I am grateful to many colleagues at the Reserve Bank for helpful comments. I am particularly indebted to Luci Ellis and Geoffrey Shuetrim for letting me use, and providing me with excellent explanations of, their respective programs for the multivariate HP filter and the state-space model estimated using a Kalman filter. All errors, however, are mine. The views in this paper are my own and do not necessarily reflect those of the Reserve Bank of Australia.
Abstract

The output gap, defined as actual less potential output, is an important variable in its own right and as an indicator of incipient changes in inflation. This paper reviews five methods of estimating it for Australian GDP data, including linear time trends, Hodrick-Prescott (HP) filter trends, multivariate HP filter trends, unobservable components models and a production function model. Estimates of the gap vary with the method used and are sensitive to changes in model specification and sample period. While gap estimates at any particular point in time are imprecise, the broad profile of the gap is similar across the range of methods examined. Inflation equations are substantially improved when any measure of the gap is included, and output gaps generally explain innovations in inflation better than output growth.

JEL Classification Numbers: E32, O56
Keywords: output gap, potential output
Table of Contents

1. Introduction 1

2. Methods of Estimating the Output Gap 2
   2.1 Linear Method 4
   2.2 Hodrick-Prescott Method 6
   2.3 Multivariate Hodrick-Prescott Method 9
   2.4 Unobservable Components Method 15
   2.5 Production Function Method 19

3. An Assessment of Estimates of the Output Gap 21

4. Conclusion 26

Appendix A: The Equations in the State-Space Model 28

Appendix B: Data Description and Sources 29

References 30
ESTIMATING OUTPUT GAPS

Gordon de Brouwer

1. Introduction

The output gap, defined as actual less potential output, is an important variable in economic and policy systems. It is important in its own right, since there are fewer jobs and lower profits than would otherwise be the case when actual output is below potential. Moreover, the output gap has a well-documented role in the theoretical and applied literature in explaining price and wage inflation. Since potential output is not directly observable, however, neither is the output gap. This paper reviews various methods of estimating output gaps with the aim of showing some of the pitfalls and complexities of the different techniques, rather than providing a definitive measure of the output gap. Indeed, the analysis in this paper underscores that estimating output gaps is fraught with uncertainty and requires considerable judgment.

Section 2 traces through some of the techniques used to estimate output gaps. It starts by setting out some definitions and reviewing the output data. The methods used to estimate the output gap include linear time trends, Hodrick-Prescott (HP) filter trends, multivariate HP filter trends, unobservable components models and a production function model. Some of these methods are univariate techniques, using particular assumptions about the time-series properties of output to identify potential output. Other methods combine these techniques with economic information about the output gap. Estimates of the output gap are shown to be sensitive to the general model used, the particular specification adopted for the model being used, and the sample period.

Section 3 assesses these output gap measures. Output gaps are estimated for two purposes. The first is to provide information about excess capacity in the economy at a particular point in time. From the perspective of monetary policy, the output gap over the forecast horizon is of most interest. The second purpose is to use a time series of the output gap in modelling exercises. For example, given that excess demand pressures are a key cause of rising inflation, the output gap can be included
in price or wage inflation equations to obtain a more precisely estimated equation and more accurate forecasts.

The analysis in Section 2 shows that, on the one hand, estimates of the output gap are imprecise and can vary considerably across estimation methods at particular points in time. On the other hand, however, there is considerable similarity in the broad time profile of the various gap estimates, suggesting that the gap relative to its past contains useful information. While it is difficult to identify the absolute size of the output gap, it is possible to infer its relative size. This underscores the need for careful analysis and practical and considered judgment. Moreover, the similarity of time profiles suggests that most gap measures may have similar explanatory power in econometric models. This is shown by comparing the estimates in a standard mark-up model of inflation for various gap measures. Including any of the gap measures canvassed in Section 2 in an inflation equation improves the explanatory power of the equation and substantially reduces prediction error. Output gap measures also tend to explain inflation in estimating equations better than output growth rates do.

2. Methods of Estimating the Output Gap

The output gap is defined as the difference between actual and potential output:

$$gap_t = y_t - y^T_t$$

where $gap$ is the output gap, $y$ is output and $y^T$ is potential output. In this form, a positive number for the gap indicates excess demand and a negative number indicates excess capacity. The output gap represents transitory movements from potential output. Given that potential output is not observed, it has to be estimated. Not surprisingly, different assumptions about potential output and different estimation methodologies yield different estimates of the output gap.

The role of the output gap in affecting wage inflation was shown by Phillips (1958). In empirical work, inflation is often characterised as a mark-up over unit labour costs and imported goods prices, with the mark-up varying over the business cycle (de Brouwer and Ericsson 1995). Wages growth, in turn, also appears to be
sensitive to the state of the business cycle and the rate of economic growth (Cockerell and Russell 1995; de Brouwer and O’Regan 1997; Debelle and Vickery 1997). This indicates that the output gap contains valuable information about movements in price and wage inflation. From a policy perspective, however, the underlying trend or potential output component should be defined in terms of a non-accelerating (or decelerating) inflation rate. While several of the simple measures examined do not do this, the multivariate and production function approaches can provide an estimate of the output gap consistent with unchanging inflation.

In this paper, output is defined as GDP(A), the broadest available measure, and is in 1989/90 prices. Figure 1 shows the log level and one- and four-quarter-ended growth rates of output from March 1980 to December 1997. This sample period
excludes the growth slowdown that followed the oil price shocks in the 1970s, and so eliminates an important structural break. It indicates that the economy contracted in 1982/83 and in 1990/91, with the former recession being deeper but the latter more protracted. These two periods would be expected to be associated with negative output gaps with similar features. We now review various methods of estimating output gaps.

