter coefficients and the spectrum of y.

Because $\sigma_y(k)$ has the properties $\sigma_y(k) = \sigma_y(-k)$ and $\sigma_y(k) = 0 \;\; \forall k > q$, we can write

$$\int_{-\pi}^{\pi} B(e^{-i\omega})\cdot e^{i\omega j}\cdot f_y(\omega)d\omega = 2\pi\cdot \sum_{k=-q}^{q} B_{j+k}\cdot \sigma_y(k) = 2\pi\cdot \vec{B}^{j\prime}\cdot \vec{\sigma}_y \qquad \text{(EQ 20)}$$

where

$\vec{B}^j = [B_{j-q}, B_{j-q+1}, B_{j-q+2}, \ldots, B_{j+q}]'$, a *2q+1* column vector

$\vec{\sigma}_y = [\sigma_y(-q), \sigma_y(-q+1), \sigma_y(-q+2), \ldots, \sigma_y(q)]'$, a $2q+1$ column vector.

We can similarly transform the right-hand side of (EQ 19), taking care to account for the truncation of $\hat{B}(e^{-i\omega})$, and obtain

$$\int_{-\pi}^{\pi} \hat{B}(e^{-i\omega}) \cdot e^{i\omega j} \cdot f_y(\omega) d\omega = 2\pi \cdot \sum_{k=-j}^{T-1-j} \hat{B}_{j+k} \cdot \sigma_y(k) = 2\pi \cdot \vec{\beta}' \cdot \vec{\sigma}_{y,j} \qquad \text{(EQ 21)}$$

where

$\vec{\beta} = [\hat{B}_0, \ldots, \hat{B}_{T-1}]'$, a length T column vector

$\vec{\sigma}_{y,j} = [\sigma_y(-j), \sigma_y(-j+1), \sigma_y(-j+2), \ldots, \sigma_y(T-1-j)]'$, a length T column vector.

This allows us to rewrite (EQ 19) compactly as

$$\vec{B}^{j\prime} \cdot \vec{\sigma}_y = \vec{\sigma}_{y,j}' \cdot \vec{\beta} \qquad \forall j = 0, \ldots, T-1 \qquad \text{(EQ 22)}$$

5.2.2 Solving the system of FOCs

In (EQ 22), $\vec{\sigma}_y$ and $\vec{\sigma}_j$ are known and determined by the data, while $\vec{B}^j$ is given by (EQ 3). This leaves us with T linear equations to solve for the T elements of $\vec{\beta}$. Stacking these T equations gives us

$$\vec{B} \cdot \vec{\sigma}_y = \bar{\Sigma}_y \cdot \vec{\beta} \qquad \text{(EQ 23)}$$

where

$\vec{B} = [\vec{B}^0, \vec{B}^1, \vec{B}^2, \ldots, \vec{B}^{T-1}]'$, a T x $2q+1$ matrix

$\bar{\Sigma}_y = [\vec{\sigma}_{y,0}, \vec{\sigma}_{y,1}, \vec{\sigma}_{y,2}, \ldots, \vec{\sigma}_{y,T-1}]'$, a T x T matrix.

The solution is

$$\vec{\beta} = \bar{\Sigma}_y^{-1} \cdot \vec{B} \cdot \vec{\sigma}_y \qquad \text{(EQ 24)}$$

provided that $\bar{\Sigma}_y^{-1}$ exists and is unique. We know that

$$\vec{\Sigma}_y = \begin{bmatrix} \sigma_y(0) & \sigma_y(1) & \sigma_y(2) & \dots & \sigma_y(T-1) \\ \sigma_y(-1) & \sigma_y(0) & \sigma_y(1) & \dots & \sigma_y(T-2) \\ \sigma_y(-2) & \sigma_y(-1) & \sigma_y(0) & \dots & \sigma_y(T-3) \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_y(-T+1) & \sigma_y(-T+2) & \sigma_y(-T+3) & \dots & \sigma_y(0) \end{bmatrix} \qquad \text{(EQ 25)}$$

$$= \begin{bmatrix} \sigma_y(0) & \sigma_y(1) & \sigma_y(2) & \dots & \sigma_y(T-1) \\ \sigma_y(1) & \sigma_y(0) & \sigma_y(1) & \dots & \sigma_y(T-2) \\ \sigma_y(2) & \sigma_y(1) & \sigma_y(0) & \dots & \sigma_y(T-3) \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_y(T-1) & \sigma_y(T-2) & \sigma_y(T-3) & \dots & \sigma_y(0) \end{bmatrix}$$

which is therefore just the covariance matrix of the vector $[y_t, y_{t-1}, y_{t-2}, \dots, y_{t-T+1}]$. Our solution is therefore applicable whenever this covariance matrix exists and is of full rank.

5.3 The Stationary ARMA(p,q) Case

If we wish to extend our analysis from the MA(q) to the ARMA(p,q) case, we can use to Wold Representation Theorem to recast the ARMA model as an infinite-order MA. Assuming that the other conditions mentioned before (EQ 8) are respected, we can continue to write our optimization problem as before and derive the same first-order conditions. The analysis then proceeds until (EQ 20), which becomes

$$\int_{-\pi}^{\pi} B(e^{-i\omega}) \cdot e^{i\omega j} \cdot f_y(\omega) d\omega = 2\pi \cdot \sum_{k=-\infty}^{\infty} B_{j+k} \cdot \sigma_y(k) = 2\pi \cdot \vec{B}^{j\prime} \cdot \vec{\sigma}_y \qquad \text{(EQ 26)}$$

where $\vec{B}^j$ and $\vec{\sigma}_y$ are now doubly-infinite-dimensional vectors. We require only that their dot-product is well defined and finite. Stacking our T equations in (EQ 23) proceeds as before, the only change being the dimensions of $\vec{B}$ and $\vec{\sigma}_y$ but not their product. The final solution in (EQ 5) is therefore unchanged.

