In search of a method for measuring the output gap of the Swedish economy

Economic, econometric and practical considerations

Göran Hjelm and Kristian Jönsson

February, 2010

WORKING PAPER NR 115, FEBRUARY 2010

PUBLISHED BY THE NATIONAL INSTITUTE OF ECONOMIC RESEARCH (NIER)

NIER prepares analyses and forecasts of the Swedish and international economy and conducts related research. **NIER** is a government agency accountable to the Ministry of Finance and is financed largely by Swedish government funds. Like other government agencies, **NIER** has an independent status and is responsible for the assessments that it publishes.

The **Working Paper** series consists of publications of research reports and other detailed analyses. The reports may concern macroeconomic issues related to the forecasts of the institute, research in environmental economics, or problems of economic and statistical methods. Some of these reports are published in their final form in this series, whereas others are previews of articles that are subsequently published in international scholarly journals under the heading of **Reprints**. Reports in both of these series can be ordered free of charge. Most publications can also be downloaded directly from the **NIER** home page.

KONJUNKTURINSTITUTET, KUNGSGATAN 12-14, BOX 3116, SE-103 62 STOCKHOLM

TEL: +46 8 453 59 00 FAX: +46 8 453 59 80

E-MAIL: KI@KONJ.SE HOMEPAGE: WWW.KONJ.SE

ISSN 1100-7818

Preface

This paper describes and evaluates measures of trend (or potential) output in order to improve the measuring and understanding of the current state of the Swedish economy. The target group of the paper is primarily policy makers and analysts in Sweden and international organisations who study the Swedish economy and give recommendations concerning appropriate stabilization policies.

The views expressed are those of the authors and do not necessarily represent the ones of neither the National Institute of Economic Research nor the Ministry of Finance.[1]

[1] We are grateful for comments made by seminar participants from the Ministry of Finance, the National Institute of Economic Research and the Riksbank (Central Bank of Sweden). All remaining errors are our own responibility.

Contents

1. Introduction

Measuring the output gap, i.e. the level of output in relation to an estimated, but unobserved, "trend"[2] output, is a very important ingredient when evaluating the current stance of an economy and forming appropriate stabilisation policy actions. Traditionally, the output gap broadly concerns the issue of identifying demand and supply imbalances and if, for example, demand exceeds supply for some time shortages in both product and labour markets eventually lead to higher price and wage inflation. As it takes time to put both fiscal and monetary policies into work, an output gap signalling the current stance of the economy and the inflation pressure in the near future would potentially be of great help. Ideally, the output gap, together with an estimation of how stabilization policies affect output, could also serve as an indicator of "how much" stabilization policy that is needed in any given phase of the business cycle in order to push the economy towards equilibrium.

As is well-known, however, there are a variety of empirical methods around to estimate the output gap and it is often rather difficult to prove which model, if any, to prefer. The more traditional methods range from univariate, a-theoretical, estimations to multivariate, theory-based, models. A common feature of methods within the traditional approach, is the focus on the macroeconometric relationship between output gaps and inflation in small, empirical models. In recent years, the development and use of so-called DSGE-models at (foremost) central banks have introduced new definitions and new interpretations of output gaps (see e.g. Galí 2008a,b). These models are often larger and have more micro foundations. The present paper is not within this line of research. In Section 2.3 we briefly comment on output gaps within the DSGE-literature. In Chapter 3, we outline the criteria by which we evaluate the different output gaps considered in this study.

Currently, there are a number of national and international organisations measuring the output gap of the Swedish economy. Internationally, the European Commission (see Cahn and Saint-Guilhem, 2007), the European Commission (see Denis et al, 2006), the International Monetary Fund and the OECD (see Beffy et al, 2007) all use variants of the so-called Production Function (PF) approach to measure the output gap of the Swedish economy. The main reason is probably that this approach has proven to be relatively easy to carry out in similar manner for many countries, which of course is a crucial criterion when doing country comparisons. Briefly, the PF-approach usually builds on a Phillips-curve framework for estimating the NAIRU[3] and univariate Hodrick-Prescott (HP) filters for estimating developments of trend Total Factor Productivity (TFP), the trend labour force and trend average working hours (see Chapter 2 for a more detailed description).

Nationally, three organisations officially estimate the output gap of Sweden; the Riksbank (Central bank of Sweden), the Ministry of finance, and the National Institute of Economic Research (NIER). The Riksbank publishes an output gap solely based on a HP-filter. The Ministry of finance and the NIER mainly use judgement to determine the NAIRU[4] and various HP-filters for evaluating the cyclical part of labour productivity, the labour force, average working hours etc. (see Chapter 2 for a more detailed description).

In sum, it appears like international and national organisations depend rather heavily on the HP-filter when measuring the current stance of the Swedish economy. Although this might not be a

---

[2] There are several other expressions (such as "potential" or "structural") for this unknown output level. We simply use "trend" here; see Chapter 2 and Chapter 5 concerning definition and labelling of various measures of "trend" output.

[3] Common methods for calculating the NAIRU are presented in Chapter 5.

[4] The NIER use equilibrium unemployment (or long run NAIRU) in their current analysis of the output gap. See Richardsson et al (2000) for a description of the differernce between the NAIRU and the long run NAIRU concepts. See also Section 5.2 for a discussion.

bad choice per se, there are well-known problems with the HP-filter. First, we have the so-called end-point problem, implying that the HP-filter, by construction, close gaps in the end of the sample (see Baxter and King, 1995). This is especially problematic as policy makers are mainly interested in the gap at the end of the sample. Second, the HP-filter embodies a minimum of economic theory. This is a major problem when communicating why a certain stabilisation policy is needed. A univariate approach is not well-suited for so-called "story-telling", which is an important part of most organisations dealing with forecasting and stabilisation policy advice (see Mestre and McAdam, 2008).

It is clear that the method chosen for calculating the output gap, at least in part, depends on the role of the organisation in question as well as on the role of the output gap in the forecasting and policy making process. As mentioned above, international organisations tend to choose a method that is possible to replicate for many countries. Organisations using trend output in their external communication probably prefer theory-based approaches (rather than mechanical filters) as such approaches simplify the "story telling" about why the gap changes from time to time. In other words, the criteria upon which the output gaps are chosen, can differ depending on how the output gap is used and communicated in a particular organisation. Besides the communication aspect, a number of criteria have been put forward in the literature. Cotis et al (2005), for example, discuss important criteria they have found in their experience of being users of output gaps at the OECD. Among the so-called core criteria they list, they prefer transparent methods with small revisions which do not have any end-point problem (see Chapter 3).

Although all of these criteria arguably are desirable, one can for example note that the ability of the output gap to forecast inflation is not included. According to Cotis et al (2005), it is central for the OECD to use the PF-approach so that one can evaluate and discuss assumptions concerning the development of trend TFP, trend work force, average working hours and so on. Others, like Camba-Mendez and Rodriguez-Palenzuela (2003), include the ability of an output gap to forecast inflation as one of their criteria.

As mentioned, the focus of an organisation probably affects the method chosen; a central bank is probably most interested in output gaps that can help to forecast inflation. When writing this paper, both authors are working at the NIER in Sweden. The criteria we list in Chapter 3 are hence influenced by our experience working side by side with forecasters and policy makers. In short, we believe that forecasters need an output gap that is theory-based and have some correlation with future developments of inflation and GDP-growth. These and other criteria chosen are discussed in Chapter 3.

The purpose of the paper is to describe and present results from a relatively large number of methods measuring the output gap of the Swedish economy. The main aim is to propose a method that could substitute the current method used by the NIER. In short, we consider:

1. Structural Vector AutoRegression (SVAR) approaches.
2. Unobserved Components (UC) approaches.
3. Multivariate HP-filters (MVHP).

Common to these approaches is that they, to different degrees, are theory-based. The choice of methods is implicitly governed by the fact that NIER's use output gaps extensively in the communication, both internally and externally. In short, "story-telling" is an important part of NIER's work. For this reason, we will not consider univariate, a-theoretical approaches in this paper. Based on a number of criteria outlined in Chapter 3, we choose a method that we believe is most useful for NIER's communication of the current stance of the Swedish economy. We also firmly believe that the chosen method could, at least, complement the methods currently used and published by other national organisations in Sweden, i.e. the Riksbank and the Ministry of Finance. For international organisations who, as described above, are concerned about international comparisons, the chosen method in this paper could complement current PF-approaches when examining the Swedish economy.

It is important to stress, however, that the choice of method in the estimation of the output gap may, as discussed above, depend on the user's preferences. Some users may solely focus on the ability of the output gap forecasting inflation. Other users may want to include exogenous information, or priors, in their estimations. Concerning the latter, we include such an approach as well (the so-called multivariate HP-filter; MVHP) in our study, where the user exogenously (based on judgement, for example) in a transparent way can impose a time series of NAIRU, trend capital utilisation etc. when estimating trend output. In short, an important second purpose of our study is to provide a set of methods for estimating the output gap in Sweden. We believe the potential drawback of different organisations using different methods by and large is outweighed by restricting the users to use their chosen method in a transparent and replicable way. For (foremost) central banks this is of utter importance as one their most important tasks is to affect expectations and facilitate agents' ability to predict the future interest path.

1.1. Outline

The rest of the paper is structured as follows. Chapter 2 briefly describes the current methods used by international and national organisations to estimate the output gap in Sweden. The latest updates of the output gap estimations of these organisations are also shown. The chapter ends with a discussion about the output gaps within the DSGE-literature. In Chapter 3, the criteria for choosing between different methods are listed and discussed. Chapter 4 describes the data and the use of both so-called quasi real-time (QRT) and full sample (FS) data. In Chapter 5, the methods considered in the paper are described and corresponding (QRT and FS) output gaps are shown together with some descriptive statistics. More technical details concerning the methods are given in Appendix. In Chapter 6, the output gaps are evaluated using the quantitative criteria listed in Chapter 3. In particular, Section 6.4 sums up the results and presents our preferred method considering both quantitative and qualitative criteria listed in Chapter 3. Chapter 7 concludes.[5] Finally, in the Appendix, the methods considered are described in greater detail and results from sensitivity analyses are presented.

1.2. Delimitations

First, we restrict ourselves to "traditional" approaches estimating the output gap; structural VAR models, UC-models and MVHP-models. We do not consider the recent methods within the DSGE-literature in which various types of output gaps can be derived and estimated. The main reason is that the profession is far from consensus concerning the appropriate definition of the output gap (see Section 2.3 for a discussion).

Based on the arguments above about the practical use of output gaps in NIER's work, including the importance of "story-telling", univariate and PF-approaches are not considered.[6] Hence, the

[5] In Hjelm (2010, only in Swedish), we discuss ways to implement the chosen method in NIER's forecasting process.

[6] As outlined in Chapter 2, the PF-approach depends rather heavily on HP-filters. The NIER used such an approach in the beginning of the 2000s, but stopped using it partly due to the disadvanteges of the HP-filter.

conclusions presented in this paper may not carry over to organisations who find univariate and/or PF-approaches useful in their work.

Finally, we do not consider so-called real-time data in this paper. Although such data would have been of great importance, it is currently not available for a reasonable number of Swedish variables. As a result, we only consider so-called quasi-real time and full sample estimates of the output gaps (see Chapter 4). We strongly suggest, however, that efforts are made by Swedish authorities making a real-time data set easily available.

2. Current methods and measures of the output gap in the Swedish economy

There are a number of international and national organisations who currently estimate, and update, an output gap of Sweden. This chapter briefly presents the current methods applied and the corresponding estimates.

2.1. International organisations

As discussed in the Introduction, international organisations tend to use the so-called Production Function (PF) approach for measuring the output gap. This is true for the European Commission (EC), the IMF and the OECD. The reason is apparently not that this method has proven to be the best one when it comes to forecasting inflation, for example. Rather, the merit of the method is probably that it has proven to be relatively easy to replicate for many countries. Another reason for applying this method internationally is arguably that the output gap is naturally divided into different "parts". This can be illustrated by assuming a Cobb-Douglas production function:

$$Y = AK^{\alpha}H^{1-\alpha}, \tag{1.1}$$

where Y is production, A is Total Factor Productivity (TFP), K is capital stock, H is total working hours and α is the capital share of total income. The output gap, expressed as $y-y^*$ (lower case letters denote logs), can be decomposed into the following "sub-gaps":

$$y-y^* = \left(a-a^*\right)+\alpha\left(k-k^*\right)+\left(1-\alpha\right)\left(h-h^*\right), \tag{1.2}$$

where:

- $\left(a-a^*\right)$ is a "TFP gap",
- $\left(k-k^*\right)$ is a "capital stock gap" (often, but not always, set to zero), and
- $\left(h-h^*\right)$ is a "hours gap" (or "labour market gap").

Hours, H, can be further decomposed using the following identity:

$$H = POP\left(\frac{LS}{POP}\right)\left(\frac{L}{LS}\right)\left(\frac{H}{L}\right), \tag{1.3}$$

where POP is working age population, LS is labour supply and L is employment, while the "hours gap" then can be decomposed as follows:

$$\begin{aligned}\left(h-h^*\right) = \left(pop-pop^*\right)+\left[\left(ls-pop\right)-\left(ls-pop\right)^*\right]+ \\ \left[\left(l-ls\right)-\left(l-ls\right)^*\right]+\left[\left(h-l\right)-\left(h-l\right)^*\right],\end{aligned} \tag{1.4}$$

where:

- $\left(pop - pop^{*}\right)$ is the "working age population gap", which, for obvious reasons, always is set to zero,
- $\left[\left(ls - pop\right) - \left(ls - pop\right)^{*}\right]$ is the "labour participation gap". Empirically, participation appears to be partly determined by business cycle conditions.
- $\left[\left(l - ls\right) - \left(l - ls\right)^{*}\right]$ is the "employment gap", which equals $-\left(u - u^{*}\right)$, where the latter is "the unemployment gap" and u^{*} is the NAIRU.
- $\left[\left(h - l\right) - \left(h - l\right)^{*}\right]$ is the "average working hours gap". Empirically, average working hours appears to be partly determined by business cycle conditions.

International organisations use hybrid PF-approaches.

- The OECD (see Beffy et al, 2007):
  - Estimates the NAIRU within an Unobserved Components (UC) setting using a state space specification and the Kalman filter for estimation. The NAIRU is extracted via a Phillips-curve relationship (see Section 5.2 for a general description of UC-models).
  - Use HP-filters to determine trend TFP, trend participation rate, trend in average working hours and trend capital services.
- The European Commission (see Denis et al, 2006):
  - Estimates the NAWRU in a UC model, relating unemployment to wage growth.[7]
  - Use HP-filters to determine trend TFP, trend participation rate and a trend in average working hours.
  - Use actual capital stock.
- The IMF:
  - Use the HP-filter on the different parts (including the NAIRU) in equations (1.2) and (1.4). Then the parts are added to an aggregate measure.[8]

The resulting output gaps of Sweden published by the international organisations above are shown in Figure 1.

2.2. National organisations

As mentioned above, three national organisations are currently measuring the output gap of the Swedish economy; the Riksbank (Central Bank of Sweden), the Ministry of Finance and the NIER. The Riksbank applies a HP-filter directly on GDP. The Ministry of Finance and the NIER divide the output gap into a "hours gap" (or "labour market gap") and a "productivity gap" as outlined in Section 2.1 above. The "productivity gap" is basically a HP-filter on the productivity series.[9] The different parts of the "hours gap" (see equation (1.4)) are treated rather similarly by the Ministry of Finance and the NIER:

---

[7] NAWRU stands for "**N**on-**A**ccelerating **W**age inflation **R**ate for **U**nemployment". For Sweden, they find that wage growth works better than inflation.

[8] Currently there is no paper available at the IMF describing their method applied to Sweden. The information in the main text stems from e-mail conversations with Jay Sanat Surti, new desk economist for Sweden.

[9] Both the Ministry of Finance and the NIER also add judgement at the end of the sample concerning the "productivity gap", partially due to the end-point problem of the HP-filter.

- The “population gap” is zero.
- The “labour force participation gap” depends on the business cycle development, foremost the unemployment rate.
- The NAIRU is based on judgement which, in turn, is based on labour market indicators, econometric estimations, price and wage developments.
- The “average working hours gap” is based on a HP-filter.

The resulting output gaps estimations of the Ministry of Finance, the Riksbank and the NIER are shown in Figure 2. One can note that the gaps of the international organisations and the Riksbank are rather similar while the gaps of the Ministry of Finance and the NIER deviate somewhat from these gaps. For example, during the 1990s, the gaps of the NIER and the Ministry of Finance imply that the economic crisis affected trend GDP to a lesser extent. One explanation for this could be a more optimistic view on how the crisis affected the level of structural unemployment. Finally, one can note that the NIER has a rather different view on the output gap in the end of the sample, 2006–2008.

In summary, the output gaps of Sweden estimated by international and Swedish organisations are heavily based on HP-filters. This is especially true for components such as TFP, labour productivity, labour force participation and average working hours. NAIRU, on the other hand, is estimated by the international organisations using economic theory as a guideline, while national organisations (except the Riksbank who does not consider the NAIRU explicitly) use judgement. Finally, it is important to note that disaggregated approaches dominate. That is, trend output is calculated by summing up the parts; working hours, productivity, labour force participation etc. No one, except the Riksbank, apply an aggregate approach that calculates trend GDP directly.

This paper aims to remedy the lack of evidence concerning output gap estimates of Sweden using methods that are characterized as being (i) aggregate, (ii) based (to various degrees) on economic theory, (iii) related to future inflation and growth developments and, (iv) based on a method that involves a limited amount of judgement. The latter aspect increases the degree of transparency and replicability.[10]

[10] It can be noted that the MVHP-filter, to a larger degree then the other models considered, involves judgement; see Section 5.3.

Figure 1 The Swedish output gap according to international organisations

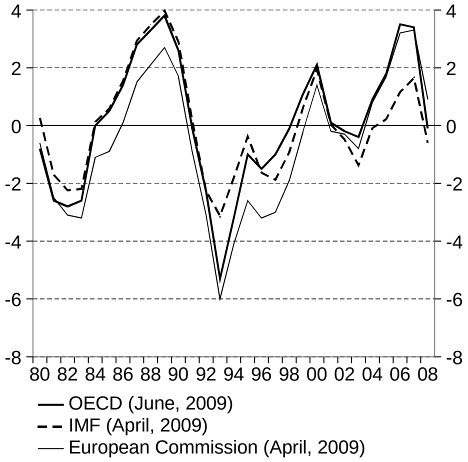
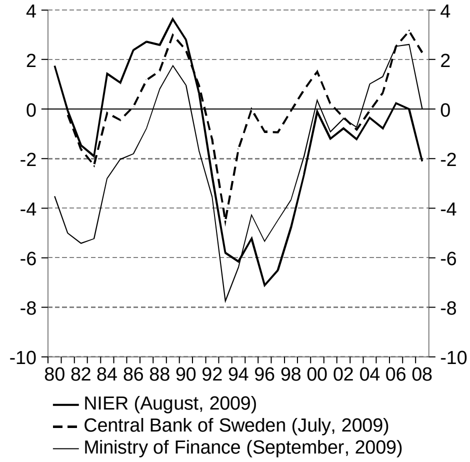


Figure 2 The Swedish output gap according to national organisations



2.3. Output gaps within the DSGE-literature

The output gap concept of the "traditional" approach outlined above and applied in this paper is different than the ones outlined within the DSGE-literature. As argued below, there are several reasons why the traditional concept of the output gap not yet should be abandoned in spite of the emerging DSGE types of gaps. In passing, one can note that the traditional types of output gaps are still widely used and developed by both national and international institutions.[11] Moreover, the "active use of flexible-price output gap in policymaking institutions remains limited" (Coenen et al, 2008, p. 2), where "flexible-price output gap" refers to the most common output gap measure put forward in the DSGE-literature.

The DSGE models are relatively new tools in business cycle analysis. These models are often based on the so-called New Keynesian framework in which real business cycle features such as rational expectations, optimizing agents and complete markets are modelled together with nominal and real rigidities (see Galí 2008a and Woodford, 2003, for textbook introductions). This new and fast growing type of models are specified and (to various degrees) already in use at several central banks including the Federal Reserve Board (see Edge et al, 2008), the European Central Bank (see Christoffel et al, 2008), the European Commission (see Ratto et al, 2006), the Bank of Canada (see Murchison and Rennison, 2006), and the Riksbank (see Adolfson et al, 2007).

The flex-price output gap is more complex concept, and has a more thorough theoretical description, than the traditional output gap concept. The standard definition of the output gap in the DSGE literature is actual output minus the output level that would prevail if price and wages would have been flexible and in the absence of so-called mark up shocks.[12] This gap focuses on the distance to actual output that monetary policy can affect and is hence, if correctly measured, the relevant one for central banks.

The flex-price level of output is affected by many shocks extracted from data when estimating DSGE-models (e.g. permanent and transitory technology shocks, foreign and import demand shocks, shocks to terms-of-trade, risk premiums, interest rates etc.). These shocks affect of course actual output as well, but as the responses differ from the flex-price output responses, the behaviour of the output gap is rather complex. Moreover, the fact that flex-price output often is allowed to be affected by all shocks but mark-up shocks, flex-price output might be as volatile (or even more volatile) than actual output. Although this is not necessarily wrong, it is arguably against today's common view of most policy makers (see Bean, 2005).[13]

Despite the differences outlined above, it is important to note that the flex-price output gap in DSGE-models is closely connected to inflation, i.e. the single most important criterion applied in this paper. Coenen et al (2008) even perform a similar out-of-sample inflation forecast contest as we carry out in this paper to evaluate output gaps of DSGE-models.

---

[11] See e.g. the list of papers presented at the 5th Eurostat Colloquium on Modern Tools for Business Cycle Analysis, in Luxembourg September-October 2008; http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1194,70264713,1194_73170612&_dad=portal&_schema=PORTAL.

[12] This is the view of, among others, Edge et al (2008), Smets and Wouters (2003) and Woodford (2003). See also Adolfson et al (2008) on different defintions of output gaps in DSGE-models.

[13] Here one can note that the SVAR approaches we consider also connect the output gap to various shocks. The output gap is in the SVAR approaches defined as actual output minus the level of output that would have prevailed in absence of a number of shocks.

There are several reasons why we believe that the traditional output gaps still should be developed and used in business cycle analysis. We agree with Coenen et al (2008) who argue that it is simply too early to introduce the DSGE-based output gaps (see also Tovar, 2008, on this issue). An active, transparent and explicit use of these types of gaps has not been seen in the publications of central banks developing these models.[14] This is probably due to a number of factors. First, the profession has not agreed upon an operational definition of the output gap which partly stems from problems associated with the identification of shocks.[15] Second, the flex-price output gap in its current form appears to be too difficult to communicate. Third, it could be the fact that central bankers (the Monetary Policy Committee as well as members of the staff) think that the DSGE-models not yet describe business cycle features good enough; this might be true both concerning the driving forces (shocks) as well as the propagation mechanisms. As an example, an arguably important indication of business cycle movements is the large variations in unemployment. This feature is arguably yet far from captured in these models. For those who do not believe in voluntary variations in unemployment, this is an unpleasant feature of the existing DSGE-models and we do not know what would happened with the DSGE-gaps and their policy prescriptions if the labour market would be modelled differently.

In conclusion, we believe that the traditional output gap measures will continue to be an important part of business cycle analysis. The DSGE-models will in all probability evolve and play an increasing role in future analysis.

[14] When it comes to DSGE-modelling, one can note that none of the three most progressive central banks publish DSGE-based output gaps in their main policy reports. The Riksbank in Sweden publishes a HP-filtered gap (see their *Monetary Policy Report*), Norges Bank publishes a judgement-based gap (see their *Monetary Policy Report*), while the Reserve Bank of New Zealand publishes a gap based on a multivariate HP-filter (see their *Monetary Policy Statement*).

[15] See Chari et al (2008) and Canova and Sala (2009) for criticism of the way shocks are identified in DSGE models.

3. Criteria for selecting output gaps in the study

As discussed in the Introduction, policy makers and economists who make use of output gaps in their analysis may have different preferences concerning the properties of the gap. For example, some experienced economists engaged in policy advice within the OECD list the following so-called core criteria (see Cotis et al, 2005):

- Consistency between economic priors and the underlying assumptions of the method.
- Transparent and easily replicated.
- Small revisions.
- Not sensitive to last observations (no end-point problem).[16]
- Precision of estimates.

The OECD often uses the output gap in their communication. The first criterion could be seen in the light of this fact; they need a method that is in line with how they believe the economy works. The rest of their criteria are straightforward. One can note that the ability of the output gap to forecast inflation is not included among their core criteria. Others, such like Camba-Mendez and Rodriguez-Palenzuela (2003), emphasize three criteria in their evaluation:

- Ability to forecast inflation.
- Small revisions.
- Correlation with standard measures of capacity utilization.

Their first criterion is the most common one in the literature, although not considered as crucial by Cotis et al (2005). As discussed in Chapter 1, the traditional view of output gaps concerns the identification of supply and demand imbalances in the economy and implications for the inflation development. Their last criterion focus on a variable often referred to when analyzing the current state of the economy. In surveys, companies are asked to evaluate how much they use of their capacity. Some find it desirable that an output gap broadly is in line with such a measure.[17]

Next, we outline the criteria for selecting output gaps in this paper. In the end of this chapter, we briefly discuss the notion of output gaps within the DSGE-literature and compare with the traditional approach.

