CENTRE FOR CENTRAL BANKING STUDIES

ECONOMIC MODELLING AND FORECASTING

Estimating the output gap and potential output for Sri Lanka: a practical guide to some trend and cycle decompositions

by

Ole Rummel
Centre for Central Banking Studies
Bank of England

12 February 2014

ole.rummel@bankofengland.co.uk

Contents[1]

[1] Sections denoted by an asterisk (*) can be skipped without loss of continuity.

1 Introduction

Many of the theoretical models we now commonly work with are designed to explain or replicate cyclical features of actual data. Indeed, understanding macroeconomic fluctuations entails the study of an economy's output relative to its trend or potential output, where the difference of the two is referred to as the output gap. Starting with Lucas (1977), macroeconomists have obtained cyclical information by eliminating the permanent component (the 'trend') of the data. The latter is typically thought to be unrelated with those features that the theoretical models are interested in explaining, including the cycle.

This throws up a quandary as most economic time series display trends or marked growth patterns that make it difficult to discern what the cyclical properties of the data are. Added to this is the fact that dynamic economic theory is silent on both the type of trend a series may display and the exact relationship between the trend and the cyclical component. As such, theoretical models have been proposed where the long-run component may be either stochastic or deterministic and may or may not be related to the cyclical component.

Another problem arises in practice since trends and cycles are usually unobservable, meaning that assumptions are needed to split observable series into these two components. Different assumptions give rise to different possible models, and it is frequently extremely difficult to formally choose among alternatives with a finite set of data. This also means that it will be impossible to use formal statistical criteria to assess the optimality of one or more of the many approaches to extracting cyclical information.

This practical guide is all about the removal of trends, the isolation of cycles and therefore the process of obtaining cyclical information from actual data. It should be obvious that the isolation of cycles is closely related to the removal of trends. As such, for a time series exhibiting cyclical deviations about a trend, the identification of the trend automatically serves to identify the cyclical deviations as well. But it may be the case that even after the separation of trend from cycle is accomplished, additional steps may be necessary to isolate cycles according to the frequency of their recurrence. For example, patterns of fluctuations in the data can recur at business-cycle frequencies, generally thought to lie between six and 24-32 quarters, as well as at seasonal frequencies.[2] In other words, unless additional steps are taken, the removal of a trend will leave seasonal fluctuations intact, and their presence can have a detrimental impact on inferences involving business-cycle behaviour.[3]

2 The output gap

Potential output – and hence the output gap – is arguably the most important unobservable variable that central banks use regularly. Not only does the growth rate of potential output shed light on the prospects for the domestic economy's economic growth, but the output gap is a key measure of inflationary pressures. In fact, Stock and Watson (1999b) argue that, if we could only choose one variable to best represent the state of the economy,

[2] Many of the business cycles reported by either the National Bureau of Economic Research (NBER) in the US or the Centre for Economic Policy Research (CEPR) in Europe have a periodicity which falls, at least approximately, in this range. In fact, the taxonomy employed by the NBER to officially date business cycles in the US describes fluctuations with periodicity between 2 and 6 years as such. In addition, conventional wisdom has it that no complete cycle (in the US) has exceeded 8 years in length. This makes variability at frequencies corresponding to cycles of 6 to 24-32 quarters of crucial economic importance.

[3] This is the reason why the literature prefers to work with seasonally-adjusted data.

that variable would be GDP. They go on to claim that '…fluctuations in aggregate output are at the core of the business cycle so that the cyclical component of real GDP is a useful proxy for the overall business cycle… (p. 15)'.

Quite generally, the output gap, y_t^g, is defined as the deviation of actual output, y_t, from its potential level, y_t^p:

$$y_t = y_t^p + y_t^g \tag{1}$$

The literature has proposed different definitions of potential output, including the maximum output that can be produced under full employment (Okun (1962)), the maximum amount of output that can be produced without causing inflationary pressures and the equilibrium level of output that would prevail in DSGE models when all prices are flexible (Congdon (2008)).

The different definitions of y_t^p demand different methods of estimating y_t^g., and any one method of estimating potential output depends on numerous assumptions that are subject to uncertainty and error. Still, the estimated output gap measure should capture a number of stylised facts in line with traditional descriptions of economic activity in the country. Moreover, for any economy, a good output gap measure should have good end-point properties (to be defined below) and be stable, that is, it should adequately characterise the current state of the economy. Furthermore, the assessment should not change substantially as new data becomes available.

In cases where one or more sectors dominate the economy, such as natural resources (mining, agriculture and fishing), additional constraints apply. In economies dependent on primary commodities, variations in commodity production are usually outside policymakers' control and are best thought of as structural changes that affect potential GDP. In this case, it is worth measuring the output gap for overall GDP and GDP with the dominant sectors excluded. As movements in sectoral GDP should appropriately be captured in the trend and not the cyclical component, a good output gap measure should yield similar results for both overall GDP and the sectoral measures of GDP.[4] In other words, a proper output gap measure should be similar for the different data series since the output gap should exclude structural movements. Indeed, this can be used as an additional criterion in evaluating output gap measures for the economy under investigation.

For reasons outlined further below, the empirical and policy-oriented literature in emerging market and low-income economies either resorts to the Hodrick-Prescott filter or a battery of output gap measures for robustness. At the same time, Billmeier (2004) concludes that an output gap rarely provides useful information on domestic inflationary pressures for inflation forecasting and that there is no single best output gap measure across (five European) countries. This finding of no model being optimal holds more generally and there is no general consensus on which method is best. Even the passage of time and increasing econometric sophistication is no help in this, as Chiu and Wieladek (2012) document that the accuracy of real-time estimates of the output gap produced by the OECD for its member countries has not improved over time.

[4] Magud and Medina (2011) analyse the differences in potential output in the natural-resource and non-natural-resource sectors in Chile and the non-linear contributions to potential growth of the different sectors.

The aim of this practical guide is to highlight potential output and output gap estimates for Sri Lanka using a variety of different techniques.[5]

We assume that an economic time series y_t, where lowercase letters denote the natural logarithm of the series in levels, can be conveniently represented by the following trend-cycle decomposition:

$$y_t = \tau_t + c_t \tag{2}$$

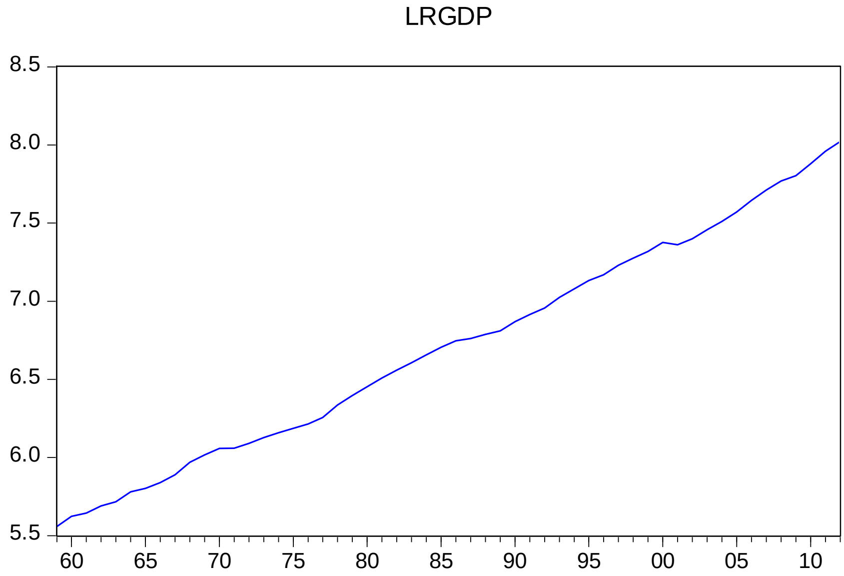
where τ_t represents the trend, which can be either deterministic or stochastic; c_t represents a possibly stochastic cyclical component, ε_t represents an irregular (error) component and τ_t and c_t are independent ('orthogonal'). Even when there is no major structural break in the economy, estimation of the output gap involves some margin of error due to measurement issues and difficulty in identifying temporary demand factors, meaning that we occasionally add an irregular (error) component, ε_t, to (1):

$$y_t = \tau_t + c_t + \varepsilon_t \tag{3}$$

3 The data

The underlying data throughout this practical guide are the natural logarithm of (seasonally-adjusted) annual real GDP for Sri Lanka over the period from 1959 to 2012, a graph of which is shown in Figure 1. For parts of the illustrations, the data are augmented by real GDP forecasts over the period from 2013 to 2018, which were taken from the IMF's October 2013 *World Economic Outlook* database.

Figure 1: Log of real (seasonally-adjusted) GDP in 2000 prices for Sri Lanka, 1959-2012



[5] The IMF (2010) suggests that if '…different approaches point to divergent assessments, a risk management approach could be given greater emphasis, with attention paid not just to the central forecast, but also to the error band around it (p. 4)'.

Working with annual data has advantages as well as disadvantages. To begin with, we should be aware of the fact that annual data may obscure a substantial amount of the cyclical variation in economic activity. On the other hand, annual observations minimise the impact of data revisions.

4 Extracting a cycle from macroeconomic time series

Canova (2007) differentiates between three different approaches to cycle extraction: statistical methods, economic methods and hybrid methods. The first approach includes procedures that have a statistical or a probabilistic justification. They generally use time-series assumptions on the observables or the trend to measure the cycle. The second approach relies on economic theory to extract the cycle. As such, the cycles we obtain using these methods have relevance only to the extent that the theoretical model used is a valid approximation to the true underlying data-generating process (DGP). The final approach uses procedures that are statistical in nature, but that have an economic justification of some sort.

The rest of the practical guide will be concerned with cycle extraction using selected methods from all three approaches.

4.1 An interlude on detrending versus filtering

Before we look at the different approaches to cycle extraction in more detail, it is worth highlighting a subtle, yet important, difference between detrending and filtering. The former involves eliminating the trend from the series, while the latter extracts the cyclical component of the series. In other words, detrending is an intermediate (or indirect) step in cycle extraction. As such, they are operationally different.

4.1.1 Detrending

The sustained upward trend in real GDP in Figure 1 might be captured by a simple linear trend. In fact, the trend was traditionally taken to be deterministic and the cyclical component was extracted as the residual of a regression of the series under observation, y_t, on a function of time. Since the trend is deterministic, the growth rate of y_t needs to be time-invariant, although the latter problem can be partially eliminated if we allow for structural breaks at pre-selected points. The latter gives rise to a segmented trend specification.

But the deterministic trend assumption is controversial, since it implies a deterministic long-run growth rate of the real economy. Adherents to the real business cycle school argue that technological advancements have permanent effects on the trend displayed by the macroeconomy. Since these technological innovations are thought to be stochastic, the trend inherent in real GDP should reflect this underlying randomness.

4.1.2 Filtering

The problem with mechanical detrending is that the trend may not be deterministic, as we have assumed above.

But in addition to detrending, time series can also be filtered to extract or eliminate special features or components.[6] In general, a filter is a function of a time series that transforms it into another series. In practice many filters are linear functions. As such, linear filters take a possibly two-sided moving average of an original set of observations, $y_t = (y_1, y_2, \ldots, y_T)$ to create a new time series, x_t:

$$x_t = \sum_{i=-k}^{l} \omega_i y_{t-i} \tag{4}$$

where the filter is defined by the two positive integers k and l and the weights $(\omega_{-k}, \ldots, \omega_0, \ldots, \omega_l)$. The weights ω_i are frequently chosen to add up to one to ensure that the level of the series is maintained. The many filters employed in the empirical literature differ in their choice of k and l and their particular specification of the ω_i's. The ideal filter has a two-sided infinite moving-average representation such that i goes from $-\infty$ (k) to ∞ (l) and where symmetry is imposed ($\omega_i = \omega_{-i}$) to avoid the filter inducing a phase shift. Such an ideal filter is clearly not feasible with a limited set of data and we will return to this point below.

4.2 Statistical decompositions

The statistical approach relies solely on output data and seeks to decompose output into potential output and a residual series, the output gap. This can be done either by simple detrending or by applying filtering techniques on the output data.

4.2.1 Removing trends ('detrending')

In the past, the representation, and thus extraction, of the cyclical component was handled in a very simple way. The trend was represented by a deterministic polynomial function of time, assumed to be independent of the cyclical component, and extracted using simple regression methods.[7]

As a first statistical method to measure the output gap, therefore, we estimate the output gap as the deviations of the output series from a simple linear trend. This is the simplest (and oldest) approach. It assumes that the trend and cycle of the series are uncorrelated and that τ_t is a deterministic process that can be approximated by (low-order) polynomial functions of time.

The appropriate way to transform a model in linear detrending is to estimate the following regression equation:

$$\begin{aligned} y_t &= \tau_t + c_t \\ &= \beta_0 + \beta_1 t + \varepsilon_t \end{aligned} \tag{5}$$

[6] A trend-cycle decomposition would be an example of extraction, while seasonal adjustment would be an example of elimination.

[7] If there is clear evidence of a structural break in the series, we need to extend the model to include a segmented linear trend.

Assuming that the series consists of trend and cycle only, subtracting the estimated values of y_t from the observed series yields estimated values of the ε_t series, which is the proxy for deviations from trend, i.e., the cycle. In other words, $\tau_t = \beta_0 + \beta_1 t$ and $c_t = \varepsilon_t$.

More generally, a time series may have a polynomial trend:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \ldots + \beta_p t^p + \varepsilon_t \qquad (6)$$

where ε_t is again a stationary process. In this case, detrending is achieved by regressing y_t on a polynomial time trend. The appropriate degree of the polynomial, p, is usually taken to be small and can be determined by standard t- and F-tests as well as the use of the information criteria like the Akaike information criterion (AIC), the Schwarz (Bayesian) information criterion (SBC) and the Hannan-Quinn information criterion (HQ). Common practice is to estimate the regression equation using the largest value of p deemed reasonable and then following a general-to-specific testing procedure (some authors specify specific-to-general instead, which I would strongly discourage based on the analysis in Lütkepohl (2007)).

In particular, if the t-statistic for β_p is zero, consider a polynomial trend of order $(p - 1)$. We continue to pare down the order of the polynomial trend until a non-zero coefficient is found. In addition, we can use F-tests to determine whether group coefficients – β_{p-i} through β_p – are statistically significant. The information criteria can be used to (re)confirm the appropriate degree of the polynomial. Remember that we want to **minimise** these criteria.

Estimating equation (6) with p = 1, 2, 3 results in the information criteria set out in Table 1.

Table 1: Information criteria of various polynomial time-trend models for real Sri Lankan GDP, 1959-2012

	$p = 1$	$p = 1, 2$	$p = 1, 2, 3$	$p = 1, 3$
Akaike information criterion	-3.51743	-4.16319	-4.21437	-4.23615
Schwarz information criterion	-3.44377	-4.0527	-4.06704	-4.12565
Hannan-Quinn information criterion	-3.48902	-4.12058	-4.15755	-4.19353

Amongst the three polynomial models, the model with p = 3 (t, t^2, t^3) minimises the information criteria. But we find that the estimated coefficient on `@trend^2` in the estimation output is not statistically significant (Table 2). Eliminating `@trend^2` from the regression equation results in a new model with information criteria given in the final column of Table 1. Of the four models considered, the trend polynomial involving first- and third-order polynomials only minimises all of the information criteria and will be the one that is chosen.

Table 2: Third-order trend polynomial model for real Sri Lankan GDP, 1959-2012

Dependent Variable: LRGDP
Method: Least Squares
Sample: 1959 2012
Included observations: 54

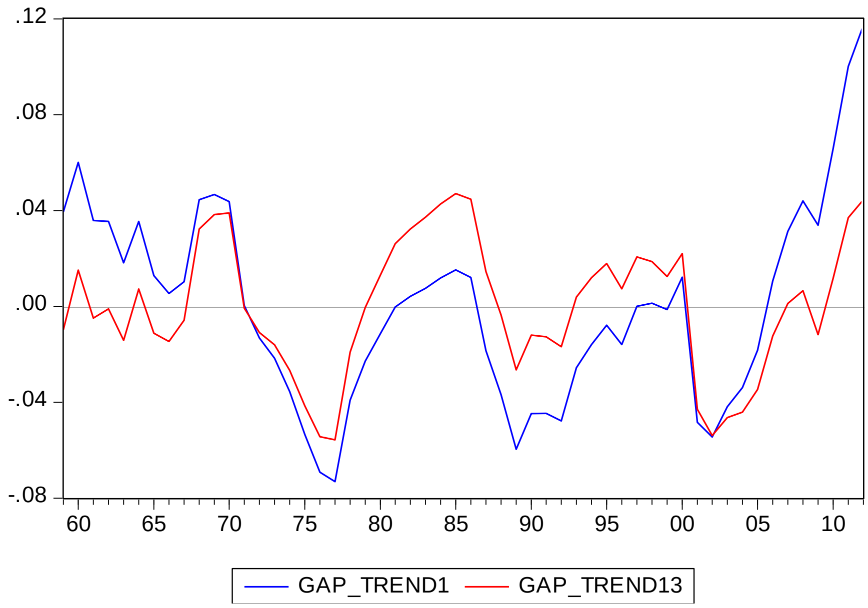
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.559006	0.014440	384.9830	0.0000
@TREND	0.042787	0.002382	17.96517	0.0000
@TREND^2	-9.21E-05	0.000105	-0.876875	0.3847

@TREND^3	2.79E-06	1.30E-06	2.147324	0.0366

R-squared	0.998489	Mean dependent var	6.711816
Adjusted R-squared	0.998398	S.D. dependent var	0.709364
S.E. of regression	0.028390	Akaike info criterion	-4.214369
Sum squared resid	0.040300	Schwarz criterion	-4.067037
Log likelihood	117.7880	Hannan-Quinn criter.	-4.157549
F-statistic	11012.95	Durbin-Watson stat	0.434341
Prob(F-statistic)	0.000000		

The different output gaps arising from the two models, i.e., the linear trend model and the first- and third-order polynomial trend model, are shown in Figure 2.

Figure 2: Output gaps using different polynomial trend functions



Linear detrending shows much larger positive output gaps at the beginning and the end of the sample than the other polynomial trend model. In particular, note the very large positive output gap in 2014, which is especially apparent with linear detrending (`gap_trend1`).[8]

An equally well-established procedure is **first-order differencing**, which relies on the assumptions that the long-run component of a series is described by a random walk without drift, that the cyclical component is stationary and that the two components are uncorrelated. In addition, we assume that y_t has a unit root which is entirely due to the long-run component of the series. In other words, the non-stationarity (or growth) in economic activity should be removed by first-differencing rather than linear detrending, making the trend a random walk without drift rather than a straight line. As a result, y_t can be represented as:

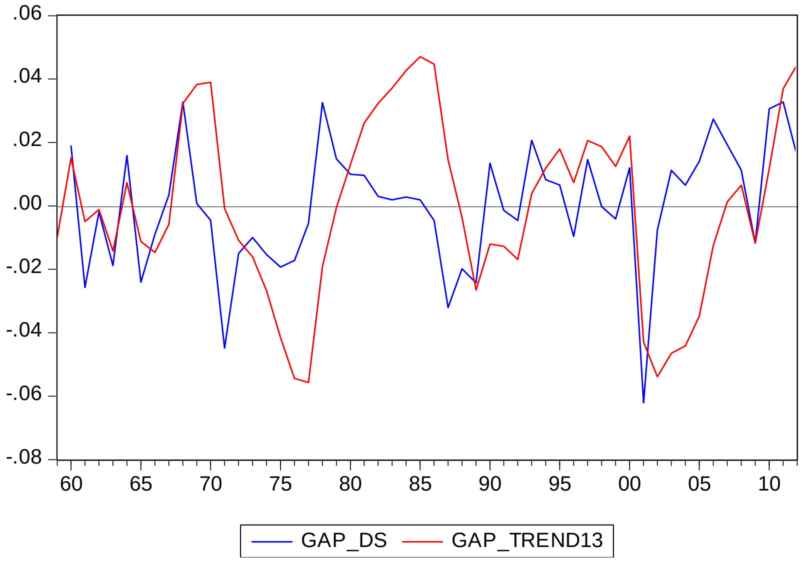
$$y_t = \tau_t + c_t$$

[8] It is difficult to ascertain this with certainty without further analysis, but this may be an illustration of Watson's (2007) remark that low-order polynomials in time yield '…unrealistic estimates of estimation errors at the end of the sample (p. 144)'.

$$= y_{t\text{-}1} + \varepsilon_t \tag{7}$$

and the trend is defined as $\tau_t = y_{t\text{-}1}$ and an estimate of c_t is obtained as $c_t = \varepsilon_t = y_t - \tau_t = y_t - y_{t\text{-}1} = \Delta y_t$. These manipulations require no estimation, so I have simply generated the first difference of `lrgdp`, denoted `dlrgdp`, using either `dlrdgp = dlog(rgdp)` or `dlrgdp = d(lrgdp)` or `dlrgdp = lrgdp - lrgdp(-1)`. In addition, I have also created a series called `lrgdp1`, which is the first lag of `lrgdp`. The output gaps for linear detrending as well as first-differencing are shown in Figure 3.

Figure 3: Output gaps – detrending versus first-differencing



In Figure 3, the cyclical component has been scaled to have a zero mean. This is achieved by including a drift (constant) term in equation (7) for estimation. In the random walk without drift specification, which is usually followed in the literature, a zero mean for the cyclical component will not necessarily be the case. While the output gap estimates using first-differencing (the blue line) have become ‘noisier’, the general movement in the output gap mirrors that of the estimates derived from (non-linear) detrending (the red line). At the same time, the blue line does not necessarily visually conform to our prevailing notion of (smooth) cyclical fluctuations.

As we have seen, the goal of approaches to remove trends from macroeconomic time series is to transform the data into mean-zero covariance-stationary stochastic processes. The definition of covariance stationarity is that a time-series process has a constant mean and variance. Under the assumption of time-invariant second moments, sample averages may be used to estimate population averages of these moments as well as functions thereof. But we should keep in mind that trend removal is not sufficient to induce covariance stationary in integrated economic time series.

I note in concluding that there are many examples in the empirical literature of ARIMA-based detrending filters, which approximately eliminate the random-walk component of the data and whose residuals are taken as an estimate of the cyclical components. For example, Chapter 1 of the influential (but now somewhat dated) textbook by Blanchard and Fischer (1989) clearly interprets the residuals of univariate ARIMA processes as the cyclical component. Another prominent application of such an approach is due to Beveridge and Nelson (1981).

