Tampere University

UNI FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

Jukka-Pekka Pyylampi

Measuring Finnish Output Gap with Structural VARs: Past & Present Estimates

Under the supervision of:

Prof. Kaisa Kotakorpi

Faculty of Management and Business, University of Tampere

Prof. Dr. Thierry Madiès

Department of Political Economy, University of Fribourg

Faculty of Management and Business
Master's Thesis
June 2021

Abstract

Jukka-Pekka Pyylampi: Measuring Finnish Output Gap with Structural VARs: Past & Present Estimates
Master's Thesis: 58 pages and 13 appendix pages
Tampere University and University of Fribourg
Master's Degree Program in Public Economics and Public Finance (MGE)
June 2021

The output gap is an economic measure that aims to describe whether an economy produces too much or too little relative to its capacity and capabilities in terms of output, reflecting a phase in a business cycle. Ideally, accurate information about the size of the gap may enable timing and sizing of stabilization measures proportionate to a cycle. The size of the gap is based on the difference between the potential and actual output of an economy. Actual output is typically compiled in a national statistical institute from observed economic activity, whereas potential output cannot be directly observed, for which there are several measures to choose from.

This thesis examines uncertainty in the output gap measures, constructed applying structural VAR-based methods to the Finnish full sample and real-time data. These measures are compared across different samples and between methods, but the primary focus is on real-time performance.

The reliability of real-time output gap estimates has been frequently questioned, as they tend to be the subject of ex-post revisions, particularly around economic turning points when the importance of the measurement tends to be emphasized. The ex-post revisions are typically associated with measurement errors in potential output, resulting from erroneous measurements in real-time. The recent empirical literature suggests that structural VAR-decompositions are able to reconstruct more reliable potential output and output gap estimates in real-time.

The results show how the output gap measures reconstructed from two types of structural VAR and the measure based on the cyclical component of the univariate Hodrick–Prescott (HP) filter form somewhat similar views about the cyclical development in Finland with full sample data. Actual output appears to be above or under potential, mainly when it is supported by the history and macroeconomic indicators. However, the size of the measures differs, particularly around economic turning points. Sometimes differences between the measures may be even several percentage points, and even the sign of the estimate may differ when actual output is close to potential.

The consistency of the output gap estimates improves over time, suggesting that ex-post revisions tend to be greater at the end of the sample. GDP data revisions do not appear to be the primary source of the revisions suggesting that these models identify too slowly whether changes in output will be long-lived, particularly around economic turning points. However, exceptionally large revisions in actual output may also affect the estimates.

All the applied output gap measures point towards greater positive output gap estimates for the period of overheating prior to the global financial crisis with the full sample than with real-time data, suggesting that these models overestimate the level of potential output during the periods of overheating. The estimates also show how the output gap measures based on structural VAR methods appear to react consistently to extreme economic turning points, but to gradual reversal of cycles are reacted with a delay. Overall, the real-time assessments do not show an indication of better real-time performance of structural VAR methods relative to the univariate HP filter with the Finnish data.

Keywords: output gap, potential output, real-time estimation, business cycles, stabilization policy

Tiivistelmä

Jukka-Pekka Pyylampi: Suomen tuotantokuilun mittaaminen rakenteellisilla VAR-malleilla: Menneet ja nykyiset arviot
Pro gradu -tutkielma: 58 sivua ja 13 liite sivua
Tampereen Yliopisto ja University of Fribourg
Master's Degree Program in Public Economics and Public Finance (MGE)
Kesäkuu 2021

Tuotantokuilu on taloudellinen mittari, joka pyrkii kuvaamaan tuottaako talous liian paljon vai vähän kapasiteettiinsa ja kyvykkyyksiinsä nähden heijastaen suhdannevaihetta. Parhaimmillaan tarkka tieto kuilun koosta voi mahdollistaa vakauttamistoimenpiteiden ajoituksen ja mitoituksen suhdannevaiheeseen. Kuilun koko perustuu talouden potentiaalisen ja toteutuneen tuotannon väliseen eroon. Toteutuneen tuotannon laatii tyypillisesti kansallinen tilastoviranomainen havaitun taloudellisen toiminnan perusteella, kun taas potentiaalista tuotantoa ei voida suoraan havaita, jonka arvioimiseksi on olemassa useita erilaisia mittareita.

Tässä tutkielmassa tarkastellaan tuotantokuilumittareihin liittyvää epävarmuutta, jotka on muodostettu rakenteellisilla VAR-malleilla, sekä lopullista että reaaliaikaista aineistoa hyödyntämällä. Kyseisiä mittareita vertaillaan otoksien ja menetelmien välillä, mutta ensisijainen painopiste on menetelmien reaaliaikaisen suorituskyvyn tarkastelussa.

Reaaliaikaisten tuotantokuiluarvioiden luotettavuutta on usein kyseenalaistettu, koska arvioihin tehdään usein jälkikäteisiä tarkistuksia erityisesti talouden käännekohtien ympärillä, jolloin mittaamisen merkitys tapaa olla korostunut. Tuotantokuiluarvioihin jälkikäteen tehdyt tarkistukset liittyvät tyypillisesti potentiaalisen tuotannon mittausvirheisiin, jotka seurausta virheellisistä reaaliaikaisista mittauksista. Viimeaikainen empiirinen kirjallisuus viittaa siihen, että rakenteelliset VAR-hajottelemat kykenevät muodostamaan luotettavampia arviota potentiaalisen tuotannosta ja tuotantokuilusta reaaliajassa.

Tulokset osoittavat kuinka kahteen rakenteelliseen VAR-menetelmään perustuvat tuotantokuilumittarit, ja Hodrick–Prescott-suotimen suhdannekomponentti, muodostavat melko samankaltaiset näkemykset Suomen suhdannekehityksestä täydellisillä otostiedoilla. Toteutunut tuotanto on potentiaalin ylä- tai alapuolella enimmäkseen silloin kun historia ja makrotaloudelliset indikaattorit tukevat sitä. Mittareiden koot kuitenkin eroavat erityisesti talouden käännekohtien ympäristössä. Joskus erot mittareiden välillä saattavat olla jopa useita prosenttiyksikköjä, ja jopa arvioiden etumerkki voi vaihdella, kun toteutunut tuotanto on lähellä potentiaalia.

Tuotantokuiluarvioiden johdonmukaisuus paranee ajan mittaan, joka osoittaa, että jälkitarkastukset ovat yleensä suurempia otoksen loppupäässä. Bruttokansantuotteen revisiot eivät ole jälkikäteistarkistuksien ensisijainen lähde, mikä viittaa siihen, että sovelletut mallit tunnistavat liian hitaasti ovatko muutokset tuotannossa pitkäikäisiä, etenkin talouden käännekohtien ympäristössä. Toteutuneen tuotannon poikkeuksellisen suuret tarkistukset saattavat kuitenkin vaikuttaa tuotantokuiluarvioiden kokoon.

Kaikki sovelletut tuotantokuilumittarit osoittavat suurempia positiivisia tuotantokuiluarvioita täydellisellä aineistolla kuin reaaliaikaisella aineistolla talouden ylikuumetessa ennen globaalia finanssikriisiä, mikä viittaa siihen, että malleilla on taipumus yliarvioida potentiaalisen tuotannon taso talouden ylikuumenemisen aikana. Arviot osoittavat myös, kuinka VAR-menetelmiin perustuvat tuotantokuilumittarit vaikuttavat reagoivan jyrkkiin talouden käännekohtiin johdonmukaisesti, mutta suhdanteen asteittaisiin kääntymisiin mallit näyttävät reagoivat viiveellä. Kaiken kaikkiaan tuotantokuilun reaaliaikaiset arvioinnit eivät osoita rakenteellisten VAR-menetelmien parempaa reaaliaikaista suorituskykyä verrattuna yksimuuttujaiseen HP-suodattimeen suomalaisella aineistolla.

Avainsanat: tuotantokuilu, potentiaalinen tuotanto, reaaliaikainen arviointi, suhdannevaihtelut, vakauttamispolitiikka

Contents

1. Introduction

The output gap is an economic measure that aims to describe whether an economy produces too much or too little relative to its capacity and capabilities in terms of output, reflecting a phase in a business cycle. Ideally, accurate information about the size of the gap may enable timing and sizing of stabilization measures proportionate to a cycle. The size of the gap is based on a difference between the potential and actual output of an economy. Actual output is typically compiled in a national statistical institute from observed economic activity. Whereas potential output cannot be directly observed, hence there are several measures to choose from. The reliability of output gap measures in real-time has been frequently questioned as the estimates tend to be the subject of ex-post revisions, particularly around economic turning points when the importance of the measurement tends to be emphasized. The ex-post revisions in the estimates are typically associated with measurement errors in potential output, resulting from erroneous measurements in real-time. The recent empirical literature suggests that structural VAR-decompositions can reconstruct more reliable estimates of potential output in real-time.

In the literature, potential output typically refers to the upper limit of output that an economy can sustainably produce in the long-run. From a theoretical point of view, the concept is usually built on Arthur Okun's (1962) seminal notion of the upper limit of output that is possible to produce without generating price pressures. The concept itself refers to the supply-side of the economy, and thus only supply shocks are generally acknowledged to have permanent effects on output, whereas temporary changes in output are driven by demand-side factors. (Coibion et al., 2018.) The size of the output gap stems from both permanent and temporary changes in output. If the level of actual output of an economy is above potential, the production capacity of an economy is overutilized, indicating growth pressures in prices as known from the Phillips Curve, while a negative output gap suggests the opposite situation (Álvarez, L. & Goméz-Loscos, A., 2017). This relationship between the output gap and prices is particularly important for the conduct of monetary policy. Another widely used measure in which the output gaps are used for representing cyclical conditions is the structural budget balance, which is used to measure the financial positioning of the government that is adjusted from the consequences of the business cycle. Hence, any uncertainty associated with the obtained size of the gap may hamper accurate sizing of the output gap-based policy recommendations (Orphanides, 2001; 2003).

Orphanides and van Norden (2002) showed how the uncertainty in real-time output gap estimates could be traced, either to a selected method or data used in the estimations. The latter refers to possible data revisions, while the former is associated with the selected method itself. For instance, several methods produce estimates that differ particularly between the end and middle of the sample, generating ex-post revisions as presented in Orphanides and van Norden (2002) and Marcellino and Musso (2011). Correspondingly, the estimates of international organizations, including OECD, IMF, and the European Commission, typically produced following production function approach, have been subject to ex-post revisions due to the uncertainty associated with the real-time estimates. (Tosetto, 2008; Kempes, 2012; Mc Morrow et al., 2015; Hernández de Cos et al., 2016; Turner et al., 2016; Kangur et al., 2019). Similar estimation approaches are also widely applied in public organizations.

One of the acknowledged difficulties in the measurement of potential output in real-time is the ability to identify whether changes in output will be permanent. Coibion et al. (2018) present evidence that the real-time estimates of potential output produced with some of the widely used methods, for example, using the HP filter and production function approach, respond too strongly to temporary disturbances in output. As an alternative, they suggest methods, such as that proposed by Blanchard and Quah (1989), which reconstruct potential output by summing disturbances in output that have had only permanent effects on output and therefore are less sensitive to temporary effects in real-time.

In 1989, Blanchard and Quah showed how a bivariate structural VAR with a long-term restriction could distinguish transitory and permanent disturbances in output, which allowed to reconstruct series of potential output from accumulated permanent disturbances. This method is known as the Blanchard and Quah (BQ) method, in which structural supply and demand shocks are identified from series of (log) real GDP and unemployment. However, the method may not produce appropriate decompositions between temporary and permanent changes in output if the demand component of output is not driven by the unemployment rate (Billmeier, 2006). Chen and Góronicka (2020) aim to address some of the limitations of the BQ method in the case of small open economies by considering a broader range of shocks, consisting of both domestic and global disturbances in output. They propose an extended SVAR model that decomposes a higher number of disturbances in output, which is achieved by increasing the variables and imposing a mixture of short and long-term restrictions, as well as sign restrictions, following the identification scheme of Forbes et al. (2018).

Coibion et al. (2018) and Chen & Góronicka (2020) present empirical evidence that the output gap estimates constructed applying structural VAR methods with appropriately identified disturbances in output are less sensitive to temporary disturbances in output and therefore, the output gap measures are subject of lower ex-post revisions. Yet, there is still relatively little empirical evidence supporting these findings. However, consistency of SVAR-based estimates has already been evidenced in previous empirical work, for example, in the euro area by Camba-Mendez & Rodriguez-Palenzuela (2003) and Mazzi et al. (2016), who explains the consistency of SVAR methods with the view that these methods are not subject to the end-point problem because those can be viewed as one-sided filters.

This thesis seeks to fill some of the empirical gap by comparing the performance of the output gap estimates produced with the previously introduced structural VAR methods against the univariate HP filter using Finnish full sample and real-time data. These measures are examined across different samples and between methods, but the primary focus is on real-time performance. First, the small open economy SVAR decompositions and output gap estimates are examined against the Finnish economic history following the example of Chen and Góronicka (2020). Secondly, the ex-post revisions in the output gap measures are investigated as initially presented in Orphanides and van Norden (2002).

In general, few exercises of this kind, particularly dealing with structural VAR methods, have been done. Similar kinds of assessments with structural VAR methods can be found, for example, from Mazzi et al. (2016) and Chen and Góronicka (2020). Yet, none of the Finnish exercises of this kind have been performed by using Finnish real-time data. Several assessments have illustrated the issue associated with the ex-post revisions in the output gap estimates through quasi-real-time estimates, which are constructed using the final data with an iterative procedure by adding one quarter or year to each estimation period, for example, in the case of the HP filter as in Kotilainen (2019). However, structural VAR methods have not typically been under examination. Although, Billmeier (2006) compares a selection of frequently used output gap measures against the Finnish economic history throughout the 1990s, which includes an examination of the structural VAR method, introduced by Blanchard and Quah (1989) using annual full sample data.

The results of this thesis show how the output gap measures based on the applied structural VAR methods and the cyclical component of the HP filter form somewhat similar views about the cyclical development in Finland with quarterly full sample data, performing reasonably well against the Finnish economic history and macroeconomic conditions. Actual output appears to be above or under potential, mainly when it is supported by the history and macroeconomic indicators. However, the size of the measures varies, particularly around economic turning points. Sometimes differences between the measures may be even several percentage points, and even the sign of the estimate may differ when actual output is close to potential.

Some inconsistencies can also be observed between macroeconomic indicators and the output gap estimates. For instance, the output gap measures based on the structural VAR methods close to the end of the sample period suggest large positive output gaps for the late 2010s, which are unconvincing against the macroeconomic conditions. This is associated on the one hand, with the ability to distinguish long-run changes in output, defined as trend component of the SVAR, and on the other, the identification of persistent and temporary shocks on output, which are reconstructing the series of potential output and the output gap. The utilized variables, such as inflation and unemployment, together with the imposed long-run and sign restrictions, may embody only a little information about the shock under inspection, particularly if the variables are persistent. For instance, findings regarding the small open economy SVAR suggest that the sign restrictions which impose a negative relationship between GDP and CPI associated with supply shocks and a positive relationship between CPI and GDP associated with demand shocks are weak during times characterized by persistent and low inflation. Furthermore, as Blanchard (2018) has noted, so far, these methods are not designed to identify supply shocks without permanent effects on output or demand shocks with long-lasting effects on output which may end up resulting in inaccurately identified shocks. Another factor that may increase the uncertainty of the estimates with Finnish data is the pronounced volatility in output, as Billmeier (2006) has stated.

The consistency of the output gap estimates improves over time, suggesting that ex-post revisions tend to be greater at the end of the sample. GDP data revisions do not appear to be the primary source of revisions in the output gap estimates, suggesting that these models identify too slowly whether changes in output will be long-lived, particularly around economic turning points. However, exceptionally large revisions in actual output, as observed around the global financial crisis (GFC), may affect the estimates. All the applied measures suggest greater positive output gap estimates for the period of overheating prior to the global financial crisis with a full sample than with real-time data, suggesting that these models overestimate the level of potential output during the periods of overheating. The results also show how the structural VAR-based output gap measures also appear to react consistently to extreme economic turning points, but to gradual reversal of cycles are reacted with a delay.

In conclusion, the greater ex-post revisions at the end of samples suggest reduced ability to identify whether the nature of change in output will be permanent in real-time, which is contrary to the findings of Coibion et al. (2018) and Chen & Góronicka (2020) as well as Mazzi et al. (2016). In the case of the SVAR methods, this appears to be resulting from the ability to single out long-term changes in output from changes associated with the cyclical component in real-time.

Overall, the reliability of the real-time estimates did not prove to be particularly good. In fact, the consistency of the real-time estimates produced with the univariate HP filter outperformed structural VAR methods in the case of Finland. However, the small open economy SVAR appear to produce more reliable estimates than the BQ method in the context of Finland. An increase in the reliability in the output gap estimates also suggest that it is possible to improve the reliability of the estimates at the end of the sample by extending the series of utilized variables with well-formed forecasts, as evidenced by filtering techniques.

This sort of examination of the output gap estimates has not been performed before, in which the primary focus is on the structural VAR methods, and so far, none of the Finnish assessments have been performed using real-time data.

The organization of the thesis is divided up into sections. Section 2 presents frequently used potential output estimation approaches. Section 3 briefly introduces the previous literature regarding measurement errors, including the Finnish literature. Section 4 discusses the empirical methodology used in this thesis, including data and estimation techniques. Section 5 examines the output gap estimates based on the small open economy SVAR decompositions against the Finnish economic history. Section 6 examines the performance of the output gap measures constructed using real-time data. Section 7 discusses the robustness of the results. Section 8 conclude the thesis.

2. Overview of Potential Output Estimation Approaches

In the literature, potential output typically refers to the upper limit of output that an economy can sustainably produce. It is being linked to production capacity and capacity utilization, as well as productivity growth. From a theoretical point of view, the concept is usually built on Arthur Okun’s (1962) seminal notion of the upper limit of output that is possible to produce without generating price pressures. Therefore, it is also connected to such economic relationships as the non-accelerating inflation rate of unemployment (NAIRU) and Okun's law. However, there are several measures of potential output to choose from, between which the nature of potential output may differ due to the varying theoretical structures and implicit assumptions. (Coibion et al., 2018.)

The nature of potential output can be viewed differently depending on the underlying factors of the measurement. Several measurement methods identify potential output as a trend component of output, representing purely statistical features of output. (Álvarez & Goméz-Loscos, 2017.) Measurement methods based on neoclassical growth theories typically identify the level of potential output as a result of long-term changes in production capacity and productivity, obtained via the natural level of capacity utilization of factor inputs and productivity growth, representing changes in the supply-side factors of an economy. Methods that identify disturbances in output construct typically potential output either from permanent disturbances that are describing changes in aggerate supply or modeling steady-state of output under (im)perfect competition and flexible prices. (Álvarez & Goméz-Loscos, 2017.; Vetlov et al., 2011)

In the literature, the measurement methods of potential output are often presented in three broad classes of approaches following Mishkin (2007). Under the first class, potential output is constructed using statistical properties and relationships of variables that rely on statistical assumptions. The second class refers to growth accounting, where the factor of inputs is plugged into the aggregate production function of an economy. The third class covers structural methods typically consisting of dynamic stochastic general equilibrium (DSGE) models and structural vector autoregression models (SVAR). Alternatively, Álvarez and Goméz-Loscos (2017) used a broader classification of approaches that classifies methods either to univariate or multivariate methods, depending on a number of utilized variables.

In the following subsections, these approaches are briefly presented as proposed by Mishkin (2007). However, the presentation is not exhaustive. The primary focus is on some of the most frequently used methods, while an increasing number of applications and techniques can be found in the literature.

2.1. Statistical approaches

Statistical approaches can be divided into univariate and multivariate methods. The univariate techniques aim to capture the trend component of output from series of real GDP, interpreted as potential output. The task is relatively simple as it does not require any theoretical structures or assumptions about the structures of an economy under inspection. In contrast, multivariate statistical methods usually contain a little more structure that is built on relationships of variables. These relationships are typically based on empirically evidenced principles, and the resulting potential output is an unobserved component between these relationships. (Mishkin, 2007.)

Univariate statistical methods use either a filtering technique or model approach to separate cyclical and trend components from a series of output. These components differ depending on how these components are formed. (Álvarez & Goméz-Loscos, 2017.) Examples of univariate methods are simple autoregression models, linear/quadradic trending models, and numerous filters that aim to extract a statistical trend component from the series of real GDP (Coibion et al., 2018). In their original simplicity, model-based univariate approaches have assumed that the long-term development of economic time series is a function of time until the stochastic trend was recognized as being better in describing economic time series. Several univariate filtering techniques, in turn, aim to capture periodicity from series of real GDP, isolating high and low-frequency movements with pre-specified frequencies, in which the latter is representing movements in potential output. (Álvarez & Goméz-Loscos, 2017.)