3.1. Quantitative and qualitative criteria

The criteria of this study are of course governed by our preferences. These are, in turn, influenced by our experiences working in an institution (the NIER) that publish forecasts and give fiscal and monetary policy advice on a quarterly basis and use the output gap in their communication.

---

[16] The inclusion of forecast to postpone the closing of gaps is not a reasonable way out of the end-point problem; see Cotis et al (2005) and Baxter and King (1995). See Kaiser and Maravall (2001) for an ARIMA-approach aiming at reducing the end-point problem.

[17] See e.g. Koske and Pain (2008), Lemoine et al (2008), Orphanides and van Norden (2004), Ross and Ubide (2001) for examples of criteria used in the literature.

We split our criteria in two groups; *quantitative* criteria which can be measured statistically (e.g. inflation forecast performance) and *qualitative* criteria which cannot be measured statistically (e.g. transparency). We do not suggest an "optimal weighting scheme" for the criteria in the paper. We do, however, weigh the quantitative criteria more than the qualitative ones (see Section 6.4). Within each group, the criterion deemed the most important is listed first, the second most important is listed second etc.

The *quantitative* criteria, together with a brief motivation, are:

Inflation forecast performance
The traditional definition of trend output is the level of output where the inflation rate is stable; if output exceeds trend output for some time, inflation begins to increase. We believe that an output gap has to have some forecast power over future inflation in order to be taken seriously in forecast and policy institutions who believe in the more "traditional" view of output gaps.[18] It is important to remember, though, that it is difficult to find strong relationships between output gaps and inflation, especially during the last 10–15 years (see e.g. Orphanides and van Norden, 2004).

Growth forecast performance
This criterion complements the first one. Everything else equal, a correctly measured negative (positive) output gap signals that output growth will rise above (below) trend growth in the future in order for the economy to return to equilibrium (i.e. close the gap). We believe that this is an important ingredient when discussing the output gap and GDP growth around the forecasting table.

Size of revision between the use of Quasi-Real Time (QRT) and Full Sample (FS) data
As outlined below in Chapter 4, we consider output gaps estimated using both QRT and FS data in the paper. Both estimates use the most recent version (or vintage) of the data series published by the statistical authorities. The difference concerns the length of the time period considered. Estimating output gaps using QRT data implies that when, for example, considering the estimate of the gap in 2001q1, data only up to 2001q1 are used. The output gap of 2001q1 when applying FS data is based on an estimation using data up to the last quarter available (e.g. 2007q4).[19] The difference between the QRT and FS gaps, respectively, depends on how a particular method re-evaluates the estimate of e.g. 2001q1 when a longer time series is allowed for.

First it is important to stress that low revision between QRT and FS gaps is not necessarily a good characteristic of an output gap. Imagine a method that generates constant, zero output gaps using both QRT and FS data. Revision is then zero, but is it a reasonable gap? Probably not. Size of revision can, however, be an important criteria; if two output gaps have similar inflation forecast performance using QRT data, but the size of revision of one of them is considerably greater, the policy maker probably prefers the output gap with small revisions and will therefore use it to a greater extent (see Cotis et al, 2005).

In the paper, we compare the revision between QRT and FS gaps for the respective models. We look into this issue from three angels. First, we compare the root of the mean squared revision and the standard deviation of the full sample gaps. Second, we look at the number of times the respective models have the same sign of the QRT and FS gaps. Third, we look at the number of times the respective models have the same sign *of the change* in the QRT and FS gaps.

---

[18] It is also worth noting that the most frequently used gap within the DSGE literature (the so-called flex-price gap) also is closely connected to the inflation development (see Section 2.3).

[19] As discussed in Chapter 4, no so-called real time data for sufficiently long period and sufficiently many variables is currently available in Sweden.

Sign and change of the gaps using QRT and FS gaps

This criterion is similar to the one above. Everything else equal, if the sign and the sign of the change in the output gap using QRT and FS data often are the same, such a method is preferred to a method that generate gaps where the sign and sign of the change of the output gaps often differ between QRT and FS data.

The *qualitative* criteria, together with a brief motivation, are:

No end-point problem

As mentioned in Chapter 1, the HP-filter suffers from the so-called end-point problem. The construction of the filter is such that the deviations between the actual level and the trend level weigh more heavily in the end of the sample compared to the middle of the sample (see Baxter and King, 1995, and Bergman et al, 2009). This implies a tendency for the filter to narrow the gap in the end of the sample. This is a major problem in policy analysis as the focus is on the last observations of the estimated gap. A frequently used remedy for this problem is to extend the sample period by inserting forecast the latest outcome of the data. This, however, induces a problem in itself since forecasts should be founded in views of the trend rather than the other way around (see footnote 16). Hence, estimation methods that do not suffer from this problem are desirable.

Transparent, replicable, easy to communicate internally and externally

Although these criteria by nature are hard to measure and compare between models, it is still important to consider these issues when finally choosing a model; can a particular model be used in practice, internally and externally?

Zero on average

This is a desirable criterion for institutions, like for example the OECD and the NIER, evaluating fiscal policy and the structural budget balances. If, for example, an output gap has a negative bias, trend GDP will be higher than actual GDP on average. This, in turn, implies that an average of the actual budget balance as a percentage of GDP will differ from an average of the structural budget balance as a percentage of trend GDP. If, as in Sweden, the government has a target of one percent actual surplus over the business cycle, the average structural surplus must be higher than one percent if the output gap has a negative bias. This causes a problem when evaluating and recommending fiscal policy actions.

The "zero-on-average"-criterion should not, however, be too much emphasised as it could potentially distort the model selection under certain circumstances. Consider for example estimation of an output gap based on Swedish data starting in the beginning of the 1990s. Given that the sharp recession was mainly caused by cyclical factors, and given that we have not yet seen a boom of corresponding magnitude, there is a considerable risk that a stringent use of the "zero-on-average"-criterion in the model selection process would lead to false conclusions concerning trend GDP and hence the output gap.

Moreover, it is important to remember that if the Phillips curve is non-linear, so that prices and wages respond more to positive gaps compared to negative ones, the output gap will on average be negative (see Bergvall and Dillén, 2005, for a simple illustration). As evidence in Eliasson (2001) rejects the existence of a linear Phillips curve in Sweden, it is probably wise not to over-emphasise the desirability of symmetric output gaps.

In Chapter 6, we evaluate the quantitative criteria above and in Section 6.4, we summarize the output gaps in relation to both quantitative and qualitative criteria and choose our preferable model.

4. Data

As mentioned in the Introduction, two types of data are used; quasi-real time (QRT) data and full sample (FS) data. The principles behind these two types of data are described below together with so-called real time data which is not yet available in Sweden.

- FS data, corresponds to the latest data available for the time series included in the study. Say that we want to study the output gap from 1976q1–2007q4. When evaluating the forecast performance of inflation in, say, 1995q1, the output gap for the period 1976q1–1994q4 is taken from an output gap estimation of the whole period, 1976q1–2007q4.

- The use of QRT data implies that we, as with FS data, still make use of the latest available data for the time series included. But, when evaluating the forecast performance of inflation in, say, 1995q1, the output gap for the period 1976q1–1994q4 is taken from an output gap estimation only consisting of data for that period, i.e. 1976q1–1994q4.

- Real-time data is the data that was available data at the time when evaluating the forecast performance is concerned. For example, when evaluating the forecast performance of inflation in, say, 1995q1, the output gap for the period 1976q1–1994q4 is taken from an output gap estimation consisting of data available at 1994q4 for the period 1976q1–1994q4.

From the description above, it is evident that the difference between FS-based gaps and QRT-based gaps is the sample used for estimation. Hence, data revisions are not taken into account.

In this paper, we evaluate the different models measuring the output gap against the criteria listed in Chapter 3 using both QRT and FS data. We put greatest weight on the QRT results as the QRT data is closer to (but still relatively far from) the type of data that is available in real time. As outlined in Chapter 3, we also compare the revisions between QRT and FS estimates of the models considered.

4.1. Symbols and definitions

In this section, we define the symbols used for the different variables used in the study. Small letters denote natural logarithms. We use quarterly data through out the paper, and the variables are seasonally adjusted.[20]

[20] All data and programmes used in the paper can be received from the authors on request.

Symbol	Description	Source
Y	GDP	Statistics Sweden
W	Real wage. Nominal wage minus core inflation (KPIX)	Statistics Sweden
P	Price level (KPIX)	Statistics Sweden
U	Unemployment rate (ILO-definition)	Statistics Sweden
π	Inflation (KPIX)	Statistics Sweden
L	Working hours	Statistics Sweden
WL/Y	Wage share	Statistics Sweden
CU	Capital utilization	Statistics Sweden
Z	Supply variables in the UC models: • Percentage change in import prices relative GDP deflator • Percentage change in the real exchange rate • Percentage change in oil price relative to core inflation	 Statistics Sweden NIER Statistics Sweden

5. Methods considered estimating output gaps

There is a large number of methods available estimating trend output and the output gap; see e.g. Dupasquier et al (1999), Orphanides and van Norden (2004) and Prioretti (2008) for surveys. For studies focusing on Swedish data; see Apel and Jansson (1999) and Cerra and Saxena (2000). In this paper, we extend the current literature on the Swedish output gap by considering a broad range of methods, evaluations and sensitivity analyses.

For reasons outlined in Chapter 1, we will not consider univariate methods. Common to all methods considered is that they are multivariate and that are all estimated using the classical approach to econometrics.[21] As mentioned in Chapter 1, three approaches are considered in the paper:

1. Structural VAR (SVAR) models.
2. Unobserved Components (UC) models.
3. Multivariate HP-filters (MVHP).

These methods are the most common ones (together with the PF-method) within the traditional approach to estimate output gaps. Within these three methods many different specifications have been suggested in the literature. For approaches 1 and 2, we consider five and seven specifications, respectively, all of which passing pre-testing evaluations. All-in-all, the current version of the paper evaluates thirteen specifications; five SVAR-models, seven UC-models, and one MVHP-model.[22] We now turn to the description of the three approaches including estimation of the respective output gaps. Technical and empirical details concerning the different approaches are outlined in Appendix.

5.1. Structural VARs

Structural Vector AutoRegression (SVAR) models are frequently considered in the literature when estimating output gaps. While many identification procedure are possible, the use of the Blanchard and Quah (1989) identification procedure (long run restrictions based on economic theory) are considered, and we do so as well in this paper.

SVAR-models with long run restrictions are implicitly or explicitly based on an economic model. In Blanchard and Quah's (1989) original paper, output was assumed to be driven by two types of shocks, supply and demand, where demand shocks were restricted not to affect output in the long run. This theoretical assumption was used to identify the two shocks in the empirical model. Funke (1997) and Cayen and van Norden (2005) are two examples using this model to calculate the output gap in West Germany and Canada, respectively, stipulating that trend output is equal to the output series that would have appeared in absence of demand shocks.

A key feature of SVAR-models is that, given the identifying assumptions, the structural shocks (such as supply and demand shocks) explaining the historic development of output can be recovered. The output gap is calculated based on the absence of one or more shocks. SVAR-models need, however, a pre-testing procedure before evaluating the output gaps against the criteria listed

[21] In the DSGE-literature, a output gaps are often estimated using a combination of calibration and Bayesian techniques. See Section 2.3 for a discussion concerning the differences in interpretation of output gaps between traditional and the DSGE approaches.

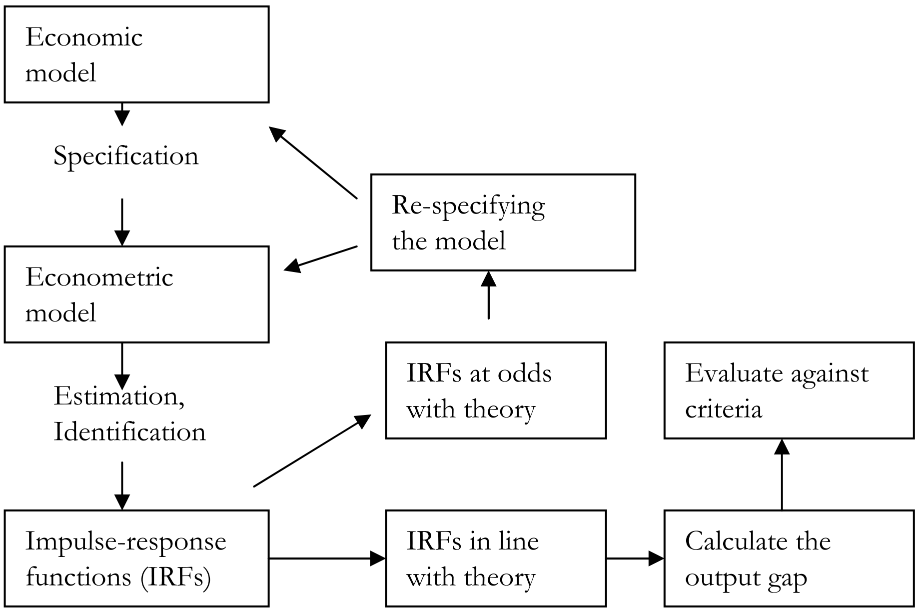
[22] As discussed and shown in Section 5.3, the MVHP-filter can be modelled in a variety of ways, depending on the preferences of the user. In this version of the paper, we just show one example of a MVHP-approach.

in Chapter 3. More specifically, we need to study the so-called impulse-response functions (IRFs). These tell us how the variables in the model respond to the identified structural shocks. If a response is at odds with theory, the model and/or the identifying assumptions are not valid. In this case, it is not worth while evaluating such a model against the criteria listed in Chapter 3. It is important to stress, however, that the process of accepting or disregarding models based on the IRFs is based on judgement and is hence somewhat arbitrarily; the identifying assumptions can not be tested statistically. We will therefore try to be as explicit as possible when evaluating the IRFs below and in Appendix.

It is also important to note that SVAR-models do not include an explicit Phillips-curve in the estimation. Hence, unlike the UC approaches outlined in Section 5.2 in this Chapter, we do not know whether the output gap from a SVAR-model correlate with inflation. Such an analysis is therefore carried out after the estimation, where the ability of forecasting inflation is one of the criterion listed in Chapter 3.

Schematically, we sum up the different stages of using SVAR-models for output gap estimation in Figure 3.

Figure 3: The SVAR-approach estimating output gaps



Among the SVAR-approaches we considered, five passed the IRF-test described above. The five passing the IRF-test are summerized below.[23]

SVAR 1

This 5-variable model comes from Fabiani et al (2001). The variables, x, the long run multiplier matix, $C(1)$, and the structural shocks, ε, of this model is given by:[24]

$$x = C(1)\varepsilon$$

$$\Leftrightarrow \tag{1.5}$$

$$\begin{bmatrix} \Delta\ln W \\ \Delta\ln Y \\ \Delta\ln P \\ U \\ WL/Y \end{bmatrix} = \begin{bmatrix} C_{11}(1) & 0 & 0 & 0 & 0 \\ C_{21}(1) & C_{22}(1) & 0 & 0 & 0 \\ C_{31}(1) & C_{32}(1) & C_{33}(1) & 0 & 0 \\ C_{41}(1) & C_{42}(1) & C_{43}(1) & C_{44}(1) & 0 \\ C_{51}(1) & C_{52}(1) & C_{53}(1) & C_{54}(1) & C_{55}(1) \end{bmatrix} \begin{bmatrix} \varepsilon^{P} \\ \varepsilon^{LS} \\ \varepsilon^{D} \\ \varepsilon^{TS_1} \\ \varepsilon^{TS_2} \end{bmatrix},$$

where W is real wage, Y is output, P is price level, U is unemployment rate and $W*L/Y$ is wage share. The structural shocks are labelled productivity (ε^P), labour supply (ε^{LS}), demand (ε^D) and two temporary supply shocks ($\varepsilon^{TS_1}, \varepsilon^{TS_2}$).[25] Trend output is defined as the output series that would have emerged in absence of demand and temporary supply shocks. That is, trend output is determined by productivity and labour supply shocks, both of which are allowed to have a long run effect on output. In terms of equation (1.5), $C_{21}(1)$ and $C_{22}(1)$ are unconstrained. One can note that as the wage share is included together with output and the real wage, it is possible to use this model to split the output gap in an labour market and productivity gap, respectively.

SVAR 2

This model is similar to SVAR 1, but four variables are included instead of five (see Fabiani et al, 2001). As a result, we get one temporary supply shock instead of two.[26] The system is:

$$\begin{bmatrix} \Delta\ln(Y/L) \\ \Delta\ln Y \\ \Delta\ln P \\ U \end{bmatrix} = \begin{bmatrix} C_{11}(1) & 0 & 0 & 0 \\ C_{21}(1) & C_{22}(1) & 0 & 0 \\ C_{31}(1) & C_{32}(1) & C_{33}(1) & 0 \\ C_{41}(1) & C_{42}(1) & C_{43}(1) & C_{44}(1) \end{bmatrix} \begin{bmatrix} \varepsilon^{P} \\ \varepsilon^{LS} \\ \varepsilon^{D} \\ \varepsilon^{TS} \end{bmatrix}, \tag{1.6}$$

where Y/L is labour productivity. Trend output is determined in the same manner as in SVAR 1. We use labour productivity instead of the real wage for two reasons. First, the IRFs look slightly better. Second, the inclusion of productivity implies that the model can split the output gap a labour market and productivity gap, respectively (see Section 2.1).

---

[23] The estimation of the SVAR-models start in 1976q1. IRFs are generally harder to interpret if starting in 1970q1.

[24] See Appendix, Section 8.1, for a description of the economic model behind the specification as well as technical details concerning identification of the structural shocks.

[25] This labelling of the shocks are later on confronted with data, i.e. the IRFs.

[26] As shown in Appendix, Section 8.1, it is somewhat hard to find a reasonable interpretation for the second temporary supply shock in the SVAR 1 model, which calls for testing a smaller system.

SVAR 3

Cerra and Saxena (2000) estimate the following 3-variable SVAR-model for Sweden:

$$\begin{bmatrix} \Delta \ln Y \\ \Delta \ln P \\ U \end{bmatrix} = \begin{bmatrix} C_{11}(1) & 0 & 0 \\ C_{21}(1) & C_{22}(1) & 0 \\ C_{31}(1) & C_{32}(1) & C_{33}(1) \end{bmatrix} \begin{bmatrix} \varepsilon^{PS} \\ \varepsilon^{D} \\ \varepsilon^{TS} \end{bmatrix}, \tag{1.7}$$

where ε^{PS} denote a supply shock that is allowed to have permanent effects on output. In this model, trend output is found by calculating the output series that would have emerged in absence of demand (ε^{D}) and temporary supply shocks (ε^{TS}). That is, trend output is determined by supply shocks allowed to have long run effects on output.

SVAR 4

This is the original Blanchard and Quah (1989) specification, used to calculate output gap by for example Cayen and van Norden (2005) and Funke (1997).

$$\begin{bmatrix} \Delta \ln Y \\ U \end{bmatrix} = \begin{bmatrix} C_{11}(1) & \\ C_{21}(1) & C_{22}(1) \end{bmatrix} \begin{bmatrix} \varepsilon^{PS} \\ \varepsilon^{D} \end{bmatrix}. \tag{1.8}$$

In this model, trend output is found by calculating the output series that would have emerged in absence of demand shocks. That is, trend output is determined by supply shocks allowed to have long run effects on output.

SVAR 5

We have also looked into some models with an explicit international linkage. It is important to stress, however, that the models above all have strong implicit international linkage. Take SVAR 2, for example. All shocks but the labour supply shock have in all probability an international ingredient. The series of demand shocks is probably highly influenced by unexpected movements in world demand as Swedish exports is more than half of GDP.

We have considered the models by Enders and Hurn (2007) and Albagli et al (2004) and only the former one pass the “IRF-test”, outlined above. Their model consists of three variables which in a Swedish setting are: OECD output, Swedish output and Swedish inflation.

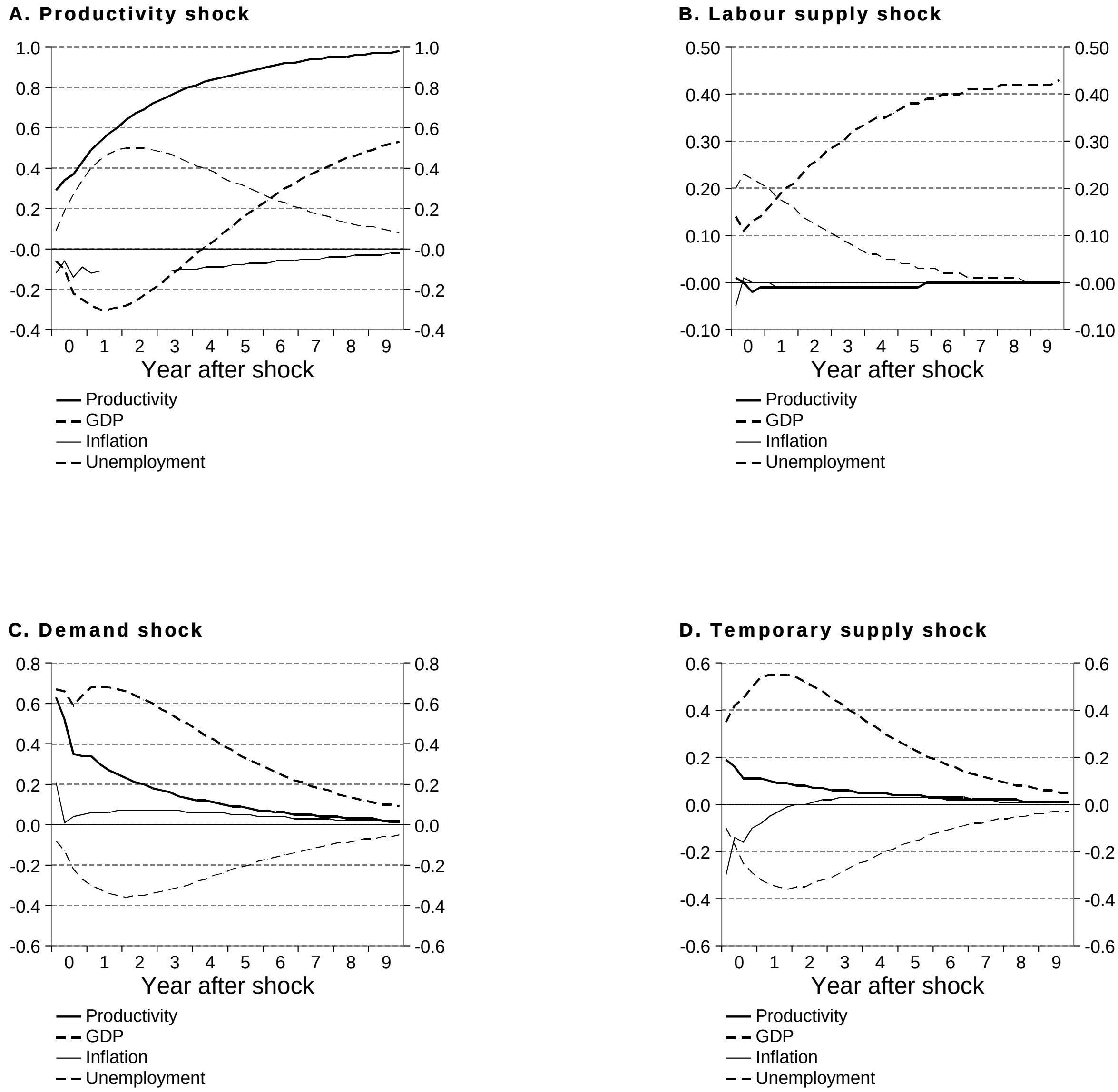
$$\begin{bmatrix} \Delta \ln Y^{OECD} \\ \Delta \ln Y^{SWE} \\ \Delta \ln P^{SWE} \end{bmatrix} = \begin{bmatrix} C_{11}(1) & 0 & 0 \\ C_{21}(1) & C_{22}(1) & 0 \\ C_{31}(1) & C_{32}(1) & C_{33}(1) \end{bmatrix} \begin{bmatrix} \varepsilon^{S,OECD} \\ \varepsilon^{S,SWE} \\ \varepsilon^{D} \end{bmatrix}, \tag{1.9}$$

In this model, trend output is found by calculating the output series that would have emerged in absence of demand shocks. That is, trend output is determined by supply shocks in the OECD as well as supply shocks in Sweden; both are allowed (but not restricted) to have long run effects on output.

EVALUATION OF IMPULSE-RESPONSE FUNCTIONS: AN EXAMPLE

In the Appendix, the impulse-response functions of the above models are shown. Here we exemplify our qualitative IRF-analysis by discussing the IRFs of the SVAR 2 model. As this model has four variables and four shocks, there are 4*4=16 IRFs to consider. They are shown shock-wise in Figure 4A–D.