4.2.2 The Beveridge-Nelson (BN) decomposition

Nelson and Plosser's (1982) seminal findings suggest that many economic time series contain a unit root and an irregular component. Having observed a series, but not the individual underlying components, is there any way to decompose the series into its constituent parts?

The answer is yes, and the approach that is widely used in this context is due to Beveridge and Nelson (1981). In fact, Beveridge and Nelson (1981) show how to decompose any ARIMA(p,1,q) model into the sum of a (pure) random walk plus drift and stationary component (i.e., the general trend plus irregular model, $y_t = \tau_t + \varepsilon_t$), and the so-called Beveridge-Nelson (BN) decomposition calculates a non-stationary (trend) and stationary (cycle) component for an integrated time series.[9]

The underlying logic of the BN filter is that the permanent value of a series y_t, P_t, is given by:

$$P_t = \mathrm{E}_t(y_\infty) \tag{8}$$

$$= E_t\left[y_t + \sum_{j=1}^{\infty} \Delta y_{t+j}\right] \tag{9}$$

$$= y_t + E_t\left[\sum_{j=1}^{\infty} \Delta y_{t+j}\right] \tag{10}$$

such that the transitory component, which is equal to the output gap if y_t is seasonally-adjusted real GDP, is given by $y_t - P_t = -E_t\left[\sum_{j=1}^{\infty} \Delta y_{t+j}\right]$. In order to compute the transitory component of y_t, we need to specify a model for Δy_t. When Δy_t is an AR(p), say, $E_t\left[\sum_{j=1}^{\infty} \Delta y_{t+j}\right]$ will be a linear function of Δy_t, Δy_{t-1}, … , Δy_{t-p+1}. In other words, the BN measure of the output gap is given by the negative of an average growth rate. This means that one will see a negative relation between the output gap and growth, such that a regression of inflation against output growth should result in a negative coefficient on the latter.

Following on from above, the BN decomposition is implemented by postulating an unrestricted ARIMA(p,1,q) model for y_t. This is based on the empirical observation that most time series can be approximated by an ARIMA(p,1,q) model. The BN decomposition thus focuses on the derivation of the trend component of the ARIMA(p,1,q) model. Once we have identified the appropriate ARMA(p,q) model for Δy_t, or, equivalently, the appropriate

[9] Morley (2011) points out that there are two ways to interpret the results from this decomposition: in one (emphasised by Watson (1986) and Morley *et al.* (2003)), the BN trend corresponds to an **estimate** of the unobserved permanent component; in the other (emphasised in the original paper by Beveridge and Nelson (1981)), the BN trend provides a **definition** of the observable permanent component. Morley concludes that the BN decomposition provides estimates of trend and cycle, making it a highly general and practical method for trend/cycle decomposition.

ARIMA(p,1,q) model for y_t, straightforward calculations result in the permanent component of the decomposition, also referred to as the BN trend. Given that the trend is postulated to be a pure random walk, the BN decomposition holds less relevance to applied economists who do not believe that the trend is a pure random walk.[10] The cyclical component is then calculated as the actual data series in levels minus the series for the BN trend (Zivot (2005)).

For example, using real US GDP data over the period from 1947 Q2 to 1985 Q4, Stock and Watson (1988) fitted the following ARMA(0,1) model to the growth rate of the natural log of real GDP:

$$\Delta y_t = 0.008 + \varepsilon_t + 0.3\varepsilon_{t-1} \qquad \varepsilon_t \sim \text{iid}(0, \sigma 2) \qquad \hat{\sigma} = 0.0106 \tag{11}$$

while Morley *et al.* (2003) fitted an ARMA(2,2) model to the same data over the period from 1947 Q1 to 1998 Q2:

$$\Delta y_t = 0.816 + 1.342\Delta y_{t-1} - 0.706\Delta y_{t-2} + \varepsilon_t - 1.054\varepsilon_{t-1} + 0.519\varepsilon_{t-2} \tag{12}$$

In many cases, we can also derive algebraic results for the shape of the BN trend. As shown in Morley (2002), if Δy_t follows the simple AR(1) process:

$$\Delta y_t = \mu + \phi(\Delta y_t - \mu) + \varepsilon_t \tag{13}$$

where $|\phi| < 1$ and $\varepsilon_t \sim$ iid $N(0, \sigma^2)$, then the BN trend and cycle are given by:

$$\tau_t = y_t + \frac{\varphi}{1-\varphi}(\Delta y_t - \mu) \tag{14}$$

and:

$$c_t = -\frac{\varphi}{1-\varphi}(\Delta y_t - \mu) \tag{15}$$

One advantage of the BN decomposition over traditional approaches is that it produces a decomposition without any assumptions on either the structure of the components or on their correlation. At the same time, we should note some interesting features of the BN decomposition. To begin with, since the two components are driven by the same shock, trend and cycle are perfectly correlated. Second, since estimates and forecasts are typically obtained from ARIMA(p,1,q) models, we need to be aware of the standard identification problems associated with ARIMA specifications. Third, since long-run forecasts of Δy_t are based on past values of y_t only, trend estimates may be very imprecise.[11] Finally, given the definition of innovations in the trend, the variability of the innovations in the trend may be

[10] Note that the canonical decomposition of Hillmer and Tiao (1982) as well as the general permanent-transitory decomposition of Quah (1992) make the assumption that the permanent component is an integrated series but not a pure random walk.

[11] Another possible implication is that the ratio of the variability of the cycle to the variability in the trend becomes arbitrarily small.

larger than the variability of the innovations in the original series. We will return to some of these points below.

At the same time, the BN filter is also not without its drawbacks. One criticism is that the output gap estimate resulting from the BN filter is not as smooth as the HP-filtered estimate (Park (1996)). In addition, while the Beveridge and Nelson (1981) decomposition provides a straightforward method to decompose any ARIMA(p,1,q) process into a temporary and permanent component, it is important to note that the BN decomposition is **not** unique. This is because the decomposition forces the innovation in the trend and stationary components to be perfectly correlated. This is important as economic theory does not always prescribe such a relationship between the two innovations. In fact, without *a priori* knowledge of the relationship between innovations in the trend and the stationary components, the decomposition of a series into a permanent and transitory component is not unique. What if we incorrectly used a BN decomposition to obtain the temporary and permanent components? Well, to begin with, there is no way for us to determine that the assumption of perfectly correlated innovations is incorrect. In other words, there is no way to identify the 'true' model – i.e., a correlation between the two innovations lying somewhere between -1 and 1 – using sample data. We will come across an equivalent representation of the BN decomposition using the state-space approach in section 4.2.3 that does away with this shortcoming.[12]

Some of these criticisms notwithstanding, we can use EView's **ARIMASel** add-in for performing an ARIMA selection routine, where the order of differencing is chosen *via* unit root tests and the AR (seasonal AR) and MA (seasonal MA) terms are chosen according to an information criterion. This EViews add-in is available from:

http://www.eviews.com/cgi/ai_download.cgi?ID=ARIMASel.aipz

The add-in is run from a `series` object, which means that we will have to double-click on `lrgdp` to open it. This series is then used as the dependent variable for the ARIMA selection process. The add-in dialog is opened by going to **Proc/Add-ins/Automatic ARIMA selection**. The selection routine requires us to specify a list of independent variables (other than the ARMA terms), a maximum order for both the AR(p) and the MA(q) terms, a maximum order of differencing for the unit-root test and which information criterion you would like to use for the selection process (Akaike, Schwarz or Hannan-Quinn). We set the maximum order for both the AR and the MA terms to 3 and the maximum difference to 1, as we think that the log of real GDP is at most I(1). The results of the ARIMA selection routine using the Schwarz information criterion are shown in Table 3.[13]

[12] Jumping ahead, Watson (1986) estimates the trend and irregular terms as unobserved components, thereby avoiding the estimation of a constrained ARIMA model. The constraints come from the fact that the assumption that the two innovations are uncorrelated places restrictions on the autoregressive and moving average coefficients of Δy_t.

[13] The mechanics of the automatic ARIMA selection routine are as follows. The procedure first detects the level of differencing to apply to the original data series by performing a KPSS unit root test on each subsequent level of differencing until the unit root test is insignificant (at the 5 per cent level). Following the selection of the order of differencing, the ARMA($p.q$) terms are selected by comparing the information criteria.

Table 3: Information criteria from the automatic ARIMA selection process

AR / MA	0.000000	1.000000	2.000000	3.000000
0.000000	-5.018011	-5.005336	-4.947904	-4.871889
1.000000	-5.016087	-4.940770	-5.001356	-4.852014
2.000000	-4.945772	-5.024003	-4.950616	-4.906158
3.000000	-4.858890	-4.926010	-4.897808	-4.784285

The selected model – which in this case is an ARMA(2,1) for the first-difference of `lrgdp` – appears in red in Table 3. Note that the preferred model once again minimises the information criterion. The Akaike information criterion also selects the ARMA(2,1) model, while the Hannan-Quinn information criterion selects an ARMA(1,2) model. This is a common occurrence in applied work, as these information criteria frequently indicate different lag orders, meaning that the 'correct' lag order can depend on the criteria or measure we use. This is typical of these tests and researchers often use the criterion most convenient for their needs.

After selecting the ARIMA(2,1) model for Sri Lankan real GDP growth data, we can use EView's **BNDecom** add-in to estimate the Beveridge-Nelson decomposition of `lrgdp`. The EViews add-in is available from:

http://www.eviews.com/cgi/ai_download.cgi?ID=bndecom.aipz

We again double-click on the `lrgdp` series in the workfile window to open it and follow the same step as before. Go to **Proc**/**Add-ins**, but this time select **Beveridge-Nelson Decomposition**. This opens a dialog box. For the problem at hand, the options should be set as shown in the following screenshot: the ARMA specification is for an ARMA(2,1) model so we set p = 2 for the AR specification and q = 1 for the MA specification. The BN decomposition automatically assumes that the series under observation is I(1). We follow Beveridge and Nelson (1981) and set s = 100. I have called the resulting trend output `bn_trend` and the cycle output `bn_cycle`. After the decomposition is completed, these two new series will appear in the workfile.

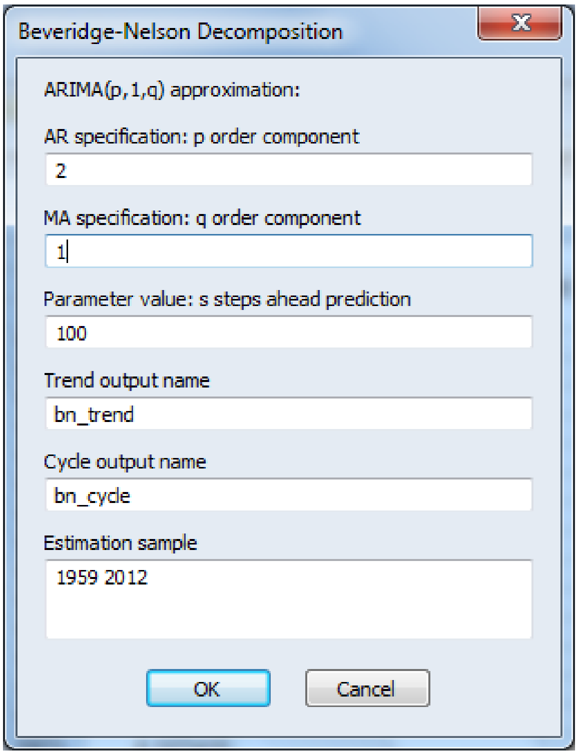
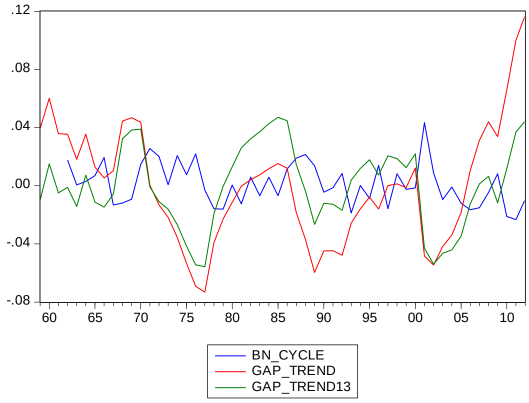


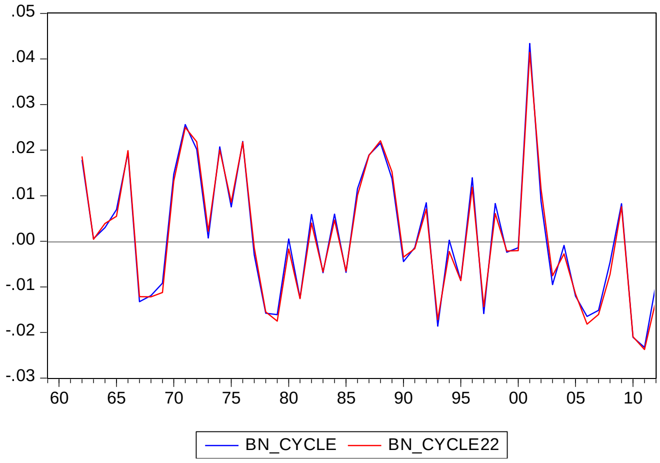
Figure 4 shows the output from the BN decomposition (`bn_cycle`) together with the two output gaps obtained from linear detrending. Note that we lose three observation points at the beginning of the sample using the BN decomposition: one on account of first differencing and another two because of the two autoregressive terms. In general, `bn_cycle` is much smaller than the detrended series. In addition, it is much more volatile than the others. Again, the blue line does not necessarily visually conform to our prevailing notion of (smooth) cyclical fluctuations.

Figure 4: Output gap estimates from detrending and the Beveridge-Nelson decomposition using an ARIMA(2,1,1) model



This result is well-known in the literature as determining the most appropriate ARIMA(p,1,q) model for real GDP growth to compute the BN decomposition is a difficult task (Campbell and Mankiw (1987)). Typically, model selection procedures favour low-order ARMA models for growth rates of real economic activity and the resulting BN cycles tend to be noisy and lack business cycle features. Oh *et al.* (2008) therefore recommend the ARIMA(2,1,2) model to compute the benchmark BN decomposition.[14] As can be seen from Figure X, though, this does not make much a difference in the case of Sri Lankan real GDP.

Figure 5: Output gap estimates from the Beveridge-Nelson decomposition using an ARIMA(2,1,1) and an ARIMA(2,1,2) model



4.2.3 Structural time-series models: the unobserved components (UC) decompositions

More recently, state-space models and Kalman filter techniques have become more widely employed in trend-cycle decompositions. These approaches treat the output gap as a latent variable and use the unobserved components (UC) model due to Harvey (1989).[15]

A basic model for representing a time series, y_t, is the additive model, also known as the classical decomposition:

$$y_t = \tau_t + \gamma_t + c_t + v_t + \varepsilon_t \tag{16}$$

which – in its most general form – models the dependent time-series variable, y_t, as consisting of a slowly-changing unobserved component, τ_t (trend), a periodic unobserved component, γ_t (seasonal), a periodically-recurring unobserved component, c_t (cycle), an unobserved autoregressive component, v_t, and an unobserved irregular component, ε_t (disturbance). It is, of course, the case that not all components need to be included at the same time. In a structural time-series (STS) or unobserved components (UC) model, the right-hand side components are modelled explicitly as stochastic processes that have a direct (semi-economic, at least) interpretation. As such, components can be deterministic functions of time (e.g., polynomials) or stochastic processes. Note the absence of any 'economic' variables on the right-hand side of equation (16). Each component of the UC model given by equation (16) is modelled in state-space form and estimated using the Kalman filter.

For example, the **trend** component, τ_t, can be modelled either deterministically as $y_t = \mu + \varepsilon_t$, with $\varepsilon_t \sim$ iid $N(0, \sigma_\varepsilon^2)$, or stochastically by a random walk plus noise, giving rise to the so-called **local level** or **random walk with noise** model:[16]

---

[14] Note that this model has an equivalent UC representation with an ARMA(2,1) model for Δyt.
[15] By latent we understand a variable that can be estimated but not observed.
[16] The model would be incomplete without a specification for the initial value: $\tau_0 \sim N(\alpha, P)$.

$$y_t = \tau_t + \varepsilon_t \qquad \varepsilon_t \sim \text{iid N}(0, \sigma_\varepsilon^2) \tag{17)}$$

$$\tau_t = \tau_{t-1} + \eta_t \qquad \eta_t \sim \text{iid N}(0, \sigma_\eta^2) \tag{18}$$

where $E(\varepsilon_t\eta_t) = 0$.[17]

In the **local level** model, only the level, τ_t, and the irregular component, ε_t, are unobserved. If the variable represents the observed log of real GDP, y_t is the sum of two unobserved components, τ_t and ε_t. The component τ_t is the state variable representing the signal of the log of real GDP in equation (17), while the transition equation (18) shows that the values of τ_t, and hence y_t, evolve according to a random walk.

The parameters to be estimated are σ_ε^2 and σ_η^2. Depending on these two parameters, the local level model nests some special trivial cases. Setting $\sigma_\eta^2 = 0$ results in the white-noise model with a constant level, $y_t \sim$ iid $N(\tau_1, \sigma_\varepsilon^2)$, while setting $\sigma_\varepsilon^2 = 0$ leads to a pure random-walk model, so that $y_t = y_{t-1} + \eta_t$. The cycle is then equal to $c_t = \varepsilon_t = y_t - \tau_t$.

The **local linear trend** model extends the local level model with a slope:

$$y_t = \tau_t + \varepsilon_t \qquad \varepsilon_t \sim \text{iid N}(0, \sigma_\varepsilon^2) \tag{19}$$

$$\tau_t = \tau_{t-1} + \beta_{t-1} + \eta_t \qquad \eta_t \sim \text{iid N}(0, \sigma_\eta^2) \tag{20}$$

$$\beta_t = \beta_{t-1} + \zeta_t \qquad \zeta_t \sim \text{iid N}(0, \sigma_\zeta^2) \tag{21}$$

where we again have to specify the initial distributions for β_1 and μ_1 and the irregular, level and slope disturbances, ε_t, η_t and ζ_t respectively, are mutually independent. In this model, β_t is the local trend rate and τ_t is the local trend itself, which is the local mean value for the observable y_t. In other words, τ_t and β_t are the level and slope of the trend respectively. The stochastic slope parameter, β_t, allows the trend to change smoothly over time. In the local linear trend model, η_t and ζ_t, account for the permanent shocks to the level of output, providing an explanation for movements in that variable. The cycle is again given by $c_t = y_t - \tau_t$. What is interesting about this model is that the change in the trend is a random walk plus noise; that is, $\Delta\tau_t$ is equal to the random walk term, β_t, plus the noise term, ζ_t.

The local linear trend model, which allows the trend level and slope to vary over time, is the standard state-space representation for handling strongly trending series and has been widely used for the purpose of trend-cycle decomposition. In essence, the trend is modelled as a random walk with drift. This model may be useful in cases where a single linear trend does not fit the data well. Smoother trends are obtained by formulating higher-order random walks, e.g., $\Delta^d\tau_t = \eta_t$.

Another advantage of this model is that it encompasses a range of other useful models. Just as in the case of the local level model, this will depend upon the values of the shock variances and initial values for β_1:

- if $\sigma_\zeta^2 = 0$, the trend is a random walk with constant drift β_1 – the model reduces to a local linear model if $\beta_1 = 0$;

[17] The assumption of independence between the innovations in the trend and cycle components is vital, but clearly not tenable in the real world. But the model given by equations (17) and (18) may not be identified without some specification of the correlation between innovations in trend and cycle.

- if **additionally** $\sigma_\eta^2 = 0$, the trend is a straight line with slope β_1 and intercept τ_1; and
- if $\sigma_\zeta^2 > 0$ but $\sigma_\eta^2 = 0$, the trend is a smooth curve, also known as an integrated random walk

The **cycle** component, c_t, following Harvey and Jäger (1993), is trigonometric in form and consists of one or more cycles defined by the pair of equations:

$$\begin{bmatrix} c_t \\ c_t^* \end{bmatrix} = \rho_c \begin{bmatrix} \cos(\lambda_c) & \sin(\lambda_c) \\ -\sin(\lambda_c) & \cos(\lambda_c) \end{bmatrix} \begin{bmatrix} c_{t-1} \\ c_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \qquad t = 1, 2, \ldots \tag{22}$$

where $0 < \rho_c < 1$ and $0 \le \lambda_c \le \pi$, where λ_c is the cycle periodicity in radians. In addition, κ_t, $\kappa_t^* \sim N(0,\sigma_\kappa^2)$, i.e., the two processes share the same variance.

The **seasonal** component, γ_t, if present, is modelled trigonometrically in a similar way to the cycle. The number of seasonal frequencies in a period, such as a year, is given by integer s. When s is even, $[s/2] = s/2$, and when s is odd, $[s/2] = (s - 1)/2$. The trigonometric seasonal form is given by:

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{jt} \tag{23}$$

where each γ_{jt} is generated by:

$$\begin{bmatrix} \gamma_{jt} \\ \gamma_{jt}^* \end{bmatrix} = \begin{bmatrix} \cos(\lambda_j) & \sin(\lambda_j) \\ -\sin(\lambda_j) & \cos(\lambda_j) \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{jt} \\ \omega_t^* \end{bmatrix} \qquad j = 1, \ldots, [s/2] \quad t = 1, \ldots, T \tag{24}$$

where $\lambda_j = 2\pi j/s$ is the frequency (in radians) and the seasonal disturbances ω_t and ω_t^* are two mutually uncorrelated, normally and independently distributed disturbances with zero mean and common variance σ_ω^2. Note that, according to equation (24), more than one seasonal term will enter into the unobserved components model. In fact, there will be two seasonal terms with quarterly data (j = 1, ... , [s/2] = 1, ... , [4/2] = 1, 2) and six seasonal terms with monthly data. For s even, the final component at $j = s/2$, equivalent to $\lambda_{(s/2)} = 2\pi(s/2)/s = \pi$, collapses to:

$$\gamma_{jt} = \gamma_{j,t-1}\cos(\lambda_j) + \omega_{jt}$$

The **autoregressive** component, v_t, has the following straightforward representation:

$$v_t = \rho_v v_{t-1} + \xi_t \qquad \xi_t \sim \text{iid } N(0, \sigma_\xi^2) \tag{25}$$

where $0 < \rho_v < 1$, while the **irregular** component, ε_t, with variance-covariance matrix, Σ_ε, is not explained in the model.