Applying the filter to ARMA processes will in practice require that we truncate the infinite sum at some point, hoping that the omitted covariance terms are sufficiently close to zero. Of course, for given parameters for the ARMA process, theoretical autocovariances may easily be calculated, so that the only limitation to the precision of our calculations are computing power and storage capacity. It would be irrealistic, however, to believe that we can infer much about autocovariances beyond the *N-1*th lag from a sample of only *N* observations.

5.4 The ARIMA(p,d,q) Case

We can try to analyse nonstationary ARIMA(p,d,q) processes by first noting that after differencing *d* times, we have ARMA(p,q) processes. It would therefore be tempting to simply assume a sample of size *T+d,* difference it *d* times, and then apply the analysis as for stationary ARMA models to the differenced data. There are two faults with this approach, both of which can be seen in our objective function (EQ 11). First, instead of attempting to match the ideal band-pass filter $B(L)$ on the nonstationary data, we are trying to match it on the differenced data, which is equivalent to setting $B(L) \cdot (1-L)^d$ as our target filter. Secondly, the power spectrum used to weight the deviations from the ideal filter would now be $f_{\Delta^d y}(\omega)$ instead of $f_y(\omega)$; it is not clear how to motivate such a choice in practice.

Instead, note that (EQ 3) implies that we can factor

$$B(L) = (1-L) \cdot b(L) = (1-L)^2 \cdot bb(L) \qquad \text{(EQ 27)}$$

$$b(L) = \sum_{j=-\infty}^{\infty} b_j \cdot L^j \text{ and } bb(L) = \sum_{j=-\infty}^{\infty} bb_j \cdot L^j \qquad \text{(EQ 28)}$$

This implies that the ideal band-pass filter will render stationary any series integrated of order no more than 2.[33] From our objective function in (EQ 11), we can see that its value will be less than infinity if and only if our one-sided filter $\hat{B}(L)$ also renders y_t stationary. We can therefore presume that the optimal one-sided filter may similarly be factored as

$$\hat{B}(L) = (1-L) \cdot \hat{b}(L) = (1-L)^2 \cdot \hat{bb}(L) \qquad \text{(EQ 29)}$$

$$\hat{b}(L) = \sum_{j=-\infty}^{\infty} \hat{b}_j \cdot L^j \text{ and } \hat{bb}(L) = \sum_{j=-\infty}^{\infty} \hat{bb}_j \cdot L^j \qquad \text{(EQ 30)}$$

This common factoring of the ideal and one-sided filters allows us to find solutions in the cases of I(1) and I(2) processes. In the first subsection, below, we detail the proof for the case of IMA(1,q) processes. Thereafter, we briefly discuss the IMA(2,q) case. Extensions to the ARIMA case should be clear from the above discussion of the ARMA case.

5.4.1 The IMA(1,q) Case

Suppose we have *T+1* observations on a process $y_t \sim IMA(1, q)$. Our minimization problem in (EQ 13) can then be rewritten as

33.See Den Haan and Sumner (2001) for a related discussion.

$$\min_{\{\hat{b}_j\}} \int_{-\pi}^{\pi} \delta(\omega) \cdot (1 - e^{-i\omega}) \cdot \delta(-\omega) \cdot (1 - e^{i\omega}) \cdot f_y(\omega) d\omega \qquad \text{(EQ 31)}$$

where now $\delta(\omega) = b(e^{-i\omega}) - \hat{b}(e^{-i\omega})$. We can also now factor

$$f_y(\omega) = f_{\Delta y}(\omega) \cdot (1 - e^{-i\omega})^{-1} \cdot (1 - e^{i\omega})^{-1} \qquad \text{(EQ 32)}$$

where $f_{\Delta y}(\omega)$ is the power spectrum for the first-difference of y_t. This means that (EQ 31) simplifies to

$$\min_{\{\hat{b}_j\}} \int_{-\pi}^{\pi} \delta(\omega) \cdot \delta(-\omega) \cdot f_{\Delta y}(\omega) d\omega \qquad \text{(EQ 33)}$$

We now have an optimization problem that is analogous to our original problem in (EQ 11); the only differences are three substitutions:

1. we use the power spectrum of Δy_t rather than of y_t.
2. the function we seek to match is now $b(L)$ rather than $B(L)$
3. the optimization is over the *T* coefficients of the differenced filter $\hat{b}_j$ rather than the *T+1* coefficients $\hat{B}_j$.

As in the original proof, we work with the spectrum of a stationary series (now Δy_t) on which we have *T* observations. We therefore arrive at an analogous first-order condition, replacing (EQ 22) with

$$\vec{b}^{j\prime} \cdot \vec{\sigma}_{\Delta y} = \vec{\sigma}_{\Delta y, j}{}' \cdot \vec{\beta}_1 \quad \forall j = 0, \ldots, T-1 \qquad \text{(EQ 34)}$$

where

$\vec{b}^j = [b_{j-q}, b_{j-q+1}, b_{j-q+2}, \ldots, b_{j+q}]'$, a *2q+1* column vector

$\vec{\sigma}_{\Delta y} = [\sigma_{\Delta y}(-q), \sigma_{\Delta y}(-q+1), \sigma_{\Delta y}(-q+2), \ldots, \sigma_{\Delta y}(q)]'$, a *2q+1* column vector

$\sigma_{\Delta y}(k)$ is the autocovariance function for Δy_t

$\vec{\sigma}_{\Delta y, j} = [\sigma_{\Delta y}(-j), \sigma_{\Delta y}(-j+1), \sigma_{\Delta y}(-j+2), \ldots, \sigma_{\Delta y}(T-1-j)]'$, a length *T* column vector.