Figure 4: Impulse-response functions of the SVAR 2 model



The first shock is labelled "productivity shock" and its dynamic effects are shown in Figure 4A. There are no restrictions on the effects that this shock has on the variables; neither in the short nor in the long run. Productivity responds positively and the long-run productivity level becomes higher due to the shock. The level of output responds negatively the first years but the long run effect is positive. Small or negative output responses to productivity shocks in the short run is a rather common finding in the literature, especially when including several shocks in the model allowed to affect output in the long run.[27] The rest of the responses of the productivity shock are in line with theory; higher productivity, lower inflation and higher unemployment.

The IRFs of the shock labelled "labour supply shock" are shown in Figure 4B. The only restriction is that the shock is not allowed to have any long run effects on the level of productivity. Output increases in both the short and the long run and unemployment rise in short and medium run. The movements in inflation and productivity are small; inflation is affected somewhat negatively in the very short run. In sum, the responses are arguably in line with the labelling of the shock.

Figure 4C shows the effects of the shock labelled as "demand shock". This shock is restricted not to have any long run effects on neither productivity nor output. Productivity, output and inflation are all affected positively by the shock in the short and medium run while unemployment falls. All these responses are arguably in line with the labelling of the shock.

Finally, Figure 4D shows the effects of the shock labelled as "temporary supply shock". This shock is restricted not to have any long run effects on neither productivity, output nor the price level. All responses are arguably in line with the labelling of the shock. A temporary positive supply shock increase productivity and output in the short and medium run while inflation and unemployment falls.

See Appendix, Section 8.1, for the IRFs of the other four SVAR models evaluated in the paper.

EXTRACTION OF STRUCTURAL SHOCKS ENRICH BUSINESS CYCLE ANALYSIS

One of the "soft" criterion listed in Chapter 3 above concerned the importance of being able to communicate, both internally and externally, the reasons behind the level and the changes of the output gap. Given the assumption of the empirical model and the labelling of the shocks, the time series structural shocks enable an exact explanation to both the level and the change in the gap when new information arrives. One can note that a larger SVAR model is potentially better as more structural shocks are recovered. One needs to be careful, however, so that the shocks make sense; data might not simply support a large model with many shocks. In short, as outlined above, the IRFs of the models considered must make sense.

OUTPUT GAPS OF THE FIVE SVAR MODELS

The SVAR-models are estimated using data 1976:1–2007:4. The output gaps are shown in Figure 5A–E. Two output gaps are shown for each model; one based on QRT data for the period 1976:1–2000:1, 1976:1–2000:2,…,1976:1–2007:4, and one based on FS data, 1976:1–2007:4. Descriptive statistics of these gaps are shown in Section 5.4 together with the output gaps of the other ap-

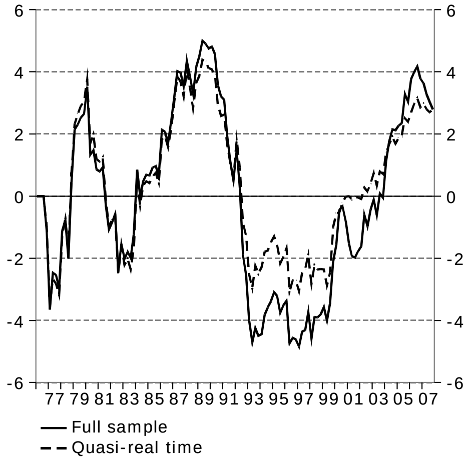
[27] See e.g. Amisano and Serati (2002), Blanchard and Quah (1989), Dolado and Jemeno (1997) and Fabiani et al (2001). One interpretation is that the model extracts productivity shocks not only due to variations in the speed new techniques are introduced in the production, but also due to variations in volume of entries and exits of firms with different productivity levels. For example, the model extracts positive productivity shocks during the crises in the 90s in Sweden which probably corresponds to the exit of firms with relatively low productivity. Output may decrease at first due to these exists while the positive effects dominates in the longer run.

proaches described below. Evaluation of the output gaps against the criteria listed in Chapter 3 is given in Chapter 6.

Figure 5: Full sample and quasi-real time (2000:1–2007:4) output gaps of the five SVAR-models

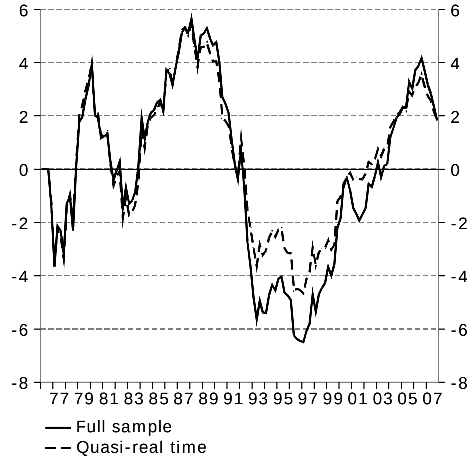
A. Output gap

Model: SVAR 1



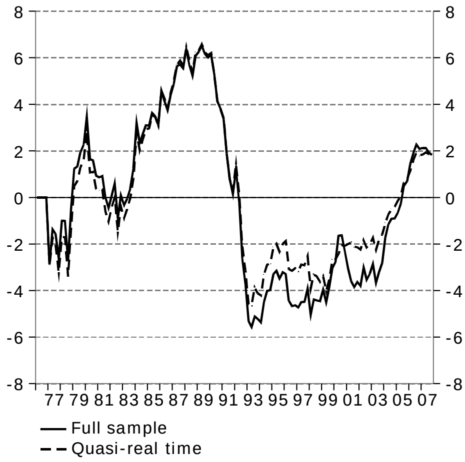
B. Output gap

Model: SVAR 2



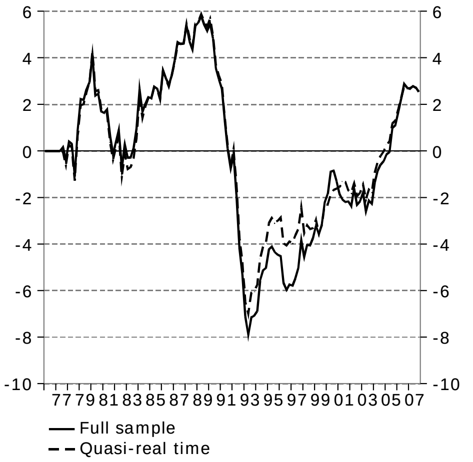
C. Output gap

Model: SVAR 3



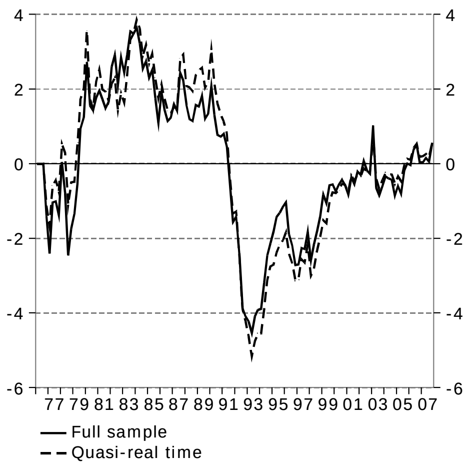
D. Output gap

Model: SVAR 4



E. Output gap

Model: SVAR 5



5.2. The unobserved components approach

In the previous section, the SVAR methodology was used to estimate the output gap. This approach use identified shocks to calculate trend levels by subtracting one or more shocks from the original series. However, it is possible to estimate the trend level in more direct way.

The fact that trend variables are inherently unobservable naturally points towards the use of unobserved components (UC) models in order to get an estimate of the potential levels. Broadly, two classes of UC models can be used for estimation of trend GDP, one class where the decomposition is based on the statistical properties of the series under consideration and one class of models where the estimation is based on macroeconomic relationships such as a Phillips curve and a Okun's law relationship. The two classes of models are described below.

This section is organized as follows. The section first includes a brief description of the building blocks of the UC models. A decomposition based on the statistical properties is presented first. Then, models including versions of the Phillips curve and the Okuns-law relationship are described. The modelling assumptions for the unobserved components are described next. The various versions of the UC models are then summarized and, finally, some estimation results are presented.

DECOMPOSITION OF A SERIES BASED ON STATISTICAL PROPERTIES

Several attempts to estimate trend GDP have originated from an UC approach that uses the statistical properties of the series under consideration in order to estimate the unobserved components. To mention one example, Watson (1986) uses an UC approach, with various assumptions on the trend component, in order to decompose (here) GDP into a permanent (or trend) component and a transitory (or cycle) component. One model considered by Watson (1986) uses the decomposition in (1.10) below.

$$\begin{aligned} y_t &= \tau_t + c_t \\ \tau_t &= \tau_{t-1} + \eta_t \quad \eta_t \sim N\left(0, \sigma_\eta^2\right) \\ c_t &= \alpha_1 c_{t-1} + \alpha_2 c_{t-2} + \varepsilon_t \quad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \end{aligned} \tag{1.10}$$

In (1.10), y_t is GDP which is decomposed into a trend component, τ_t, and a cyclical component, c_t. The two latter components are unobserved and the modelling assumptions for these are given by the last two equations in (1.10).

As it stands, the decomposition of the time series y_t does not relate the trend component or the cyclical component to any system of economic variables. In this sense the decomposition is similar to the HP filter. A loose theoretical foundation of the model can however be found using a simple multiplicator-accelerator model for GDP. This simple UC model will be evaluated in the current paper. However, the same principle approach will also be utilized in other models, but then in a context that allows the extracted components to occur macroeconomic relationships, such as a Phillips curve and Okun's law relationship (see below). By letting the estimated unobserved components occur in economics relationships, the estimates and economic interpretation of the components becomes more conformable with the notion of e.g. trend GDP or the NAIRU.

USING MACROECONOMIC RELATIONSHIPS IN UC MODELS TO ESTIMATE TREND OUTPUT AND THE NAIRU

Since trend variables such as trend GDP or NAIRU are inherently unobserved quantities, a natural approach when estimating them is to formulate an UC model that relate the unobserved variables to observed quantities such as unemployment and inflation. Two macroeconomic relationships, that link the unobserved quantities to observed variables, are often considered when estimating trend GDP. These relationships are a version of the Phillips curve and Okun's law, respectively. The former relationship is most often used to tie together observed inflation and the unemployment gap while the latter is used to establish relationship between the unemployment gap and the output gap. However, a Phillips curve relationship that directly ties together trend GDP and inflation has also been considered (see e.g. Kuttner, 1994).

It is worth emphasizing that the trend unemployment rate extracted using the UC models below, is called NAIRU (with no qualifying adjective). This is the same NAIRU concept used by international organisations like the OECD. In the literature, two other NAIRU concepts appear from time to time; *short-run NAIRU* and *long-run NAIRU* (see Richardson et al, 2000). It should be noted that the trend unemployment differs across model in terms of interpretation. More specifically, some of the NAIRU measures from the UC models capture a longer-term concept of equilibrium unemployment while measures from e.g. SVAR-models capture a shorter-term measure.[28]

A version of the Phillips curve

In order to extract a trend level of GDP, Kuttner (1994) suggests a model relating output gap, z_t, to the change in inflation, $\Delta\pi_t$, as in (1.11).[29]

$$\Delta\pi_t = \mu_\pi + \beta z_t + \varepsilon_t + \delta_1\varepsilon_{t-1} \tag{1.11}$$

In (1.11), ε_t and ε_{t-1} are error terms that captures the time series properties of inflation rate. A consequence of modelling inflation as a first difference is that the trend level of GDP that is extracted represents the level of GDP that is consistent with constant inflation. The output gap is then modelled, together with the Phillips curve expression, as in (1.10). The state space representation of the UC model is described in Appendix.

A large difference between this specification and the other specifications considered in this paper is that the relationship above relates the output gap to inflation, while the other UC models considered link inflation to the unemployment gap and the unemployment gap to the output gap. The unemployment-based Phillips curves are described next.

Another version of the Phillips curve

Besides being linked to the output gap, inflation is often modelled as a function of the unemployment gap, i.e. the deviation between the actual unemployment and the non-accelerating inflation rate of unemployment, NAIRU. By specifying the Phillips-curve in this way, the observable variables inflation and the unemployment rate can be used to estimate the unobserved component, i.e. the NAIRU. The underlying intuition of this approach is that of Phillips (1958), Phelps (1968) and Friedman (1968) but the model specification chosen in the current paper takes the form of the Gordon (1997) triangle model of inflation. The triangle model of inflation basically states that the

---

[28] It is, however, possible to obtain a short-term NAIRU from the UC models. See e.g. Batini and Greenslade (2003).

[29] Several specifications similar to the one considered by Kuttner (1994) was considered prior to the selection of the presented specification.

inflation rate can be modelled by three main components, lagged inflation, the unemployment gap and supply components. The specification used in the current paper is given in (1.12).

$$\pi_t = \Phi(L)\pi_{t-1} + \Gamma(L)\left(U_t - U_t^n\right) + \Upsilon(L)Z_t + \delta D_t + \varepsilon_t \qquad (1.12)$$

In (1.12), $\Phi(L)$, $\Gamma(L)$ and $\Upsilon(L)$ are lag polynomials that stipulate how lagged inflation, π_{t-1}, unemployment gap, $U_t - U_t^n$, and supply variables, Z_t, affect inflation contemporaneously and in the future. Furthermore, D_t in (1.12) is a dummy variable that is intended to capture the slowdown in inflation that occurred in Sweden in the middle of the 1990s. Finally, ε_t is a stochastic disturbance that captures inflation changes not accounted for by the lagged inflation terms, the excess demand terms, the supply variables and the dummy.

The lagged inflation terms in (1.12) are included in order to capture inertia in the inflation process. This inertia can be caused by several factors. A first factor is genuine autocorrelation in the inflation process due to e.g. price-adjustment delays caused by non-continuous price updating. Another factor, expressing economic behaviour, is expectation formation. Since it is possible that adaptive expectations are rational (see Muth, 1960), the lagged inflation terms in the Phillips curve could be a result of rational behaviour on behalf of the agents. Regardless of the causes of inertia, lagged inflation terms are included in the Phillips curve.

The inclusion of the labour-market tightness measure, $U_t - U_t^n$ where U_t^n is the NAIRU, allows for identification of the unobserved component, the NAIRU, by associating increases in inflation with a smaller unemployment gap.

In order for the Phillips-curve relationship in (1.12) to capture the effect that unemployment deviations from the NAIRU have on inflation, it is important to account for inflation changes that do not stem from a change in the unemployment rate or the NAIRU. Such inflation changes can for example be caused by supply shocks. Consider for example a shock to the oil price. Such a shock can cause prices to increase even though no effects on unemployment or the NAIRU have been realized. In absence of supply variables in the triangle model of inflation, supply shocks could imply a shift in NAIRU even though labour-market tightness has remained unchanged. Hence, there is an important role for variables that captures supply shocks in the inflation model in (1.12). Having concluded that supply variables are important to include in the inflation relationship, the choice of variables has to be specified.

Among the supply variables in the Phillips curve relationship, we include the change in the oil price (measured in Swedish crowns) relative core inflation (KPIX), the change in the relative price of imports[30] and the change in the real exchange rate weighted according to the KIX index. These variables are similar to the ones used by Apel and Jansson (1999).

Besides the variables that were mentioned above, an important question that has to be addressed is how to treat productivity. If variations in productivity growth mainly are due to temporary technology shocks not affecting the long run productivity level, the use of productivity growth as a supply proxy would be warranted in the Phillips Curve. A temporary productivity slowdown caused by a temporary technology shock would otherwise, if not controlled for, just like a temporarily higher oil price, increase the estimated NAIRU as inflation rise which is not desirable. If, on the other hand,

[30] The relative price of imports is calculated using the GDP deflator.

variations in productivity growth mainly are driven by technology shocks with permanent effects on the productivity level, one should not include productivity growth among the supply proxies. The reason for this is that the effects of these types of shocks are captured by the unemployment gap measure.[31] According the SVAR 2 model outlined above in Section 5.1, only a negligible part of the variation in productivity growth is due to temporary supply shocks while the lion's part is due to demand and permanent technology shocks. Hence, productivity is not included in the Phillips curve.

Using a Philllips curve in order to obtain an estimate of the NAIRU does not directly imply an estimate of trend GDP. In order to establish this link, Okun's law is often used.

Using Okun's law

To obtain an estimate of the trend level of GDP, NAIRU is liked to trend GDP through Okun's law. This relationship stipulates that the unemployment gap is inversely related to the output gap with the relative magnitude of one third (see e.g. Moosa, 1997). That is when the output gap increases with 1 percentage point, the unemployment gap decreases with approximately one third of a percentage point. The basic relationship, where the output gap is explained with the unemployment gap, is given in (1.13).

$$\begin{aligned} y_t^c &= \beta_1 U_t^c + \beta_2 U_{t-1}^c + \eta_t \Rightarrow \\ y_t &= y_t^p + \beta_1 U_t^c + \beta_2 U_{t-1}^c + \eta_t \end{aligned} \tag{1.13}$$

In (1.13), U_t^c denotes the cyclical component of unemployment as before while y_t^p denotes trend output. Based on the reasoning above, that a 1 percentage point change in the output gap reduces the unemployment gap by a third of a percentage point, we should expect to find that β_1 is approximately –3 (for further findings regarding the Okun coefficient, see e.g. Moosa, 1997).

Given the Phillips curve and the Okun's law relationship, a set of assumptions for the unobserved components is needed. The modelling assumptions of these quantities are presented next.

MODELLING THE UNOBSERVED COMPONENTS

In order to model the unobserved components, several approaches can be considered. For example, when estimating the NAIRU, Staiger, Stock and Watson (1997) considers approaches that take the unobserved quantity to be constant, constant between break points or evolving as a random walk. The latter approach is also taken by Apel and Jansson (1999) for trend GDP and the NAIRU. Since the random walk assumption for the unobserved components is rather flexible, this approach is maintained in the current paper. Hence, the trend level of output and the NAIRU are modelled as in (1.14) below.

$$\begin{aligned} y_t^p &= \alpha + y_{t-1}^p + v_t^{yp} \\ U_t^n &= U_{t-1}^n + v_t^{Un} \end{aligned} \tag{1.14}$$

From (1.14) it is apparent that trend GDP evolves according to a random walk plus drift while the NAIRU evolves according to a pure random walk.

[31] If permanent productivity shocks affect NAIRU positively or negatively is debated in the literature; see e.g. Ball and Mankiw (2002) and Blanchard and Katz (1997). For our purposes, the sign of the effect on the NAIRU is not relevant; we simly extract NAIRU from the data without any restrictions on the effects of various shocks.

In (1.14) it is assumed that the drift in trend GDP is constant through time. To relax this assumption, and allow for a time-varying drift in trend GDP, another specification for trend production is also considered. This specification is given by (1.15) below.

$$\begin{aligned} y_t^p &= \alpha_{t-1} + y_{t-1}^p + v_t^{yp} \\ \alpha_t &= \alpha_{t-1} + v_t^{\alpha} \end{aligned} \tag{1.15}$$

In the model specifications that are considered in this paper, Apel and Jansson (1999) are followed and the cyclical part of unemployment is modelled as an unobserved component that, together with the NAIRU, makes up total unemployment. An AR(2) specification, as in (1.16), is adopted for the cyclical part of unemployment.

$$U_t^c = \rho_1 U_{t-1}^c + \rho_2 U_{t-2}^c + v_t^{Uc} \tag{1.16}$$

SUMMARY OF THE UC MODELS

Given these specifications of above, the UC models can be cast in state-space form and estimated by using the Kalman filter. A complete description of the state-space form together with the estimates that are obtained are presented in Appendix, Section 8.2.

In Table 1, we give a summary of the UC models evaluated in the paper. In the Appendix, estimates and standard errors on central parameters of the various models are displayed.

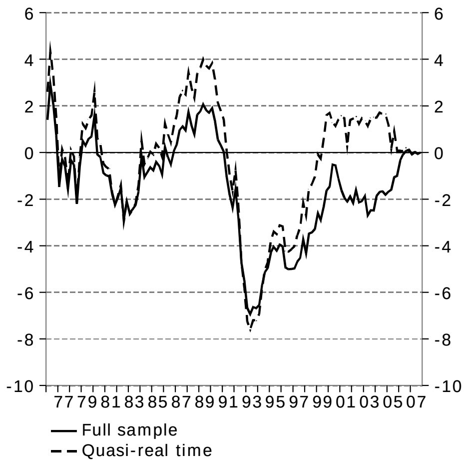
Table 1: Description of UC models.

	Description
UC 1	Watson (1986), trend GDP is a random walk plus time-varying drift (RW-TVD)
UC 2	Watson (1986), trend GDP is a random walk plus constant drift (RW-D)
UC 3	Kuttner (1994); see (1.11).
UC 4	Apel and Jansson (1999), inflation, trend GDP is RW-TVD.
UC 5	Apel and Jansson (1999), change in inflation, trend GDP is RW-TVD.
UC 6	Apel and Jansson (1999), inflation, trend GDP is RW-D.
UC 7	Apel and Jansson (1999), change in inflation, trend GDP is RW-D.

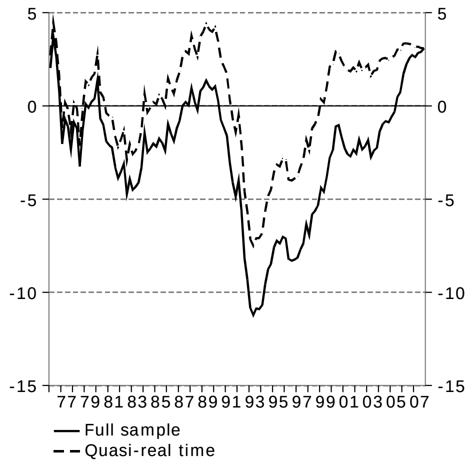
The output gaps of the seven UC models, using both QRT and FS data, are shown in Figure 6A–G.

Figure 6: Full sample and quasi-real time (2000:1–2007:4) output gaps of the seven UC-models

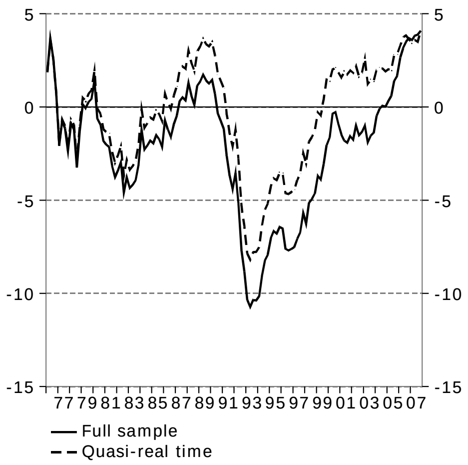
A. Output gap
Model: UC 1



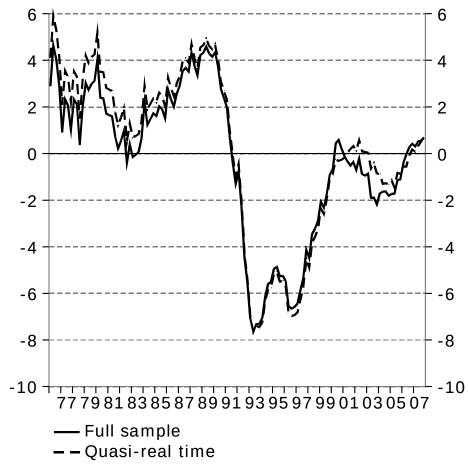
B. Output gap
Model: UC 2



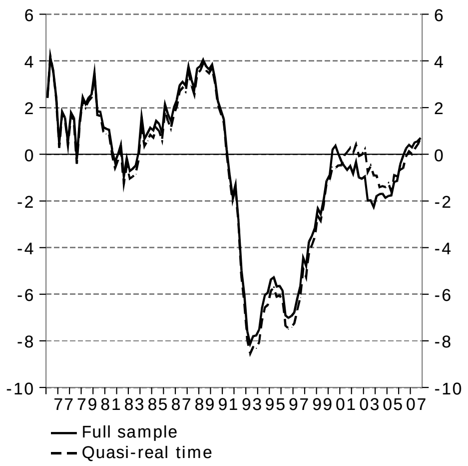
C. Output gap
Model: UC 3



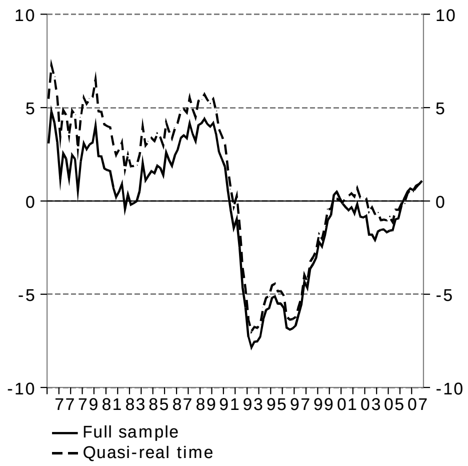
D. Output gap
Model: UC 4



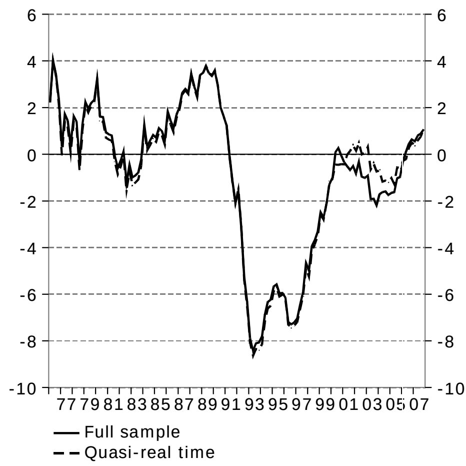
E. Output gap
Model: UC 5



F. Output gap
Model: UC 6



G. Output gap
Model: UC 7



5.3. Multivariate HP-filters

As discussed above, two serious drawbacks with the HP filter is its inability to incorporate economic information in the trend extraction procedure and the end-point problem, respectively. To overcome these drawbacks, Laxton and Tetlow (1992) suggested a methodology, the so-called multivariate Hodrick-Prescott (MVHP) filter, that incorporates economic information in the trend extraction process. More specifically, besides weighting the variability of the cyclical component and the change in the trend growth, these authors suggested that residuals from macroeconomic relationships that involve the output gap should be considered as well.[32]

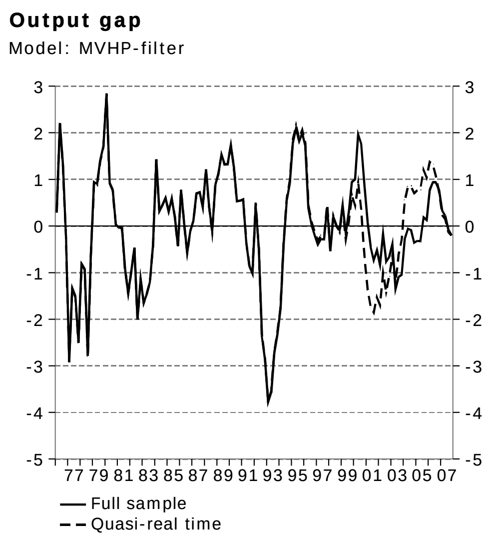
In order to incorporate economic information as suggested by Laxton and Tetlow (1992), three macroeconomic relationships are considered for inclusion in the MVHP filter. The first relationship is the connection between the output gap and a capacity utilization gap. Second, a Phillips curve type relationship is considered. This relationship states how inflation is related to the output gap. Finally, an Okun's law relationship, that relates the output gap to the unemployment gap, is included. These three relationships then give rise to three different residual series which are weighted together with the original HP loss function components in order to add up to a composite loss function.