As already mentioned, the basic idea behind the unobserved components model is to provide structural equations for the components of the trend-cycle decomposition given by equation (16) that have a direct interpretation. As shown by Harvey (1989), cycles estimates obtained by UC decompositions enjoy certain optimality properties. As a result, this approach has become quite popular in the time-series literature. In fact, UC specifications to extract cyclical components are generally preferred to ARIMA representations for two reasons:

- there is no guarantee that an ARIMA(p,d,q) model identified by standard methods will have those features that the underlying series is thought the exhibit; and
- many of the ARIMA models favoured by applied researchers for the purpose of cycle extraction, such as the ARIMA(0,1,1) model, fail to forecast certain long-run components

The UC decomposition is characterised by two basic features. First, a researcher specifies a flexible structure for the trend, cycle and other features of the data, such as seasonals or irregulars.[18] Second, given the hypothesised structure, the data is allowed to select the characteristics of the components. Subsequent diagnostic testing and inference can be used to examine what is left unexplained.

Unobserved component models have a long history in trend-cycle decompositions, going back to Harvey (1985), Watson (1986) and Clark (1987), who used both the local level as well as the local linear trend model for estimating potential output and the output gap.

Watson's (1986) variant of the **local level** model is made up of two unobserved components, which are a stochastic trend, τ_t, and a stochastic cycle, c_t, which has been specified to follow an AR(2) process:

$$y_t = \tau_t + c_t \tag{26}$$

$$\tau_t = \beta + \tau_{t-1} + \eta_t \qquad \eta_t \sim \text{iid N}(0,\sigma_\eta^2) \tag{27}$$

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + \zeta_t \qquad \zeta_t \sim \text{iid N}(0, \sigma_\zeta^2) \tag{28}$$

Following the definition above, the trend will be an integrated random walk trend which, when estimated, tends to be relatively smooth. Having been specified as an stationary AR(2) process, the cycle is now clearly defined and will **not** be given by $c_t = y_t - \tau_t$.

In contrast, Harvey (1985) and Clark (1987) considered the **local linear trend** model with the stochastic cycle following an ARIMA(2,0,0) model:

$$y_t = \tau_t + c_t \tag{29}$$

$$\tau_t = \tau_{t-1} + \beta_{t-1} + \eta_t \qquad \eta_t \sim \text{iid N}(0, \sigma_\eta^2) \tag{30}$$

$$\beta_t = \beta_{t-1} + \omega_t \qquad \omega_t \sim \text{iid N}(0, \sigma_\omega^2) \tag{31}$$

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + \zeta_t \qquad \zeta_t \sim \text{iid N}(0, \sigma_\zeta^2) \tag{32}$$

[18] These structures or unobserved components in turn imply an ARIMA representation for y_t which is more complicated than the one typically selected with standard (Box-Jenkins) methods. The reduced form of the local local linear trend model, for example, is an ARIMA(0,2,2) model.

where η_t, ω_t and ζ_t are independent white-noise processes. The original specification in Clark (1987) was broader in the sense that it allowed the cycle to have a finite polynomial in the lag operator, L, such that equation (32) becomes $\Phi(L)c_t = \zeta_t$. Equation (32) specifies the stationary cyclical component as a finite autoregression rather than a more general ARMA process with moving-average terms. This is done for two reasons. First, the elimination of moving average terms facilitates estimation. Second, in many empirical applications, a simple AR(2) model fits c_t fairly well, obviating the need to include additional moving-average terms.

Equation (29) represents a decomposition of economic activity, y_t, into two additive components, the trend, τ_t and the cycle, c_t. Note that an irregular or white-noise term, ε_t, where $\varepsilon_t \sim$ iid $N(0, \sigma_\varepsilon^2)$ could be added to equation (29). In so doing, we can test for the importance of the irregular term (measurement error) by testing the statistical significance of its variance. If the estimated variance, $\hat{\sigma}_\varepsilon^2$, is not statistically different from zero, the irregular term can be dropped.[19]

The non-stationary trend component in equation (29) is modelled as a local approximation to a linear trend (the local linear trend model). Innovations in the level of τ_t are given by η_t, while innovations in its first difference or growth rate are given by ω_t. Allowing a variable growth rate ($\sigma_\omega > 0$) for the trend component is another testable hypothesis.

Estimation of the parameters σ_η^2, σ_ζ^2, σ_κ^2, ρ, λ_c and σ_ε^2 is done by maximum likelihood using the Kalman filter, using the prediction error decomposition of the latter. Once this has been accomplished, estimates of the trend, cyclical and irregular components are obtained from a smoothing algorithm which revises recursive estimates. As such, the smoothed estimates of the cycle can serve as a measure of the output gap.

We note in passing that the UC model is able to deal with structural breaks in the underlying series. Assuming the trend component in the economic time series to be modelled was segmented, i.e., it contained a break at one or more points, the smooth trend of the local linear trend model would adapt to it.

Univariate Kalman filter estimations can be improved upon by adding macroeconomic information. An early application was provided by Clark (1989), where the cyclical movement in output is measured using a bivariate UC model, where output and unemployment (or alternatively inflation) each have their own trend component, but the cyclical component is common to the two series.[20] Assume that the log of real GDP, y_t, contains a stochastic trend, τ_{yt}, and a stationary cyclical component, c_{yt}. The unemployment rate, u_t, similarly has a trend component, τ_{ut}, and a stationary cyclical component, c_{ut}. The model is then:

$$y_t = \tau_{yt} + c_{yt} \tag{33}$$

$$\tau_{yt} = \delta + \tau_{y,t-1} + v_t \qquad v_t \sim \text{iid } N(0, \sigma_v^2) \tag{34}$$

$$c_{yt} = \phi_1 c_{y,t-1} + \phi_2 c_{y,t-2} + \varepsilon_t \qquad \varepsilon_t \sim \text{iid } N(0, \sigma_\varepsilon^2) \tag{35}$$

[19] In essence, this means that short-run irregularities in real GDP are apparently filtered out when the data are constructed.

[20] One advantage of using the inflation rate over unemployment is its closer linkage to the derived estimate of the output gap, whose definition is based on stable inflation. Cyclically high inflation rates would be correlated with the cyclical component of output through c_{yt}.

$$u_t = \tau_{ut} + c_{ut} \tag{36}$$

$$\tau_{ut} = \tau_{u,t\text{-}1} + \eta_t \qquad \eta_t \sim \text{iid N}(0,\, \sigma_\eta^{\;2}) \tag{37}$$

$$c_{ut} = \beta_0 c_{yt} + \beta_1 c_{y,t\text{-}1} + \beta_2 c_{y,t\text{-}2} + \zeta_t \qquad \zeta_t \sim \text{iid N}(0,\, \sigma_\zeta^{\;2}) \tag{38}$$

where all errors are white noise and the cyclical component of unemployment, c_{ut}, is assumed to be a function of current and past transitory components of output.

In fact, papers such as Kuttner (1994), Planas and Rossi (2004), Domenech and Gomez (2006), Planas *et al.* (2008) and Harvey (2011) use UC in models linking inflation with the output gap. Popular macroeconomic relations that are added to univariate state-space models are a Phillips curve, an IS curve and a relationship representing Okun's law. The main motivation in these papers is to use the information in inflation to obtain better estimates of the output gap.

4.2.4 Regime-shifting decompositions: the Markov-switching model

Of particular interest to economists is the apparent tendency of many economic variables to behave quite differently during economic downturns, when underutilisation of factors of production rather than their long-run tendency to grow governs economic dynamics. In fact, Mitchell (1927) described asymmetry as a regularity of the business cycle. This is the idea that expansions are fundamentally different from recessions in their duration and in the abruptness of changes in growth.

This empirical observation underscores the appeal of the Markov-switching model, which is a non-linear model given by:

$$y_t = c_{s_t} + \sum_{i=1}^{p} \beta_{i,s_t} y_{t-i} + \varepsilon_t \tag{39}$$

where $\varepsilon_t \sim$ iid(0, $\sigma_{s_t}^2$) and s_t is a Markov-chain taking values in the set {1, … , m} with transition matrix P. The model essentially implies that there are m different regimes in the economy regulated by an unobserved Markov chain. Equation (39) has the appearance of a linear AR(p) model but for the subscript s_t on the intercept, the autoregressive coefficients and the residual variance. Individually, the model of the economy in each regime is given by an AR(p) process with regime-specific intercept, autoregressive coefficients and residual variance. Taken together, though, the combination of the different linear models results in a non-linear model. In order to estimate and calibrate a switching regime model to associate growth and growth volatility for different unobserved states of the economy, we have to decide on the number of regimes existing in the economy. In the case of potential output and the output gap, it seems natural to set the number of regimes equal to three: recession, sustainable growth and overheating economy. The definition of sustainable growth will define the pattern for potential growth and the output gap since, under sustainable growth, macroeconomic policy should aim to be neutral and, if this is the case, the economy should be growing at potential.

The model can be estimated by maximum likelihood using the filter suggested by Hamilton (1989) and an empirical application of this approach can be found in Johnson (2013).

4.3 Hybrid decompositions

As discussed above, filtering is different from detrending, as filters are operators that carve out particular frequencies of the spectrum of data.[21] For that reason, one of the preferred methods for detrending data involves the use of filters designed to separate the cycle from the trend, whilst allowing for a slowly evolving trend. A number of these filters employ economic theory to justify the particular frequencies of the data that the filters use for extracting a cycle from the data.

4.3.2 The Hodrick-Prescott filter

Suppose you observe the values y_1 through y_T and want to decompose the series into a trend, τ_t, and stationary component $c_t = y_t - \tau_t$. Probably the most widely-used method to decompose a GDP series into trend and cycle is the Hodrick-Prescott (1997) (HP) filter, which allows for time variation in the trend estimate.[22] Its main strength lies in its simplicity.[23]

It is a two-sided linear filter (to be defined below) that computes the smoothed series τ_t of y_t by minimising the variance of y_t around τ_t subject to a penalty that constrains the second difference of τ_t. The idea underlying this approach is that because of self-equilibrating forces, actual output fluctuates around potential output over time. The HP decomposition is characterised by two basic features, which are that the trend and the cycle are assumed to be uncorrelated, and that the trend is assumed to be stochastic while moving 'smoothly' over time. The latter is understood to mean that the trend can change over time as long as the changes are not abrupt. By penalising variations in the second difference (i.e., the acceleration) of the trend, Hodrick and Prescott (1997) operationalize the concept of 'smoothness'. In order to apply a statistical filter, it is typically necessary to specify specific parameters in advance, such as the desired degree of smoothness of the extracted trend, in this case denoted by λ. With y_t representing the raw series at time t and τ_t representing the (time-varying) trend of the series, the HP filter (indirectly) extracts a trend which minimises the following loss function (with respect to τ_t):

$$\min_{\tau_t}\left\{\sum_{t=1}^{T}(y_t-\tau_t)^2+\lambda\sum_{t=2}^{T-1}\left[(\Delta^2\tau_t)^2\right]\right\}$$

[21] Still, this has not prevented the empirical literature from using the terms detrending and filtering interchangeably for the process of extracting cycles, even though the two processes are quite distinct.
[22] Canova (1998, p. 485) also argues that 'The popularity of the HP filter among applied macroeconomists results from…[the fact that the] implied trend line resembles what an analyst would draw by hand through the plot of the data'. In addition, Canova (1999) points out that the HP-extracted business-cycle movements for the US resemble the NBER-backed recession definitions.
[23] Mise *et al.* (2005) show that the HP filter is optimal for a wide class of time series.

$$= \min_{\tau_t}\left\{\sum_{t=1}^{T}(y_t-\tau_t)^2+\lambda\sum_{t=2}^{T-1}\left[((\tau_{t+1}-\tau_t)-(\tau_t-\tau_{t-1}))^2\right]\right\}$$

$$= \min_{\tau_t}\left\{\sum_{t=1}^{T}(y_t-\tau_t)^2+\lambda\sum_{t=2}^{T-1}\left[(\tau_{t+1}-2\tau_t+\tau_{t-1})^2\right]\right\} \qquad (40)$$

where T is the sample size. The first term of equation (40) represents the sum of the squared deviations of the trend from the actual data, and captures how close the extracted trend is to the underlying data series. The second term of equation (40) represents the extent to which the slope of the trend changes between successive observations in such a way that the lower this value, the 'smoother' the trend. Equation (40) is meant to capture the trade-off between the extent to which a trend matches the data and the smoothness of the trend.[24]

The weight that is placed on the smoothness of the trend relative to deviations from the actual data is captured by the smoothing parameter, λ. The parameter λ determines the importance of having a smoothly evolving growth component: with $\lambda = 0$, no weight is placed on the smoothness of the trend and all variation in y_t will be assigned to the trend component, meaning that the extracted trend will simply be equal to the original series; as $\lambda \to \infty$, the trend is assigned to be maximally smooth, that is, the HP trend converges to a simple linear trend. In general, λ is specified to strike a compromise between these two extremes.

The standard values of λ are $\lambda = 100$ for annual data, $\lambda = 1600$ for quarterly data and $\lambda = 14400$ for monthly data. The typical value of λ for quarterly data, equal to $\lambda = 1600$, implies that the standard error of the cycle is 40 times larger than the standard error of the second difference of the trend. This standard value for λ retains cycles with an average duration of 4 to 6 years in the data. In practical terms this means that cycles longer than 6-7 years are attributed to the trend. We should note that the choice of λ is not innocuous. In fact, implicit estimates of λ obtained using BN or UC decompositions are only in the range [2,8].[25]

The unobservable output gap (denoted `hp_cycle`) is computed using the natural logarithm of real GDP (`lrgdp`), which has been de-trended using the Hodrick-Prescott (HP) filter. To smooth the series `lrgdp` using the HP filter, double-click on it to open the series and choose **Proc/Hodrick-Prescott Filter…**, which opens up the Hodrick-Prescott Filter dialog box. We can provide a name for both the smoothed and the cycle series. EViews will suggest a name for the smoothed series, but I have chosen to call them `hp_trend` and `hp_cycle` respectively. EViews has recognised the fact that the data are annual, so has specified a value of 100 for the smoothing parameter, λ, automatically. You may specify any other value for λ directly or use the frequency power rule of Ravn and Uhlig (2002) – we will do so a bit later on to assess the robustness of the results to the choice of λ. Upon clicking **OK**, EViews will display a graph of the filtered series together with the original series (Figure X). We also note that the workfile now contains the additional two series that we have created.

Alternatively, the EViews command when using the command line is either:

```
lrgdp.hpf(100) hp_trend @ hp_cycle
```

[24] The trade-off is that a very smooth trend will only loosely match the original data series, while a trend that matches the original data series very closely will display a much greater degree of volatility.

[25] It has been argued that if the shocks to the economy are primarily demand-driven, then potential output does not move closely with actual output data and a higher degree of smoothing should be used in the filter. The opposite applies in the case of supply shocks, when a lower degree of smoothing is appropriate.

when you want to extract both the smoothed and the cycle series (which will be given the name of `hp_trend` and `hp_cycle` in the workfile respectively) or simply:

```
lrgdp.hpf(100) hp_trend
```

when you only want the smoothed series, after which you obviously need to define the output gap as the difference between the two series:

```
hp_cycle = lrgdp - hp_trend
```

Figure 6: Sri Lankan output gap (calculated using the Hodrick-Prescott filter and the Beveridge-Nelson decomposition), 1959 - 2012

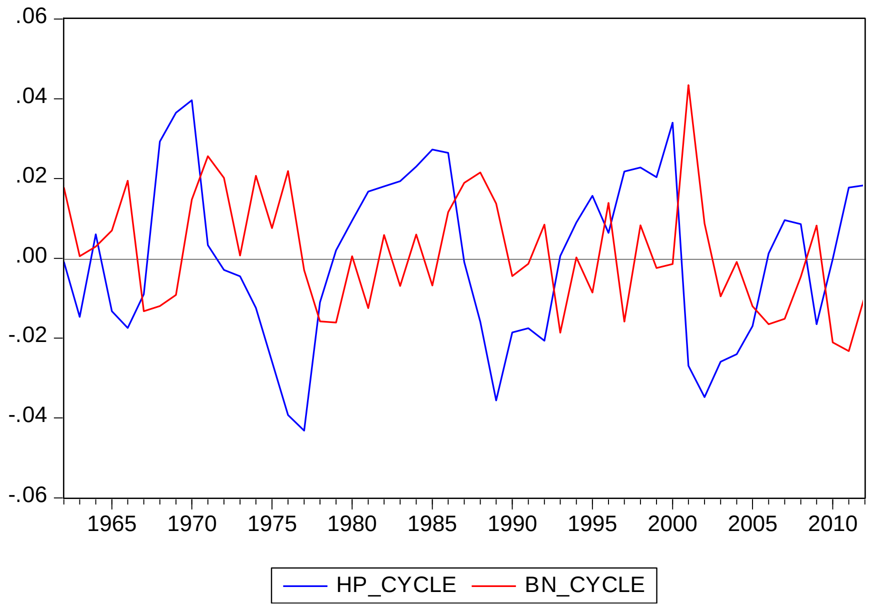


Figure 6 shows that the output gap measure has been quite volatile, fluctuating in a band from roughly -4 to 4 per cent since 1959. At the same time, the estimates are smoother than those derived from a BN decomposition, underscoring the earlier argument that BN-derived output gap estimates are more volatile than the HP-derived counterparts. We can see that the output gap is relatively large and positive at the end of the sample period in 2012. We will return to this point below.

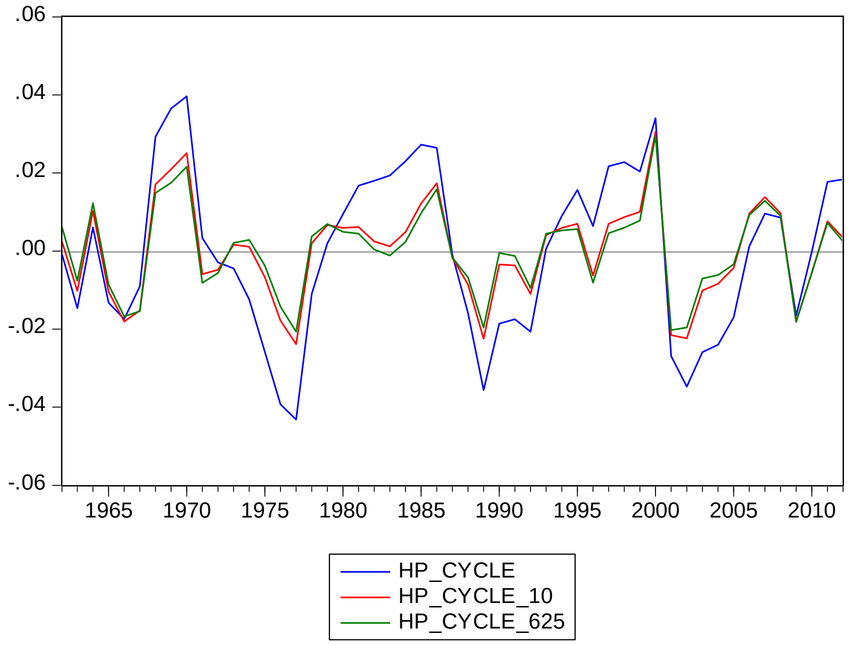
In essence, the value of λ is set *a priori* to isolate those cyclical fluctuations which belong to the specific frequency band the researcher wants to investigate. The standard value for λ for quarterly data is 1600, which has been calibrated on US GDP data.[26] As such, it might not be an appropriate parameter value for emerging market and developing economies where the trend might be a lot less smooth. Canova (1998) investigated this issue and

[26] The value of 100 (14400) for annual (monthly) data is due to Backus and Kehoe (1992), who suggested an adjustment of the value of λ by multiplying the standard (quarterly) value of 1600 with the square of the frequency of observations relative to quarterly data. For example, the relative frequency is 3 for monthly data and 1/4 for annual data, which makes the corresponding values of the smoothing parameter equal to $\lambda = 1600/4^2 = 100$ and $\lambda = 1600 \times 3^2 = 14400$ for annual and monthly data respectively.

recommended assessing the robustness of the results by using three different values for λ for quarterly data, which are 8, 40 and 1600.[27]

Alternatively, we can use the data-dependent method due to Ravn and Uhlig (2002), who suggested using the fourth power of the change in observation frequency for data observed at a different frequency.[28] Their frequency power rule uses the number of periods per year divided by 4, raised to a power and multiplied by 1600. On the basis of their analysis in both the frequency- and the time-domain, Ravn and Uhlig (2002) recommend a power value of 4. For example, yearly data are observed only one-fourth as often as quarterly data. Hence, for yearly data $\lambda = 1600/4^4 = 6.25$ is recommended. Note that a power rule of 2 recovers the original Hodrick Prescott values for λ.[29] Overall, rather than the value of $\lambda = 100$ traditionally used for annual data, Ravn and Uhlig (2002, p. 371) cite values of $\lambda = 6.25$ and $\lambda = 10$ (suggested by Hassler *et al.* (1992) and Baxter and King (1999)) as potentially being more appropriate for that particular sampling frequency. Figure 7 shows the estimated output gaps for different values of λ.

Figure 7: Sri Lankan output gap (calculated using the Hodrick-Prescott filter with λ = 6.25, 10 and 100), 1959 - 2012



The estimated output gaps show the sensitivity of the results to different values of λ. Using the smaller values of $\lambda = 6.25$ or $\lambda = 10$ shows that the output gap at the end of the period is not only much smaller, but also in the process of closing.

---

[27] As we will see later on, the HP filter is equivalent to the smoothed trend from an unobserved components model with a stochastic trend component and a random irregular term. The signal-to-noise ratio, that is, the ratio of the variance of the trend shock over the variance of the irregular shock, is equal to the inverse of the HP smoothing constant (Harvey and Jäger (1993), Harvey and Trimbur (2008)). This means that a smaller λ is associated with a higher signal-to-noise ratio.

[28] For completeness sake, I should also mention the method of optimal filtering due to Pedersen (2001), which is meant to provide optimal values for λ. Ross and Ubide (2001) also discuss alternative approaches for determining the parameter λ endogenously.

[29] For yearly data and a power rule of 2, $\lambda = 1600/4^2 = 100$.

But its sensitivity to the value of the smoothing parameter, it is also worth pointing out that the Hodrick-Prescott filter in EViews is generally – and erroneously – employed as a one-sided filter, meaning that it is used with historical data up to time t (today). It is, however, the case that the HP-filter uses observations at time $t + i$, $i > 0$ to construct the current time point t. Stopping the sample at the current date means that only current and past states, rather than past, current **and** future states influence the current observation.