$$\vec{\beta}_1 = [\hat{b}_0, \ldots, \hat{b}_{T-1}]'$$, a length *T* column vector

These may be stacked and solved for the optimal $b(L)$, given by

$$\overrightarrow{\beta_1} = \vec{\Sigma}_{\Delta y}^{-1} \cdot \vec{b} \cdot \vec{\sigma}_{\Delta y} \qquad \text{(EQ 35)}$$

where

$$\vec{b} = [\vec{b}^0, \vec{b}^1, \vec{b}^2, \ldots, \vec{b}^{T-1}]'$$, a *T x 2q+1* matrix

$$\bar{\Sigma}_{\Delta y} = [\vec{\sigma}_{\Delta y, 0}, \vec{\sigma}_{\Delta y, 1}, \vec{\sigma}_{\Delta y, 2}, \ldots, \vec{\sigma}_{\Delta y, T-1}]'$$, a *T x T* matrix.

While (EQ 35) determines $\overrightarrow{\beta_1}$, this gives us only *T* conditions with which to identify the *T+1* coefficients $\hat{B}_j$. The remaining condition is (EQ 29), which implies that the sum over *j* of $\hat{B}_j$ must equal zero.

5.4.2 The IMA(2,q) Case

The proof for the IMA(2,q) case with *T+2* observations proceeds analogously to the IMA(1,q) case; we again transform the original minimization problem into one involving the power spectrum of a stationary series. This means working throughout the proof with *T* observations on $\Delta^2 y$, its power spectrum $f_{\Delta^2 y}(\omega)$ and autocovariance function $\sigma_{\Delta^2 y}(k)$, the 2nd difference of the ideal band-pass filter $\hat{b}b(L)$ and recovering the optimal 1-sided filter for 2nd-differenced data $bb(L)$.

The basic intuition should by now be clear. Because we know that the ideal 2-sided filter will give us a stationary series, we can transform the optimization problem into an equivalent problem using suitably differenced data and a suitably differenced ideal filter. This approach may break down if we try to go beyond the I(2) case since the ideal filter is no longer sure to give a stationary filtered series. Fortunately, the vast majority of economic time series do not require models with orders of integration larger than 2.[34]

34.An alternative approach would be to change the definition of the ideal filter to ensure stationarity.

6.0 Annex: Historical Antecedents

The tools of frequency-domain analysis and signal processing theory have been available for over a quarter-century. The idea of deriving an optimal filter defined as in (EQ 11) is an obvious one, and the derivation of its closed-form solution is not difficult. Despite this, it is barely mentioned or discussed in modern econometrics or macroeconomics.[35] Christiano and Fitzgerald (1999) derive the optimal filtered estimated at each point in the sample for an IMA(1,q) process.[36] They make no specific references to previous derivations of this type. However, they refer to Stock and Watson (1998)'s time-domain procedure of forecasting future values of output to "pad" the sample and then applying the Baxter-King filter, but do not clearly state its relationship.[37]

However, Koopmans (1974) gives the closed-form solution for a closely related problem.[38] He considers the case of two weakly-stationary processes $\{x_t, y_t\}$, where we observe only y_t but we know their cross spectral density $f_{xy}(\omega)$ as well as their spectral densities $f_x(\omega), f_y(\omega)$. (Recall that knowing $f_{xy}(\omega)$ is equivalent to knowing their cross-covariance function $C_{xy}(k)$.) We wish to construct the linear filter $B_v(L)$ which minimizes

$$E([x_{t+v} - B_v(L) \cdot y_t]^2) \qquad \text{(EQ 36)}$$

subject to the restriction that $B_v(L)$ is a polynomial in only non-negative powers of L.

Koopmans' proof relies on an underlying intuition based on projections. The optimal one-sided filter is simply the projection from the space spanned by x_{t+v} into the space spanned by past and present values of y_t, $H_y^-(t)$. This can be broken down into two projection steps; the first being the projection of x_t onto the space spanned by past, present and future values of y_t (H_y), then projecting that result onto $H_y^-(t)$.[39]

He begins by noting that the optimal *two-sided* filter has a transfer function given by

$$B(\omega) = f_{xy}(\omega)/f_y(\omega) \qquad \text{(EQ 37)}$$

35.One reason may be that most introductory treatments of such techniques assume (1) that the data being analysed are stationary, and (2) that our data series are very long. Another reason may be that economists are confused about how to apply such techniques to nonstationary data; see Den Haan and Sumner's critique of Harvey and Jaeger (1993) and Cogley and Nason (1995) [*Ibid.*, section 3.2, esp. p. 11-12.]

36.They assert that the proof could extended to ARIMA (p,1,q), but that this would be "tedious." Schleicher (2001) provides just such an extension, with a proof which is considerably more elegant than the original proof in Christiano and Fitzgerald (1999).

37.This is the method used in the BPFILTER.SRC routine provided by Estima for RATS; the routine is Alan Taylor's translation of Watson's code for Stock and Watson (1998).

38.See Koopmans (1974), particularly Section 7.6 "Linear Filtering in Real Time," pp. 249-252 and Section 5.5 "Bivariate Spectral Parameters", esp. example 5.5 pp 147-148. Again, there is no mention of the origin of the derivation, presumably because it was considered too well-known or trivial a result.

39.The projection may be broken into two steps because $H_y^-(t)$ is a subspace of H_y.

Applying the optimal one-sided filter $B_V(L)$ to our available series $\{y_1, \ldots, y_T\}$ will give the same result as applying the doubly-infinite optimal two-sided filter $B(L)$ to a doubly-infinite "padded" series $\{\tilde{y}_t\}$ where $\tilde{y}_t = y_t$ for $t = 1, \ldots, T$, $\tilde{y}_t = \hat{y}_t$ otherwise. $\hat{y}_t$ is simply the optimal forecast/backcast value of y_t given the available sample from 1 to T. Stock and Watson (1998)'s approach is therefore an approximation of this which truncates the forecast/backcast at some finite values.