2.1 Linear Method

The simplest way to estimate the output gap is to calculate potential output using a linear trend, which, when estimated using logged quarterly GDP(A) from 1980 to 1997, yields the following equation:

\[ y_t = 1046 + 0.0077trend \]

(0.01) (0.0001)

\[ R^2 = 0.98 \]

This estimates trend growth in output over the past couple of decades to be about 3.1 per cent a year. The output gap implied from this trend estimate is shown in the left-hand panel of Figure 2. The gap profile will be discussed in Section 3. A general criticism of this method (and one which also applies to other techniques) is that the estimate of the gap depends on the sample period. For example, the bars in the right-hand panel of Figure 2 show the estimated output gap for the last period in the sample – the December quarter 1997 – as the starting date for the estimation moves forward in time from March 1975 to December 1984. Output is estimated to be above potential at December 1997 for all the sample lengths examined, but the spread in the range of estimates is one per cent. The selection of starting and end-points in a linear regression matters. For example, when the sample starts at the lowest point in a recession, which is 1982 here, the slope of the straight line fitting the series is steeper, making the gap between actual and potential output at the end of the sample smaller. This highlights the importance of starting the regression at a period when the economy is basically in balance. The sensitivity of gap estimates to the sample period also underscores the general uncertainty of gap estimates.
The assumption that potential output grows at a constant rate is also difficult to accept. Output growth can be decomposed into growth of labour productivity and of labour inputs, which in turn can be decomposed into changes in population, labour force participation and average hours worked. There is no compelling reason for these components to be constant over time, especially when an economy has undergone considerable structural reform, and so a more general, time-varying, approach is preferable.

This is also reflected in the time series properties of the gap estimate. If output follows a deterministic trend, then the residual from removing that trend should be a stationary series. But if output is an integrated series of order one, and hence follows a stochastic trend, then the residual from removing a linear trend is still non-stationary. This would violate the usual assumption that the output gap is a mean-reverting variable, such that ‘shocks’ to it do not persist. There is a large and mixed literature on whether output follows a deterministic trend, possibly with breaks, or a stochastic trend (Diebold and Senhadji 1996). Table 1 presents some relevant statistics for GDP(A) from 1980 to 1997. Rows 1 and 2 indicate that output is better characterised as following a stochastic trend, although the test has low power. Rows 3 and 4 indicate that the output gap estimated using a linear trend is
non-stationary, similar to the results of Hodrick and Prescott (1997). This suggests that a different detrending procedure is preferred.

<table>
<thead>
<tr>
<th>Table 1: ADF Test for Unit Roots in Output and the Output Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
</tr>
<tr>
<td>GDP(A)</td>
</tr>
<tr>
<td>Δ GDP(A)</td>
</tr>
<tr>
<td>Linear trend gap</td>
</tr>
<tr>
<td>Δ linear trend gap</td>
</tr>
</tbody>
</table>

Notes: T-ratio in parentheses. Estimated with one lag of the dependent variable on the right-hand side. The critical t-ratios for the constant, trend and lagged level variable are contained in Fuller (1976), with * and ** indicating significance at the conventional 5% and 1% levels.

2.2 Hodrick-Prescott Method

One such detrending procedure is that suggested by Hodrick and Prescott (1997) which contains the linear trend as a special case.\(^1\) The Hodrick-Prescott (HP) filter sets the potential component of output to minimise the loss function, \(L\),

\[
L = \sum_{t=1}^{S} \left( y_t - y_t^T \right)^2 + \lambda \sum_{t=2}^{S-1} \left( \Delta y_{t+1}^T - \Delta y_t^T \right)^2
\]  

(3)

where \(\lambda\) is the smoothing weight on potential output growth and \(S\) is the sample size. Changing this weight affects how responsive potential output is to movements in actual output. As the smoothing factor approaches infinity, the loss function is minimised by penalising changes in potential growth, which is done by making potential output growth constant (i.e. a linear trend growth rate). As the smoothing factor approaches zero, the loss function is minimised by eliminating the difference between actual and potential output, which is done by making potential output equal to actual output. These features are captured in the left-hand panel in Figure 3, which shows estimates of the output gap for three smoothing factors, 6000, 1600 and 100, with a lower smoothing factor producing a ‘smaller’ estimate of the gap.

\(^1\) The working paper version was written in 1980 but was not published until 1997.
The advantage of the HP filter is that it renders the output gap stationary over a wide range of smoothing values (Hodrick and Prescott 1997) and it allows the trend to change over time. But it also has the distinct disadvantage that the selection of the smoothing weight is arbitrary, and that this matters to the estimate. For example, consider the estimate of the gap at December 1997 on the left-hand panel of Figure 3. For high smoothing, the estimate indicates output is above potential, but for moderate or low smoothing, the estimate suggests output is below potential. Indeed, for estimates using $\lambda$ in the range 100 to 2,000, the gap is negative, but for values over this, the gap is positive. This method is not useful for identifying the absolute value of the output gap at a particular point in time.

