FAQ: How do I estimate the output gap?

Fabio Canova *
BI Norwegian Business School, CAMP and CEPR

November 4, 2022

Abstract

I investigate the properties of gaps and potentials in a variety of New Keynesian DSGE models and their relationship with estimates obtained with standard approaches. Theoretical gaps display low frequency variations, have similar frequency domain representation as potentials, and are correlated with them. Theoretical transitory and permanent fluctuations display similar features, but are uncorrelated. All existing approaches generate distortions in the estimates. Gaps are best estimated with a polynomial filter; transitory fluctuations with a differencing approach. Explanations for the outcomes are given. I design a statistical procedure reducing the biases of existing methods.

Key words: Gaps and potentials, DSGE models, labor share, Butterworth filter.

JEL Classification: C31, E27, E32.

*I thank Jim Hamilton, Robert Hodrick, Harald Uhlig, Todd Walker, Christian Matthes, Paolo Gelain, Jesper Lindé, Elman Mertens, Evi Pappa and the participants of seminars at the Riksbank, ECB, University of Cardiff, University of Indiana, Bank of Canada, Cleveland Fed, Dutch National Bank, Texas A&M, UC3M, University of Cyprus, Bank of Finland, First Macro Dolomiti Conference, Castelrotto, and the WGEM workshop, Paris for comments and suggestions.

1 Introduction

Since Woodford [2003] academic economists have been keenly interested in the dynamics of certain latent (star) variables, such as the output potential, the natural rate of interest, or the NAIRU, as they contain information about the first best that an economy can achieve. They have also been concerned with the changes in the properties of these variables that extraordinary events, such as the 2008 financial crisis or the COVID-19 pandemic, may have produced. Policymakers, on the other hand, care about deviations of actual variables from their star counterparts (the gaps) as they characterize inefficiencies that policy may want to eliminate. Thus, they may want respond with monetary policy actions, e.g., to a widening output gap, or use output gap estimates to predict the current level of inflation, via a Phillips curve, or the unemployment rate, via Okun's law.

Unfortunately, while star quantities are central in macroeconomic discussions, they are unobservable. In addition, their measurement is elusive, as it requires counterfactuals involving the identification of shocks and the quantification of particular distortions. For example, the output potential is typically defined as the level of output prevailing absent nominal frictions once monetary and markup shocks are eliminated. Because one needs to eliminate the effects of frictions such a measurement exercise is naturally performed within the context of an estimated (calibrated) structural model. Nevertheless, for measurement exercises, statistical procedures are generally preferred, as they are considered less prone to misspecification of the structural relationships and thus more robust. However, existing statistical approaches typically employ identification assumptions which are inconsistent with the information that New Keynesian (NK) structural models used to discuss the dynamics of stars and gaps variables provide. Hence, different users draw conclusions about the empirical properties of star and gap variables using concepts and measurements that may have little to do with their theoretical counterparts. In practice, e.g. the term output gap is employed to refer to the difference between actual output and its permanent component, defined as the level prevailing in the long run; or the difference between actual output and its trend, where a family of statistical concepts could be used to define the latter. But potential output might not be trending in a statistical sense; and may feature both permanent and transitory swings.

The lapse between theory and measurement undermines both academic investigations and clouds policy advises. In recent years macro-economists have argued about what the data tells us about potential output, see Coibon, Gorodnichenko, and Ulate [2018], the properties of the natural rate of interest, see Laubach and William [2016] or Del Negro, Giannone, Giannoni, and Tambalotti [2019], the dynamics of NAIRU, see Crump, Giannoni, and Sahini [2019] or the unemployment gap

Crump, Giannoni, and Sahini [2022] using statistical models which can not be mapped into the fundamental equations present in NK theories. In addition, given the variety of statistical latent variable extraction tools, researchers also disagree on the timing, the dynamics and the features of the estimated deviations from star variables. Thus, they contend about how to measure cyclical fluctuations, see Hamilton [2018], whether "the trend drives the cycle" or the reverse is true, see Aguiar and Gopinath [2007], and Heathcote, Perri, and Violante [2020]; on the role of permanent and transitory disturbances for short and long run macroeconomic fluctuations, see Schmitt-Grohe and Uribe [2019], Jorda', Singh, and Taylor [2020]; or the theories consistent with cyclical facts, see Angeletos, Collard, and Dellas [2020].

The contribution of the paper. This paper investigates on the relationship between theoretical notions of potential and gaps (or permanent and transitory fluctuations) and the estimates recoverable with statistical approaches using laboratory economies. I first demonstrate in a simple analytical example why existing procedures fail to deliver estimates with the right population properties. I then conduct a Monte Carlo exercise to rank procedures and quantify the distortions, employing a number of structural NK models as data generating process (DGP) and a variety of procedures to split the simulated data into two latent components. Finally, I propose a statistical method, that could be used to tighten the match between theory based and empirical based concepts, which has smaller biases.

In a structural model, potential and gaps are well-defined objects. As mentioned, potentials are the equilibrium outcomes obtained eliminating nominal frictions, markup and monetary disturbances, and the gaps are the deviations between the level and the potential of the variables. Similarly, the permanent component of a variable is what the model produces when certain disturbances display a unit root, and the transitory component is the difference between the level and the permanent component. Thus, when a structural model is used as DGP, one can rank estimation approaches and examine the reasons for why distortions occur.

Since even the most sophisticated NK models are likely to have important empirical deficiencies, one may wonder whether any general conclusion can be derived from such an exercise. Indeed, robust conclusions be obtained because what makes statistical methods inadequate are features intrinsic to dynamic models and their solution procedure and not particular choices made by researchers. In fact, in a dense class of dynamic models, gaps and potentials are correlated, display similar persistence, have low and business cycle frequency components which account for similar portion of their total variance. In addition, gaps display important low frequency variations and potential relevant business cycle variations. Permanent and transitory fluctuations also display similar characteristics but are uncorrelated by construction.

These robust features of gaps and potentials are in turn determined by three generic properties of dynamic models: the linear approximation of the solution for the potentials and the gaps has a VAR representation; certain disturbances drive both gaps and potentials; when these shocks are persistent, both latent variables display the typical "Granger spectral shape". While the spectral similarities in potentials and gaps can be made larger when the DGP possesses features enhancing low frequency components in the gaps, such as Beaudry, Galizia, and Portier [2020], stretching out the effect of cyclical shocks in the medium run, such as Gertler and Comin [2006], or making equilibrium adjustments lumpy, such as Thomas [2002], no elaborated dynamic equilibrium model will make the component uncorrelated, display different persistence and significantly change the spetral shape of potentials and gaps. In addition, while relative contribution of gaps to the total variance at low and business cycle frequencies depends on the frictions present in the economy, the potentials will always account for a non-negligible portion of the variance of the observable at business cycle frequencies and the gaps for a non-negligible portion of the variance at low frequencies, contrary to what most statistical approaches assume.

I employ numerous extraction methods, covering well the set of procedures commonly used and employ numerous statistics to compare them. Because many approaches features free parameters and some require parameter estimation, I also examine whether the outcomes are affected by alternative choices of the free parameters or when the sample size changes.

To avoid misunderstandings, it is important to highlight what the paper does not. This paper is not concerned with finding the best way to isolate a stationary component from potentially non-stationary data for the estimation of a stationary structural model (a question studied, e.g. in Canova [2014]); nor about the properties of cyclical fluctuations or of parameter estimates when the DGP features trend-stationary or difference-stationary characteristics, see Cogley and Nason [1995] or Gorodnichenko and Ng [2010]. It is also not about constructing robust stylized facts, as for example discussed Canova [1998] and Canova [1999]; or the properties of certain filters, see **?** for a recent example. While all these questions are of interest, they are unrelated to the issues discussed here which arise because statistical estimation procedures employ identification assumptions which are inconsistent with the basic features of dynamic NK models.

The related literature To the best of my knowledge a systematic analysis of the features of model-based potentials and gaps and of the mapping between theoretical concepts and estimation outcomes is absent from the literature. Christiano, Trabant, and Walentin [2011] simulated data from a search-and-matching friction model and studied the performance of the Hodrick and Prescott (HP) and of an optimal two-sided filter in capturing model-based output potential. Hodrick [2020] simulated data from a number of time series models and compared the ability of HP, band pass (BP), and local projection filters to separate permanent from transitory fluctuations. Because the models considered are of reduced form type, his work nicely complements the analysis conducted here.

Beaudry et al. [2020] showed that there are interesting medium term fluctuations in hours, because they have a length that does not fit the typical definition of cyclical fluctuations, and typically disregarded in the analysis. They argue that these fluctuations are crucial to better understand the type of cyclical models consistent with the data; see Kulish and Pagan [2019] for a critical view. Lubik, Matthes, and Verona [2019] argue that many macroeconomic time series display important medium term fluctuations, casting doubts on the evaluation exercise conducted by Angeletos et al. [2020], which solely focuses at business cycle frequencies.

The punchline. The paper has two simple conclusions. If the data has been generated by the class of NK models macro-economists and policymakers employ to interpret the dynamics of star variables, toexamine aggregate fluctuations and to produce out-of-sample predictions, the available toolkit of latent variable estimation procedures is inappropriate. Surprisingly, the oldest (Polynomial) and the simplest (Growth differencing) procedures turn out to be the least distorting.

Furthermore, the practice of defining as cyclical those fluctuations with 8-32 quarters periodicity needs considerable refinement. When standard dynamic macroeconomic models are taken seriously, filters constructed to extract cycles with such periodicity in mind or evaluation exercises comparing the performance of cyclical models at those frequencies produce severe inferential distortions. Models generate gaps (transitory components) where low frequency (32-64 quarters) fluctuations are as or more important than business cycle (8-32 quarters) fluctuations; at the same time, they produce potentials (permanent components) with considerable business cycle variations. Thus, focusing on 8-32 quarters fluctuations overestimate the variability of the gaps (transitory components) at business cycle frequencies and underestimate their variability in the low frequencies. These errors alter the sequence and the number of turning points, the amplitude and the duration of expansions and recessions, and the ability to predict interesting macroeconomic trade-offs.

Where to go next. Given that empirical approaches are inadequate, one may decide structurally estimate the model assumed to have generated the data, and construct model-based estimates of gaps and potentials, in the spirit of Christiano et al. [2011], Justiniano, Primiceri, and Tambalotti [2013], or Furlanetto, Gelain, and Taheri-Sanjani [2021]. When misspecification is a concern, the composite likelihood approach of Canova and Matthes [2021] could be used to robustify inference.

While structural models are popular, statistical procedures are likely to continue to be used for latent variable estimation in the foreseeable future, perhaps benchmarking the outcomes with an estimated model, as e.g. in Croitorov, Hristov, McMorrow, Pfeiffer, Roeger, and Vandermuellen [2019]. Bearing this in mind, I design a simple univariate statistical approach, which is consistent with the information NK models provide. The procedure is flexible, can be rigged to produce estimated gaps with arbitrary correlation pattern with potentials, and important low frequency variations; and potentials with interesting business cycle frequency variations. The procedure uniformly outperforms the best extraction procedures in the Monte Carlo experiments; and, once used in practice, generate output gaps matching the spectral properties of the US labor share.

The rest of the paper. The next section provides a simple model indicting why existing statistical approaches will fail. Section 3 discusses the design of the experiments and shows properties of potentials and gaps in a number of models. Section 4 lists the estimation procedures. Section 5 presents the statistics used; and section 6 summarizes of outcomes. Section 7 interprets the results and designs an alternative estimation procedure for gap extraction. Section 8 concludes. The on-line appendix has a number of complementary tables and figures.

2 Why statistical estimation fails: some intuition

To highlight why there is a mismatch between theoretical measures of potentials and gaps and the etsimates obtained with existing statistical procedures, consider a simple NK model, see Gali [2015]

$$\begin{aligned} \pi_t &= \beta E_t \pi_{t+1} + k\tilde{y}_t && (1)\\ \tilde{y}_t &= -\frac{1}{\sigma}(i_t - E_t\pi_{t+1} - r_t^n) + E_t\tilde{y}_{t+1} && (2)\\ r_t^n &= \rho + \sigma\psi_y^n E_t(\Delta a_{t+1}) && (3)\\ i_t &= \rho + \phi_\pi \pi_t + \phi_y \tilde{y}_t + v_t && (4) \end{aligned}$$

where v_t is a monetary policy shock; a_t a technology shock; $\tilde{y}_t = y_t - y_t^n$ is the output gap and y_t^n the output potential; β is the discount factor, $\rho = -\log(\beta)$, σ is the coefficient of constant relative risk aversion, $\psi_y^n = \frac{1+\psi}{\sigma(1-\alpha)+\psi+\alpha}$;$\psi$ is the inverse of the Frish elasticity, $(1-\alpha)$ labor exponent in the production function. Using the method of undetermined coefficients the solution for $(\tilde{y}_t, y_t^n)$ is

$$\begin{aligned} \tilde{y}_t &= -(1-\beta\rho_a)\sigma\psi_y^n(1-\rho_a)\Lambda_a a_t - (1-\beta\rho_v)\Lambda_v v_t \\ y_t^n &= \psi_y^n a_t + \gamma_y^n \end{aligned} \tag{5}$$

where

$$\Lambda_v = \frac{1}{(1-\beta\rho_v)[\sigma(1-\rho_v)+\phi_y]+k(\phi_\pi-\rho_v)} > 0 \tag{6}$$

$$\Lambda_a = \frac{1}{(1-\beta\rho_a)[\sigma(1-\rho_a)+\phi_y]+k(\phi_\pi-\rho_a)} > 0 \tag{7}$$

$$\gamma_y^n = \frac{-(1-\alpha)(\mu - log(1-\alpha))}{\sigma(1-\alpha)+\psi+\alpha} > 0 \tag{8}$$

Equation (5) highlights three important facts. The technology shock a_t enters both the solution for the output gap $\tilde{y}_t$ and for the output potential y_t^n. Hence, the two latent components will be generally correlated. In addition, when a_t is persistent, both the output gap and the output potential will be persistent. Finally, if the persistence of a_t is larger than the persistence of v_t, $\tilde{y}_t$ and y_t^n will display similar distribution of the variance by frequency.

While in this simple model there are only two shocks, in more complicated models there would be many disturbances entering the solution of $\tilde{y}_t$ and y_t^n (e.g. investment specific, government spending, preference shocks, etc.) and other disturbances only affecting the gap (e.g. markup shocks). Thus, also in more complicated models, potentials and gaps will be correlated, their persistence will be driven by the more persistent shock entering in both, and the spectral shape will be similar, provided that the disturbances entering only the gap are less persistent than those affecting both latent components, a very mild condition, generally satisfied by the parameterizations used in the literature. Notice also that, since government spending and the preference shocks enter the solution of both the potential and the gap when there are more than two shocks, it is not generally true that output potential is driven only by "supply" shocks.

There is one interesting special case when, in the above model, the results will be different. If a_t has a unit root $\rho_a = 1$, the output gap $\tilde{y}_t$ is driven only by monetary shocks and output potential y_t^n only by technology shocks. Thus, in this case, the components will be uncorrelated, supply shocks drive the potential and demand shocks drive the gap. In addition, the potential will be integrated and the gap stationary, but may still display similar distribution of the variance by frequency if v_t are persistent. Nevertheless, in models with more than two shocks, even when $\rho_a = 1$, the two components will be correlated and both "demand" and "supply" shocks may drive the potential.

In a world with unit roots, a permanent-transitory decomposition is an interesting alternative to a potential-gap decomposition. In the simple model we consider, the two decompositions are identical, i.e.

$$y_t^T = y_t - y_t^P = -(1-\beta\rho_a)\Lambda_v v_t = y_t - y_t^n \quad (9)$$

$$y_t^P = y_t^n + (1-\beta\rho_a)\sigma\Psi_y^n(1-\rho_a)\Lambda_a a_t = y_t^n \quad (10)$$

where y_t^T is the transitory component and y_t^P the permanent component. Note that the orthogonality of two components will still hold by construction in larger models with a larger number of shocks. Still, if there are more than two disturbances and stationary shocks are highly persistent, the transitory component will be highly persistent and, away from the zero frequency, the permanent and the transitory components will have similar distribution of the variance by frequency. Thus, also in this case, the main features we emphasize hold.

The above discussion can be applied to any star and gap variable generated by a NK model. For example, the solutions for the the nominal interest rate gap $\tilde{i}_t$ and the natural rate of interest (r_t^n) in the simple model are

$$\tilde{r}_t = -\rho_a[\sigma\psi_y^n(1-\rho_a)k\Lambda_a a_t - [(1-\beta\rho_v)(1-\rho_v)\sigma - \rho_v k]\Lambda_v v_t \quad (11)$$

$$r_t^n = -\sigma\psi_y^n(1-\rho_a)a_t \quad (12)$$

Since the technology shock affects $\tilde{r}_t$ and r_t^n, the same issues emerge also in this case.

This simple example highlights a few important characteristics of NK dynamic equilibrium models. Potentials and gaps are generally correlated; they have similar persistence; and, at the relevant frequencies, similar distribution of the variance. When a unit root is present and a permanent-transitory decomposition is employed, the two latent components will be uncorrelated, but their persistence and the distribution of the variance by frequency will still be similar, provided that transitory shock are sufficiently persistent.

These features hold in population (and thus they are independent of the sample size available for estimation); concern any star and gap variable in NK models generate, and appear regardless of the details of the underlying economy. For example, as shown in the next section, adding financial frictions or introducing mechanisms allowing to stretch the effect of transitory shocks in the long run does not change these qualitative features. They will also emerge in models with different or no microfundations as long as their linear approximation has a VAR solution (state space solution if there are endogenous state variables) and the shocks driving the economy are persistent.