The Hodrick-Prescott (HP) filter is one of the most used univariate filtering techniques for measuring potential output. It builds on an optimization problem between the compatibility of the trend with the actual output and the smoothness of the trend. The cyclical component at time t depends on the present, past, and future values of real GDP. The minimization problem of the HP filter is typically written in the form:

$$\min_{\tau}(\sum_{t=1}^{T}(y_t - g_t)^2 \lambda \sum_{T=2}^{T-1}[(g_{t+1} - g_t) - (g_t - g_{t-1})]^2) \quad (1)$$

The minimization builds on two parts; fitting the trend to the GDP series represented in the first part of the equation and determining a degree of smoothness. The degree of smoothing must be chosen via parameter λ, which penalizes the changes in trend output. The higher is the chosen value, and the greater is the penalty. (Hodrick & Prescott, 1981.) For instance, the smoothing parameter has been presented to pass fluctuations less than eight years, when it is set to 1600 with quarterly data (Hodrick & Prescott 1997).

The HP filter is easy to produce with GDP data of any country, enabling an internationally comparable estimate of potential output (Blagrave et al., 2015; Álvarez & Goméz-Loscos, 2017). The underlying assumption is that potential output varies smoothly over time, and the smoothing parameter determines the degree of smoothness (Álvarez & Goméz-Loscos, 2017). However, the pre-selected smoothing parameter may seem arbitrary if an ideal selection of smoothing would be based on data. (Hamilton, 2018)

Another problem that limits the use of the HP filter is the end-of-sample bias. The most recent estimates produced with the HP filter tend to be subject to greater adjustments because the cyclical component is built from the past, present, and future values of output. Therefore, many practitioners of this method mitigate the problems at the end of the sample by extending the time series of output with well-formed forecasts. (Blagrave et al., 2015, Hamilton, 2018) In addition to these factors of uncertainty, the HP filter has also been evidenced to produce spurious dynamic relations. (Cogley & Nason, 1995; Hamilton, 2018).

Multivariate statistical methods introduce additional variables into the process of measurement, which are assumed to contain information on the development of potential output. Multivariate filters and unobserved component models have been typically seen representing multivariate statistical approaches, in which potential output is as an unobserved component, between the statistical relationships of variables, such as inflation, unemployment, and output. These methods often embody slightly more structure, as they may utilize economic relationships built on empirical principles of macroeconomics, such as the Okun's Law and the Phillips curve. (Álvarez & Goméz-Loscos, 2017.) One of the early examples utilizing a relationship between the output gap and inflation was introduced by Kuttner (1994). More examples that utilize a connection between inflation, unemployment, and the output gap can be found from Benes et al. (2010) and Blagrave et al. (2015). Furthermore, some of the approaches have also introduced financial variables into models, aiming to utilize information associated with increasing asset prices. (Borio et al., 2014; Melolinna et al., 2016)

One of the advantages of the multivariate filters and the UC models is indeed the ability to incorporate additional variables and lag structures into a model. These augmentations can be introduced quite easily and may eventually enhance the output gap estimates. For instance, some of the UC-based estimates have been found to be more consistent relative to other methods. (Marcellino & Musso, 2011; Kangur et al., 2019; Sariola, 2019.)

Overall, statistical approaches are typically known from loose theoretical structures. At the same time, it has been one of the debated issues of these approaches. For instance, the statistical trend component of output obtained using statistical methods, particularly in the case of univariate filters, represents purely statistical features in the series of output. Hence, the link to the theoretical concept of potential output is thin, leaving a vacuum for economic and theoretical reasoning. However, multivariate statistical methods may overcome some of the limitations by adding more structure by introducing variables that may embody information on the development of potential out and the output gap. (Blagrave et al., 2015.)

2.2. Production function approaches

Production function approaches are built on a relationship between output and factors of inputs, expressed in an aggregate production function of an economy and thus are linked to neoclassical growth theories. The method itself can be thought of as a way of accounting growth through factor inputs that are plugged into the aggregate production function of an economy. (Mishkin, 2007). The level of potential output is produced by entering the level of each input that meets their "normal" capacity utilization rate, and thus are reflecting the supply potential of an economy (Havik et al., 2014, Álvarez & Goméz-Loscos, 2017). In other words, the utilization rate of labor and capital input should be at their natural rate, and the total factor of productivity should meet its level of long-term growth. Hence, the challenge is to construct “normal” utilization rates of inputs as well as the total factor of productivity. Furthermore, choosing and solving a production function usually requires a handful of assumptions. (Havik et al., 2014.)

First, one must choose a functional form of a production function and specifications for factor inputs depending on it to describe the production of an economy (Álvarez & Goméz-Loscos, 2017). For instance, the European Commission applies the Cobb-Douglas production function for its member states, in which potential output is obtained through a combination of factor inputs, labor, and capital, multiplied by total factor productivity (TFP), that captures all other factors (Havik et al., 2014). Another more general functional form that has been applied in the measurement of potential output is the constant elasticity substitution (CES) production function, which was first introduced by Solow in 1956. The CES production function distinguishes changes in labor augmenting productivity and capital augmenting productivity. The constant substitution elasticity between labor and capital is the key parameter characterizing economic efficiency and output growth in the CES production functions.

The Cobb-Douglas production function is a currently widely applied functional form among public institutions and international organizations. Potential output is obtained through a mixture of factor inputs, consisting of the total factor of productivity (A), Labor (L), Capital Stock (K). The function itself is based on the work of Paul Douglas and Charles Cobb, and the general functional form can be written as follows:

$$Y = AL^{\alpha}K^{1-\alpha} \qquad (2)$$

α and $1-\alpha$ are the output elasticities of labor and capital. The Cobb-Douglas production function used by the European Commission (EC) can be written:

$$Y = (U_L L E_L)^{\alpha}(U_K K E_K)^{1-\alpha} = (E_L^{\alpha} E_K^{1-\alpha})(U_L^{\alpha} U_K^{1-\alpha})L^{\alpha}K^{1-\alpha} \qquad (3)$$

where the total factor of productivity is

$$TFP = (E_L^{\alpha} E_K^{1-\alpha})(U_L^{\alpha} U_K^{1-\alpha}) \qquad (4)$$

Y represents the potential output, L is labor, and K is the capital stock. These are corrected for the degree of excess capacity $(U_L^{\alpha} U_K^{1-\alpha})$ and adjusted the level of efficiency $(E_L^{\alpha} E_K^{1-\alpha})$. The total factor of productivity (TFP) summarizes a technological level of factor inputs as well as a degree of utilization. (Havik et al., 2014.)

Calculating potential output using the Cobb-Douglas function presented above requires a range of assumptions, including assumptions about returns to scale and factor price elasticity. The methodology of the EC assumes that returns to scale are constant, and the factor price elasticity is set to one. The elasticities of labor and capital on output are estimated from the wage share by adding assumptions about perfect competition. The potential use of capital stock in factor inputs can be defined by the full utilization of the standing capital. The series capital stock itself is used as such as it represents the total capacity of the economy. The labor input is calculated from the trend labor force, non-accelerating wage rate of unemployment NAWRU, and trend hours worked. The natural efficiency of factors of inputs is obtained through the relationship between the TFP cycle and the degree of capacity utilization by using a bivariate Kalman filter. (Havik et al., 2014.) After the natural level of the factor of inputs are determined, the potential output can be summarized as follows:

$$Y^P = (L^P E_L^T)^{\alpha}(K E_K^T)^{1-\alpha} \quad (5)$$

One of the key advantages of the production function approach is that the growth of potential output is driven by a range of factors. However, a great deal of uncertainty may be associated with the estimated factors of inputs that are plugged into the aggregate production function, which may eventually affect the estimates of potential output. For example, the estimates on the level of the natural rate of unemployment may be subject to structural changes, measurement errors, and inconsistencies in the data. (Mishkin, 2007.)

The equations in production function approaches are usually completed with the assumptions regarding on dynamics of the unobserved components and country-specific economic structures. (Havik et al., 2014.) Therefore, production function approaches typically require a detailed description of underlying production factors of an economy, and thus a single functional form may not be as suitable for economies with differing characteristics.

2.3. Structural approaches

Structural approaches are typically built on the Keynesian view of cyclical fluctuations. Potential output is obtained through the identification of disturbances in output that are representing shocks hitting the economy. These models are typically either used to project the economy or are known from empirical research of macroeconomics. Broadly speaking, potential output is typically constructed either from permanent disturbances in output that are representing changes in aggerate supply or by modeling an efficient level of output under (im)perfect competition and flexible prices. (Álvarez & Goméz-Loscos, 2017.; Vetlov et al., 2011) Methods that are built on structural vector autoregression models reconstruct potential output from historical contributions of shocks by summing persistent shocks on output over time. Dynamic stochastic general equilibrium (DSGE) models typically model the level of potential output through the identification of shocks hitting the economy, which are identified by calibrating a model where rigidities are characteristic (Álvarez & Goméz-Loscos, 2017). Once the shocks are identified, the effects of chosen shocks can be removed, and rigidities can be eased to obtain the level of potential output (Andrés et al., 2005).

In the DSGE models, the potential output can be interpreted either a) as a long-term growth of output defined by persistent technology shocks, b) as an efficient level of output that prevails when prices and wages are fully flexible under perfect competition, or c) as a natural level of output under imperfect competition with fully flexible prices and wages (Vetlov et al., 2011). Real and nominal rigidities are usually characteristic features for DSGE models, denoting that prices, wages, or capital do not immediately adjust to identified shocks. Hence, transitory deviations from the potential output are a consequence of delayed adaption in prices and nominal wages. (Vetlov et al., 2011.; Newby & Orjasniemi, 2012.)

Methods that are built on structural vector autoregression models typically aim to describe changes in potential output through structural shocks hitting the economy. The identification scheme of shocks is typically built on the Keynesian view of cyclical fluctuations, in which changes in output are decomposed to temporary and permanent disturbances. The latter characterize changes in aggregate supply, whereas temporary disturbances illustrate changes in the demand-side of output. Hence, the changes in aggregate supply represent movements in potential output, and thus changes in potential output can be calculated from the initial period by summing disturbances that have had permanent effects on output. (Álvarez & Goméz-Loscos, 2017.)

The Blanchard & Quah (BQ) method, which was first introduced in 1989 by Olivier Blanchard and Danny Quah, builds on a bivariate structural vector autoregression (SVAR) model with a long-run restriction that identifies two sorts of disturbances in output, temporary and permanent. The model itself is built on the Keynesian framework and Stanley Fisher's nominal wage contracting theory. The method plugs (log) real GDP growth (Δy_t) and unemployment rate (u_t) into a structural vector autoregression (SVAR) model, in which structural shocks ε_t^s and ε_t^d are assumed to be uncorrelated. The bivariate SVAR system that follows stationary process can be written in matrix form:

$$\begin{pmatrix} \Delta y_t \\ u_t \end{pmatrix} = \begin{pmatrix} Init_{0,1} \\ Init_{0,2} \end{pmatrix} + \begin{pmatrix} \beta_{1,11} & \beta_{1,12} \\ \beta_{1,21} & \beta_{1,22} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \beta_{2,11} & \beta_{1,21} \\ \beta_{2,21} & \beta_{1,22} \end{pmatrix} \begin{pmatrix} \Delta y_{t-2} \\ u_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^d \end{pmatrix} \quad (6)$$

$$\Sigma_e = \begin{pmatrix} var(\varepsilon_t^s) & cov(\varepsilon_t^s, \varepsilon_t^d) \\ cov(\varepsilon_{t,}^s \varepsilon_t^d) & var(\varepsilon_t^d) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (7)$$

where $Init_j$ is a vector of initial values, B_j is a (2x2) matrix of coefficients for lags j=0,1,2. The identification of disturbances is based on the assumption that demand disturbances are assumed to have short-run effects on output and unemployment, which eventually disappear in the long-run, while only supply disturbances are assumed to have long-run effects on output. This is achieved by imposing a zero long-run restriction on the cumulative effect of demand disturbances ε_t^d on (log) real GDP growth (Δy_t) on the impact matrix:

$$\begin{pmatrix} \Delta y_{t,t+\infty} \\ u_{t,t+\infty} \end{pmatrix} = \begin{pmatrix} d_{11} & 0 \\ d_{21} & d_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^d \end{pmatrix} \quad (8)$$

The contributions of each shock at each point in time can be obtained by computing historical decompositions of the VAR. The BQ method decomposes (log) real GDP growth (Δy_t) to the initial long-run rate of GDP y_{init}, the growth rate of GDP due to permanent disturbances $\sum_{i=1}^{t} \varepsilon_t^s$, and rate of GDP due to temporary disturbances $\sum_{i=1}^{t} \varepsilon_t^d$, which can be written as follows:

$$\Delta y_t = y_{init} + \sum_{i=1}^{t} \varepsilon_t^s + \sum_{i=1}^{t} \varepsilon_t^d \quad (9)$$

The accumulated permanent disturbances represent movements in potential output, while the output gap can be defined through the cumulative temporary disturbances. Hence, potential output Δy_t^{pot} can be defined as:

$$\Delta y_t^{pot} = y_{init} + \sum_{i=1}^{t} \varepsilon_t^s \quad (10)$$

The identification strategy of the BQ method simply allows either transitory or persistent effects on output. The underlying assumption is that temporary changes are driven mainly by demand-side factors, while permanent changes are driven by supply-side factors. However, the reliability of this simplified assumption has been questioned whether it is consistent with the true nature of shocks, and are the shocks appropriately identified (Coibion et al., 2018.) Blanchard (2018) notes the method may inaccurately identify shocks that may affect potential output temporarily instead of permanent effects, which are misleadingly identified as demand shocks. Likewise, some shocks identified as demand shocks may affect potential output for some time, which will not show in the estimates of potential output. On the other hand, the findings of Blanchard (2015) also suggest that demand-triggered recessions may have persistent effects on output. Also, the reliability of the long-run restrictions in some cases has already been questioned by Faust and Leeper (1997).

One of the challenges and strengths of this kind of approach is indeed the coordination between economic theory and identified shocks. More specifically, whether the imposed restrictions identify disturbances in output appropriately while being consistent with economic theory and economic reasonableness. The BQ method may not sufficiently describe disturbances in output if the cyclical demand component of output is not driven by the unemployment rate. In such a case, temporary effects on output are not, in fact, demand-driven. (Billmeier, 2006.)

However, one of the features that can be considered as an advantage of the BQ method when measuring potential output is that it does not restrict potential output to a random walk process. This enables more dynamics and gradual adjustments of output to technology shocks. (Álvarez & Goméz-Loscos, 2017.) Another advantage of the BQ method is the reliability of the estimates at the end of the sample (Billmeier, 2006). These methods can be viewed as one-sided, which is one of the explaining factors of the consistency in SVAR methods in real-time. (Mazzi et al., 2016) The findings of Coibion et al. (2018) also suggest that the real-time estimates of potential output produced using approaches, such as the BQ method, are characterized by the lesser sensitivity to temporary changes in output.

3. Review of the Previous Literature

3.1. Ex-post revisions and measurement errors

The intrinsic uncertainty is present in the measurement of both potential output and the output gap, particularly in real-time. Measurement errors become visible in ex-post revisions in the estimates that are typically resulting from erroneous measurements in real-time. The sources of unreliability in the output gap estimates have been illustrated comprehensively in the seminal paper of Orphanides van Norden (2002). They used a selection of frequently used detrending methods to illustrate how the uncertainty associated with real-time estimates of the output gap can be traced either to a selected method that is used for the construction of potential output or to data used in the estimations. Their result shows how the ex-post revisions in the output gap estimates were primarily related to model problems at the end of the sample rather than data revisions. These methods underestimated the output gaps, particularly in the peaks of the cycles.

Orphanides and van Norden (2002) also discussed three sources of uncertainty that may cause ex-post revisions from the real-time estimates: a) data used in estimations may get revised, b) extending time series may change the observed position in a business cycle, c) arrival of new data may force us to revise the used model. The sources of uncertainty may be interlinked and, therefore, cannot be necessarily completely become isolated from one another. For instance, the uncertainty associated with a selected method is also linked to the uncertainty associated with extending time series because several methods may produce estimates that differ particularly between the end and middle of the sample.

The uncertainty associated with the end of the sample was documented already by Kuttner (1994) and St-Amant and van Norden (1998). Orphanides (1998; 2000) brought up the measurement errors in the official output gap estimates. Many authors have also examined the issue in the context of the policy implications, including Kuttner (1992), Orphanides (1998, 2001), and Smets (1998). For instance, Smets (1998) and Orphanides (2001) examine real-time policy recommendations based on Taylor's rule applying real-time data. Furthermore, Orphanides (2003a, 2003b) has elaborated the issue of measurement errors in the context of activist stabilization policy.

After introducing the underlying factors of the uncertainty, Orphanides van Norden (2002) has shown the way for several studies and assessments regarding the unreliability of the output gap estimates. Marcellino and Musso (2011) showed how the ex-post revisions in the Euro area output gap estimates are largely associated with the end-of-sample problem in both univariate and multivariate methods. However, their result also suggests that some of the UC models may produce more reliable estimates in real-time. They also note that the estimates produced with the multivariate models do not necessarily consistently outperform the estimates produced with the univariate methods. Guisinger et al. (2018) compare five trend measures and the CBO measure of the U.S. and finds highlighted measurement errors between the measures around turning points and at the edge of the sample when using real-time data. However, they also show that these differences do not influence the policy perception constructed using a typical Taylor rule.

Another strand of the literature examines the uncertainty in the output gap estimates of international organizations that are mainly produced with the production function approach. For example, Tosetto (2008) and Turner et al. (2016) examine estimates of the OECD produced with the production function approach and find large (and persistent) revisions in the real-time estimates of the OECD countries. Turner et al. (2016) also showed how the reliability of the OECD output gap estimates could be improved by introducing additional cyclical adjustments, especially for labor efficiency.

Part of the empirical work has been focused on comparing the estimates produced by different institutions. Kempes (2012) finds downward bias in the estimates of the 15 EU member states produced by the EC, IMF, and OECD. Virkkola (2013) shows how the real-time estimates of the European Commission (EC) and the IMF have sometimes differed significantly. The findings of Mc Morrow et al. (2015) suggest that the production function methodology of the EC outperforms the previously used HP filter in real-time, particularly in economic turning points. They also highlight the fact that the bulk of the problem in the most frequently used measures of potential output, including the production function methodology of the EC, is the ability to capture economic upturns in real-time. Hernández de Cos et al. (2016) points out that the direction of revisions in the estimates of the EC is procyclical with respect to the current state of the economy. Kangur et al. (2019) find that most measurement errors in the estimates of the IMF are associated with mismeasurement of recessions, which typically can be observed as a limited decline in potential output.

Overall, the reliability of the estimates has also been found to improve. Edge & Rudd (2016) and Champagne et al. (2018) point out how revisions in output gap estimates of the Federal Reserve Board and the Bank of Canada's staff have decreased over time. Champagne et al. (2018) explain the improved reliability through the staff experience, the development of new methods, and the increased amount of real-time data.

Previous empirical research has also presented evidence that some of the estimation methods are less sensitive to real-time mismeasurement, and part of the methods have been specifically developed to encounter the issue associated with mismeasurement. For instance, Borio et al. (2014) and Melolinna et al. (2016) aim to mitigate the real-time measurement errors by introducing financial indicators into their models. The underlying idea is that financial information may deliver information on increasing asset prices. (Borio et al., 2014). Blagrave et al. (2015) increase the reliability of the real-time estimates by adding inflation and growth expectations into a multivariate filter. Alichi et al. (2017, 2018, 2019) tackle the measurement errors by introducing alternative specifications and model extensions that utilize capacity-utilization rate, labor market hysteresis, and monetary policy rule.

One of the acknowledged difficulties in the measurement of potential output in real-time is the ability to identify whether changes in output will be permanent. Coibion et al. (2018) present evidence that real-time estimates of potential output produced with some of the widely used methods, such as the HP filter and production function approach, respond too strongly to transitory shocks. As an alternative, they suggest methods, such as Blanchard and Quah (1989), that reconstruct potential output from distinguished permanent effects on output and therefore being less sensitive to temporary effects in real-time. Chen and Góronicka (2020) further investigate this matter by applying several filtering techniques and structural VAR methods to the U.K. data. They also propose an extended SVAR that builds on the identification scheme constructed by Forbes et al. (2018), that allows a higher number of disturbances in output and meets some of the characteristic features for a small open economy. The results also suggest that structural VAR methods outperform filtering techniques in real-time due to the lesser sensitivity to transitory shock, which appears in lower ex-post revisions in the estimates. However, the reliability associated with SVAR-based output gap measures has been discussed before. Camba-Mendez & Rodriguez-Palenzuela (2003) and Mazzi et al. (2016) compare a range of output gap measures for the euro area and show that the consistency of SVAR-based output gap measures is high relative to filtering techniques. They consider that one of the explaining factors of the consistency in SVAR methods in real-time is the view that these methods are one-sided and therefore are not subject to the end-point problem.