Not only does the size of the gap vary with the smoothing weight, however, but so too do the relative scale and timing of peaks and troughs in output. For example, the left-side panel of Figure 3 shows that relatively high smoothing weights position output well above potential in 1989 compared with 1985, but a low smoothing weight reverses this relative scale. Similarly, a low smoothing weight marks 1991 as the time when output was most below potential while the higher smoothing weights mark 1992 as that point. In other words, information about the cycle changes with the smoothing weight.
If the smoothing parameter matters to the outcome, there has to be a clear method to select the smoothing parameter if the approach is to be useful. The smoothing parameter is conventionally set to 1600 in empirical work. The ‘appropriate’ value depends on the relative size of the variances of the shocks to permanent and transitory components to output (Hodrick and Prescott 1997). Hodrick and Prescott (1997) chose the value of 1600 for US GNP on the basis of their assessment of the relative size of shocks to that series. Guay and St Amant (1996) present Monte Carlo evidence, however, that a smoothing parameter of 1600 is only appropriate under implausible joint assumptions about the relative importance of demand and supply shocks and about the persistence of cycles. There are other doubts about the accuracy of the HP filter in decomposing a time series into trend and cycle. For example, there is evidence that the accuracy of the decomposition varies over different data-generating processes and different data sets (King and Rebelo 1993; Harvey and Jaeger 1993).

A problem common to most estimates of potential output is that they can change as new data observations come to hand (which is separate to the issue of the estimate changing as data are revised). This also happens with the HP filter since it contains leads and lags of output in the loss function. It is only a symmetric two-sided filter at the middle of the sample, with end-points having more leverage. Using a sample of 128 observations, for example, St Amant and van Norden (1997) show that observations at the centre of the sample receive at most a six per cent weight while the last observation accounts for 20 per cent of the weight. The end-point problem implies that estimates of the gap at the end of the sample may be subject to substantial revision as new data come to hand, the period which is of most interest to policy-makers. This is illustrated in the right-side panel of Figure 3, where the sample is progressively extended using consensus forecasts of GDP for 1998 and 1999, with output growth projected to decelerate to about 3 per cent by the end of 1998 and then pick up to 3½ per cent by the end of 1999. As the sample period is extended using this information, end-points are revised substantially, although (at least for the forecasts here) not more than when the smoothing parameter is changed. The size of the revision also depends on the cycle: the revision to the gap is larger at turning points than when the economy is at potential.

---

2 The Consensus forecasts at March 1998 for year-on-year growth for the eight quarters in 1998 and 1999 are 4.1, 3.1, 2.8, 3.1, 3.1, 3.3, 3.4 and 3.5 per cent.
2.3 Multivariate Hodrick-Prescott Method

Linear detrending and the HP filter are statistical tools which do not use economic or structural information. There are economic relationships (the Phillips curve and Okun’s Law) and economic indicators (capacity utilisation), however, which also contain information about the supply side of the economy and the stage of the business cycle. Accordingly, Laxton and Tetlow (1992) proposed an extension to the HP filter which incorporates such economic information. Potential output, for example, can be defined as the series which minimises the loss function:

\[
L = \sum_{t=1}^{S} \left( y_t - y_t^T \right)^2 + \lambda \sum_{t=2}^{S-1} \left( \Delta y_{t-1}^T - \Delta y_t^T \right)^2 + \sum_{t=1}^{S} \mu_t \epsilon_{\pi,t}^2 + \sum_{t=1}^{S} \beta_u \epsilon_{u,t}^2 + \sum_{t=1}^{S} \psi_{cu,t}^2
\]

where \( \epsilon \) is a residual from a regression, the subscripts \( \pi, u \) and \( cu \) indicate a Phillips curve equation, Okun’s Law equation and capacity utilisation equation respectively, and \( \mu, \beta \) and \( \psi \) are possibly time-varying weights. The residuals are from the following equations:

\[
\pi_t = \pi_t^T + A(L)(y_t - y_t^T) + \epsilon_{\pi,t} \quad (5)
\]

\[
u_t = nairu_t - B(L)(y_t - y_t^T) + \epsilon_{u,t} \quad (6)
\]

\[
u_t = cu_t^T + C(L)(y_t - y_t^T) + \epsilon_{cu,t} \quad (7)
\]

Equation (5) is a Phillips curve, stating that inflation will be above expectations when output is above the (non-accelerating inflation) level of potential output. Equation (6) draws on Okun’s Law, with the unemployment rate below the non-accelerating inflation rate of unemployment (NAIRU) when output is above potential. Equation (7) draws on a partial indicator of supply capacity, stating that capacity utilisation is above trend when output is above potential (Conway and Hunt 1997). These relationships, the data for which are shown in Figure 4, contain information about the output gap. The multivariate filter in Equation (4) sets potential output to minimise a weighted average of deviations of output from potential, changes in the potential rate of growth and errors in the three conditioning structural relationships. By conditioning the HP filter estimate of the output gap on
this information, a more precise estimate of potential output, and hence the output gap, should be obtained.

Figure 4: Conditioning Information for the Multivariate HP Filter

Note: Recessions, defined as two or more quarters of successive contraction in output, are shown as the shaded area.

To proceed with this technique, Equations (5), (6) and (7) need to be specified. To obtain initial values for these equations, potential output is determined by the univariate HP filter value for $\lambda = 1600$. To specify the Phillips curve in Equation (5), it is necessary to specify the expectations formation process. As a first
approximation, it is assumed that expected inflation depends on lags of past consumer and imported price inflation (Debelle and Stevens 1995):

\[
\pi_t = 0.05 \pi_{t-1} + 0.45 \pi_{t-2} + 0.04 \pi^*_t + 0.07(y_t - y^*_t) \\
\text{RSS} = 0.00058
\]

where \(\pi\) and \(\pi^*\) are quarterly underlying inflation and import price inflation, the coefficients on the various inflation terms sum to one (a restriction which is not rejected by the data), RSS is the sum of squared residuals, and standard errors are shown in parentheses. In terms of using this equation to help identify potential output, this result suggests that current inflation contains information about the current gap.