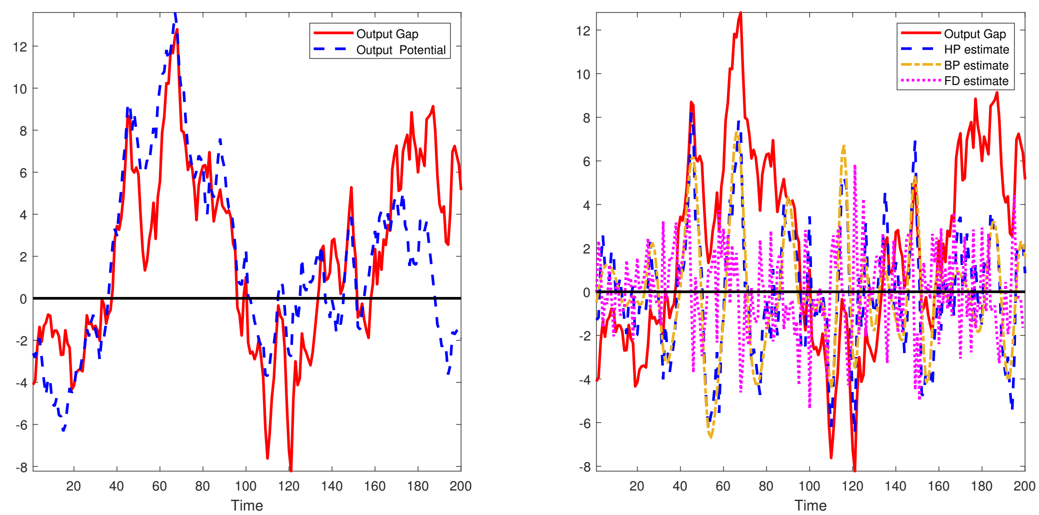

Figure 1: Gap and potential output, SW model and a few estimates

To illustrate the points in the starkest possible manner figure 1 plots one time series for $(\tilde{y}_t, y_t^n)$ simulated from one of the medium scale models used in the experimental exercise (the Smets and Wouters [2007] model). The time paths are clearly similar; the correlation between the two components is around 0.8; and their AR(1) coefficients are, respectively, 0.96 and 0.97. Figure 2 confirms that the distribution of the variance by frequency is similar.

Statistical filters. The large majority of the statistical procedures employed to separate two latent components from one observable time series employ two identifying assumptions: a) the components are uncorrelated, and b.1) the persistence of one component is larger than the persistence of the other or b.2) the distribution of the variance by frequency differs. For example, the Hodrick and Prescott (HP) filter assumes a) and b.2) while the Blanchard and Quah VAR decomposition assumes a) and b.1). Because identification assumption are inconsistent with the DGP, it is just by sheer luck that the estimated gap will resemble the gap relevant for the policy discussion.

To illustrate, the second panel of figure 1 plots the true output gap and estimates obtained with three leading statistical approaches: HP and Band pass (BP) filters, and a first difference (FOD) filter. Needless to say, all estimate fail to capture the low frequency movements present in the theoretical gap. As the second panel of figure 2 indicates, the mistakes are due to the fact that these filters display squared gain functions which considerably deviate from the theoretical squared gain

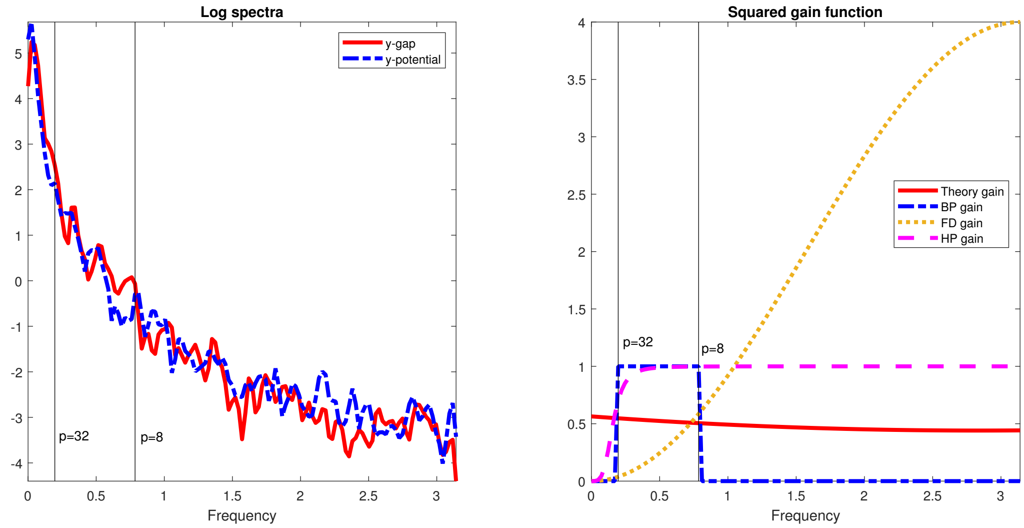

Figure 2: Log spectral density of gap and potential output, SW model and a few squared gain function.

function, which is roughly constant across frequencies.

3 The design of the experiments

I use a number of NK general equilibrium models as the DGP for the controlled experiments I run. In a NK model potentials and gaps are well defined concepts; a set of equations characterizing the optimality conditions for the potential economy - that is, the economy without nominal frictions, markup and monetary disturbances - can be added to the optimality conditions of the gap problem; and the solution for both type of variables is jointly obtained. Level variables are then calculated as the sum of the gaps and the potentials.

The features of gaps and potentials in the baseline model. The baseline setup I use is the standard closed economy model popularized by Smets and Wouters [2007] (SW) with real frictions (habit in consumption and investment adjustment costs), nominal frictions (price and wage stickiness and indexation), a Taylor rule for interest rate determination, and seven structural disturbances (to Total factor productivity (TFP), to investment, to government expenditure, to the Taylor rule, to the price and wage markups, and to the risk premium), all of which are assumed to be stationary. This model is the natural benchmark for three reasons: it has a good fit to the data of many countries;

it has been used to analyze policy trade-offs and optimal policy decisions, see e.g. Justiniano et al. [2013]; many policy institutions use versions of this model for out-of-sample forecasting exercises. The baseline parameters I use are SW posterior mean estimates.

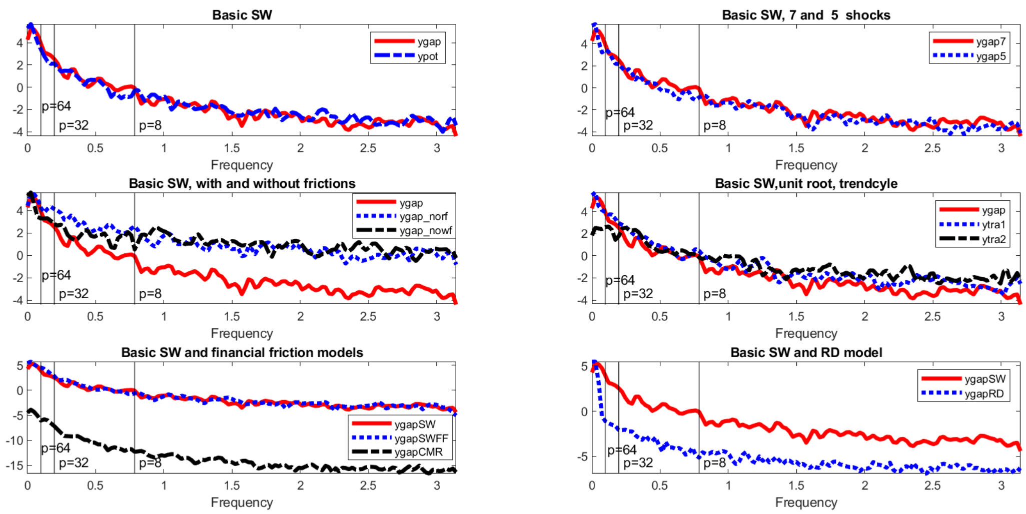

Figure 3: Log spectral densities, output gaps and output potentials, different models

The first panel of figure 3 plots the population log spectral density of the output gap and of the output potential obtained from the SW model. Note that the low frequency variability in the gap (32-64 quarters) is larger than its business cycle variability (8-32 quarters); the gap and the potential have similar spectral shape; and they almost equally account for the variance of output at low and business cycle frequencies. Furthermore, because TFP, investment, preference and government expenditure disturbances drive both the gap and the potential, the two latent components are correlated. These features are not unique to output. Figure C.1 in the on-line appendix shows that the log spectral properties of the gaps and the potentials of consumption, investment, hours, real wage, real return to capital, and capacity utilization are very similar to the ones of output.

Tinkering with the specification. The three qualitative features I emphasize above do not depend on the parameter values assumed. As long as the parameter are drawn from the estimated posterior distribution and results averaged, the log two spectral densities will still be similar to each other.

In the baseline specification, a number of shocks drive the fluctuations in the potentials and in the gaps of the endogenous variables. Because these shocks are highly persistent, the conclusion that the potentials and the gaps have similar spectral features and are correlated follows as a corollary. What if some shocks, for example, investment and government spending disturbances, are absent? Would that make the two latent variables sufficiently different, given that markup and monetary policy disturbances only enter the gaps? The second panel of figure 3 shows that this is not the case: having a highly autocorrelated TFP shock is enough, given that the disturbances entering only in the gap have smaller persistence.

The third panel of figure 3 indicates that the spectral properties of the latent variables are not driven by specific bells and whistles present in a structural model. Eliminating real frictions (habit in consumption and investment adjustment costs) or wage rigidities does not change the relative properties of the output gap. Compared with the baseline specification, the gap displays more variance at the low and business cycle frequencies meaning that the rigidities typically used to fit wage inflation and the investment dynamics also decrease output gap variability. Note that eliminating these rigidities also produces a similar upward shift in the log spectral density of the potential at the same frequencies. Thus, the relative contribution to the variance of the process is unchanged, the spectral shape of the two latent components is similar and the average coherence high (0.78).

Making technology shocks permanent Assuming that all the disturbances are stationary may appear to unrealistic to many econometricians. Common wisdom suggests that certain real variables should display an upward trend. To ensure that this is the case, I alternatively consider a setup with a unit root in TFP. In this case, the levels of output (Y), consumption (C), investment (I) and real wages (W) will be on a balanced growth path. Hours (H), return to capital (RK), capacity utilization (CapU) will instead be stationary. Panel 4 of Figure 3 shows that when transitory shocks have large persistence, the spectrum of the output gap and of transitory output are alike and have qualitatively similar distribution of the variance by frequency (compare "ygap" with "ytra1"). It is also worth examining what happens to the spectral shape of the latent variables when, as in Aguiar and Gopinath [2007], "the trend generates the cycle", i.e., the permanent TFP disturbance accounts for a large portion of the data variability at business cycle frequencies. To produce this pattern, I decrease by 3/4 the persistence of the transitory disturbances. The relative importance of transitory low frequency fluctuations decreases (compare "ytra1" and "ytra2"). On the other hand, the importance of the permanent component at business cycle frequencies is magnified. Still, both transitory and permanent components have more variability in the low than in the business cycle frequencies. Hence, also in this case, the basic properties we emphasize hold true.

Gaps and potentials in other models Although the SW model is popular, it leaves out features that may be important to explain the data. For example, it does not account for search and matching frictions, it does not take into consideration the relationship between the real and the financial side of the economy, nor has built-in endogenous mechanisms making transitory shocks important in the medium-long run. Because there are many alternative models one could consider, for illustration, I show what happens to the properties of the output gap when NK models with financial frictions or R&D are used [1]. In models with financial frictions, I define potentials in the same way I have done without them: they characterize the economy without nominal frictions and with markup and the monetary disturbances set to zero. Thus, financial frictions (and risk shocks if present) affect both the potentials and the gaps.

Panel 5 of figure 3 compares the spectral properties of gaps in the baseline specification (SW); in the baseline specification with financial frictions (SWFF), see Del Negro, Giannoni, and Schorfheide [2015]; and in the specification with risky contracts (CMR), see Christiano, Motto, and Rostagno [2014]. Although in the CMR model the gap displays uniformly lower log spectral density, since transitory disturbances have considerably smaller estimated variability than in the baseline SW model, the presence of financial frictions does not change the spectral features of the gaps. If anything, the relative importance of low frequency components grows larger.

What happens to the spectral density of the gaps when the model features endogenous mechanisms increasing the persistence of exogenous shocks? Any mechanism boosting the persistence and/or creating larger amounts of "medium term" volatility, will generally an increase the proportion of the total variance due to the gaps and their importance relative to potentials in the low frequencies. Hence, adding such a mechanism is likely to make the features we emphasize more extreme, in the sense that gaps may have more variability in the low frequencies, and potentials more variability at business cycle frequencies, but it will leave them correlated. Panel 6 of figure 3 plots the spectral properties of the output gaps in the baseline SW model and in a simple NK model with R&D, in the spirit of Gertler and Comin [2006], where shocks to government physical and R&D investments drive the fluctuations. In this model TFP endogenously grows because R&D investment affects labor productivity. Thus, iid productivity disturbances may generate persistent TFP movements. As expected, the spectral shape of the output gap is unchanged relative to the baseline specification

[1] I have also experimented with the Beaudry et al. [2020] model. Because the mechanism producing the peak in the spectral density at the low frequency in hours affect the two latent components, the same conclusions obtain: gaps display low frequency variations, gaps and potentials have similar spectral shapes, and are correlated.

and the proportion of output gap variance in the low frequency is still larger than the proportion at business cycle frequencies. Note that now, most of the gap variance is now concentrated in the very low frequencies [2].

Table 1: Relative variances

Output	All frequencies	Low frequencies	BC frequencies	Own variance low frequencies	Own variance BC frequencies
Gap (SW stationary)	0.58	0.66	0.50	0.16	0.07
Gap (SW 5 shocks)	0.64	0.65	0.91	0.10	0.04
Gap (SW, no real frictions)	0.74	0.78	0.94	0.19	0.24
Gap (SW, no wage rigidities)	0.41	0.60	0.85	0.10	0.13
Gap (SWFF)	1.06	0.91	1.01	0.22	0.05
Gap (CMR)	< 0.01	< 0.01	< 0.01	0.15	0.02
Gap (R&D)	0.005	0.001	0.007	0.009	0.003
Transitory (SW unit root)	0.01	0.80	0.64	0.18	0.10
Transitory (SW trendcycle)	< 0.01	0.61	0.36	0.22	0.44

Notes: SW stationary is the baseline SW model, SW 5 shocks is the baseline SW model without investment and government spending shocks, SW no real frictions is the baseline SW model without habit in consumption and investment adjustment costs, SW no wage rigidities is the baseline SW model without wage indexation and the wage Philips curve. SWFF is the SW model with financial frictions; CMR is the model of Christiano et al. [2014]; R&D is a NK model with research and development. SW unit root is SW model with permanent TFP disturbances; SW trendcycle is the SW model with unit root in TFP and low persistence of transitory disturbances. The first three columns report the fraction of the output variance due to the gap (transitory component) , on average, across all, low, or business cycle frequencies. The last two columns report the fraction of the variance of the output gap (transitory output) at low (32-64 quarters) and at business cycles (8-32 quarters) frequencies. Numbers may exceed 1 when the latent components are correlated.

Summary. Table 1 summarizes the properties of the output gap (transitory output) for the DGPs I consider. Equilibrium models featuring highly persistent TFP disturbances and the definition of potential I employ will make gaps and potentials correlated, generate latent components with similar spectral distribution of the variance, and produce large amounts of low frequency variability in the gaps and of business cycle frequency variability in the potentials. The persistence of the disturbances is also key to make transitory and permanent components share similar spectral features.

[2] All the DGPs considered have disturbances with constant moments. When there is, e.g. , stochastic volatility, none of the features I have emphasized is altered, provided that a third order approximation to the solution is used to simulate data. Stochastic volatility increases the portion of the variance of gaps and potentials in the low frequencies, but it does not alter the relative importance of the two latent components at low and business cycle frequencies.

The table also suggests that the theoretical gain function needed to extract the output gap is either roughly uniform across frequencies or has a hump at business cycle frequencies. (see first three columns of table 1). As discussed next, none of the existing estimation procedures generates estimated gain functions with these features. When a unit root is present, the permanent component explains the majority of the data fluctuations as the height of its spectral density in the very low frequencies is of an order of magnitude larger. Still even in this case, the portion of the variance in the low frequencies due to transitory fluctuations, and the portion of the variance at business cycle frequencies due to permanent fluctuations are non-negligible.

4 Latent variable estimation procedures

There are numerous procedures one could employ to estimate two latent components from one observable time series (typically named "trend" and "cycle"). I focus on the most commonly used in the macroeconometric literature. I do not consider production function methods, because estimates of the long run values of the inputs need demographics, participation rates, and other slow moving variables that can not be produced with my DGPs. The procedures differ in many dimensions. Some are statistical and others have economic justification; some are univariate and others multivariate; some require parameter estimation and others do not.