3.2. The literature in the context of Finland

In the Finnish literature, little empirical work has been done comparing the measures of potential output and the output gap using real-time data. Ex-post revisions and measurement errors are typically examined through quasi-real-time estimates as a part of a survey of several estimation approaches or as a part of a more comprehensive examination of a single method. The most widely applied estimation approaches have been under examination more frequently. For instance, the production function methodology of the European Commission has gained much attention in the Finnish literature due to its relevance within the EU's fiscal framework. Moreover, several assessments have reviewed a selection of estimation approaches, sometimes consisting of illustration of ex-post revisions through the quasi-real-time estimates.

Billmeier (2006) compares a selection of frequently used output gap measures against the Finnish economic history in the 1990s and argues that one of the factors that increase the uncertainty of the estimates is the pronounced volatility in output. The results also show how many of the output gap measures do not have the ability to predict inflation in an out-of-sample simulation of a Phillips curve. However, the output gap measures based on the BQ decomposition, and a frequency domain approach, are the most capable in this exercise. As a conclusion, Billmeier (2006) suggests that further refinement of the output gap measures, for example, a larger VAR with quarterly data, could improve the Finnish output gap estimates. Melolinna (2010) outlines a few simple output gap measures and the concept of the output gap in the context of Finland. The illustrative part shows how the output gap estimates produced with the HP filter have changed over time for Finland and in the Euro area. Kotilainen (2019) overviews different potential output and the output gap measures in the context of Finland. The measures produced with statistical approaches are examined against the Ministry of Finance estimates, produced with the production function methodology of the European Commission. The review shows how quasi-real-time estimates produced with the selected statistical methods revise over time. The full sample estimates differed from quasi-real-time estimates, particularly around the global financial crisis.

The European Commission (EC) production function methodology has gained much attention in the Finnish literature. This is mainly because the output gap estimates are plugged into the measures of structural budget balance (Mourre et al., 2014). Hence, several studies have examined the fiscal policy implications of the output gap measures of the EC in the context of Finland. Kuusi (2015) and Huovari et al. (2017) examine the performance of the output gap measures of the EC through the factor of inputs in the context of the structural budget balance. Their findings highlight the uncertainty associated with the estimates of NAWRU. Kuusi (2015) also examines the EC's output gap estimates using quasi-real-time estimates of NAWRU.

Jysmä et al. (2019) seek to develop the production function methodology of the EC in the context of Finland. They apply the constant elasticity of substitution (CES) function to the Finnish economy with filtration of the potential output using the Sequential Monte Carlo method. Their results suggest that the more general CES function would be better in describing production factors for Finland. However, real-time uncertainty in the NAWRU and labor augmenting productivity remains high, causing uncertainty in the estimates.

Most of the publications of the Bank of Finland have mainly focused on the most frequently used estimation approaches, but also on the Bank of Finland's DSGE model called AINO. Haavio (2008) shows how the Cobb-Douglas production function and the HP filter construct fairly similar views about the cyclical development in Finland, whereas the output gap measures produced with the Bank of Finland's DSGE model prove to be smaller in absolute value. Newby and Orjasniemi (2012) also compare output gap estimates based on the Aino model and production function approach and illustrate how the estimates based on the Aino revise more as the time series expands. Sariola (2019) applies a somewhat different approach by introducing a model that combines features from unobserved components (UC) models and the production function approach. The results show how the quasi-real-time estimates mostly consistent with the ex-post estimates of the IMF and the EC, even in extreme turning points. The UC model-based output gap measure suggests that potential output growth slowed down in the aftermath of the global financial crisis when actual output dropped below potential and remained there for almost a decade.

The previous empirical research that can be found from the Finnish literature examines measurement errors typically from the quasi-real-time estimates. Yet, none of the Finnish exercises were performed with actual real-time data. Therefore, the estimates do not consider data revisions and thus do not reveal the complete error in real-time estimates due to the lack of real-time data. Instead, these exercises are frequently carried out using the final data with an iterative procedure by adding one quarter or year to each estimation period, defined as quasi-real-time estimates. However, genuine real-time estimates cannot often be constructed due to the fact that real-time data may not be available. On the other hand, ex-post revisions in the estimates have been found to be mainly associated with the model mismeasurement at the end of the sample rather than data revisions, which is why exercise with actual real-time data may not be necessary.

The previous literature suggests that the unreliability of the output gap estimates appears to be high, particularly in economic turning points, as already shown in Orphanides and van Norden (2002). Measurement errors often occur as underestimations of the output gaps, resulting from the overestimation of potential output around cyclical peaks as initially showed in Orphanides and van Norden (2002), or inability to predict recessions, as presented in Kangur et al. (2019). Ex-post revisions in the output gap estimates are typically associated with measurement errors at the end of the sample, rather than data revisions, according to the findings of Orphanides and van Norden (2002) and Marcellino and Musso (2011). However, the previous literature has often approached the matter via statistical approaches and production functions rather than structural approaches. Therefore, findings of Coibion et al. (2018) and Chen & Góronicka (2020), as well as Mazzi et al. (2016), are fascinating as they suggest that the estimates of potential output and the output gap reconstructed using structural VAR decompositions are less sensitive to temporary disturbances in output as well as the end of the sample problem, and thus are more reliable in real-time. Yet, there is still relatively little empirical evidence supporting these findings.

4. Empirical Methodology

The key interest of this thesis is to examine the performance of the output gap measures constructed applying structural VAR methods. These estimates are examined against the Finnish economic history, but the primary focus is on real-time performance. More specifically, to what extent these measures tend to be subject to ex-post revisions resulting from measurement errors in real-time. For this purpose, two structural VAR methods and the univariate HP filter are applied to the Finnish full sample and real-time data following Orphanides and van Norden (2002).

The empirics consist of two parts. First, the small open economy SVAR, which builds on the identification scheme constructed by Forbes et al. (2018), is applied to the Finnish data following Chen and Góronicka (2020). The resulting output gap estimates, shock contributions, and quarter-on-quarter (log) GDP growth decompositions are examined against the Finnish economic history from 1964Q4 to 2020Q2. Because of the sensitivity to structural breaks, the complete dataset is divided into three sub-datasets, following significant structural changes in the Finnish economy. Secondly, the performance of three output gap measures is compared using the Finnish full sample and real-time data, constructed with the small open economy (SOE) SVAR, the Blanchard & Quah (BQ) SVAR method, and the HP filter.

The ex-post revisions in the estimates are examined as initially presented in Orphanides and van Norden (2002). For this purpose, real-time and quasi-real-time estimates are constructed using the Finnish data from 1996Q3 and 2020Q2. The real-time estimates are constructed performing estimations using the data vintage available at each point in time starting from 2008Q1 until 2020Q2. In this section, alternative real-time estimates are also constructed by removing a few of the estimates from the end of the samples, defined as delayed real-time estimates. These estimates can be considered being constructed with "perfect forecasts," i.e., the original published real-time is used in the estimations, but only t-2, t-4 or t-8 estimates have been used.

4.1. Identification schemes and estimation techniques

The applied identification scheme in the small open economy SVAR proposed by Chen and Góronicka (2020) follows the identification scheme constructed by Forbes (2018). Six structural shocks on output are identified using the following variables: real GDP growth, consumer price inflation (CPI) at constant tax, interest rate, changes in the effective exchange rate, import price inflation, and changes in foreign export prices. Structural shocks are identified by combining sign restrictions and a mixture of short- and long-run restrictions, as shown in Table 1.

Table 1. SOE SVAR: Identification Restrictions

	FI Supply shock	FI Demand shock	Monetary policy shock	Exogenous exchange rate shock	Persistent global shock	Transitory global shock
			Short-run restrictions			
FI GDP growth	+	+	-			
FI CPI	-	+	-	-		
FI Interest rate			+	-		
FI Nominal EER			+	+		
FI Import prices						
World (ex-FI) prices	0	0	0	0	+	+
			Long-run restrictions			
FI GDP growth		0	0	0		0
FI CPI						
FI Interest rate						
FI Nominal EER						
FI Import prices						
World (ex-FI) prices	0	0	0	0		

Note:

1) +/- sign restricts the impulse response of the variable either to be positive or negative

2) 0 restricts the impulse response of the variable to be zero

Source: Author's modifications, Chen & Góronicka (2020), Forbes (2018).

The restrictions in the identification scheme aim to restrict other than domestic supply-side shocks and persistent global shocks, to have persistent long-run effects on output. The distinction between domestic and global shocks is achieved by using zero restrictions; domestic developments do not affect export prices, while shocks with global origin may affect prices of export and the domestic development of the economy. A sign restriction that builds a negative relationship between CPI and GDP represents domestic supply shocks. It is constructed for the first two periods. Domestic demand shocks are identified through a positive correlation between GDP and CPI. (Forbes et al., 2018) However, a domestic demand-driven appreciation of the domestic exchange rate or interest rate is not imposed as in Chen & Góronicka (2020) and Forbes (2018). These are based on the assumption that domestic demand shocks in Finland do not affect the exchange rate of the euro or the policy rate of the European Central Bank. The Finnish exchange rate has also been mainly fixed for the period before the currency union, except between 1992 and 1998, when the exchange rate of the Finnish markka floated. Therefore, the assumption is also used for the periods before the currency union. As in Chen & Góronicka (2020) global shocks are allowed to have permanent effects on global output, but it is not imposed. As a result, domestic demand shocks with long-run effects on domestic output are nested.

The imposed sign restrictions in the extended SVAR have been used extensively (Fry & Pagan, 2011). For instance, a negative relationship between CPI and GDP posed for the first two periods aims to ensure that identified the domestic shocks that lead to a permanent shift in the level output are sourced from supply-side factors. A positive relationship between domestic demand shocks, CPI, and GDP is based on the assumption that positive demand shocks lead to an increase in prices and output. Shocks originated from the changes in monetary policy are identified through lower interest rates associated with depreciation of nominal exchange rate and increase of CPI and GDP. An exogenous appreciation of the exchange rate, in turn, leads to a decrease in CPI. (Forbes et al., 2018.)

One of the challenges and strengths of this kind of approach is the coordination between economic theory and identified shocks. More specifically, whether the imposed restrictions identify shocks hitting on output appropriately while being at the same time consistent with economic theory and economic reasonableness. (Forbes, 2018) The identification schemes of the BQ method (1989) and Forbes et al. (2018) distinguish permanent and temporary changes in output. However, an over-strict limit between temporary and permanent disturbances in output may end up producing shocks that are not in line with their true nature. Blanchard (2018) notes that some demand shocks may, in fact, affect potential output for a while instead of permanent effects.

The extended SVAR is estimated following Forbes et al. (2018), with two lags for each estimation period using Bayesian methods with Minnesota-style priors, as presented in Binning (2013). The estimations are carried out using MATLAB codes build on the replication files of Forbes et al. (2018). The lag lengths for each subsample period are selected according to Akaike information criteria (AIC). The estimations are based on a Gibbs sampling procedure, and the final 1000 repetitions are saved and used. The restrictions are simultaneously imposed on each shock for two periods during each quarter. The short- and long-run algorithm is based on the work of Rubio-Ramirez et al. (2010) and extended by Binning (2013) for under-identified models. The cumulative impulse responses for each estimation period can be found in the annex.

The BQ method is estimated following Blanchard and Quah (1989). The estimations are carried out by using the MATLAB replication codes written by Ambrogio Cesa-Bianchi, which are designed for the VAR Toolbox and include computations for impulse response functions, historical and forecast error variance decompositions. Real GDP data is transformed into logs, and the unemployment rate is detrended. The model is estimated using 3 and 5 lags which are selected according to the Akaike information criterion (AIC). The first sample period between 1996Q3 and 2020Q2 is estimated with three lags. The second from 1980Q1 until 1996Q2 with 5. The cumulative impulse responses for real GDP growth and the unemployment rate can be found in the annex.

The HP filter is estimated using the value 1600 for the smoothing parameter, as recommended when using quarterly data. The estimations are performed by using MATLAB. The end-of-sample bias was not mitigated with the extended samples. More information regarding the robustness of the results can be found in section 7.

4.2. Data

The complete dataset used in this thesis is divided into three sub-datasets considering significant structural changes in the Finnish economy. At the beginning of the first sample period, the Finnish economy was heavily regulated, which continued until the end of the 1970s. The period of economic regulation was followed by gradual liberalization of the economy towards the free market conditions. Finland joined the European Union in 1995 and later became part of the currency union. These eras form subsamples, defined in this thesis as follows: a period of economic regulation (1964Q1-1979Q4), liberalization (1980Q1-19962), and currency union (1996Q3-2020Q2).

The extended SVAR model is estimated using quarterly data for Finland and the rest of the world over the period from 1964Q1 through 2020Q2 using the following six variables: real GDP growth, CPI inflation, monetary policy rate, changes in the nominal effective exchange rate index (NEER), import price inflation, and changes in foreign export prices. The estimations are performed in three parts following the subsamples.

The period of currency union contains real GDP data that is seasonally adjusted by Statistics Finland. A version of harmonized consumer price index is used, including the contribution from VAT changes (cf. Chen & Góronicka 2020). Domestic monetary policy is a combined series of Bank of Finland's (BoF) tender rate from 1996Q3 until 1998Q4 and European Central Bank's MRO/Deposit rate from 1999Q1 to 2020Q2. The nominal effective exchange rate index (NEER) is collected from the Bank of International settlements (BIS). The basket of the narrow index is used that consists of 27 economies. The collected monthly data is transformed into quarterly averages. It is also worth mentioning that the Finnish markka was the account currency until the beginning of 1999. However, the Finnish markka was pegged to the EU Exchange Rate Mechanism (ERM) in 1996, until it was converted to the euro from the beginning of 1999. Import price inflation is described using an import price deflator that is formed from a seasonally adjusted import series collected from Statistics Finland. Export prices are constructed from the annual world CPI series weighted by the Finnish export share. The world CPI and the export share are collected from the IMF database. The annual point figure for the world CPI is used for each quarter within a year.

For the period of economic liberalization, real GDP data are seasonally adjusted by Statistics Finland. The harmonized consumer price index is chained backward using the headline consumer price index. Domestic monetary policy is measured with a combination of the monthly average of BoF intraday credit rate from 1980q1 until 1992q2 and BoF tender rate from 1992q3 until 1996q2. The narrow nominal effective exchange rate index (NEER) is collected from the Bank of International settlements (BIS). Import price deflator is formed from seasonally adjusted import series collected from Statistics Finland. Export prices are constructed from the annual world CPI series weighted by the Finnish export share. The world CPI and the export share are collected from the IMF database.

For the period of economic regulation, neither data for import deflator nor world CPI were available and thus alternative data was required for describing import price inflation and export prices. Therefore, both import and export prices are replaced with their counterparts in the producer price index (PPI), collected from Statistics Finland. Domestic monetary policy is measured using monthly averages of the central bank's lending rate for commercial banks from 1964q1 until 1975q1, and a monthly average of BoF intraday credit rate between 1970q1 and 1979q4. Otherwise, the applied data is consistent with the period of economic liberalization.

The quarterly series of real GDP and import are extended backward by using historical quarterly national accounts for the period between 1964 and 1990[1]. Transformation of quarterly (log) differences are used except for the interest rate. A downward trend in the series of interest rates is treated by using detrended series. Real GDP is transformed to logs in the estimations of potential output and the output gap.

In section six, the estimations of the small open economy SVAR are performed for all subsample periods by using the alternative data for import and export (i.e., import and export prices are drawn from PPI). The bivariate structural VAR is estimated as in Blanchard and Quah (1989), using the extended series of quarterly real GDP data and annual unemployment rates. Annual employment rates are copied to each quarter within the same year. The unemployment rate is collected from Statistics Finland for the period of the currency union. The series is extended for the period of economic liberalization using year-on-year estimates constructed by the International Labor Organization (ILO). The univariate HP filter is estimated separately for each subsample using the extended quarterly real GDP data. Problems associated with the end-of-sample are not mitigated with forecasts.

5. Extended SVAR: Quarterly Growth Decompositions and Past Output Gaps

At first, the small open economy SVAR that is built on the identification scheme constructed by Forbes et al. (2018) is applied to the Finnish full sample data following the example of Chen and Góronicka (2020). The method's performance is examined by running separate regressions to construct quarterly growth decompositions, series of the output gap estimates, and shock contributions from alternative combinations of zero and sign restrictions. Both the quarterly growth decompositions and the output gap estimates are examined against the Finnish economic history from 1964Q4 to 2020Q2. However, a few of the observations are lost around the structural breaks at the

Compilation methods and classifications of historical quarterly national accounts differ in many respects from the methods and classifications currently in use (see Official Statistics of Finland (OSF) (2013).

beginning of each subsample period. Quarter-on-quarter (log) GDP growth is decomposed to the rate of growth in the long-term and the growth rates due to the shocks introduced in Table 1.

Two different types of output gap measures with distinguishing interpretations of potential output are constructed by using the small open economy SVAR. The first measure is reconstructed by summing the persistent effects on output resulting from domestic supply and persistent global shocks, as in Chen and Góronicka (2020). Potential output can be affected by persistent global shocks which can be sourced, for example, from technology shock that reduces the cost of imports and increases domestic productivity, or from the changes in relative factor intensity, resulting from the development of foreign relative factor prices. (Chen and Góronicka, 2020) The second measure is based on the SVAR interpretation discussed, for example, by Mazzi et al. (2016), in which historical contributions of shocks are representing the total deviation from the SVAR trend, defined as potential output. The cyclical component representing the output gap is reconstructed by accumulating historical contributions of all shocks, which can be viewed as representing driving factors of the output gap.

Figure 1 shows the pattern of annual real GDP growth rates in Finland between 1975 and 2019. Positive rates of growth continued from the second half of the 1970s until the early 1990s depression. In the early 1990s depression economic activity decreased sharply, and the effects on the Finnish economy were long-lasting. After recovering from the depression, the Finnish economy started to grow rapidly. However, the growth slowed down due to the international recession in the 2000s but continued to grow shortly after. The severe second meltdown was triggered by the global financial crisis (GFC) in the second half of the 2000s. The growth recovered rapidly, but it was followed by several unfavorable years of negative growth until the second half of the 2010s, characterized by the Euro Crisis. In the second half of the 2010s, the growth started accelerating again.

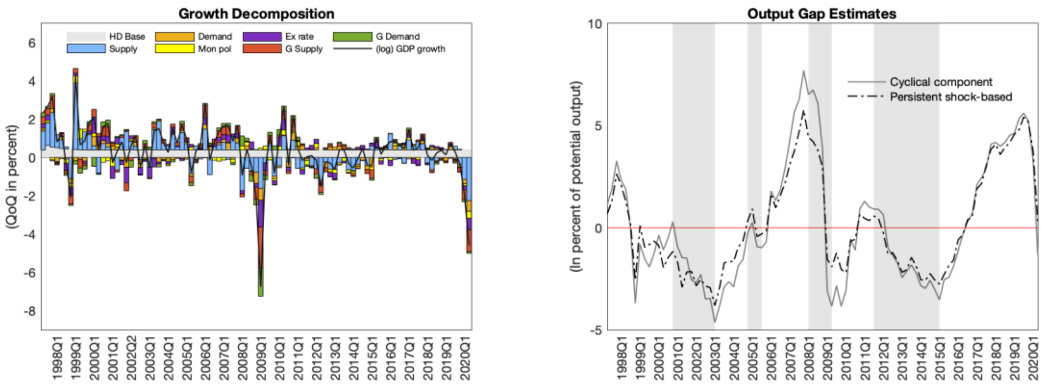
Figure 1. Annual Real GDP growth in Finland (1975-2019)

Source: Statistics Finland

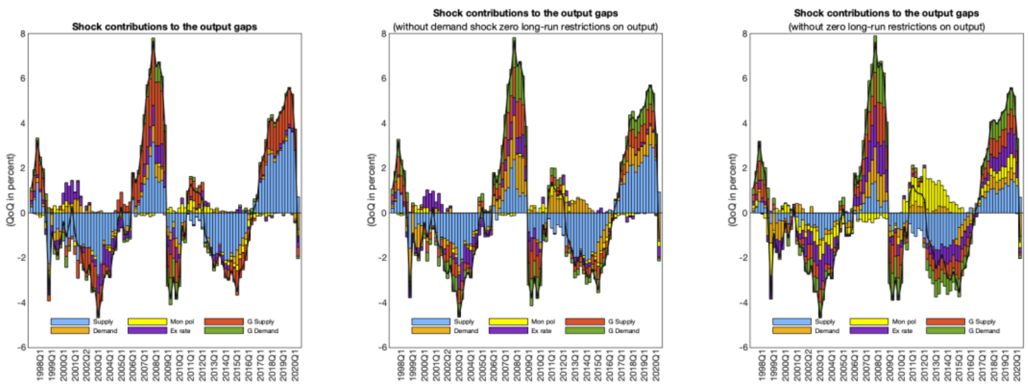
Figure 2 presents the historical decompositions of (log differenced) GDP growth for the period of the currency union, constructed using the small open economy SVAR. It suggests that persistent effects on output, defined as domestic and global supply shocks, have been driving sources of Finnish economic growth. The relative shares of persistent shocks are highlighted, particularly during the cycle peaks and troughs. Figure 2 also plots two series of the output gap estimates calculated by running regressions for (log) GDP growth. The persistent shock-based output gap estimates appear to differ from the cyclical component, particularly during the period of overheating prior to the GFC and in the trough after the sharp decline. The reference turning points shown in the figure have been collected from the Composite Leading Indicators (CLI) of the OECD. The cyclical peaks and troughs in the series of output gap estimates appear to be widely consistent with the CLI. The output gap estimates suggest that the Finnish economy experienced two periods of substantial overheating prior to the GFC and in the late 2010s prior to the global pandemic caused by the COVID-19. However, the latter overheating is unconvincing against the macroeconomic conditions, which will be discussed later. These periods were followed by sharp declines in growth.