The relationship to the problem at hand can be understood by considering the latent variable x_t. Ideally, we would like to estimate $x_t =$ the (unobserved) output gap. This is possible if we know $C_{xy}(k)$, its autocovariance function with observed output. This will be the case if we postulate a structural model, as in the structural unobserved components used by Harvey (1985), Kuttner (1994), Gerlach and Smets (1997), Kaiser and Maravall (2001), etc. The optimal estimates we obtain from the above correspond to the solutions these authors obtain using the standard recursive Wiener-Kolmogorov filtering and smoothing equations. Of course, results may be highly sensitive to the specification of the structural model, as noted by Morley, Nelson and Zivot (1999).

The band-filtering approach remains agnostic about the structural model by choosing x_t to be the optimally band-filtered component of observed output. Since the form of the optimal filter $B(L)$ is known, $x_t = B(L) \cdot y_t$ and $f_{xy}(\omega) = B(\omega) \cdot f_y(\omega)$. We can therefore construct the optimal finite-sample approximation of x_t as a function of only the dynamic properties of output. In doing so, however, we ignore the fact that this definition of x_t is only an approximation of the output gap we ultimately care about. To see this, note that as our sample becomes large, our estimation error for x_t in the middle of our sample tends asymptotically towards zero in this approach, whereas in the unobserved approach we have asymptotic standard errors associated with our smoothed estimates.

6.1 Koopmans' Proof

Since y_t is weakly stationary, it has a Wold representation

$$y_t = \sum_{j=0}^{\infty} c_j \cdot \xi_{t-j} \quad \text{(EQ 38)}$$

where ξ_t is white noise and we define $C(\omega)$ to the transfer function associated with the above filter. The spectrum of y_t is therefore $f_y(\omega) = C(\omega) \cdot f_\xi(\omega)$. If $B(\omega)$ is (as above) the transfer function of the optimal symmetric band-pass filter, then since we seek to estimate x_{t+v}, we need to replace it with $B(\omega) \cdot e^{i\omega v}$ to account for the offset of *v* periods. It therefore follows that the optimal two-sided estimates of x_t are given by

$$\hat{x}_v(t) = \int_{-\pi}^{\pi} e^{i\omega t + v} \cdot B(\omega) \cdot C(\omega) \cdot f_\xi(\omega) = \int_{-\pi}^{\pi} D(\omega) \cdot f_\xi(\omega) \quad \text{(EQ 39)}$$

where

$$D(\omega) = \sum_{-\infty}^{\infty} d_j \cdot e^{i\omega j}$$

The one-sided problem simply replaces future values of ξ_t with their expectation, which is zero. This means that the optimal one-sided estimate is given by

$$\tilde{x}_v(t) = \int_{-\pi}^{\pi} D_v(\omega) \cdot f_{\xi}(\omega) \qquad \text{(EQ 40)}$$

$$\text{where } D_v(\omega) = \sum_{o}^{\infty} d_j \cdot e^{i\omega j} \qquad \text{(EQ 41)}$$

This derivation has its counterpart in the derivation for the MA(q) presented above in Section 5.2. (EQ 5) shows us that we are projecting the optimal filter $\vec{B}$ into the space spanned by our T observations on y_t via the $\vec{\sigma}_y$ and Σ_y terms. Note that there are effectively two distinct truncations occurring; one due to the limited persistence of the underlying MA process and another due to the finite sample size.[40]

40.The former is easy to overlook in Koopmans' treatment as he assumes a possibly infinite MA representation.

7.0 Bibliography

Baxter and King (1999), “Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series”, *Review of Economics and Statistics*, 81(4): 575-93.

Burns, Arthur F. and W. C. Mitchell (1947) *Measuring Business Cycles*, NBER, New York.

Cayen, Jean-Philippe and Simon van Norden (2002) “La fiabilité des estimations de l'écart de production au Canada”, *Bank of Canada working paper* 2002-10

Christiano, Lawrence C. and Terry J. Fitzgerald (1999) “The Band Pass Filter,” *Federal Reserve Bank of Cleveland working paper* 99-06, also published as NBER working paper 7257.

Cogley, T. and J. M. Nason (1995) “Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research,” *Journal of Economic Dynamics and Control*, 19, 253-278.

Den Haan, Wouter J. and Steven Sumner (2001) “The Comovements Between Real Activity and Prices in the G7” *University of California San Diego Department of Economics Working Paper*, 2001-05, 33 p.

Gerlach, Stefan and Frank Smets (1997) “Output Gaps and Inflation: Unobservable-Components Estimates for the G-7 Countries." Bank for International Settlements mimeo, Basel

Gomez, Victor (2001) “The Use of Butterworth Filters for Trend and Cycle Estimation in Economic Time Series,” *Journal of Business and Economic Statistics*, 19(3), 365-73.

Guay, Alain and Pierre St. Amant (1997) *Do Mechanical Filters Provide a Good Approximation of Business Cycles?*, Bank of Canada Technical Report No. 78, Ottawa.

Harvey, Andrew C. (1985) “Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 3, 216-227.

Harvey, Andrew C. and A. Jaeger (1993) “Detrending, Stylized Facts and the Business Cycle,” *Journal of Applied Econometrics*, 8, 231-247.

Hodrick, Robert and Edward Prescott (1997) “Post-War Business Cycles: An Empirical Investigation.” *Journal of Money, Credit and Banking*, 29:1, 1-16. (Working paper version cited as Hodrick and Prescott 1977.)

Kaiser, Regina and Agustin Maravall (2001) *Measuring Business Cycles in Economic Time Series,* forthcoming.