Univariate approaches. The first approach is the oldest and maintains that the trend is deterministic and uncorrelated with the cycle. Thus, the latter can obtained as the residual of a regression on a polynomial trend. I use a quadratic polynomial and run the regressions variable by variable. The results obtained with this approach are denoted in the tables by the acronym POLY

The second approach is the Hodrick and Prescott filter. Here the trend is assumed to be stochastic but smooth and uncorrelated with the cycle. The latter is the difference between the level of the series and the Hodrick and Prescott trend, which is obtained via the ridge estimator:

$$\tilde{y} = (H'H + \lambda Q'Q)^{-1}H'y \tag{13}$$

where λ is a smoothing parameter, $y = (y_1, \ldots y_t)$ the observable series, $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_t, \tilde{y}_{t+1}, \tilde{y}_{t+2})$, the trend, $H = (I_{t\times t}, 0_{t\times 2})$ and $Q_{t\times(t+2)} = \begin{bmatrix} 1 & -2 & 1 & 0 & \ldots & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \ldots & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & \ldots & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \ldots & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \ldots & \ldots & 1 & -2 & 1 \end{bmatrix}$. I set $\lambda = 1600$

and denote the results with the acronym HP. For sensitivity, I also consider $\lambda = 51200$ (acronym: HPa), a value close to the BIS recommendations, see Borio [2012].

The third approach assumes that the trend is stochastic, displays at least one unit root, and it is uncorrelated with the cycle. I consider two separate sub-cases: one where the cycle is obtained by short differencing (one period); and one where it is obtained by a long differencing (24 periods). I denote the results with the acronyms FOD and LD, respectively. In sensitivity analysis, I also consider 4 and 16 periods differencing operators (acronyms: FODa, LDa)

The fourth approach permits the trend to be stochastic and to display both stationary or non-stationary features, but assumes that its variability is entirely located in the low frequencies of the spectrum. To extract cycles with 8-32 quarters periodicity, I use the band pass filter implementation of Christiano and Fitzgerald [2003], which employs asymmetric and time-dependent weights. I denote the results with the acronym BP. For sensitivity, I also examine the trigonometric version of the filter, see Corbae, Ouliaris, and Phillips [2002] (acronym: Trigo), the stationary, truncated symmetric version of Baxter and King [1999] (acronym: BK), and a version of the baseline filter designed to extract cycles with 8-64 quarters periodicity (acronym: BPa).

As alternative, I have considered a wavelet filtering approach, recently popularized by Lubik et al. [2019]. Wavelet are one-sided MA filters, where the length of the MA polynomial depends on the cycles being extracted. For example, to obtain cycles with a 8-32 quarters periodicity, a MA(16) is used but to extract cycles with a 32-64 quarters periodicity, a MA(32) is employed. Wavelets have some intuitive advantages over band pass filters, as they work in time domain and the number of MA terms is finite. I denote the results obtained with this approach with the acronym Wa.

The fifth approach is based on local projections and follows Hamilton [2018]. Here the trend is the medium term predictable component of a variable and the cycle is assumed to be uncorrelated with the trend. The latter is obtained as the residual of a regression of each variable at $t + m$ on current and up to d lags of the variable. I set m=8, d=4, and report results under the acronym Ham. In the sensitivity analysis, I consider the alternative of m=12 and d=2 (acronym: Hama)

The sixth approach uses a state space formulation of the latent variable problem. It assumes that the trend is a random walk with drift; that the cycle is an AR(2) process, and allows the innovations in the trend and the cycle to be correlated. No error is present in the measurement equation. Using a flat prior on the parameters, I compute posterior distributions using a MCMC approach, as in Grant and Chan [2017]. The reported properties are computed averaging trends and cycles estimates over retained draws. I denote the results with the acronym UC. In the sensitivity analysis I consider a bivariate filter with output and capacity utilization as observables (acronym UCbiv).

Multivariate approaches. Given the general equilibrium nature of the DGP, univariate approaches are inefficient as they disregard, for example, the fact that cyclical components have similar features (since they are driven by the same disturbances) or the presence of balance growth when TFP has a unit root. The next two procedures account for these possibilities.

The first, based on Beveridge and Nelson [1981]' decomposition, defines the trend as the predictable long run component of a vector of variables. Cycles are the difference between the vector of observable variables and the estimated trends. By construction, trends and cycles are perfectly correlated, as they are driven by the the same reduced form innovations of a vector autoregression on lags of the relevant variables. I run the decomposition unrestricted, that is, without the signal-to-noise prior of Kamber, Morley, and Wong [2018], because the DGP is already a low order VAR.

The second procedure follows Blanchard and Quah [1989] and still uses a vector autoregression to compute the innovations. However, the decomposition uses identified disturbances to separate the two latent components: the trend is driven by supply disturbances and the cycle by both supply and demand disturbances. In the implementation I use, the vector autoregression includes output growth and hours for both approaches. I denote the results with the acronyms BN and BQ. In the sensitivity analysis, I consider a trivariate VARs with output growth, consumption to output, and investment to output ratios (acronyms: BNa, BQa).

5 Summary statistics

It is difficult to select a single statistic conveying reliable information about the performance of different approaches; and different researchers may be comfortable with different ones. Here I summarize the outcomes of the experiments with a number of indicators and aggregate the outcomes using a simple counting measure. The indicators are computed averaging results using 100 data replications to wash out simulation uncertainty.

For each approach, I first compute the root mean square error (RMSE), calculated as the difference between the true gaps (transitory components) and the estimates. Second, I report the contemporaneous correlation between the true gaps (transitory components) and the estimates, the first order autocorrelation and the variability of the estimates, benchmarking them with those of the true gaps (transitory components). I do not report contemporaneus cross-correlations as the results for this statistic can be inferred from autocorrelation and variabilties.

I compute these four statistics for 9 series the model generates: output, consumption, investment, return to capital (real rate), hours, real wages, capacity utilization, the nominal rate and inflation (the latter serie has no potential). I also extract a factor from the data, apply the filtering procedures, and compare RMSE, variability, auto and contemporaneous correlations of the filtered series to the those of the factor, computed using the true gaps (transitory components).

Third, I compute turning points in the filtered data and compare their number, the average duration and amplitude of expansions and recessions with those of the true gaps (transitory components). Because in the baseline specification, the model is solved linearly around the steady state, durations and amplitudes are roughly symmetric across business cycle phases in the simulated gaps (transitory components). This may not necessarily be the case in the estimates.

Policymakers are interested in gaps because they may help to understand the current state of the economy or to predict variables of interest, such as inflation via a Phillips curve, or employment with a Okun law. For this reason, I also compute two additional set of statistics. The first measures the RMSE in real time, focusing attention on the last 12 periods of each sample; the second compares the variance of the prediction error in the regressions implied by the true output gap (transitory output) and those implied by the estimates. Letting y^i_{t-j} denote either the true output gap (transitory output) or the estimated one at $t-j$ and letting m=1,4, the predictive regressions take the form:

$$\pi_{t+m} = \alpha_0 + \sum_{j=1}^{2} \alpha_j \pi_{t-j} + \sum_{j=1}^{3} \beta_j y^i_{t-j} + e_{t+m} \tag{14}$$

$$H_{t+m} = \alpha_0 + \sum_{j=1}^{2} \alpha_j H_{t-j} + \sum_{j=1}^{3} \beta_j y^i_{t-j} + v_{t+m} \tag{15}$$

where π_t is inflation and H_t is hours worked.

6 The results

Table 2 summarizes the results for the SW baseline DGP with stationary shocks (top panel) and with a unit root in TFP (bottom panel). It reports, for each average statistics, the number of times across variables a procedure ranks first. The counting measure assigns a one to the best procedure (0.5 if there is a tie) and zero to the others. To be clear, in the RMSE row a 5 in the Poly column means that the RMSE of that approach is the smallest for 5 variables. Totals are computed equally weighting all statistics. Tables A5-A11 in the on-line appendix give the details for each statistics [3].

[3] For each variable (factor) and for each procedure, table A5 reports the average RMSEs; table A6 the average real time RMSEs; table A7 the average contemporaneous correlation, table A8 the average AR1 coefficient; and table A9 the average variability; table A10 the average number of turning points, the average duration and amplitude of recessions and expansions for output and the factor; and table A11 the average difference between the true variance of the prediction error of Phillips curve and Okun law regressions and the estimated one.

Table 2: Summary results across variables, DGP: Stationary SW, T=750

Statistic	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
						Gap					
RMSE	5	3						1	0.5	0.5	8
Correlation	9								0.5	0.5	8
AR1	4			3		3					6
Variability	4			2		3	1				
TP	1.5	5	2	1.5							3
RT-RMSE		1				3	2	3	0.5	0.5	8
PC	2										2
OL				1			1				
Total	25.5	9	2	8.5	0	9	4	4	1.5	1.5	35
						Transitory					
RMSE			9					1			
Correlation											
AR1	4					5.5		0.5			1
Variability	3			6			1				1
TP	4	4		2							3
RT-RMSE			4					6			
PC				1						1	
OL				2							
Total	11	4	13	11	0	5.5	1	7.5	0	1	5

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24 order differencing, UC is unobservable component filtering, BP is band pass filtering, Wa is wavelet filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. RMSE is the root mean square error, Correlation the contemporaneous correlation with the true series, AR1 the first autoregressive coefficient, Variability the variability of the series, TP the number of turning points, the duration and the amplitude of expansions and recessions, RT-RMSE is the real time root MSE, PC is Philips curve predictions, OL the Okun law predictions. In each row the ranking is over 9 series (two series for BN and BQ) and one factor. For TP where the ranking is for output and the factor. Numbers are computed summing the top ranks, equally weighting all variables; when there are ties, each get a value of 0.5.

The Polynomial approach is the least distorting when measuring gaps and it is superior as far RMSE, contemporaneous correlation, AR1, variability, and Phillips curve regressions are concerned. The next approaches in the ranking are appropriate for certain statistics (the HP filter for turning points (TP) detection; the Wavelet filter for TP detection, AR1 coefficient and variability) but seem much less suited than the Polynomial filter to globally capture the features of the gaps. Three other aspects of the top panel of table 2 are worth emphasizing. First, the long difference filter is superior to the Hamilton filter. Hamilton [2018] shows that his filter is close to a m-period difference filter. Hence, the horizon of the local projection $m = 8$ is probably too short, confirming the presence of important low frequency variability in the gaps. Second, the commonly used UC approach, is competitive only in terms of RMSE; for other statistics, it ranks in the middle of the pack. Finally, even discounting the fact that they concern only two of the variables, the performance of BN and BQ procedures is quite poor.

In absolute terms, biases in gap estimates are important. For example, for output, the average RMSE exceeds 5 percent, which is larger than the average RMSE produced by a random walk (3.80). Moreover, the real time RMSEs are 50 percent larger than the real time RMSEs produced by a forecast that uses $T - 12$ value for all successive 12 periods. Distortions are considerable also in terms of volatility and correlation measures; for instance, the largest contemporaneous correlation between estimated and true output gap is only 0.65.

The FOD filter ranks first when estimating the transitory components, closely followed by the Polynomial and LD filters, but its superiority stems entirely from RMSE measures. Relative to gaps extraction, there is a deterioration of the performance of HP and Hamilton filters and an improvement in the performance of the UC filter. The improvement for the latter comes from the real time RMSEs, and involves primarily output, consumption and investment. On the other hand, the performance of BN and BQ is still deficient. Given the popularity in the literature of the BQ procedure, the next section takes a close look at why this is the case. The band pass filter is poor for all statistics and all series. In comparison, the Wavelet filter does a more reasonable job. Because both aim at capturing the fluctuations located in the central part of the spectrum, the wavelet filter is less distorting because it induces a smaller approximation error in the low frequencies.

Quantitatively, when extracting transitory components the distortions are larger. Interestingly, almost all estimates poorly correlate with the true transitory components. This explains why, in comparison with gaps measurement, RMSEs tend to be larger (for instance, for transitory output it is greater than 6 percent).

In general, many approaches underestimate the variability of both the gaps and the transitory components. The exceptions are the Polynomial, the BN and BQ procedures. For the former this is due to the nature of the squared gain function, as discussed in section 6. The excess variability produced by BN and BQ procedures is less expected. Section 6 also discusses why this is the case.

To conclude, somewhat unexpectedly, the crudest and the oldest procedure, the Polynomial filtering, is the least damaging when characterizing gaps. A similarly crude differencing filter ranks well when extracting transitory components. Larger information sets or fancier econometrics do not help.

Table 3: Summary results for output, across statistics, different DGPs, T=750

DGP	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
Gap (SW stationary)	6	3	1	1			2	1			7
Gap (SWFF)	9	2	2			1					7
Gap (CMR)	3	2	1	2	2	3	1				
Gap (SW5)	5	4		1	1	1				2	10
Gap (No real frictions)	6	3	1	1		1	2				6
Gap (No wage rigidities)	5.5	5.5					3				2
Gap (R&D)	3	1	4		2	1	2			1	3
Transitory (SW unit root)	6.5	3	3.5	1							4
Transitory (SW trend cycle)	1.5	4.5	5				1		1	1	4
Total	45.5	28	17.5	6	5	7	11	1	1	4	43

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24 order differencing, UC is unobservable component differencing, BP is band pass filtering, Wa is wavelet filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. SW stationary is the stationary Smets and Wouter model; SWFF is the Smets and Wouter model with financial frictions; CMR is the Christiano, Motto, and Rostagno model; SW5 is the stationary Smets and Wouter model with 5 shocks; No real friction is the SW model without habit and investment adjustment costs; No wage frictions is the SW model with flexible wages and no wage indexation; R&D is a NK model with R&D. SW unit root is the Smets and Wouter model with a unit root in TFP; and SW trend cycle is the Smets and Wouter model with a unit root in TFP and low persistence for the transitory shocks. Numbers are computed summing the top ranks, equally weighting all statistics; when there are ties each get a value of 0.5.

Focusing on the output gap. Table 2 gives a bird-eye view of how different procedures perform in extracting gaps or transitory components, on average across variables for one DGP (the SW model). Table 3 summarizes the results on average across statistics for output gaps (transitory output) produced in the 9 DGPs considered in section 2. Here a 6 in the POLY column means that the procedure listed is best for 6 out of the 14 statistics for that DGP.

The Polynomial approach does well in replicating output gap dynamics when the DGP is a model of the Smets and Wouter variety and, when financial frictions are present, it clearly dominates all others. Given that the Polynomial approach has been discredited in the literature for leaving near non-stationary dynamics in filtered series, these results deserve further discussion, see section 6.

The HP filter ranks well, both when the disturbances are highly persistent and when they are not. The performance of the HP filter is striking, in light of the criticisms raised in the literature; see recently Hamilton [2018]. Finally, the performance of the UC approach is generally poor.

Thus, the Polynomial procedure seems the most robust for output gap extraction, and it maintains his relative superiority across DGPs. The HP filter is also relatively robust and ranks on top for a number of statistics, when both gaps and transitory components are generated by differetn models.

Sensitivity. I have repeated the exercise with the baseline DGP varying the free parameters of the procedures, adjusting certain filters to capture low frequency variability, or adding information to multivariate approaches.

The absolute performance of HP and BP filters can be improved by choosing a higher λ or a lower limit for the frequency band (see tables B.1-B.7 in the on-line appendix). However, the Polynomial procedure still ranks first. The alternative Hamilton filter produces estimated cycles whose frequency distribution is more in line with those of the true gaps (transitory components) but this does not change its relative position in the ranks. Quantitatively speaking, the gains obtained by optimizing the parameters of the filters are small.

Perhaps more interesting are the results when the sample size is smaller. Because certain procedures require parameter estimation while others do not, estimation uncertainty may affect the ranking. Table 4 summarizes the results for T=150 and the baseline SW DGP (tables B.8-B.14 in the on-line appendix provide the details). For gaps, the Polynomial approach is still superior but its performance worsens. Instead, procedures which do not require parameter estimation, such as HP or LD, improve in the overall ranking. Still, the snapshot of tables 2 is roughly maintained.

The conclusions are somewhat affected instead when measuring the transitory components. The FOD filter, which came on top in table 2, now looses its superiority and the LD filter becomes the least distorting, with the UC approach lagging third. The LD filter dominates when measuring the AR1 and the volatility of the transitory component, while the FOD filter is still superior in terms of RMSE measures. For the two regressions policymakers care about, the sample size makes little difference, and the LD filter is most appropriate to construct regressors in the predictions equations.

Table 4: Summary results across variables, DGP: Stationary SW, T=150

Statistic	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Gaps										
RMSE	4	4						1	0.5	0.5	7
Correlation	7						2		0.5	0.5	10
AR1	3			4.5		2.5					4
Variability	5	1		3					0.5	0.5	2
TP	0.5	3.5	2	1		2	1				2
RT-RMSE		3				4	2		0.5	0.5	8
PC								2			2
OL				2							
Total	19.5	11.5	2	10.5	0	8.5	6	4	1.5	1.5	33
	Transitory										
RMSE			4		1			5			
Correlation											
AR1	0.5			7		1.5			0.5	0.5	3
Variability	3			7							7
TP	2	2		3	1.5			1.5			
RT-RMSE			5		3			2			
PC				2							
OL			1	0		1					2
Total	5.5	2	10	19	5.5	2.5	0	8.5	0.5	0.5	13

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24 order differencing, UC is unobservable component filtering, BP is band pass filtering, Wa is wavelet filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. RMSE is the root mean square error, Correlation the contemporanoeus correlation with the true series, AR1 the first autoregressive coefficient, Variability the variability of the series, TP the number of turning points, the duration and the amplitude of expansions and recessions, RT-RMSE is the real time root MSE, PC is Phillips curve predictions, OL is Okun law predictions. In each row the ranking is over 9 series and one factor, except for TP where the ranking is for output and the factor. Numbers are computed summing the top ranks, equally weighting all variables; when there are ties each get a value of 0.5.

I have also conducted an experiment where the benchmark SW model is simulated using a second order solution. In this case, procedures which assume a linear, parametric structure are penalized relative to procedures which non-parametrically split the data into two latent variables. The magnitude of the distortions is, on the whole, larger; but the ranking presented in tables 2 and 3 is unaffected.