From the late 1990s until the mid- 2000s, actual output remained mainly under the potential. The negative gap suggests that the production capacity of the Finnish economy was underused, and the growth was primarily achieved without generating inflationary pressures. In the second half 2000s, actual output sharply increased above potential prior to the sharp decline triggered by the GFC. However, the persistent shock-based output gap measure shows a limited decline, for example, in comparison to the mid-2010s. After the meltdown, actual output recovered briefly above potential at the beginning of the 2010s but fell back during the Euro Crisis. Again, the decline appears to be more severe than experienced in the GFC in the persistent shocks-based output gap measure, which one may find hard to prove. Both of the output gap measures close to the end of the sample period also suggest that the Finnish economy experienced considerable overheating in the second half of the 2010s the meltdown caused by the COVID-19, which are unconvincing, against the macroeconomic conditions. However, at the time of writing, turning points for this cycle were still unclear, and the estimates close to the end of the sample may be subject to revisions as the time series expands.

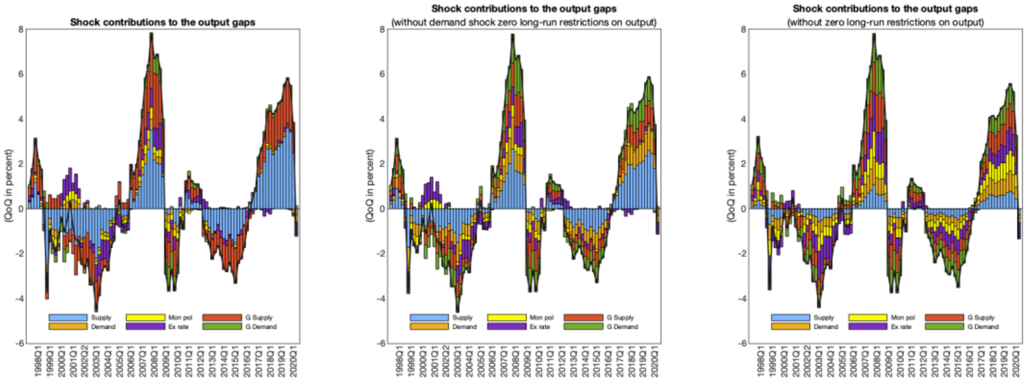
Figure 2. SOE SVAR: Quarterly Growth Decompositions and Output Gap Estimates for the Period of Currency Union



Shock contributions



Without sign restrictions



Source: Author's calculations and modifications following the example of Chen and Góronicka (2020).

Figure 2 also plots six separate combinations of zero and sign restrictions for the cumulative contributions of structural shocks to the growth path of (log) GDP, based on the small open economy SVAR. The first decomposition of shock contributions is based on the identifications scheme presented in Table 1. The decomposition shows how supply shocks and persistent global shock are the driving factors of the cyclical component. In the second decomposition, demand shocks are also allowed to have long-run effects on output. As a result, the weight of domestic supply and global shocks decreases, while contributions of domestic and global demand shock increase. In the third decomposition, all zero long-run restrictions are removed, allowing all shocks to have permanent effects on output increasing shock contributions of monetary policy and exogenous exchange rate shocks.

The estimations were also performed separately without any sign restrictions. Together with the long-run restrictions, the contributions of monetary policy and exogenous exchange rate shock increased, while the contributions of domestic demand shocks decreased, indicating the importance of sign restrictions for assigning temporary increases in output to domestic demand shocks. It also shows how the contributions of domestic supply shocks decrease, while persistent global shocks increase, particularly in the negative cycle in the mid-2010s, when not imposing any sign restrictions. Overall, when comparing the shock contributions, with and without sign restrictions, the shock decomposition appears significantly different, particularly around the mid-2010s, and even the sign of impulse responses. For instance, the decompositions also without long-run restrictions on output, changes signs of shock contributions on output resulting from demand and monetary policy shocks in the first half of the 2010s.

How the estimates behave in relation to the Finnish economic history and macroeconomic conditions during the period of a currency union? The Finnish economy grew rapidly from the mid-1990s until the 2000s. The headline quarterly year-on-year (YoY) GDP growth was around 4,8 percent on average. The annual inflation based on the headline consumer price index was on average 1,1 percent during that period. Unemployment was still high in the aftermath of the 1990s depression, around 12,9 percent. In the first half of the 2000s, growth started to slow down but reached an average rate of growth, around 3,2 percent on average. Inflation rose to 1,7 percent, and the unemployment rate declined to 9,2 percent on average. Inflation rose over 2 percent again in 2000 and 2001, reaching the level of 3,4 and 2,6 percent. In the latter half of the 2000s, which is characterized by both rapid growth and the global financial crisis, the growth rate declined to 1 percent on average per quarter. Unemployment continued to fall average to 7.5 percent during that period.

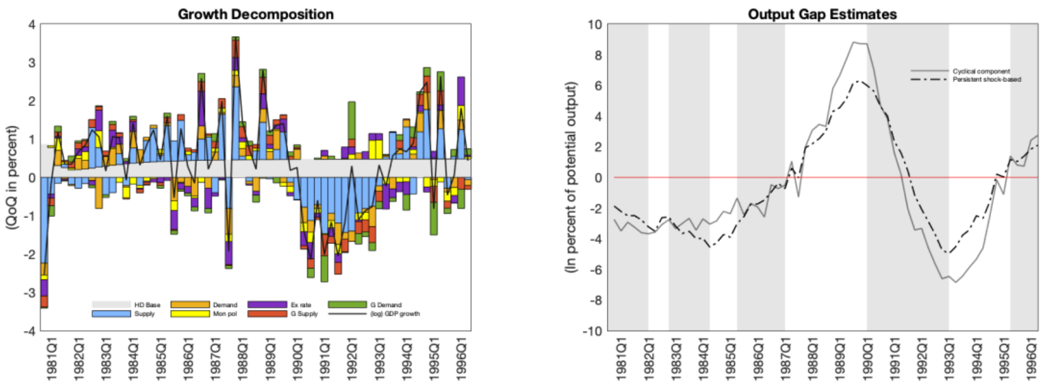
The quarterly (YoY) growth peaked over 6 percent prior to the GFC. Inflation accelerated over 2 percent again, reaching first the level of 2,5 and then 4,1 percent in 2007 and 2008, suggesting a positive output gap. At the same time, unemployment fell to the first 6,9 and then 6,4 percent. During the crisis, growth was almost minus 9 percent less than the corresponding quarter of the year before. The Finnish economy recovered shortly after, during 2010 and 2011. Inflation first reached 3,8 percent and then fell to 2,8 percent during those years, pointing towards a positive output gap. In the aftermath of the GFC, the Finnish economy also experienced a series of negative asymmetric shocks, including a collapse in the electronics and paper industry, which slowed down the recovery to the pre-crisis levels (Suni & Vihriälä, 2016).

The slow period of growth characterized by the Euro Crisis continued throughout the first half of the 2010s, quarterly (YoY) at the level of 0,7 percent on average. Unemployment, in turn, increased to 8,2 percent on average during the 2010s. The highest unemployment rate in the 2010s was experienced in 2015, around 9,4 percent. At the same time, inflation turned negative, to -0,2 percent, suggesting a negative output gap. An average growth rate accelerated again in the second half of the 2010s, reaching a rate of 1.7 percent, after which growth turned negative as a result of the COVID-19 pandemic in 2020. However, inflation remained persistently low at no more than 1.1 percent in the second half of 2010, suggesting a limited positive output gap at most, while the output gap measure based on the SOE SVAR decompositions suggests large positive output gaps which are unconvincing in this context.

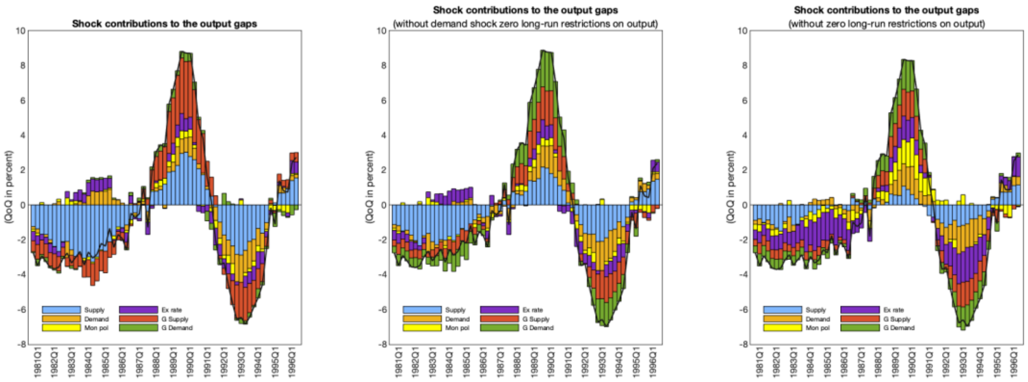
Figure 3 presents the historical decompositions of (log) GDP growth during the period of economic liberalization. It shows that most of the growth is driven by domestic supply shocks, particularly in the 1990s Finnish depression. The output gap estimates plotted in Figure 3 also appear to be broadly in line with the reference turning points of the CLI. The output gap measures suggest that the Finnish economy experienced substantial overheating in the latter half of the 1980s, which was followed by a sharp decline in the beginning 1990s. The persistent shock-based output gap estimates appear to differ from the cyclical component, particularly around the peak prior to the 1990s Finnish depression and in the trough after the sharp decline. For background information, the overheating of the economy was a consequence of several factors, including improved terms of trade, liberalization of financial markets, increased capital supply, excessive credit growth, and increased asset prices. Eventually, the early 1990s Finnish Depression was triggered by a series of negative shocks, including high interest rates, strong currency, banking crisis, and the collapse of the Soviet Union. (Kiander, 1996; Vihriälä, 1997; Honkapohja & Koskela, 1999; Jonung et al., 2009; Gorodnichenko, 2012; Gulan et al., 2014)

The pattern of the output gap estimates shown in Figure 3 shows how actual output remained below potential in the first half of the 1980s. It suggests a favorable period of growth without price pressures. However, in the late 1980s, actual output increased above potential, and the Finnish economy overheated prior to the upcoming 1990s Depression, which was followed by a downfall, shifting actual output back below potential output. In the second half of the 1990s, actual output climbed again over potential, pointing towards increasing price pressures.

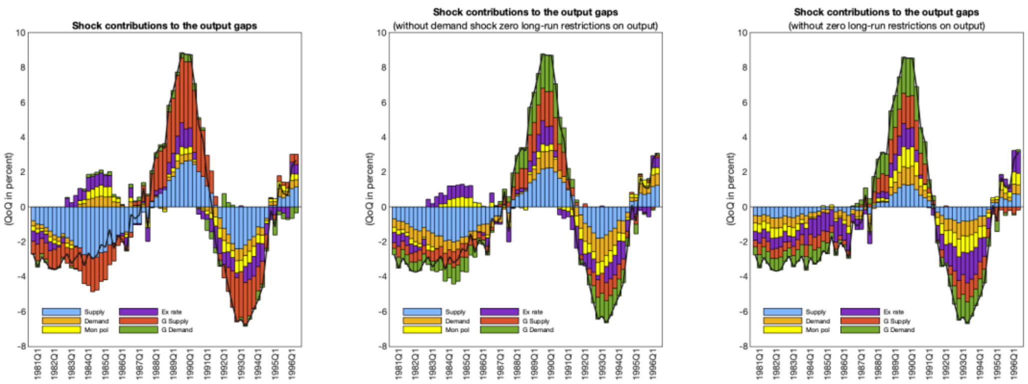
Figure 3. SOE SVAR: Quarterly Growth Decompositions and Output Gap Estimates for the Period of Economic Liberalization



Shock contributions



Without sign restrictions



Source: Author's calculations and modifications following the example of Chen and Góronicka (2020).

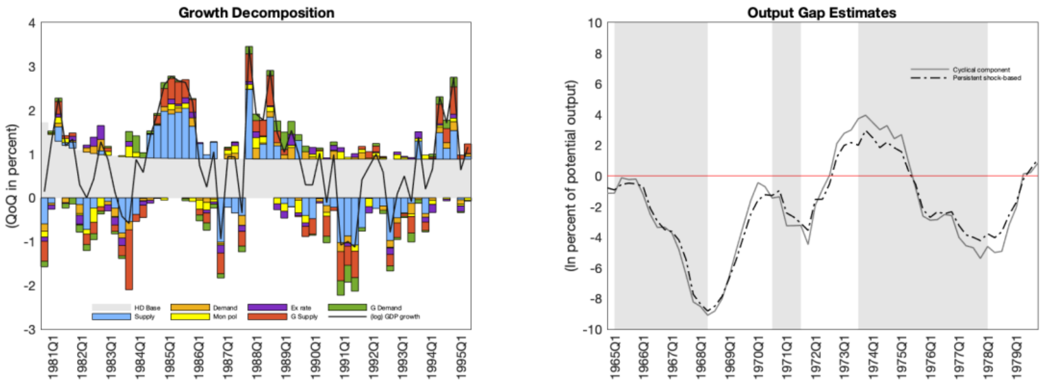
Figure 3 also plots six decompositions of shock contributions for the period of economic liberalization, using alterative combinations of zero and sign restrictions. Similarly, domestic supply shocks and persistent global shocks are the driving factors of the cyclical component in the first specification. When allowing demand shocks to have long-run effects on output, the contributions of global and domestic demand shocks increase, particularly during the period of overheating in the late 1980s and during the 1990s Finnish depression. Without any long-run restrictions, the contributions of monetary policy and exchange rate shocks increase, while the relative contributions of domestic supply shocks decrease most. If not imposing any sign restrictions, contributions of monetary policy, exogenous exchange rate, and global supply shocks increase, while contributions of domestic supply and demand shocks continue to decrease. Otherwise, when comparing the shock contributions with and without sign restrictions, shock contributions appear to show somewhat similar decompositions.

How were macroeconomic conditions during the period of economic liberalization? The Finnish economy grew rapidly from in the late mid-1980s, over five percentage points during several consecutive quarters based on year-on-year (YoY) quarterly changes of the historical quarterly GDP data. Correspondingly, during the 1990s depression, the quarterly growth declines were minus five percentage points for several subsequent quarters. The annual inflation increased from over 3 percent to over 6 percent during the second half of the 1980s based on the headline consumer price index, suggesting a large positive output gap. However, inflation fell shortly after the crisis was triggered in the 1990s. At the beginning of the decade, unemployment was slightly over three percentage points. As a result of the crisis, unemployment rose to 16.6 percent by 1994 and remained above 10 percent until the end of the 1990s.

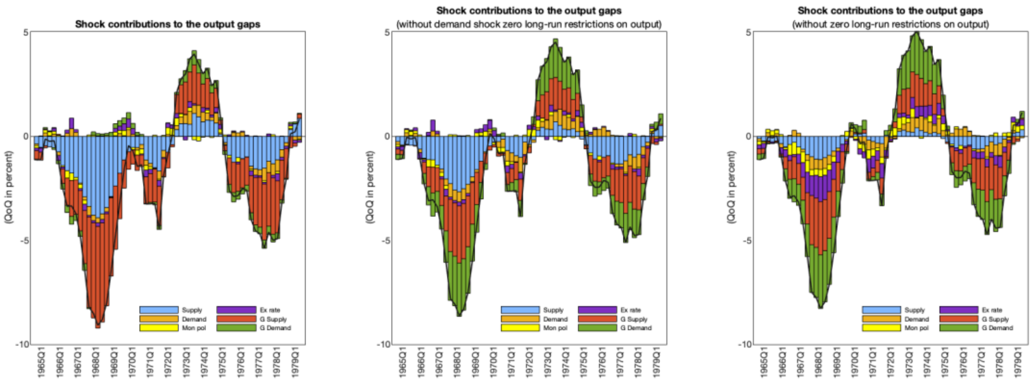
Figure 4 shows the historical decompositions of GDP growth and the output gap measurements during the period of economic regulation. For this period, the alternative dataset is used (i.e., import and export prices are derived from PPI). It shows that peaks and troughs are characterized by domestic supply-side shocks but also by persistent global shocks, particularly during the expansion and contraction phases. From a historical point of view, the growth of the Finnish economy slowed down between 1966 and 1968 due to domestic cost inflation and the international recession. The devaluation of the Finnish markka was carried out in 1967. The stabilization of the Finnish economy in the late 1960s was followed by a faster period of growth which ended up in a brief recession in 1971 because of tight economic policies and a strike in the metal industry. In the following years, the Finnish economy started to grow rapidly, fueled by an international upturn, which was also followed by high inflation. The period of rapid growth continued until the 1970s oil crisis. The severity of the recession was mitigated by increased exports to the Soviet Union. The period of slow growth continued until the mid-1980s. (Hjerppe, 1989.)

The output gap measures in Figure 4 suggest that actual output remained below potential from the mid-1960s until the first half of the 1970s. Strikingly large negative output gaps can be observed in the late 1960s, close to nine percentage points. However, in the first half of the 1970s, actual output rose above potential, suggesting increased inflationary pressures. A few years later, actual output dropped back below potential, and price pressures eased during the 1970s oil crisis. Actual output remained below potential until the late 1970s. The cyclical component and persistent shock-based output gap estimates are broadly in line. However, slight differences can be observed around the peaks and troughs in the 1970s.

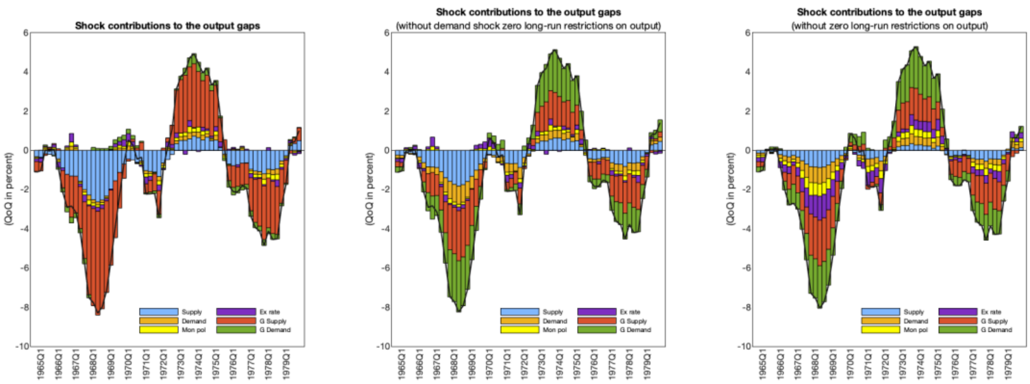
Figure 4. SOE SVAR: Quarterly Growth Decompositions and Output Gap Estimates for the Period of Economic Regulation



Shock contributions



Without sign restrictions



Source: Author's calculations and modifications following the example of Chen and Góronicka (2020).

In Figure 4, the decompositions of shock contributions respond correspondingly to the long-run restrictions in the period of economic regulation. The latter two specifications with loosened long-run restrictions appear to highlight global demand and supply shocks as an important contributor to cyclical development. Again, if not imposing any sign restrictions, contributions of monetary policy, exogenous exchange rate, and global supply shocks increase, while contributions of domestic supply and demand shocks decrease. When comparing the shock contributions, with and without sign restrictions, the decompositions appear to be rather similar.