This is quite difficult to infer from the equation (40). It might be easier to see from differentiating the final line of equation (40) with respect to τ_t and re-arranging the first-order conditions for minimisation. This operation yields the Euler equations for each time period:

$$(y_1 - \tau_1) = \lambda(\tau_3 - 2\tau_2 + \tau_1) \quad \text{for } t = 1$$

$$(y_2 - \tau_2) = \lambda(\tau_4 - 4\tau_3 + 5\tau_2 - 2\tau_1) \quad \text{for } t = 2$$

$$\ldots$$

$$(y_t - \tau_t) = \lambda(\tau_{t+2} - 4\tau_{t+1} + 6\tau_t - 4\tau_{t\text{-}1} + \tau_{t\text{-}2}) \quad \text{for } t = 3, \ldots, T\text{-}2 \tag{41}$$

$$\ldots$$

$$(y_{T\text{-}1} - \tau_{T\text{-}1}) = \lambda(\text{-}2\tau_T + 5\tau_{T\text{-}1} - 4\tau_{T\text{-}1} + \tau_{T\text{-}3}) \quad \text{for } t = T\text{-}1$$

$$(y_T - \tau_T) = \lambda(\tau_T - 2\tau_{T\text{-}1} + \tau_{T\text{-}2}) \quad \text{for } t = T$$

Writing the first-order conditions this way illustrates the end-point problem extremely clearly. Note that the first-order conditions for $t = 1, 2$ as well as $t = T\text{-}1, T$ are different from the first-order conditions for the remaining observations. Let us consider the issue at the end of the sample. From equation (41) we can see that the observations for $t = 3, \ldots, T\text{-}2$ involve terms in τ_{t+2} and τ_{t+1}. Towards the end of the sample, these observations fall outside the sample range, which means that the final two first-order conditions have been adjusted and only contain contemporaneous as well as two lagged observations.[30]

As a result, the assessment of the difference between the trend and the actual data might change substantially when new data is added. This makes the HP filtered series a problematic output gap measure as it fails to correctly measure the **current** state of the economy relative to potential and hence lead to inappropriate policy decisions.[31] As suggested in Mise *et al.* (2005), the end-point problem with the HP-filter can be avoided if a sufficient number of GDP forecasts are available that can be included in the underlying data series.[32] Sufficient in this context generally refers to 28 quarters or seven years. We therefore augment the real GDP data for Sri Lanka, which end in 2012, with forecast values from the IMF October 2013 *World Economic Outlook* database, which take the total sample of the data out to 2018. By having data beyond the end of the sample period of actual data, the endpoint problems associated with the HP-filter are avoided (Mise *et al.* (2005)). Note

[30] Note that the same is true, in 'reverse, so to speak, for the first two observations in the sample.

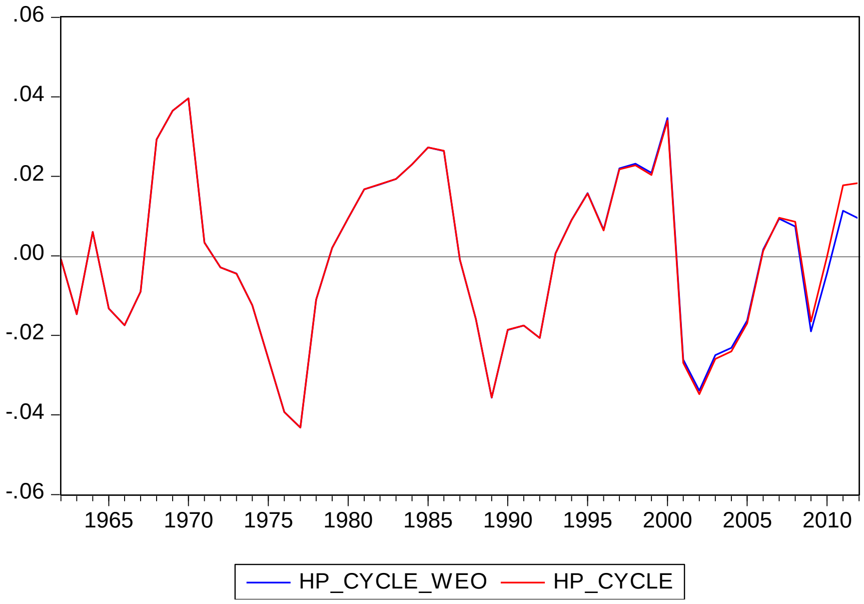
[31] An apocryphal description of just such a policy error on account of a misleading potential output estimate is given in Orphanides (2002). He argues that the Fed believed the output gap in the 1970's to be much more negative than it really was, which led policymakers to take action that overheated the economy and contributed to an inflationary surge.

[32] Watson (2007) finds similar results when estimating the trend with a band-pass filter.

that the end-point problem can be serious. The main source of revisions in output gap estimates is due to the pervasive unreliability of end-of-sample estimates, be it using the HP filter or the BP filters (Orphanides and van Norden (2002)). The question of whether end-point problems are serious or not has been partially addressed in Watson (2007), who concluded that future values were critical to accurately separate the trend from the fluctuations for series showing serially correlated fluctuations around a slowly evolving trend.

Figure 8 shows the results of the standard HP filter using data up to and including 2012 (the red line) and the standard HP filter using data up to and including 2018. The current, i.e., 2012, estimate of the output gap is much smaller than the naïve HP filtered series.

Figure 8: Sri Lankan output gap (calculated using the Hodrick-Prescott filter with augmented real GDP data up to 2018), 1959 - 2012



This finding highlights some of the advantages and disadvantages of the HP filter. Let us start with the disadvantages first.

As is obvious from the definition of the HP filter above, the method lacks a close link to economic theory, such that the HP filter abstracts from any potential relevant economic information other than the output series itself. In other words, the filter does not exploit the interactions between unemployment, output and inflation. As a purely statistical technique designed to smooth actual GDP data in order to estimate potential output, it may be slow in 'discovering' structural breaks. As a result, this entails the risk of **overestimating** potential GDP if the sample period ends with a sudden and large loss of output. Harvey and Jäger (1993), amongst others, report that the uncritical use of mechanical detrending with the HP filter can lead investigators to report spurious cyclical behaviour and show how particular quantitative features of the business cycle are not robust to the choice of detrending.[33] Finally, the reliability of estimates of the output gap in real time is studied in Orphanides and

[33] This is particularly the case for series that are either integrated or driven by deterministic trends. In particular, applying the standard HP filter to a random walk produces detrended observations which have the characteristic of a business cycle for quarterly data (Harvey and Jäger (1993, p. 234).

van Norden (2002), who find highly volatile estimation errors for real-time estimates of the same magnitude as the output gap itself. Attribution of the error to various causes indicates that the end-of-sample problem is the most important reason for the estimation error.

The advantages of the HP-filter are that it implicitly allows potential output growth to vary over time and that it can extract the **same** trend from a set of variables. This will not be the case for the BN decomposition, which will not yield the same trend when applied to each series individually.

In summary, the HP filter is a mechanical device which defines the extracted cycles *via* the choice of λ. The standard values of λ (λ = 100 for annual data, λ = 1600 for quarterly data and λ = 14400 for monthly data) have all been calibrated to the mean length of domestic US cycle data, which makes it unlikely that they will correspond exactly to typical business-cycle frequencies in other countries. For example, if a country has cycles with an average length of 9 years, the mechanical application of the HP filter will move these cycles to the trend.

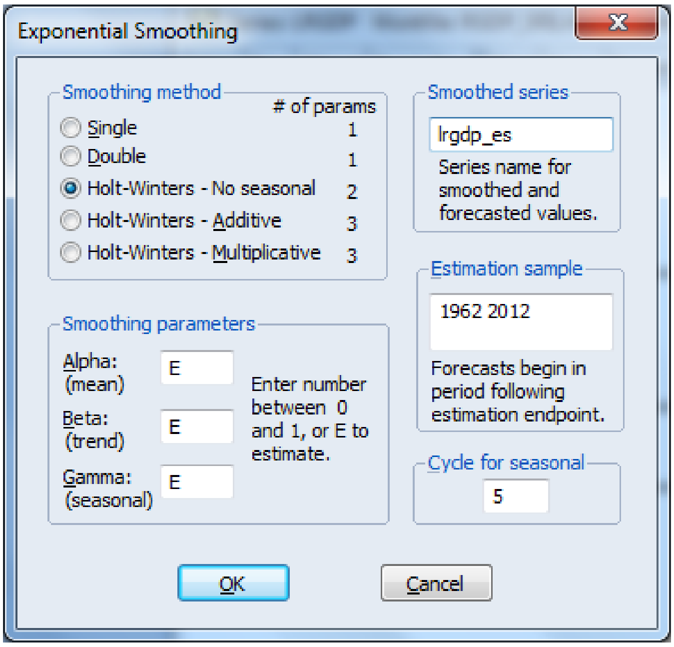
4.3.3 The exponential smoothing (ES) filter

The application of the exponential smoothing filter for (business) cycle extraction has a long history, with Lucas (1980) being an early application. In contrast to the HP filter, which penalises the acceleration of the trend (equation (40)), the ES filter penalises only the change in the trend. The rest of the penalty function remains unchanged, even though the first-oder conditions for minimisation will obviously also be different:

$$\min_{\tau_t}\left\{\sum_{t=0}^{T}(y_t-\tau_t)^2+\lambda\sum_{t=0}^{T-1}\left[(\Delta\tau_t)^2\right]\right\} = \min_{\tau_t}\left\{\sum_{t=0}^{T}(y_t-\tau_t)^2+\lambda\sum_{t=0}^{T-1}\left[(\tau_t-\tau_{t-1})^2\right]\right\} \qquad (42)$$

The ES and the HP filters are broadly similar (Figure X). The main difference is that the ES filter picks up trends with longer periodicity.

To perform exponential smoothing, double-click on the `lrgdp` series to open it and select **Proc/Exponential Smoothing...**. We use the **Holt-Winters – No seasonal** as this method is appropriate for series with a linear time trend and no seasonal variation. Name the smoothed series `lrgdp_es` and estimate all parameters over the period 1962-2012. Leave the remaining settings at their default values, as shown in the following screenshot.



When you click **OK**, EViews displays the results of the smoothing procedure, reproduced in Table 4.

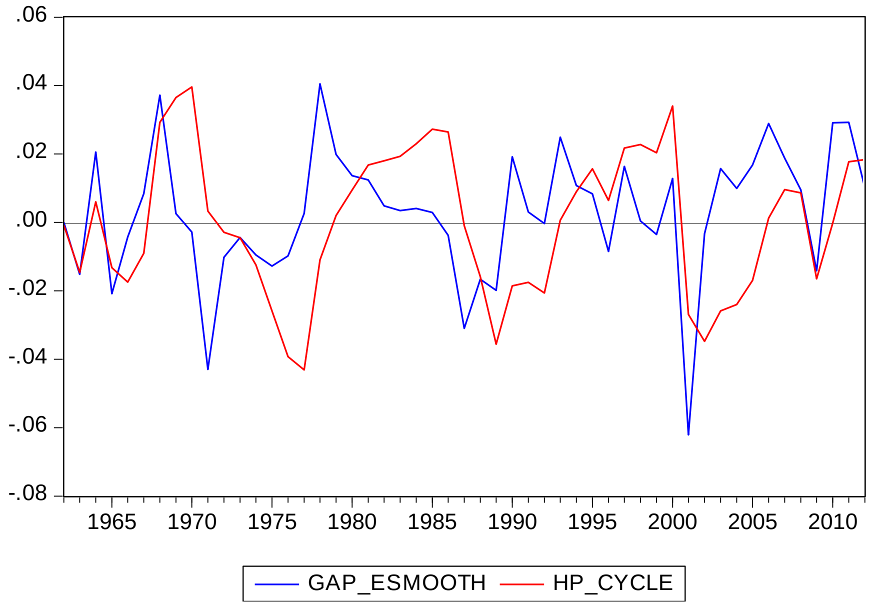
Table 4: Results from exponential smoothing

Sample: 1962 2012
Included observations: 51
Method: Holt-Winters No Seasonal
Original Series: LRGDP
Forecast Series: LRGDP_ES

Parameters:	Alpha	1.0000
	Beta	0.0700
Sum of Squared Residuals		0.018468
Root Mean Squared Error		0.019030
End of Period Levels:	Mean	8.022005
	Trend	0.052750

The first part displays the estimated (or specified) parameter values, the sum of squared residuals and the root mean squared error of the forecast. The values for α and β in this example mean that the smoothed series is calculated as $y_{t+k} = \alpha + \beta k$. The output gap is then calculated as `gap_esmooth = lrgdp – lrgdp_es`, and displayed in Figure 9 with the standard HP filter for comparison.

Figure 9: Output gaps from exponentially-smoothed series and HP filter, 1959-2012



4.3.4 The moving average (MA) filter

But the approach with the longest history in business-cycle analysis, going all the way back to the seminal analysis by Burns and Mitchell (1946), is the MA filter. The latter is defined by a polynomial $B(L)$, which is either one- or two-sided. The former operates on p lags of y_t while the latter operates on p lags as well as p leads of y_t. We will not discuss MA filters in the context of this practical guide, as it has long been known that the mechanical application of moving-average filters can create a wide range of undesirable effects in the data (Fishman (1969)). Furthermore, Osborn (1995) looked at the consequences of applying moving-average detrending in the context of integrated processes and also found instances of spurious cyclicality.

4.3.5 The band-pass (BP) filter

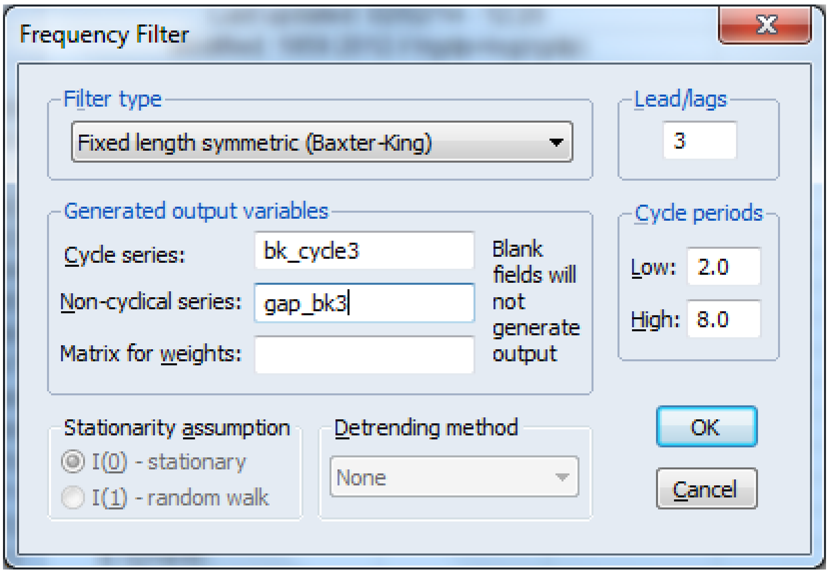
The point of departure for the band-pass (BP) filter is that any time series can be thought of as being built up from a set of underlying sub-processes whose individual frequency components exhibit cycles of varying frequency. This approach to modelling (economic) time series is called the frequency domain. The other side of the argument is that you can, in principle, take any time series y_t and figure out what fraction of its variance is generated from those sub-processes that cycle within a given frequency range. One way to classify filters is then according to the frequencies that are allowed to pass through and those that are blocked. A high-pass filter lets through only the high-frequency components, while a low-pass filter allows through the trend or growth frequencies. The business-cycle frequency, for example, is generally acknowledged to lie between 6 and 32 quarters, so a business-cycle pass filter should allow through frequencies ranging from 6 to 32 quarters.

The band-pass filter then is designed to shut down all fluctuations outside of a chosen frequency band between cycles with periods between P_L and P_U. For business-cycle applications, P_L and P_U are generally set equal to 6 and 24-32 respectively. The appeal of BP filters is due to the fact that they make the notion of business cycles operational by selecting fluctuations in a pre-specified range, say 6 to 24-32 quarters. While it is not feasible to implement the ideal band-pass filter, as doing so requires an infinite number of observations of the unfiltered series as input, finite-sample approaches have been developed by Baxter and King (1999) and Christiano and Fitzgerald (2003) in the time domain and Corbae and Ouliaris (2006) in the frequency domain.[34]

In EViews, the different band-pass filters are available as a series **Proc**, meaning that they are invoked by opening the series under observation and going to **Proc/Frequency Filter…**. EViews makes three band-pass filters available:

- fixed length symmetric (Baxter-King) – this is EView's default;
- fixed length symmetric (Christiano-Fitzgerald); and
- full length asymmetric (Christiano-Fitzgerald)

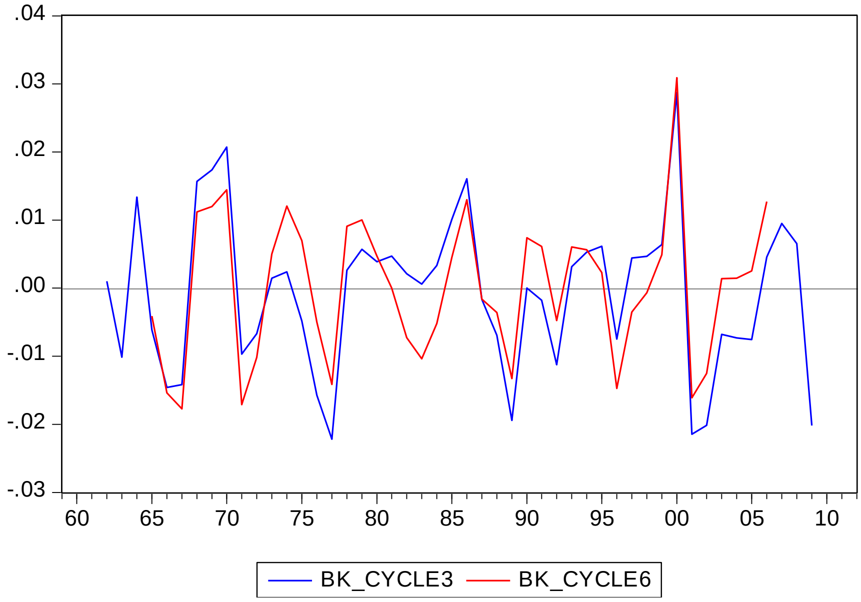
Let us start with the BK filter. Double click on lrgdp to open the series, go to **Proc/Frequency Filter…** and the dialog box for the fixed length symmetric Baxter-King filter appears. There are only a few options associated with the Baxter-King band-pass filter. As such, we need to select the frequency length (i.e., the lags and leads) for the moving average and the low and high values for the cycle period (P_L, P_U). By default, EViews will fill the required dialog boxes with reasonable default values that are based on the type of the workfile. If need be, they can of course by manually adjusted. Moreover, we have the option of naming the series objects that will contain the saved output for the cyclical and non-cyclical components. The **Cycle series** will be a series object containing the filtered series (cyclical component), while the **Non-cyclical series** is simply the difference between the actual and the filtered series. Note that the options for **Cycle periods** correspond to a frequency filter between 2 and 8 years. The following screenshot shows the required input for applying the BK filter to `lrgdp`:



[34] Broadly speaking, band-pass filters are designed to eliminate high and low frequency movements in the data using a (weighted) two-sided symmetric moving average of the data, where cycles in a 'band', given by a specified lower and upper bound, are 'passed' through (i.e., extracted) and the remaining cycles outside the band are 'filtered' out. In so doing, they isolate the cyclical component of a time series by specifying a range for its duration.

We then repeat the exercise, but set the number of **Lead/lags** to 6 to assess the sensitivity of the results to the choice of leads and lags. The results for the cycle series using `lrgdp` and both 3 and 6 leads/lags are shown in Figure 10.

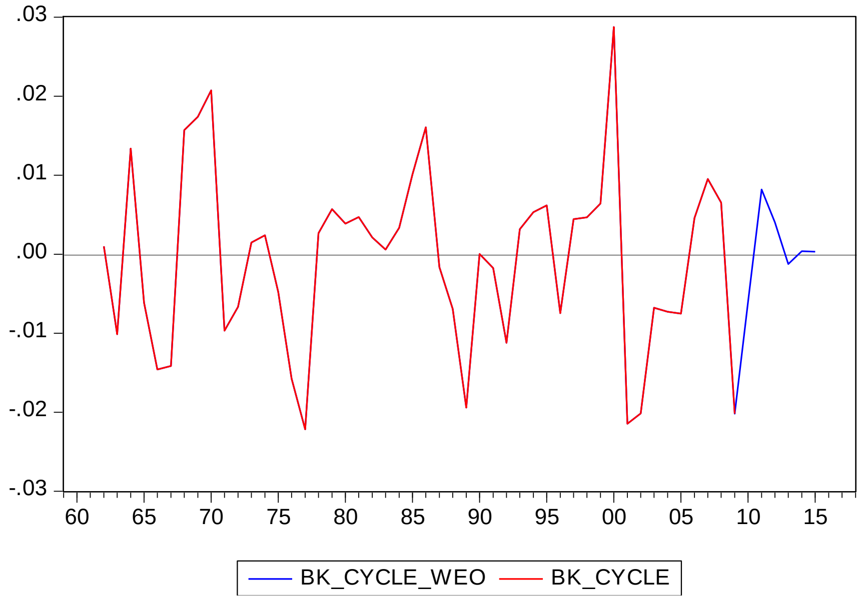
Figure 10: Baxter-King band-pass estimates of the Sri Lankan output gap



It turns out that band-pass filters are also prone to end-point problems. The fixed-length versions of the BK and CF filters use the same number of leads and lag terms for every weighted moving average. In other words, if the filtered series is extracted using *p* leads and lags of the observed data, we will lose *p* observations both at the beginning and the end of the original sample. As a result, Watson (2007) also recommends 'padding' the data into the future (and past) using forecasts and backcasts of the series, although the error in band-pass filter estimates arises from the use of forecasts of future values of y_t. As mentioned in the context of the HP filter above, because of the end-point problem, any two-sided symmetric filter will not be appropriate for policy analysis, which mainly focuses on assessing the current state of the economy.

The results for the cycle series using both the original data series, `lrgdp`, as well as the augmented data series, `lrgdp_weo`, are shown in Figure 11.