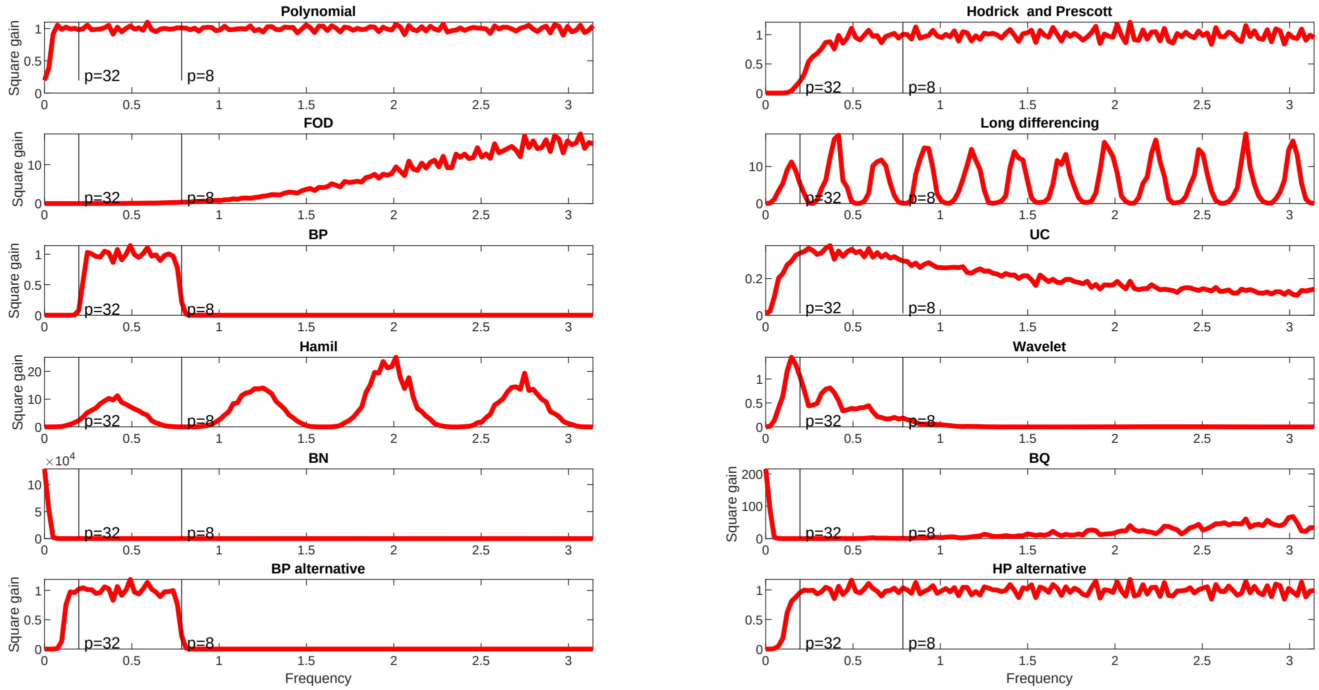

Figure 4: Estimated squared gain functions for output: selected filters.

7 The estimated gain functions

To understand the outcomes reported in tables 2 and 3, I employ the estimated squared gain function which measures, frequency by frequency, how the variance of the output of a procedure relates to the variance of the input. As mentioned, the theoretical squared gain functions are different from zero, bounded above and roughly constant at least across low and business cycle frequencies. Thus, deviations of the estimated from the theoretical squared gain functions help us to understand the distortions of each procedure produces. Figure 4 plots the estimated squared gain functions for 12 procedures (Polynomial, HP, BP, Hamilton, FOD, LD, BN, BQ, UC, Wave, HPa and BPa) obtained with one realization of the output series, when all disturbances are stationary [4].

The squared gain of the polynomial approach is zero at the zero frequency and one everywhere else. Thus, the procedure removes very long run variability and leaves the rest of the spectrum of the input series untouched. This means that the estimated gap has roughly the same volatility features as the original series and the frequency distribution of the variance is similar.

The FOD approach has a squared gain with the familiar shape: it attributes all the variability of the original series in the low and business cycle frequencies to the potential while the gap captures,

[4]There is no loss of generality focusing on one realization because the spectral properties of output are similar across replications.

primarily, very high frequency variability. The squared gain of the HP filter displays its well-known high-pass features. The filter when applied to the estimation of the gap eliminates low frequency variability, keeps the high frequency variability unchanged and, at business cycle frequencies, smoothly eliminates power, when moving from cycles of 8 to 32 quarters. The band pass filter, on the other hand, knocks out low and high frequency variability and passes the business cycle frequencies almost unchanged. Because the sample is finite, there is some compression, in the sense that at some business cycle frequencies, the squared gain is less than one.

Perhaps more interesting is the squared gain of the Hamilton filter. Since Hamilton [2018] only briefly discusses it (see footnote 18), it worth to explicitly spell out its features. Because the projection equation uses y_{t+m} as dependent variable, the gain function is zero at $m/2$ separate frequencies. Between these frequencies, it is bell-shaped and, at the vertex, its height exceeds 10. Hence, while at certain frequencies the input variability is eliminated, at others, it is multiplied by a factor of 10 or more. In other words, it emphasizes frequencies of the spectrum not necessarily connected with meaningful cycles creating excess variability at 'uninteresting' frequencies. Hamilton [2018] mentions that the cycles his approach produces are similar to those of a LD filter. Figure 4 confirms the statement: the LD squared gain has, qualitatively, the same features as the Hamilton squared gain but, because I take a 24-quarter difference, the LD filter also emphasizes low frequency variability.

The BN and the BQ filters have qualitatively similar squared gain functions, despite different identification assumptions. Because the squared gain is large at very low frequencies, estimated cycles display strong low frequency variations, which are absent in the theory-based gaps, and considerable persistence.

The UC and the Wavelet filters have gains functions which differ from the others. In particular, the squared gain of the UC filter is less than one at all frequencies and has a vertex of around 0.3 in the low frequencies. Thus, the approach seems to recognize that gaps have power in the low frequencies and display variability at all frequencies. However, the shape of the function depends on the estimated parameters. As figure 5 highlights, the estimated squared gain for consumption differs and magnifies the variance at business cycle frequencies. The Wavelet filter also displays a squared gain function with no zeros except at the zero frequency. However, it increases the variability of the original series at low frequencies.

Figure 4 also plots the gain functions of the alternative HP and BP filters (HPa, BPa), both of which are designed to capture cycles with 8 to 64 quarters periodicity. Clearly, by changing the

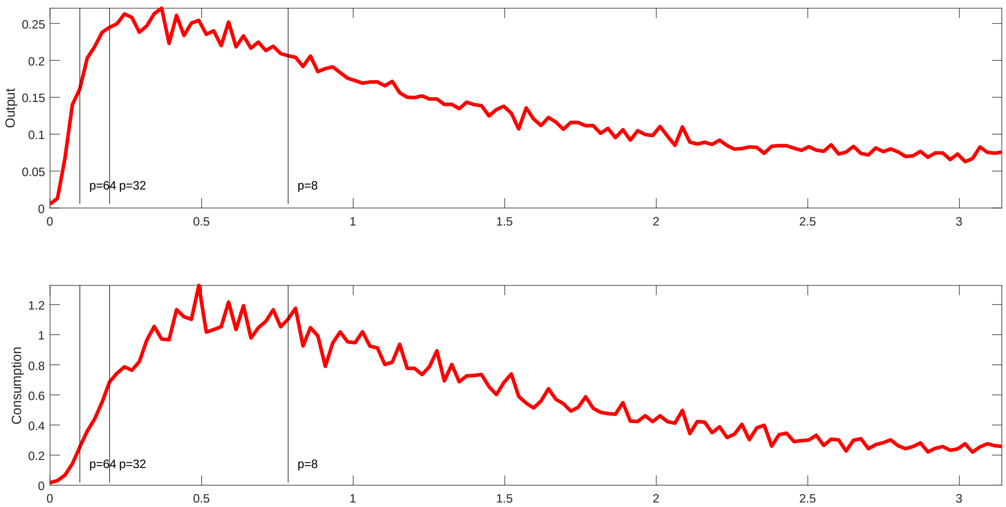

Figure 5: Estimated squared gain functions: UC filter

parameters, one can shifts a portion of the low frequency variations of the data to the gap estimate. Still, these changes do not reduce the squared gain at business cycle or high frequencies. Thus, all the variability at these frequencies is still erroneously attributed to the gap and this accounts for the fact that the alternative and the original HP and BP filters have similar position in the rankings.

Although the quantitative details change when the TFP disturbance has a unit root, the shapes of the estimated squared gain functions are roughly similar (see figure C.2 in the on-line appendix). Hence, the characteristics of the filters more than the properties of the shocks matter in determining the properties of the estimates of the latent variables.

Discussion. Figure 4 helps to understand why the Polynomial approach is the least distorting for gaps. Since the estimated squared gain is one at almost all frequencies, there is a general overestimation of the true gaps variance but the persistence and variance share by frequency are roughly matched. Thus, the ups and downs of theoretical and estimated gaps are similar as far as timing and duration, see figure 6. A similar argument holds when considering transitory components. However, because they do not have a constant variance share by frequency, distortions are larger.

Because in the DGPs, potentials explain an important portion of the variance at business cycle frequencies and gaps explain a large fraction of the variance at low frequencies, the HP (BP) filter, which attribute most (all) of the low frequency variations to the trend and most (all) the business cycle variations to the cycle, distorts the frequency distribution of the variance of the latent variables. Hence, the persistence of the gaps is underestimated and the filtered series have ups and downs that do not generally match those of theory-based gaps in terms of timing, duration, and amplitude. When

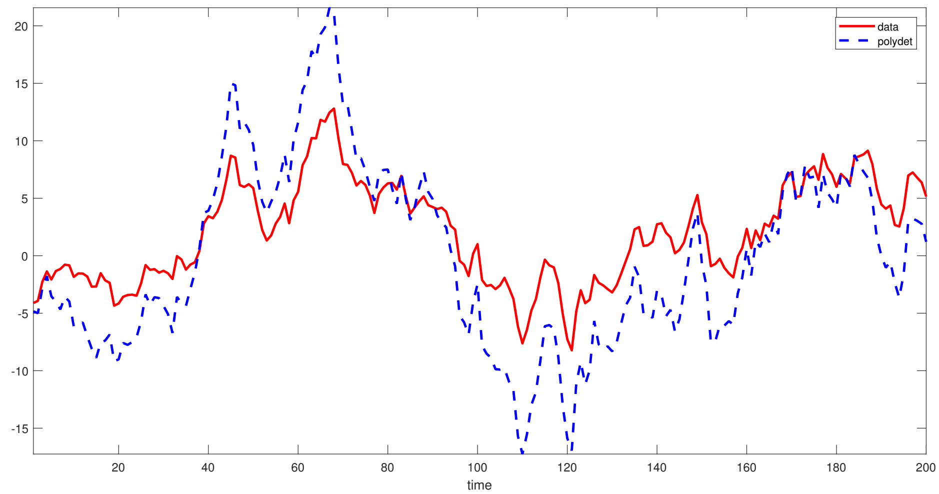

Figure 6: Output gap and polynomial estimate

TFP has a unit root, distortions remain because both filters fail to recognize that the permanent component has significant variations at business cycle frequencies.

Since the HP filter is the leading procedure to extract cycles in international institutions (e.g. BIS or OECD), further discussion is warranted. The standard HP filter uses a smoothing parameter of $\lambda = 1600$, which typically is interpreted as indicating that the standard deviation of the cycle is 40 times the standard deviation of the second difference of the trend. Hamilton [2018] criticizes this choice of λ suggesting that estimates of the ratio obtained in state space models that approximate a one-sided HP filter are much smaller. When I compute the range of theoretical λ values, obtained by taking the variability of the gaps (transitory components) to the second difference of the potential (permanent components) across simulated series, I find that, indeed, they are much smaller than 1600 and in the range [3,24]. Still, gaps display quite a lot of low frequency variations and only when $\lambda = 51200$ these variations become part of the estimated gap. Note that $\lambda = 51200$ is close to the value typically used to extract financial cycles and that, with such a λ, the absolute performance of HP filter improves. Hence, when the two latent components have similar spectral properties, are correlated, and the gaps are not iid, λ does not have the standard interpretation given in the literature; and the use of a state space approach to estimate it leads to misleading conclusions.

Hamilton [2018] suggests to employ local projections to extract cyclical fluctuations. My exercise shows that when the data is generated by standard NK equilibrium models, the filter ranks low, both for gaps and transitory components. A number of reasons may explain this unexpected outcome.

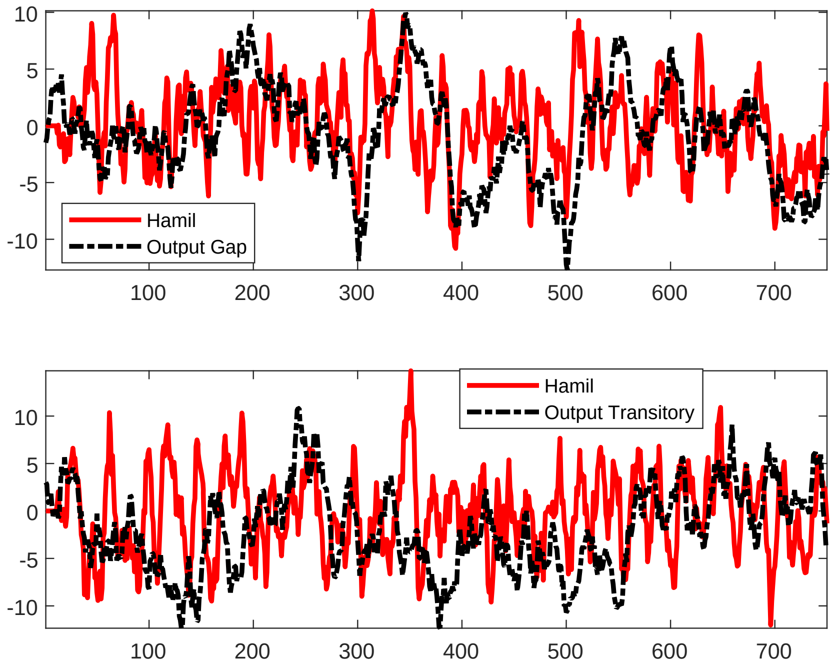

Figure 7: True output gap, true transitory output and Hamilton estimates.

First, gaps and potentials are correlated while the projection equation assumes that they are not. Second, even though the approach does not impose unit roots, unit roots are in fact removed (these are the zeros in the estimated squared gain function). Because the simulated data displays, at most, one unit root, overdifferencing is an issue. While theoretically important, these two reasons do not explain the poor performance of the approach because the LD filter, which errs in exactly the same way, has a much better score. Thus, the projection filter is poor because it does not recognize the presence of low frequency gaps variations and attributes them to the estimated potential. In addition, it magnifies the importance of certain high frequency variations, which have little economic interpretation. Figure 7 illustrates the difficulties of the approach in replicating the dynamics of the output gap (transitory output) in one of the simulation of the SW model.

VAR-based decompositions perform well for hours but not for output and the distortions in terms of RMSE, correlation with the true component, or variability are large, both in absolute and in relative terms. This is true when all shocks are stationary and when TFP has a unit root. Under stationary, the estimated VAR is misspecified (output is overdifferenced), explaining, in part, the inferior performance of BN and BQ approaches for output gap extraction. When TFP has a unit root, the overdifferencing problem disappears but both approaches are still poor, because the estimated gaps display considerable low frequency variations which are absent from the true gaps.

Coibon et al. [2018] argued that a BQ decomposition can be used to measure the dynamics of potential output. My results do not support their choice: the estimated permanent component overstates the dynamics of potential fluctuations if a NK DSGE model generates the data. Figure 8, which plots true potential and permanent output components in one simulation together with the estimated BQ permanent components, clearly shows the problem. Estimates displays too much low frequency variations (there are very long drifts in the ups and downs) and too little medium-business cycle variations.

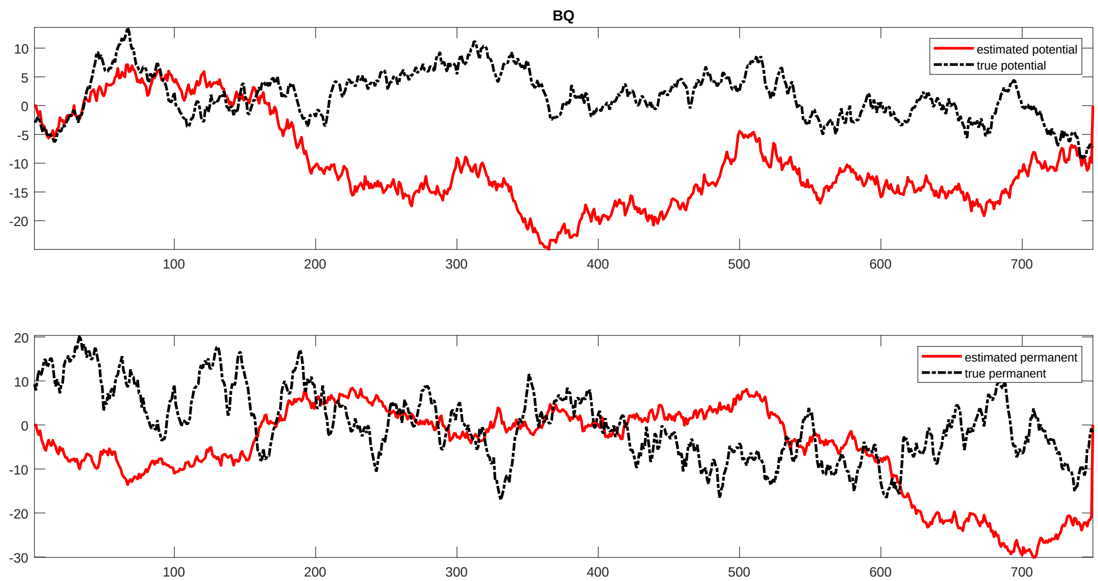

Figure 8: True potential, true permanent component and BQ estimates.