Koopmans, L. H. (1974) *The Spectral Analysis of Time Series*, Academic Press, 366 p.

Kuttner, Kenneth N. (1994) “Estimating Potential Output as a Latent Variable," *Journal of Business and Economic Statistics*, 12(3), 361-68.

Morley, James, Charles Nelson and Eric Zivot (1999) *Why are Beveridge-Nelson and Unobserved-Component Decompositions of GDP so different?* University of Washington mimeo.

Orphanides and van Norden (1999) “The Reliability of Output Gap Estimates in Real Time” *FEDS* discussion paper, 45 p.

Orphanides and van Norden (2002) “The Unreliability of Output Gap Estimates in Real Time” *The Review of Economics and Statistics*, 84(4) 569-83.

Ouliaris, Sam and P. C. B. Phillips (1995) *Coint 2.0: GAUSS Procedures for Cointegrated Regressions,* distributed by Aptech Systems.

Pedersen, Torben Mark (1998) “The Hodrick-Prescott Filter, The Slutzky Effect, and the Distortionary Effect of Filters,” University of Copenhagen, *Institute of Economics Discussion Paper* 98/09.

Pollock, D. S. G. (2000) “Trend Estimation and De-trending via Rational Square-Wave Filters,” *Journal of Econometrics*, 99(2), 317-34.

St. Amant and van Norden (1997) *Measurement of the Output Gap: A Discussion of Recent Research at the Bank of Canada*, Bank of Canada Technical Report No. 79.

Schleicher, Christoph (2001) “Approximate Bandpass Filters for ARIMA Models” University of British Columbia mimeo.

Setterfield, M. A., D. V. Gordon and L. Osberg (1992) “Searching for a Will o' the Wisp: An Empirical Study of the NAIRU in Canada,” *European Economic Review,* 36(1), 119-36.

Staiger, Douglas, James H. Stock and Mark W. Watson (1997) “The NAIRU, Unemployment and Monetary Policy” *Journal of Economic Perspectives*, 11:1, 33-49.

Stock and Watson (1998) “Business Cycle Fluctuations in U.S. Macroeconomic Time Series.” *NBER Working Paper* No. 6528, 83 p., prepared for *The Handbook of Macroeconomics*, edited by John B. Taylor and Michael Woodford.

8.0 Tables

TABLE 1. Estimated ARMA Models

Series	Q_t	π_t	Y_t
# AR Parameters	1	1	1
ρ	0.401	0.904	0.882
# MA Parameters	0	0	2
$\sum\theta_i$			0.380
$\sigma^2 \times 10^6$	10.38	3.45	26.18
# Observations	49	92	50
Frequency	Quarterly	Monthly	Quarterly

TABLE 2. Optimal Filter for Productivity Growth Trends

Statistic	ARMA Model	Kernel Model
1) Variance of Raw Series (x 10^6)[a]	12.37	17.90
2) Variance of Ideally Filtered Series (x 10^6)[b]	1.78	2.81
3) Variance of Optimally Filtered Series (x 10^6)[c]	0.99	2.08
4) MSE of Filtered Estimate (x 10^6)[d]	0.79	0.74
5) Correlation with Ideal Estimate[e]	0.745	0.859
6) N/S ratio[f]	0.799	0.355
7) Mean Phase Lag (Quarters)[g]	3.862	2.435

a. Differences between the ARMA and kernel estimates of the variance are due to approximations in the construction of the theoretical autocorrelations of the ARMA model; the kernel estimates precisely match the sample variance of the series.

b. Calculated as the integral of the spectral density over the interval [-*u*,*u*].

c. Calculated as the integral of the spectral density over the interval $(-\pi, \pi]$.

d. Calculated as (2) - (3).

e. Calculated as $\sqrt{(3)/(2)}$.

f. Calculated as (4) / (3).

g. Calculated as the integral of the product of the phase lag and the spectral density of the filtered series.

TABLE 3. Optimal Filter for Inflation Trends[a]

Statistic	ARMA Model	Kernel Model
1) Variance of Raw Series (x 10^6)	18.86	25.05
2) Variance of Ideally Filtered Series (x 10^6)	11.00	16.01
3) Variance of Optimally Filtered Series (x 10^6)	8.63	12.80
4) MSE of Filtered Estimate (x 10^6)	2.37	3.21
5) Correlation with Ideal Estimate	0.886	0.894
6) N/S ratio	0.275	0.251
7) Mean Phase Lag (Months)	3.33	0.68

a. See footnotes for Table 2.

TABLE 4. Optimal Filter for Business Cycles[a]

Statistic	ARMA Model	Kernel Model
1) Variance of Raw Series (x 10^6)	211.3	224.2
2) Variance of Ideally Filtered Series (x 10^6)	60.72	92.51
3) Variance of Optimally Filtered Series (x 10^6)	38.02	63.53
4) MSE of Filtered Estimate (x 10^6)	22.70	28.98
5) Correlation with Ideal Estimate	0.791	0.829
6) N/S ratio	0.597	0.456
7) Mean Phase Lag (Quarters)	-2.43	-0.70

a. See footnotes for Table 2.

TABLE 5. Reconstructed Noise-Signal Ratios from Orphanides and van Norden (1999)

Model	Errors[a] (1)	Final[b] (2)	Noise/Signal = (1)/(2)
Clark	1.11	2.11	0.53
Harvey-Jaeger	1.22	1.55	0.79
Watson	1.16	3.44	0.34

a. Figures are the reported standard deviations for Final - Quasi-Final revisions, taken from Orphanides and van Norden (1999) Table 4, p. 36.

b. Figures are the reported standard deviations for Final output gaps, taken from Orphanides and van Norden (1999) Table 1, p. 32.