Figure 11: Baxter-King band-pass estimates of the Sri Lankan output gap

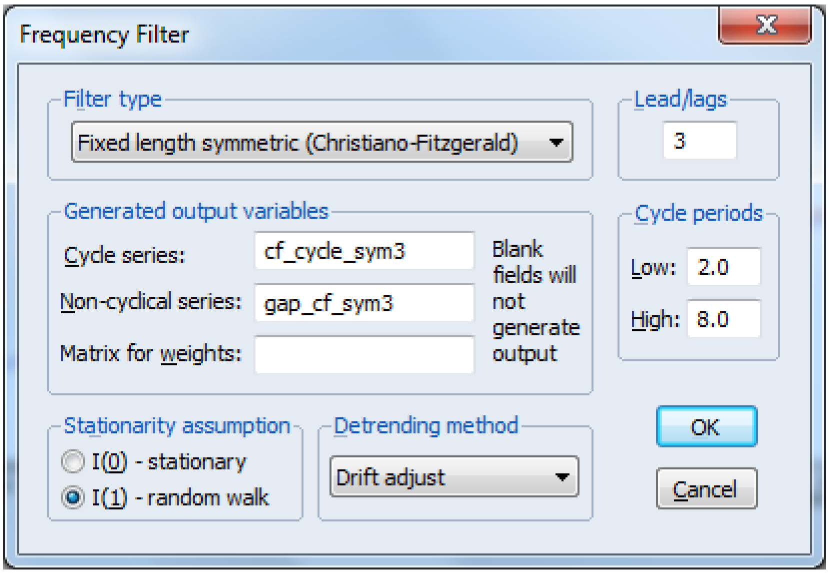


The red line, which is the output gap estimate based on data up to 2012, and the blue line, which shows the output gap using data until 2018. We note two things. To begin with, the red line starts in 1962 and ends in 2009, as we lose three observations On account of the lags and leads respectively. Two, the exact juxtaposition of the red and the blue line shows that the symmetric filter is time-invariant since the moving average weights depend only on the specified frequency band and do not use the data. This is why the output gap estimates over the historical data period overlap. The benefit that arises from using an augmented data set is that it will allow us to generate output gap estimates at the end of the sample. This is because a filtered series computed using *p* leads and lags (*p* = 3 and 6 in our case) will lose *p* observations at both the beginning and the end of the series.[35]

End-point problems will also affect the symmetric version of the Christiano-Fitzgerald band-pass filter. In addition to the filters we have come across for the Baxter-King band-pass filter, the Christiano-Fitzgerald filters provide you with additional options for handling trending data. This is the case for both the symmetric and the asymmetric versions of the CF filter. The first additional setting involves the **Stationarity assumption**, which requires us to specify whether the series under investigation is presumed to be an I(0) covariance stationary process or an I(1) unit-root process. Prior to applying the CF filters, we can choose to either demean or detrend the covariance stationary data or to demean, detrend or remove the drift using the adjustment suggested by Christiano and Fitzgerald (2003). For the purposes of this analysis, we will assume that `lrgdp` is a non-stationary series.

The dialog box for the symmetric CF filter with `lrgdp` being I(1) is as follows:

[35] On the other hand, the BK filter fulfils the stability condition mentioned in Section 2.



The output of the fixed length symmetric Christiano-Fitzgerald band-pass filter for both p = 3 (the blue line) and p = 6 (the red line) is displayed in Figure 12. The obvious end-point problems with the use of leads and lags are again quite apparent, as is the sensitivity of the estimated output gaps to different values of p.

Figure 12: Symmetric Christiano-Fitzgerald band-pass estimates of the Sri Lankan output gap

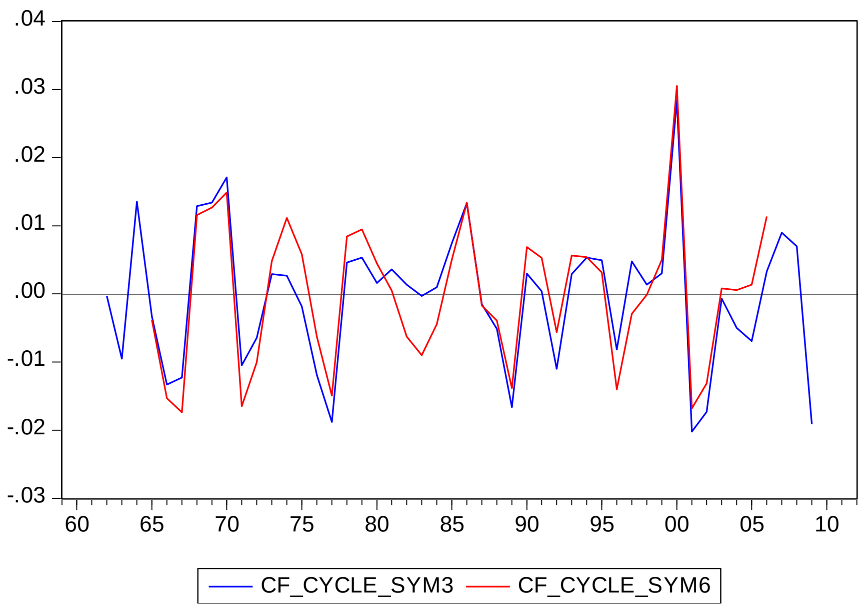
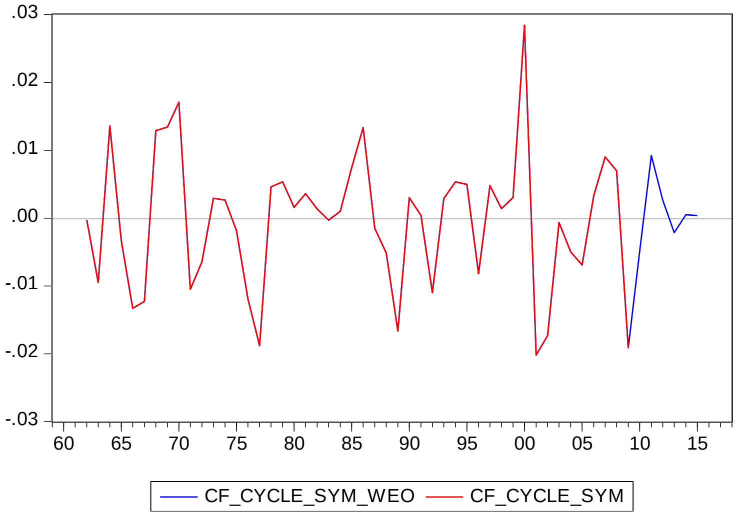


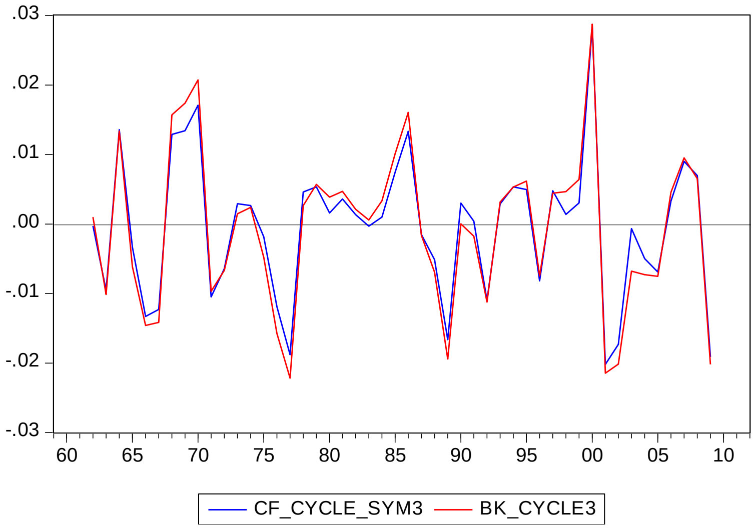
Figure 13 shows the estimated output gap using the augmented real GDP data out to 2018. Just as in the case of the symmetric Baxter-King band-pass filter, the exact juxtaposition of the red and the blue line shows the symmetric CF band-pass filter to be time-invariant. Once again, this is because the moving average weights depend only on the specified frequency band and do not use the data. This is why the output gap estimates over the historical data period overlap. The benefit that arises from using an augmented data set is that it will allow us to generate output gap estimates at the end of the sample.

Figure 13: Symmetric Christiano-Fitzgerald band-pass estimates of the Sri Lankan output gap



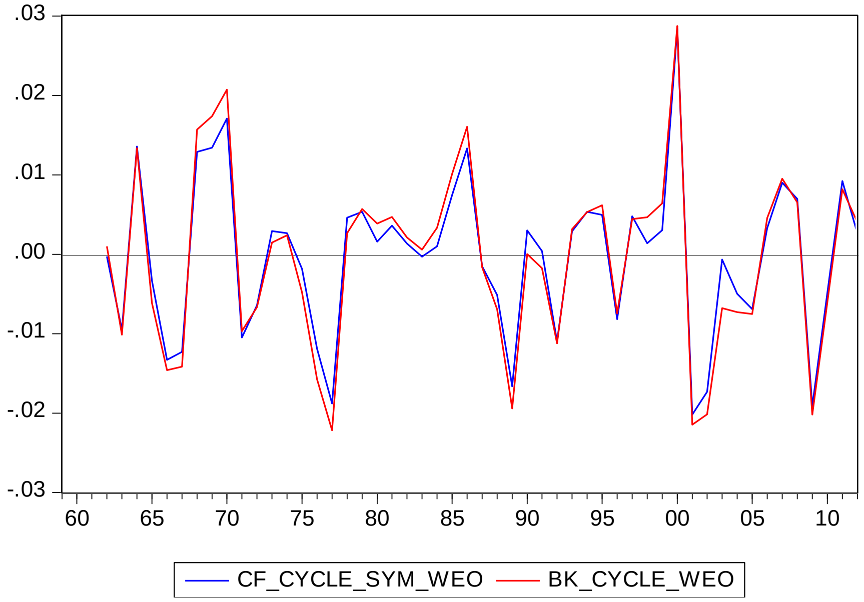
Even though the resulting BK and CF output gap estimates look the same, they are very slightly different, as Figure 14 shows. The slight differences arise because the fixed-length symmetric BK and CF band-pass filters differ in the choice of the objective function used to select the moving-average weights.

Figure 14: Symmetric Baxter-King and Christiano-Fitzgerald band-pass estimates of the Sri Lankan output gap



Regardless of the slight differences, both band-pass filters show a declining output gap in 2011 and 2012 using the augmented data up to and including 2018 (Figure 15).

Figure 15: Symmetric Baxter-King and Christiano-Fitzgerald band-pass estimates of the Sri Lankan output gap (augmented data)

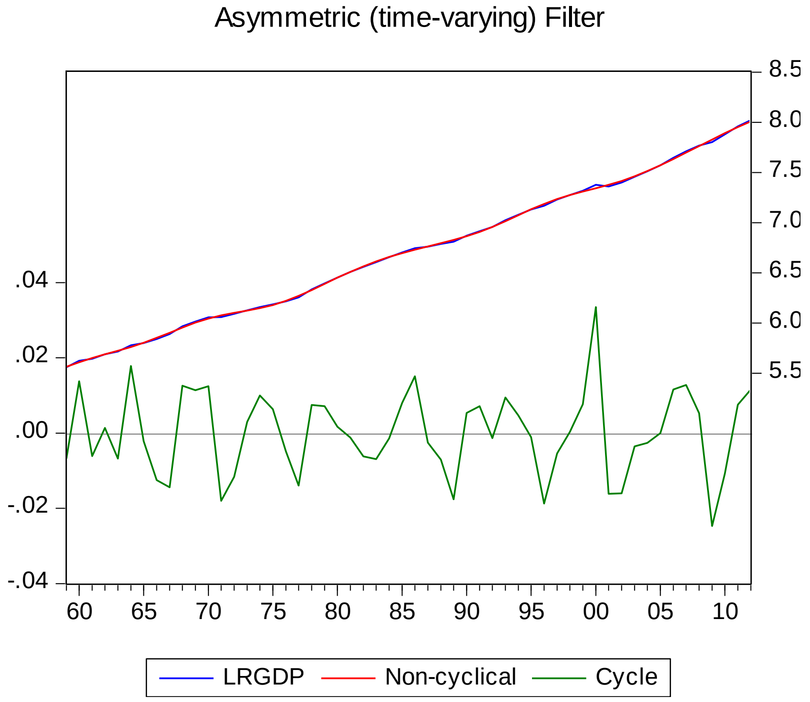


In addition to a symmetric band-pass filter, Christiano and Fitzgerald (2003) (CF) also developed a filter that is non-stationary and **asymmetric**, but still depends on the time-series properties of the underlying data. The asymmetric version of the CF filter is quite general and also has the advantage that it solves beginning and end-of-sample problems. Remember that the underlying idea is that the ideal band-pass filter decomposes a time series into different frequency components through a linear transformation of the data that leaves the components for a specified frequency band intact and eliminates all other components. In theory, the ideal filter requires an infinite data set, so an approximation is required. In the case of the CF filter, the authors construct an approximation to the ideal band-pass filter which is optimal when the underlying raw data follow a random walk (hence the non-stationarity mentioned above). In practice, the end-point problem is resolved by padding the series with random-walk forecasts. In contrast to the earlier symmetric filters, the asymmetric filtered series can therefore be computed to the ends of the original sample.

But the cost of this generality is that other *a priori* choices have to be made, in particular whether y_t is stationary or integrated and what its serial correlation properties are. Reassuringly perhaps, Christiano and Fitzgerald (2003) indicate that, in practice, an approximation obtained by arbitrarily assuming that y_t is a random walk works well for a variety of macroeconomic time series.

The dialog box for the asymmetric CF filter with `lrgdp` assumed to be I(1) is as before, but as this filter is 'full length', it uses all of the observations in the sample so that the **Lead/Lag** option is irrelevant (and indeed greyed out). Part of the output from applying the asymmetric CF filter to `lrgdp` is shown in Figure 16 below.

Figure 16: Asymmetric (full length) Christiano-Fitzgerald band-pass estimates of the Sri Lankan output gap



Both the BK and the CF filter operate with the time-domain representation of the filter. But it is also possible to work in the frequency domain. The advantage of the latter approach is that no approximation, i.e., arbitrary truncation involving *p* lags and leads, is needed and no loss of data at the beginning or the end of the sample is involved. But two major drawbacks still exist. One, the definition of the cyclical component now depends on the sample size.[36] This means that when new information arrives, the measurement of the cyclical component needs to be changed for all *t*. Two, since the spectrum of y_t is undefined at the zero frequency when the series is non-stationary, a stationary transformation is required before the spectrum is computed. A preliminary step is therefore to decide whether a deterministic or a stochastic trend should be removed.

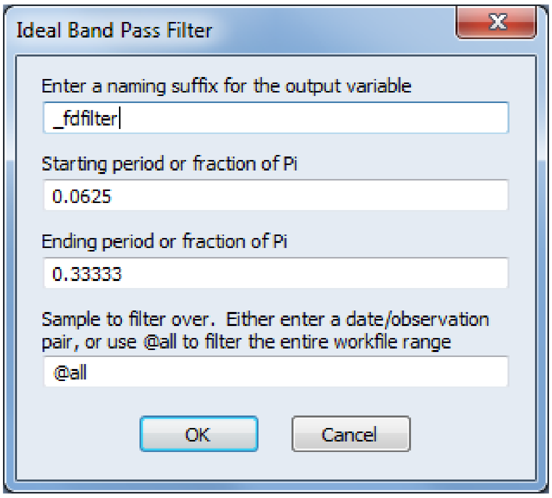
A band-pass filter in the frequency domain that does not suffer from the latter problem has been proposed by Corbae and Ouliaris (CO) (2006). In contrast to other frequency-domain (band-pass) filters, the CO filter operates on the level of a time series, even if the series has unit-root, i.e., non-stationary, components and the variance of the series is increasing in the sample size.

We can use EView's **FDFilter** add-in to estimate the business-cycle component of a time series, the output gap or the long-run trend implicit in the series. The EViews add-in is available from:

http://www.eviews.com/cgi/ai_download.cgi?ID=FDFilter.aipz

[36] Technically speaking, this is because Fourier frequencies are functions of *T*.

As before, we select `lrgdp` by double-clicking on it. If the **FDFilter** add-in is installed correctly, it should now be available in the **Add-ins** menu. We open the **Corbae-Ouliaris FD** filter item, which opens the following dialog box:



As can be seen from the screenshot above, we have to set four different items. The first thing is the name for the output series. The **FDFilter** add-in attaches a suffix to the original data series. The default setting is `_fd`, such that if the original series is called `lrgdp`, the filtered series will appear as `lrgdp_fd` in the workfile. I have decided to give the filtered series the marginally more informative name of `lrgdp_fdfilter`. The next two boxes concern the starting and ending periods, which can be either given as periods (or the number of observations) or as fractions of π. In the discussion of the band-pass filters above we used P_L and P_U to set the boundaries for the filter to pass through. In particular, 6 periods corresponds to 2/6 = 0.3333 radians, while 32 periods correspond to 2/32 = 0.0625 radians.[37] Periodicities and frequencies are obviously inversely related, such that periodicities in the range (P_L, P_U) correspond to frequencies in the range $(2\pi/P_U, 2\pi/P_L)$.

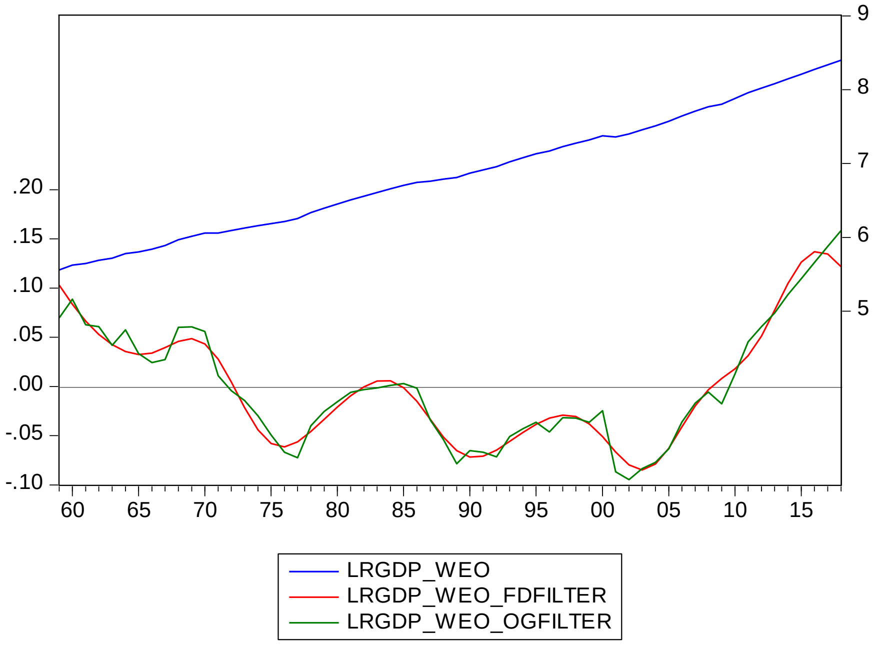
The default setting for the filter in EViews extracts the classic business-cycle frequencies between 6 (P_L) and 32 (P_U) periods. If you prefer working with observations (or periods), you can set the starting period to 6 and the ending period to 32 directly. The final dialog box allows us to set the sample range for the filter. We invoke the filter upon pressing **OK**. The result will should be a time-series object called `lrgdp_fdfilter` in the workfile that contains the business-cycle component of `lrgdp` – except for the fact that we have an insufficient number of observations to carry out the CO filter. I have therefore re-done the filter with `lrgdp_weo` which, if you remember, adds another six annual observations at the end. It appears that 54 observations are too small for the **FDFilter** add-in to work.[38] Let me stress that I have used the larger sample only to get the **FDFilter** add-in to work and not for any other reasons. In particular, I have **not** extended the sample size to deal with any end-point problems, as the CO filter is immune to them.

The EViews instructions accompanying the **FDFilter** add-in suggest that the output gap implicit in `lrgdp` can be obtained using a setting of (0.025, 1.0), which corresponds to

[37] Remember that 6 periods are deemed to be the lower bound for business-cycle frequencies, while 32 periods are deemed to be the upper bound for business-cycle frequencies.
[38] Note that 52 periods would translate into 13 years of quarterly data.

the business cycle (6, 32) plus some irregular variation (32, $+\infty$) in the time series. I have given this series the suffix `_ogfilter`. Invoking the filter as above, the results for the different procedures are given in Figure 17 below.

Figure 17: Frequency-filtered output gap estimates, 1959 – 2018



As can be seen from Figure 17 above, the output gap series in green is slightly more volatile than the business-cycle component (the red line). This is to be expected as we have allowed the output gap to include some irregular variation.

The results show a large and falling positive output gap for the first 15 years of the sample, followed by a predominantly negative output gap for the next three decades or so. The output gap appears have briefly closed in the mid-1980's. Since 2010, the output gap is positive and increasing.

4.3.6 Filters and the state-space representation

Many of the filters discussed above have an equivalent state-space representation. Morley (2002) showed this correspondence for the BN decomposition, Harvey and Jäger (1993) for the HP-filter and Harvey and Trimbur (2003) for filters in general.

Consider the BN decomposition for y_t. It should be obvious from equation (10), say, that the calculation of the exact BN trend and cycle is complicated by the presence of an infinite sum in the long-run forecast. But exact calculations become relatively straightforward if the BN forecasting model can be cast into state-space form. The definition of the BN trend in equation (10) identifies the permanent component as a pure random walk and this result can be used to link the BN decomposition with traditional UC model with random-walk trends.

Assuming a known Gaussian ARMA structure for the first difference, Δy_t, the level y_t can be thought of as made up of a permanent (unobserved) component, denoted τ_t, and a transitory (unobserved) component, denoted c_t, in the following UC representation:

$$y_t = \tau_t + c_t \tag{43}$$

$$\tau_t = \beta + \tau_{t-1} + \eta_t \quad \eta_t \sim \text{iid } N(0, \sigma_\eta^2) \tag{44}$$

$$\phi_p(L)c_t = \theta_q(L)\zeta_t \quad \zeta_t \sim \text{iid } N(0, \sigma_\zeta^2) \tag{45}$$

$$\text{Cov}(\eta_t, \zeta_t) = \rho_{\eta\zeta} \tag{46}$$

where β is the mean growth rate of y_t and $\phi_p(L)$ and $\theta_q(L)$ are lag polynomials capturing the AR(p) and MA(q) structure in the transitory component respectively.[39] According to equations (43) to (46), the permanent component follows a random walk with drift (equation (44)), the transitory component follows a stationary ARMA process with a mean of zero and the error processes associated with the permanent and transitory components are correlated. This set-up constitutes the standard treatment of trend-cycle decomposition in the state-space framework (Proietti (2006)).

The important role of the correlation coefficient, $\rho_{\eta\zeta}$, is investigated in Morley (2011), who shows that the UC representation of the model only holds if $|\rho_{\eta\zeta}| < 1$. If this is the case, the BN trend provides an optimal estimate of the permanent component under the assumption that it follows a random walk and that the unconditional expectation of c_t is zero. Using the state-space representation, the hypothesis that $|\rho_{\eta\zeta}| = 1$ is of course testable. Clark (1985) as well as Harvey and Jäger (1993) suggest specifying $p = 2$, which allows the cycle process to be periodic.