To estimate Equation (6), the NAIRU is assumed to be the series estimated by Debelle and Vickery (1997), shown in the middle panel of Figure 4. Equation (6) is:

\[
(u_t - \text{nairu}_t) = -0.66(y_t - y^*_t) \\
\text{RSS} = 0.04
\]

As expected, unemployment falls when current demand in the economy is strong relative to potential.

The capacity utilisation equation is estimated using the ACCI-Westpac survey measure of capacity utilisation for manufacturing firms, shown in the bottom panel of Figure 4. The profile of capacity utilisation matches the profile of the recessions. There are a number of problems associated with using survey data as an indication of supply and demand imbalances, including whether firms clearly distinguish between labour and capital constraints, whether firms have a uniform definition of capacity and whether what they define to be ‘normal’ varies over the cycle. Moreover, when surveys are restricted to a particular sector, as is the case with the ACCI-Westpac measure, it is unclear how representative their responses are of the

---

3 This survey is reported as a net balance of respondents who think that they are working above their normal capacity.
broader economy. For example, spare capacity widened during 1995 due to particular problems in the manufacturing sector.

The estimated equation for capacity utilisation is:

\[
u_t = -0.028 + 9.93(y_t - y^p_t)\]

(0.02) (1.26)

\[RSS = 195\]  

(10)

This suggests that current capacity utilisation contains information about the current state of the output gap.

Using these specifications for the Phillips curve, Okun’s Law and capacity utilisation equations, the loss function in Equation (4) can be minimised to solve for potential output. The following iterative procedure is used to estimate potential output. As a first step, the conditioning equations are estimated using the univariate HP filter – Equations (8), (9) and (10) above – and the residuals are used to minimise the loss function in Equation (4) and solve for potential output. The output gap is computed and the coefficient on the output gap in the structural equations is re-estimated.4 The loss function is minimised with these new residuals, and the coefficient on the output gap in the structural equations is recomputed. This procedure, similar to that outlined in Conway and Hunt (1997), is continued until the changes in the output gap coefficients satisfy a convergence criterion.

The issue of the weighting of the various components in the loss function also needs to be resolved. As is apparent from comparing the sum of squared residuals (RSS) for the three conditioning structural equations above, the residuals from the capacity utilisation equation are considerably larger than for the Phillips curve and Okun’s Law equations. If the residuals are assigned equal weights in the loss function, then the loss function is essentially minimised by setting the gap to minimise the residual in the capacity utilisation equation (and, indeed, including a gap term calculated this way substantially reduces the residual sum of squares in the capacity utilisation

4 To make the estimation process simpler, the procedure maintains the original slope coefficients for all variables in the conditioning equations apart from the output gap. This assumes orthogonality between regressors. The equations were re-estimated with the final multivariate HP gap measure with only minimal change in the slope coefficients on the non-gap terms.
equation). Similarly, the sum of squared deviations of output from trend is larger than for the Phillips curve and Okun’s Law equations, which implies that if the weight on the deviations of output from potential are also not weighted, the loss function puts all the weight on the univariate HP filter components, making the multivariate and univariate gap estimates almost identical.

Accordingly, at each stage in the iteration, the sum of squared residuals for each of the conditioning equations is weighted by its size relative to the sum of squared residuals of output less potential. For example, the weight on the inflation equation residuals, \( \mu \), is equal to \( \frac{\sum_{t=1}^{S} (y_t - y_t^T)^2}{\sum_{t=1}^{S} e_{\pi,t}^2} \) from the previous iteration (or, in the case of the first iteration, the starting values). The larger is the sum of squared residuals in the conditioning equation relative to the sum of squared output gaps, the smaller is its weight in the loss function. The weight changes with each iteration. The top panel of Figure 5 shows the gap estimated after convergence (which takes about 250 iterations).

While the profiles of the gap estimated by the univariate and multivariate HP filters are very similar, such that they tend to move in the same direction, the size of the gap is very different for each of the filters. When the estimate of potential output is conditioned on additional economic information, the output gap is typically negative, indicating that actual has been below potential output for most of the past two decades except for the early and late 1980s.

This is explained by the unemployment equation, which, as shown in Figure 4, indicates that the unemployment rate has been above the estimated NAIRU for most of this period. The persistent wedge between the unemployment rate and the estimate of the NAIRU used here indicates that potential output has been higher than actual output.\(^5\) The bottom panel of Figure 5 shows the estimate of the gap

---

\(^5\) The unemployment equation embodies the assumption that output is at potential when expected inflation equals actual inflation, rather than the assumption that output is at potential when inflation is steady. The measures of expected inflation used by Debelle and Vickery to estimate the NAIRU have been systematically above actual inflation, and this gives rise to the wedge between the unemployment rate and the NAIRU. Only when actual and expected
when the unemployment-output relation is excluded from the procedure. Since there is not a persistent wedge in the Phillips curve and capacity utilisation equation in the way that they have been estimated, potential output in this case is not estimated to have been systematically above actual output. The restricted set of conditioning information indicates that output was further below potential than indicated by the univariate HP filter in the early 1990s and in 1997, consistent with the weakness in capacity utilisation and declining inflation at these times.

The upshot of this experiment is that judgment matters. The estimate of the gap depends on the conditioning relationships, how those relationships are defined, and how the information is weighted in the loss function. When the output gap is conditioned on information which contains a constant wedge between the inflation are equal, is the unemployment rate at the NAIRU. Other measures of expected inflation may yield different estimates of the NAIRU, and so this number is illustrative only.
unemployment rate and the NAIRU, there is a substantial negative output gap. Once this wedge is removed, the gap narrows considerably. The way the conditioning relationships are modelled is of critical importance to the estimate of the output gap. Moreover, instead of assuming that the weights for the conditioning variables are constant over the sample period, the modeller can change the weights to reflect his or her assessment of the relative importance of the conditioning information at different times. Similarly, estimates of the gap towards the end of the sample will change when forecasts are included.