The intuition underlying the MVHP is as follows. If the extracted HP gap for GDP is such that it renders large residuals in any of the three relationships described above, then the composite loss function for the MVHP filter increases and the extracted trend GDP is changed in order to mitigate the increase in the loss function. This will of course alter the contribution of the different components of the composite loss function, but optimization over the extracted trend component yields the trend GDP that taken together minimizes the composite loss function. A detailed description of the MVHP filter is given in Appendix, Section 8.3.

We consider only one type of MVHP filter in this paper; see Figure 7. As is evident from the above discussion (and the more detailed outline in Appendix, Section 8.3), the MVHP filter can be varied in numerous ways, both when it comes to which variables/relationships to include and the weighting scheme in the loss function. The method is hence well suited for users adding their preferred variables and weights. For example, it is straightforward to impose a time series of NAIRU and/or "normal" capital utilization in the optimization procedure and get an estimate of the output gap that is consistent with these priors. The preferences (as well as change in the preferences) of the user can be stated in a transparent way. We believe, though, that when deciding on the parameters of the MVHP filter, the decision should also depend on explicit criteria listed by the user.

[32] The MVHP-filters have become rather popular in recent years at policy institutions; see e.g. Antonicova and Hucek (2005), Benes and N'Diaye (2004), Bignasca and Rossi (2007) and Karagedikli and Plantier (2005) for applications.

Figure 7: Full sample and quasi-real time (2000q1–2007q4) output gap of the MVHP-model



5.4. Descriptive statistics of quasi-real time and full sample output gaps

In this section, we show some descriptive statistics of the output gaps shown in Figure 5–Figure 7. In Chapter 6, we evaluate the output gaps against the criteria listed in Chapter 3.

To begin with, Table 2 shows the correlation between the output gaps considered; correlation between full sample (FS) gaps is shown below the diagonal and quasi-real time (QRT) gaps is shown above the diagonal. First one can note the very high correlation between the UC 1–3 and UC 4–7 models, respectively. The SVAR 1–SVAR 4 models also display rather high correlation, while the correlation with SVAR 5 is lower. The MVHP filter is not highly correlated with any of the models.

Table 2: Correlation between full sample (below the diagonal) and quasi-real time (above the diagonal) output gaps

	S 1*	S2	S3	S4	S5	UC1	UC2	UC3	UC4	UC5	UC6	UC7	MVHP
SVAR 1	**1,00**	0,96	0,85	0,85	0,61	0,72	0,76	0,76	0,64	0,70	0,60	0,72	0,52
SVAR 2	0,97	**1,00**	0,89	0,90	0,73	0,74	0,76	0,75	0,73	0,78	0,70	0,79	0,46
SVAR 3	0,89	0,92	**1,00**	0,94	0,77	0,65	0,62	0,59	0,73	0,74	0,71	0,74	0,49
SVAR 4	0,90	0,94	0,96	**1,00**	0,85	0,76	0,74	0,71	0,86	0,87	0,85	0,88	0,48
SVAR 5	0,60	0,71	0,75	0,78	**1,00**	0,68	0,61	0,58	0,85	0,84	0,84	0,83	0,36
UC 1	0,83	0,86	0,83	0,92	0,69	**1,00**	0,96	0,93	0,81	0,89	0,77	0,90	0,48
UC 2	0,82	0,83	0,70	0,84	0,61	0,94	**1,00**	0,99	0,72	0,82	0,67	0,85	0,50
UC 3	0,82	0,82	0,67	0,82	0,58	0,91	1,00	**1,00**	0,68	0,79	0,62	0,82	0,48
UC 4	0,76	0,83	0,83	0,93	0,75	0,96	0,86	0,82	**1,00**	0,99	1,00	0,97	0,26
UC 5	0,77	0,84	0,82	0,93	0,75	0,97	0,88	0,85	1,00	**1,00**	0,97	1,00	0,30
UC 6	0,76	0,83	0,82	0,92	0,74	0,96	0,87	0,83	1,00	1,00	**1,00**	0,95	0,25
UC 7	0,78	0,84	0,81	0,92	0,74	0,97	0,90	0,87	0,99	1,00	1,00	**1,00**	0,32
MVHP	0,44	0,39	0,43	0,42	0,38	0,52	0,45	0,45	0,37	0,38	0,35	0,37	**1,00**

Remark: "S1", "S2",…,"S5" stand for "SVAR1", "SVAR2",…, "SVAR5". The time period here for comparison is 1976q1–2007q4 for all models.

Table 3 shows the averages of the QRT and FS gaps together with the variance of trend GDP growth. The average gaps of the SVAR-models are zero in the limit but can differ from zero in small samples. This is confirmed in Table 3; all SVAR-specifications (QRT and FS) have relatively small asymmetries.

The UC-models are not constrained to be zero on average within the sample, and there are also rather large differences between the averages among the UC-specifications. First one can note that the asymmetry differs between the QRT and FS specifications. The FS averages of UC 1–3 are negative and of considerable magnitude, whereas the asymmetry for UC 4–7 are greatest when applying QRT data. UC 4 has the lowest asymmetry among the UC models when considering FS and QRT data together. Finally, the MVHP model is constrained to be zero on average in large samples which is confirmed in the table.

Concerning the variance of trend GDP growth, it is clear from Table 3 that the SVAR-models have higher variance than UC and MVHP models. Greater variance in trend output implies, by definition, lower variance in the output gap.

Table 3: Descriptive statistics of full sample (FS) and quasi-real time (QRT) output gaps.

	Average, FS	Average, QRT	Variance, FS	Variance, QRT
SVAR 1	-0,05	0,34	0,12	0,19
SVAR 2	-0,05	0,36	0,12	0,18
SVAR 3	-0,18	0,09	0,09	0,13
SVAR 4	-0,20	0,06	0,18	0,20
SVAR 5	-0,06	0,02	0,38	0,43
UC 1	-1,52	-0,31	0,04	0,10
UC 2	-2,61	0,24	0,00	0,02
UC 3	-2,16	-0,37	0,00	0,03
UC 4	-0,28	0,13	0,04	0,07
UC 5	-0,67	-0,80	0,04	0,07
UC 6	-0,32	0,89	0,05	0,06
UC 7	-0,83	-0,85	0,05	0,06
MVHP	-0,03	-0,05	0,08	0,10

Remark: **Average** denotes the average value of the respective gaps. **Variance** denotes the variance of trend output growth of the respective models. Higher variance in trend output implies lower variance of the output gap. The time period here for comparison is 1976q1–2007q4 for all models.

6. Evaluation of the output gaps

In this chapter, the quantitative criteria listed in Chapter 3 are evaluated. We start in Section 6.1 by presenting the method used for evaluating the inflation and growth forecast performance. The results are given in Section 6.2. In Section 6.3, we evaluate the gaps against the remaining quantitative criteria. We choose our preferable model in Section 6.4. In Section 6.5 we show some further properties of the chosen model.

6.1. Inflation and growth forecast contest: Method

As discussed in Chapter 3, two important criteria for discriminate between the different output gaps are their ability to forecast inflation and GDP growth. In this section, we describe the method by which the forecast performance is to be evaluated.

First, it is important to note that the exercise here is not to try to find "the best" model of forecasting inflation. For such a task, one probably needs to include unit labour costs, exchange rates, inflation expectations etc. Instead, we carry out a comparative analysis in which the *relative* merits of the different output gaps are evaluated. It is though worth noting that the ability of output gaps to predict future inflation has decreased the last ten, fifteen years, perhaps due to the lower variability of inflation; see e.g. Fisher et al (2002) and Orphanides and van Norden (2004).

BENCHMARK SPECIFICATION

In the following, we discuss inflation but the same principles carry over to the growth forecast contest as well, which we comment on in the end of this section. We compare our output gap specifications with the following simple autoregressive model of inflation:

$$\pi_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i D_{i,t} + \sum_{j=1}^{p} \beta_j \pi_{t-j} + \varepsilon_t, \tag{1.17}$$

where π is inflation, measured at quarterly basis. Two definitions of inflation are included in the analysis; core inflation (*KPIX*) and core inflation excluding energy prices (*KPIXEE*). D_i is a dummy variable[33] and ε is an error term. Two lags, $p = 2$, are applied for both inflation rates based on LM-tests of both the whole sample (1976q1–2007q4) and the shortest sample considered in the forecast competition (1976q1–1999q4).

Three forecast horizons are considered; 2, 4 and 6 quarters ahead. Forecasted values are subsequently included in equation (1.17). For example, when forecasting *t+n* the forecast equation becomes:

$$\widehat{\pi}_{t+n} = \widehat{\alpha}_0 + \sum_{i=1}^{k} \widehat{\alpha}_i D_{i,t+n} + \sum_{j=1}^{p} \widehat{\beta}_j \widehat{\pi}_{t+n-j}, \tag{1.18}$$

[33] For the two types of inflation measures considerd (core inflation and core inflation excluding energy), a single dummy is applied for capturing the change in exchange rate regime in 1992q3 (i.e. $k = 1$ in equation (1.17)). The dummy is zero 1976q1–1992q2. It is equal to 1/16 1993q3 and increase by 1/16 every quarter until it reaches one 1996q3. This is to approximate the succussive shift in inflation and inflation expectations due to the new exchange rate and inflation target regime in the first half of the 1990s.

where " ^ " indicates estimated parameters and forecasted variables, respectively.

ADDING OUTPUT GAPS TO BENCHMARK SPECIFICATION

The output gaps are included in two related ways in order to investigate their ability to predict future inflation. First, building on equation (1.17), it is simply included in levels:

$$\begin{aligned}\pi_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i D_{i,t} + \sum_{j=1}^{p} \beta_j \pi_{t-j} + \\ \chi\left(y_{t-s} - y^*_{t-s}\right) + \varepsilon_t,\end{aligned} \quad (1.19)$$

where y is the natural logarithm of GDP, y^* is the trend counterpart; $s = 2,3,...,6$ are tested.[34] Second, we also test to include the change in the output gap due to the potential existence of so-called speed-limit effects (see Lown and Rich, 1997, and Ross and Ubide, 2003, on this). That is, building on equation (1.19) we have:

$$\begin{aligned}\pi_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i D_{i,t} + \sum_{j=1}^{p} \beta_j \pi_{t-j} + \\ \chi\left(y_{t-s} - y^*_{t-s}\right) + \delta\left[\left(y_{t-s} - y^*_{t-s}\right) - \left(y_{t-s-r} - y^*_{t-s-r}\right)\right] + \varepsilon_t,\end{aligned} \quad (1.20)$$

where $r = 2,3,4$ are tested. Hence, each output gap is tested using 20 combinations for each forecast horizon ($n = 2,4,6$).[35] That is, we test the models both with and without the inclusion of a "change in the gap" variable. The combination which gives the lowest forecast error on average for each horizon is chosen for a particular model and compared with the counterparts of the competing models.

PRINCIPLES OF THE FORECAST CONTEST

The evaluation period is 2000q1–2007q4. The model in question, (1.17), (1.19) or (1.20), is first estimated for the period 1976q1–1999q4. Then forecasts of 2, 4 and 6 quarters ahead are calculated. For inflation, the principles behind updating of right-hand-side variables were shown in equation (1.18) above. Due to the lag structure, actual values of the output gaps can be used in many cases when calculating the inflation forecasts. In some cases, however, when the forecast horizon is greater than the lag length of the output gaps (i.e., using the above notation, when $n > s$), forecasts of the output gaps are needed. These are based on the estimated AR(1)-model so that the forecast of the output gap equals:[36]

$$\left(y_{t+n} - y^*_{t+n}\right) = \hat{\phi}\left(y_{t+n-1} - y^*_{t+n-1}\right). \quad (1.21)$$

---

[34] Coincident output gap ($s = 0$) and output gap lagged once ($s = 1$) are not considered as we search for an output gap that can help policy makers stabilising the economy. As a result, we want the output gap to signal inflation pressures with at least two quarters notice.

[35] For each model we first have 5 combinations ($s = 2,3,...,6$) when the change in the gap is excluded. In addition, we have 15 combinations when the change in the gap is included, i.e. all combinations of $s = 2,3,...,6$ and $r = 2,3,4$ are considered.

[36] Alternatively, one can use forecast of the output gap from the different models applied in the paper. See Appendix, Section 8.4, for results.

Then, the coefficients of the inflation model in question is re-estimated up to 2000q1, including a re-estimation of the output gaps (i.e. using QRT-data of the gaps) followed by inflation forecasts up to 2, 4 and 6 quarters. This process continues up to 2007q2 for the two quarter forecast (forecasting 2007q4); up to 2006q4 for the four quarter forecast (forecasting 2007q4) and, finally, up to 2006q2 for the six quarter forecast (forecasting 2007q4). All-in-all, there are 29 two quarter forecasts, 27 four quarter forecast and 25 six quarter forecasts for each combination of explanatory variables considered.

FORECAST EVALUATION METRIC

We apply the Root Mean Squared Error (RMSE) for measuring forecast error. For model m and F number of forecasts, we have:

$$RMSE_m = \sqrt{\frac{1}{F}\sum_{f=1}^{F}\left(\hat{\pi}_{m,f} - \pi_{m,f}\right)^2}. \tag{1.22}$$

This metric is normalized by the RMSE of the benchmark model in equation (1.17), $RMSE_{bm}$, yielding a *relative RMSE* statistic which is lower than 1 if the output gap model outperforms the benchmark AR-model of inflation:

$$\text{Relative RMSE} = \frac{RMSE_m}{RMSE_{bm}} = \frac{\sqrt{\frac{1}{F}\sum_{f=1}^{F}\left(\hat{\pi}_{m,f} - \pi_{m,f}\right)^2}}{\sqrt{\frac{1}{F}\sum_{f=1}^{F}\left(\hat{\pi}_{bm,f} - \pi_{m,f}\right)^2}}, \tag{1.23}$$

where subscript bm stands for "benchmark", i.e. equation (1.17).

The above procedure is almost exactly the same for the GDP growth forecast contest. The only difference is that $s = 1, 2, ..., 6$; i.e. $s = 1$ is considered (see footnote 34).

SOME IMPORTANT PRINCIPLES OF THE EVALUATION OF THE MODELS

When evaluating the output gap measures, one of the quantitative criteria is output the ability of the gap to improve the inflation forecasts when used in a univariate AR model for inflation. This approach enables a comparison of the gap measures across the models used in the current paper. An alternative approach, that wouldn't enable such a comparison, would be use inflation forecasts that are produced directly from the models. As mentioned above, in actual forecasting much more detailed models of inflation are used. This would, however, limit the possible models to models that have an explicit inflation equation. Also, such an approach would put focus on the inflation forecasting ability of a specific model. Since the main purpose of this paper is to compare output gap measures from a wide variety of models, separate forecasting models for inflation are considered.[37]

[37] However, some guidance to the inflation forecasting ability of the SVAR models is given in Appendix, Section 8.4, where the inflation forecasts from these models are presented The reason for doing this exercise for the SVAR models is that these models have a fairly rich structure concerning inflation determination. This is not true for the UC and MVHP models.

WHAT ABOUT SIGNIFICANCE TESTS OF THE FORECAST CONTESTS?

Due to the fact that the benchmark model is nested in the other forecasting models considered, it is not straightforward to perform significance tests of superior forecasting ability. The use of significance tests could be useful in the current context since they can help in avoiding that wrong conclusions are drawn as a consequence of pure chance. However, since it is not straightforward to perform these significance tests, some other criterion that reduces the risk of drawing unwarranted conclusions is of interest.[38]

In Appendix, Section 8.4, we compare the forecasting performance of the chosen s and r - combination (all in all, 20 combinations) of each of the 13 models considered (see equation (1.20)) along with the *average* forecasting performance of the 20 different combinations of each of the 13 models, respectively. If the gap measure employed indeed provides information regarding the future inflation or GDP growth, then we of course expect to see that the performance of the best forecasting model is good relative to the baseline model considered. However, if the average performance over the 20 combinations is good as well, then it seems less likely that the favourable result is due to the random outcome that a specific model happens to perform good while the gap measure completely lack information regarding the future inflation and GDP growth. Hence, in the Appendix (Section 8.4) the forecasting performance of the chosen s and r - combination of each of the 13 models is presented together with the average forecasting performance over the 20 combinations of each model. It turns out that the best models concerning inflation and GDP forecasts also have good results for the 20 combinations tested of the respective models.

6.2. Inflation and growth forecast contest: Results

RESULTS OF THE INFLATION FORECAST CONTEST

The results when using core inflation are shown in Table 4 and ditto for core inflation, excluding energy prices, are shown in Table 5. Relative RMSE values calculated according to equation (1.23) are shown.[39,40]

Although results using both QRT and FS data are shown, most weight is put on QRT data (see the sum-up in Section 6.4). In short, the results in Table 4 and Table 5 can be summarized as follows:

- Adding output gaps to an AR-model of inflation in general improves the forecast performance of inflation; i.e. the relative RMSE values are below 1.
- Looking at the “average-columns”, it is clear that the output gap models perform better when excluding energy prices; i.e. the relative RMSE values are generally lower in Table 5 relative to Table 4.
- More information is not always better on average; the relative RMSE statistics are smaller as often as they are larger for the QRT gaps compared to the FS gaps; especially for core inflation, excluding energy (see the “Average”-column).

---

[38] The use of significance tests when ranking models is also debated in the literature, see Armstrong (2007). If the objective is to find the best model, it is rather simple; choose the one with the lowest relative RMSE. As outlined in Chapter 3, we consider several aspects/criteria when choosing model in this paper. These can not be evaluated and weighted using a formal significance test.

[39] The specific s and r chosen (see equations (1.19) and (1.20)) for each output gap can be received from the authors on request.

[40] In order to get a grip of the inflation forecasts of the gap-augmented AR-models (see Section 6.1), the forecasts of three best performing models are shown in Figure 8.

- Overall, the UC 4–7 perform somewhat better than the alternatives. The UC 1–3 models perform rather badly.
- Among the SVAR-models, SVAR 1–2 perform best.
- The MVHP-model performs rather badly; the relative RMSE statistics for the QRT gaps are rather close to one.

Table 4: Relative RMSE statistics based on quasi-real time (QRT) and full sample (FS) data. Dependent variable: Quarterly core inflation (*KPIX*). Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Horizon (quarters)							
	2		4		6		Average	
	QRT	FS	QRT	FS	QRT	FS	QRT	FS
SVAR 1	0,97	0,97	0,94	0,94	0,90	0,92	0,94	0,94
SVAR 2	0,98	0,98	0,95	0,96	0,91	0,93	0,95	0,96
SVAR 3	0,97	0,97	0,95	0,96	0,94	0,95	0,96	0,96
SVAR 4	0,96	0,96	0,96	0,95	0,94	0,93	0,95	0,95
SVAR 5	0,99	1,00	0,99	1,01	1,01	1,03	1,00	1,01
UC 1	0,97	0,96	0,98	0,94	0,97	0,89	0,97	0,93
UC 2	0,98	0,97	1,00	0,97	0,99	0,94	0,99	0,96
UC 3	0,98	0,97	1,00	0,97	0,99	0,94	0,99	0,96
UC 4	0,96	0,95	0,92	0,91	0,87	0,86	0,91	0,91
UC 5	0,96	0,95	0,92	0,92	0,88	0,87	0,92	0,91
UC 6	0,96	0,95	0,91	0,91	0,84	0,87	0,90	0,91
UC 7	0,96	0,95	0,93	0,92	0,89	0,88	0,93	0,92
MVHP	0,98	0,96	0,99	0,93	0,98	0,90	0,98	0,93

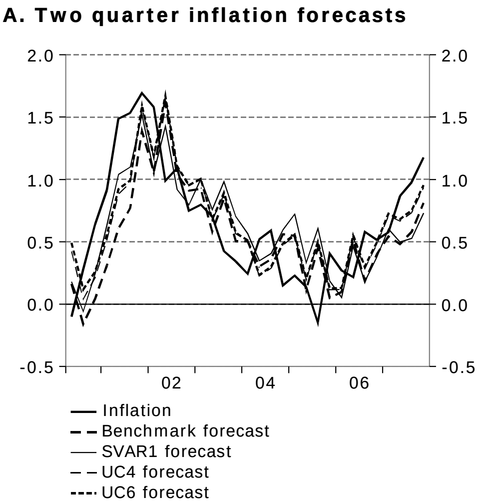
Remark: See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of inflation, see equation (1.17).

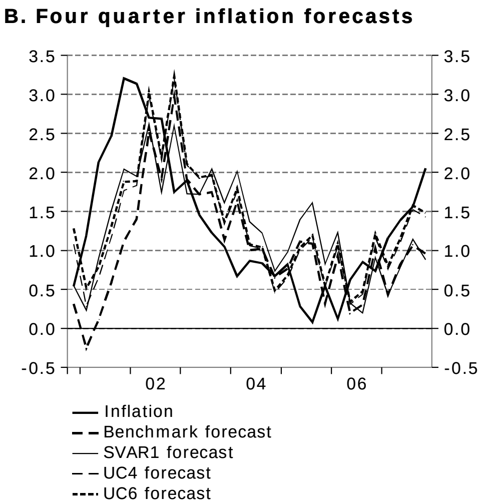
Table 5: Relative RMSE statistics based on quasi-real time (QRT) and full sample (FS) data. Dependent variable: Quarterly core inflation, excluding energy (*KPIXEE*). Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

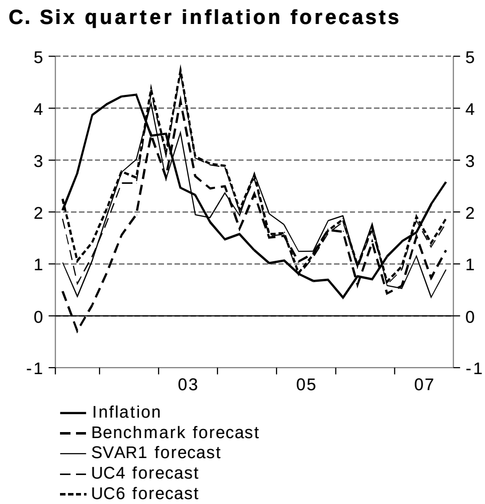
	Horizon (quarters)							
	2		4		6		Average	
	QRT	FS	QRT	FS	QRT	FS	QRT	FS
SVAR 1	0,89	0,93	0,86	0,90	0,84	0,88	0,86	0,90
SVAR 2	0,91	0,93	0,86	0,90	0,83	0,88	0,87	0,90
SVAR 3	0,98	0,99	0,95	0,96	0,92	0,95	0,95	0,97
SVAR 4	0,94	0,93	0,93	0,92	0,92	0,92	0,93	0,93
SVAR 5	0,96	0,99	0,94	0,98	0,94	0,99	0,95	0,99
UC 1	0,92	0,89	0,92	0,86	0,94	0,85	0,93	0,87
UC 2	0,93	0,89	0,94	0,88	0,97	0,88	0,94	0,88
UC 3	0,93	0,90	0,94	0,89	0,96	0,89	0,94	0,89
UC 4	0,86	0,86	0,82	0,83	0,81	0,84	0,83	0,84
UC 5	0,87	0,87	0,84	0,84	0,84	0,84	0,85	0,85
UC 6	0,85	0,86	0,79	0,82	0,76	0,82	0,80	0,83
UC 7	0,87	0,86	0,84	0,84	0,85	0,84	0,85	0,85
MVHP	0,96	0,90	0,96	0,87	0,96	0,86	0,96	0,88

Remark: See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of inflation, see equation (1.17).