The BN and the BQ decompositions are still poor when TFP has a unit root because the estimated VAR is misspecified, as the deformation problems studied in Canova and Ferroni [2022] are important. In particular, the VAR model is bivariate (trivariate) while the DGP has seven disturbances; and not all the states of the DGP enter the empirical model. Thus, seven shocks are compressed into two (three) innovations; and some VAR innovations became serially correlated, even when a generous lag length is used. Deformation problems increase the persistence of the estimated shocks and alter the correlation structure between the estimated and actual shocks. Indeed, the estimated BQ transitory output is more persistent than the true transitory output; and the correlation between the TFP disturbances and the estimated supply shocks is low (0.43) because the latter captures a number of stationary demand disturbances.

When the sample is short (T=150), standard problems estimating long term quantities in small samples are added, see Erceg, Guerrieri, and Gust [2005]. Since a long lag length is needed to reduce states omission in the VAR, parameter estimation may be further compromised, making the credibility of latent estimates problematic.

Ramey and Zubairy [2018] used estimates of potential output to scale down the variables, prior to the computation government spending multipliers. While they use a polynomial approach and thus minimize the distortions when the DGP belongs the class of DSGE models I consider, it is generally unwise to employ latent variable estimates to compute multipliers since inference depends on the quality of the preliminary potential output estimates one has available.

8 A class of filters for DSGE-based gap extraction

Because existing procedures make assumptions which are inconsistent with the structural features of the class of NK models considered, inferential distortions are large. When the latent variables display roughly similar proportion of the variance at low and business cycle frequencies, standard filters generate biases. Different filters carve the spectrum differently but they attribute most of the low frequency variations to the trend, and almost all of the business cycle variations to the cycle, muting the persistence of the estimated cycles, altering the sequence of turning points, and the properties of amplitudes and durations.

Given these shortcomings, how should then one approach the measurement of gaps? One possibility is to take a structural model, estimate its parameters by conventional likelihood methods and, with mean or modal estimates and some initial conditions, generate model-based estimates of the latent quantities, as in Christiano et al. [2011],Justiniano et al. [2013] or Furlanetto et al. [2021]. Clearly, if the sample is short or the prior insufficiently tight, latent variable estimates are likely be very noisy. Furthermore, if the model is misspecified, estimates of the latent quantities become unreliable. While not much can be done about small samples, model misspecification can be taken care, in part, with the approach of Canova and Matthes [2021], which robustifies the measurement of latent quantities using the information provided by a variety of models.

Still, most researchers may prefer to be agnostic about the process generating the data. In addition, when selecting a filter, they may be only willing to take into account the frequency domain features emphasized in section 2, rather than other, more model specific characteristics. Can one design a procedure that takes into account the features of New Keynesian models and produce latent variable estimates which are uniformly superior to those of standard procedures?

Engineers extensively employ Butterworth filters in their work, because they are flexible, one-sided, have a convenient ARMA representation, and have uniform squared gain functions across the frequencies of interest. For my purpose, they are useful because one can design a filter in the class ensuring that the estimated cycle features significant low frequency variations and the estimated trend significant business cycle variations. Figure 9 shows the squared gain function of a number of Butterworth filters as function of three free parameters: the polynomial order (n) used to filter the data (reported are n=1,2,4); the cutoff point ω, where the squared gain declines (reported are 0.95π, 0.75π and 0.50π); and the scale G_0, determining the height of the squared gain (reported are $G_0 = 1, 0.4$). Clearly, depending on n, ω, G_0 a variety of shapes are possible. Note that the HP and the BP filters are special high pass and band pass Butterworth filters, when $G_0 = 1$.

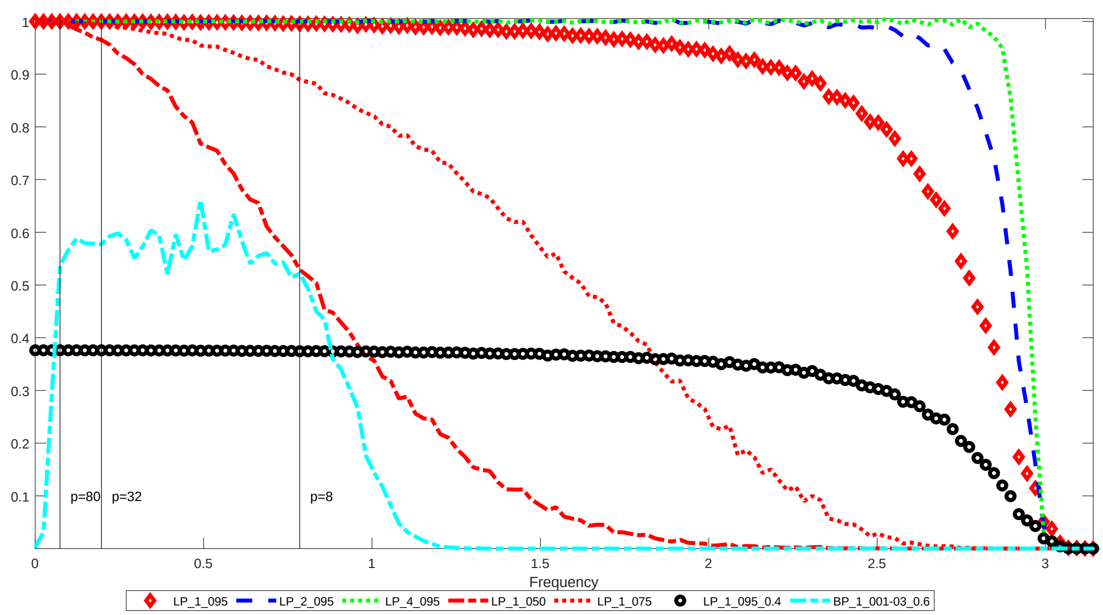

Figure 9: Squared gain functions, various Butterworth filters.

For gaps estimation, the relevant squared gain function is the black one with circles, since it has a uniform height of about 0.4 up to $\omega = 2/3\pi$. Thus, when this filter is used in practice, a portion of the low frequency variations of the input series goes to the estimated cycle and a portion of the business cycle frequency variations to the estimated trend.

Figure 10 plots one simulation of the gaps produced by the baseline SW model, together with the estimates obtained with a Butterworth filter, where for all variables, n=1,$\omega = 0.95\pi$, $G_0 = 0.4$. Figure C.3 in the on-line appendix presents the true and the estimated log spectral densities for gaps. Compared with figures 7 or 8, the match is clearly superior: the estimates capture fairly well the gap movements and replicate the distribution of variance of the true process at low and business cycle frequencies. This pattern is not specific to gaps. Figures C4-C5 in the on-line appendix show that the same holds true for 8 transitory series generated by one simulation of the model, when a Butterworth filter with n=1, $\omega = 0.004$ and $G_0 = 1$ is employed.

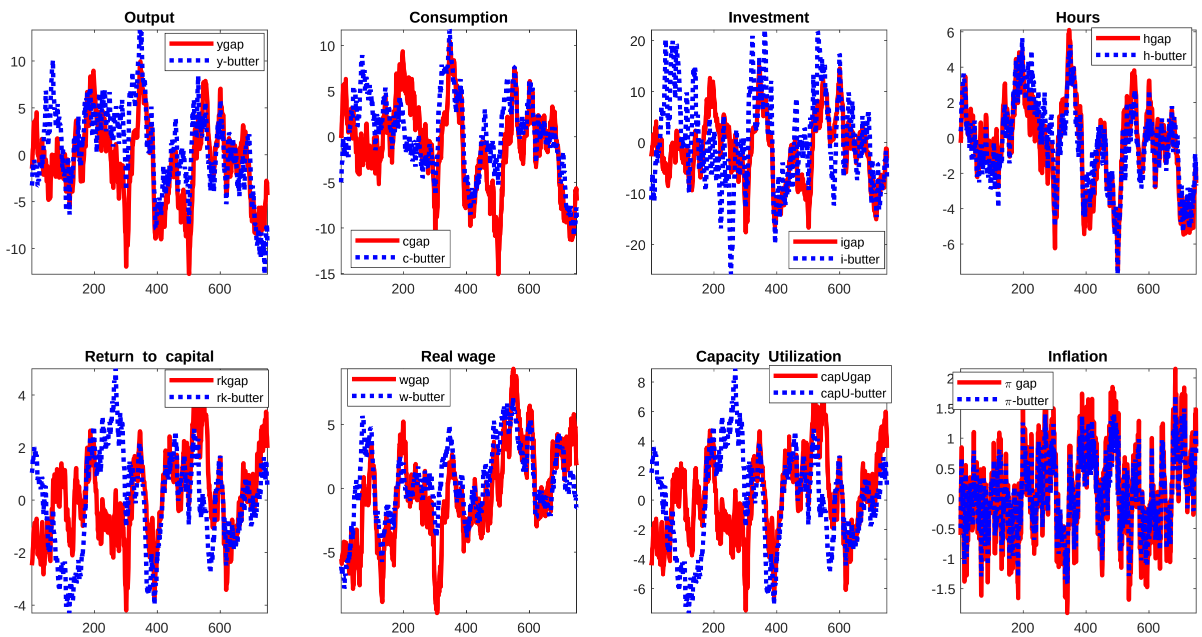

Figure 10: True gaps and Butterworth estimates.

The last column of tables 2-4 reports the number of times Butterworth estimates improves on the best performing method across series or statistics; the last column of tables A5-A11 gives the average values for each statistics across Monte Carlo simulations. The tables confirm the visual impression figure 10 provides: Butterworth filtered estimates are closer to the true gaps than the previously selected best approach in 37 out of 66 cases when T=750 and in 33 out of 66 cases when T=150 for the SW model. For output gap the estimates they are superior across DGPs in 35 cases. Thus, a careful design of a Butterworth filter may help to reduce the distortions.

The performance for transitory components is less impressive, as it is difficult to twist a Butterworth filter to produce a humped shaped squared gain function matching the patterns present in table 1. To do so, G_0 needs to be a function of ω. Still, even with this handicap, the approach does better than the previously selected best approach in over 1/3 of the cases.

Matching the labor share In New Keynesian models the labor share is often used in place of the output gap to derive e.g. the New Keynesian Philips curve, see e.g. Gali [2015], or testable empirical implications of the theory, see e.g. Gali and Gertler [1999]. Thus, one can use the labor share as a testing ground, to empirically find the procedure which produces output gap estimates "closer" to the labor share.

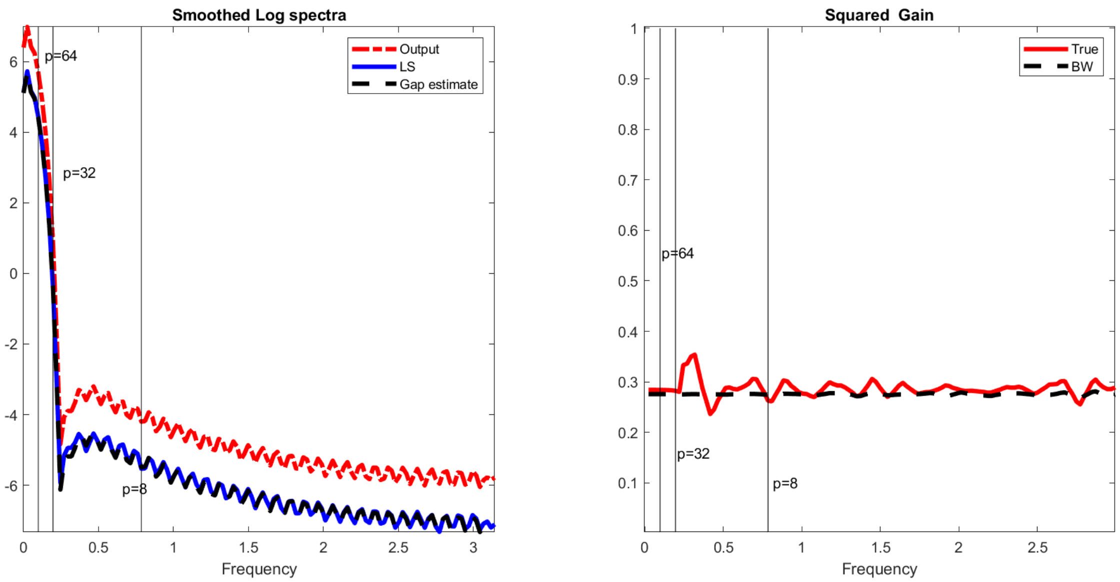

Figure 11: Log spectra and gain functions.

The first panel of figure 11, which plots the log spectrum of US output and of the US labor share, shows that indeed the two series are visually related at all frequencies and display the same shape. Interestingly, the proportion of the variance the two series displays at all frequencies is roughly similar (10 percent in the low frequencies, 0.06 percent at the business cycle frequencies) which implies that the ratio of the two spectral densities is roughly constant at about 0.3 at all frequencies (see red continuous line in second panel). Thus, if the labor share is a good estimate of the output gap, estimates of the latter should be persistent and must display low frequencies components that are more important than business cycle components. A Butterworth filter, whose gain function replicates the constant ratio of the two spectral densities, is depicted by the black dashed line in the second panel. By construction, such a filter produces a gap estimate whose spectral features match fairy well the spectral features of the labor share (compare black dashed line with blue solid line in the first panel). It is clear from previous discussion that other statistical approaches will produce a log spectral density for the estimated output gap which is different from the one of the labor share. Still one can ask how bad these approaches will do relative to the gap estimate produced by the Butterworth filter.

Table 5: Matching the US labor share, summary results

Statistic	RMSE(all)	RMSE(low)	RMSE(bc)	Corr(all)	Corr(low)	Corr(bc)	Persistence	Variance
POLY	0.1500	0.0217	0.0157	0.4423***	0.1900	0.5845**	0.0019	0.0300
HP	0.0561	0.0072	0.0127	0.4430***	0.2712	0.5824	-0.0007	-0.0052
FOD	0.0560	0.0064	0.0076	0.4364	0.1991	0.5796	-0.0002	-0.0088
LD	7.9553	0.2814	0.0427	0.1886	0.1558	0.2750	0.2639	1.5606
BP	0.0554	0.0064*	0.0139	0.2398	0.0988	0.5659	-0.0007	-0.0061
Wa	2.5607	0.0900	0.0220	0.1349	0.1521	0.3344	0.0843	0.4918
Ham	0.1588	0.0236	0.0377	0.3746	0.2614	0.5307	0.0000*	0.0266
UC	0.1002	0.0093	0.0074**	0.3530	0.1521	0.5621	0.0017	0.0054
BN	56.0487	2.2333	0.4567	0.2214	0.4534*	0.3260	1.7990	11.0246
BQ	2.4840	0.1221	0.0459	0.3097	0.3506	0.4810	0.0811	0.5715
BW	0.0631*	0.0095	0.0074**	0.4405***	0.1900	0.5845**	0.0000*	-0.0023*

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24 order differencing, UC is unobservable component filtering, BP is band pass filtering, Wa is wavelet filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. RMSE(all) is the root mean square error between the gap estimate and the labor share at all frequencies and RMSE(low) and RMSE(bc) the root mean square errors at low and business cycle frequencies; Corr(all) is the average coherence (correlation) between gap estimate and the labor share at all frequencies and corr(low) and corr(bc) the average coherences at low and business cycle frequencies. Persistence is the difference is the persistence of the estimated gap and of the labor share (measured by the height of the spectral density at the zero frequency) and variability is the differences in the variabilities of the estimated gap and of the labor share. A '*' indicates the procedure which matches best the statistic. A tie is indicated by multiple '*'.

Table 5 show a few summary statistics computable from the spectral densities of the estimated gaps and of the labor share. A Butterworth filter is superior to the other procedures in terms of overall RMSE, overall correlation with the labor share, persistence and variability. The UC filter is doing reasonably well overall, confirming its ability to assign low frequency components to the output gap mentioned earlier. The other approaches seem capable of matching some statistic at some frequencies, but overall they lag behind.

9 Conclusions

This paper shows that if the data has been generated by the class of NK models macro-economists and policymakers employ to interpret aggregate fluctuations and to compute out-of-sample predictions, the available toolkit of latent variable decompositions is inadequate. If one has to choose, the oldest (Polynomial) and the simplest (Differencing) approaches are the least distorting.

Distortions obtain because in a variety of NK models, gaps are correlated with potentials and have substantial low frequency variability; potentials (permanent components) have considerable business cycle variability, and the frequency distribution of the variance of gaps and potentials is roughly similar. In this situation, filters that vertically carve the spectrum by frequencies are unsuitable; as are methods that fail to recognize that low frequency variations in the gaps (transitory components) are equally or even more important than business cycle variations. The polynomial approach is the least distorting when measuring gaps because, away from the zero frequency, it leaves the frequency distribution of the data variance unchanged. Differencing works better for transitory fluctuations, because a portion of the low frequency variations enter in the estimate of the transitory component.

Given the unsatisfactory performance of popular procedures, care should be employed in using their output to evaluate the state of the economy, to forecast inflation or unemployment, or to provide policy recommendations. The warning is even more important when the sample is short, the approach requires parameter estimation, and real time estimates are needed.