2.4 Unobservable Components Method

Given that potential output and the gap are unknown, it makes sense to use statistical techniques which can decompose a time series into unobservable components, such as that set out by Watson (1986). For example, output ($y$) can be decomposed into a permanent ($y_p$) and a transitory component ($z$):

$$y_t \equiv y_p^t + z_t$$  \hspace{1cm} (11)

where the permanent and transitory components correspond to the potential and gap terms in Equation (1). Permanent, or potential, output is assumed to follow a random walk with drift:

$$y_p^t = \mu^y + y_p^{t-1} + \epsilon^y_t$$  \hspace{1cm} (12)

where $\mu^y$ is a drift term and $\epsilon^y_t \sim N(0, \sigma^2_y)$. Since the estimation is from 1980 to 1997, and hence excludes the structural break in growth in the 1970s, the drift term is assumed to be constant. The output gap is assumed to follow an AR(2) process:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \epsilon^z_t$$  \hspace{1cm} (13)

where $\epsilon^z_t \sim N(0, \sigma^2_z)$ and the stationarity conditions hold.

This can be written in state-space form (Kuttner 1994; Gerlach and Smets 1997). Using the terminology of Harvey (1981), the vector of observed variables – a single
variable, output, in the example described above – is denoted as $Y$, while the vector of unobserved state variables – the trend growth rate, permanent output and the output gap in this case – is denoted by $\alpha$. The evolution of the observed variables is described in the measurement equation as a function of the unobserved state variables:

$$Y_t = Z\alpha_t + d_t + Se_t$$

(14)

where $Z$ is a matrix of coefficients, $d$ is a matrix of exogenous variables and $e$ is a vector of white-noise errors weighted by $S$. The time series processes for the unobserved state variables are set out in the transition equation:

$$\alpha_t = T\alpha_{t-1} + c_t + \eta e_t$$

(15)

where $T$ is a matrix of coefficients, $c$ is a matrix of exogenous variables and $e$ is a vector of white-noise errors weighted by $\eta$. Appendix A sets out the model more explicitly.

Estimates of the parameters of the model and the unobserved state variables can be obtained by maximising the likelihood function:

$$\log \Lambda = -\frac{NS}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{S} \log |F_t| - \frac{1}{2} \sum_{t=1}^{S} v_t'F_t^{-1}v$$

(16)

using the Kalman filter, where $N$ is the number of observed variables, $S$ is the sample size, $v$ is the prediction error matrix and $F$ is the mean square error matrix for the prediction errors. The initial values for the estimation were obtained using the HP filter with $\lambda = 1600$ and are shown in the first column of Table 2. The final estimated coefficients from this model are shown in the second column of Table 2, and the top panel of Figure 6 shows the estimate of the output gap.
| Initial estimates (HP model, \( \lambda = 1600 \)) | UC model of output and inflation | \( \mu_y \) | 0.0078 | 0.0077 | 0.0077 |
| \( \phi_1 \) | 1.15 | 1.52 | 1.26 |
| \( \phi_2 \) | -0.35 | -0.61 | -0.42 |
| \( \mu^\pi \) | -0.0001 | – | 0.0000 |
| \( \delta_1 \) | 0.50 | – | 0.96 |
| \( \delta_2 \) | 0.46 | – | 0.00 |
| \( \beta \) | 0.07 | – | 0.05 |
| Standard error (\( \varepsilon^y \)) | 0.0014 | 0.0060 | 0.0030 |
| Standard error (\( \varepsilon^\pi \)) | 0.0080 | 0.0051 | 0.0043 |
| Standard error (\( \varepsilon^\beta \)) | 0.0029 | – | 0.0013 |

Notes: Standard errors in brackets.

This estimate is a univariate decomposition of output. Kuttner (1994) has extended the model to include information about the output gap contained in the Phillips curve, assuming that inflation is below the expected rate of inflation when output is below potential (a negative gap). Gerlach and Smets (1997) have used this approach to estimate output gaps for the G7 economies. Following this, Equation (17) sets out an equation for the Phillips curve assuming that inflation expectations depend on past values of inflation of consumer price and imported goods and services:

\[
\pi_t = \mu^\pi + \delta_1 \pi_{t-1} + \delta_2 \pi_{t-2} + (1 - \delta_1 - \delta_2) \pi_{t-1}^* + \beta \pi_t + \varepsilon_t^\pi
\]  

(17)

where the error is also white noise. A constant term is included for generality but is expected to be zero. Appendix A also sets out the equations for the general state-space form when inflation is modelled this way and is included in the model.
Initial parameter estimates were obtained by estimating Equation (17) using the HP estimate of the output gap, and these are also presented in the first column of Table 2. The unobservable components estimate of the output gap which conditions on both the output and inflation data is shown in the bottom panel of Figure 6, and the coefficient estimates are set out in the third column of Table 2. The profile of the output gap is very similar when information about inflation is used, but the size of the gap varies at times. For example, demand conditions appear relatively weaker in the late 1980s when inflation is included, presumably because the Accord dampened wage rises, which are an important driving force in the inflation process. Similarly, demand conditions appear a little stronger in 1994 than suggested by just looking at output, presumably because inflation accelerated strongly at this time.

**Figure 6: Unobserved Components Estimate of the Output Gap**

Using output data only...