Figure 8: Two, four and six quarter inflation forecasts of some models.







RESULTS OF THE GROWTH FORECAST CONTEST

The results of the growth forecast contest are presented in Table 6. Relative RMSE values calculated according to equation (1.23) are shown. As for inflation, most weight is put on results using QRT data. In short, the results in Table 6 can be summarized as follows:

- The SVAR 5 model performs well (0,92 in the "Average", "QRT" column). The SVAR 1–4 models perform rather badly; the relative RMSE statistics are above or just slightly below one. The same is true for the MVHP model when applying QRT data.
- The UC 4 model performs best (0,91 in the "Average", "QRT" column). However, as for the inflation forecasts above, the UC 4–7 models have very similar results.
- When using FS data, the MVHP model performs well. We believe, however, that one should be very cautious with this result. The MVHP-filter is a variant of the HP-filter. Using the HP-filter, or variants such as the MVHP-filter, to decompose a variable in a trend and a cycle component can introduce spurious cycles in the estimated gap (see e.g. Jaeger, 1994). Hence, caution should be taken when judging the HP-based gap measures based on the growth forecasting models.

Table 6: Relative RMSE statistics based on quasi-real time (QRT) and full sample (FS) data. Dependent variable: Quarterly GDP growth. Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Horizon (quarters)							
	2		4		6		Average	
	QRT	FS	QRT	FS	QRT	FS	QRT	FS
SVAR 1	1,04	1,03	1,06	1,06	1,10	1,11	1,07	1,07
SVAR 2	1,00	1,02	0,98	1,02	0,99	1,04	0,99	1,03
SVAR 3	1,00	1,00	0,99	1,00	0,99	1,00	0,99	1,00
SVAR 4	0,99	0,99	1,00	1,00	1,01	1,00	1,00	1,00
SVAR 5	0,96	0,99	0,92	0,99	0,89	0,99	0,92	0,99
UC 1	0,99	0,95	0,96	0,89	0,96	0,87	0,97	0,90
UC 2	1,00	0,99	0,97	0,97	0,98	0,98	0,98	0,98
UC 3	1,00	0,99	0,97	0,98	0,98	0,99	0,98	0,99
UC 4	0,95	0,94	0,90	0,87	0,88	0,83	0,91	0,88
UC 5	0,96	0,95	0,90	0,87	0,89	0,84	0,92	0,89
UC 6	0,96	0,95	0,91	0,88	0,91	0,85	0,93	0,89
UC 7	0,96	0,96	0,91	0,89	0,91	0,86	0,93	0,90
MVHP	1,03	0,91	1,03	0,85	1,05	0,77	1,03	0,85

Remark: See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of GDP-growth, see equation (1.17).

We do not suggest any formal weighting scheme for the results of the inflation and growth forecast contests in Table 4–Table 6. A simple arithmetic mean of the results in the "average" QRT-column in the three tables implies that UC models 4 and 6 are the best ones, but there are in principle no important differences between the UC 4–7 models. No SVAR-model performs among the best when considering both inflation and growth forecast. For example, SVAR 1–2 performs well concerning inflation forecast but performs badly for growth. The opposite is true for SVAR 5.

We now turn to the results concerning the remaining quantitative criteria. In Section 6.4 we sum up the results concerning both quantitative and qualitative criteria and choose our preferred model.

6.3. Evaluation of the remaining quantitative criteria

In addition to the inflation and growth forecast performance, three other quantitative criteria were listed in Chapter 3. All of them concerned the relationship between the QRT and FS gaps;

(i) size of revision,
(ii) the number of times the QRT gap 2000q1–2007q4 has the same sign as the FS gap, and, finally,
(iii) the number of times the change in the QRT gap 2000q1–2007q4 has the same sign as the *change* in the FS gap.

Everything else equal, policy makers are arguable more willing to use an output gap which is revised less than a competing gap. As UC 4–7 models were the best on average in the inflation and growth forecast contests (especially when using QRT-data), we pay special attention to these models when considering the remaining quantitative criteria examined in this section.

Table 7 concerns the above mentioned remaining quantitative criteria. The first column considers the correlation between the QRT and FS gaps of the respective models (although this is not a formal criterion). It is clear that the SVAR-models have higher correlation between QRT and FS gaps on average compared to the UC-models and the MVHP-model, although the difference to the UC 2–3 models is modest. Among the UC-models (4–7) that performed best in the inflation and growth forecast competitions in Section 6.2, UC 6 has highest correlation coefficient (0,86) and UC 5 the lowest (0,69).

The "Revision" column shows the root of the mean square revision divided by the standard deviation of the FS gap. It turns out that the SVAR 5 model get the lowest revision; 0,08. The UC-models (4–7) that performed best in the inflation and growth forecast competitions in Section 6.2, are very similar, ranging between 0,21–0,23.

The "Same sign" column shows the frequency with which the FS and QRT gaps have the same sign 2000q1–2007q4. Here it appears like the SVAR-models have higher frequencies than the other models, although UC 4–7 and MVHP have relatively high numbers compared to UC 1–3. Among the UC-models (4–7) that performed best in the inflation and growth forecast competitions in Section 6.2, UC 5 and UC 6 has slightly higher frequency coefficient (0,72) than UC 4 (0,63) and UC 7 (0,69).

The last column of Table 7 shows the frequency with which the FS and QRT gaps have the same sign of the change in the gap. Here SVAR 5 has the highest frequency (0,97) and UC 1 have the lowest (0,58). Among the UC-models (4–7) that performed best in the inflation and growth forecast competitions in Section 6.2, UC 6 has higher frequency coefficient (0,77) than UC 4 the lowest (0,61).

Table 7: Descriptive statistics of full sample and quasi-real time output gaps. Time period for comparison: 2000:1–2007:4.

	Correlation	Revision	Same sign	Same change
SVAR 1	0,97	0,33	0,78	0,77
SVAR 2	0,99	0,21	0,88	0,87
SVAR 3	0,97	0,27	0,97	0,87
SVAR 4	0,98	0,13	0,94	0,81
SVAR 5	0,95	0,08	0,91	0,97
UC 1	-0,73	1,25	0,19	0,58
UC 2	0,93	1,01	0,31	0,81
UC 3	0,95	0,71	0,44	0,81
UC 4	0,72	0,21	0,63	0,61
UC 5	0,69	0,22	0,72	0,68
UC 6	0,86	0,21	0,72	0,77
UC 7	0,76	0,23	0,69	0,74
MVHP	0,59	0,71	0,75	0,87

Remark: **Correlation** denotes the correlation between the full sample and quasi-real time gaps. **Revision** is the root of the mean square revision and the standard deviation of the full sample gap. **Same sign** denotes the frequency with which the full sample and quasi-real time gaps have the same sign. **Same change** denotes the frequency with which the change in the full sample and quasi-real time gaps has the same sign of the change in the gap. The time period here for comparison is 1976q1–2007q4 for all models.

As a final comparison between the QRT and FS gaps of the respective models, the bias and efficiency of the QRT estimates are evaluated in Table 8. The following equation has been estimated for each model:

$$GAP_t^{QRT} = \alpha + \beta * GAP_t^{FS}. \qquad (1.24)$$

If $\alpha \neq 0$, the QRT estimate is biased and if $\beta \neq 1$, the QRT estimate is inefficient (see Koske and Pain, 2008). Before considering the results, it is worth noting that the evaluation period considered, 2000:1–2007:4, is rather short.

As can be seen in Table 8, the QRT estimates of the majority of the models have a significant positive bias (see the first "t-value" column). One can note, however, that several UC-models (UC 4, 5, and 7) appear not to have any bias. Concerning efficiency, the results in the last column in Table 8 implies that the QRT estimates of all models (but MVHP) are, using standard confidence levels, inefficient. That is, when the FS gap alters one percentage point, the QRT gaps alter on average significantly less.

Finally, one can note that among the UC-models (4–7) that performed best in the inflation and growth forecast competitions in Section 6.2, only the UC 6 model has biased estimates (t-value of 3,58). The QRT estimates are, however, inefficient for all UC 4–7 models (see the second "t-value" column).

Table 8: Bias and efficiency of quasi-real time estimates

	Constant (α)	t-value ($\alpha = 0$)	Slope (β)	t-value ($\beta = 1$)
SVAR 1	0,67	11,73	0,58	-16,65
SVAR 2	0,59	12,19	0,70	-13,54
SVAR 3	0,21	2,41	0,72	-7,92
SVAR 4	0,16	2,33	0,93	-2,08
SVAR 5	0,03	1,33	0,86	-2,81
UC 1	0,23	1,59	-0,56	-16,60
UC 2	2,67	72,12	0,24	-42,27
UC 3	2,23	47,00	0,38	-27,39
UC 4	-0,02	-0,27	0,49	-5,97
UC 5	-0,09	-0,85	0,48	-5,58
UC 6	0,23	3,58	0,56	-7,24
UC 7	0,09	1,02	0,51	-6,07
MVHP	-0,07	-0,47	0,76	-1,22

Remark: See equation (1.24). The time period here for comparison is 1976q1–2007q4 for all models.

6.4. Choosing our preferred model

In this section, we choose our preferred model based on the evaluation of the quantitative criteria in Sections 6.2–6.3 above as the qualitative criteria listed in Section 3.1.

QUANTITATIVE CRITERIA

As discussed in Chapter 3, we think the inflation and growth forecast performance are the most important quantitative criteria, at least for an organisation like the NIER. We also believe that one should focus more on the results using QRT data as this type of data is more similar to the data at hand when analysing the state of the economy.

It is clear from Table 4 and Table 5 that the UC 4–7 models on average outperforms the competitors when it comes to *inflation forecast performance*. This is true for all horizons and both inflation measures considered when using QRT data. Among these four UC models, UC 6 is best, although the differences are small. One can also note that SVAR 1 and SVAR 2 have rather similar inflation forecast performance as UC 4–7 using QRT data.

When it comes to *growth forecast performance*, again, UC 4–7 is best on average and show rather similar results when using QRT data. UC 4 is best but the differences between are small among these four UC-models. One can note, though, that SVAR 1 and SVAR 2, which performed well in the inflation forecast, perform noticeable worse here. The opposite is true for SVAR 5; it performs well in the growth forecast but not in the inflation forecast.

Taking a simple average of the “average” QRT-columns in Table 4–Table 6, it turns out that UC 4 and UC 6 get the lowest value (0,88) while UC 5 and UC 7 get the second lowest value (0,90). It is clear that the UC 4–7 models over all perform best and that the differences between these models are small.

Focusing on UC 4–7, Table 7 shows that the QRT and FS gaps of the UC 6 model overall appears to be more similar (high "correlation", low "revision", high share of "same sign" and "same change"). The UC 6 model is however the only model among UC 4–7 which has a biased estimate of the QRT gap (see Table 8).

All-in-all, we interpret the results and comparison of the quantitative criteria as favouring the UC 4–7 models. Among these, the UC 6 model is perhaps the one that does best on average although the differences are small. It is worth noticing, however, that when inspecting the QRT and FS estimates of the output gap of the UC 6 model (see Figure 6F in Section 5.2), this model revises estimates 1976:1–1990:1 much more compared to, especially, the UC 5 and UC 7-models. Moreover, although the evaluation period in the paper has been 2000:1–2007:4, we have also looked at how the models behave up to the latest data point, 2009:2, which includes the period of severe financial and economic turbulence. Again, the UC 6 model (together with the UC 4 model) is rather volatile – both in terms of revision and in the parameter estimates. The parameters of the UC 5 and UC 7 models, on the other hand, are rather stable and significant during the period 2007:4–2009:2.[41] As this period is characterized by great macroeconomic turmoil, we see this as additional strength of these models.[42]

QUALITATIVE CRITERIA

As we think the quantitative criteria more important than the qualitative ones, we focus on the UC 4–7 models which performed best according to the quantative analysis above. Concerning the first criterion, "no end-point-problem", this is not an obvious problem for any of the UC-models.

The second criterion, "transparent, replicable, easy to communicate internally and externally", is of course not straightforward to analyze. First, it is important to note that the structure of the UC 4–7 models is similar at a "bird's eyes view". They all relate the difference between unemployment and NAIRU to inflation and they all relate the output gap to the unemployment gap via an Okun's law relationship (see Section 5.2 for more details). Due to these economic relationships, the output gap of the UC 4–7 models are all possible to communicate. Additionally, one can note that the Phillips curve specification of the UC 5 and UC 7 models are very similar to the ones used by the OECD (see Gianella et al, 2009, for a recent study).[43] This may improve transparency and comparisons.

When comparing the UC 5 and UC 7 models, the results are similar in many respects. Econometrically, the difference is that the former allows for a time-varying drift specification for trend GDP while the latter model assumes a random walk with constant drift specification. It turns out that the UC 7 specification has some problems with autocorrelation which is not the case for UC 5.

CHOSEN MODEL AND POSSIBLE FUTURE MODIFICATIONS

Based on the quantitative and qualitative assessment above, we choose UC 5 as our preferred model. As outlined above, there is a close race between UC 4–7 concerning the quantitative criteria. However, due to more stable and stronger significance of central parameters in the Phillips curve as well as less tendency to revise history, the UC 5 and UC 7 models are preferred to the UC 4 and UC 6 models. The choice between UC 5 and UC 7 is not that important as they are similar in struc-

---

[41] See the Appendix, Section 8.2, for parameter estimates and standard errors based on the evaluation period.

[42] The difference between the UC 4, UC 6 and UC 5, UC 7, is that the latter two impose dynamic homogeneity in the Phillips curve. As mentioned, this specification yields more stable and significant paramters, which is a favourable property; see Section 5.2.

[43] The Phillips curve in UC 5 and UC 7 assumes so-called dynamic homogeneity and models changes in inflation while UC 4 and UC 6 models the level of inflation.

ture and generate similar evaluation results. On econometric grounds, UC 5 may be slightly more preferable. Hence, based on the criteria and evidence of this paper, UC 5 is our choice.

It is important to emphasis, though, that in our procedure to choose a model above, we have kept the same model specification for the respective models the whole evaluation period (2000q1–2007q4). That is, we did not alter lag lengths, explanatory variables and so on. In the practical use of the chosen model, one might further elaborate with the model. For example, one could continuously check the model specification up to the latest data available. Perhaps other lag lengths and/or supply variables in the Phillips curve are suitable in the future. Moreover, the financial crises 2008–2009, including the historical drop in GDP 2008q4, affects estimates to a non-negligible extent. It might be worthwhile testing to account for this drop by inclusion of dummies. In short, when taking this model into practice in a forecasting environment, some "fine tuning" is probably needed. It is difficult, though, to forecast(!) the type of adjustments that might be worthwhile to carry out.

6.5. Some further properties of the chosen output gap

In the evaluation of the methods above, we focused on the period 2000q1–2007q4 and the different gaps were plotted for the period 1976q1–2007q4 in Figure 5–Figure 7. In Figure 9, the output gap of the chosen model is plotted up to (at the time of writing) the latest published data, 2009q2, together with 90% confidence bands. According to the chosen model, the output gap is significantly negative 2009q2. The uncertainty around the estimate is, however, large. Figure 10 shows the NAIRU estimate up to 2009q2. Figure 11 shows the close relationship between the output gap and the unemployment gap as implied by the model. In some periods, however, there are some differences, especially during the current downturn and the severe downturn in the 1990s. Due to the very rapid negative movements in output during these periods, we believe that companies did not have time to adjust the number of working hours accordingly. As a result, labour hoarding increased during these periods which can account for a part of the difference between the output gap and the unemployment gap.[44]

In Figure 12, the output gap on yearly basis (1990–2009) for the chosen model is plotted together with the counterparts of OECD, IMF and European Commission.[45] Figure 13 shows a comparison with current estimates of Swedish agencies. Finally, Figure 14 shows the corresponding estimates of trend growth of the different methods.[46]

[44] Increased labour hoarding is sometimes labelled as a "negative productivity gap".

[45] For the two unpublished quarters (2009q3–2009q4), the forecast from the respective organisation is applied. Differernces for, foremost 2009, can therefore depend on both the method applied and the forecasts.

[46] As shown in Hjelm (2010), the chosen output gap corresponds well to some standard survey-based measures of resource utilization, like capacity utilization and business confidence indicators.

Figure 9: Output gap of the chosen model 1990:1–2009:2

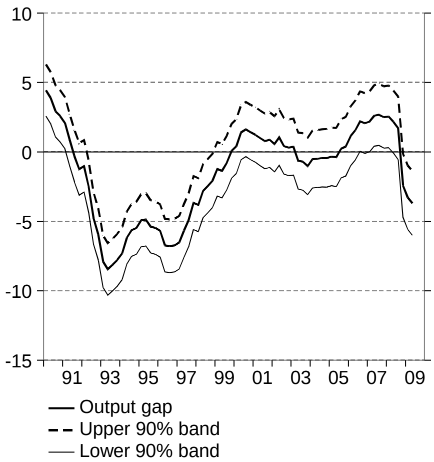


Figure 10: NAIRU of the chosen model 1990:1–2009:2

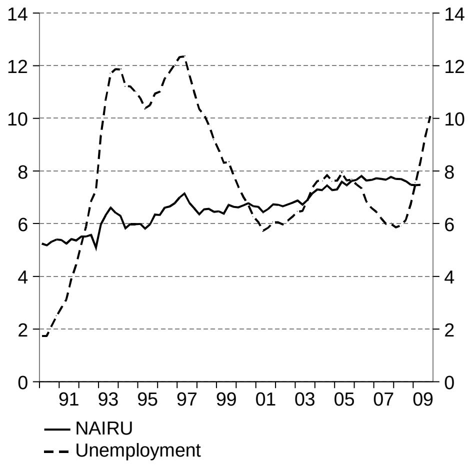


Figure 11: Output gap and unemployment gap of the chosen model 1990:1–2009:2

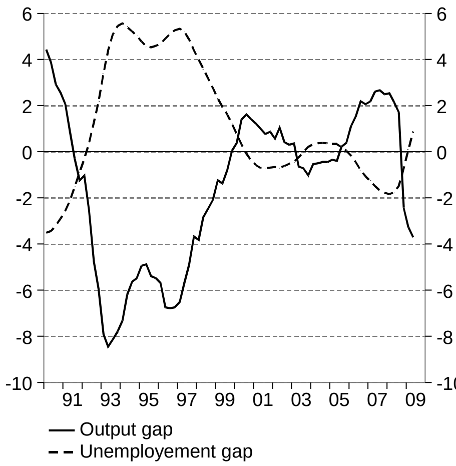


Figure 12: International comparisons of output gaps 1980–2009

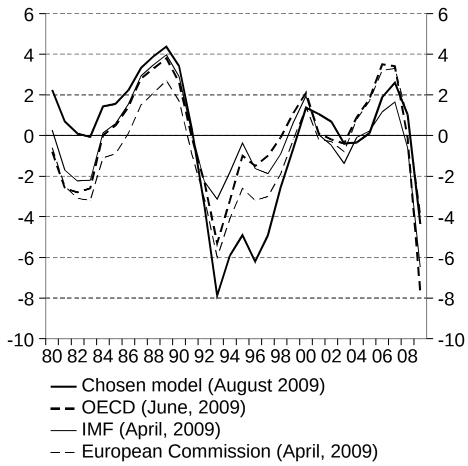


Figure 13: Swedish comparisons of output gaps 1980–2009

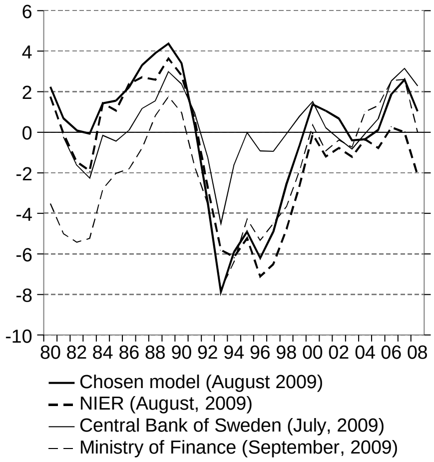
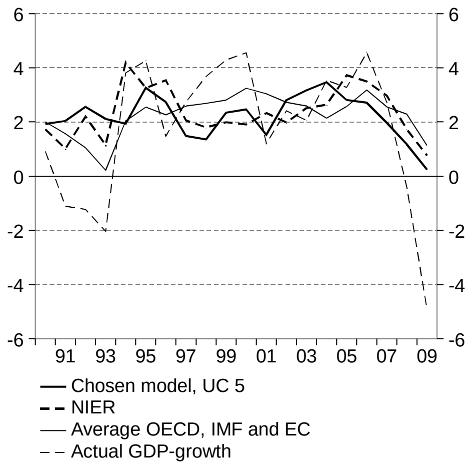


Figure 14: Trend GDP growth 1990–2009



Note: EC = European Commission.

7. Conclusions

Measuring the current stance of the economy is a difficult task. The estimates are in general very uncertain (i.e. the confidence bands are large) and it is often hard to discriminate between the merits of different approaches. Still, in all real time discussions concerning appropriate stabilisation policies, the question whether the economy is above or below some "trend" value arises. In many policy oriented organisations, the deviation from this "trend" is in focus; both in their internal process producing forecasts and in their external communication. In this paper, we put forward a measure of "trend" output (and, hence, an output gap) of the Swedish economy based on a number of criteria we believe are both relevant and useful when considering appropriate stabilisation policies.

The definition and extraction of this "trend" varies, both in the literature and in practice. We follow the "traditional" approach, which, in some way or another, tries to identify the level of output that is consistent with a stable inflation rate. Within this approach many different models and many different criteria have been put forward. We argue that it is reasonable that different organisations have different criteria when calculating their respective output gaps, as long as the method is transparent and replicable. We also note in the paper that explicit application of the rather new approach to output gaps put forward in the DSGE-literature still is limited, which calls for further use and development of the "traditional" approach.

We suggest a number of quantitative and qualitative criteria we believe is relevant and useful for discriminate between different methods measuring the output gap. Among these, we believe the ability of an output gap to improve forecasts of inflation and GDP growth should have the greatest weight, although we also consider the size of revision between the use of so-called quasi real time and full sample data.

We find that an unobservable components (UC) approach (called UC 5 in the paper) over all fulfil the criteria best relative to the alternative models considered in the paper, although it was a close race with some other UC models. It is important to note that our estimate of trend GDP (and hence our chosen output gap) is not "optimal" from any welfare point of view. It should simply be considered as the level of output which is consistent with a stable inflation rate. In the chosen model, trend GDP is extracted using two main economic relations; a Phillips curve and Okun's law. The model performs well compared to its competitors when it comes to inflation and growth forecasts. The revisions of the models trend GDP is also rather modest. Moreover, central parameters in the Phillips curve and Okun's law are rather stable and more significant compared to its competitors.

We believe that our preferred method could complement or substitute the current methods applied by national organisations. We also believe that our preferred method could complement the current methods applied by international organisations on Swedish data; "complement" and not "substitute" as an obvious important criterion of international organisations is to use the same method for all countries studied.

8. Appendix

Two main topics are outlined in this Appendix. First, in Sections 8.1–8.3, more technical and empirical details are given concerning the SVAR, UC and MVHP models. Second, in Section 8.4, we carry out various sensitivity analyses referred to in the text. More specifically, results from the following sensitivity analyses are presented.

- Inflation and growth forecast when using the average forecast of the 20 combinations tested for each of the 13 models.
- Results of inflation and growth forecast contest when using forecast of trend GDP from the models. (In the main text, we used a simple AR(1)-structure to calculate forecast of the output gap).
- Inflation forecast of the SVAR models.
- Combined forecasts: Do they perform better?

8.1. Structural VARs

As discussed in Section 5.1, we apply SVAR-models according to the scheme outlined in Figure 3; specification of economic model, estimation (including identification) of econometric model, evaluation of impulse-response functions (IRFs), calculation of output gap and, finally, evaluation of the output gap against criteria listed in Chapter 3. In this section, we first briefly present an economic model used to interpret shocks in some of the SVAR models applied in the paper. Then we derive the identification procedure used. Finally, we show the IRFs of SVAR-models 1, 3, 4 and 5 used in the paper; IRFs of SVAR-model 2 was shown in Figure 4A–D in the main text.

ECONOMIC MODEL

SVAR-models are generally not very well founded in theory. In the majority of applications of SVAR-models used for estimating output gaps, the theoretical basis is often only mentioned in passing. As a result, the types of shocks driving the results are more or less stipulated. We do not have a solid theoretical basis for our SVAR-models either. We do however, use the model by Fabiani et al (2001) outlined below, to interpret our identification scheme and our results.