One may argue that the class of structural models I consider is unrealistic or that one can build structural models which satisfy the assumptions of some commonly used in statistical approaches and proceed with business as usual. This argument disregards the fact that the features that make existing approaches inadequate obtain also in models with additional or different frictions and different organizing principles, as long as disturbances are persistent. Thus, unless one is willing to dismiss the majority of existing medium scale macroeconomic models, one must find a different way out of the conundrum.

I have suggested two potential solutions. One is to compute estimates of the gaps (transitory components), conditional on a model and the estimated parameters. While the exercise is straightforward, some researchers may feel uneasy with the approach, given that even complex models are far away from the DGP; and apparently innocent estimation choices may impair inference, see e.g. Canova [2014]. The alternative is to design estimation procedures which take into account the features the data is likely to display. I have described a class of Butterworth filters that can be rigged to produce latent components that uniformly dominate existing ones for gaps estimation; and are competitive for transitory components estimation.

Two additional implications of my study are worth mentioning. While it is standard to think of economic and financial cycles as having different length, in the sense that the largest share of the variances is located at different frequencies of the spectrum, see e.g. Borio [2012], the fact that models with or without financial constraints have similar features and that gaps and transitory components have considerable low frequency power in both types of models suggests that, perhaps, it is the insistence of macro-economists on focusing attention on cycles of 8 to 32 quarters that has given a misleading impression that the two cycles are different. I will explore this issue in future work.

There has been an industry over the last 30 years collecting stylized business cycle facts, both to inform the construction of realistic models of aggregate fluctuations and to test them, see Angeletos et al. [2020] for a recent example. These exercises generally consider fluctuations with 8 to 32 quarters periodicity. Given the results I present, it is perhaps desirable to switch attention to a broader range of freqeuncies or, at the minimum, account for the fact that most of the data variance is not located at business cycles frequencies. This will help researchers to better understand what kind of models are consistent with the data.

References

Mark Aguiar and Gita Gopinath. Emerging market business cycles: The trend is the cycle. *Journal of Political Economy*, 115(1):69–102, 2007.

Marios Angeletos, Fabrice Collard, and Harris Dellas. Business cycle anatomy. *American Economic Review*, 110:3030–3070, 2020.

Marianne Baxter and Robert King. Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics*, 81:575–593, 1999.

Paul Beaudry, Dana Galizia, and Frank Portier. Putting the cycle back into business cycle analysis. *American Economic Review*, 110:1–47, 2020.

Stephen Beveridge and Charles R. Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, 7:151–174, 1981.

Olivier Jean Blanchard and Danny Quah. The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, 79:655–73, 1989.

Claudio Borio. The financial cycle and macroeconomics: what have we learned? Technical report, BIS working paper 2012, 2012.

Fabio Canova. Detrending and business cycle facts. *Journal of Monetary Economics*, 41:475–540, 1998.

Fabio Canova. Reference cycle and turning points: A sensitivity analysis to detrending and dating rules. *Economic Journal*, 109:126–150, 1999.

Fabio Canova. Bridging dsge models and the raw data. *Journal of Monetary Economics*, 67:1–15, 2014.

Fabio Canova and Filippo Ferroni. Mind the gap! stylized dynamic facts and structural models. *American Economic Journals: Macroeconomics*, 14(4):104–135, 2022.

Fabio Canova and Christian Matthes. Dealing with misspecifaction in structural macroeconometric models. *Quantitative Economics*, 12:313–350, 2021.

Larry Christiano, Matthias Trabant, and Karl Walentin. Dsge models for of monetary policy analysis. *Handbook of Monetary Economics, editors B. M. Friedman and M. Woodford, Elsevier B.V. North-Holland*, (3A):285–367, 2011.

Lawrence Christiano and Terry J. Fitzgerald. The band pass filter. *International Economic Review*, 44:435–465, 2003.

Lawrence Christiano, Roberto Motto, and Massimo Rostagno. Risk shocks. *American Economic Review*, 104:27–65, 2014.

Tim Cogley and James Nason. 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, 1995.

Olivier Coibon, Yuriy Gorodnichenko, and Mauricio Ulate. The cyclical sensitivity in estimates of potential output. *Brookings Papers on Economic Activity*, Fall, 2018.

Dean Corbae, Sam Ouliaris, and C.B. Phillips, Peter. Band spectral regression with trending data. *Econometrica*, 70(3):1067–1109, 2002.

Olga Croitorov, Atanas Hristov, Kieran McMorrow, Phillipp Pfeiffer, Werner Roeger, and Valerie Vandermuellen. The eu cam and the relationship between output gaps, cyclical indicators, and inflation. Technical report, European Commission, DG ECFIN report, 2019.

Stefano Crump, Richard K.and Eusepi, Marc Giannoni, and Aysegul Sahini. A unified approach to measuring u*. *Brookings papers on Economics Activity*, Spring:143–214, 2019.

Stefano Crump, Richard K.and Eusepi, Marc Giannoni, and Aysegul Sahini. The unemployment-inflation trade-off revisited: the phillips curve in covid times. Technical report, NBER working paper 29785, 2022.

Marco Del Negro, Mark Giannoni, and Frank Schorfheide. Inflation in the great recession and in new keynesian models. *American Economic Journals: Macroeconomics*, 7:168–196, 2015.

Marco Del Negro, Domenico Giannone, Marc Giannoni, and Andrea Tambalotti. *Journal of International Economics*, (118):248–262, 2019.

Christopher Erceg, Luca Guerrieri, and Christopher Gust. Can long run restrictions identify technology shocks? *Journal of the European Economic Association*, 3:1237–1278, 2005.

Francesco Furlanetto, Paolo Gelain, and Marzie Taheri-Sanjani. Output gap, monetary policy trade-offs, and financial frictions. *Review of Economic Dynamics*, 41:52–70, 2021.

Jordi Gali. *Monetary Policy, Inflation and the Business Cycle*. Princeton University Press, 2015.

Jordi Gali and Mark Gertler. Inflation dynamics: A structural economeytric investigation. *Journal of Monetary Economics*, 44(2):195–222, 1999.

Mark Gertler and Diego Comin. Medium term business cycles. *American Economic Review*, 96: 523–551, 2006.

Yuriy Gorodnichenko and Serena Ng. Estimation of dsge models when data are persistent. *Journal of Monetary Economics*, 57:325–340, 2010.

Angelia Grant and Joshua Chan. A bayesian model comparison for trend-cycle decompositions of output. *Journal of Money Credit and Banking*, 49:525–552, 2017.

James Hamilton. Why you should never use the hodrick and prescott filter. *Review of Economics and Statistics*, 100:831–843, 2018.

Jonathan Heathcote, Fabrizio Perri, and Gianluca Violante. The rise of us earning inequality: does the cyce drive the trend? *Review of Economic Dynamics*, 37:181–204, 2020.

Robert Hodrick. An exploration of trend-cycle decomposition methodologies in simulated data. Technical report, NBER working paper 26750, 2020.

Oscar Jorda', Sanjay Singh, and Alan Taylor. The long run effects of monetary policy. Technical report, San Francisco Federal Reserve., 2020.

Alejandro Justiniano, Giorgio Primiceri, and Andrea Tambalotti. Is there a trade-off between output and inflation stabilization. *American Economic Journals: Macroeconomics*, 5:1–31, 2013.

Gunes Kamber, James Morley, and Benjamin Wong. Intuitive and reliable estimates of the output gap from a beveridge and nelson filter. *Review of Economic and Statistics*, 100:550–566, 2018.

Mariano Kulish and Adrian Pagan. Turning points and oscillatory cycles. Technical report, University of Sydney, manuscript, 2019.

Thomas Laubach and John William. Measuring the natural rate redux. *Business Economics*, 51: 57–67, 2016.

Thomas Lubik, Christian Matthes, and Fabio Verona. Assessing us output fluctuations across time and frequencies. Technical report, Fed of Richmond working paper 19-6, 2019.

Valerie Ramey and Sarah Zubairy. Goverment spending multipliers in good times and in bad: Evidence from us historical data. *Journal of Political Economy*, 126:850–901, 2018.

Stephanie Schmitt-Grohe and martin Uribe. Exchange rates and uncovered interest differentials: the role of permanent monetary shocks. Technical report, Colombia University, 2019.

Frank Smets and Rafael Wouters. Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97:585–606, 2007.

Julia Thomas. Is lumpy investment relevant for business cycles. *Journal of Political Economy*, 110: 508–534, 2002.

Michael Woodford. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, 2003.

On-line Appendix

Table A5: Average RMSE, DGP: SW, T=750

Variable	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
						Gap					
Y	5.23	5.07	5.38	6.93	5.20	5.33	5.25	**4.81**	104.17	13.58	4.32(*)
C	**5.43**	5.69	5.94	6.76	5.81	5.79	5.72	5.69	NaN	NaN	4.43(*)
I	11.19	**9.26**	9.31	15.49	9.50	10.69	10.47	10.30	NaN	NaN	9.12(*)
H	1.67	2.50	2.77	3.04	2.60	2.40	2.16	2.26	**1.39**	**1.39**	0.99(*)
RK	2.22	**1.53**	1.68	2.47	1.60	1.76	1.76	1.68	NaN	NaN	2.09
W	**4.20**	4.87	5.02	5.63	4.95	4.93	4.89	4.90	NaN	NaN	3.17(*)
CapU	3.96	**2.73**	3.00	4.40	2.85	3.14	3.14	3.01	NaN	NaN	3.72
π	**0.41**	0.67	0.80	0.80	0.72	0.67	0.50	0.72	NaN	NaN	0.17(*)
R	**0.40**	0.68	0.87	0.87	0.78	0.72	0.50	0.89	NaN	NaN	0.20(*)
Factor	**1.57**	2.14	2.25	2.15	2.19	2.00	2.03	2.24	NaN	NaN	1.35(*)
						Transitory					
Y	8.64	6.03	**5.82**	8.56	5.98	6.88	7.06	6.08	156.00	17.58	9.49
C	9.62	6.44	**6.33**	8.95	6.40	7.32	7.18	6.55	NaN	NaN	7.93
I	17.53	14.15	**13.48**	19.63	14.08	15.88	16.38	15.44	NaN	NaN	16.84
H	4.00	3.22	**3.10**	4.28	3.18	3.50	3.69	3.12	22.83	22.83	3.77
RK	3.48	2.74	**2.70**	3.56	2.73	3.01	2.99	2.75	NaN	NaN	3.22
W	17.18	5.05	**5.00**	6.81	5.03	5.68	5.50	5.16	NaN	NaN	6.11
CapU	6.20	4.86	**4.81**	6.33	4.86	5.37	5.31	4.92	NaN	NaN	5.74
π	1.02	0.87	**0.83**	1.14	0.86	0.93	1	0.93	NaN	NaN	0.99
R	1.14	0.99	0.98	1.28	0.94	1.02	1.12	**0.94**	NaN	NaN	1.10
Factor	2.52	2.16	**2.13**	2.55	2.15	2.28	2.28	2.20	NaN	NaN	2.37

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. The RMSE is computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table A6: Average real time RMSE, DGP: SW, T=750

Variable	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Gap										
Y	6.62	5.00	5.28	7.34	4.98	5.25	5.31	**4.61**	20.96	14.90	4.55(*)
C	7.28	**5.74**	5.92	7.25	5.77	6.02	5.95	5.86	NaN	NaN	4.29(*)
I	12.76	9.11	9.03	17.36	9.03	11.28	10.90	**8.65**	NaN	NaN	8.63(*)
H	2.31	2.55	2.77	3.01	2.55	2.31	2.21	2.51	**1.43**	**1.43**	1.01(*)
RK	2.58	1.83	1.93	2.46	1.80	**1.76**	1.95	1.78	NaN	NaN	1.98
W	6.24	5.12	5.29	6.22	5.13	5.20	5.25	**4.59**	NaN	NaN	3.27(*)
CapU	4.60	3.26	3.44	4.39	3.22	**3.13**	3.48	3.17	NaN	NaN	3.53
π	0.64	0.67	0.80	0.88	0.70	0.70	**0.50**	0.77	NaN	NaN	0.17(*)
R	0.63	0.70	0.87	0.90	0.78	0.70	**0.50**	0.96	NaN	NaN	0.20(*)
Factor	2.33	2.24	2.30	2.32	2.25	**2.09**	2.13	2.31	NaN	NaN	1.33(*)
	Transitory										
Y	10.23	6.20	**6.12**	8.31	6.23	6.89	7.06	5.80	31.55	21.93	7.00
C	11.61	6.72	**6.69**	9.26	6.70	7.58	7.28	6.01	NaN	NaN	7.34
I	19.58	14.81	**14.33**	20.38	14.81	16.83	17.38	14.36	NaN	NaN	17.04
H	4.59	3.39	3.36	4.15	3.41	3.64	3.91	**3.02**	23.73	23.73	3.73
RK	4.17	3.09	3.06	3.75	3.09	3.31	3.36	**3.00**	N	NaN	3.23
W	8.85	4.66	**4.63**	6.71	4.67	5.42	5.08	5.10	NaN	NaN	5.25
capU	7.42	5.50	5.46	6.68	5.51	5.89	5.98	**5.28**	NaN	NaN	5.76
π	1.10	0.88	0.86	1.14	0.86	0.94	1.05	**0.85**	NaN	NaN	0.95
R	1.28	1.03	1.04	1.36	0.98	1.08	1.23	**0.91**	NaN	NaN	1.11
Factor	3.13	2.41	2.42	2.86	2.41	2.58	2.57	**2.28**	NaN	NaN	2.48

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. The RMSE is computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table A7: Average contemporaneous correlations, DGP: SW, T=750

Variable	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Gaps and Filtered variables										
Y	**0.65**	0.34	0.03	0.38	0.25	0.41	0.44	0.44	-0.33	0.25	0.71(*)
C	**0.65**	0.33	0.06	0.36	0.23	0.37	0.38	0.31	NaN	NaN	0.72(*)
I	**0.59**	0.27	0.04	0.34	0.19	0.37	0.44	0.43	NaN	NaN	0.65(*)
H	0.86	0.42	0.06	0.5	0.31	0.54	0.68	0.67	**0.93**	**0.93**	0.93(*)
RK	**0.53**	0.41	0.04	0.38	0.29	0.4	0.36	0.22	NaN	NaN	0.55(*)
W	**0.67**	0.26	0.01	0.31	0.16	0.32	0.3	0.06	NaN	NaN	0.77(*)
CapU	**0.53**	0.41	0.04	0.38	0.29	0.4	0.36	0.16	NaN	NaN	0.55(*)
π	**0.93**	0.61	0.17	0.57	0.48	0.61	0.84	0.57	NaN	NaN	1.00(*)
R	**0.94**	0.67	0.29	0.59	0.48	0.6	0.88	0.36	NaN	NaN	1.00(*)
Factor	**0.71**	0.31	0.04	0.43	0.2	0.45	0.41	0.16	NaN	NaN	0.79(*)
	Transitory and Filtered variables										
Y	0.01	0	0	0.01	0	0.01	-0.01	0.01	0.05	0	0.01
C	0.01	0	-0.01	-0.01	0	-0.02	-0.02	0.01	NaN	NaN	0
I	0.01	0	-0.01	0	0	0	-0.01	0	NaN	NaN	0
H	-0.01	0	0	0.01	0	0.01	0	0	-0.02	-0.02	0
RK	0.01	0	-0.01	-0.01	0	-0.01	-0.01	0	NaN	NaN	0.01
W	-0.02	0.01	-0.01	-0.02	0	-0.02	-0.02	0	NaN	NaN	-0.01
CapU	0.01	0	-0.01	-0.01	0	-0.01	-0.01	-0.01	NaN	NaN	0.01
π	0.03	0	0	0.01	0	0.01	0.01	0.01	NaN	NaN	0.01
R	0.01	0	0	0	0	0	0	0	NaN	NaN	0
Factor	0.03	0	0	0	0	0	0	0.01	NaN	NaN	0.01

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. Correlations are computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table A8: Average AR1 coefficient, DGP: SW, T=750

Variable	True	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
		Gaps and filtered variables										
Y	0.98	**0.98**	0.83	0.21	0.96	0.93	**0.98**	0.89	0.89	1	0.97	0.98(*)
C	0.98	**0.99**	0.83	0.23	**0.97**	0.93	**0.99**	0.89	0.88	NaN	NaN	0.98(*)
I	0.99	**0.98**	0.89	0.51	0.97	0.93	**0.98**	0.91	0.98	NaN	NaN	0.98(*)
H	0.98	0.95	0.78	0.11	0.93	0.93	**0.97**	0.89	0.9	0.96	0.96	0.96
RK	0.97	0.99	0.83	0.21	**0.97**	0.94	0.99	0.89	0.93	NaN	NaN	0.99
W	0.98	0.99	0.82	0.12	**0.98**	0.94	0.99	0.87	0.94	NaN	NaN	0.99
CapU	0.97	0.99	0.83	0.21	**0.97**	0.94	0.99	0.89	0.91	NaN	NaN	0.99
π	0.94	**0.93**	0.78	0.16	0.9	0.92	0.96	0.88	0.83	NaN	NaN	0.94(*)
R	0.83	**0.80**	0.54	-0.21	0.76	0.91	0.95	0.77	0.47	NaN	NaN	0.85(*)
Factor	0.98	**0.99**	0.8	0.14	0.97	0.93	**0.99**	0.89	0.93	NaN	NaN	0.99(*)
		Transitory and filtered variables										
Y	0.97	**0.98**	0.83	0.2	0.96	0.93	**0.98**	0.89	0.89	0.99	0.98	0.97(*)
C	0.99	**0.99**	0.84	0.26	0.98	0.93	**0.99**	0.89	0.87	NaN	NaN	0.98
I	0.98	0.97	0.89	0.5	0.97	0.93	**0.98**	0.91	**0.98**	NaN	NaN	0.97
H	0.96	**0.95**	0.78	0.1	0.93	0.93	**0.97**	0.89	0.9	0.96	0.96	0.94
RK	0.99	0.98	0.83	0.19	0.97	0.94	**0.99**	0.89	0.93	NaN	NaN	0.98
W	0.99	**0.99**	0.82	0.14	0.98	0.94	**0.99**	0.87	0.94	NaN	NaN	0.98
CapU	0.99	0.98	0.83	0.19	0.97	0.94	**0.99**	0.89	0.93	NaN	NaN	0.98
π	0.94	**0.93**	0.78	0.15	0.91	0.92	0.97	0.88	0.84	NaN	NaN	0.91
R	0.84	**0.81**	0.54	-0.21	0.76	0.91	0.95	0.76	0.41	NaN	NaN	0.77
Factor	0.99	0.98	0.8	0.12	0.96	0.93	**0.99**	0.88	0.93	NaN	NaN	0.97