As shown in Harvey and Jäger (1993), the filter proposed by Hodrick and Prescott (1997) is equivalent to the smoothed trend obtained from a particular specification of the local linear trend model given by equations (47) to (49). As such, it may be rationalised as the optimal estimator of the trend component in the following structural time-series model:

$$y_t = \tau_t + \varepsilon_t \quad \varepsilon_t \sim \text{iid N}(0, \sigma_\varepsilon^2) \tag{47}$$

$$\tau_t = \tau_{t-1} + \beta_{t-1} \tag{48}$$

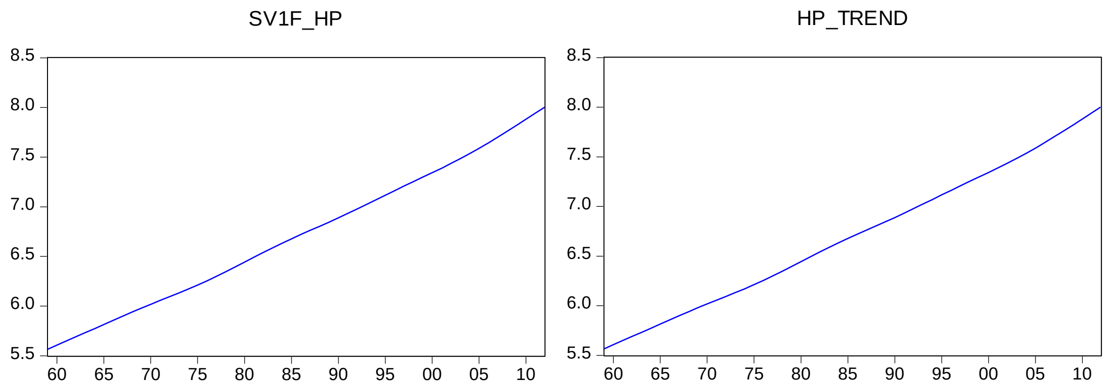
$$\beta_t = \beta_{t-1} + v_t \quad v_t \sim \text{iid N}(0, \sigma_v^2) \tag{49}$$

with the special restriction that $\sigma_v^2 = q\sigma_\varepsilon^2$, where q = 1/14400, 1/1600 and 1/100 for monthly, quarterly and annual data respectively. The HP estimate of the cyclical component is then simply given by the smoothed irregular component.

The state-space object `ss_hp_filter` contains the state-space representation of the local linear trend model corresponding to the HP filter. After estimation, we can recover the first smoothed state series from the estimated model, which I have called `sv1f_hp`. This series corresponds to the smoothed trend. Figure 18 shows that it is exactly equal to the Hodrick-Prescott trend obtained from running the Hodrick-Prescott filter procedure on `lrgdp` in EViews, which we have called `hp_trend` above.

[39] As we saw in Section 4.2.3 above, in more elaborate implementations the rate of drift, β, is itself allowed to evolve as a random walk and sometime an additional irregular term is added. The extent to which results are sensitive to changes in model specification are an empirical matter.

Figure 18: Smoothed state series from UC model versus HP-filtered trend (Sri Lankan real GDP, 1959-2012)



In this particular case, attempts to estimate the ratio q by applying maximum likelihood to equations (47) to (49) produced very small values of q, leading to the cycle effectively picking up most of the movement in the stationary part of the series. This means that unless equations (47) to (49) are a reasonable model for the series under investigation, q must be fixed in order to obtain sensible results. This is illustrated in the state-space object called `ss_hp_filter_est`, which leaves q unrestricted. The estimated state-space object is reproduced in Table 5 below.

Table 5: Estimated local linear trend model for Sri Lankan real GDP, 1959-2012

Sspace: SS_HP_FILTER_EST
Method: Maximum likelihood (Marquardt)
Sample: 1959 2012
Included observations: 54
Convergence achieved after 14 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	-8.852144	0.430657	-20.55498	0.0000
C(2)	0.552027	0.415599	1.328269	0.1841
	Final State	Root MSE	z-Statistic	Prob.
SV1	8.092943	0.018991	426.1530	0.0000
SV2	0.069582	0.018677	3.725637	0.0002
Log likelihood	111.3385	Akaike info criterion		-4.049575
Parameters	2	Schwarz criterion		-3.975909
Diffuse priors	2	Hannan-Quinn criter.		-4.021165

As can be seen from Table 5, the (insignificant) estimate of `c(2)` is equal to 0.55, meaning that the estimated q = 1/`c(2)` = 1.81.

Finally, Harvey and Trimbur (2003) propose an alternative procedure for approximating band-pass filters based on the UC model. An appealing feature of their proposal is that the end-of-sample problem is easily handled by the Kalman filter.

4.4 Economic decompositions

A third approach to extracting the cyclical component from data is guided by economic models. This methodology is based on the idea of temporary and permanent shocks. As such, economic decompositions generally define the trend as the component of the series that is driven by **permanent** shocks. The other feature that all economic decompositions have in common is that they use vector autoregressions (VARs) for what is in essence a permanent-temporary decomposition. In the case of structural VARs (SVARs), the (minimal) use of identification restrictions is meant to result in shocks with permanent or transitory features. But we should not forget that, as discussed in Canova (1998), '…all economic-based decompositions are, at best, attempts to approximate unknown features of a series and therefore subject to specification errors (p. 477)'.

4.4.1 The Blanchard and Quah (1989) (BQ) decomposition

The statistical filters discussed so far rely exclusively on the information provided in a single time series to identify the output gap. But we can also take recourse to multivariate output gap measure. The most prominent **economic** trend-cycle decomposition is based on the seminal paper on long-run identification restrictions for structural VARs (SVARs) by Blanchard and Quah (1989) (BQ). It is based on the traditional Keynesian and neoclassical synthesis, which identifies potential output with the aggregate supply capacity of the economy and cyclical fluctuations with changes in aggregate demand. The appeal of this approach lies in its compatibility with a wide range of theoretical models. In a bivariate model, structural supply and demand shocks are identified by assuming that the former have a permanent impact on output, while the latter are only allowed to have a temporary effect. In particular, two types of – uncorrelated – structural disturbances are postulated, which can affect two time series in the original paper: the log of real output and the unemployment rate. Two assumptions then identify the structural shocks:

- neither disturbance has long-run effects on the time-series used in the estimation, more specifically the first differences of the original series – in other words, growth rates are stationary; and
- disturbances to the growth rate of log real GDP may have long-run effects on the level of the series, while disturbances to the unemployment rate are restricted to not having long-run effects on the level of output

In light of the likely effects of the two structural disturbances, it seems natural to label them as supply and demand shocks.

Since the long-run is considered, the variables that enter the VAR have to be stationary.[40] Note that if some of the variables are I(1), then it is possible, if other variables are I(0), to decompose the I(1) variable into two components: a permanent and a transitory component. Thus, the Blanchard and Quah decomposition is an alternative form of conducting Beveridge and Nelson (1981) decompositions. There are some similarities between the BQ and multivariate BN decompositions. To begin with, the trend is a random walk in both cases. The difference is that while trend and cycle are driven by orthogonal

[40] If a structural break is present in the data, the appropriate transformation is to demean the data before applying the BQ decomposition.

shocks, they are driven by the same combination of shocks in the BN decomposition. This means that the disturbances in the BQ decomposition have some vague economic interpretation, which will not be the case for those of a multivariate BN decomposition.

The BQ decomposition in EViews proceeds as follows. A vector autoregression (VAR) consists of $t = 1, \ldots, T$ observations on a set of n endogenous macroeconomic variables $y_t = (y_{1t}, \ldots, y_{nt})'$, such that y_t is a $(n \times 1)$ vector containing T observations on n time series [41] We have an underlying **structural** system of equations of the form:

$$Ay_t = C(L)y_t + Bu_t \tag{50}$$

where the structural shocks u_t are normally distributed, i.e., $u_t \sim N(0, \Sigma)$, where Σ is generally assumed to be a diagonal matrix, usually the identity matrix, such that $u_t \sim N(0, \mathrm{I})$. Unfortunately, we cannot estimate this equation directly due to identification issues. Instead, we estimate an unrestricted VAR of the form:

$$y_t = A^{-1}C(L)y_t + A^{-1}Bu_t \tag{51}$$

Rearrangement of this equation yields:

$$y_t = (\mathrm{I} - A^{-1}C(L))^{-1}A^{-1}Bu_t \tag{52}$$

Equation (52) shows how the random (stochastic) shocks affect the long-run levels of the variables. If we define a matrix $M = (\mathrm{I} - A^{-1}C(L))^{-1}A^{-1}B$, the aggregate effect of a shock u is given by matrix M. Hence, if we assume that the (long-run) cumulative effect of a sub-shock u_i on a variable y_j is zero, then the column i and row j element of matrix M should be zero. For example, suppose you have a two-variable VAR where you want to restrict the long-run response of the second endogenous variable, y_{2t}, to the first structural shock, u_{1t}, to be zero, then $m_{21} = 0$:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ 0 & m_{22} \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{53}$$

In other words, knowing the values of the matrix M tells us something about matrices A and B. Due to the number of restrictions required, EViews also imposes the restriction that matrix A is the identity matrix.[42] It then uses matrix M to estimate matrix B.

Long-run restrictions in EViews can be specified either in matrix form (where the matrix M is entered) or in text form. We will consider a bivariate VAR using the growth rate of real GDP and the inflation rate. We then impose the long-run restriction that only supply shocks have permanent effects on real GDP and estimate the impulse response function associated with this identification scheme.

[41] Note that for the BQ decomposition to be meaningful, at least one of the variables needs to be non-stationary, since I(0) variables do not have a permanent component. If stationarity of the second variable is not given, we may need to apply some transformation on the second series to render it stationary.
[42] The expression for the long-run response in equation (52) involves the inverse of A. Since EViews requires all restrictions to be linear in the elements of A and B, the A matrix must be the identity matrix when you specify a long-run restriction.

We use inflation rather than the unemployment rate for several reasons. To begin with, it may be the case that unemployment data are not very informative due to high formal (un)employment and underemployment. If this should be the case, the inflation rate can be used instead of the unemployment rate as the additional macroeconomic variable to real GDP.[43] Two, in the case of Sri Lanka, unemployment rate data is only available after 1990, while inflation rate data goes back to 1959. Finally, unemployment has been falling steadily since 1990, while the inflation has been fluctuating around a mean value of 8.5 per cent since 1959.

We estimate a reduced-form VAR with `dlrgdp` and `inflation` over the period from 1959 to 2012, impose the long-run restriction that only supply shocks have permanent effects on real GDP and estimate the impulse response functions of the identification scheme. A VAR with one lag using the first difference of real GDP and inflation as well as a constant shows no lower-order autocorrelation at conventional significance levels. The VAR is stable, but the residuals are non-normal (this is left as an optional exercise). The estimation results are given in Table 6.

Table 6: VAR(1) estimation results for `dlrgdp` and `inflation`, 1959 –2012

Vector Autoregression Estimates
Sample (adjusted): 1961 2012
Included observations: 52 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	DLRGDP	INFLATION
DLRGDP(-1)	0.226542	27.51490
	(0.14030)	(38.1704)
	[1.61473]	[0.72084]
INFLATION(-1)	0.000151	0.457614
	(0.00045)	(0.12237)
	[0.33580]	[3.73948]
C	0.034341	3.618858
	(0.00746)	(2.02913)
	[4.60443]	[1.78345]
R-squared	0.057558	0.244846
Adj. R-squared	0.019091	0.214023
Sum sq. resids	0.017858	1321.841
S.E. equation	0.019090	5.193876
F-statistic	1.496298	7.943709
Log likelihood	133.6060	-157.9088
Akaike AIC	-5.023308	6.188799
Schwarz SC	-4.910736	6.301371
Mean dependent	0.046116	8.867481
S.D. dependent	0.019275	5.858504
Determinant resid covariance (dof adj.)		0.009537
Determinant resid covariance		0.008468
Log likelihood		-23.51291

[43] If administered prices account for a large percentage of the CPI basket, it is more appropriate to use a market-determined inflation measure excluding administered prices. In many cases, however, these data are not available and this avenue cannot be pursued.

Akaike information criterion	1.135112
Schwarz criterion	1.360255

Using the estimated VAR as well as the long-run restriction that supply shocks have no long-run effect on GDP, we can extract a transitory, i.e., cyclical, component of GDP.

We now have to create the long-run impact matrix, *M*:

$$\begin{pmatrix} dlrgdp_t \\ \inf lation_t \end{pmatrix} = \begin{pmatrix} m_{11} & 0 \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} demand\_shock_t \\ \sup ply\_shock_t \end{pmatrix}$$

Imposing this (long-run) restriction can be done either in text or matrix format. To impose this restriction using the text format type:

```
@lr1(@u2)=0
```

To impose the restriction using the matrix format, create the long-run matrix, which I have called `matrix_lr` in the workfile:

$$M = \begin{pmatrix} m_{11} & 0 \\ m_{21} & m_{22} \end{pmatrix} = \begin{pmatrix} NA & 0 \\ NA & NA \end{pmatrix}$$

In other words, the long-run response of the first variable (real output) to the second structural shock (a supply shock) is zero. : To impose the restriction above in text format, we select **Proc and Estimate Structural Factorisation…** from the VAR window menu. In the SVAR options dialog, select **Text** (or **Matrix** as appropriate). The output can be found in Table 7.

Table 7: SVAR(1) estimation results for `dlrgdp` and `inflation`, 1959 – 2012 (long-run restrictions)

Structural VAR Estimates
Sample (adjusted): 1961 2012
Included observations: 52 after adjustments
Estimation method: method of scoring (analytic derivatives)
Convergence achieved after 6 iterations
Structural VAR is just-identified

Model: Ae = Bu where E[uu']=I
Restriction Type: long-run text form
Long-run response pattern:

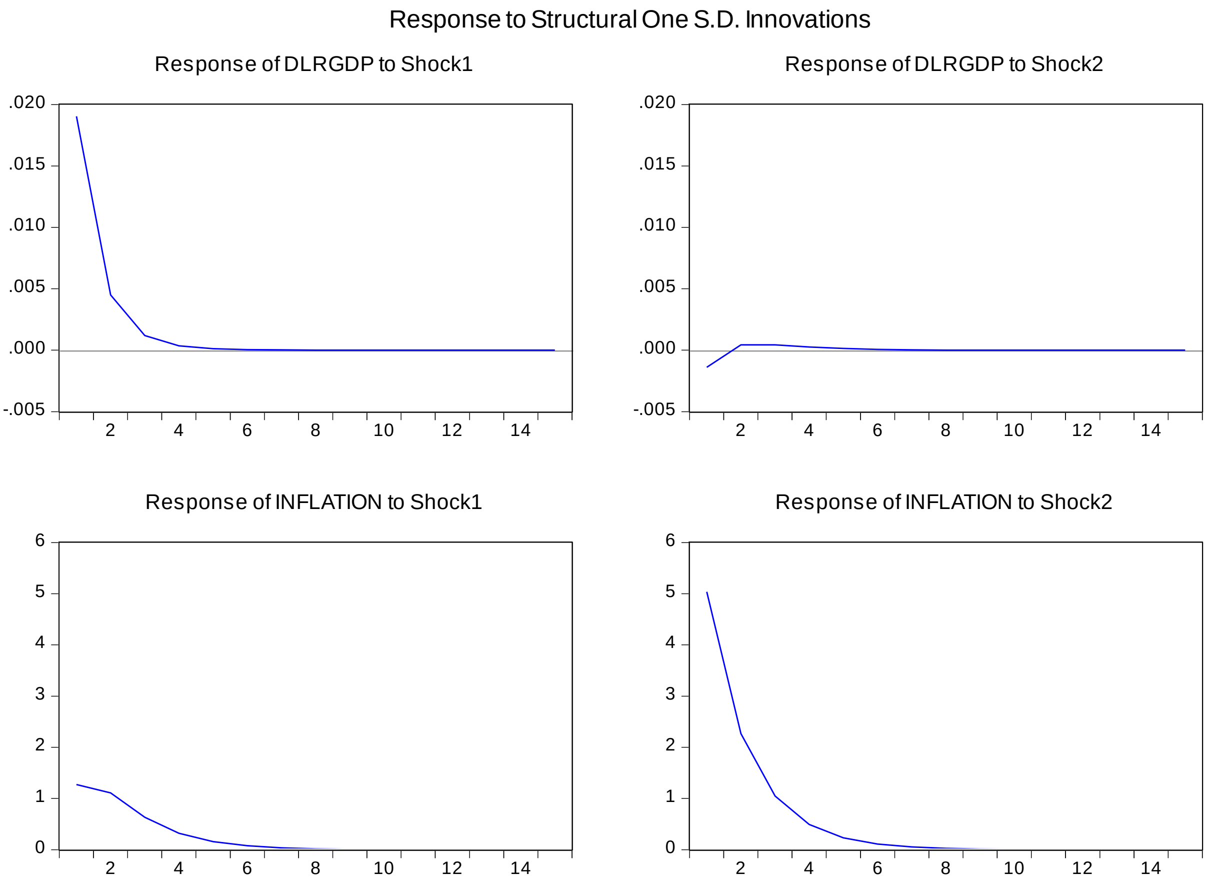
C(1)	0
C(2)	C(3)

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	0.025324	0.002483	10.19804	0.0000
C(2)	3.629467	1.335808	2.717058	0.0066
C(3)	9.284470	0.910417	10.19804	0.0000
Log likelihood	-26.60292			

Estimated A matrix:		
	1.000000	0.000000
	0.000000	1.000000
Estimated B matrix:		
	0.019039	-0.001402
	1.271787	5.035763

In terms of impulse response functions, we are now no longer considering EView's default option of a Cholesky decomposition. In order to tell EViews this, we have to go to **View/Impulse Responses...**, check the box for **Accumulated Responses** and select the **Structural Decomposition** option on the **Impulse Definition** tab. Note that EViews does not give a name to the shocks but labels them sequentially. In our case, the shocks are referred to as `Shock1` and `Shock2`. In addition, it is not possible to get EViews to calculate standard error bands around the impulse response functions automatically. The (accumulated) impulse responses associated with this identification are given in Figure 19.

Figure 19: Accumulated impulse response functions



`Shock1` is the demand shock, whereas `Shock2` is the supply shock. Thus, a demand shock increases real GDP growth (as expected?) and inflation. A supply shock leaves output unaffected and increases inflation.

From the estimated structural VAR, how can we generate the fundamental shocks, i.e., the u_t's in equation (50), using EViews? To generate the fundamental shocks, we use the equation $A\varepsilon_t = Bu_t$, where ε_t is the error or residual from the VAR regression which has been generated, and matrices A and B come from the estimated structural VAR. The fundamental shocks are then simply: $\hat{u}_t = B^{-1} A \varepsilon_t$.

We generate the fundamental (structural) shocks from the long-run SVAR by selecting **Proc** and **Make Residuals**. EViews will automatically generate series named `resid??` in the same ordering as the VAR estimate. In this case we assume that they are `resid01` and `resid02`. Change the names as appropriate – I have called them `lr_shock1` and `lr_shock2`. Create two matrices, calling them `mat_lr_a` and `mat_lr_b` by typing in the command window:

```
matrix(2,2) mat_lr_a
matrix(2,2) mat_lr_b
```

and enter the estimated coefficient values from the estimated long-run SVAR, i.e.:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

for the entries of `mat_lr_a` (remember that EViews assumes the A matrix to be the identity matrix, cf. footnote [18]) and:

$$\begin{pmatrix} 0.019039 & -0.001402 \\ 1.271787 & 5.035763 \end{pmatrix}$$

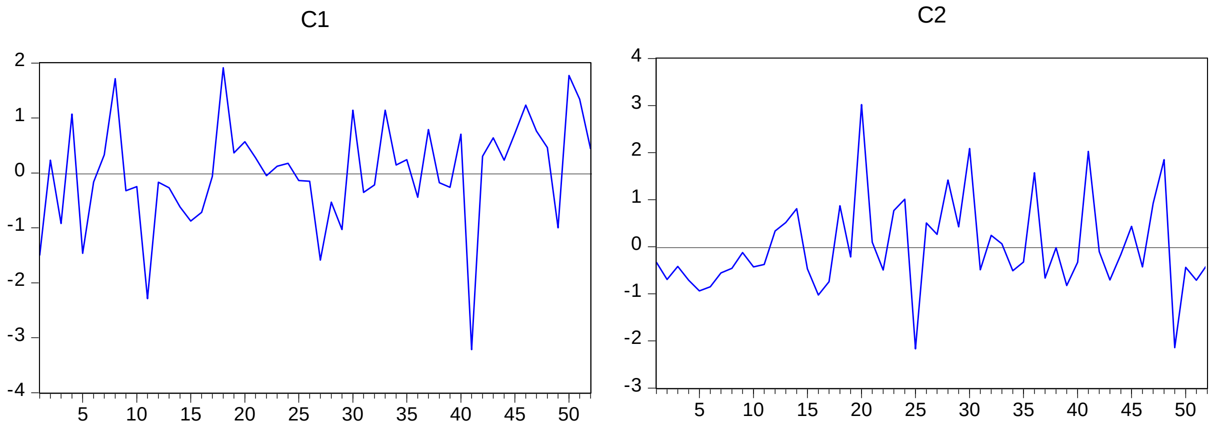
for `mat_lr_b`.

Type in the command window (or create a programme and run it):

```
group resgroup lr_shock1 lr_shock2
matrix resmatrix=@convert(resgroup)
matrix resfund=@transpose(@inverse(mat_lr_b) * mat_lr_a *
@transpose(resmatrix))
show resfund
```

Select **View, Graph, Line** to see a plot of the fundamental (structural) errors (Figure 20).

Figure 20: Fundamental shocks



Getting the underlying series out of the matrix called `resfund` is a bit convoluted. At the moment, the matrix `resfund` is made up of two columns, the first of which is the demand shock (`C1` in Figure 20) and the second of which is the supply shock (`C2` in Figure 20). Using the following five lines in the command window will generate two series, called `demand_shock` and `supply_shock` that extricate the two structural shocks from the matrix `resfund`:

```
series demand_shock
series supply_shock
group g1 demand_shock supply_shock
sample s1 1961 2012
mtos(resfund,g1,s1)
```

The derivation of the output gap requires setting a starting point at which the gap is closed. Potential output is then associated with cumulated supply shocks (`supply_shocks`), whereas the output gap reflects cyclical swings in aggregate demand.[44] Interestingly, we can use the BQ output gap to construct a corresponding measure of potential output. The log of potential output, `ln(potential GDP)`, is defined as:

$$\mathrm{ln(potential\ GDP)} = \mathrm{ln(actual\ GDP)} - (\mathrm{ln(gap/100)} + 1) \qquad (53)$$

Based on these estimates, we frequently find that potential output for developing countries is almost as volatile as the (real) GDP series itself. Indeed, Aguiar and Gopinath (2007) document that shocks to trend growth, rather than transitory fluctuations around a stable trend, are the primary source of fluctuations in emerging market economies.