The equations of the model in Fabiani et al (2001) are the following:

$$y_t = \phi\left(d_t - p_t\right) + a\vartheta_t \qquad \text{Demand} \qquad (1.25)$$

$$y_t = n_t + \vartheta_t \qquad \text{Supply} \qquad (1.26)$$

$$p_t = w_t - \vartheta_t + \beta u_t + \mu_t \qquad \text{Price setting} \qquad (1.27)$$

$$l_t = \alpha E_{t-1}(w_t - p_t - \vartheta_t) + \tau_t \qquad \text{Labour supply} \qquad (1.28)$$

$$w_t = E_{t-1}(p_t + \vartheta_t) + k_t - \sigma E_{t-1} u_t \qquad \text{Wage setting} \qquad (1.29)$$

$$u_t = l_t - n_t \qquad \text{unemployment (definition)} \qquad (1.30)$$

$$\vartheta_t = \vartheta_{t-1} + \varepsilon_t^s \qquad \text{Productivity shock} \qquad (1.31)$$

$$\mu_t = \lambda\mu_{t-1} + \varepsilon_t^m, \quad |\lambda| < 1, \qquad \text{Mark-up shock} \qquad (1.32)$$

$$\tau_t = \tau_{t-1} + \varepsilon_t^l \qquad \text{Labour supply shock} \qquad (1.33)$$

$$k_t = \rho k_{t-1} + \varepsilon_t^w, \quad |\rho|<1, \qquad \text{Wage-push shock} \tag{1.34}$$

$$d_t = d_{t-1} + \varepsilon_t^d \qquad \text{Demand shock} \tag{1.35}$$

A central assumption for wage setting in the model is that it is carried out in the beginning of the period; that is before the realization of all shocks, but ε^w. This is represented by the expectation operator "E" in equation (1.29)). Combining (1.27) and (1.29) gives the following expression:

$$E_{t-1}u_t = \frac{\lambda\mu_{t-1} + k_t}{\sigma - \beta}. \tag{1.36}$$

Using (1.36), (1.28) can be rewritten as:

$$l_t = \tau_t - \left(\frac{\alpha\beta}{\sigma-\beta}\right)\left(k_t + \lambda\mu_{t-1}\right). \tag{1.37}$$

Using (1.26) and substitute (1.25) for y and (1.27) for p generates the following expression for employment, n:

$$n_t = \phi\left(d_t + \vartheta_t - \mu_t - w_t - \beta u_t\right) + (a-1)\vartheta_t. \tag{1.38}$$

Equations (1.37) and (1.38) can now be substituted into (1.30) to get an expression for unemployment:

$$u_t = \left(\frac{1}{1-\phi\beta}\right)\left[\tau_t + \phi w_t - \left(\frac{\alpha\beta}{\sigma-\beta}\right)\left(k_t + \lambda\mu_{t-1}\right) - (a+\phi-1)\vartheta_t - \phi d_t\right]. \tag{1.39}$$

In order to express unemployment as a function of only parameters and exogenous processes, we have to solve for the nominal wage, w. This is achieved by taking expectations of (1.39), and let these equal (1.36) and then solve for w:

$$w_t = \left(\frac{1}{\phi}\right)\left[(a+\phi-1)\vartheta_{t-1} + \phi d_{t-1} + \left(\frac{\alpha\beta}{\sigma-\beta} + \frac{1-\phi\beta}{\sigma-\beta}\right)\left(k_t + \lambda\mu_{t-1}\right) - \tau_{t-1}\right]. \tag{1.40}$$

By putting (1.40) into (1.39), unemployment becomes a function of the parameters and exogenous processes of the model:

$$u_t = \left(\frac{1}{1-\phi\beta}\right)\left[\varepsilon_t^l - \phi\varepsilon_t^d - \left(a+\phi-1\right)\varepsilon_t^s\right] + \left(\frac{1}{\sigma-\beta}\right)\left(k_t + \lambda\mu_{t-1}\right). \tag{1.41}$$

If $|\rho|<1$, k is stationary which implies that unemployment is stationary. The nominal wage in (1.40) depends on productivity (ϑ), demand (d) and labour supply (τ) in the long run. It is now straightforward to derive how the rest of the variables in the models are affected by the five exogenous processes ((1.31)–(1.35)) in the long run.

The price level (p) is given by equation (1.27) and is in the long run determined by the same mechanisms as nominal wages, i.e. productivity, demand and labour supply. The price level is though not affected in the long run by the processes of labour market and mark-up shocks (k resp. μ) as they are stationary (see equation (1.38))

The real wage ($w-p$) is given by rewriting equation (2.3):

$$(w_t - p_t) = \vartheta_t - \mu_t - \beta u_t. \tag{1.42}$$

The real wage is according to (1.42) driven only by productivity (ϑ) as both mark-up (μ) and unemployment (u) is stationary.

Using equations (1.26) and (1.30), *GDP* (y) can be written as:

$$y_t = l_t - u_t + \vartheta_t. \tag{1.43}$$

According to (1.37), labour supply (l) is only driven by labour supply shocks (τ) in the long run. GDP is driven by these shocks as well as productivity (ϑ) as unemployment (u) is stationary.

The wage share ($W*N/P*Y$) can be written as:

$$\begin{aligned}
& w_t + n_t - p_t - y_t = w_t + n_t - (w_t - \vartheta_t + \mu_t + \beta u_t) - (n_t + \vartheta_t) \\
& = -\mu_t - \beta u_t \\
& \Leftrightarrow \\
& \frac{W_t * N_t}{P_t * Y_t} = \exp(-\mu_t - \beta u_t).
\end{aligned} \tag{1.44}$$

The wage share is hence stationary as both the mark-up (μ) as unemployment (u) is stationary.

The economic model outlined above implies ten long run restrictions. Labour supply shocks have no long run effects on real wages (restriction (i), see (1.42)); demand shocks have no long run effects on neither real wages, nor GDP (restrictions (ii)-(iii), see (1.42)–(1.43)); labour supply and mark-up shocks have no long run effects on neither real wages (restrictions (iv)-(v), see (1.42)), GDP (restrictions (vi)-(vii), see (1.43)) nor the price level (restrictions (viii)-(ix), see (1.27) and (1.40)); finally, mark-up shocks have no long run effects on unemployment (restriction (x), see (1.41)).

These ten restrictions are used in the SVAR 1 model. The SVAR 2 model builds on the economic model above but without mark-up shocks.[47] As is shown in Figure 15E below, the IRFs of the shock labelled as a "mark-up shock" in the SVAR 1 model are somewhat hard to interpret. It is hence not clear what type of shocks that is extracted from the data which motivates the SVAR 2 specification. As shown in Figure 4A–D in the main text, the IRFs of the SVAR 2 model are reasonable.

IDENTIFICATION

The reduced form shocks generated in the estimation of VAR models have no economic interpretation. By adding identifying assumptions (restrictions), the reduced form shocks become structural

[47] The first variable in this system is labour productivity, see equation (1.6) in the main text. Fabiani et al (2001) use here the real wage in their four variable application. The use of labour productivity, which yield very similar results, is govern by the potential use of deviding the output gap into a productivity and labour market gap (see discussion in Section 2.1).

enabling an economic interpretation. These structural shocks govern the calculation of the output gaps of the SVAR-models in the main text. For example, in the SVAR 5 model, trend output is given by the output series that would have been the case in absence of demand, wage-push and mark-up shocks. Below, we briefly describe how the identification is achieved.

The moving average representation of a VAR-model is given by:

$$\begin{aligned} x_t &= \gamma + R(L)\nu_t \\ &= \gamma + R_0\nu_t + R_1\nu_{t-1} + \ldots + R_\infty \nu_{t-\infty}, \end{aligned} \qquad (1.45)$$

where $R_0 = I$ and, for example, $x' = [\Delta(w-p) \quad \Delta y \quad \Delta p \quad u \quad WN/PY]$ in the SVAR 5 model. The structural VAR (SVAR) counterpart is:

$$\begin{aligned} x_t &= \gamma + C(L)\varepsilon_t \\ &= \gamma + C_0\varepsilon_t + C_1\varepsilon_{t-1} + \ldots + C_\infty \varepsilon_{t-\infty}, \end{aligned} \qquad (1.46)$$

Equations (1.45) and (1.46) imply a linear relationship between the reduced and structural forms:

$$\varepsilon_t = C(L)^{-1} R(L)\nu_t. \qquad (1.47)$$

We assume that there is a non-singular matrix S such that $S\varepsilon_t = \nu_t$. By comparing equations (1.45) and (1.46), we have that $C_0 = S, C_1 = R_1 S, C_2 = R_2 S$, i.e., $C(L) = R(L)S$. Hence, we only need to find C_0 to extract $C(L)$ (note that $R(L)$ is known from the estimation of the original VAR model). One can also note that:

$$R(1)C_0 = C(1), \qquad (1.48)$$

where $C(1) = C_0 + C_1 + \ldots + C_\infty$ (same analogy for $R(1)$). When estimating the VAR model, we get the variance-covariance matrix $E\left[\nu_t \nu_t^{'}\right] = \Sigma$. By assuming that the structural shocks are orthogonal and normalizing the variances to one, we have $E\left[\varepsilon_t \varepsilon_t^{'}\right] = I$. As $C_0\varepsilon_t = \nu_t$, we have that:

$$C_0 C_0^{'} = \Sigma. \qquad (1.49)$$

Combinding (1.48) and (1.49) we get:

$$C(1)C(1)' = R(1)\Sigma R(1)'. \qquad (1.50)$$

Let H be a lower triangular Choleski decomposition of the right hand side of (1.50) such that $C(1) = H$. By using (1.48), C_0 is given by:

$$C_0 = R(1)^{-1} H. \qquad (1.51)$$

The time series of structural shocks ν_t can then be calculated by using $C(L) = R(L)C_0$ and (1.47). These can then be used in equation (1.46) in order to calculate series including some shocks, excluding others.

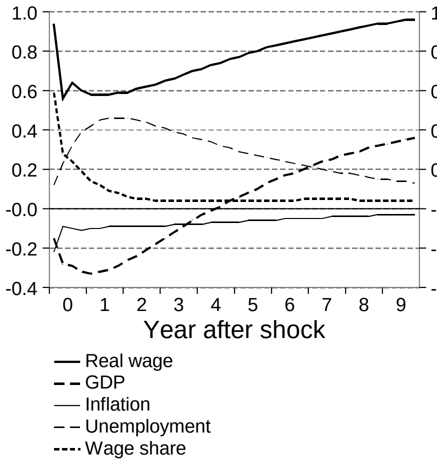
IMPULSE RESPONSE FUNCTIONS (IRFS)

In Figure 15–18 below, the IRFs for the SVAR 1, SVAR 3, SVAR 4 and SVAR 5 models are shown (see Section 5.1 for the IRFs of the SVAR 2 model).[48]

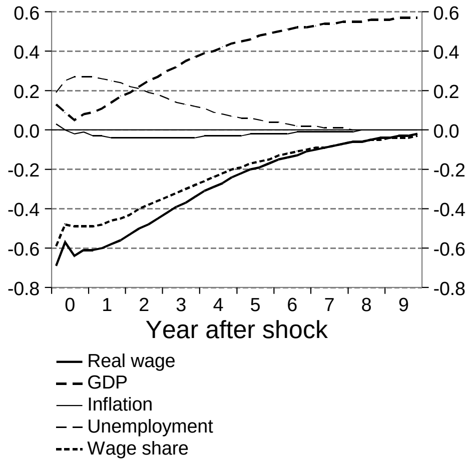
[48] In addition to the five SVAR-models discussed in the paper, we have evaluated several other SVAR-approaches. As the IRFs of these models did not make economic sense, these models were excluded from the paper.

Figure 15: Impulse-response functions of the SVAR 1 model[49]

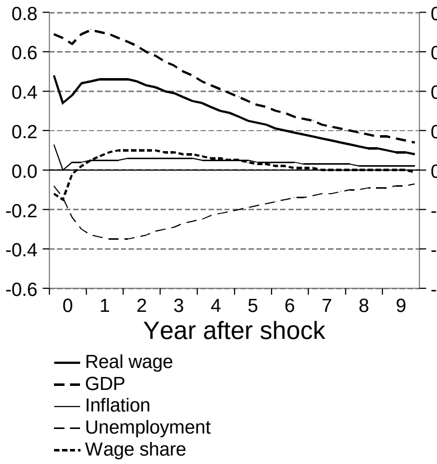
A. Productivity shock



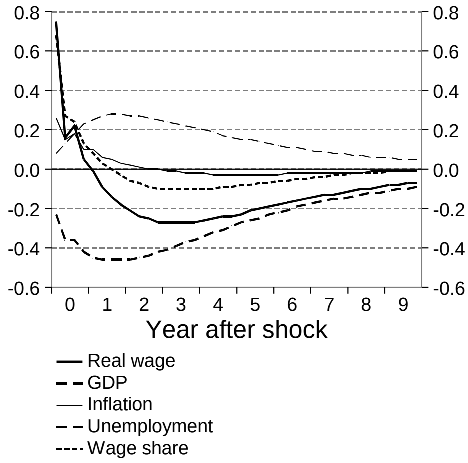
B. Labour supply shock



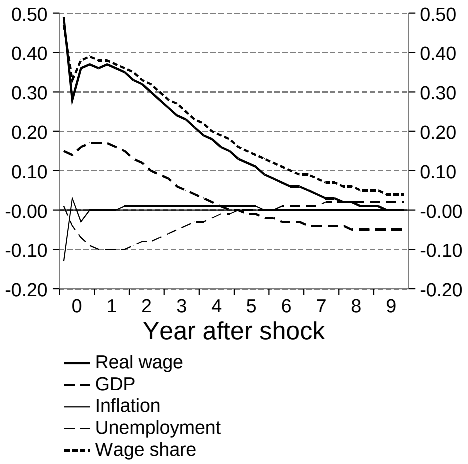
C. Demand shock



D. Wage push shock



E. (Negative) mark up shock(?)

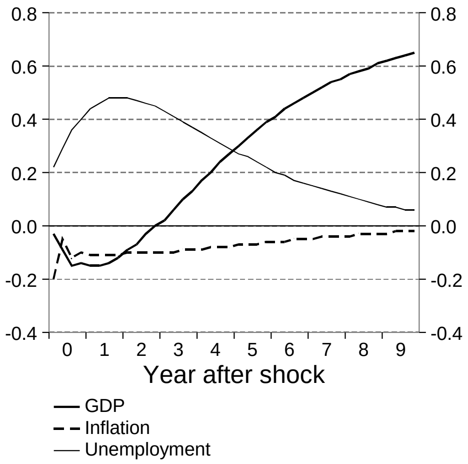


[49] According to the economic model, the fifth shock (shock "E" in the figure) is a (here, negative) mark up shock. The support from this labelling is though not that strong from the impulse-responses.

Figure 16: Impulse-response functions of the SVAR 3 model

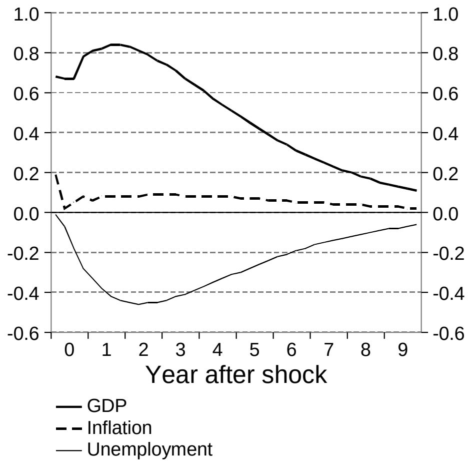
A. Produktivitety shock

Model: SVAR 3



B. Demand shock

Model: SVAR 3



C. (Negative) temporary supply shock

Model: SVAR 3

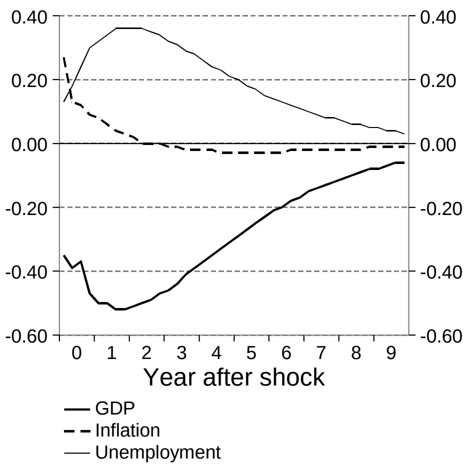


Figure 17: Impulse-response functions of the SVAR 4 model

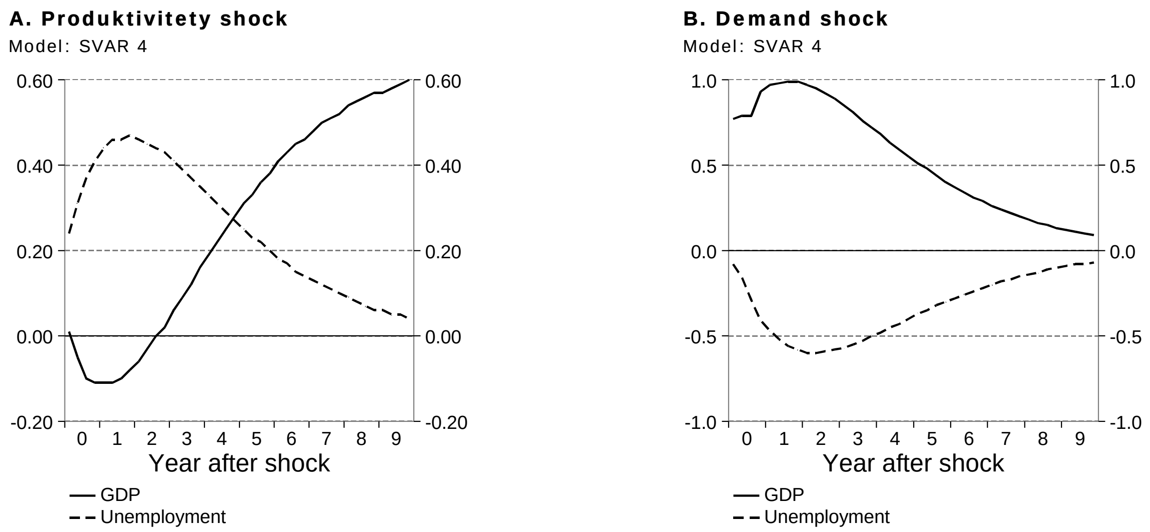
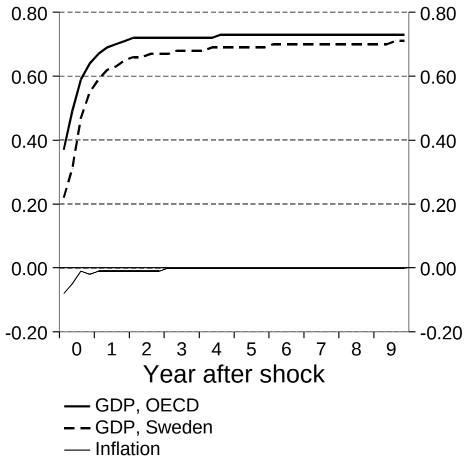
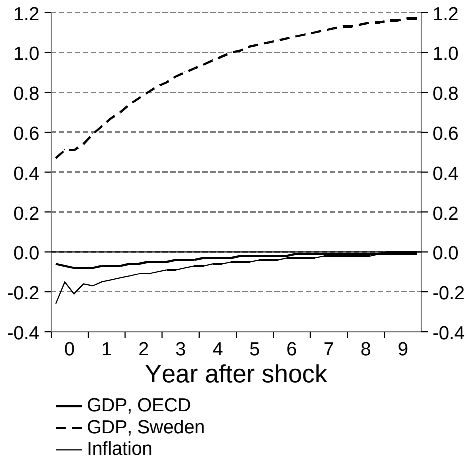


Figure 18: Impulse-response functions of the SVAR 5 model

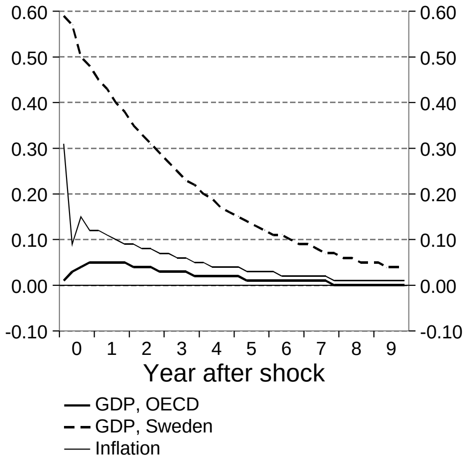
A. International supply shock
Model: SVAR 5



B. Swedish supply shock
Model: SVAR 5



C. Demand shock
Model: SVAR 5



8.2. The unobserved components models in state-space form

WATSON (1986)

The model that is used for extracting trend GDP using a purely statistical tool is given by the system in (1.52) below.

$$
\begin{aligned}
y_t &= y_t^p + y_t^c \\
y_t^p &= \alpha + y_{t-1}^p + v_t^{yp} \\
y_t^c &= \rho_1 y_{t-1}^c + \rho_2 y_{t-2}^c + v_t^{yc}
\end{aligned}
\tag{1.52}
$$

Letting y_t^p, y_t^c and y_{t-1}^c be the state variables that are organized into the 3×1 vector ξ_t, and letting $Y_t = y_t$, the state space formulation of the model is given by:

$$
\begin{aligned}
Y_t &= H\xi_t + e_t^m,\; e_t^m \sim N(0,Q) \\
\xi_t &= \tilde{\mu} + F\xi_{t-1} + e_t^s,\; e_t^s \sim N(0,R)
\end{aligned}
\tag{1.53}
$$

with

$$
H = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \qquad Q = 0
$$

$$
\tilde{\mu} = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix} \qquad F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho_1 & \rho_2 \\ 0 & 1 & 0 \end{bmatrix} \quad R = \begin{bmatrix} \sigma_{v^{yp}}^2 & 0 & 0 \\ 0 & \sigma_{v^{yc}}^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\tag{1.54}
$$

Given the state space formulation, the Kalman filter is used to obtain the maximum likelihood estimates of the parameters. The parameter estimates that are obtained when the model is estimated using data for the period 1970q1–2007q4 is presented in Table 9.

Table 9: Estimation results, central parameters, UC 1–UC 3 models

Parameter	UC1 (Watson, 1996)	UC2 (Watson, 1996)	UC3 (Kuttner, 1994)
β	-	-	0,001 (0,001)
ρ_1	0,77 (0,13)	0,81 (0,12)	0,78 (0,10)
ρ_2	0,23 (0,13)	0,17 (0,12)	0,19 (0,10)

Remark: Heteroscedastic consistent standard errors are given in parentheses for UC 1 and UC 2. Standard errors of UC 3 are computed using the cross-product of the Jacobian, due to problems calculating the covariance matrix using Heteroscedastic consistent standard errors. All calculations are using the Constrained maximum likelihood estimation package, version 2.0, in Gauss.

From the estimated parameters in Table 9 it is clear that the persistence of the cyclical component is very high, reflecting among other things the long-lasting economic slow-down during the 1990s.

KUTTNER (1994)

Instead of relying entirely on the statistical properties of the time series that is to be decomposed, Kuttner (1994) includes the output gap in a Phillips curve-type relationship. The unobserved components system that is the basis for the implementation of the Kuttner (1994) relationship is presented in (1.55) below:

$$\begin{aligned} y_t &= y_t^p + y_t^c \\ \Delta\pi_t &= \mu_\pi + \beta y_t^c + \varepsilon_t + \delta_1 \varepsilon_{t-1} \\ y_t^p &= \alpha + y_{t-1}^p + \nu_t^{yp} \\ y_t^c &= \rho_1 y_{t-1}^c + \rho_2 y_{t-2}^c + \nu_t^{yc} \end{aligned} \tag{1.55}$$

Putting the moving average term of the second equation in the state vector, and augmenting Y_t with $\Delta\pi_t$, the state space formulation of the model is obtained with by modifying the matrices in (1.54) accordingly.