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. The AR1 coefficient is computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table A9: Average variability, DGP=SW, T=750

Variable	True	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
		Gaps and filtered variables										
Y	23.62	38.81	4.07	1.56	42.88	3.45	15	**16.62**	6.83	4632.6	81.43	31.08
C	28.77	36.04	2.14	0.82	**29.97**	1.78	10.41	9.93	4.14	NaN	NaN	31.38
I	69.37	171.48	26.61	6.85	235.27	24.01	**84.51**	98.16	139.45	NaN	NaN	129.97
H	6.43	**6.82**	1.39	0.65	9.39	1.15	3.28	4.46	1.26	8.06	8.06	4.98
RK	2.48	6.3	0.4	0.15	6.05	0.33	**2.1**	1.88	0.3	NaN	NaN	5.21
W	20.85	**20.83**	0.9	0.38	17.78	0.68	6.13	4.67	0.6	NaN	NaN	18.04
CapU	7.86	19.98	1.25	0.48	19.17	1.03	**6.64**	5.96	0.8	NaN	NaN	16.52
π	0.58	0.47	0.14	0.07	**0.68**	0.12	0.24	0.39	0.25	NaN	NaN	0.35
R	0.69	**0.59**	0.24	0.22	0.88	0.15	0.27	0.51	0.17	NaN	NaN	0.42
Factor	5.05	**3.13**	0.21	0.09	2.8	0.17	0.97	0.93	0.39	NaN	NaN	2.7
		Transitory and filtered variables										
Y	28.94	42.79	4.07	1.56	42.04	3.38	14.72	**17.11**	6.16	7649.1	140.59	24.2(*)
C	33.58	54.45	2.32	0.86	**38.29**	1.86	13.22	11.65	3.99	NaN	NaN	23.65
I	160.43	**137.45**	25.7	6.62	203.33	23.23	73.7	89.49	100.49	NaN	NaN	110.4
H	8.29	6.96	1.35	0.64	**9.19**	1.1	3.21	4.5	1.15	12.03	12.03	5.25
RK	6.32	5.14	0.38	0.15	**5.33**	0.31	1.85	1.72	0.3	NaN	NaN	3.37
W	20.43	26.62	0.93	0.39	**19.92**	0.7	6.85	5.06	0.74	NaN	NaN	12.45
CapU	20.02	16.3	1.2	0.46	**16.88**	0.97	5.86	5.45	1.01	NaN	NaN	10.69
π	0.57	0.46	0.14	0.06	**0.67**	0.11	0.23	0.37	0.25	NaN	NaN	0.37
R	0.69	**0.58**	0.23	0.21	0.87	0.15	0.27	0.5	0.15	NaN	NaN	0.49
Factor	4.52	**2.11**	0.17	0.08	1.98	0.14	0.69	0.72	0.44	NaN	NaN	1.18

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RR the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. The variability is computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table A10: Average number of turning points, durations, and amplitudes, DGP: SW, T=750

Variable	True	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
		Output Gap and filtered output										
Number TP	125.87	123.4	**124.78**	109.11	124.51	129.74	135.95	112.77	130.24	147.91	111.78	128
DurE	5.67	5.8	**5.73**	6.68	5.79	5.5	5.22	6.39	5.52	4.5	6.49	5.62(*)
DurR	5.71	5.81	**5.75**	6.49	5.75	5.5	5.27	6.32	5.49	5.19	6.36	5.57
AmpE	-2.49	-3.62	**-3.57**	-4.69	-5.09	-1.43	-1.47	-5.67	-1.48	-16.91	-6.47	-2.67(*)
AmpR	2.49	3.62	**3.57**	4.69	5.09	1.43	1.47	5.67	1.48	16.91	6.47	2.67(*)
		Factor Gap and filtered factor										
Number TP	120.07	**120.44**	122.63	103.61	120.51	130.47	135.31	113.42	130.78	NaN	NaN	124.97
DurE	6.01	5.96	5.84	6.87	**5.98**	5.48	6.28	5.25	5.46	NaN	NaN	5.71
DurR	5.92	**5.93**	5.84	7.05	**5.93**	5.46	6.39	5.31	5.63	NaN	NaN	5.75
AmpE	-1.08	-0.83	-0.82	**-1.12**	-1.17	-0.3	-1.36	-0.32	-0.49	NaN	NaN	-0.61
AmpR	1.08	0.83	0.82	**1.12**	1.17	0.3	1.36	0.32	0.49	NaN	NaN	0.61
		Output Transitory and filtered output										
Number TP	125.57	123.94	**125.69**	109.32	125.35	130.63	137.32	112.3	130.72	150.61	111.23	123.88
DurE	5.66	5.76	**5.72**	6.56	5.73	5.44	5.19	6.35	5.6	4.18	6.52	5.76
DurR	5.75	**5.79**	5.66	6.55	5.68	5.49	5.2	6.45	5.37	5.32	6.38	5.79(*)
AmpE	-3.4	**-3.62**	-3.58	-4.72	-5.11	-1.41	-1.47	-5.74	-1.44	-22.65	-5.59	-3.62(*)
AmpR	3.4	**3.61**	3.57	4.71	5.1	1.41	1.47	5.73	1.44	22.65	5.59	3
		Factor transitory and filtered factor										
Number TP	121.2	120.79	**121.56**	105.4	121.8	131.03	113.62	135.26	126.26	NaN	NaN	120.69
DurE	5.9	5.96	**5.87**	6.87	5.93	5.45	6.25	5.25	5.72	NaN	NaN	5.96
DurR	5.92	**5.92**	5.94	6.76	5.86	5.44	6.38	5.32	5.76	NaN	NaN	5.92(*)
AmpE	-0.98	-0.74	-0.73	-1	**-1.05**	-0.27	-1.17	-0.29	-0.52	NaN	NaN	-0.74
AmpR	0.98	0.74	0.73	1	**1.05**	0.27	1.17	0.29	0.52	NaN	NaN	0.74

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output and Factor is the first principal component of the nine series. DurE and DurR are the durations of expansions and recessions; AmpE and AmpR the amplitudes of expansions and recessions. Statistics are computed averaging over 100 data replications. In bold is the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table A11: Average Phillips curve and Okun law predictions, DGP: SW, T=750

Step ahead	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Phillips curve prediction: Output Gap										
1	**0.25**	0.26	0.27	0.37	0.26	0.32	0.36	0.27	0.36	0.4	0.21(*)
4	**1.41**	1.5	1.67	1.91	1.67	1.55	1.87	1.53	2.14	2.23	1.04(*)
	Phillips curve prediction: Transitory Output										
1	0.21	0.45	0.19	0.14	0.43	0.15	0.14	0.19	0.13	**0.1**	0.21
4	0.79	0.93	0.6	**0.49**	2.89	0.7	0.52	0.77	0.52	0.58	0.68
	Okun law prediction: Output Gap										
1	0.24	1.24	0.24	0.19	3.79	0.21	**0.19**	1.21	0.21	0.26	0.44
4	3.67	33.84	2.49	**1.75**	45.94	2.74	2.23	10.88	2.43	2.04	3.13
	Okun law prediction: transitory Output										
1	0.32	0.31	0.33	**0.30**	0.31	0.32	0.31	0.37	0.34	0.33	0.32
4	3.02	2.81	2.76	**2.68**	3.05	2.73	2.85	3.27	3.21	2.89	2.93

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. The Phillips curve and the Okun law predictions are regression of the form: $x_{t+m} = \alpha_0 + \alpha_1 x_t + \sum_{j=1}^{3} \beta_j y_{t-j}$ where y_{t-j} is the true gap (transitory) or the estimated one, $x_t = \pi_t$ (inflation) or H_t (hours) and m=1,4. Reported the difference in variance of the prediction error between each procedure and the true prediction error, averaged over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table B.1: Average RMSE, DGP: SW, T=750

Variable	HPa	LDa	BPa	Hama	BNa	BQa	FODa	Trigo	BK	UCbiv
	Gap									
Y	4.93	6.54	5.13	5.20	5.39	5.29	5.71	5.21	5.21	17.75
C	5.49	6.47	5.72	5.63	3.51	3.51	6.03	5.82	5.81	NaN
I	9.76	14.54	9.95	10.81	8.67	8.67	10.99	9.52	9.50	NaN
H	2.30	3.01	2.47	1.99	NaN	NaN	2.90	2.61	2.61	NaN
RK	**1.50**	2.20	1.57	1.81	NaN	NaN	1.78	1.60	1.60	NaN
W	4.69	5.43	4.88	4.78	NaN	NaN	5.11	4.95	4.94	NaN
CapU	**2.67**	9.93	2.80	3.23	NaN	NaN	3.18	2.85	2.84	13.04
π	0.61	0.80	0.67	0.46	NaN	NaN	0.80	0.72	0.72	NaN
R	0.61	0.87	0.72	0.46	NaN	NaN	0.87	0.78	0.78	NaN
Factor	2.01	2.15	2.72	1.91	NaN	NaN	2.24	2.20	2.20	NaN
	Transitory									
Y	6.45	8.10	6.31	7.53	5.83	5.83	6.42	5.98	5.97	10.92
C	6.77	8.19	6.64	7.66	5.84	5.84	6.68	6.40	6.40	NaN
I	15.17	18.91	14.92	17.13	14.10	14.10	15.15	14.09	14.08	NaN
H	3.39	4.12	3.33	3.82	NaN	NaN	3.46	3.18	3.18	NaN
RK	2.86	3.33	2.81	3.11	NaN	NaN	2.83	2.73	2.73	NaN
W	5.26	6.23	5.17	5.83	NaN	NaN	5.20	5.03	5.03	NaN
CapU	5.09	5.93	5.01	5.55	NaN	NaN	5.04	4.86	4.86	10.65
π	0.92	1.12	0.90	1.01	NaN	NaN	0.96	0.86	0.86	NaN
R	1.04	1.27	0.99	1.14	NaN	NaN	1.12	**0.94**	**0.94**	NaN
Factor	2.21	2.45	2.19	2.35	NaN	NaN	2.20	2.15	2.15	NaN

Notes: HPa is modified Hodrick and Prescott filtering, FODa is fourth order differencing, LDa is 16th order differencing, BPa is modified band pass filtering, Hama is modified local projection detrending, BNa and BQa are trivariate Beveridge and Nelson and Blanchard and Quah decompositions, Trigo and BK are the trigonometric and the Baxter and King versions of a BP filter, UCbiv is bivariate UC filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. Statistics are computed averaging over 100 data replications. For C and I BNa and BQa refer to the comparison of filtered C/Y and I/Y with the data. In bold are cases where the reported statistic improves the best result presented in table 5.

Table B.2: Average real time RMSE, DGP: SW, T=750

Variable	HPa	LDa	BPa	Hama	BNa	BQa	FODa	Trigo	BK	UCbiv
	Gap									
Y	5.02	6.66	4.73	5.26	5.28	5.28	5.70	5.10	4.98	**2.93**
C	5.88	6.88	5.67	5.95	3.66	3.66	6.15	5.81	5.77	NaN
I	9.64	16.10	9.29	11.33	8.53	8.53	11.24	9.34	9.03	NaN
H	2.47	3.01	2.40	2.01	NaN	NaN	2.93	2.58	2.55	NaN
RK	1.81	2.25	1.76	1.92	NaN	NaN	2.02	1.82	1.80	NaN
W	5.19	5.84	4.98	5.07	NaN	NaN	5.48	5.19	5.13	NaN
capU	3.23	4.04	3.14	3.42	NaN	NaN	3.60	3.24	3.22	**1.93**
π	0.66	0.87	0.67	**0.47**	NaN	NaN	0.81	0.72	0.70	NaN
R	0.66	0.90	0.72	**0.45**	NaN	NaN	0.88	0.80	0.78	NaN
Factor	2.22	2.25	2.17	**2.00**	NaN	NaN	2.31	2.26	2.25	NaN
	Transitory									
Y	6.38	7.90	6.26	7.47	6.12	6.12	6.54	6.21	6.23	**2.74**
C	6.86	8.27	**6.67**	7.62	7.04	7.04	6.97	6.70	6.70	NaN
I	15.37	19.92	15.42	18.59	15.18	15.18	15.93	14.79	14.81	NaN
H	3.45	4.13	3.40	4.00	NaN	NaN	3.62	3.41	3.41	NaN
RK	3.12	3.55	3.14	3.53	NaN	NaN	3.13	3.09	3.09	NaN
W	4.74	5.89	4.75	5.48	NaN	NaN	4.73	4.68	3.67	NaN
capU	5.56	6.32	5.59	6.29	NaN	NaN	5.57	5.50	5.51	**1.43**
π	0.91	1.14	0.88	1.06	NaN	NaN	1.01	0.88	0.86	NaN
R	1.06	1.30	1.01	1.23	NaN	NaN	1.19	1.01	0.98	NaN
Factor	2.45	2.67	2.41	2.60	NaN	NaN	2.48	2.40	2.41	NaN

Notes: HPa is modified Hodrick and Prescott filtering, FODa is forth order differencing, LDa is 16th order differencing, BPa is modified band pass filtering, Hama is modified local projection detrending, BNa and BQa are trivariate Beveridge and Nelson and Blanchard and Quah decompositions, Trigo and BK are the trigonometric and the Baxter and King versions of a BP filter, UCbiv is bivariate UC filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. Statistics are computed averaging over 100 data replications. For C and I BNa and BQa refer to the comparison of filtered C/Y and I/Y with the data. In bold are cases where the reported statistic improves the best result presented in table 6.

Table B.3: Average contemporaneous correlation, DGP: SW, T=750

Variable	HPa	LDa	BPa	Hama	BNa	BQa	FODa	Trigo	BK	UCbiv
	Gap									
Y	0.46	0.35	0.36	0.53	0.03	0.03	0.15	0.24	0.25	-0.01
C	0.44	0.32	0.32	0.46	0.81	0.81	0.16	0.22	0.23	NaN
I	0.38	0.31	0.29	0.51	0.73	0.73	0.14	0.19	0.19	NaN
H	0.58	0.45	0.46	0.76	NaN	NaN	0.22	0.31	0.31	NaN
RK	0.5	0.35	0.39	0.44	NaN	NaN	0.16	0.29	0.29	NaN
W	0.39	0.25	0.26	0.4	NaN	NaN	0.07	0.15	0.16	NaN
capU	0.5	0.35	0.39	0.44	NaN	NaN	0.16	0.29	0.29	0.03
π	0.71	0.55	0.6	0.88	NaN	NaN	0.38	0.47	0.48	NaN
R	0.76	0.57	0.59	0.9	NaN	NaN	0.44	0.48	0.48	NaN
Factor	0.46	0.38	0.32	0.52	NaN	NaN	0.15	0.2	0.2	NaN
	Transitory									
Y	0	0	0	0	0	0	0	0	0	0.07
C	0	-0.02	0	-0.02	-0.05	-0.05	-0.01	0	0	NaN
I	0	0	0	-0.01	0.01	0.01	-0.01	0	0	NaN
H	0	0.01	0	0	NaN	NaN	0	0	0	NaN
RK	0	-0.01	0	0	NaN	NaN	-0.01	0	0	NaN
W	0	-0.01	0.01	-0.03	NaN	NaN	-0.01	0	0	NaN
CapU	0	-0.01	0	0	NaN	NaN	-0.01	0	0	-0.01
π	0.01	0.01	0.01	0.01	NaN	NaN	0	0	0	NaN
R	0	-0.01	0	-0.01	NaN	NaN	0	0	0	NaN
Factor	0	0	0	0	NaN	NaN	0	0	0	NaN

Notes: HPa is modified Hodrick and Prescott filtering, FODa is forth order differencing, LDa is 16th order differencing, BPa is modified band pass filtering, Hama is modified local projection detrending, BNa and BQa are trivariate Beveridge and Nelson and Blanchard and Quah decompositions, Trigo and BK are the trigonometric and the Baxter and King versions of a BP filter, UCbiv is bivariate UC filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. Statistics are computed averaging over 100 data replications. For C and I BNa and BQa refer to the comparison of filtered C/Y and I/Y with the data. In bold are cases where the reported statistic improves the best result presented in table 7.