Figure 1:
Gain^2 for Ideal and Approximate Band-Pass Filters

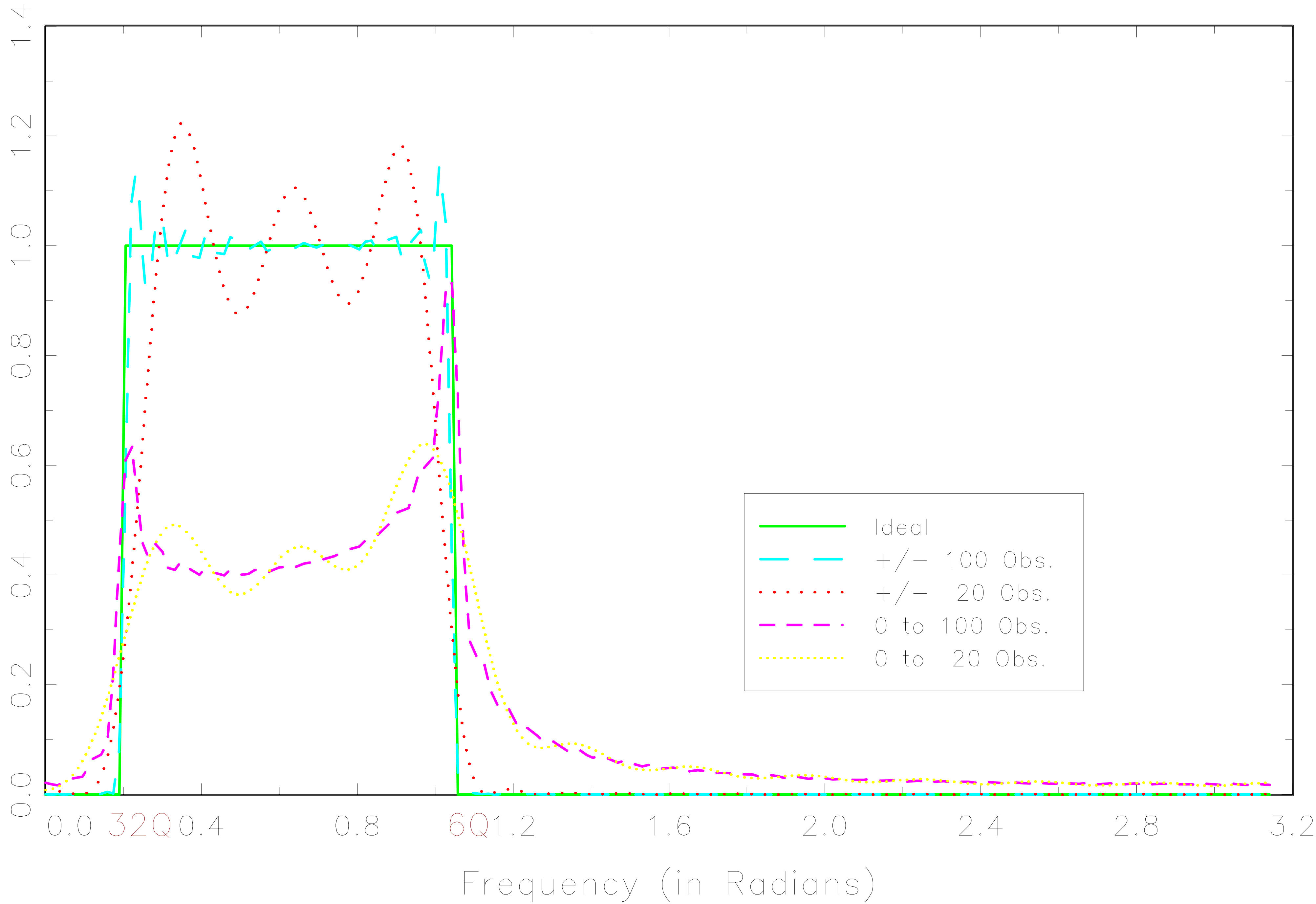