A 'larger' version of the BQ decomposition that uses more information was suggested by King *et al.* (1991) and is known as the King, Plosser, Stock and Watson (KPSW) decomposition. While the procedures are quite similar, more information is used in the KPSW decomposition to estimate the trend, including cointegration restrictions and a larger

[44] We can use EView's `@cumsum(supply_shock, s)` command for that purpose, which calculates the cumulative sum of the values in `supply_shock` from the start of the workfile/sample. Writing the command as `@cumsum(supply_shock, s)` allows you to set the cumulative sum over a sample period defined by *s*, such as `sample s = first_period last_period`.

number of variables. In addition, the KPSW approach easily generalises to larger systems, while the BQ decomposition is primarily designed for bivariate models.[45] But this approach is not readily available in EViews and would require a tailor-made EViews program to make it operational.

4.4.2 The production-function approach

On the theoretical side, we can use a production function approach, although the data requirements often turn out to be unrealistic for emerging market and developing economies. The production function approach decomposes GDP using employment and capital stock data, assuming a certain technology. It is standard in the literature to assume that the economy is characterised by a Cobb-Douglas production function assuming constant returns to scale technology:

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha} \tag{54}$$

where Y_t is output, K_t and L_t are capital and labour services respectively, while A_t is the contribution of technology or total factor productivity (TFP). The output elasticities, equal to α for capital and $(1 - \alpha)$ for labour services, add up to one reflecting constant returns to scale. For simplicity, labour input is frequently defined as the number of employees in the economy.

Because capital input is not available, it is usually generated using the perpetual inventory model, given by:

$$K_t = (1 - \delta)K_{t-1} + I_t \tag{55}$$

where I_t is investment and the (annual) depreciation rate δ is usually set to either 0.05 or 0.10. The initial capital stock is then computed as $K_0 = I^*/(g + \delta)$, where I^* is benchmark investment, calculated as the average proportion of investment in total GDP, while g is the average growth rate of the economy over the sample period. Hence, based on these parameters, the initial capital stock is derived as:

$$K_0 = (\bar{I}/\bar{Y})Y_0/(g+\delta) \tag{56}$$

Since TFP is not directly observable, it is computed by inverting the technological process given by equation (54). In other words, we calculate TFP as the residual determinant of output that is not explained by labour and capital using:

$$A_t = \frac{Y_t}{K_t^{\alpha} L_t^{1-\alpha}} \tag{57}$$

or:

$$a_t = y_t - \alpha k_t - (1 - \alpha) l_t \tag{58}$$

[45] Note, however, that BQ decompositions with more than two variables exist, although it becomes more difficult to unequivocally impose the necessary long-run restrictions.

where lower-case letters indicate variables in logs. This results in a series for TFP which can be combined with the other inputs from equation (54) to decompose GDP growth. Somewhat more elaborate versions of TFP involve the following functional form:

$$\ln(Y_t) = \alpha\ln(P_t \times PR_t \times (1 - U_t) \times H_t) + (1 - \alpha)\ln(K_t) + \varepsilon_t \tag{59}$$

where Y_t is output, P_t is the working-age population, PR_t is the labour participation rate, U_t is the unemployment rate, H_t is the number of hours worked per worker, K_t is the capital stock and α is the average labour share over the sample period. TFP from equation (59) is once again derived as the equation's residual, ε_t.

Some authors prefer to purge the effects of a varying capacity utilisation rate on TFP by modifying the original production-function approach in equation (59) as:

$$\ln(Y_t) = \alpha\ln(P_t \times PR_t \times (1 - U_t) \times H_t) + (1 - \alpha)\ln(CU_t \times K_t) + \tilde{\varepsilon}_t \tag{60}$$

where CU_t is the capacity utilisation rate and $\tilde{\varepsilon}_t$ is **adjusted** TFP. It is well known that TFP fluctuations account for a considerable part of output fluctuations over the business cycle. At the same time, though, the estimated TFP series is very sensitive to changes in the intensity of capital utilisation.[46]

With these results in hand, we can now compute the output gap using the production function approach. In terms of this approach, output will be at its potential if the rates of capacity utilisation are normal, that is, labour input is consistent with the natural rate of unemployment and technological progress (as captured by total factor productivity or TFP) is at its trend level. More specifically, we use equation (54) with trends for all the variables. The final step in this approach involves removing cyclical factors from adjusted TFP and the factor inputs. This is done by using equilibrium values of the variables in equation (54), which are again most simply derived through filtering techniques. The standard HP filter is used to generate those trends assuming a smoothness parameter, λ, of 100 for annual data. The assumption of constant returns to scale is maintained and elasticities are calibrated as α = 0.5. Starting from the new production function in terms of HP-filtered series, denoted by an asterisk, we have:

$$Y_t^* = A_t^*(K_t^*)^\alpha(L_t^*)^{1-\alpha} \tag{61}$$

Potential labour input, L^*, could be the level of employment consistent with the (time-varying) natural rate of unemployment, UR^*:

$$L_t^* = LF_t(1 - UR_t^*) \tag{62}$$

where LF_t is the labour force. The natural rate of unemployment can be derived in a number of ways, for example, by HP-filtering the observed unemployment rate. While measured TFP tends to be rather volatile, in part reflecting fluctuations in the capacity utilisation rate, it usually displays a clear trending behaviour that can be readily extracted using a simple HP

[46] This results from the fact that estimated TFP relies on the total capital stock and not the flow of services actually provided by the stock at any point in time, while, in reality, capital utilisation varies considerably over the business cycle.

filter or a regression of measured TFP on a time trend. The resulting TFP trend smoothes away cyclical fluctuations in capacity utilisation and can therefore be used in the production function to estimate a series for potential output.

The output gap is then calculated as a percentage of potential output as either:

$$Gap_t = 100(Y_t - Y_t^*)/Y_t^* \tag{63}$$

or the logarithmic approximation of equation (63):

$$gap_t = 100(\ln Y_t - \ln Y_t^*) \tag{64}$$

Some shortcomings of this approach include the dependence on a number of crucial assumptions, such as the (constant) shares of capital and labour, and the function form of the production relationship, including the number of input factors and the specification of returns to scale. We should also not forget that we are still using the HP-filter for the purposes of calculating trend values of the factor inputs. Problems in obtaining trend estimates of real GDP are thus simply transferred to trend estimates of factor inputs.

4.4.3 The output gap and cointegration analysis

Gradzewicz and Kolasa (2005) present a cointegration approach to estimating the output gap for Poland. The calculation of the output gap proceeds in two ways, both involving the concept of cointegration:

- the first method is based on a two-factor production function that is estimated in a cointegrated vector error-correction mechanism (VECM) system, in which potential GDP is calculated as the product resulting from the maximum level of production inputs; while

- the second method employs a permanent-transitory decomposition of real GDP, using the long-term restrictions in the VECM that arise endogenously by cointegrating relationships – these cointegrating relationships are based both on the long-term production function and on the permanent-income hypothesis

As pointed out by the authors, cointegration analysis and the use of the Johansen methodology connect the two methods highlighted above. As such, the two approaches have a more solid grounding in economics than the statistical and hybrid decompositions discussed so far. The Johansen methodology is used to calculate a Cobb-Douglas production function, assuming that the TFP_t variables can be approximated by an exponential trend, i.e., a linear trend after taking the logarithm. This assumption was made in order to employ the standard assumptions concerning trends in the Johansen procedure. Using Polish data, Gradzewicz and Kolasa (2005) find one statistically significant cointegrating relationship between Y_t, L_t and K_t.[47] The resulting estimated long-term relationship between GDP, labour input and capital was found to be:

[47] Labour input (L_t) was assumed to be equal to the number of employed persons according to the Labour Force Survey (LFS). Capital (K_t) was assumed to be equal to the gross value of fixed assets in the domestic economy. As outlined in the main text, TFP_t was approximated by an exponential trend.

$$\hat{Y}_t = L_t^{0.57} K_t^{0.43} e^{0.29+0.01t} \qquad (65)$$

Note that this estimated relationship of the long-term production function does not suffer from the fact that trend values for labour input and capital have been calculated using the HP-filter.

An alternative approach to determining potential output, used both in economic theory and empirical research, is to treat it as the long-term GDP trend. The methodology employed by the authors employs the permanent-transitory decomposition described in Yang (1988). The dynamic system is composed of three variables: real GDP, the number of employed persons and capital. All three variables are I(1), meaning that economic theory defines a cointegrating relationship existing between these three variables as the Cobb-Douglas production function described above. The system was therefore estimated as a VECM and the restriction of constant returns to scale successfully imposed on the parameters of the cointegration relationship. Using this approach, the long-run relationship was given by:

$$\hat{Y}_t = L_t^{0.493} K_t^{0.507} e^{-0.055+0.0009t} \qquad (66)$$

One advantage of this approach is that the estimated output gap is, by design, stationary. In order to validate some of their findings, the authors also considered a permanent-transitory decomposition based on the permanent income hypothesis (PIH). In particular, the PIH imposes different economic assumptions that the Cobb-Douglas production function. In this case, the character of the cointegration restrictions is based on consumption theory. The three variables in the VECM are real GDP, consumption and the short-term real interest rate. This specification may therefore be more readily estimable if data on the capital stock or the gross value of fixed assets in the domestic economy are not available. In the system thus defined, the long-term relationship is defined as a stationary consumption-to-GDP ratio.

In summary, the alternative methods of estimating the output gap using cointegration analysis differ with regard to the concept and method of calculation. Potential GDP estimated using the production-function approach and the Johansen methodology can be regarded as reflecting the supply side of the economy. This means that the level of GDP corresponds to long-term inputs of the factors of production. On the other hand, methods based on a permanent-transitory decomposition of GDP use long-term cointegrating relationships between macroeconomic variables and yield potential GDP estimates that are a product of accumulated shocks.

4.4.4 Dynamic factor models

The final approach we consider in this and the next Section is an (approximate) dynamic factor model. These models exploit the fact that there may be a source of common fluctuations in a vector of economic time series. In particular, the approach assumes that a low-dimensional vector of unobservable variables (the so-called 'factors') drives the co-movements across variables in the vector of economic time series. These factors can be considered as an exhaustive summary of the information contained in a (large) data set, such that they may provide a better assessment of 'real' or 'nominal' conditions in the economy than any particular single variable. Another way to think about the unobserved factors is that they reflect theoretically motivated concepts such as ‘economic activity’, ‘price pressures’ or ‘credit conditions’ that cannot easily be represented by one or two series but rather are reflected in a wide range of economic variables.

The advantages from this approach is that it uses a lot more information than the other methods consider so far and that, in theory, this series should be robust to revisions in the sense that these are now due to unbiased measurement errors or idiosyncratic news (Giannone *et al.* (2008)).[48]

One of the many ways in which dynamic factor models have been used is to provide real-time estimates of the output gap. A recent application of this approach to calculating the Norwegian output gap can be found in Aastveit and Trovik (forthcoming). I will therefore highlight both the theoretical as well as the practical aspects of calculating the output gap *via* the common component series.

This approach extracts a common component in GDP from a large set of related macroeconomic indicators. Suppose we have a ‘large’ set of N informational or background (zero-mean stationary) time series, X_t, where $K + M \ll N$, which is related to the unobservable factors, F_t, and – sometimes, but not always – to some observable variables, Y_t: How large is large in this context? Boivin and Ng (2006) find that factors extracted from as few as 40 (pre-screened) series often yield satisfactory or even better results than using a much larger informational set, X_t. As argued by Stock and Watson (2009), the factors are still estimated consistently by principal components even if there is some time variation in the loading parameters.

We assume that the large set of data, X_t, can be described by an (approximate) dynamic factor model. Let:

$$X_t = \Lambda(F_t, Y_t) + e_t = \Lambda^f F_t + \Lambda^y Y_t + e_t \tag{67}$$

where the Λ^i are the factor loadings, X_t is a $(T \times N)$ panel data matrix which contains a range of macroeconomic and financial variables – the individual X_{it} need to be **stationary** and **standardised**.

As an aside, note that Stock and Watson (2002) refer to equation (68) – without the observable variables, Y_t – as a **dynamic factor model**:

$$X_t = \Lambda(F_t) + e_t = \Lambda^f F_t + e_t \tag{68}$$

The number of factors, r, is typically much smaller than the number of variables, n, thus ensuring a parsimonious model. The factors evolve over time according to the following VAR:

$$F_t = AF_{t-1} + Bu_t \tag{69}$$

where A is a $(r \times r)$ parameter matrix (with all the roots of $\det(\mathrm{I}_r - Az)$ outside the unit circle), B is a $(r \times q)$ matrix of full rank q, where q is the number of shocks in the economy, i.e., the dimension of u_t. We also assume that the common shocks, u_t, follow a white-noise process

[48] Another advantage arises in forecasting applications. Stock and Watson (1999a) and Bernanke and Boivin (2003) have demonstrated how useful factor methods are for forecasting inflation. Banerjee *et al.* (2006) show that dynamic factor models are a reasonable alternative forecasting tool in the face of short spans of reliable time series.

and that $Q = \mathrm{E}(Bu_t(Bu_t)')$. In this model, r larger than q would capture the lead and lag relationships between common factors and common shocks. Together, equations (67) and (69) define a state-space representation of an (approximate) dynamic factor model (Forni *et al.* (2005)).[49]

The number of factors in factor models is usually exogenously determined, usually on the basis of subjective criteria such as *a priori* knowledge, visual inspection of a scree plot and the proportion of variance explained. At the same time, formal information criteria to determine the optimal number of factors also exist (Bai and Ng (2002)).

Following Giannone *et al.* (2008), equations (67) and (69) are estimated in a two-step procedure. First, OLS is used to estimate parameters on the principal components from the balanced part of the data set (i.e., the data set up to the latest date for which there exist observations on all variables). These parameters and factors are used as initial values in a Kalman-filter re-estimation of the now possibly non-orthogonal factors. Moreover, the unbalanced part of the panel data set can now be incorporated through the use of the Kalman filter. Missing observations are interpreted to have an infinitely large noise-to-signal ratio. The latter is achieved by parameterising the variance of the idiosyncratic component of the missing observations to infinity at the end of the sample.

Having obtained an estimate of the factors conditional on all available information up to time t, GDP growth is estimated as a simple projection, meaning that quarterly GDP growth is regressed on the factors using OLS. This approach implicitly assumes that the common factors capture the dynamic interactions among the dependent variables as well as the dynamics in GDP.

In the next step, the estimated GDP growth series is transformed to levels. Finally, we obtain an estimate of the output gap by detrending the estimated GDP series in levels.

4.4.5 Survey data on cyclical indicators

Each of the methodologies discussed so far has relied on a specific economic statistical model of potential output and hence spare capacity in the economy. But additional evidence on the output gap is provided by **survey-based** measures of spare capacity. Broadly speaking, while these surveys frequently vary in scope and design, they all contain specific questions on whether or not companies have free capacity to expand production of their goods and services. In addition, some surveys may also include questions on whether or not companies are facing difficulty recruiting staff. The value of these surveys arises only if the survey data cover a full economic cycle, so survey time series data that only cover a relatively short time period are of limited use. It should be kept in mind, though, that the surveys capture significant parts of the domestic economy, but not necessarily all of it.

The UK's Office for Budget Responsibility (OBR) (2011a, 2011b) has added this approach to its toolkit for calculating the UK's output gap.[50] It uses a wide range of contemporaneous indicators of the cyclical position of the economy to inform an estimate of the current size of the output gap. I have been provided with annual data on nine capacity utilisation surveys in industry from 1988 to 2008, to which I have added the deviation of the annual unemployment rate from its HP-filtered trend over the same period, yielding a full set

---

[49] It is an approximate dynamic factor model since the idiosyncratic terms in equation (67) are allowed to be weakly correlated (Forni *et al.* (2000), Stock and Watson (2002)).

[50] Indeed, Goldman Sachs (2013) have recently used the same approach to calculate output gaps for the euro-area.

of ten so called cyclical indicators.[51] I then followed the methodology as outlined in Chapter 4 of OBR (2011a), which involves **principal components** analysis. In a nutshell, principal component analysis specifies a number of different linear combinations of the underlying variables (in this case the ten cyclical indicators) which are:

- uncorrelated (orthogonal); and
- contain the maximum variance

The first principal component is the linear combination which has the greatest variance, the second principal component is a different linear combination which has the second-highest variance and so on. In the example at hand, the first principal component, i.e., a particular linear combination of some underlying cyclical indicator series, is interpreted as a proxy for the output gap. This goes back to an assumption that the output gap is the most important common determinant of the cyclical indicators.

Table 8 sets out the variables used in this illustrative application of principal components analysis, along with the derived weights for each indicator variable. One advantage of principal components analysis is that it can be used to combine a variety of different types of indicators. As with any series, survey data may be subject to limitations as an indicator of the degree of spare capacity or excess demand.[52] For that reason, it is helpful to complement survey data with non-survey cyclical information. For that reason, we have added the deviation of the unemployment rate from an HP-trend estimate of the non-accelerating inflation rate of unemployment (NAIRU) to nine indicators of capacity utilisation in the Sri Lankan industrial sector.

Table 8: Principal component variable information and weights

Variable	Sample period for standardisation	Weights
Capacity utilisation in industry: food, beverage and tobacco products	1988 – 2008	0.076
Capacity utilisation in industry: textile, wearing apparel and leather	1988 – 2008	0.145
Capacity utilisation in industry: wood and wood products	1988 – 2008	0.257
Capacity utilisation in industry: paper and paper products	1988 – 2008	0.440
Capacity utilisation in industry: chemical, petroleum, rubber and plastic products	1988 – 2008	-0.440
Capacity utilisation in industry: non-metallic mineral products	1988 – 2008	-0.296
Capacity utilisation in industry: basic metal products	1988 – 2008	-0.455
Capacity utilisation in industry: fabricated metal products	1988 – 2008	0.239
Capacity utilisation in industry: manufactured products	1988 – 2008	0.448
Deviation of unemployment rate from HP-trend estimate of the NAIRU	1990 – 2008	-0.029

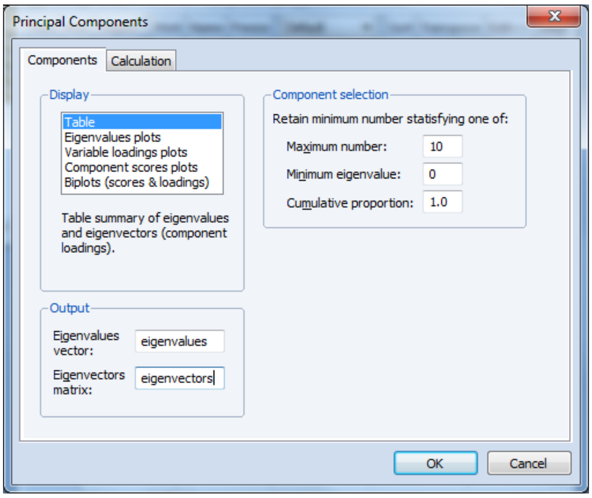
Notes: Sample period for standardisation (second column) refers to the sample period used to calculate means and standard deviations. The weights (loadings) in the third column are used to construct the first principal component. The square of the principal component weights add up to one.

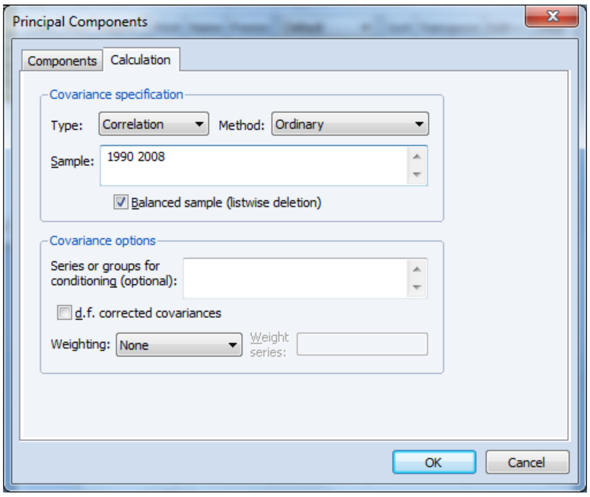
[51] Ideally, I would have also liked survey measures of recruitment difficulties and average earnings growth.

[52] Many survey measures report the balance of firms operating above or below capacity, but not the extent of spare capacity within firms. Bank of England (2011) has a discussion of some of the issues in interpreting survey measures of spare capacity.

The EViews workfile **prin_comp_sri_lanka.wf1** contains the above ten principal component indicator variables, which have been standardised to have a mean of zero and standard deviation of one over the sample period in the second column. Note that it is not always necessary to use the entire available time series. For indicators with a relatively short time span it may not be appropriate to use the entire sample if the starting point occurs during a period of either elevated or depressed activity, as this may introduce a cyclical bias in the long-term average.[53]

In the next step, we create a group (called `gall`) out of all the ten variables. In the `gall` window, we extract the principal components by going to **View/Principal Components…**, which brings up the **Principal Components** dialog box. This box has two panes. On the **Components** pane, we give meaningful names to the vector of eigenvalues and the matrix of eigenvectors such that we store them for later use. I have called them `eigenvalues` and `eigenvectors`. On the **Calculation** pane, we use the 'Correlation' type when calculating principal components. This forces EViews to use the correlation matrix to extract the principal components and implies that the data does not need to be standardised (why?). The eigenvectors are saved in a matrix called eigenvectors (which will be used later on). We set the sample period to 1990 2008 as the unemployment rate series only starts in 1990. The following two screenshots show the two panes of the **Principal Components** dialog box.





Pressing **OK** brings up the results of the principal components analysis in tabular form, the top portion of which is reproduced in Table 9 below. We can see that the first principal component explains just about 41 per cent of the variation in the entire dataset. The **Proportion** column in Table 9 shows that the marginal contribution of additional principal components decreases: the second principal component on its own explains 23 per cent of the variation in the entire dataset, the third principal component on its own explains another 15 per cent of the variation in the entire dataset and so on. We can see that the final three principal components all explain less than 1 per cent of the variation in the entire dataset.