Figures 11, 12, and 13 that can be found in the Appendix show impulse responses for the shocks presented in (log) GDP growth decompositions, constructed using the specification presented in Table 1. The impulse responses for the period of regulation and liberalization show how domestic supply shocks cause a permanent increase in output, while at the same time, prices move in the opposite direction. Similarly, impulse responses of global supply and demand shocks are showing comparable properties. Furthermore, both a loosening monetary policy and an appreciation of the nominal effective exchange rate lead to a temporary fall in output. However, when comparing the impulse responses to the period of the currency union, the sign of the impulse response of domestic supply shock on GDP and CPI in the period of currency, shown in Figure 11, differs from the period of liberalization and regulation. Fry and Pagan (2007) have shown that the inability to indentify the sign of the impulse response may be associated with the loose information embodied in the restrictions. One of the explanations that may cause the uncertainty is low and persistent inflation, which is characteristic in the 2010s.

The growth decompositions for each subsample period, based on the identification scheme presented in Table 1, build a view in which supply-side shocks drive growth which is in line with the results of Chen & Góronicka (2020). The relative shares of the other shocks were lower throughout the estimation periods. This in line with the idea that increased production capacity and technological progress maintain growth in the long-run. The relative contributions of the shocks differ to some degree between the subsample periods, which can be observed through the decompositions and impulse responses that can be found in the annex. Some of these differences may be possible to explain against different economic conditions and structural changes in the Finnish economy.

The relative contributions of shocks are highly dependent on the selected long-run restrictions, while the effects on the SVAR trend and the total amount of shock contributions appear to rather below. If only the supply and persistent global shocks are allowed to have long-run effects on output, most of the contributions of shocks are accumulated to these shocks. This appears to be central for the potential output that is reconstructed from the domestic supply and persistent global shocks. However, loosening the zero long-run restrictions may allow more intuitive decompositions of the cyclical component, defined as the output gap. Allowing domestic demand and global demand shocks to have long-run effects on output, the contributions of demand shocks increase, which can be interpreted as hysteresis, particularly around the cycle peaks and troughs. This would be in line with the findings of Blanchard (2015) regarding demand-triggered recessions. If not imposing any zero long-run restrictions on output, the contributions of monetary policy shocks and exogenous exchange rate shocks increase, which one may find much more difficult to explain.

The sign restriction appears not to be critical for the output gap estimates as the long-run restrictions already appear to assign most of the shock contributions to domestic and global supply shocks. However, when not imposing sign restrictions, the relative contributions of domestic supply shocks decrease, and contributions of global supply shocks decrease. The sign restrictions also appear to be important for assigning temporary increases in output to domestic demand disturbances, for example, instead of monetary policy and exogenous exchange rate shocks. However, the contributions of these shocks were rather low with the long-run restrictions. Interestingly, when comparing the shock contributions, with and without sign restrictions, the changes were rather moderate in the period of liberalization and regulation relative to the period of the currency union, during which persistency of low inflation and interest rates have been characteristic. For instance, the decompositions also without long-run restrictions on output change the signs of the shock contributions on output resulting from demand and monetary policy shocks in the 2010s. These findings suggest that the sign restrictions imposing a positive relationship between CPI and GDP, associated with demand shocks, contain only a little information. Similarly, the sign restriction that sets a link between monetary policy shocks and a negative relationship between GDP and CPI appears not to be strong.

The output gap measures reconstructed from the SOE SVAR decompositions build a view that is somewhat consistent with the Finnish economic history and macroeconomic conditions. Actual output appears to be above or under potential, mostly when it is supported by the history and macroeconomic indicators. However, some inconsistencies can be observed between the macroeconomic indicators and the output gap estimates. For instance, the output gap measures close to the end of the sample period points towards considerable overheating in the second half of the 2010s prior to the meltdown caused by the COVID-19, which are unconvincing against the macroeconomic conditions. Also, the size of the drop in the GFC appears to be low to the recession experienced in the first half of the 2000s and during the Euro Crisis, particularly in the persistent shock-based output gap estimates.

The inconsistencies between the output gap measures and macroeconomic conditions appear sourced from several factors depending on the point of view. When looking at the persistent shock-based output gap estimates, most of the shock contributions are assigned to domestic and global supply shocks. Therefore, the output gap estimates appear to be associated with the ability to distinguish long-run changes in output, defined as the trend component of the SVAR, from changes associated with the cyclical component. For example, the alternative specifications of the identifying restrictions had only minor effects on the trend component and on the total amount of shock contributions. However, one may also find it troubling that most of the shocks are indeed assigned to domestic supply or persistent global shocks.

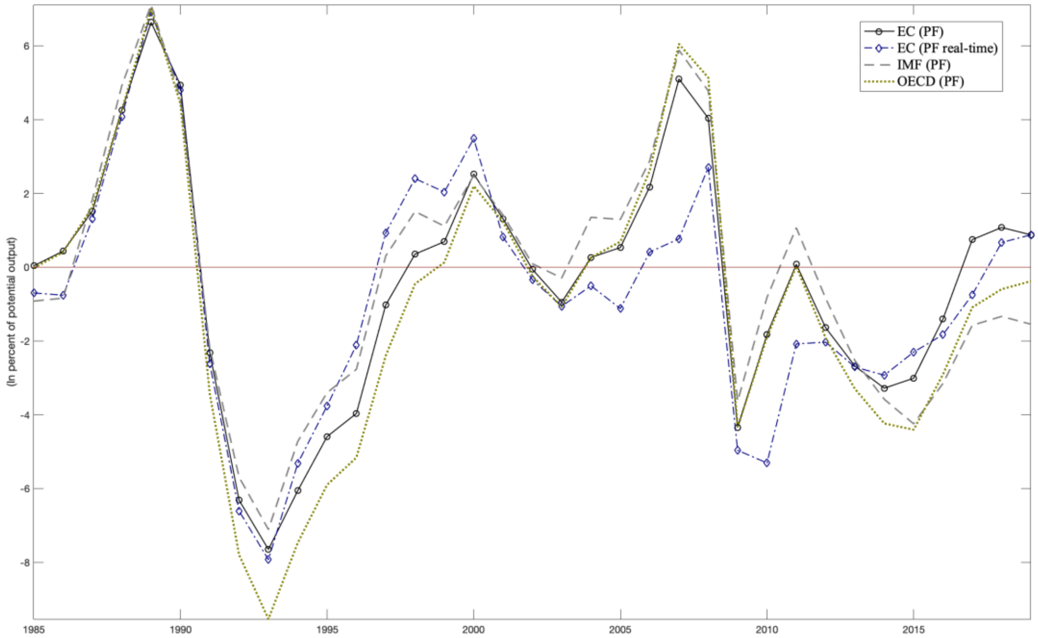
The persistent shock-based output gap measures reconstruct rather similarly sized output gap estimates that the cyclical components of the SVAR. The identification scheme presented in Table 1 assigns most of the shock contributions to domestic supply and persistent global shocks, while the long-run changes associated with the trend component of the SVAR remain almost unchanged. If not imposing zero-long-run or sign restrictions, only minor differences can be observed between the SVAR trends and the total amount of shock contributions. Therefore, the long-run restrictions appear to be the most critical restrictions for the output gap estimates reconstructed from the accumulated domestic supply and persistent global shock. Furthermore, the sign restrictions may contain little information during times, for example, characterized by low and persistent inflation and interest rates. These findings also suggest that the critical arguments of Blanchard (2018) regarding inaccurately identified supply and demand shocks in the case of the BQ method appear to be relevant when applying this method. These methods are not designed to identify supply shocks without permanent effects on output or demand shocks with long-lasting effects on output, which one may find problematic. Furthermore, the utilized variables, such as inflation and interest rate, together with the imposed restrictions, may embody only a little information about the shock under inspection, particularly if the variables are persistent.

6. Comparing Performance of Past and Present Output Gap Estimates

Next, three output gap measures are constructed using full sample and real-time data. The size of the output gap estimates is compared between samples to examine the performance of the output measures, primarily in real-time. The estimates are compared in two parts, in which the previously used small open economy SVAR method is compared against the BQ method and the univariate HP filter. At first, special attention is paid to the size of estimates based on full sample data, particularly around economic turning points and at the end of the sample. Secondly, the reliability of the estimates is examined using real-time data following Orphanides and van Norden (2002). More specifically, series of output gap estimates are constructed performing estimations using the data vintage available at each point in time. These estimates are examined against the full sample estimates, produced using the last vintage of the total estimation period. Finally, measurement errors in real-time estimates are constructed using a difference between real-time and full sample estimates.

At first, for illustrative purposes, Figure 5 plots the annual output gap estimates for Finland produced by the European Commission, IMF, and OECD between 1985 and 2019. The estimates have been produced using the production function (PF) approach using annual data. These estimates are based on actual data, and forecasts for ongoing or future years have not been used. The annual real-time estimates of the European Commission (EC) are based on Spring vintages from 2005 to 2020. Overall, the estimates of international organizations are close to each other, forming a similar picture about the cyclical position of the Finnish economy. However, we can also observe that the annual real-time estimates of the EC differ most, illustrating the uncertainty associated with annual real-time estimates. The uncertainty appears to be present, specifically around economic turning points. For instance, the annual real-time estimates of the EC suggest lower positive output gaps around the peak prior to the global financial crisis (GFC). Furthermore, the estimates of the EC at the end of the sample have a different sign compared to the estimates produced by the IMF and the OECD.

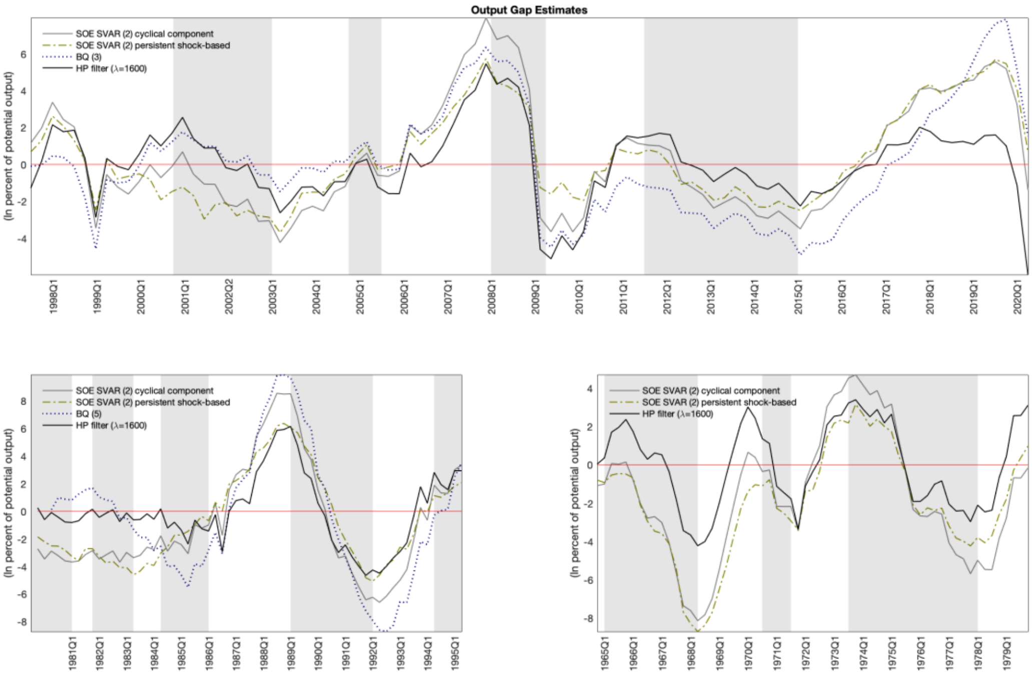
Figure 5. Annual Output Gap Estimates of the International Organizations



Source: European Commission (EC), IMF, and author.

Figure 6 plots the output gap measures based on the small open economy (SOE) SVAR, the BQ method, and the univariate HP filter. The output gap measures have been constructed separately for each subsample period. The output gap measures form rather similar patterns of cyclical development in all subsample periods. Thus, the size of the measures varies, particularly in economic turning points. At the end of the last sample period, the difference is highlighted, being several percentage points between the output gap measures. The structural VAR-based measures show large positive output gaps in the late 2010s, suggesting overheating of the economy and high inflationary pressures, whereas the output gap measure based on the cyclical component of the HP filter shows much lower positive output gap rates. The average annual estimate of the international organizations, shown in table 2, is -0,4 percent for 2019. Also, inflation has remained low in the late 2010s, reaching the level of 1,1 percent at maximum. In this context, the structural VAR-based output gap measures are unconvincing. However, the estimates close to the end of the sample may revise when the time series expands.

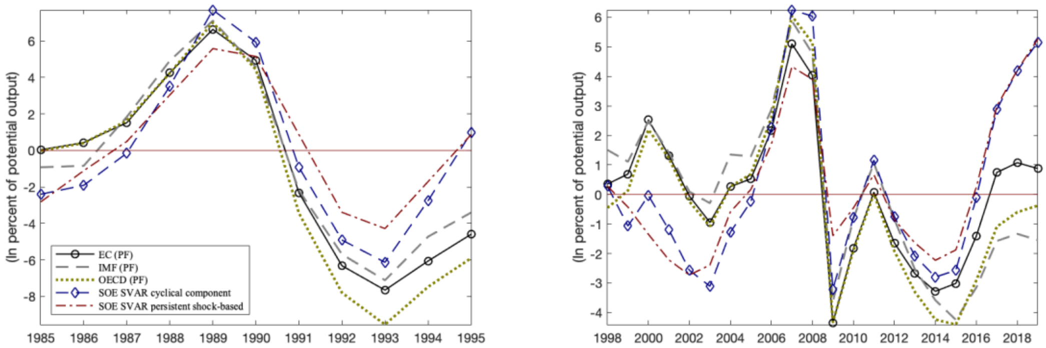
Figure 6. Comparison of the Quarterly Output Gap Estimates



Source: Authors' calculations

Figure 7 shows a comparison of the annual averages of the output gap estimates based on the small open economy (SOE) SVAR to the annual estimates of the international organizations. The average estimate of the international organizations for the peak prior to the global financial crisis (GFC) is about 5,7 percent in 2007. The average of quarterly estimates based on persistent shocks of the SOE SVAR meets the level precisely, and the average of HP filter estimates is quite close, at a level of 5,5 percent, shown in table 2. The estimates based on the cyclical component of the SVAR are at the level of 8 percent, producing the highest estimate, whereas the BG estimates point towards a positive gap of 6,4 percent. Interestingly, in the trough of the GFC, the persistent shocks-based output gaps of the SOE SVAR decomposition suggests a limited decline and lower negative output gaps in absolute value relative to the other output gap measures. In 2009 the average estimates of the international organizations were -4,1 percent, whereas the persistent shocks-based output gaps show a negative output gap of -1,8 percent, on average. The other quarterly output gap measures are within a radius of 0,5 percentage points from the average of international organizations. The deviation between estimates can also be observed, particularly around the international recession and the Euro crisis in the first half of the 2000s and the 2010s. The output gaps based on the BG method remain negative around the brief recovery in 2011, while the SOE SVAR and the HP filter point towards positive output gaps.

Figure 7. Comparison with the Annual Estimates of International Organizations



Source: Authors' calculations

There is also a large deviation between estimates in the late 1980s when the Finnish economy was experiencing serious overheating. The average estimates of the international organizations suggest a positive gap of 6,9 percent. The cyclical component of the SOE is at the level of 7,7 percent, being closest to the average of international organizations. The persistent shock-based measure indicates a lower positive output gap, 5,6 percent. The HP filter suggests the lowest positive gap of 5 percent, and the BG method the highest, 8,8 percent. Similarly, a large deviation between the size of the measures can be observed during the Finnish 1990s depression. The average of the estimates of the international organizations is -8,1 percent. The BQ method points towards a negative gap of -8,4 percent, being the closest to the average estimate of international organizations. The HP filter points towards a most limited decline -4 percent.

Table 2. Comparison of the Annual Estimates

Year	The average of annual estimates of EC, IMF, and OECD	SOE SVAR (2) Cyclical Component	SOE SVAR (2) Persistent Shock-based	BQ (3)	HP filter (λ=1600)
1989	6,9	7,7	5,6	8,8	5,0
1993	-8,1	-6,1	-4,3	-8,4	-4,0
2003	-0,8	-3,1	-2,4	-0,7	-1,8
2007	5,7	8,0	5,7	6,4	5,5
2009	-4,1	-3,7	-1,8	-4,3	-4,6
2011	0,4	1,0	0,6	-1,3	1,7
2015	-3,9	-1,9	-1,4	-4,0	-1,3
2019	-0,4	5,2	5,5	7,9	1,0

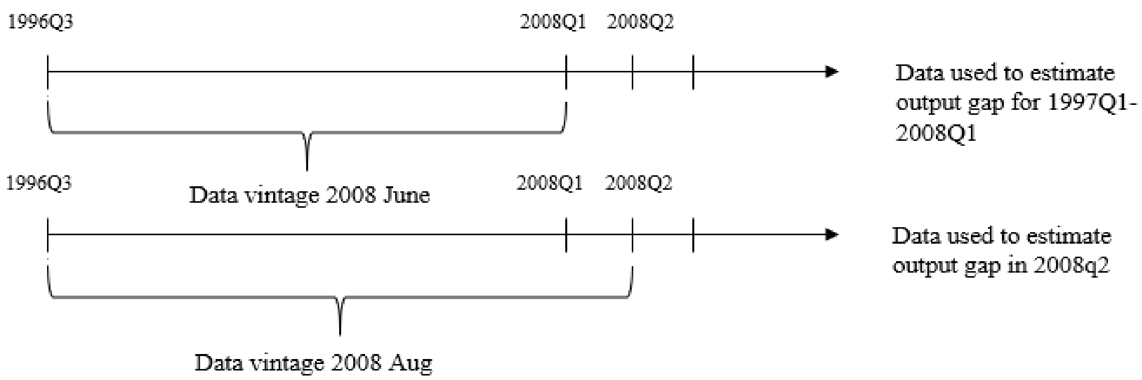
Note: Annual averages of the output gap estimates in severe cycle peaks and troughs.

Source: European Commission (EC), IMF, and author.

Overall, the output gap measures constructed with the structural VAR decompositions and the HP filter using full sample data show how all the applied methods form rather similar views about the cyclical development in Finland. However, the size of the measures differs, particularly around economic turning points. Sometimes differences between the measures may be even several percentage points, and even the sign of the estimate may differ when actual output is close to potential. For instance, the output gap measures based on structural VARs close to the end of the sample also suggest large positive output gaps for the late 2010s, suggesting high inflationary pressures. In contrast, the output gap measure based on the cyclical component of the HP filter show much lower and more credible positive gap rates in comparison to the estimates of international organization and macroeconomic indicators. However, the estimates at the end of the sample period may be revised when the time series expands.

6.1. Revisions in output gap measures

The real-time reliability of output gap estimates can be evaluated using revisions in output gap estimates. The revisions in the output gap estimates can be formed using real-time data, following Orphanides and van Norden (2002). Next, the real-time estimates are constructed, performing estimations using the data vintage available at each point in time starting from 2008Q1 until 2020Q2[2]. Separate regressions are performed for each quarter, and only the real GDP series is replaced using the first vintage published for each period. Real GDP vintages have been collected from archives of Statistics Finland. The first real-time output gap estimate starts from 2008Q1. The estimates before 2008Q1 are based on data released in June 2008.



Source: Author's calculations following the example of Chen and Góronicka (2020).

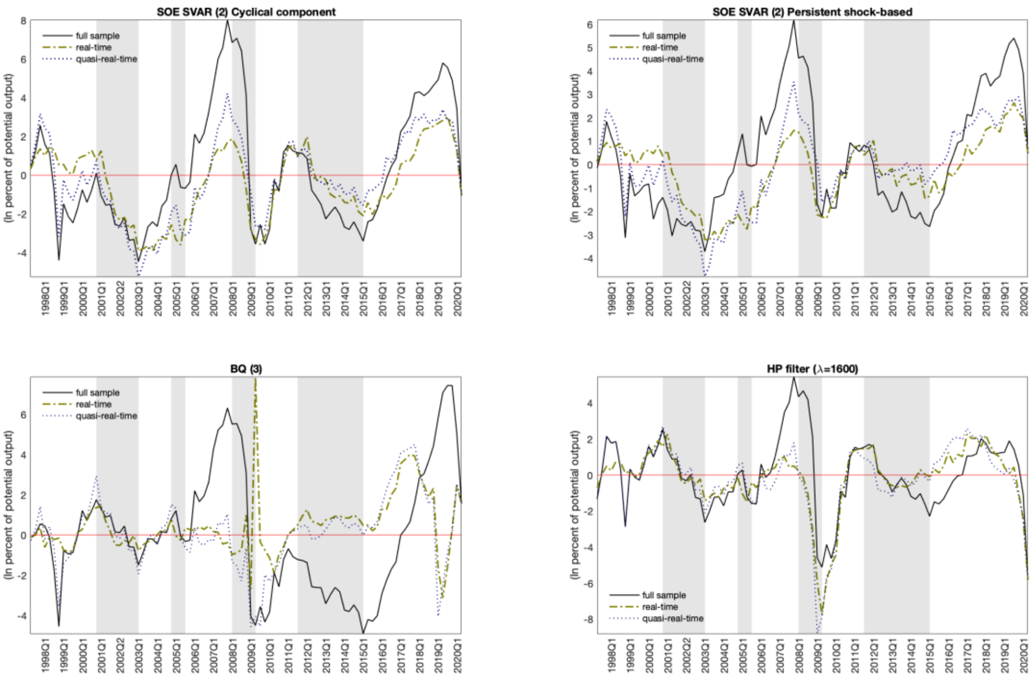
In this section, the small open economy SOE SVAR is estimated by using the alternative dataset (i.e., import and export prices are derived from PPI) as it facilitates regressions. The output gap estimates constructed using the alternative dataset follow quite closely to the estimates constructed using the original dataset. However, some differences can be observed between the size of the estimates. The comparison of the output estimates constructed using the original and the alternative dataset can be found in the annex. Furthermore, the alternative dataset also changes impulse responses and growth decompositions to some extent, which should bear in mind, when applying the method. Series of quasi-real-time estimates are formed by running models separately at each point in time, using the full sample data. The difference between real-time and quasi-real-time estimates illustrates revisions in output gap estimates associated with data revisions. The difference between quasi-real-time and full sample estimates embodies revisions in output gap measures associated with the model performance. (Orphanides and van Norden, 2002.)