... and adding information about inflation.
2.5 Production Function Method

An alternative, ‘structural’, approach is to estimate the output gap as actual output less potential output calculated on the basis of an aggregate production function (see, for example, Laxton and Tetlow 1992; OECD 1994; Giorno et al. 1995; De Masi 1997; Fisher et al. 1997). For example, suppose that output can be characterised as a Cobb-Douglas production function:

\[ y_t = tfp_t + \alpha l_t + (1-\alpha)k_t \]  

(18)

where \( y \) is output, \( tfp \) is total factor productivity, \( l \) is effective labour and \( k \) is the capital stock, \( \alpha \) is the labour share of income, and variables are in logarithms. If the inputs are equilibrium values, then the production function provides an estimate of potential output and hence the output gap.

This is typically made operational in the following way. First, using historical values of the labour share of income (\( \alpha = 0.57 \)), total factor productivity is estimated as output less the weighted sum of effective labour and physical capital inputs:

\[ tfp_t = y_t - \alpha l_t - (1-\alpha)k_t \]  

(19)

Output is GDP(A). Effective labour is defined as full-time-equivalent employment multiplied by the ratio of the trend participation rate to the actual participation rate. Full-time equivalent employment is used since this corrects for changes in ‘labour intensity’ associated with an increase in the share of part-time jobs. It is calculated as full-time employment plus 15/40 times part-time employment. Employment also changes over the cycle because of changes in the participation rate, and the effect of this is removed by multiplying full-time-equivalent employment by the ratio of trend to actual participation. The trend in the participation rate is estimated as a linear trend from 1980 to 1997. The estimate of capital is the published capital stock, interpolated to give a quarterly series. The top panel of Figure 7 plots the natural log of indices of output, labour and capital.

A trend is then fitted to the residual, total factor productivity, in order to obtain an estimate of trend productivity to be used in calculating potential output. There are many techniques to calculate a trend. Since productivity growth changes over time, a simple linear trend is inappropriate. An HP filter is used to estimate trend total
factor productivity, and is shown in the second panel of Figure 7 with the estimate of total factor productivity from Equation (19). While use of the HP filter is arbitrary, it is no less so than alternative methods such as estimated or judgmental linear break-in-trend models. This estimate indicates that total factor productivity has been growing around 1.8 per cent a year since 1994.

**Figure 7: A Production Function Estimate of the Output Gap**

Potential output can then be estimated as trend total factor productivity plus the weighted sum of full-employment effective labour and the capital stock. Full-employment labour is assumed to be that level of employment consistent with non-accelerating inflation, and we use Debelle and Vickery’s (1997) estimate of the NAIRU for this calculation:

\[
l^T = l f (1 - nairu)(p^T / p)(ftequiv / N)^T
\]  

(20)
where \(lf\) is the labour force, \(nairu\) is the non-accelerating-inflation rate of unemployment, \(p\) is the participation rate, \(ftequiv\) is full-time equivalent employment, \(N\) is employment and a superscript \(T\) indicates a (linear) trend. The trend in the ratio of full-time-equivalent employment to total employment is used instead of the actual ratio in order to abstract from cyclical variation. The output gap is actual less potential output. The bottom panel of Figure 7 shows this estimate, together with a one standard error confidence band based on the standard error from the estimate of trend total factor productivity. These standard errors are one way to indicate the imprecision of estimates of the output gap. To the extent that there is uncertainty about the full-employment labour force and the capital stock (which is certainly the case), the ‘true’ standard error is larger (Fisher \textit{et al.} 1997).

The production function approach has been criticised on a number of grounds – that the economic structure changes, often in a way that is not known, that the method still uses simple detrending procedures (such as the HP filter) which have an important bearing on the gap estimate, that a range of \textit{ad hoc} assumptions are made about potential labour and capital, and that the data on the capital stock are of poor quality (Laxton and Tetlow 1992; Conway and Hunt 1997). These criticisms are valid but most of them also apply to other methods, and the production function approach has a strong intuitive appeal and is widely used (Giorno \textit{et al.} 1995; De Masi 1997).

3. \textbf{An Assessment of Estimates of the Output Gap}

Figure 8 shows estimates of the output gap for each of the five methods examined in Section 2. There are two striking features of these gap estimates. The first is that, at each point in time, there is a range of gap estimates and these sometimes give contradictory indications about whether there is excess capacity in the economy. For example, the linear trend model, some specifications of the univariate and multivariate HP filter models and the unobservable components model indicate that output was above potential at the end of 1997, while other specifications of the HP model and the production function approach indicate the opposite. Indeed, the multivariate HP filter conditioned on the particular NAIRU equation used indicates a large negative gap. The second feature is that all the output gap profiles are similar. Indeed, the correlation coefficients between the various gap measures range
between 0.86 and 0.98. This section explores some implications of these two features.

A common characterisation of forward-looking monetary policy is that policymakers set the short-term nominal interest so that its real interest rate component is above its ‘neutral’ value when inflation is expected to be above target and output is expected to be above potential over the forecast horizon. The methods canvassed in the previous section give different and conflicting answers about the output gap at particular points in time, suggesting that it is difficult to uncover the absolute size of the output gap at a point in time. But the result that the profiles of the gap are broadly similar indicates that it is sensible to assess the relative size of the output gap.

---

gap at a particular point in time by comparing the current estimate of the gap to its recent history and to past peaks and troughs.