[53] On a more technical note, the cyclical indicators should be symmetrically distributed over time.

Table 9: Principal components analysis: eigenvalues

Principal Components Analysis
Sample: 1990 2008
Included observations: 19
Computed using: Ordinary correlations
Extracting 10 of 10 possible components

Eigenvalues: (Sum = 10, Average = 1)

Number	Value	Difference	Proportion	Cumulative Value	Cumulative Proportion
1	4.102241	1.793523	0.4102	4.102241	0.4102
2	2.308717	0.806791	0.2309	6.410958	0.6411
3	1.501926	0.729520	0.1502	7.912884	0.7913
4	0.772407	0.228790	0.0772	8.685291	0.8685
5	0.543617	0.182658	0.0544	9.228908	0.9229
6	0.360959	0.126295	0.0361	9.589867	0.9590
7	0.234664	0.143761	0.0235	9.824531	0.9825
8	0.090904	0.033391	0.0091	9.915435	0.9915
9	0.057513	0.030461	0.0058	9.972948	0.9973
10	0.027052	---	0.0027	10.00000	1.0000

The bottom portion of EView's principal component output yields the weights for the first principal component (`PC 1`) that are reproduced in Table 10 above.

Table 10: Principal components analysis: eigenvectors

Eigenvectors (loadings):

Variable	PC 1
BASIC_METAL	-0.455296
CHEMICAL	-0.400295
FABRICATED	0.238581
FOOD	0.076118
MANUFACTURED	0.448432
NON_METALLIC	-0.296114
PAPER	0.439737
TEXTILE	0.144530
UNEMPLOYMENT	-0.028965
WOOD	0.256679

One attribute of principal components analysis is that it is a purely statistical technique that is not easily embedded into an economic framework. One upshot of this is that some of the factor loadings in Table 10 are negative, meaning that some cyclical indicator variables are actually subtracted from the linear combination that makes up the first principal component. The output from principal components analysis can be written as:

$$PC1_t = \sum_{i=1}^{10} \theta_i Z_{it} \tag{70}$$

where Z_{it} (i = 1, …, 10) denotes the (standardised) cyclical indicator and θ_i the respective weight (or factor loading) associated with each indicator, as given in Table 10. Equation (70) underscores the fact that the first principal component represents a linear combination of the (ten) underlying cyclical indicators. A note on the number of required cyclical indicators. In light of the analysis in Boivin and Ng (2006), who find that factors extracted from as few as 40 (pre-screened) series often yield satisfactory or even better results than using a much larger informational set, the ten cyclical indicators used in this practical guide are therefore on the small side and should only serve as an illustration of how this approach works. At the same time, the OBR, which pioneered this approach, only uses 24 cyclical indicators.

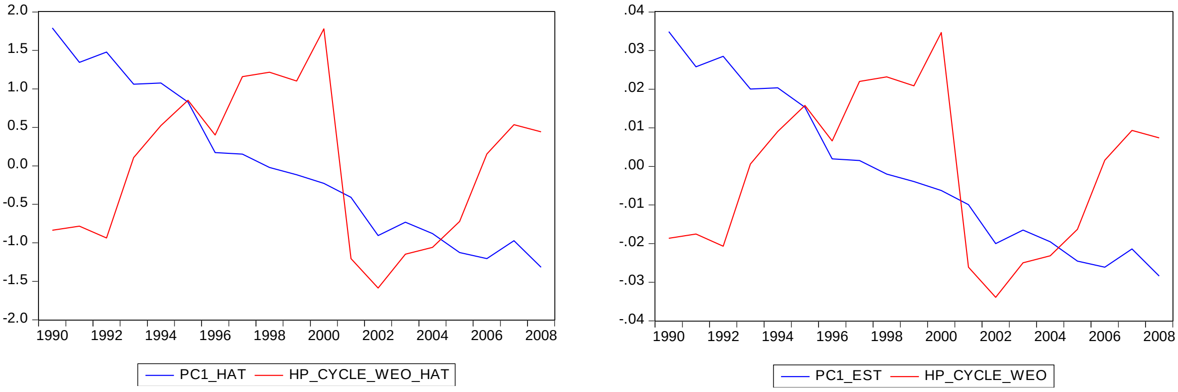
At the moment, though, the units of X_t and the output gap estimate are not the same. To make the results of the principal components analysis meaningful, the series $PC1_t$ is itself standardised using the mean and standard deviation of $PC1_t$ from 1990 to 2008. To translate the estimate to a proxy for the output gap, this standardised series is then scaled to the mean and standard deviation of the HP-filtered historical output gap series from 1990 to 2008. In particular, if μ_{HP} and σ_{HP} denote the mean (usually close to zero) and standard deviation of the HP-filtered output gap series from 1990 to 2008, the scaled version of $PC1_t$ is obtained as:

$$\hat{PC1}_t = \mu_{HP} + \sigma_{HP}\overline{PC}\hat{1}_t \qquad (71)$$

where $\overline{PC}1_t$ is the standardised version of $PC1_t$. It is not uncommon for quarterly versions of this output gap estimator to be extremely volatile, in which case (centred) moving averages can be applied to the series to adjust for its volatility.

Figure 21 shows the estimated output gap from the cyclical indicators (the blue line) as well as the corresponding HP-filtered estimate of the output gap over the same period (the red line). The left-hand panel shows the two series in standardised format, while the right-hand side panel shows the two series scaled to match the mean and standard deviation of the HP-filtered output gap estimate. On the basis of the (very small) number of cyclical indicators, as summarised in the first principal component, we would conclude that the output gap has seen a steady deterioration over the period from 1990 to 2008.

Figure 21: HP-filtered output gap and cyclical indicator output gap, 1990-2008



5 Summary and conclusions

Each of these approaches makes different assumptions regarding the nature of the trend and cycle components of the underlying series. In particular, the HP filter allows control over the smoothness of the cycle component, while the BN decomposition and approaches using the Kalman filter to do not. As such, the BN decomposition assumes that the trend is dominant and that the cycle is noisy, while Kalman filter approaches assume the opposite. Band-pass filters, on the other hand, allow control of the frequency of the cycle, in other words, the length of the business cycle.

We have estimated potential output using a variety of different approaches, none of which can lay claim to being 'the' correct method. One aim of using a battery of different methods is to obtain a range of values that we can deem reasonable.

References and further reading

Aastveit, K A and Trovik, T (forthcoming), 'Estimating the output gap in real time: a factor model approach', *Quarterly Review of Economics and Finance*.

Aguiar, M and Gopinath, G (2007), 'Emerging market business cycles: the cycle is the trend', *Journal of Political Economy*, Vol. 115, No. 1, pages 69-102.

Backus, D K and Kehoe, P J (1992), 'International evidence on the historical properties of business cycles', *American Economic Review*, Vol. 82, No. 4, pages 864-888.

Bai, J and Ng, S (2002), 'Determining the number of factors in approximate factor models', *Econometrica*, Vol. 70, No. 1, pages 191-221.

Bai, J and Ng, S (2007), 'Determining the number of primitive shocks in factor models' *Journal of Business & Economic Statistics*, Vol. 25, No. 1, pages 52-60.

Banerjee, A, Marcellino, M and Masten, I (2006), 'Forecasting macroeconomic variables for the new member states', Chapter 4 in Artis, M, Banerjee, A and Marcellino, M (eds), *The central and eastern European countries and the European Union*, Cambridge, Cambridge University Press. http://www.eui.eu/Personal/Marcellino/22.pdf.

Bank of England (2011), *Inflation report*, August. http://www.bankofengland.co.uk/publications/Documents/inflationreport/ir11aug.pdf.

Baxter, M and King, R G (1999), 'Measuring business cycles: approximate band-pass filters for economic time series', *Review of Economics and Statistics*, Vol. 81, No. 4, pages 575-93.

Benes, J and N'Diaye, P (2004), 'A multivariate filter for measuring potential output and the NAIRU: Application to the Czech Republic', *IMF Working Paper WP/04/45*. http://www.imf.org/external/pubs/ft/wp/2004/wp0445.pdf.

Bernanke, B S and Boivin, J (2003), 'Monetary policy in a data-rich environment', *Journal of Monetary Economics*, Vol. 50, No. 3, pages 525-46.

Beveridge, S and Nelson, C R (1981), 'A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle" ', *Journal of Monetary Economics*, Vol. 7, No. 2, pages 151-74.

Blanchard, O J and Fischer, S, *Lectures on macroeconomics*, Cambridge, Massachusetts, MIT Press.

Blanchard, O J and Quah, D T (1989), 'The dynamic effects of aggregated demand and supply disturbances', *American Economic Review*, Vol. 79, No. 4, pages 655-73.

Boivin, J and Ng, S (2006), 'Are more data always better for factor analysis?', *Journal of Econometrics*, Vol. 132, No. 12, pages 169-94.

Billmeier, A (2004), 'Ghostbusting: which output gap measure really matters?', *IMF Working Paper WP/04/146*. http://www.imf.org/external/pubs/ft/wp/2004/wp04146.pdf.

Burns, A F and Mitchell, W (1946), *Measuring business cycles*, New York, National Bureau of Economic Research.

Campbell, J Y and Mankiw, N G (1987), 'Permanent and transitory components in macroeconomic fluctuations', *American Economic Review*, Vol. 77, No. 2, pages 111-17.

Canova, F (1998), 'Detrending and business cycle facts', *Journal of Monetary Economics*, Vol. 41, No. 3, pages 475-512.

Canova, F (1998), 'Detrending and business cycle facts: a user's guide', *Journal of Monetary Economics*, Vol. 41, No. 3, pages 533-40.

Canova, F (1999), 'Does detrending matter for the determination of the reference cycle and selection of turning points?', Economic *Journal*, Vol. 109, No. 452, pages 126-50.

Canova, F (2007), *Methods for applied macroeconomic research*, Princeton, Princeton University Press.

Chiu, A and Wieladek, T (2012), 'Did output gap measurement improve over time?', *Bank of England External MPC Unit Discussion Paper No. 36*. http://www.bankofengland.co.uk/research/Documents/externalmpcpapers/extmpcpaper0036.pdf.

Christiano, L J and Fitzgerald, T J (2003), 'The band pass filter', *International Economic Review*, Vol. 44, No. 2, pages 435-65.

Clark, P K (1987), 'The cyclical component of US economic activity', *Quarterly Journal of Economics*, Vol. 102, No. 4, pages 797-814.

Clark, P K (1989), 'Trend reversion in real output and unemployment', *Journal of Econometrics*, Vol. 40, No. 1, pages 15-32.

Congdon, T (2008), 'Two concepts of the output gap', *World Economics*, Vol. 9, No. 1, pages 147-75.

Corbae, P D and Ouliaris, S (2006), 'Extracting cycles from nonstationary data', Chapter 6 in Corbae, P D, Durlauf, S N and Hansen, B E (eds), *Econometric theory and practice: frontiers of analysis and applied research*, Cambridge, Cambridge University Press, pages 167-177.

Cotis, J-P, Elmeskov, J and Mourougame, A (2005), 'Estimates of potential output: benefits and pitfalls from a policy perspective', Chapter 2 in Reichlin, L (ed), *Euro area business cycles: stylised facts and measurement issues*, London, CEPR. http://www.oecd.org/eco/outlook/23527966.pdf.

DeJong, D N and Dave, C (2007), *Structural macroeconometrics*, Princeton, Princeton University Press.

Domenech, R and Gomez, V (2006), 'Estimating potential output, core inflation and the NAIRU as latent variables', *Journal of Business & Economic Statistics*, Vol. 24, No. 3, pages 354-65.

Fishman, O S (1969), *Spectral methods in econometrics*, Cambridge, Massachusetts, Harvard University Press.

Forni, M, Giannoni, D, Lippi, M and Reichlin, L (2009), 'Opening the black box: structural factor models with large cross-sections', *Econometric Theory*, Vol. 25, No. 5, pages 1319-47.

Forni, M, Hallin, M, Lippi, M and Reichlin, L (2000), 'The generalised dynamic-factor model: identification and estimation', *Review of Economics and Statistics*, Vol. 82, No. 4, pages 540-54.

Forni, M, Hallin, M, Lippi, M and Reichlin, L (2005), 'The generalised dynamic factor model: one-sided estimation and forecasting', *Journal of the American Statistical Association*, Vol. 100, No. 471, pages 830-40.

Gerlach-Kristen, P (2011), 'The global output gap: measurement issues and regional disparities', *BIS Quarterly Review*, June, pages 29-37. http://www.bis.org/publ/qtrpdf/r_qt1106e.pdf.

Giannone, D, Reichlin, L and Small, D (2008), 'Nowcasting: the realtime informational content of macroeconomic data', *Journal of Monetary Economics*, Vol. 55, No. 4, pages 665-76.

Goldman Sachs (2013), 'Measuring European output gaps in real time', *European Economics Analyst Issue No: 13/31*, 12 September. www.gs.com.

Gradzewicz, M and Kolasa, M (2005), 'Estimating the output gap in the Polish economy: the VECM approach', *IFC Bulletin*, Vol. 20. http://www.bis.org/ifc/publ/ifcb20.pdf.

Hamilton, J D (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, Vol. 57, No. X, pages 357-84.

Harvey, A C (1985), 'Trends and cycles in macroeconomic time series', *Journal of Business & Economic Statistics*, Vol. 3, No. 3, pages 216-27.

Harvey, A C (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge, Cambridge University Press.

Harvey, A C (2011), 'Modelling the Phillips curve with unobserved components', *Applied Financial Economics*, Vol. 21, No. 1-2, pages 7-17.

Harvey, A C and Jäger, A (1993), 'Detrending, stylized facts and the business cycle', *Journal of Applied Econometrics*, Vol. 8, No. 3, pages 231-47.

Harvey, A C and Trimbur, T M (2003), 'General model-based filters for extracting trends and cycles in economic time series', *Review of Economics and Statistics*, Vol. 85, No. 2, pages 244-55.

Harvey, A C and Trimbur, T M (2008), 'Trend estimation and the Hodrick-Prescott filter', *Journal of the Japan Statistical Society*, Vol. 38, No. 1, pages 41-49.

Hassler, J, Lundvik, P, Persson, T and Söderlind, P (1994), 'The Swedish business cycle – stylised facts over 130 years', Chapter 1 in Bergström, V and Vredin, A (eds), *Measuring and interpreting business cycles*, Oxford, Clarendon Press.

Hillmer, S C and Tiao, G C (1982), 'An ARIMA-model-based approach to seasonal adjustment', *Journal of the American Statistical Association*, Vol. 77, No. X, pages 63-70.

Hodrick, R J and Prescott, E C (1997), 'Postwar US business cycles: an empirical investigation', *Journal of Money, Credit and Banking*, Vol. 29, No. 1, pages 1-16.

International Monetary Fund (IMF) (2010), 'United Kingdom: selected issues paper', *IMF Country Report No. 10/337*. http://www.imf.org/external/pubs/ft/scr/2010/cr10337.pdf.

Johnson, C A (2013), 'Potential output and the output gap in Central America, Panama and the Dominican Republic', *IMF Working Paper WP/13/145*. http://www.imf.org/external/pubs/ft/wp/2013/wp13145.pdf.

King, R G, Plosser, C I, Stock, J H and Watson, M W (1991), 'Stochastic trends and economic fluctuations', *American Economic Review*, Vol. 81, No. 4, pages 819-40.

Konuki, T (2008), 'Estimating potential output and the output gap in Slovakia', *IMF Working Paper WP/08/275*. http://www.imf.org/external/pubs/ft/wp/2008/wp08275.pdf.

Kuttner, K N (1994), 'Estimating potential output as a latent variable', *Journal of Business & Economic Statistics*, Vol. 12, No. 3, pages 361-8.

Laubach, T and Williams, J C (2003), 'Measuring the natural rate of interest', *Review of Economics and Statistics*, Vol. 85, No. 4, pages 1063-70.

Lucas, R E (1977), 'Understanding business cycles', *Carnegie-Rochester Conference Series on Public Policy*, No. 5, pages 7-29.

Lucas, R E (1980), 'Methods and problems in business cycle theory', *Journal of Money, Credit and Banking*, Vol. 12, No. 4, pages 696-715.

Lütkepohl, H (2007), 'General-to-specific or specific-to-general modelling? An opinion on current econometric terminology', *Journal of Econometrics*, Vol. 136, No. 1, pages 319-24.

Magud, N and Medina, L (2011), 'The Chilean output gap', *IMF Working Paper WP/11/2*. http://www.imf.org/external/pubs/ft/wp/2011/wp1102.pdf.

Mise, E, Kim, T-H and Newbold, P (2005), ' On supoptimality of the Hodrick-Prescott filter at time series endpoints', *Journal of Macroeconomics*, Vol. 27, No. 1, pages 53-67.

Mishkin, F S (2007), 'Estimating potential output', 21 May. http://www.federalreserve.gov/newsevents/speech/mishkin20070524a.htm.

Morley, J C (2002), 'A state-space approach to calculating the Beveridge-Nelson decomposition', *Economics Letters*, Vol. 75, No. 1, pages 123-27.

Morley, J C (2011), 'The two interpretations of the Beveridge-Nelson decomposition', *Macroeconomic Dynamics*, Vol. 15, No. 3, pages 419-39.

Morley, J C, Nelson, C R and Zivot, E (2003), 'Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?', *Review of Economics and Statistics*, Vol. 85, No. 2, pages 235-43.

Murray, C J (2003), 'Cyclical properties of Baxter-King filtered time series', *Review of Economics and Statistics*, Vol. 85, No. 2, pages 472-76.

Nelson, C and Plosser, C I (1982), 'Trends and random walks in macroeconomic series', *Journal of Monetary Economics*, Vol. 10, No. X, pages 139-62.

Office for Budget Responsibility (2011a), 'Estimating the output gap', *Briefing paper No. 2*. http://budgetresponsibility.org.uk/pubs/briefing+paper+No2+FINAL.pdf.

Office for Budget Responsibility (2011b), 'Estimating the UK's historical output gap', *Working paper No. 1*. http://budgetresponsibility.org.uk/pubs/WorkingPaperNo1-Estimating-the-UKs-historical-output-gap.pdf.

Oh, K H, Zivot, E and Creal, D (2008), 'The relationship between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics', *Journal of Econometrics*, Vol. 148, No. 2, pages 207-19.

Okun, A M (1962), 'Potential GNP: its measurement and significance', *Cowles Commission/Foundation Paper 190*. http://cowles.econ.yale.edu/P/cp/p01b/p0190.pdf.

Orphanides, A (2002), 'Monetary policy rules and the great inflation'. *American Economic Review*, Vol. 92, No. 2, pages 115-20.

Orphanides, A and van Norden, S (2002), 'The unreliability of output-gap estimates in real time', *Review of Economics and Statistics*, Vol. 84, No. 4, pages 569-83.

Osborn, D R (1995), 'Moving average detrending and the analysis of business cycles', *Oxford Bulletin of Economics and Statistics*, Vol. 57, No. 4, pages 547-558.

Park, G (1996), 'The role of detrending methods in a model of real business cycles', *Journal of Macroeconomics*, Vol. 18, No. 3, pages 479-501.

Pedersen, T M (2001), 'The Hodrick-Prescott filter, the Slutzky effect and the distortionary effect of filters', *Journal of Economic Dynamics and Control*, Vol. 25, No. 8, pages 1081-101.

Perron, P (1989), 'The great crash, the oil shock and the unit root hypothesis', *Econometrica*, Vol. 57, No. 6, pages 1361-1401.

Planas, C and Rossi, A (2004), 'Can inflation data improve the real-time reliability of output gap estimates?', *Journal of Applied Econometrics*, Vol. 19, No. 1, pages 121-33.

Planas, C, Rossi, A and Fiorentini, G (2008), 'Bayesian analysis of the output gap', *Journal of Business & Economic Statistics*, Vol. 26, No. 1, pages 18-32.

Proietti, T (2006), 'Temporal disaggregation by state-space methods: dynamic regression methods revisited', *Econometrics Journal*, Vol. 9, No. 3, pages 357-72.

Ravn, M O and Uhlig, H (2002), 'On adjusting the Hodrick-Prescott filter for the frequency of observations', *Review of Economics and Statistics*, Vol. 84, No. 2, pages 371-75.

Quah, D T (1992), 'The relative importance of permanent and transitory components: identification and some theoretical bounds', *Econometrica*, Vol. 60, No. 1, pages 107-18.

Ross, K and Ubide, A (2001), 'Mind the gap: what is the best measure of slack in the euro area?', *IMF Working Paper WP/01/203*. http://www.imf.org/external/pubs/ft/wp/2001/wp01203.pdf.

Stock, J H and Watson, M W (1988), 'Variable trends in economic time series', *Journal of Economic Perspectives*, Vol. 2, No. 3, pages 147-74.

Stock, J H and Watson, M W (1999a), 'Forecasting inflation', *Journal of Monetary Economics*, Vol. 44, No. 2, pages 293-335.

Stock, J H and Watson, M W (1999b), 'Business cycle fluctuations in US macroeconomic time series', Chapter 1 in Taylor, J B and Woodford, M (eds), *Handbook of macroeconomics, volume I, part A*, Amsterdam, North-Holland, pages 3-64.

Stock, J H and Watson, M W (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association*, Vol. 97, No. 460, pages 1167-79.

Stock, J H and Watson, M W (2009), 'Forecasting in dynamic factor models subject to structural instability', Chapter 7 in Castle, J L and Shephard, N (eds), *The methodology and practice of econometrics: a festschrift in honour of David F Hendry*, Oxford University Press, Oxford, pages 173-205.

Stock, J H and Watson, M W (2010), 'Dynamic factor models', Chapter 2 in Clements, M P and Hendry, D F (eds), *The Oxford handbook of economic forecasting*, Oxford University Press, Oxford, pages 35-60.

Watson, M W (1986), 'Univariate detrending methods with stochastic trends', *Journal of Monetary Economics*, Vol. 18, No. 1, pages 49-75.

Watson, M W (2007), 'How accurate are real-time estimates of output trends and gaps?', *Federal Reserve Bank of Richmond Economic Quarterly*, Vol. 93, No. 2, pages 143-61. https://www.richmondfed.org/publications/research/economic_quarterly/2007/spring/pdf/watson.pdf.

Yang, M (1998), 'On identifying permanent and transitory shocks in VAR models', *Economics Letters*, Vol. 58, No. 2, pages 171-75.

Zivot, E (2005), 'Trend-cycle decompositions', http://faculty.washington.edu/ezivot/econ584/notes/trendcycle.pdf.