2 The compilation of data vintages from 2008Q1 to 2013Q4 is based on the European system of national and regional accounts (ESA 1995). The vintages from 2014Q1 onwards have been compiled according to the ESA 2010. The review did not have a considerable effect on the quarterly cyclical or seasonal variations of gross domestic product (see Official Statistics of Finland (OSF) (2014b).

Figure 8 plots the real-time and quasi-real-time estimates against the output gap estimates based on full sample data. Both the real-time estimates differ from the full sample estimates, particularly around the period of overheating prior to the global financial crisis (GFC). Most of the time, real-time and quasi-real-time estimates appear be broadly consistent with each other by each method. None of the four output gap measures capture large positive output gaps prior to the global financial crisis (GFC) in real-time. Both of the output gap measures based on the small open economy SVAR show very similar development of the real-time estimates. The quasi-real-time estimates of these measures appear to capture the upswing prior to the GFC to some extent, suggesting that poor performance of these methods before the GFC is associated with both data revisions and model performance. After the crisis burst, the real-time estimates of the small open economy SVAR and the HP filter follow the full sample estimates closely during the meltdown and recovery phase. However, the size of the negative gaps between the full sample and the real-time estimates differ, especially in the case of the HP filter.

Overall, the SOE SVAR appear to react to the ongoing business cycle with a delay as the full sample estimates, and the real-time estimates appear to differ, particularly around economic turning points and during gradual reversals of the cycles. The real-time performance of the BQ method appears to be the lowest as the real-time estimates appear to differ most relative to the full sample estimates. The real-time estimates of the HP filter appear to follow more closely to the full sample estimates than the structural VAR-based measures.

Figure 8. Real-time Output Gap Estimates



Source: Authors' calculations.

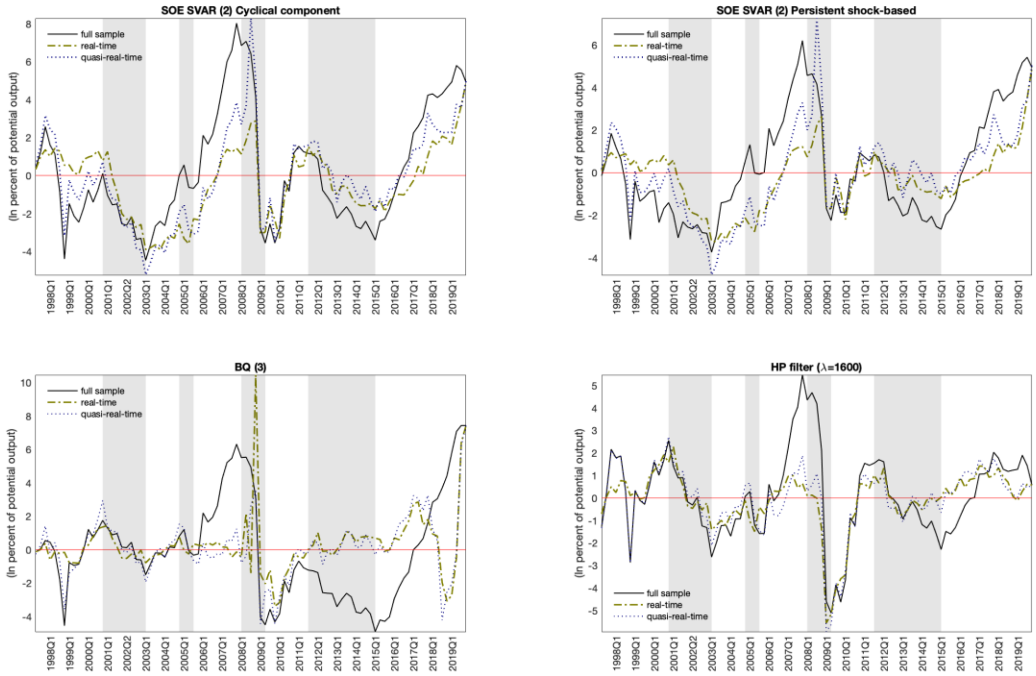
Notes: Figure 7 presents output gap estimates for Finland constructed using real-time and full sample data (1997Q2-2008Q1 =>, and 1997Q2–2020Q2).

Similar growth paths of quasi-real-time and real-time estimates indicate that the output gap revisions are primarily associated with the model performance at the end of the sample. Hence, all the models appear to adjust the estimates to some extent at the end of the sample period as the time series expands. In order to verify whether the estimates adjust over time, the real-time exercise is performed by removing a few of the produced output gap estimates from the end of the samples. The estimates are constructed similarly as before, but a few of the estimates are removed from the end of the samples, and only the last remaining estimate is used. The data vintage published in June 2008 is used to construct a series of the real-time output gaps until 2007Q3. The estimates constructed by using the data vintage published in 2008 August are utilized until 2007Q4, and so on. These estimates can be considered being constructed with "perfect forecasts," i.e., the original published real-time data vintage is used in the estimations, while only the output gap estimates of t-2, t-4, or t-8 of are used for each quarter. In this thesis, these estimates are defined as delayed real-time estimates.

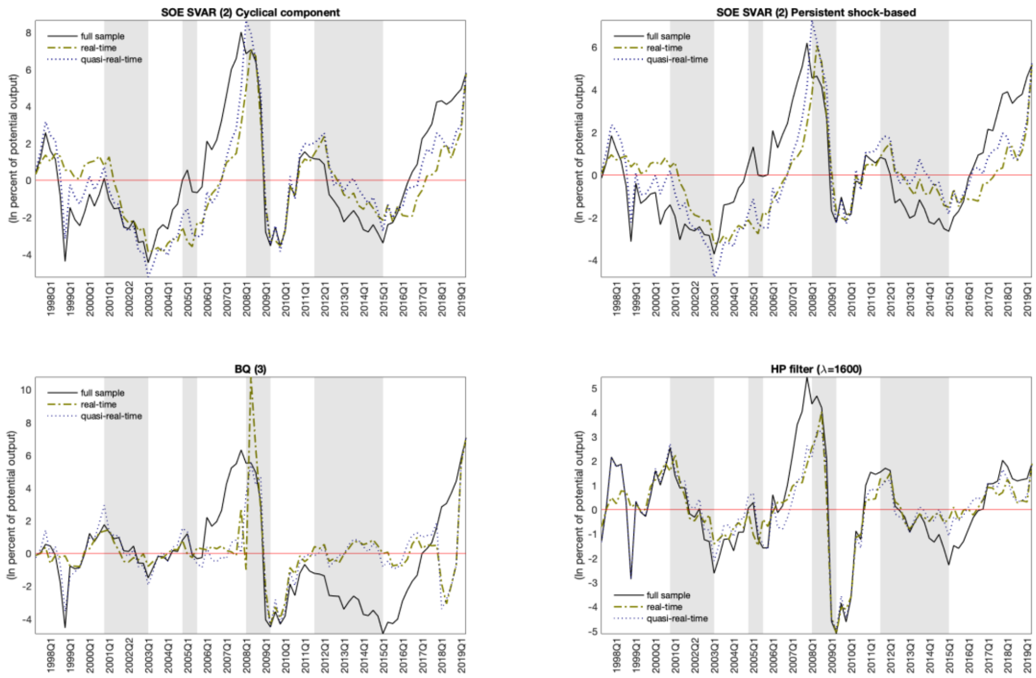
Figure 9 plots the delayed real-time output gap estimates showing that the estimates improve after a few quarters from the first publication. Series of the output gap estimates based on SVAR methods show sharp peaks prior to the global financial crisis with "the extended samples" after the final two quarters have been removed from the end of the real-time estimates. After the removal of four quarters, the measures based on small open economy SVAR decompositions and the HP filter start to follow the estimates based on full sample data more closely, indicating that these estimates indeed improve over time. The estimates based on the BQ method also show gradual improvement. Finally, the delayed real-time estimates that are based on t-8 are showing very similar development to the estimates based on full sample data. However, the output gap measures based on the SOE SVAR still appear to react to the upturn phase prior to the GFC with a delay. The delayed real-time and full sample estimates are also still showing differences during the unfavorable years in the 2010s. Again, the HP filter appears to be the most consistent.

Figure 9. Delayed Real-time Output Gap Estimates

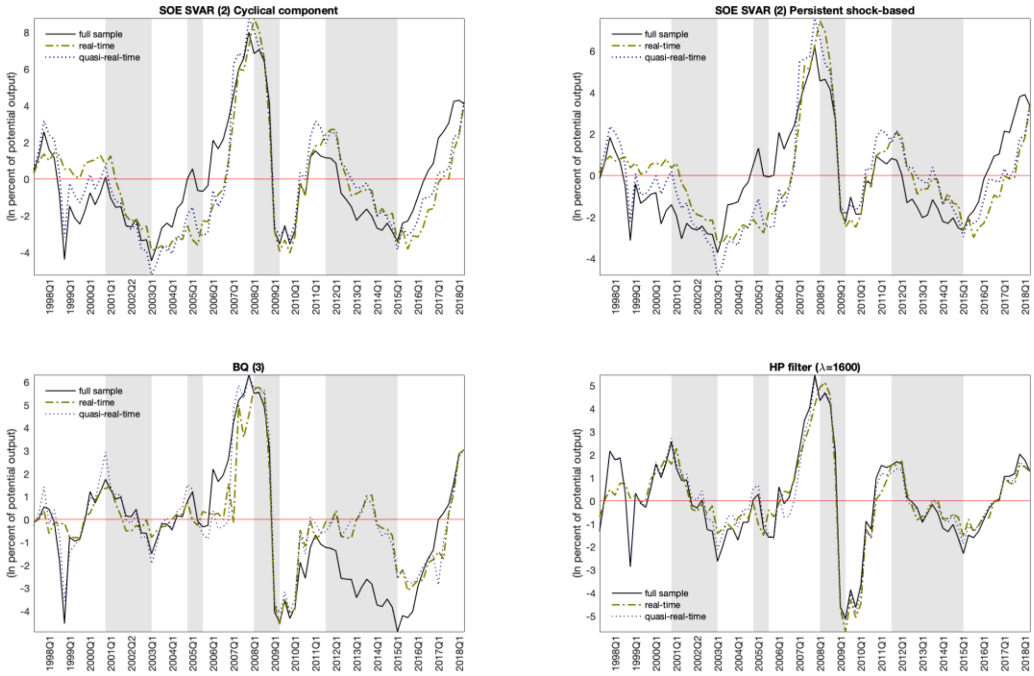
t-2 real-time estimates



t-4 real-time estimates



t-8 real-time estimates



Source: Authors' calculations.

Notes: Figure 8 presents output gap estimates for Finland constructed using real-time and full sample data (1997Q2–2019Q4, 1997Q2–2019Q2, 1997Q2–2018Q2).

The consistency of the estimates improves over time in all the applied methods. The end of the sample is a generally acknowledged problem for many filtering techniques, including the HP filter. However, these results also suggest that structural VAR methods have difficulties identifying whether the change in output will be permanent in real-time, particularly around economic turning points. The delayed real-time estimates show that the estimates in all the applied methods improve remarkably after a few quarters have passed from the first publication of the GDP data. However, GDP data revisions do not appear to be the primary source of revisions in the output gap estimates. Although, exceptionally large revisions in actual output may greatly affect the estimates, as observed in the output gap estimates based on the small open economy SVAR around the global financial crisis.

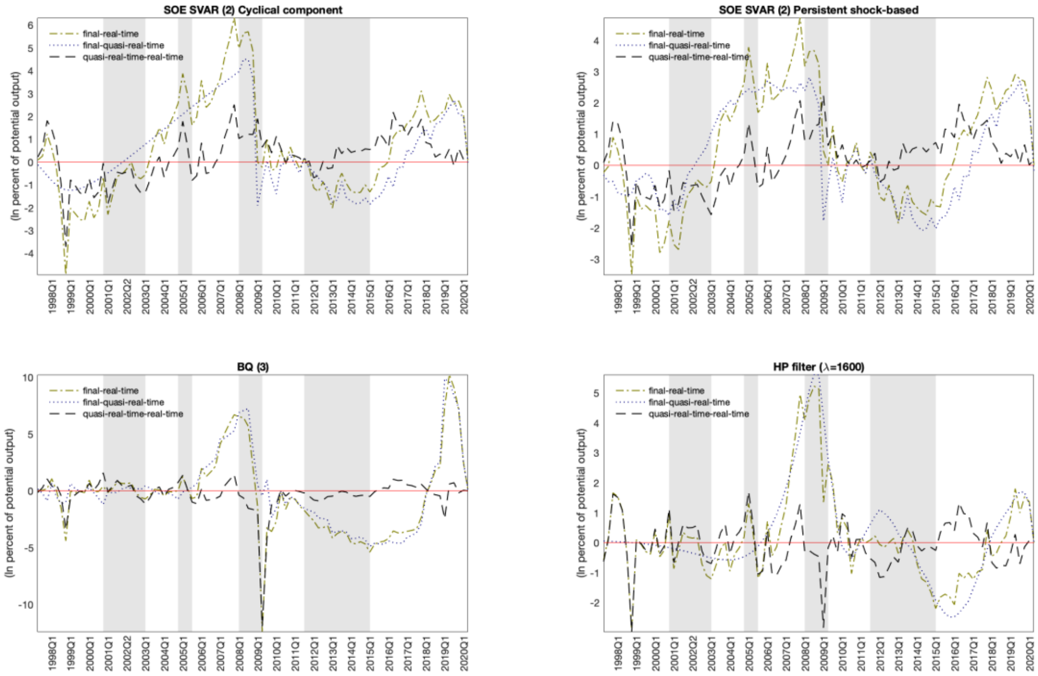
Overall, the HP filter appears to produce the most consistent output gap estimates. Both of the output gap measures based on the SOE SVAR show very similar development of the output gap estimates suggesting that real-time performance in these methods is primarily associated with the ability to single out the long-run changes, defined as trend component, and the changes associated with the cyclical component. The real-time performance of the BQ method appears to be the lowest. One possible way to mitigate bias in real-time estimates is to extend the time series of utilized variables with well-formed forecasts, as the delayed estimates suggest. However, the delayed real-time estimates give an over-optimistic view because the uncertainty typically associated with the actual forecast will eventually transmit to the estimates of the output gap.

6.2. Source and size of revisions

The output gap estimates produced in the previous section allow to further investigate revisions in the output gap estimates, as shown in Orphanides and Norden (2002). The changes in the estimates can be decomposed into data relating and model associated revisions, using the differences between real-time and full sample estimates. The model-associated revisions can be separated through a difference between final and quasi-real-time estimates, while data revisions associated with changes in output gap estimates can be obtained, deducting real-time estimates from quasi-real-time estimates.

In Figure 10, the difference between final and real-time estimates illustrates the total revision in output gap estimates. The model-associated output gap revisions appear to follow quite closely to total revisions in all the applied methods, confirming the view that the revisions in the output gap estimates are primarily associated with the model performance. However, the revisions around the global financial crisis appear to be associated with both data revisions and model performance, particularly in the case of the small open economy SVAR. The largest revisions in the output gap estimates take place around the global financial crisis, but large revisions can also be observed around the economic turning point in the mid-2010s, suggesting that revisions tend to increase specifically around economic turning points.

Figure 10. Revisions in Output Gap Estimates



Source: Sources: Authors' calculations.

Notes: Figure 9 presents revisions in output gap estimates constructed from output gap estimates based on real-time and full sample data (1997Q2–2020Q2).

Table 3 presents descriptive statistics for the revisions in the output gap estimates. The revisions in the output gap estimates indeed decrease over time, which be observed from the decreasing absolute values of the performance indicators. The HP filter appears to outperform structural VAR methods in real-time, having the lowest values in each performance indicator, including root mean squares (RMS) and the lowest revisions on average. This appears to be the case with the delayed real-time estimates. Based on the revisions, both of the output gap measures based on the SOE SVAR appear to be more reliable in real-time than the BQ method in the context of Finland. These measures outperform the BQ method in each performance indicator. The BQ method appears to have the highest revisions during the sample period.

Table 3. Output Gap Revision Statistics

Method	MEAN	SD	RMS	MIN	MAX	MAR
Real-time estimates						
SOE SVAR (2) Cyclical Component						
Final-real-time	0,66	2,15	2,25	-4,94	6,32	1,76
Final-quasi-real-time	0,43	1,71	1,76	-1,90	4,50	1,45
Quasi-real-time-real-time	0,23	1,01	1,04	-3,72	2,51	0,78
SOE SVAR (2) Persistent Shock-based						
Final-real-time	0,53	1,83	1,90	-3,50	4,73	1,58
Final-quasi-real-time	0,34	1,53	1,57	-2,09	2,81	1,36
Quasi-real-time-real-time	0,19	0,88	0,90	-2,60	2,24	0,70
BQ (3)						
Final-real-time	-0,40	3,51	3,53	-12,37	10,21	2,50
Final-quasi-real-time	-0,10	3,29	3,29	-4,89	9,85	2,29
Quasi-real-time-real-time	-0,30	1,50	1,53	-11,92	1,58	0,77
HP filter (λ=1600)						
Final-real-time	0,21	1,54	1,55	-2,97	5,23	1,05
Final-quasi-real-time	0,26	1,60	1,62	-2,49	5,61	1,10
Quasi-real-time-real-time	-0,05	0,78	0,78	-2,97	1,70	0,59
Delayed estimates t-2						
SOE SVAR (2) Cyclical Component						
Final-real-time	0,60	2,14	2,23	-4,94	6,58	1,76
Final-quasi-real-time	0,35	1,61	1,65	-1,86	4,17	1,37
Quasi-real-time-real-time	0,26	1,16	1,18	-3,72	5,53	0,83
SOE SVAR (2) Persistent Shock-based						
Final-real-time	0,46	1,84	1,90	-3,50	4,97	1,58
Final-quasi-real-time	0,26	1,55	1,57	-3,08	2,91	1,35
Quasi-real-time-real-time	0,19	1,00	1,02	-2,60	5,01	0,73
BQ (3)						
Final-real-time	-0,29	3,04	3,06	-7,36	8,53	2,16
Final-quasi-real-time	-0,17	2,84	2,84	-4,59	7,90	1,98
Quasi-real-time-real-time	-0,12	1,19	1,19	-6,95	4,04	0,72
HP filter (λ=1600)						
Final-real-time	0,20	1,39	1,41	-2,97	4,82	0,96
Final-quasi-real-time	0,15	1,15	1,16	-1,78	4,11	0,80
Quasi-real-time-real-time	0,05	0,66	0,67	-2,97	1,70	0,48

Delayed estimates t-4						
SOE SVAR (2) Cyclical Component						
Final-real-time	0,42	1,96	2,00	-4,94	5,14	1,61
Final-quasi-real-time	0,26	1,52	1,54	-1,81	3,84	1,29
Quasi-real-time-real-time	0,16	1,03	1,04	-3,72	3,75	0,77
SOE SVAR (2) Persistent Shock-based						
Final-real-time	0,29	1,78	1,81	-3,50	3,87	1,53
Final-quasi-real-time	0,18	1,52	1,53	-2,70	2,89	1,32
Quasi-real-time-real-time	0,11	0,91	0,92	-2,60	3,48	0,70
BQ (3)						
Final-real-time	-0,33	2,48	2,50	-5,24	6,51	1,74
Final-quasi-real-time	-0,27	2,34	2,35	-4,14	6,28	1,59
Quasi-real-time-real-time	-0,06	1,09	1,09	-5,27	4,56	0,67
HP filter (λ=1600)						
Final-real-time	0,09	0,99	0,99	-2,97	3,61	0,71
Final-quasi-real-time	0,07	0,79	0,79	-1,20	2,83	0,56
Quasi-real-time-real-time	0,02	0,65	0,65	-2,97	1,70	0,45
Delayed estimates t-8						
SOE SVAR (2) Cyclical Component						
Final-real-time	0,15	1,69	1,70	-4,94	3,87	1,37
Final-quasi-real-time	0,10	1,34	1,34	-1,75	3,16	1,13
Quasi-real-time-real-time	0,04	0,97	0,97	-3,72	3,00	0,74
SOE SVAR (2) Persistent Shock-based						
Final-real-time	0,03	1,63	1,63	-3,50	3,79	1,35
Final-quasi-real-time	0,04	1,40	1,40	-2,04	2,82	1,20
Quasi-real-time-real-time	-0,01	0,85	0,85	-2,60	2,79	0,67
BQ (3)						
Final-real-time	-0,38	1,49	1,54	-4,39	4,38	1,08
Final-quasi-real-time	-0,49	1,29	1,38	-3,53	2,85	0,93
Quasi-real-time-real-time	0,12	0,90	0,90	-3,40	4,99	0,54
HP filter (λ=1600)						
Final-real-time	0,00	0,70	0,70	-2,97	1,66	0,52
Final-quasi-real-time	-0,03	0,34	0,34	-0,58	0,98	0,27
Quasi-real-time-real-time	0,03	0,65	0,65	-2,97	1,70	0,46

* Mean absolute revision (MAR)

Source: Author's calculations following the example of Orphanides, A. & van Norden, S. (2002).