Some methods are also better than others in gauging the current size of the output gap. An assumption embodied in the univariate techniques is that the output gap is close to zero on average, but this is not the case with multivariate procedures and the production function approach. If the inflation rate has been reduced over the period under consideration, then the average output gap should not be zero. In Australia’s case, for example, underlying inflation was reduced from an average 8 per cent from 1980 to 1990, to an average 2½ per cent since then, suggesting that univariate techniques will tend to understate the size of the output gap. Multivariate techniques do not force the output gap to average zero as long as the conditioning processes allow for a persistent wedge in those relationships due to a negative output gap (for example, the gap causing a wedge between the unemployment rate and the NAIRU or between actual and expected inflation). Similarly, the production function does not force the gap to average zero and so may provide a more accurate estimate of the size of the gap at particular points in time (although, as emphasised in Section 2.5, there is still uncertainty associated with this estimate).

The shortcomings of the various gap models indicate that judgment is needed in assessing model predictions: models cannot be used mechanistically but form one part of a broad set of information. Mistakes can, of course, be made, but the effects of misjudgments about the output gap are reduced when policy has clear targets. For example, as explained in de Brouwer and O’Regan (1997) in the context of inflation targeting, if policy-makers err in their judgment about the output gap, and set policy too tight or loose, then inflation forecasts will contain systematic errors and inflation will deviate from the target rate and invoke a policy response.

Moreover, the result that the profile of the output gap is broadly consistent across estimation methods indicates that the difference between the measures is essentially a scale effect, which has an important implication for model building. Economic theory and modelling suggest that the output gap can help predict consumer price and wage inflation, implying that the various gap measures can be ranked by how well they help explain price and wage inflation. A simple mark-up model of
inflation, written in error-correction format and based on de Brouwer and Ericsson (1995), is used as a benchmark model to assess the performance of the different output gap estimates. The equation is:

$$\Delta p_t = \alpha_0 - \alpha_1 p_{t-1} + \alpha_2 ulc_{t-1} + \alpha_3 ip_{t-1} + \alpha_4 \Delta ulc_t + \alpha_5 gap_{t-3}$$ (21)

where \( p \) is underlying consumer prices, \( ulc \) is unit labour costs, \( ip \) is import prices (adjusted for tariffs), \( gap \) is the output gap and \( \alpha_1 = \alpha_2 + a_3 \). The third lag of the gap is used since this fits the equation best in all cases. The results are presented in Table 3.

### Table 3: Inflation Equations and the Output Gap, 1980–97

<table>
<thead>
<tr>
<th></th>
<th>No gap</th>
<th>4-quarter-ended growth in output</th>
<th>Linear gap ( \lambda = 1 )</th>
<th>HP gap</th>
<th>MV HP gap</th>
<th>Unobserved components gap</th>
<th>Production function gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_0 )</td>
<td>0.0065**</td>
<td>-0.0078</td>
<td>-0.007**</td>
<td>0.0069**</td>
<td>0.0101*</td>
<td>0.0071*</td>
<td>0.0083**</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0047)</td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.001)</td>
<td>(0.0005)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>( -\alpha_1 )</td>
<td>-0.08**</td>
<td>-0.15**</td>
<td>-0.09**</td>
<td>-0.09**</td>
<td>-0.08**</td>
<td>-0.09**</td>
<td>-0.09**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.05**</td>
<td>0.10**</td>
<td>0.05**</td>
<td>0.05**</td>
<td>0.04**</td>
<td>0.05**</td>
<td>0.05**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>0.03**</td>
<td>0.05**</td>
<td>0.04**</td>
<td>0.04**</td>
<td>0.04**</td>
<td>0.04**</td>
<td>0.04**</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>( \alpha_4 )</td>
<td>0.19**</td>
<td>0.21**</td>
<td>0.12**</td>
<td>0.12**</td>
<td>0.10**</td>
<td>0.11**</td>
<td>0.11**</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>( \alpha_5 )</td>
<td>-</td>
<td>0.024***</td>
<td>0.083**</td>
<td>0.112**</td>
<td>0.092**</td>
<td>0.118**</td>
<td>0.104**</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.014)</td>
<td>(0.019)</td>
<td>(0.016)</td>
<td>(0.019)</td>
<td>(0.017)</td>
<td></td>
</tr>
</tbody>
</table>

| R-bar-sq       | 0.870   | 0.884                           | 0.906                         | 0.909   | 0.908     | 0.911                     | 0.912                  |
| Q(18)          | 0.08    | 0.11                            | 0.56                          | 0.55    | 0.39      | 0.47                      | 0.28                   |
| \( \alpha_1 = \alpha_2 + a_3 \) | 0.27    | 0.23                            | 0.49                          | 0.43    | 0.53      | 0.81                      | 0.94                   |

Notes: * and ** indicate significant at the 5% and 1% level. The Q(18) statistic is the marginal significance of the Box-Pierce test that all 18 autocorrelation coefficients of the error are zero. The linear homogeneity restriction is imposed and the marginal significance that the restriction holds is reported in the final row.

The explanatory power of the equation is good even without an output gap since inflation is a very persistent process. But including any of the gap measures improves the fit of the equation, generally by about four per cent in this case, and
eliminates serial correlation in the residuals. Moreover, all the output gap measures explain inflation better than just including the current four-quarter-ended output growth. The gap measure that explains output the best is that derived using the production function method outlined in Section 2.5, but this is marginal.

Regardless of the particular measure of the output gap, the movement in the gap over time substantially helps to predict inflation, with the scale effect reflected in the constant and the slope coefficient on the output gap. This is important. In the first place, in empirical models of inflation like Equation (21), the inflationary impulse from the output gap depends on where the output gap is relative to its average value, rather than where it is relative to zero. The various estimates of the gap reported in Table 3 have different averages, and this is reflected in the different constant terms in the estimated equations. Second, in empirical models of inflation like Equation (21), the effect of different cyclical amplitudes in estimates of the output gap on the inflation forecast are ‘washed out’ by different slope coefficients. Estimates of the output gap which produce greater amplitude in the cycle, for example, will tend to have smaller regression coefficients on the output gap in an inflation equation. The ‘fit’ of the equation is similar, even if the estimated parameters vary.