APEL AND JANSSON (1999)

Given the unobserved components model that were considered above, a state-space reformulation is useful for estimation of the model parameters. To this end, once again consider the UC model from Section 5.2. Including an identity that simply states that the NAIRU plus the cyclical part of unemployment equals total unemployment, the following measurement equations are obtained:

$$\begin{aligned} U_t &= U_t^n + U_t^c \\ y_t &= y_t^p + \beta_1 U_t^c + \beta_2 U_{t-1}^c + \eta_t \\ \pi_t &= \Phi(L)\pi_{t-1} + \Gamma(L)\left(U_t - U_t^n\right) + \Upsilon(L) Z_t + \delta D_t + \varepsilon_t \end{aligned} \tag{1.56}$$

Now suppose that the three variables U_t, Y_t and π_t are organised into the 3×1 vector $Y_t = \left(U_t, y_t, \pi_t\right)'$ while the unobserved quantities y_t^p, u_t^n, u_t^c and u_{t-1}^c are organised into the 4×1 vector $\xi_t = \left(y_t^p, u_t^n, u_t^c, u_{t-1}^c\right)'$. Furthermore, suppose that the dummy variables together with the elements of $\Phi(L)\pi_{t-1}$ and $\Upsilon(L)Z_t$ are stacked into a $1\times K$ vector $X_t = \left(\pi_{t-1}, ..., \pi_{t-1-i}, Z_t, ..., Z_{t-j}, D_t\right)$ with $K = \text{i+j+3}$ where i and j are the orders of the lag polynomials $\Phi(L)$ and $\Upsilon(L)$. Assuming that the order of $\Gamma(L)$ is at most two and that disturbances are normally distributed, the linear state-space representation of this UC model is given by (1.57).

$$\begin{aligned} Y_t &= H\xi_t + AX_t + e_t^m,\ e_t^m \sim N(0,Q) \\ \xi_t &= \tilde{\mu} + F\xi_{t-1} + e_t^s,\ e_t^s \sim N(0,R) \end{aligned} \tag{1.57}$$

In (1.57), H, A, F, Q and R are parameter matrices that take the form given by (1.58).

$$H=\begin{bmatrix}0&1&1&0\\1&0&\beta_1&\beta_2\\0&0&\gamma_1&\gamma_2\end{bmatrix}\qquad A=\begin{bmatrix}0_{1\times K}\\0_{1\times K}\\\theta_{1\times K}\end{bmatrix}\qquad Q=\begin{bmatrix}0&0&0\\0&\sigma_\eta^2&0\\0&0&\sigma_\varepsilon^2\end{bmatrix}$$

$$\tilde{\mu}=\begin{bmatrix}\mu\\0\\0\\0\end{bmatrix}\qquad F=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&\rho_1&\rho_2\\0&0&1&0\end{bmatrix}\qquad R=\begin{bmatrix}\sigma^2_{\nu^{yp}}&0&0&0\\0&\sigma^2_{\nu^{Un}}&0&0\\0&0&\sigma^2_{\nu^{Uc}}&0\\0&0&0&0\end{bmatrix} \tag{1.58}$$

In (1.56), the parameters that enters for the lagged inflation terms, for the supply-side variables and for the dummy variable are collected in to the vector $\theta_{1\times K}$. Building on this baseline model, extensions that allow for time-varying drift in the growth of trend GDP are constructed.

Given the state-space representation of the UC model, estimation by maximum likelihood using the Kalman filter is straightforward. In Table 10, the essential parameter estimates are presented for the baseline model.

Table 10: Estimation results, central parameters, UC 4–UC 7 models

Parameter	UC4	UC5	UC6	UC7
β_1	-3,64 (0,76)	-3,68 (1,15)	-3,57 (0,79)	-3,66 (0,98)
β_2	2,54 (0,73)	2,60 (1,15)	2,47 (0,76)	2,56 (0,98)
γ_1	-0,13 (0,21)	-0,30 (0,21)	-0,11 (0,19)	-0,28 (0,12)
γ_2	0,10 (0,21)	0,29 (0,20)	0,07 (0,20)	0,27 (0,12)
ρ_1	1,87 (0,04)	1,88 (0,07)	1,87 (0,04)	1,88 (0,07)
ρ_2	-0,88 (0,05)	-0,89 (0,07)	-0,88 (0,05)	-0,89 (0,07)

Remark: Heteroscedastic consistent standard errors are given in parentheses for UC 5 and UC 7. Standard errors of UC 4 and UC 6 are computed using the cross-product of the Jacobian, due to problems calculating the covariance matrix using Heteroscedastic consistent standard errors. All calculations are using the Constrained maximum likelihood estimation package, version 2.0, in Gauss. γ_1 and γ_2 denote the first and the second term of the lag polynomial $\Gamma(L)$.

8.3. Multivariate HP-filter

In order to extract the MVHP trend component, the loss function, L, in (1.59) below is minimized with respect to the trend sequence τ_t.

$$
\begin{aligned}
& L=\sum_{t=1}^{T}(y_t-\tau_t)^2+\sum_{t=2}^{T-1}(\Delta\tau_{t+1}-\Delta\tau_t)^2+\gamma_1\sum_t\left(\varepsilon_t^{CU}\right)^2+\gamma_2\sum_t\left(\varepsilon_t^{PC}\right)^2+\gamma_3\sum_t\left(\varepsilon_t^{OL}\right)^2 \\
& \left(CU_t-CU_t^p\right)=\left(y_t-y_t^p\right)+\varepsilon_t^{CU} \\
& \pi_t=\sum_{i=1}^{3}\alpha_i\pi_{t-i}+\beta\left(y_t-y_t^p\right)+\varepsilon_t^{PC} \\
& \left(U-U_t^n\right)=\frac{1}{3}\left(y_t-y_t^p\right)+\varepsilon_t^{OL}
\end{aligned}
\tag{1.59}
$$

CU_t denotes rate of capacity utilization. The residuals that are obtained from the capacity utilization relationship, the Phillips curve and the Okun's law relationship, are obtained either by stipulating the parameters that occur in the expression or by estimation. For the capacity utilization relationship and the Okun's law relationship, the parameters are stipulated. The parameters chosen are those presented in (1.59). For the Phillips curve on the other hand, the parameters are estimated. Finally, the parameters that states how much weight the various residuals are to have in the composite loss function, i.e. that parameters γ_1, γ_2 and γ_3, are stipulated and set to two. Finally, the trend components of the capacity utilization rate and the NAIRU are obtained by applying the HP filter. In all of these choices, Laxton and Tetlow (1992) are followed. As mentioned in the main text, all these choices can be made by the user in a transparent way.

8.4. Sensitivity analysis

In this final section of the Appendix, we show results of some sensitivity analyses referred to in the main text. In the end of this section, we sum up the main conclusions from this exercise.

INFLATION AND GROWTH FORECAST WHEN USING AN AVERAGE OF THE 20 COMBINATIONS TESTED FOR EACH OF THE MODELS

As discussed in the main text (see Section 6.1) it is far from straightforward to perform significance tests of the inflation and growth forecast contests. In order to remedy this problem to, at least, some extent, we have calculated the average relative RMSE of the 20 combinations considered for each of the 13 models. By considering the average over all forecasting combinations for the respective models, it is possible to shed light on to what extent a superior forecasting ability is attributable to pure chance relative to a better gap measure. If the forecasting ability for a specific model, as well as the average forecasting ability over the 20 combinations for that model, are good then is likely that the gap measure contains information regarding the future development of inflation and growth. On the other hand, if the average forecasting ability over the 20 combinations produce forecasts that not as good, then it could be that the forecasting properties of the specific model are due to chance. Hence, the RMSE for the individual forecasting models are provided together with the average RMSE for the 20 forecasting models.

The UC 4–UC 7 models performed overall best when considering inflation and growth forecasts, see Section 6.2. The single best combination of each model is shown in the last column in Table 11, stemming from Table 4 in the main text. As can be seen in Table 11, the average of the 20 combinations of the UC 4–UC 7 models are all among the lowest, together with SVAR 3 and SVAR 4. This is also true for UC 4–UC 7 when considering core inflation, excluding energy; see Table 12.

Table 11: Relative RMSE statistics based on quasi-real time data.
Average of the 20 combinations tested for each model.
Dependent variable: Quarterly core inflation (*KPIX*).
Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Average of 2, 4 and 6 quarters	
	Average of 20 combinations	Best combination[1]
SVAR 1	1,00	0,94
SVAR 2	1,01	0,95
SVAR 3	0,99	0,96
SVAR 4	0,99	0,95
SVAR 5	1,04	1,00
UC 1	1,03	0,97
UC 2	1,04	0,99
UC 3	1,04	0,99
UC 4	0,99	0,91
UC 5	0,99	0,92
UC 6	0,98	0,90
UC 7	0,99	0,93
MVHP	1,01	0,98

Remark: See equation (1.23) for calculation of the relative RMSE statistics.
A value below 1,0 implies better forecast performance than an AR-model of inflation, see equation (1.17). [1]From Table 4.

Table 12: Relative RMSE statistics based on quasi-real time data.
Average of the 20 combinations tested for each model.
Dependent variable: Quarterly core inflation, excluding energy (*KPIXee*).
Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Average of 2, 4 and 6 quarters	
	Average of 20 combinations	Best combination[1]
SVAR 1	0,96	0,86
SVAR 2	0,97	0,87
SVAR 3	1,00	0,95
SVAR 4	0,98	0,93
SVAR 5	1,02	0,95
UC 1	0,99	0,93
UC 2	1,00	0,94
UC 3	1,00	0,94
UC 4	0,92	0,83
UC 5	0,94	0,85
UC 6	0,90	0,80
UC 7	0,96	0,85
MVHP	0,97	0,96

Remark: See equation (1.23) for calculation of the relative RMSE statistics.
A value below 1,0 implies better forecast performance than an AR-model of inflation, see equation (1.17). [1]From Table 5.

When it comes to the growth forecast evaluation, the broad picture is the same as above. The average of the 20 combinations of the UC4–UC 7 models is among the best together with SVAR 3 and SVAR 4 (see column "Average of 20 combinations" in Table 13). One can note, though, that the SVAR 5 model clearly outperform the competitors (value 0,98). The different lag structures of the 20 combinations in the SVAR 5 model do apparently not affect the growth forecast performance much as the value of the model's best combination is just slightly lower (0,92).

Table 13: Relative RMSE statistics based on quasi-real time data. Average of the 20 combinations tested for each model. Dependent variable: Quarterly GDP growth. Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Average of 2, 4 and 6 quarters	
	Average of 20 combinations	Best combination[1]
SVAR 1	1,45	1,07
SVAR 2	1,23	0,99
SVAR 3	1,09	0,99
SVAR 4	1,09	1,00
SVAR 5	0,98	0,92
UC 1	1,15	0,97
UC 2	1,18	0,98
UC 3	1,18	0,98
UC 4	1,09	0,91
UC 5	1,10	0,92
UC 6	1,09	0,93
UC 7	1,11	0,93
MVHP	1,16	1,03

Remark: See equation (1.23) for calculation of the relative RMSE statistics.
A value below 1,0 implies better forecast performance than an AR-model of GDP growth, see equation (1.17). [1]From Table 6.

RESULTS OF INFLATION AND GROWTH FORECAST CONTEST WHEN USING FORECAST OF POTENTIAL GDP FROM THE RESPECTIVE MODELS

As discussed in the main text, Section 6.1, we do not focus on the different models' ability to forecast trend GDP in our evaluation. Instead, we apply a simple estimated AR(1)-structure to carry out forecasts of the output gaps when needed in the forecast competitions. The main reason for this is that we put forward a measure of the output gap that primarily should be used to evaluate the current state of the economy (i.e. up to the latest data point published). Future developments of trend GDP is driven by several factors not explicitly included our models, such as demographics. This being said, it may be of a more general interest to investigate how the models perform in the inflation and growth forecast contests *if* forecasts of trend GDP (and thereby the output gap) of the respective models are used instead of the autoregressive ditto used in the main text.

The SVAR models could potentially benefit from the inclusion of output gap forecast as these models predicts a dynamic pattern of potential GDP growth, stemming from the permanent shocks up to the latest data point published. The forecast of potential GDP growth of the UC and MVHP models is, on the other hand, simply equal to the last estimated potential growth in the sample. As the AR(1) coefficient is high in the output gap forecasts in the main text, the difference here for the UC and MVHP-filters are small. As can be seen in Table 14 and Table 15, the SVAR-models benefit from including trend GDP forecasts in the inflation forecast contest (comparing the two right columns). The opposite is true in the growth forecast contest; see Table 16.

Table 14: Relative RMSE statistics based on quasi-real time data when using forecast of potential GDP of the respective models.
Dependent variable: Quarterly core inflation (*KPIX*).
Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Horizon				
	2	4	6	Average using model forecast of the output gap	Average using AR- forecast of the output gap[1]
SVAR 1	0,94	0,89	0,84	0,89	0,94
SVAR 2	0,98	0,93	0,91	0,94	0,95
SVAR 3	0,97	0,94	0,87	0,92	0,96
SVAR 4	0,97	0,92	0,89	0,93	0,95
SVAR 5	0,94	0,92	0,85	0,91	1,00
UC 1	0,97	0,98	0,98	0,98	0,97
UC 2	0,98	1,00	1,00	1,00	0,99
UC 3	0,98	1,00	1,00	0,99	0,99
UC 4	0,96	0,92	0,87	0,92	0,91
UC 5	0,96	0,92	0,89	0,92	0,92
UC 6	0,96	0,91	0,85	0,91	0,90
UC 7	0,96	0,93	0,89	0,93	0,93
MVHP	0,98	0,99	0,98	0,98	0,98

Remark: See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of inflation, see equation (1.17). [1]From Table 4.

Table 15: Relative RMSE statistics based on quasi-real time data when using forecast of potential GDP of the respective models.
Dependent variable: Quarterly core inflation, excluding energy (*KPIXEE*).
Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Horizon				
	2	4	6	Average using model forecast of the output gap	Average using AR- forecast of the output gap[1]
SVAR 1	0,90	0,86	0,84	0,86	0,86
SVAR 2	0,93	0,89	0,86	0,89	0,87
SVAR 3	0,85	0,90	0,89	0,88	0,95
SVAR 4	0,91	0,89	0,86	0,89	0,93
SVAR 5	0,94	0,90	0,86	0,90	0,95
UC 1	0,92	0,92	0,94	0,93	0,93
UC 2	0,93	0,94	0,96	0,94	0,94
UC 3	0,93	0,94	0,96	0,94	0,94
UC 4	0,86	0,82	0,80	0,83	0,83
UC 5	0,87	0,83	0,83	0,84	0,85
UC 6	0,85	0,79	0,76	0,80	0,80
UC 7	0,87	0,84	0,84	0,85	0,85
MVHP	0,96	0,96	0,96	0,96	0,96

Remark: The best model (i.e. the lowest relative RMSE) for each column is shown in bold.
See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of inflation, see equation (1.17). [1]From Table 5.

When it comes to growth forecasts, the use of model based forecasts of the output gap still benefit the results of the SVAR models (foremost SVAR 3–5).

Table 16: Relative RMSE statistics based on quasi-real time data when using forecast of potential GDP of the respective models. Dependent variable: Quarterly GDP growth. Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Horizon				
	2	4	6	Average using model forecast of the output gap	Average using AR- forecast of the output gap[1]
SVAR 1	1,09	1,19	1,19	1,16	1,07
SVAR 2	1,11	1,16	1,16	1,14	0,99
SVAR 3	1,07	1,08	1,03	1,06	0,99
SVAR 4	1,08	1,07	1,06	1,07	1,00
SVAR 5	1,06	1,10	1,06	1,07	0,92
UC 1	0,99	0,96	0,98	0,98	0,97
UC 2	1,00	0,98	1,00	0,99	0,98
UC 3	1,00	0,98	1,00	0,99	0,98
UC 4	0,95	0,90	0,89	0,91	0,91
UC 5	0,96	0,90	0,90	0,92	0,92
UC 6	0,96	0,91	0,91	0,93	0,93
UC 7	0,96	0,91	0,91	0,93	0,93
MVHP	1,03	1,03	1,05	1,03	1,03

Remark: See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of GDP-growth, see equation (1.17). [1]From Table 6.

INFLATION FORECAST OF THE SVAR MODELS

As discussed in the main text, Section 6.1, we use the same structure of the inflation forecast equation when evaluating the models considered in the paper. The reason is to isolate the contribution of the respective output gaps to predict future inflation, given the same information (i.e. lagged inflation and a dummy variable; see equation (1.17)).

It may be of some interest, however, to see how the SVAR-models *own* inflation forecasts perform relative to the AR-benchmark; especially when consider the ranking within these models. Four of the five SVAR-models contain core inflation (only SVAR 4 does not include core inflation). Inflation forecasts of these models have the potential to perform better compared to using the forecast model in the main text (see equation (1.20)) as the inflation forecast from the SVAR-model may benefit from this forecast of the other variables in the SVAR model (such as GDP growth).

Three types of inflation forecasts from the SVAR-models are shown in Table 17. The first column ("AR-model of inflation") only repeats the results of Table 4 in the main text for comparison. The second column ("SVAR unrestricted") shows the inflation forecast of the respective SVAR-models relative to the same autoregressive benchmark use elsewhere in the paper (see equation (1.17)). As can be seen in Table 17, the relative RMSE is this column considerable higher (i.e. worse) compared to the first column; especially at longer horizons. Why? The main reason is that the inflation forecast in the (here labelled "unrestricted") SVAR-models, by definition, goes to the mean of the sample (see Lütkepohl, 2005). As the mean of inflation is above 4 percent in the sample considered, the inflation forecasts of the SVAR models for the period 2000:1–2007:4 naturally becomes worse the longer horizon considered. To remedy this problem, we have also performed inflation forecasts were we restrict the inflation forecast of the SVAR models to go to the inflation target of 2 percent.[50] As can be seen in the last column of Table 17, the SVAR models performs considerable better compared to the alternatives. The SVAR 3 models performs best, although there is a rather close race with SVAR 1 and SVAR 2.

Table 17: Relative RMSE statistics based on quasi-real time data for SVAR models when using different forecast models (see the 'Remark' below the table).
Dependent variable: Quarterly core inflation (*KPIX*).
Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	HORIZON: 2 quarter		
	AR-model of inflation[1]	**SVAR, unrestricted**	**SVAR, restricted**
SVAR 1	0,97	1,08	0,92
SVAR 2	0,98	1,00	0,90
SVAR 3	0,97	0,99	0,90
SVAR 5	0,99	1,00	0,97
	HORIZON: 4 quarter		
SVAR 1	0,94	1,27	0,90
SVAR 2	0,95	1,08	0,88
SVAR 3	0,95	1,06	0,85
SVAR 5	0,99	1,05	0,98
	HORIZON: 6 quarter		
SVAR 1	0,90	1,45	0,84
SVAR 2	0,91	1,14	0,81
SVAR 3	0,94	1,15	0,79
SVAR 5	1,01	1,10	0,97
	AVERAGE		
SVAR 1	0,94	1,27	0,89
SVAR 2	0,95	1,07	0,87
SVAR 3	0,96	1,07	0,85
SVAR 5	1,00	1,05	0,97

Remark: See equation (1.23) for calculation of the relative RMSE statistics. See the text above the table for explanations of the different columns. [1] From Table 4.

COMBINED FORECASTS: DO THEY PERFORM BETTER?

It is well-known that combining *forecasts* from several models often improve forecast compared to using single models; see e.g. Ailfi and Timmermann (2005), Bacchini et al (2008) and Granger and Yuan (2004). As shown in a Monte Carlo study by Rennison (2004), combining *output gaps* (in his case: SVAR and HP models) improves the correlation between the true and the estimated output gaps. In this section, we briefly examine if an equal weighting scheme of the various output gaps improves on the inflation and growth forecasts presented above in Section 6.2. The short answer is: no.

Three types of combinations of gaps are tested:

- All 13 gaps.

---

[50] We do so by adjusting the constant (ν) in the VAR model by asigning appropriate values of the moving average coefficient, or mean, (μ); $\nu = \left(I - \widehat{A}_1 - \widehat{A}_2 - \ldots - \widehat{A}_p\right)\mu$. Apart from the inflation rate, the growth rate is assumed to settle at about two percent.

- Picking out the two of the best SVAR-models and two of the best UC-models. Below we combine SVAR 2, SVAR 4, UC 4 and UC 6.
- Picking out the two best models overall; UC 4 and UC 6.

For each of the above combinations, equal weighting is applied. The results for core inflation, core inflation excluding energy and GDP growth are shown in Table 18–Table 20. In none of the cases considered, the combined gaps outperform the best single models on average.

Table 18: Relative RMSE statistics based on quasi-real time (QRT) data. Dependent variable: Quarterly core inflation (KPIX). Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Horizon (quarters)			
Type of combination	**2**	**4**	**6**	**Average**
All 13 models	0,97	0,95	0,93	0,95
SVAR 2, SVAR 4, UC 4, UC 6	0,96	0,93	0,90	0,93
UC 4, UC 6	0,96	0,92	0,88	0,92
Single models				
SVAR 2	0,98	0,95	0,91	0,95
SVAR 4	0,96	0,96	0,94	0,95
UC 4	0,96	0,92	0,87	0,91
UC 6	0,96	0,91	0,84	0,90

Remark: See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of inflation, see equation (1.17).

Table 19: Relative RMSE statistics based on quasi-real time (QRT) data. Dependent variable: Quarterly core inflation, excluding energy (KPIXEE). Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Horizon (quarters)			
Type of combination	**2**	**4**	**6**	**Average**
All 13 models	0,90	0,89	0,90	0,90
SVAR 2, SVAR 4, UC 4, UC 6	0,88	0,86	0,86	0,87
UC 4, UC 6	0,86	0,83	0,83	0,84
Single models				
SVAR 2	0,91	0,86	0,83	0,87
SVAR 4	0,94	0,93	0,92	0,93
UC 4	0,86	0,82	0,81	0,83
UC 6	0,85	0,79	0,76	0,80

Remark: See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of inflation, see equation (1.17).

Table 20: Relative RMSE statistics based on quasi-real time (QRT) data. Dependent variable: Quarterly GDP growth. Evaluation period: 2000:1–2007:4. First estimation period 1976q1–1999q4.

	Horizon (quarters)			
Type of combination	**2**	**4**	**6**	**Average**
All 13 models	0,98	0,96	0,96	0,96
SVAR 2, SVAR 4, UC 4, UC 6	0,98	0,98	0,98	0,98
UC 4, UC 6	0,96	0,90	0,89	0,92
Single models				
SVAR 2	1,00	0,98	0,99	0,99
SVAR 4	0,99	1,00	1,01	1,00
UC 4	0,95	0,90	0,88	0,91
UC 6	0,96	0,91	0,91	0,93

Remark: See equation (1.23) for calculation of the relative RMSE statistics. A value below 1,0 implies better forecast performance than an AR-model of GDP growth, see equation (1.17).

SUMMING UP THE SENSITIVITY ANALYSIS

The main conclusions from this section are as follows:

- When considering the average of the 20 forecasting combinations of each of the 13 models, the ranking between the models are unaltered. More specifically, the ranking in the inflation and growth forecast contests is similar both when using the best combination for each model and when using the average of the 20 combinations for each of 13 models.

- When using forecast of trend GDP from the respective models (instead of an AR(1)-model in the main text), the SVAR-models improve their performance in the inflation forecast contest although the best UC-models (UC 4-UC 7) still dominates. When it comes to GDP-forecast, the results of the SVAR-models deteriorate while this is not the case for the best UC-models (UC 4-UC 7).

- The inflation forecasts of the SVAR-models improve on the inflation forecasts of the AR-model extended with the SVAR-gaps. This is not surprising as the SVAR-models are generally much richer (includes more variables) compared to the AR-specification in the baseline inflation equation. However these results can not be compared with the inflation forecasts of the other models shown in the paper as they all are based on an AR-model of inflation. In practice, far more complicated models of inflation forecasts are used. The focus in the paper is to evaluate the *relative* merits of the different gaps concerning inflation- and growth forecasts.

- Finally, combining different estimates of the output gaps considered in the paper does not improve the inflation- and growth forecast performance.

9. References

Adolfson, M., Lassén, S., Lindé, J., Svensson, L.E.O. (2008): "Optimal monetary policy in an operational medium-sized DSGE model". *Working paper nr. 225,* the Riksbank (Central bank of Sweden).

Adolfson, M., Lassén, S., Lindé, J., Villani, M. (2007): "RAMSES – a new general equilibrium model for monetary policy analysis". *Economic Review 2*, the Riksbank (Central bank of Sweden).

Ailfi, M., Timmermann, A. (2005): "Persistence in forecast performance and conditional combination strategies". *Journal of Econometrics*, Vol. 132, Issues 1–2, pp. 31–53.

Albagli, E., García, P., Restrepo, J. (2004): "Labor market rigidity and structural shocks: An open-economy approach for international comparisons". *Working paper 263*. Central Bank of Chile.

Amisano, G., Serati, M. (2002): "What goes up sometimes stays up: shocks and institutions as determinants of unemployment persistence", *Linc Papers No. 111*.

Antonicova, Z., Hucek, J. (2005): "Multivariate filter with unobserved components". Central bank of Slovenia, *BIATEC Vol. 8, No. 6*.

Apel, M., Jansson, P. (1999): "System estimates of potential output and the NAIRU". *Empirical Economics, Vol. 24, No. 3*.

Armstrong, J.S. (2007): "Significance tests harm progress in forecasting". *International Journal of Forecasting, Vol. 23*, pp. 321–327.

Bacchini, F., Ciammola, A., Iannaccone, R., Marini, M. (2008): "Combining forecasts for producing flash estimates of Euro area GDP". Conferernce paper presented at the 5th Eurostat Colloquium on Modern Tools for Business Cycle Analysis, September-October 2008.

Ball, L., Mankiw, N.G. (2002): "The NAIRU in theory and practice". *Journal of Economic Perspectives, Vol. 16, No. 4*, pp. 115–136.

Batini, N., Greenslade, J. (2003): "Measuring the UK short-run NAIRU". *Discussion paper No. 12*. Bank of England.

Baxter, M., R.G. King (1995): "Measuring business cycles: approximate band-pass filters for economic time series". *NBER Working paper No. 5022*.