Overall, the reliability of the output estimates did not prove to be particularly good in real-time, particularly around economic turning points, which can be observed through the revisions in output gap estimates. In fact, the reliability of the real-time estimates produced with the univariate HP filter outperformed structural VAR methods in the case of Finland. For example, all the output gap measures based on the full sample data point towards greater positive gaps than the real-time estimates for the upturn phase before the global financial crisis (GFC), suggesting an overestimation of potential output in real-time. These findings regarding the HP filter are in line with the quasi-real-time estimates presented in Kotilainen (2019). Overall, the results are broadly in line with the previous literature in which the revisions in the output gap measures are found to increase around the economic turning points and at the end of the sample, for example, as already presented in Orphanides and van Norden (2002) and Marcellino and Musso (2011). However, these findings are not based on the structural VAR methods, which are typically found more consistent in real-time.

When comparing the real-time performance of the output gap estimates through the output gap revisions, the HP filter outperformed the structural SVAR methods, being the most consistent in real-time. Both of the output gap measures based on the SOE SVAR show very similar development of the output gap estimates. The BQ method seems to be the most inconsistent with the Finnish data, showing similar real-time properties. Moreover, it is also important to note that data revisions do not appear to be the primary source of revisions in the selected output gap measures, which is in line with the results of Orphanides and van Norden (2002) and Marcellino and Musso (2011). However, exceptionally large revisions in actual output may also affect the estimates, as observed around the global financial crisis in the case of small open economy SVAR.

Overall, these results suggest that structural SVAR based output gap measures at the end of a sample perform poorly compared to those in the middle, as the ex-post revisions appear to be greater at the end of the sample, which is contrary to the findings of Coibion et al. (2018) and Chen & Góronicka (2020) as well as Mazzi et al. (2016). The output gap measures reconstructed using the small open economy SVAR, and the BQ decompositions appear to identify too slowly whether changes in output will be long-lived in the case of Finland. In the applied structural VAR methods, low real-time performance is on the one had associated with the real-time ability to single out long-run changes in output, defined as trend component of the SVAR, and on the other, the identification of persistent and temporary shocks on output, which are reconstructing the series of potential output and the output gap. For instance, the utilized variables, such as inflation and unemployment, together with the imposed long-run and sign restrictions, may embody only a little information about the shock under inspection, particularly if the variables are persistent. Another factor that may increase the uncertainty of the estimates with Finnish data is the pronounced volatility in output, as Billmeier (2006) has stated. However, the delayed estimates show how the estimates improve remarkably after a few quarters have passed from the first publication of the GDP data, suggesting that it may be possible to improve the reliability of these measures at the end of the sample by extending series of variables with well-formed forecasts, as known from the case of filtering techniques.

7. Robustness

The robustness of the results has been checked using the alternative specification and estimation periods for each the applied model. The regressions of the small open economy SVAR with full sample data were performed with 2, 4, and 8 lag lengths. These procedures did not have a significant effect on the output gap estimates. The alternative dataset changes impulse responses to some extent, which should be kept in mind when applying the method with this dataset. These differences did not have a significant effect on potential output and the output gap estimates with full sample data. However, some differences between the size of the estimates can be observed. The comparison of the output estimates constructed using the original and the alternative dataset for each subsample period can be found in the annex.

The regressions for the BQ SVAR were performed with 2, 3, and 5 lag lengths. The estimated potential output and output gaps changed drastically in each selected lag length. In addition, quarterly unemployment rates were first considered but rejected because the results changed significantly due to greater unemployment variation. The impulse responses of the variables using the first difference of log GDP can be found in the annex.

The structural VARs were also estimated using alternative starting and ending points using a) the whole sample period starting from 1964q1 to 2020q2, b) a sample starting from 1964q1 to 1996q2, c) 1980q1 to 2020q1, d) from 1991q1 until 2020q2. The results pointed to considerably different growth decompositions and the output gap estimates, indicating considerable vulnerability to the structural changes in the Finnish economy, which cannot be isolated from the economic time series.

Therefore, it is essential to recognize these changes and deal with them appropriately when using structural VAR methods.

The problems associated with the end of the sample in the case of the HP filter were not mitigated with forecasts. However, the delayed real-time estimates can be assumed to represent extended samples with "perfect" forecasts.

Nevertheless, the results should not be taken without caution. The small number of observations in subsamples might have affected the result of this thesis, and re-evaluation should be reconsidered when the time series expands. The results also suggest that these methods may change the observed business cycle position as the time series expands. Therefore, the estimates can be expected to revise, particularly close to the end of the sample period.

8. Conclusion

This thesis examines the performance of the output gap measures, constructed by applying structural autovector regression models to the Finnish full-sample and real-time data. The output gap measures are examined across samples and methods, but the primary focus is on ex-post revisions and real-time performance. The reliability of the output gap measures is investigated using real-time data following Orphanides and van Norden (2002). The consistency of SVAR-based estimates has already been proven in previous empirical work, for example, in the euro area by Camba-Mendez & Rodriguez-Palenzuela (2003) and Mazzi et al. (2016). More recent empirical literature, based on Coibion et al. (2018) and Chen & Góronicka (2020), suggests that potential output estimates reconstructed from structural VAR decompositions are less sensitive to measurement errors in real-time due to the lesser sensitivity to temporary disturbances in output. These findings are based on the well-known Blanchard and Quah (1989) method and the small open economy SVAR proposed by Chen and Góronicka (2020) which is built on the identification scheme of Forbes et al. (2018). In this thesis, the performance of these output gap measures is examined and compared against the univariate HP filter in the context of Finland.

The output gap measures produced applying the structural VARs, and the HP filter to the full sample data form somewhat similar views about the cyclical development in Finland. Actual output appears to be above or under potential, mainly when it is supported by the history and macroeconomic indicators. However, the size of the measures varies, particularly around economic turning points. Sometimes differences between the measures may be even several percentage points, and even the sign of the estimate may differ when actual output is close to potential. The applied structural VAR-based output gap measures appear to perform reasonably well against the Finnish economic history and macroeconomic indicators. However, some inconsistencies can be observed between the indicators and output gap estimates. For example, the output gap measures based on structural VAR methods close to the end of the sample suggest large positive output gaps for the late 2010s, which are unconvincing against macroeconomic conditions.

The results suggest that the reliability of the real-time output gap estimates produced with the structural VAR methods appears to be relatively low, particularly around economic turning points, which is in contra to findings of Coibion et al. (2018) and Chen & Góronicka (2020) as well as Mazzi et al. (2016). In fact, the real-time estimates produced with the univariate HP filter outperformed structural VAR methods in the context of Finland. The ex-post revisions resulting from erroneous measurements in real-time appears to be more significant at the end of the samples. For example, all the applied methods suggest greater positive output gap estimates for the period of overheating prior to the global financial crisis (GFC) with full sample data than with real-time data, suggesting that these models tend to overestimate potential output during the periods of overheating in real-time. The results also show how the real-time estimates of the small open economy SVAR appear to follow the full sample estimates closely during the meltdown and recovery phase in the GFC, suggesting that the measure reacts to extreme economic turning points consistently, but to gradual reversal of cycles are reacted with a delay. These results are also broadly in line with the previous literature in which the revisions in the output gap measures are found to increase around the economic turning points and at the end of the sample, for example, as already presented in Orphanides and van Norden (2002) and Marcellino and Musso (2011). However, these findings are not based on the structural VAR methods, which are typically found to be more consistent in real-time.

The performance of the estimates significantly improves after a few quarters have passed from the first publication of the GDP data. The data revisions do not appear to be the primary source of revisions in output gap estimates, suggesting that all the applied models identify too slowly whether the changes in output will be permanent. However, exceptionally large revisions in actual output may also affect the estimates, as observed around the global financial crisis in the case of small open economy SVAR. The delayed real-time estimates also suggest that it may be possible to improve the reliability of the estimates at the end of the sample by extending the series of variables with well-formed forecasts, as known from the case of filtering techniques. However, this requires forecasted values for each variable, while large revisions between forecasted values and upcoming actual data may eventually end up revising the output gap estimates.

In the applied structural VAR methods, low real-time performance is associated on the one hand, to the ability to distinguish long-run changes in output, defined as trend component of the SVAR, and on the other, the identification of persistent and temporary shocks on output, which are reconstructing the series of potential output and the output gap. The utilized variables, such as inflation and unemployment, together with the imposed long-run and sign restrictions, may embody only a little information about the shock under inspection if the variables are persistent. For instance, findings regarding the small open economy SVAR suggest that the sign restrictions which impose a negative relationship between GDP and CPI associated with supply shocks and a positive relationship between CPI and GDP associated with demand shocks are weak during times characterized by persistent and low inflation. Furthermore, as Blanchard (2018) has noted, so far, these methods are not designed to identify supply shocks without permanent effects on output or demand shocks with long-lasting effects on output which may end up resulting inaccurately identified shocks. Another factor that may increase the uncertainty of the estimates with Finnish data is the pronounced volatility in output, as Billmeier (2006) has stated.

This thesis shows how the measurement errors in real-time are still an enduring problem in measures of potential output and the output gap. Methods that aim to strictly distinguish transitory and permanent effects on output, such as structural VAR decompositions, may not be less sensitive to measurement errors if long-term changes in output remain ambiguous in real-time and for subsequent quarters.

Further real-time investigation of structural VAR-based potential output and output gap measures with the Finnish data remains yet to be done with tests that were beyond the scope of this thesis. For instance, it may still be worth examining the ability of the real-time estimates to predict inflation, or how the real-time estimates of potential output respond to previously identified shocks, as presented in Coibion et al. (2018) and Chen and Góronicka (2020). However, the ability to single out long-term changes in output from changes associated with a cyclical component appears to be the main concern associated with structural VAR methods in real-time with Finnish data. Nevertheless, the output gap measures based on structural VAR decompositions may still be worth considering when measuring potential output and the output gap. However, the unreliability of real-time estimates should be taken seriously, particularly around economic turning points. For instance, misperceived development of the output gaps around economic turning points may eventually lead to incorrectly measured policy recommendations, and when implemented, intended countercyclical policies end up having unintended consequences. Hence, the reliability of the estimates should be confirmed, preferably using multiple methods and tests.

References

Alichi, A., Bizimana, O., Laxton, D., Tanyeri, K., Wang, H., Yao, J. and Zhang, F. (2017). Multivariate Filter Estimation of Potential Output for the United States. IMF Working Paper, 106.

Alichi, A., Al-Mashat, R., Avetisyan, H., Benes, J., Bizimana, O., Butavyan, A., Ford, R., Ghazaryan, N., Grigoryan, V., Harutyunyan, M., Hovhannisyan, A., Hovhannisyan, E., Karapetyan, H., Kharaishvili, M., Laxton, D., Liqokeli, A., Matikyan, K., Minasyan, G., Mkhatrishvili, S., Nurbekyan, A., Orlov, A., Pashinyan, B., Petrosyan, G., Rezepina, Y., Shirkhanyan, A., Sopromadze, T., Torosyan, L., Vardanyan, E., Wang, H., & Yao, J. (2018). Estimates of Potential Output and the Neutral Rate for the U.S. Economy. IMF Working Paper, 152.

Andrés, J., Lopez-Salido, J. & Nelson, E. (2005). Sticky-Price Models and the Natural Rate Hypothesis. Journal of Monetary Economics 52(5), 1025–53.

Álvarez, L. & Goméz-Loscos, A. (2017). A Menu on Output Gap Estimation Methods. Documentos de Trabajo, 1720, Banco De España.

Benes. J., Clinton, K., Garcia-Saltos, R., Johnson, M., Laxton, D., Manchev, P. & Matheson, T. (2010). Estimating Potential Output with a Multivariate Filer. IMF Working Paper, 285.

Billmeier, A. (2004). Measuring a roller coaster: Evidence on the Finnish output gap. IMF Working Paper, 57.

Binning, A. (2013). Unidentified SVAR Models: A Framework for Combining Short and Long-run Restrictions with Sign-restrictions. Norges Bank Monetary Policy Working Paper, 14.

Blagrave, P., Garcia-Santos, R., Laxton, D. & Zhang, F. (2015). A Simple Multivariate Filter for Estimating Potential Output. IMF Working Paper, 79.

Blanchard, O. & Quah, D. (1989). The Dynamic Effects of Demand and Supply Disturbances. American Economic Review, 79(4), 655-673.

Blanchard, O. (2018). Olivier Blanchard provides a brief reaction to Real-Time Estimates of Potential GDP, by Coibion, Gorodnichenko, and Ulate. Center on Budget and Policy Priorities, January.

Borio, C., Disyatat, P., & Juselius, M. (2014). A Parsimonious Approach to Incorporating Economic Information in Measures of Potential Output. BIS Working Paper, 44.

Camba-Mendez, G. & Rodriguez-Palenzuela, D. (2003). Assessment Criteria for Output Gap Estimates. Economic Modelling. 20(3), 529-562.

Champagne, J., Poulin-Bellisle, G. & Sekkel, R. (2018). The Real-Time Properties of the Bank of Canada's Staff Output Gap Estimates. Journal of Money, Credit and Banking, 50(6), 1167-1188.

Chen, J., & Górnicka, L. (2020). Measuring output gap: Is it worth your time? IMF Working Papers, 20(24).

Cogley, T., & Nason, J. M. (1995). Effects of the Hodrick-Prescott filter on trend and difference stationary time series. Implications for business cycle research. Journal of Economic Dynamics and Control, 19, 253–278.

Coibion, O., Gorodnichenko, Y. & M. Ulate, (2018). The Cyclical Sensitivity in Estimates of Potential Output. Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, 49(2), 343-441.

Guisinger, A., Owyang, M. & Shell, H. (2018). Comparing Measures of Potential Output. Federal Reserve Bank of St. Louis Review, 100(4), 297-316.

Gorodnichenko, Y., Mendoza, E. & Tesar, L. (2012). The Finnish Great Depression: From Russia with Love. American Economic Review, 102 (4), 1619–44.

Gulan, A., Haavio, M. & Kilponen, J. (2014). Kiss Me Deadly: From Finnish Great Depression to Great Recession. Bank of Finland Discussion Paper.

Edge, R. & Rudd, J. (2016). Real-Time Properties of the Federal Reserve's Output Gap. The Review of Economics and Statistics, 98(4), 785-791.

Eurostat (2013). Handbook on quarterly national accounts, Eurostat Manuals and Guidelines, 2013 Edition.

Faust, J., & Leeper, E. M. (1997). When do long-run identifying restrictions give reliable results? Journal of Business and Economic Statistics, 15(3), 345–353.

Forbes, K., Hjortsoe, I. & Nenova, T. (2018). The shocks matter: Improving our estimates of exchange rate pass-through. Journal of International Economics, 114, 255-275.

Fry, R. & Pagan, A. (2007). Some Issues in Using Sign Restrictions for Identifying Structural VARs. National Centre for Econometric Research, Working Paper, 14.

Fry, R. & Pagan, A. (2011). Sign Restrictions in Structural Vector Autoregressions: A Critical Review. Journal of Economic Literature, 49(4), 938-60.

Gali, J. (1999). Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations? American Economic Review, 89(1), 249-271.

Hamilton, J.D., (2018). Why You Should Never Use the Hodrick-Prescott Filter. The Review of Economics and Statistics, 100(5), 831-843.

Harvey, A. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Havik, K., McMorrow, K., Orlandi, F., Planas, C., Raciborski, R., Roeger, W., . . . VanderMeulen, V. (2014). The production function methodology for calculating potential growth rates and output gaps (No. 535). Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Hernández de Cos, P., Lacuesta, A. & Moral-Benito, E. (2016). An exploration of real-time revisions of output gap estimates across European countries. Banco de España Occasional Paper, 1605.

Hjerppe, R, (1989). The Finnish Economy 1860-1985 – Growth and Structural Change. Bank of Finland, Government Printing Centre.

Hodrick, R. & E. Prescott, (1997). Postwar U.S. Business Cycles: An Empirical Investigation. Journal of Money, Credit, and Banking, 29(1), 1-16.

Honkapohja, S. & Koskela, E (1999). The economic crisis of the 1990s in Finland. Economic Policy, 14(29), 399–436.

Jonung, L, Kiander, J & Vartia, P. (2009). The Great Financial Crisis in Finland and Sweden. The Nordic Experience of Financial Liberalization, Edward Elgar Publishing Inc.

Jysmä, S., Kiema, I., Kuusi, T., & Lehmus, M. (2019). The Finnish potential output: Measurement and medium-term prospects. Publications of the Government´s Analysis, Assessment and Research Activities, 13.

Kiander, J & Vartia, P. (1996). The great depression of the 1990s in Finland. Finnish Economic Papers, 9(1), 72–88.

Kuttner, K., (1992). Monetary Policy with Uncertain Estimates of Potential Output. Economic Perspectives, Federal Reserve Bank of Chicago, 16, 2-15.

Kuttner, K. (1994). Estimating Potential Output as a Latent Variable. Journal of Business and Economic Statistics, 12(3), 361–68.

Kuusi T. (2015). Alternatives for measuring structural budgetary position. Publication Series of the Government's Analysis, Assessment and Research Activities, 5.

Marcellino, M., & Musso, A. (2010). Real time estimates of the euro area output gap: Reliability and forecasting performance. European Central Bank Working Paper Series, 1157.

Marcellino, M. & Musso, A. (2011). The reliability of real-time estimates of the Euro-Area output gap. Economic Modelling, 28, 1842-1856.

Melolinna, M. (2010). Euroalueen ja Suomen tuotantokuilu. BoF Online, Suomen Pankki, 4.

Melolinna, M. & Tóth, M. (2016). Output Gaps, Inflation and Financial Cycles in the United Kingdom. Bank of England Staff Working Paper, 585.

Mishkin, F. (2007). Estimating Potential Output. Speech given at Conference on Price Measurement for Monetary Policy, Federal Reserve Bank of Dallas, Dallas, May 24.

Mourre, G., Astarita, C., & Princen, S. (2014). Adjusting the budget balance for the business cycle: The EU methodology (No. 536). Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Mazzi, G., Mitchell, J., & Moauro, F. (2016). Structural vector autoregressive (SVAR) based estimates of the euro area output gap: theoretical considerations and empirical evidence. Publications Office of the European Union.

Newby, E., & Orjasniemi, S. (2012). Potentiaalisen tuotannon arviointimenetelmiä. BoF Online, Bank of Finland.

Okun, A. (1962). Potential GNP: Its Measurement and Significance. American Statistical Association–Proceedings of the Business and Economics Statistics Section, 98–104.

Orphanides, A., (2001). Monetary policy rules based on real-time data. American Economic Review, 91(4), 964-985.

Orphanides, A., (2003a). Monetary policy rules with noisy information. Journal of Monetary Economics, 50(3), 605–631.

Orphanides, A., (2003b). The quest for prosperity without inflation. Journal of Monetary Economics, 50(3), 633–663.

Orphanides, A. & van Norden, S. (2002). The Unreliability of Output-Gap Estimates in Real Time. The Review of Economics and Statistics, 84(4), 569-583.

Official Statistics of Finland (OSF) (2013): Quarterly national accounts [e-publication]. Historical time series of Quarterly National Accounts. Helsinki: Statistics Finland. Access method: http://www.stat.fi/til/ntp/ntp_2013-02-02_men_001.html, 26.5.2021.