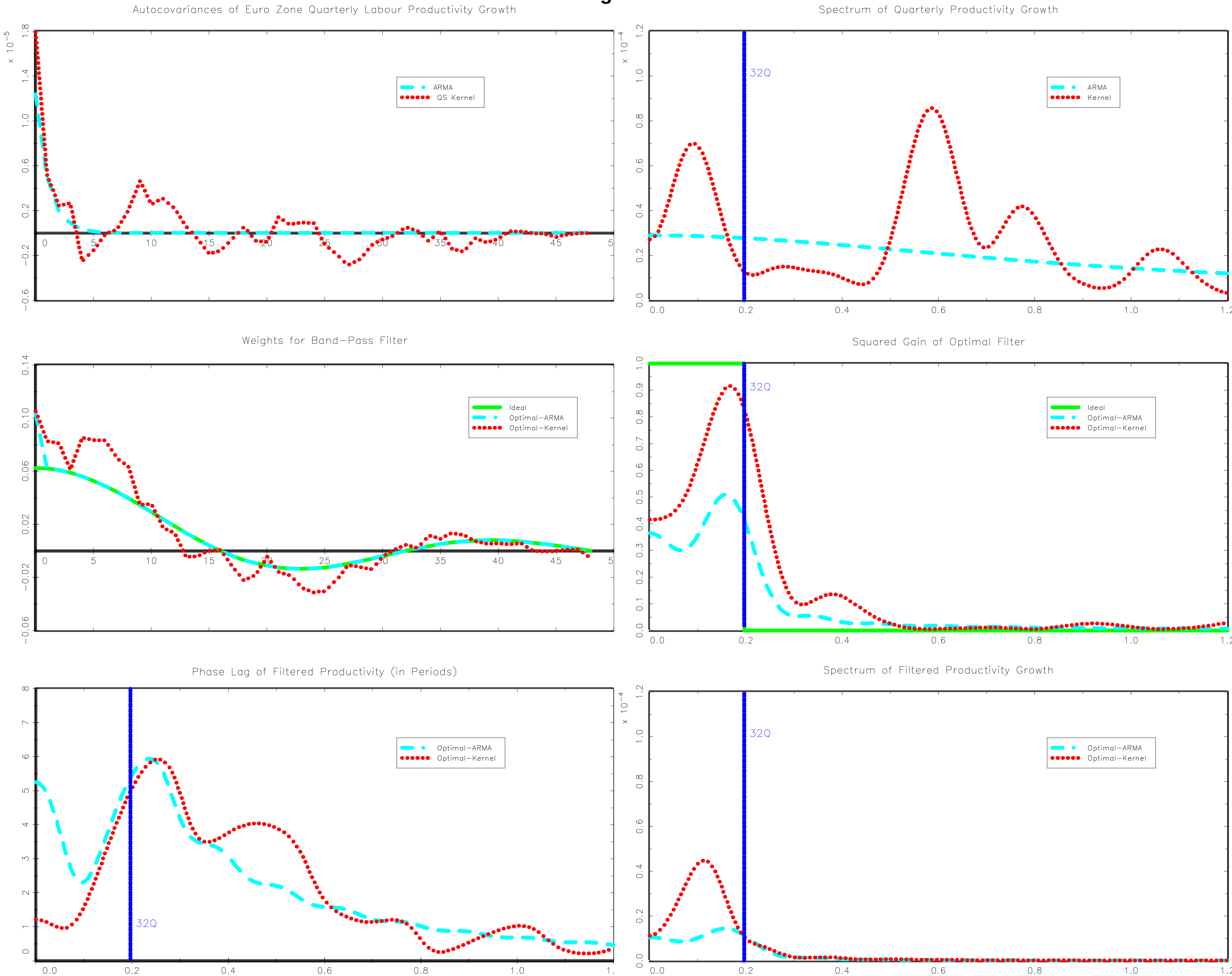
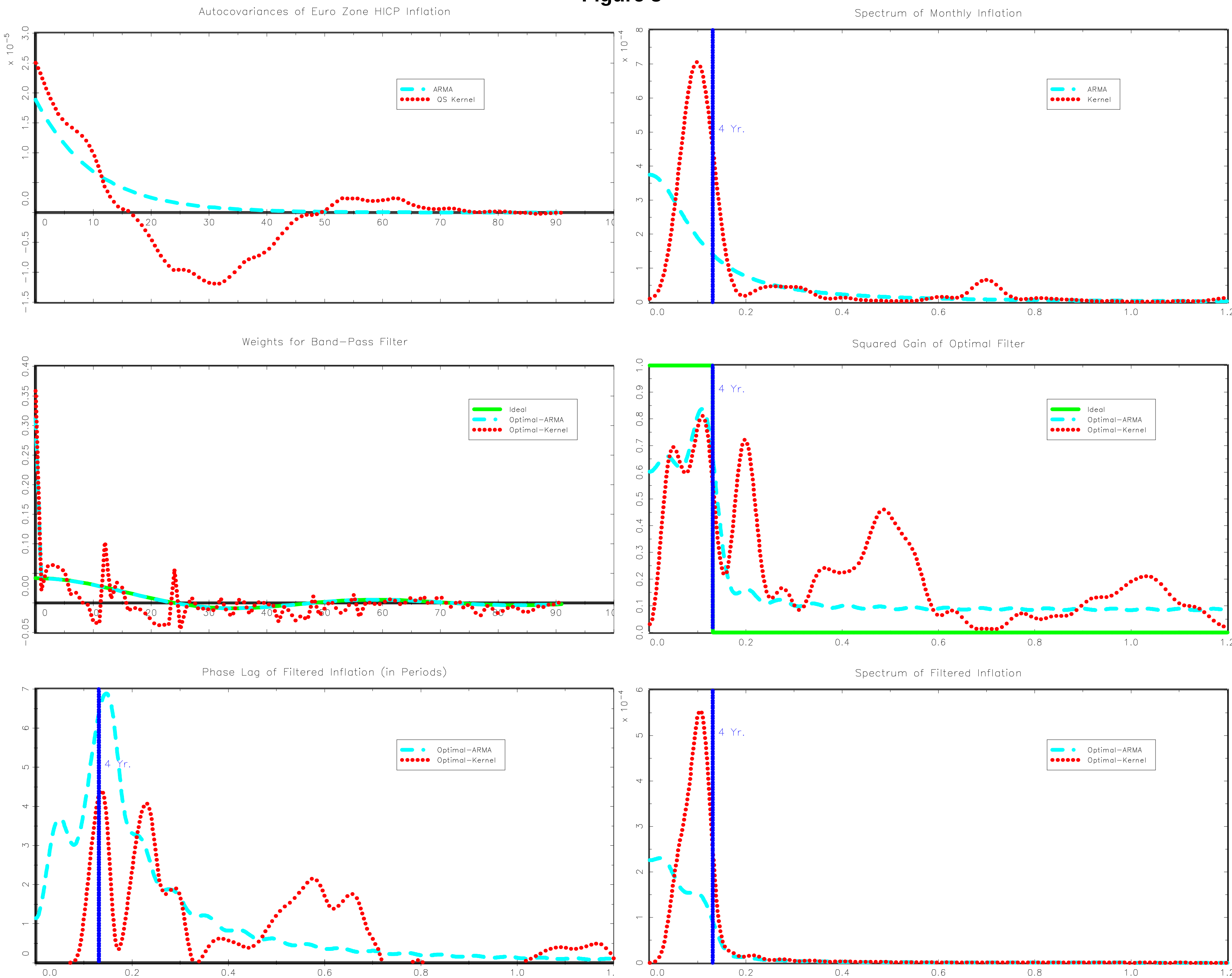