Given that the state of demand is an important source of innovations to price and wage inflation, the inflation prediction error is expected to be considerably higher when the gap is excluded. Table 4 shows how the various gap measures can improve out-of-sample prediction of inflation, reporting the RMSE and absolute mean prediction error for March 1996 to December 1997 when the inflation equations are estimated to the December quarter of 1995. The predictions are based on the actual values of all the independent variables, including the output gap measures. On this basis, the multivariate HP filter and production function measures perform best in predicting inflation, almost halving the inflation prediction error, at least after the fact.

---

7 The output gap also influences unit labour cost growth, so the improvement in fit is even greater when current growth in unit labour costs is excluded from the equation. When this is done, the relative ranking of gap measures does not change.
8 This holds for both total GDP(A) and non-farm GDP(A).
9 Each of the gap estimates is based on information over the full sample period and so this test overstates the out-of-sample predictive capacity of the output gap. We do not pursue this here.
Table 4: Prediction Errors

<table>
<thead>
<tr>
<th>RMSE</th>
<th>No gap</th>
<th>4-quarter-ended growth in output</th>
<th>Linear gap</th>
<th>HP gap</th>
<th>MV HP gap</th>
<th>Unobserved components gap</th>
<th>Production function gap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0030</td>
<td>0.0022</td>
<td>0.0021</td>
<td>0.0026</td>
<td>0.0015</td>
<td>0.0021</td>
<td>0.0016</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.0028</td>
<td>0.0020</td>
<td>0.0019</td>
<td>0.0023</td>
<td>0.0014</td>
<td>0.0019</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

Note: RMSE is the root mean square error.

4. Conclusion

Policy-makers are interested in the output gap for two reasons. First, excess capacity indicates that the economy could produce more and could generate more jobs. Second, excess capacity is an important determinant of price and wage outcomes, and so contains key information for monetary policy. This paper reviewed five techniques used to estimate the output gap – linear time trends, Hodrick-Prescott (HP) filter trends, multivariate HP filter trends, unobservable components models and a production function model. It shows that the estimates of the output gap depend on the estimation method selected, assumptions made in estimation, and the sample period, making it difficult to determine the absolute size of the output gap at a particular point in time. This difficulty is made worse when the output gap is estimated over periods which include a disinflation (as occurred in Australia in the early 1990s). Over such periods, the output gap should be negative on average, but univariate statistical techniques force the gap to average zero. Output gap methods which rely on this assumption yield positively biased estimates of the current output gap. This is not necessarily the case with more complex techniques, since they allow estimates of the gap to be conditioned on this information. But the way the conditioning relationships which underlay these models are structured is of critical importance to the estimate of the output gap, which underscores the need for careful analysis and judgment.

The profiles of the output gap produced by these different techniques, however, are broadly similar. This suggests that the relative size of the gap can be determined by comparing the gap at the date of interest with the past profile of the gap. It also
indicates that the gap measure can be useful for model-based analysis, since the differences between gap measures will be captured by adjustment to the constant and slope coefficients. For example, using a simple mark-up model of inflation, all of the output gap measures examined improve the fit of the inflation equation and substantially reduce _ex post_ prediction error. The output gap performs this task better than output growth, indicating that the degree of excess capacity in the economy is an important determinant of inflation.
Appendix A: The Equations in the State-Space Model

Given the characterisation of output, Equations (14) and (15) can be written as:

\[
\begin{bmatrix} y_t \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t^y \\ y^p_t \\ z_t \\ z_{t-1} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mu_t^y \\ y^p_t \\ z_t \\ z_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1}^y \\ y^p_{t-1} \\ z_{t-1} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \end{bmatrix} \quad \text{(A.1)}
\]

When inflation is included in the model, in the way specified in Equation (17), the state-space form in Equations (14) and (15) is:

\[
\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \delta_1 & \delta_2 \end{bmatrix} \begin{bmatrix} \mu_t^y \\ \mu_t^\pi \\ y^p_t \\ \pi_t \\ z_t \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ (1-\delta_1-\delta_2)\pi_{t-1} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \end{bmatrix} \quad \text{(A.3)}
\]

\[
\begin{bmatrix} \mu_t^y \\ \mu_t^\pi \\ y^p_t \\ z_t \\ z_{t-1} \\ \theta_t \\ \omega_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_1 & \phi_2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \beta & 0 & \delta_1 & \delta_2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1}^y \\ \mu_{t-1}^\pi \\ y^p_{t-1} \\ z_{t-1} \\ z_{t-2} \\ \theta_{t-1} \\ \omega_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \end{bmatrix} \quad \text{(A.4)}
\]
Appendix B: Data Description and Sources


**Inflation:** Treasury underlying consumer price inflation, *Consumer Price Index*, ABS Cat. No. 6401.0.

**Import prices:** *Import Price Index*, ABS Cat. No. 6414.0.

**Expected Inflation:** Melbourne Institute survey of households’ expected inflation one year ahead, Melbourne Institute, University of Melbourne.

**Unemployment rate:** *Labour Force*, ABS Cat. No. 6203.0.

**NAIRU estimate:** Debelle and Vickery (1997).

**Capacity utilisation:** ACCI-Westpac survey of manufacturing firms.

**Labour force:** *Labour Force*, ABS Cat. No. 6203.0.

**Capital stock:** *Capital Stock*, ABS Cat. No. 5221.0.
References