Official Statistics of Finland (OSF) (2014a): Quarterly national accounts [e-publication]. Methodological description of Quarterly National Accounts. Helsinki: Statistics Finland. Access method: http://www.stat.fi/til/ntp/2014/ntp_2014_2014-10-13_men_001_en.html, 26.5.2021.

Official Statistics of Finland (OSF) (2014b): Quarterly national accounts [e-publication]. Helsinki: Statistics Finland. Access method: http://www.stat.fi/til/ntp/ntp_2014-07-11_uut_001_en.html, 26.5.2021.

Mc Morrow, K., Roeger, W., Vandermeulen, V. & Havik, K. (2015). An assessment of the relative quality of the Output Gap estimates produced by the EU's Production Function Methodology. European Commission, Directorate General for Economic and Financial Affairs European Economy, Discussion Paper, 020.

Solow, R. (1956). A Contribution to the Theory of Economic Growth. The Quarterly Journal of Economics, 70(1), 65-94.

Tosetto, E., (2008). Revisions of Quarterly Output Gap Estimates for 15 OECD Member Countries, OECD Statistics Directorate.

Turner, D., Cavalleri, M.C., Guillemette, Y., Kopoin, A., Ollivaud, P. & Rusticelli, E., (2016). An investigation into improving the real-time reliability of OECD output gap estimates. OECD Economics Department Working Papers, 1294.

Sariola, M. (2019). An unobserved components model for Finland – estimates of potential output and NAWRU. BoF Economics Review, Bank of Finland, 2.

Suni, P, & Vihriälä, Vesa (2016). Finland and Its Northern Peers in the Great Recession. ETLA Reports, 49.

Vetlov, I., Hlédik, T, Jonsson, M, Kucsera, H. & Pisani, M. (2011). Potential output in DSGE models. European Central Bank Working Paper Series, 1351.

Virkola, T. (2014). Real-time measures of output gap and fiscal policy stance. ETLA Raportit, 37.

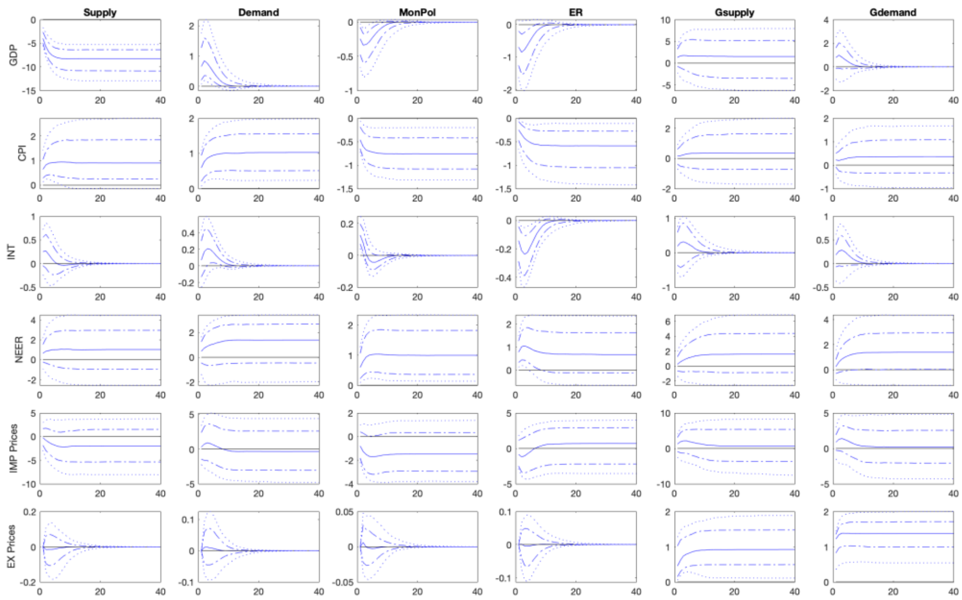
Vihriala, Vesa, (1997). Banks and the Finnish Credit Cycle 1986–1995. Bank of Finland Studies, E:7, Bank of Finland.

Appendix

Table 4. Data Sources and Variable Definitions

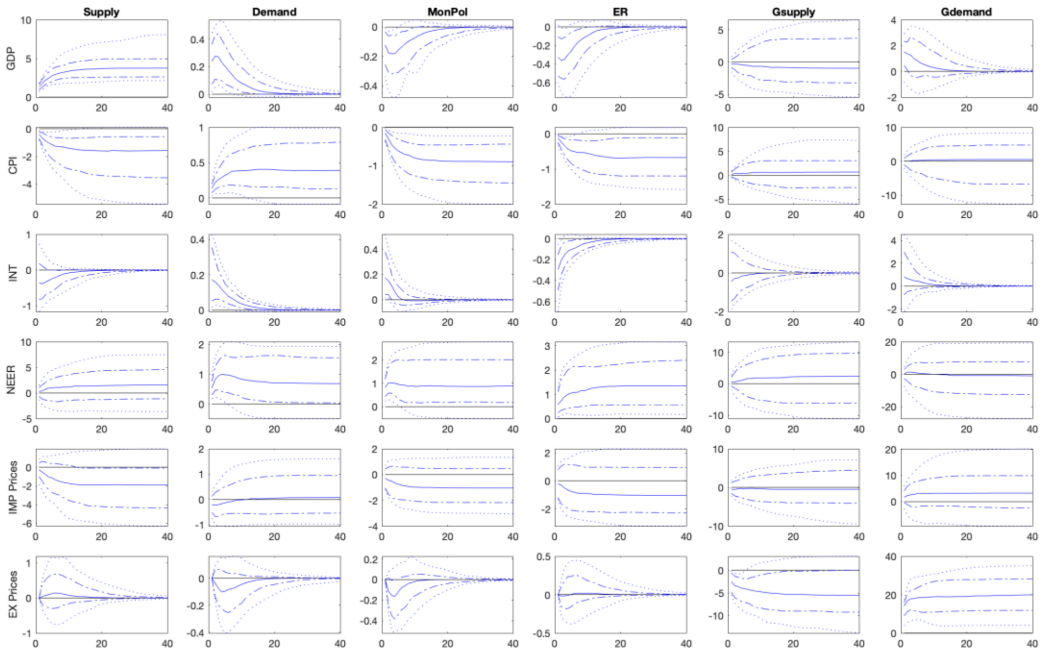
Variable	Source 1.	Comment	Source 2.	Comment
real GDP	Statistics Finland	Quarterly national accounts		
rate of unemployment	Statistics Finland	Annual rates	ILO	Annual estimates
consumer price index	Statistics Finland	Monthly consumer price index: Harmonized Index of Consumer Prices (HICP) from 1996Q3 to 2020Q2; Headline Consumer Price Index from 1964Q1 to 1996Q2		
interest rate	Bank of Finland	Monthly averages of central bank lending rate for commercial banks from 1964Q1 to 1975Q1; Monthly average of BoF intraday credit rate from 1975Q1 to 1992Q2; Tender rate from 1992Q1 to1998Q4; ECB MRO/deposit rate from 1996Q1 to 2020Q2.		
nominal exchange rate index	BIS	Narrow Indices		
import prices	WEO	FI imports of goods and services price deflator	Statistics Finland	Producer price index (PPI): Import prices
export prices	WEO	World CPI weighted by FI export share	Statistics Finland	Producer price index (PPI): Import prices
reference turning points	OECD	OECD Composite Leading Indicators: Reference Turning Points and Component Series		

Figure 11. SOE SVAR: Impulse Responses for the Period of Currency Union



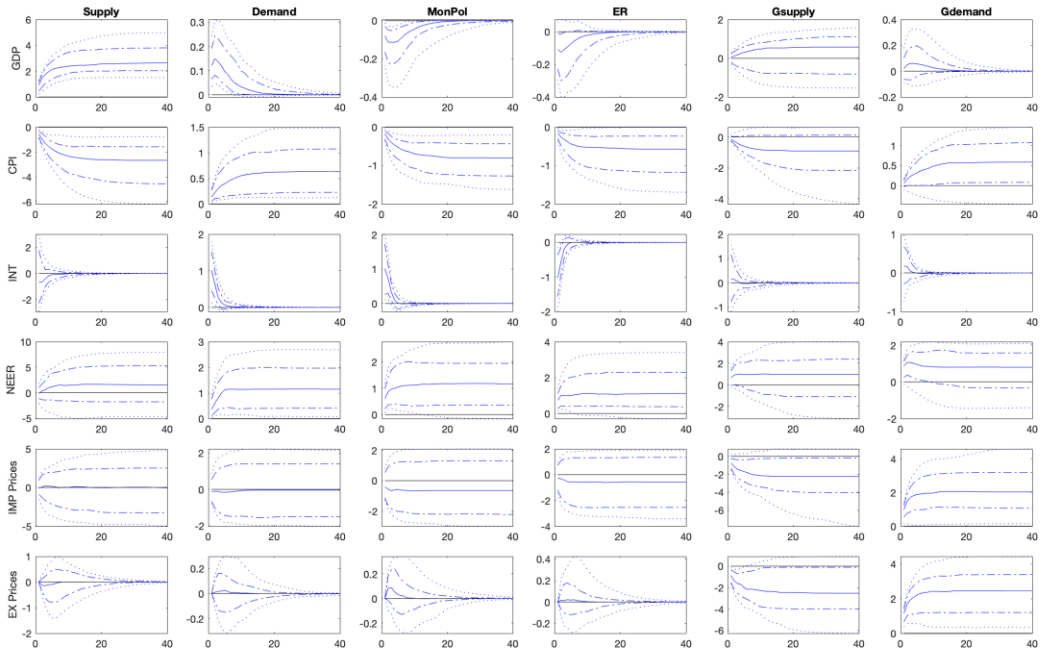
Note: The figure plots the median impulse responses with the 68% and 90% confidence intervals for each variable to the respective shock, constructed using the first (log) difference of variables, except for the interest rate that is already detrended.

Figure 12. SOE SVAR: Impulse Responses for the Period of Economic Liberalization



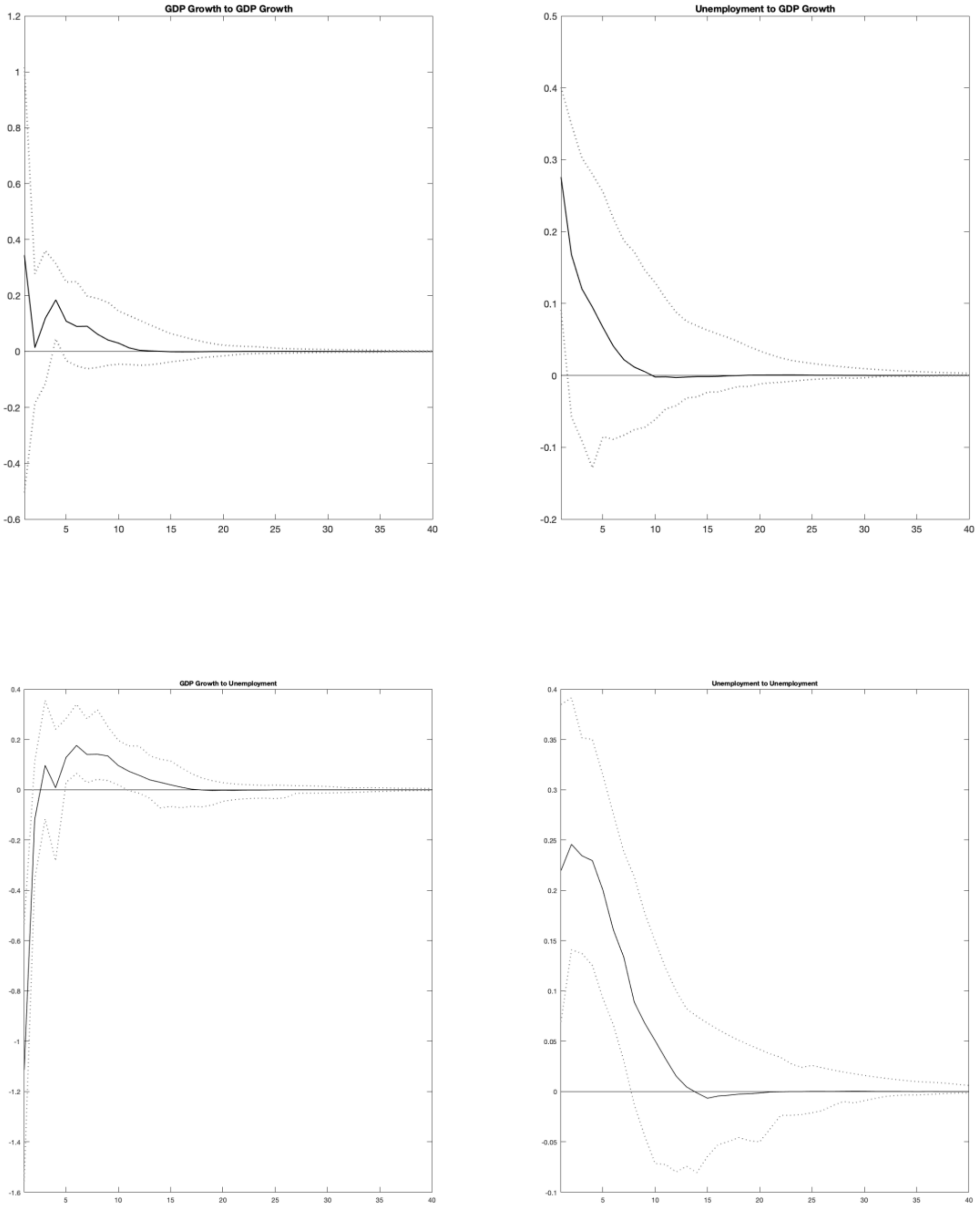
Note: The figure plots the median impulse responses with the 68% and 90% confidence intervals for each variable to the respective shock, constructed using the first (log) difference of variables, except for the interest rate that is already detrended.

Figure 13. SOE SVAR: Impulse Responses for the Period of Economic Regulation



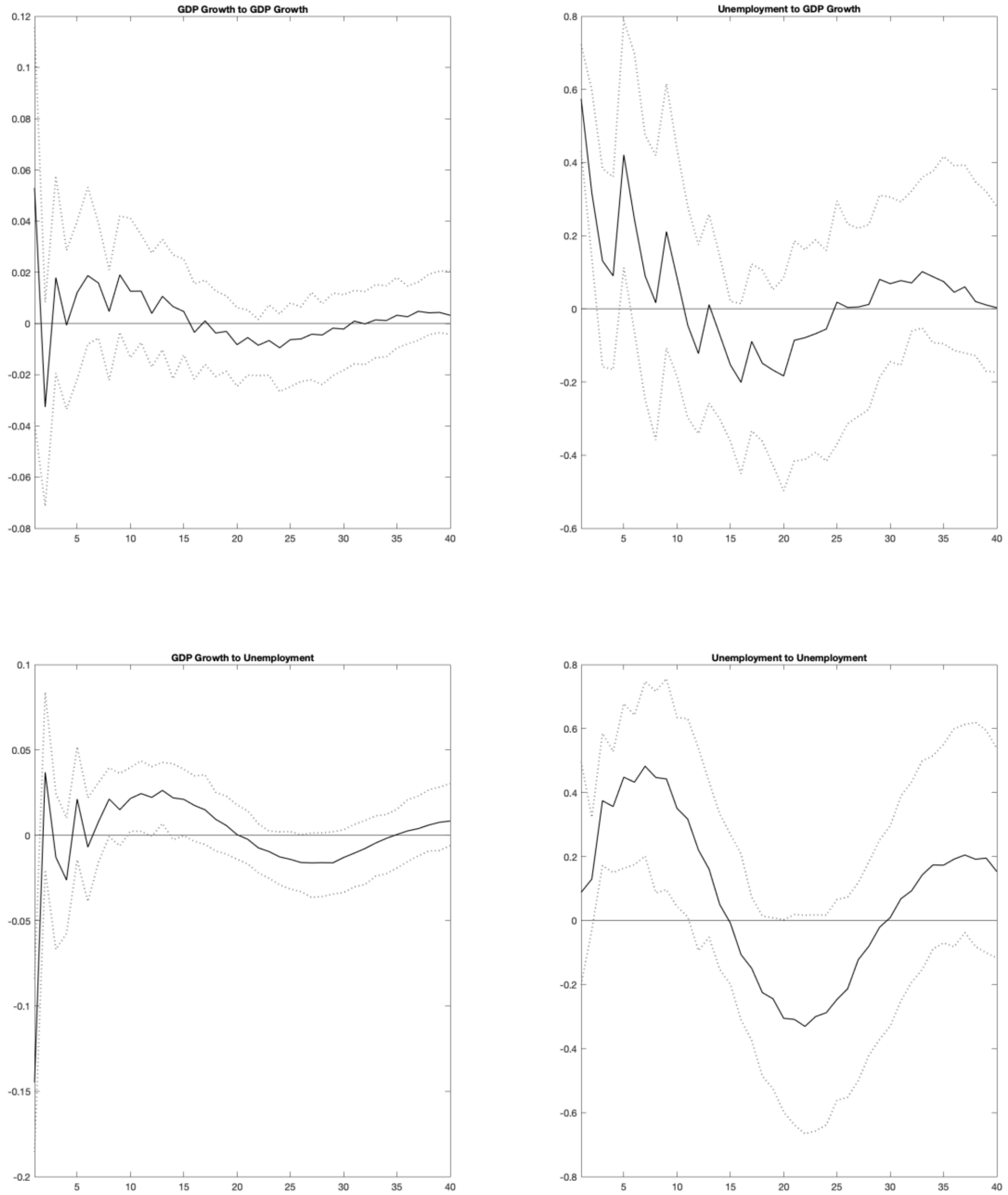
Note: The figure plots the median impulse responses with the 68% and 90% confidence intervals for each variable to the respective shock, constructed using the first (log) difference of variables, except for the interest rate that is already detrended.

Figure 14. BQ SVAR: Impulse Responses for the Period of Currency Union



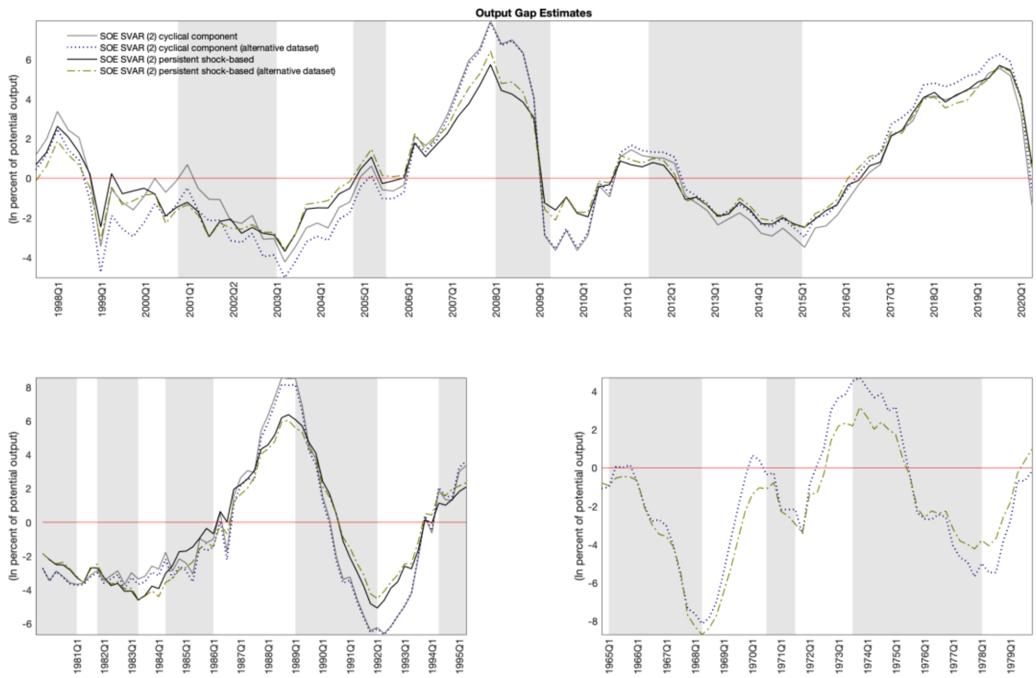
Note: The figure plots the impulse responses for the variables using the first difference of log GDP.

Figure 15. BQ SVAR: Impulse Responses for the Period of Economic Liberalization



Note: The figure plots the impulse responses for the variables using the first difference of log GDP.

Figure 16. SOE SVAR: Comparing Original and Alternative dataset



Notes: Figure 15 presents the output gap estimates for Finland constructed with the original and the alternative dataset using full sample data (1997Q2–2019Q4, 1997Q2–2019Q2, 1997Q2–2018Q2). The original dataset follows the dataset proposed by Chen and Góronicka (2020). In the alternative dataset, import and export prices are derived from PPI.