Figure 2



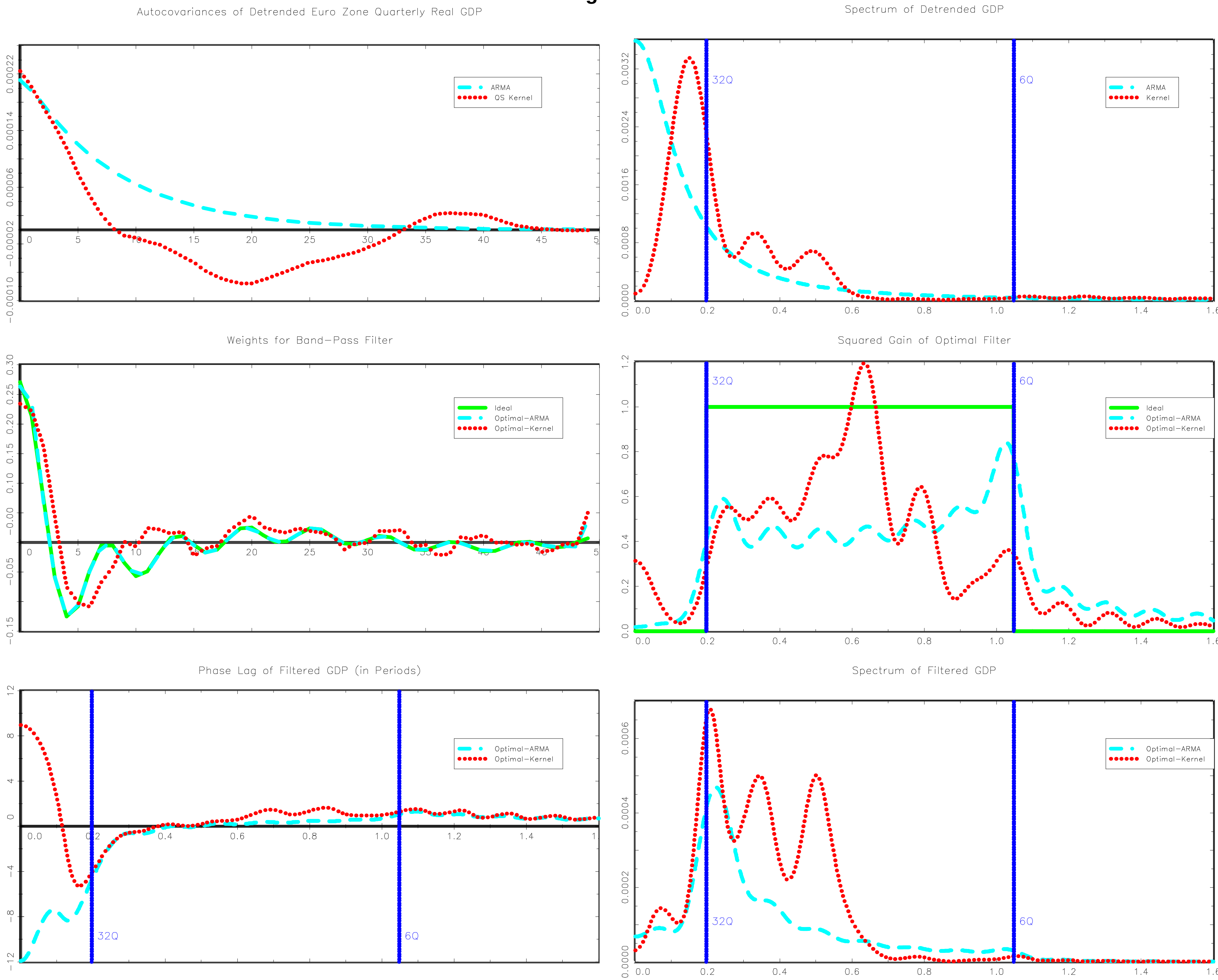
Autocovariances of Euro Zone Quarterly Labour Productivity Growth
ARMA
QS Kernel
Spectrum of Quarterly Productivity Growth
32Q
ARMA
Kernel
Weights for Band-Pass Filter
Ideal
Optimal-ARMA
Optimal-Kernel
Squared Gain of Optimal Filter
32Q
Ideal
Optimal-ARMA
Optimal-Kernel
Phase Lag of Filtered Productivity (in Periods)
32Q
Optimal-ARMA
Optimal-Kernel
Spectrum of Filtered Productivity Growth
32Q
Optimal-ARMA
Optimal-Kernel

Figure 3



Autocovariances of Euro Zone HICP Inflation
x 10^−5
ARMA
QS Kernel
Spectrum of Monthly Inflation
x 10^−4
ARMA
Kernel
4 Yr.
Weights for Band−Pass Filter
Ideal
Optimal−ARMA
Optimal−Kernel
Squared Gain of Optimal Filter
Ideal
Optimal−ARMA
Optimal−Kernel
4 Yr.
Phase Lag of Filtered Inflation (in Periods)
Optimal−ARMA
Optimal−Kernel
4 Yr.
Spectrum of Filtered Inflation
x 10^−4
Optimal−ARMA
Optimal−Kernel
4 Yr.

Figure 4



Autocovariances of Detrended Euro Zone Quarterly Real GDP
ARMA
QS Kernel
Spectrum of Detrended GDP
ARMA
Kernel
32Q
6Q
Weights for Band−Pass Filter
Ideal
Optimal−ARMA
Optimal−Kernel
Squared Gain of Optimal Filter
Ideal
Optimal−ARMA
Optimal−Kernel
32Q
6Q
Phase Lag of Filtered GDP (in Periods)
Optimal−ARMA
Optimal−Kernel
32Q
6Q
Spectrum of Filtered GDP
Optimal−ARMA
Optimal−Kernel
32Q
6Q