Bean, C. (2005): "Comments on: "Separating the business cycle from other economic fluctuations". In: *The Greenspan era: Lessons for the fut*ure. Federal Reserve Bank of Kansas City."

Beffy, P-O., Ollivaud, P., Richardson, P., Sédillot, F. (2006): "New OECD methods for supply-side and medium-term assessments: A capital services apporach". *Working paper No. 482*, OECD.

Benes, J., N'Diaye, P. (2004): "A multivariate filter for measuring potential output and the NAIRU: Application to the Czech Republic". *IMF Working paper No. 45*.

Bergman, U.M., Juel, S., Steigum, E. (2009): "Monetary policy and financial turmoil". Norges Bank Watch 2009. Centre for Monetary Economics. Norwegian School of Management.

Bergvall, A., Dillén, M. (2005); "Några funderingar kring negativa produktionsgap, icke-linjära Phillipskurvor och stabiliseringspolitik". (In Swedish only). Manuscript. NIER.

Bignasca, F., Rossi, E. (2007): "Applying the Hirose-Kamada filter to Swiss data: Output gap and exchange rate pass-through estimates". Swiss National Bank, manuscript, October.

Blanchard, O.J., Katz, L. (1997): "What we know and do not know about the natural rate of unemployment". *Journal of Economic Perspectives, Vol. 11, Issue 1*, pp. 51–72.

Blanchard, O.J., Quah, D. (1989): "The dynamic effects of aggregate demand and supply distrubances". *American Economic Review, Vol. 79, No. 4*, pp. 655–673.

Cahn, C., Saint-Guihem, A. (2007): "Potential output growth in several industrialized countries: A comparison". *Working paper No. 828*, European Central Bank.

Camba-Mendez, G., Rodriguez-Palenzuela, D. (2003): "Assessment criteria for output gap estimates". *Economic Modelling, Vol. 20, No. 3*, pp. 529–562.

Canova, F., Sala, L. (2009): "Back to square one: Identification issues in DSGE models". *Journal of Monetary Economics No. 56*, pp. 431–449.

Cayen, J.P., van Norden, S. (2005): "The reliability of Canadian output gap estimates". *The North American Journal of Economics and Finance. Vol 16, No. 3*, pp 373–393.

Cerra, V., Saxena, S.C. (2000): "Alternative methods of estimating potential output and the output gap: an application to Sweden". *IMF Working paper No. 59*.

Chari, V.V., Kehoe, P.J., McGrattan, E.R. (2008): "New Keynesian Models: Not yet useful for policy analysis". *Research Department Staff Report 409*. Federal Reserve Bank of Minneapolis.

Christoffel, K., Coenen, G., Warne, A. (2008): "The new area-wide model of the Euro area: Specification, estimation results and properties". *Working paper No. 944*, European Central Bank.

Coenen, G., Smets, F., Vetlov, I. (2008): "Estimation of Euro area output gap using the NAWM". Manuscript, August.

Cotis, J-P., Elmeskov, J., Mourougane, A. (2003): "Estimates of potential output: benefits and pitfalls from a policy perspective". Manuscript. OECD.

Denis, C., Grenouilleau, D., McMorrow, K., Röger, W. (2006): "Calculating potential growth rates and output gaps – A revised production function approach". *Economic papers No. 247*, European Commission.

Dolado, J.J., Jemeno, J.F. (1997): "The causes of Spanish unemployment: a structural VAR approach". *European Economic Review Vol. 41*, pp. 1281–1307.

Dupasquier, C., Guay, A., St-Amant, P (1999): "A survey of alternative methodologies for estimating potential output and the output gap". *Journal of Macroeconomics, Vol.21, No. 3*, pp. 577–595.

Edge, R.M., Kiley, M.T., Laforte, J-P. (2008): "Natural rate measures in an estimated DSGE model of the U.S. economy". *Journal of Economic Dynamics and Control. Vol 32*, pp. 2512-2535.

Eliasson, A-C. (2001): "Is the short-run Phillips curve nonlinear? Empirical evidence for Australia, Sweden and the United States". *Working paper No. 124*. Sveriges Riksbank (Central bank of Sweden).

Enders, W. and Hurn, S. (2007): "Identifying aggregate demand and supply shocks in a small open economy". *Oxford Economic Papers,* Vol.59, No. 3, pp.411–429.

Fabiani, S., Locarno, A., Oneto, G., Sestito, P. (2001): "The sources of unemployment fluctuations: An empirical application to the Italian case. *Labour Economics, Vol. 8, No. 2*, pp. 259–289. "

Fisher, J.D., Chin, T.L., Ruilin, Z. (2002): "When can we forecast inflation?" Federal Reserve Bank of Chicago *Economic Perspectives, No. 1*, pp. 30–42.

Friedman, M. (1968): "The role of monetary policy". *American Economic Review, Vol. 58*, pp. 1–17.

Funke, M. (1997): "Supply potential and output gaps in West German manufacturing". *International Journal of Forecasting, Vol. 13*, pp 211–222.

Galí, J. (2008a): *Monetary policy, inflation, and the business cycle.* Princeton University Press. United States.

Galí, J. (2008b): "The New Keynesian Approach to Monetary Policy Ananlysis: Lessons and New Directions". Manuscript, Febuary 6th.

Gianella, C., Koske, I., Rusticelli, E., Chatal, O. (2009): "What drives the NAIRU? Evidence from a panel of OECD Countries". *Working paper No. 649*. OECD.

Gordon, R. (1997): "The Time-Varying NAIRU and its implications for economic policy". *Journal of Economic Literature, Vol. 11, No. 1*, pp. 11–32

Granger, C.W.J., Jeon, Y. (2004): "Thick modeling". *Economic Modelling*, vol. 21, issue 2, pp.323–343.

Hjelm, G. (2010): "Projekt BNP-gap: Avrapportering och tankar kring eventuell implementering". Memo. National Institute of Economic Research. (Only in Swedish).

Jaeger, A. (1994): "Mechanical Detrending by Hodrick-Prescott Filtering: A Note". *Empirical Economics, Vol. 19*, issue 3, pp. 493–500.

Karagedikli, Ö., Plantier, L.C. (2005): "Do multivariate filters have better revision properties? An empirical analysis". Manuscript, June.

Koske, I., Pain, N. (2008): "The usefulness of output gaps for policy analysis". *OECD Working paper No. 621*.

Kuttner, K.N. (1994): "Estimating potential output as a latent variable". *Journal of Business and Economic Statistics, Vol. 12, No. 3*, pp. 361–368.

Laxton, D., Tetlow, R. (1992): "A simple multivariate filter for the measurement of potential output". *Technical Report No. 59*, Bank of Canada.

Lemoine, M., Mazzi, G.L., Monperrus-Veroni, P, Reynes, R. (2008) "Real time estimation of potential output and output gap for the Euro-area: Comparing production function, unobserved components and SVAR approaches". Paper presented at the 5th Eurostat colloquium on modern tools for business cycle analysis, Luxembourg September-October 2008,

Lown, C., Rich, R. (1997): "Is there an inflation puzzle?" *Federal Reserve Bank of New York Economic Policy Review*, December, pp. 51–59.

Lütkepohl, H. (2005): *New introduction to multiple time series econometrics*. Springer-Verlag. Germany.

Kaiser, R., Maravall, A. (2001): *Measuring business cycle in economic time series*. Lecture notes in statistics 154, New York. Springer-Verlag.

Mestre, R., McAdam, P. (2008): "Is forecasting with large models informative? Assessing the role of judgement in macro-economic forecasts". *ECB Working paper No. 950*.

Moosa, I.A. (1997). "A Cross-country Comparison of Okun's Coefficient". *Journal of Comparative Economics, Vol. 24, No. 3,* pp. 335–356.

Murchison, S., Rennison, A. (2006): "ToTEM: The Bank of Canada's new quarterly projection model". *Technical report 97*, Bank of Canada.

Muth, J. (1960): "Optimal properties of exponential weighted forecasts". *Journal of American Statistical Association, Vol. 55*, pp. 290–306.

Orphandides, A., van Norden, S. (2004): "The reliability of inflation forecasts based on output gap estimates in real time". *FEDS Working paper No. 68*.

Phelps, E. (1968): "Money-wage dynamics and labor market equilibrium". *Journal of Political Economy, Vol. 76*, pp. 678–711.

Phillips, A.W. (1958): "The relationship between unemployment and the rate of change of money wages in the United Kingdom, 1861–1957". *Economica, Vol. 25*, pp. 283–299.

Proietti T. (2008), Structural Time Series Models for Business Cycle Analysis. To appear in the Handbook of Econometrics: Vol. 2, Applied Econometrics, Part 3.4., ed. T. Mills and K. Patterson, Palgrave, London, 2008.

Ratto, M., Roeger, W., Veld, J. (2006): "Fiscal policy in an estimated open-economy model for the EURO area". *Economic papers No. 266*. European Commission.

Richardson, R., Boone, L., Giorno, C., Meacci, M., Rae, D., Turner, D. (2000): "The concept, policy use and measurement of structural unemployment". *OECD Working paper No. 250*.

Ross, K., Ubide, A. (2001): "Mind the gap: what is the best measure of slack in the Euro area?" *IMF Working paper No. 203*.

Staiger, D., Stock, J.H., Watson, M.W. (1997): "The NAIRU, unemployement, and monetary policy". *Journal of Economic Perspectives, Vol. 11*, pp. 33–51.

Tovar, C.E. (2008): ”DSGE models and central banks”. *BIS Working paper No. 258*.

Watson, M.W. (1986): “Univariate detrending methods with stochastic trends.” *Journal of Monetary Economics, Vol. 18*, pp. 49–75.

Woodford, M. (2003): *Interest and prices: Foundations of a theory of monetary policy.* Princeton University Press. New Jersey.

10. Titles in the Working Paper Series

No	Author	Title	Year
1	Warne, Anders and Anders Vredin	Current Account and Business Cycles: Stylized Facts for Sweden	1989
2	Östblom, Göran	Change in Technical Structure of the Swedish Economy	1989
3	Söderling, Paul	Mamtax. A Dynamic CGE Model for Tax Reform Simulations	1989
4	Kanis, Alfred and Aleksander Markowski	The Supply Side of the Econometric Model of the NIER	1990
5	Berg, Lennart	The Financial Sector in the SNEPQ Model	1991
6	Ågren, Anders and Bo Jonsson	Consumer Attitudes, Buying Intentions and Consumption Expenditures. An Analysis of the Swedish Household Survey Data	1991
7	Berg, Lennart and Reinhold Bergström	A Quarterly Consumption Function for Sweden 1979-1989	1991
8	Öller, Lars-Erik	Good Business Cycle Forecasts- A Must for Stabilization Policies	1992
9	Jonsson, Bo and Anders Ågren	Forecasting Car Expenditures Using Household Survey Data	1992
10	Löfgren, Karl-Gustaf, Bo Ranneby and Sara Sjöstedt	Forecasting the Business Cycle Not Using Minimum Autocorrelation Factors	1992
11	Gerlach, Stefan	Current Quarter Forecasts of Swedish GNP Using Monthly Variables	1992
12	Bergström, Reinhold	The Relationship Between Manufacturing Production and Different Business Survey Series in Sweden	1992
13	Edlund, Per-Olov and Sune Karlsson	Forecasting the Swedish Unemployment Rate: VAR vs. Transfer Function Modelling	1992
14	Rahiala, Markku and Timo Teräsvirta	Business Survey Data in Forecasting the Output of Swedish and Finnish Metal and Engineering Industries: A Kalman Filter Approach	1992
15	Christofferson, Anders, Roland Roberts and Ulla Eriksson	The Relationship Between Manufacturing and Various BTS Series in Sweden Illuminated by Frequency and Complex Demodulate Methods	1992
16	Jonsson, Bo	Sample Based Proportions as Values on an Independent Variable in a Regression Model	1992
17	Öller, Lars-Erik	Eliciting Turning Point Warnings from Business Surveys	1992
18	Forster, Margaret M	Volatility, Trading Mechanisms and International Cross-Listing	1992
19	Jonsson, Bo	Prediction with a Linear Regression Model and Errors in a Regressor	1992
20	Gorton, Gary and Richard Rosen	Corporate Control, Portfolio Choice, and the Decline of Banking	1993
21	Gustafsson, Claes-Håkan and Åke Holmén	The Index of Industrial Production – A Formal Description of the Process Behind it	1993
22	Karlsson, Tohmas	A General Equilibrium Analysis of the Swedish Tax	1993

		Reforms 1989-1991	
23	Jonsson, Bo	Forecasting Car Expenditures Using Household Survey Data- A Comparison of Different Predictors	1993
24	Gennotte, Gerard and Hayne Leland	Low Margins, Derivative Securitites and Volatility	1993
25	Boot, Arnoud W.A. and Stuart I. Greenbaum	Discretion in the Regulation of U.S. Banking	1993
26	Spiegel, Matthew and Deane J. Seppi	Does Round-the-Clock Trading Result in Pareto Improvements?	1993
27	Seppi, Deane J.	How Important are Block Trades in the Price Discovery Process?	1993
28	Glosten, Lawrence R.	Equilibrium in an Electronic Open Limit Order Book	1993
29	Boot, Arnoud W.A., Stuart I Greenbaum and Anjan V. Thakor	Reputation and Discretion in Financial Contracting	1993
30a	Bergström, Reinhold	The Full Tricotomous Scale Compared with Net Balances in Qualitative Business Survey Data – Experiences from the Swedish Business Tendency Surveys	1993
30b	Bergström, Reinhold	Quantitative Production Series Compared with Qualiative Business Survey Series for Five Sectors of the Swedish Manufacturing Industry	1993
31	Lin, Chien-Fu Jeff and Timo Teräsvirta	Testing the Constancy of Regression Parameters Against Continous Change	1993
32	Markowski, Aleksander and Parameswar Nandakumar	A Long-Run Equilibrium Model for Sweden. The Theory Behind the Long-Run Solution to the Econometric Model KOSMOS	1993
33	Markowski, Aleksander and Tony Persson	Capital Rental Cost and the Adjustment for the Effects of the Investment Fund System in the Econometric Model Kosmos	1993
34	Kanis, Alfred and Bharat Barot	On Determinants of Private Consumption in Sweden	1993
35	Kääntä, Pekka and Christer Tallbom	Using Business Survey Data for Forecasting Swedish Quantitative Business Cycle Varable. A Kalman Filter Approach	1993
36	Ohlsson, Henry and Anders Vredin	Political Cycles and Cyclical Policies. A New Test Approach Using Fiscal Forecasts	1993
37	Markowski, Aleksander and Lars Ernsäter	The Supply Side in the Econometric Model KOSMOS	1994
38	Gustafsson, Claes-Håkan	On the Consistency of Data on Production, Deliveries, and Inventories in the Swedish Manufacturing Industry	1994
39	Rahiala, Markku and Tapani Kovalainen	Modelling Wages Subject to Both Contracted Increments and Drift by Means of a Simultaneous-Equations Model with Non-Standard Error Structure	1994
40	Öller, Lars-Erik and Christer Tallbom	Hybrid Indicators for the Swedish Economy Based on Noisy Statistical Data and the Business Tendency	1994

		Survey	
41	Östblom, Göran	A Converging Triangularization Algorithm and the Intertemporal Similarity of Production Structures	1994
42a	Markowski, Aleksander	Labour Supply, Hours Worked and Unemployment in the Econometric Model KOSMOS	1994
42b	Markowski, Aleksander	Wage Rate Determination in the Econometric Model KOSMOS	1994
43	Ahlroth, Sofia, Anders Björklund and Anders Forslund	The Output of the Swedish Education Sector	1994
44a	Markowski, Aleksander	Private Consumption Expenditure in the Econometric Model KOSMOS	1994
44b	Markowski, Aleksander	The Input-Output Core: Determination of Inventory Investment and Other Business Output in the Econometric Model KOSMOS	1994
45	Bergström, Reinhold	The Accuracy of the Swedish National Budget Forecasts 1955-92	1995
46	Sjöö, Boo	Dynamic Adjustment and Long-Run Economic Stability	1995
47a	Markowski, Aleksander	Determination of the Effective Exchange Rate in the Econometric Model KOSMOS	1995
47b	Markowski, Aleksander	Interest Rate Determination in the Econometric Model KOSMOS	1995
48	Barot, Bharat	Estimating the Effects of Wealth, Interest Rates and Unemployment on Private Consumption in Sweden	1995
49	Lundvik, Petter	Generational Accounting in a Small Open Economy	1996
50	Eriksson, Kimmo, Johan Karlander and Lars-Erik Öller	Hierarchical Assignments: Stability and Fairness	1996
51	Url, Thomas	Internationalists, Regionalists, or Eurocentrists	1996
52	Ruist, Erik	Temporal Aggregation of an Econometric Equation	1996
53	Markowski, Aleksander	The Financial Block in the Econometric Model KOSMOS	1996
54	Östblom, Göran	Emissions to the Air and the Allocation of GDP: Medium Term Projections for Sweden. In Conflict with the Goals of SO_2, SO_2 and NOX Emissions for Year 2000	1996
55	Koskinen, Lasse, Aleksander Markowski, Parameswar Nandakumar and Lars-Erik Öller	Three Seminar Papers on Output Gap	1997
56	Oke, Timothy and Lars-Erik Öller	Testing for Short Memory in a VARMA Process	1997
57	Johansson, Anders and Karl-Markus Modén	Investment Plan Revisions and Share Price Volatility	1997
58	Lyhagen, Johan	The Effect of Precautionary Saving on Consumption in Sweden	1998
59	Koskinen, Lasse and Lars-Erik Öller	A Hidden Markov Model as a Dynamic Bayesian Classifier, with an Application to Forecasting Business-Cycle Turning Points	1998

60	Kragh, Börje and Aleksander Markowski	Kofi – a Macromodel of the Swedish Financial Markets	1998
61	Gajda, Jan B. and Aleksander Markowski	Model Evaluation Using Stochastic Simulations: The Case of the Econometric Model KOSMOS	1998
62	Johansson, Kerstin	Exports in the Econometric Model KOSMOS	1998
63	Johansson, Kerstin	Permanent Shocks and Spillovers: A Sectoral Approach Using a Structural VAR	1998
64	Öller, Lars-Erik and Bharat Barot	Comparing the Accuracy of European GDP Forecasts	1999
65	Huhtala , Anni and Eva Samakovlis	Does International Harmonization of Environmental Policy Instruments Make Economic Sense? The Case of Paper Recycling in Europe	1999
66	Nilsson, Charlotte	A Unilateral Versus a Multilateral Carbon Dioxide Tax - A Numerical Analysis With The European Model GEM-E3	1999
67	Braconier, Henrik and Steinar Holden	The Public Budget Balance – Fiscal Indicators and Cyclical Sensitivity in the Nordic Countries	1999
68	Nilsson, Kristian	Alternative Measures of the Swedish Real Exchange Rate	1999
69	Östblom, Göran	An Environmental Medium Term Economic Model – EMEC	1999
70	Johnsson, Helena and Peter Kaplan	An Econometric Study of Private Consumption Expenditure in Sweden	1999
71	Arai, Mahmood and Fredrik Heyman	Permanent and Temporary Labour: Job and Worker Flows in Sweden 1989-1998	2000
72	Öller, Lars-Erik and Bharat Barot	The Accuracy of European Growth and Inflation Forecasts	2000
73	Ahlroth, Sofia	Correcting Net Domestic Product for Sulphur Dioxide and Nitrogen Oxide Emissions: Implementation of a Theoretical Model in Practice	2000
74	Andersson, Michael K. And Mikael P. Gredenhoff	Improving Fractional Integration Tests with Bootstrap Distribution	2000
75	Nilsson, Charlotte and Anni Huhtala	Is CO_2 Trading Always Beneficial? A CGE-Model Analysis on Secondary Environmental Benefits	2000
76	Skånberg, Kristian	Constructing a Partially Environmentally Adjusted Net Domestic Product for Sweden 1993 and 1997	2001
77	Huhtala, Anni, Annie Toppinen and Mattias Boman,	An Environmental Accountant's Dilemma: Are Stumpage Prices Reliable Indicators of Resource Scarcity?	2001
78	Nilsson, Kristian	Do Fundamentals Explain the Behavior of the Real Effective Exchange Rate?	2002
79	Bharat, Barot	Growth and Business Cycles for the Swedish Economy	2002
80	Bharat, Barot	House Prices and Housing Investment in Sweden and the United Kingdom. Econometric Analysis for the Period 1970-1998	2002
81	Hjelm, Göran	Simultaneous Determination of NAIRU, Output Gaps and Structural Budget Balances: Swedish Evi-	2003

		dence	
82	Huhtala, Anni and Eva Samalkovis	Green Accounting, Air Pollution and Health	2003
83	Lindström, Tomas	The Role of High-Tech Capital Formation for Swedish Productivity Growth	2003
84	Hansson, Jesper, Per Jansson and Mårten Löf	Business survey data: do they help in forecasting the macro economy?	2003
85	Boman, Mattias, Anni Huhtala, Charlotte Nilsson, Sofia Ahlroth, Göran Bostedt, Leif Mattson and Peichen Gong	Applying the Contingent Valuation Method in Resource Accounting: A Bold Proposal	
86	Gren, Ing-Marie	Monetary Green Accounting and Ecosystem Services	2003
87	Samakovlis, Eva, Anni Huhtala, Tom Bellander and Magnus Svartengren	Air Quality and Morbidity: Concentration-response Relationships for Sweden	2004
88	Alsterlind, Jan, Alek Markowski and Kristian Nilsson	Modelling the Foreign Sector in a Macroeconometric Model of Sweden	2004
89	Lindén, Johan	The Labor Market in KIMOD	2004
90	Braconier, Henrik and Tomas Forsfält	A New Method for Constructing a Cyclically Adjusted Budget Balance: the Case of Sweden	2004
91	Hansen, Sten and Tomas Lindström	Is Rising Returns to Scale a Figment of Poor Data?	2004
92	Hjelm, Göran	When Are Fiscal Contractions Successful? Lessons for Countries Within and Outside the EMU	2004
93	Östblom, Göran and Samakovlis, Eva	Costs of Climate Policy when Pollution Affects Health and Labour Productivity. A General Equilibrium Analysis Applied to Sweden	2004
94	Forslund Johanna, Eva Samakovlis and Maria Vredin Johansson	Matters Risk? The Allocation of Government Subsidies for Remediation of Contaminated Sites under the Local Investment Programme	2006
95	Erlandsson Mattias and Alek Markowski	The Effective Exchange Rate Index KIX - Theory and Practice	2006
96	Östblom Göran and Charlotte Berg	The EMEC model: Version 2.0	2006
97	Hammar, Henrik, Tommy Lundgren and Magnus Sjöström	The significance of transport costs in the Swedish forest industry	2006
98	Barot, Bharat	Empirical Studies in Consumption, House Prices and the Accuracy of European Growth and Inflation Forecasts	2006
99	Hjelm, Göran	Kan arbetsmarknadens parter minska jämviktsarbetslösheten? Teori och modellsimuleringar	2006
100	Bergvall, Anders, Tomas Forsfält, Göran Hjelm, Jonny Nilsson and Ju-	KIMOD 1.0 Documentation of NIER´s Dynamic Macroeconomic General Equilibrium Model of the Swedish Economy	2007

	hana Vartiainen		
101	Östblom, Göran	Nitrogen and Sulphur Outcomes of a Carbon Emissions Target Excluding Traded Allowances - An Input-Output Analysis of the Swedish Case	2007
102	Hammar, Henrik and Åsa Löfgren	Explaining adoption of end of pipe solutions and clean technologies – Determinants of firms' investments for reducing emissions to air in four sextors in Sweden	2007
103	Östblom, Göran and Henrik Hammar	Outcomes of a Swedish Kilometre Tax. An Analysis of Economic Effects and Effects on NOx Emissions	2007
104	Forsfält, Tomas, Johnny Nilsson and Juhana Vartianinen	Modellansatser i Konjunkturinstitutets medelfristprognoser	2008
105	Samakovlis, Eva	How are Green National Accounts Produced in Practice?	2008
107	Forslund, Johanna, Per Johansson, Eva Samakovlis and Maria Vredin Johansson	Can we by time? Evaluation. Evaluation of the government's directed grant to remediation in Sweden	2009
108	Forslund, Johanna Eva Samakovlis, Maria Vredin Johansson and Lars Barregård	Does Remediation Save Lives? On the Cost of Cleaning Up Arsenic-Contaminated Sites in Sweden	2009
109	Sjöström, Magnus and Göran Östblom	Future Waste Scenarios for Sweden on the Basis of a CGE-model	2009
110	Österholm, Pär	The Effect on the Swedish Real Economy of the Financial Crisis	2009
111	Forsfält, Tomas	KIMOD 2.0 Documentation of changes in the model from January 2007 to January 2009	2009
112	Österholm, Pär	Improving Unemployment Rate Forecasts Using Survey Data	2009
113	Österholm, Pär	Unemployment and Labour-Force Participation in Sweden	2009
114	Jonsson, Thomas and Pär Österholm	The Properties of Survey-Based Inflation Expectations in Sweden	2009
115	Hjelm, Göran and Kristian Jönson	In search of a method for measuring the output gap of the Swedish economy. Economic, econometric and practical considerations	2010