Table B.4: Average AR1 coefficient, DGP: SW, T=750

Variable	True	HPa	LDa	BPa	Hama	BNa	BQa	FODa	Trigo	BK	UCbiv
	Gap										
Y	0.98	0.92	0.95	0.96	0.94	0.21	0.21	0.83	0.93	0.93	0.84
C	0.98	0.93	0.96	0.97	0.93	0.98	0.98	0.84	0.93	0.93	NaN
I	0.99	0.94	0.96	0.96	0.95	0.98	0.98	0.88	0.93	0.93	NaN
H	0.98	0.88	0.92	0.95	0.92	NaN	NaN	0.78	0.93	0.93	NaN
RK	0.97	0.93	0.96	**0.97**	0.93	NaN	NaN	0.84	0.94	0.94	NaN
W	0.98	0.94	0.97	0.97	0.92	NaN	NaN	0.84	0.94	0.94	NaN
CapU	0.97	0.93	0.96	**0.97**	0.93	NaN	NaN	0.84	0.94	0.94	0.78
π	0.94	0.86	0.9	**0.95**	0.91	NaN	NaN	0.77	0.92	0.92	NaN
R	0.83	0.67	0.75	0.94	0.79	NaN	NaN	0.57	0.91	0.91	NaN
Factor	0.98	0.92	0.96	0.97	0.93	NaN	NaN	0.82	0.93	0.93	NaN
	Transitory										
Y	0.97	0.92	0.95	0.96	0.93	0.2	0.2	0.82	0.93	0.93	0.96
C	0.99	0.94	0.97	0.97	0.93	0.98	0.98	0.85	0.93	0.93	NaN
I	0.98	0.94	0.96	0.96	0.95	0.98	0.98	0.87	0.93	0.93	NaN
H	0.96	0.87	0.92	0.96	0.92	NaN	NaN	0.77	0.93	0.93	NaN
RK	0.99	0.93	0.96	0.97	0.93	NaN	NaN	0.83	0.94	0.94	NaN
W	0.99	0.94	0.97	0.97	0.92	NaN	NaN	0.84	0.94	0.94	NaN
CapU	0.99	0.93	0.96	0.97	0.93	NaN	NaN	0.83	0.94	0.94	0.94
π	0.94	0.86	0.9	**0.95**	0.91	NaN	NaN	0.77	0.92	0.92	NaN
R	0.84	0.67	0.74	0.94	0.79	NaN	NaN	0.57	0.91	0.91	NaN
Factor	0.99	0.91	0.95	0.96	0.93	NaN	NaN	0.8	0.93	0.93	NaN

Notes: HPa is modified Hodrick and Prescott filtering, FODa is forth order differencing, LDa is 16th order differencing, BPa is modified band pass filtering, Hama is modified local projection detrending, BNa and BQa are trivariate Beveridge and Nelson and Blanchard and Quah decompositions, Trigo and BK are the trigonometric and the Baxter and King versions of a BP filter, UCbiv is bivariate UC filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. Statistics are computed averaging over 100 data replications. For C and I BNa and BQa refer to the comparison of filtered C/Y and I/Y with the data. In bold are cases where the reported statistic improves the best result presented in table 8.

Table B.5: Average variability, DGP: SW, T=750

Variable	True	HPa	LDa	BPa	Hama	BNa	BQa	FODa	Trigo	BK	UCbiv
					Gap						
Y	23.62	9.55	33.07	7.64	**22.67**	1.55	1.55	8.87	3.5	3.45	276.56
C	28.77	5.55	21.42	4.27	14.67	13.32	13.32	4.85	1.81	1.78	NaN
I	69.37	**59.88**	195.72	50.47	127.94	94.69	94.69	55.95	24.4	24.01	NaN
H	6.43	2.63	8.1	2.13	5.36	NaN	NaN	3.02	1.17	1.15	NaN
RK	2.48	1.12	4.28	0.81	**2.84**	NaN	NaN	0.9	0.33	0.33	NaN
W	20.85	2.92	12	2.09	7.6	NaN	NaN	2.11	0.7	0.68	NaN
CapU	7.86	3.54	13.57	2.57	**8.99**	NaN	NaN	2.85	1.05	1.03	127.04
π	0.58	0.23	**0.65**	0.2	0.43	NaN	NaN	0.31	0.12	0.12	NaN
R	0.69	0.33	0.86	0.23	0.54	NaN	NaN	0.52	0.15	0.15	NaN
Factor	5.05	0.53	2	0.41	1.35	NaN	NaN	0.47	0.17	0.17	NaN
					Transitory						
Y	28.94	9.38	**32.94**	7.56	23.87	1.55	1.55	8.82	3.42	3.38	261.54
C	33.58	6.46	26.44	4.88	18.27	21.93	21.93	5.36	1.88	1.86	NaN
I	160.43	55.57	178.23	48	114.37	104.02	104.02	53.11	23.55	23.23	NaN
H	8.29	2.53	**7.89**	2.07	5.5	NaN	NaN	2.92	1.11	1.1	NaN
RK	6.32	1.03	3.87	0.78	2.54	NaN	NaN	0.85	0.31	0.31	NaN
W	20.43	3.1	13.2	2.24	8.49	NaN	NaN	2.22	0.71	0.7	NaN
CapU	20.02	3.26	12.25	2.49	8.04	NaN	NaN	2.68	0.99	0.97	89.91
π	0.57	0.22	**0.63**	0.19	0.41	NaN	NaN	0.3	0.11	0.11	NaN
R	0.69	0.32	0.84	0.23	0.53	NaN	NaN	0.51	0.15	0.15	NaN
Factor	4.52	0.41	1.5	0.32	1.03	NaN	NaN	0.38	0.14	0.14	NaN

Notes: HPa is modified Hodrick and Prescott filtering, FODa is forth order differencing, LDa is 16th order differencing, BPa is modified band pass filtering, Hama is modified local projection detrending, BNa and BQa are trivariate Beveridge and Nelson and Blanchard and Quah decompositions, Trigo and BK are the trigonometric and the Baxter and King versions of a BP filter, UCbiv is bivariate UC filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. Statistics are computed averaging over 100 data replications. For C and I BNa and BQa refer to the comparison of filtered C/Y and I/Y with the data. In bold are cases where the reported statistic improves the best result presented in table 9.

Table B.6: Average number of turning points, durations, and amplitudes, DGP: SW, T=750

Variable	True	HPa	LDa	BPa	Hama	BNa	BQa	FODa	Trigo	BK	UCbiv
	Output Gap										
Number of TP	125.87	123.44	124.31	128.28	119.83	109.11	109.11	141.49	129.37	129.74	129.2
DurE	5.67	5.77	5.75	5.55	5.99	6.68	6.68	5.03	5.52	5.5	4.93
DurR	5.71	5.83	5.77	5.58	6	6.49	6.49	5.06	5.52	5.5	6.27
AmpE	-2.49	-3.61	-5.08	-1.45	-4.95	-4.69	-4.69	-5.22	-1.43	-1.43	-21.68
AmpR	2.49	3.61	5.08	1.45	4.95	4.69	4.69	5.22	1.43	1.43	- 21.69
	Factor Gap										
Number of TP	120.07	120.69	121.06	128.56	118.84	NaN	NaN	138.45	129.98	130.44	NaN
DurE	6.01	5.93	5.9	5.57	6.05	NaN	NaN	5.16	5.49	5.49	NaN
DurR	5.92	**5.93**	5.96	5.53	6.05	NaN	NaN	5.16	5.49	5.46	NaN
AmpE	-1.08	-0.83	-1.16	-0.31	-1.25	NaN	NaN	-1.19	-0.3	-0.3	NaN
AmpR	1.08	0.83	1.16	0.31	1.25	NaN	NaN	1.19	0.3	0.3	NaN
	Transitory output										
Number of TP	125.57	123.94	124.84	128.87	119.01	109.32	109.32	142.5	130.56	130.63	127.76
DurE	5.66	5.79	5.8	5.53	6.09	6.56	6.56	5.03	5.46	5.44	5.5
DurR	5.75	**5.76**	5.68	5.55	5.97	6.55	6.55	5	5.47	5.49	5.81
AmpE	-3.4	**-3.61**	-5.1	-1.45	-5.17	-4.72	-4.72	-5.25	-1.41	-1.41	-12.25
AmpR	3.4	**3.61**	5.1	1.45	5.18	4.71	4.71	5.25	1.41	1.41	12.27
	Transitory factor										
Number of TP	121.2	120.91	120.96	129.39	118.07	NaN	NaN	140.58	131.44	131.04	NaN
DurE	5.9	5.97	5.98	5.52	6.12	NaN	NaN	5.07	5.44	5.45	NaN
DurR	5.92	5.91	5.89	5.52	6.03	NaN	NaN	5.11	5.43	5.44	NaN
AmpE	-0.98	-0.74	-1.05	-0.27	-1.07	NaN	NaN	-1.06	-0.27	-0.27	NaN
AmpR	0.98	0.74	1.05	0.27	1.08	NaN	NaN	1.06	0.27	0.27	NaN

Notes: HPa is modified Hodrick and Prescott filtering, FODa is forth order differencing, LDa is 16th order differencing, BPa is modified band pass filtering, Hama is modified local projection detrending, BNa and BQa are trivariate Beveridge and Nelson and Blanchard and Quah decompositions, Trigo and BK are the trigonometric and the Baxter and King versions of a BP filter, UCbiv is bivariate UC filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. Statistics are computed averaging over 100 data replications. For C and I BNa and BQa refer to the comparison of filtered C/Y and I/Y with the data. In bold are cases where the reported statistic improves the best result presented in table 10.

Table B.7 Average Phillips curve and Okun law predictions, DGP: SW, T=750

Variable	HPa	LDa	BPa	Hama	BNa	BQa	FODa	Trigo	BK	UCbiv
	Phillips curve prediction: Output Gap									
1	0.21	0.36	0.25	0.34	0.27	0.27	0.32	0.26	0.26	0.32
4	1.48	1.81	1.88	1.74	1.67	1.67	1.69	1.68	1.67	1.6
	Phillips curve prediction: Transitory Output									
1	0.36	0.15	0.39	0.13	0.19	0.19	0.16	0.42	0.43	0.15
4	1.03	0.54	3.21	0.5	0.6	0.6	0.62	2.87	2.89	0.73
	Phillips curve prediction: Output Gap									
1	0.55	0.21	4.67	0.21	0.24	0.24	0.23	3.75	3.79	4.91
4	17.25	1.99	55.2	2.56	2.49	2.49	2.27	45.59	45.94	17.28
	Okun law prediction: Transitory Output									
1	0.32	0.31	0.32	0.31	0.33	0.33	0.31	0.31	0.31	0.34
4	2.85	2.69	3.21	2.97	2.76	2.76	2.63	3.03	3.05	2.95

Notes: HPa is modified Hodrick and Prescott filtering, FODa is forth order differencing, LDa is 16th order differencing, BPa is modified band pass filtering, Hama is modified local projection detrending, BNa and BQa are trivariate Beveridge and Nelson and Blanchard and Quah decompositions, Trigo and Bk are the trigonometric and the Baxter and King versions of a BP filter, UCbiv is bivariate UC filtering. The Phillips curve and the Okun law predictions are regression of the form ; $x_{t+m} = \alpha_0 + \alpha_1 x_t + \sum_{j=1}^{3} \beta_j y_{t-j}$ where y_{t-j} is the true gap (transitory) or the estimated one, $x_t = \pi_t$ (inflation) or H_t (hours) and m=1,4. Reported is the difference in variance of the prediction error between each procedure and the true prediction error, averaged over 100 data replications. In bold cases where the reported statistic improves the best result presented in table 11.

Table B.8: Average RMSE, DGP: SW, T=150

Variable	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Gap										
Y	5.22	**5.13**	5.45	7.18	5.24	5.46	5.29	10.22	33.8	11.86	4.47(*)
C	**5.65**	5.79	6.03	7.09	5.89	6.00	5.85	10.62	NaN	NaN	4.72(*)
I	10.58	**9.27**	9.37	15.84	9.44	10.85	10.38	12.61	NaN	NaN	9.48
H	2.34	2.56	2.81	2.98	2.64	2.38	2.29	9.47	**1.89**	**1.89**	1.03(*)
RK	1.62	**1.58**	1.73	2.57	1.64	1.88	1.78	1.64	NaN	NaN	2.18
W	4.44	4.59	4.75	5.55	4.66	4.75	4.67	**4.35**	NaN	NaN	3.32(*)
CapU	2.88	**2.81**	3.09	4.58	2.91	3.34	3.17	4.77	NaN	NaN	3.86
π	0.64	0.71	0.83	0.81	0.75	0.69	**0.6**	0.76	NaN	NaN	0.17(*)
R	**0.63**	0.71	0.88	0.87	0.79	0.73	0.59	0.76	NaN	NaN	0.02(*)
Factor	**2.00**	2.12	2.28	2.56	2.18	2.08	2.09	2.17	NaN	NaN	1.57(*)
	Transitory										
Y	6.87	6.14	**5.93**	8.82	6.07	7.02	6.95	9.37	43.00	10.98	7.09
C	7.27	6.66	**5.56**	8.96	6.62	7.45	7.21	10.08	NaN	NaN	7.43
I	16.23	14.51	13.87	20.08	14.41	16.36	16.24	**13.15**	NaN	NaN	16.74
H	3.53	3.21	**3.10**	4.41	3.16	3.58	3.60	7.09	7.30	7.30	3.65
RK	3.01	2.74	2.70	3.52	2.72	3.01	2.92	**1.77**	NaN	NaN	3.09
W	5.79	5.33	5.29	6.96	5.37	5.92	5.65	**4.85**	NaN	NaN	6.04
CapU	5.36	4.88	4.81	6.28	4.85	5.36	5.20	**4.62**	NaN	NaN	5.51
π	0.93	0.87	**0.84**	1.14	0.86	0.92	0.95	0.88	NaN	NaN	0.96
R	1.05	0.99	0.99	1.29	0.95	1.02	1.09	**0.92**	NaN	NaN	1.09
Factor	2.50	2.17	2.09	3.24	**2.14**	2.55	2.47	2.81	NaN	NaN	2.62

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. The RMSE is computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table B.9: Average real time RMSE, DGP: SW, T=150

Variable	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Gap										
Y	6.08	5.98	6.21	6.54	6.00	**5.15**	5.78	10.96	18.24	16.43	4.34(*)
C	6.76	6.65	6.84	6.33	6.65	**5.86**	6.69	11.04	NaN	NaN	4.6(*)
I	11.60	**10.51**	10.56	15.76	10.52	10.74	10.85	16.94	NaN	NaN	8.83(*)
H	3.07	2.99	3.2	3.00	3.01	2.52	2.44	9.07	**1.84**	**1.84**	1.01(*)
RK	1.95	**1.55**	1.64	2.86	1.56	1.97	1.87	2.71	NaN	NaN	2.28
W	4.92	4.75	4.81	5.19	4.76	**4.48**	4.80	5.4	NaN	NaN	3.19(*)
CapU	3.48	**2.76**	2.91	5.09	2.77	3.51	3.32	6.1	NaN	NaN	4.07
π	0.82	0.80	0.93	0.89	0.82	0.77	**0.70**	0.82	NaN	NaN	0.2(*)
R	0.82	0.80	0.96	0.93	0.86	0.80	**0.67**	0.88	NaN	NaN	0.22(*)
Factor	3.19	2.83	2.86	3.08	2.81	**2.72**	2.86	3.05	NaN	NaN	1.95(*)
	Transitory										
Y	6.75	6.26	**6.22**	8.75	6.26	7.05	7.04	11.07	21.43	14.74	7.26
C	7.29	6.63	**6.59**	9.45	6.64	7.71	7.22	11.27	NaN	NaN	7.49
I	15.21	13.69	**13.16**	19.42	13.69	15.82	15.53	16.71	NaN	NaN	15.64
H	3.35	3.16	3.15	4.06	**3.13**	3.34	3.54	7.85	7.83	7.83	3.43
RK	3.12	2.77	2.76	3.79	2.74	3.16	3.01	**2.32**	NaN	NaN	2.83
W	6.16	5.68	5.66	7.63	**5.68**	6.48	5.96	6.13	NaN	NaN	6.34
CapU	5.55	4.94	4.92	6.74	**4.89**	5.64	5.30	5.02	NaN	NaN	5.05
π	0.91	0.85	**0.83**	1.12	0.84	0.92	0.97	0.86	NaN	NaN	0.95
R	1.05	0.95	0.95	1.34	0.91	1.05	1.15	**0.86**	NaN	NaN	1.14
Factor	2.9	2.53	**2.52**	3.39	2.53	2.83	2.83	2.68	NaN	NaN	2.69

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. The RMSE is computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table B.10: Average contemporaneous correlations, DGP: SW, T=150

Variable	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Gaps and Filtered variables										
Y	**0.59**	0.45	0.04	0.48	0.33	0.52	0.54	0.13	-0.58	0.15	0.74(*)
C	**0.58**	0.45	0.08	0.46	0.33	0.48	0.48	0.11	NaN	NaN	0.76(*)
I	**0.51**	0.37	0.04	0.45	0.27	0.49	0.51	0.29	NaN	NaN	0.66(*)
H	0.75	0.53	0.09	0.62	0.41	0.68	0.77	0.1	**0.93**	**0.93**	0.93(*)
RK	**0.6**	0.5	0.03	0.42	0.37	0.45	0.43	0.34	NaN	NaN	0.61(*)
W	**0.57**	0.4	0.02	0.44	0.26	0.45	0.39	0.13	NaN	NaN	0.74(*)
CapU	**0.6**	0.5	0.03	0.42	0.37	0.45	0.43	0.25	NaN	NaN	0.61(*)
π	0.87	0.74	0.22	0.65	0.6	0.72	**0.91**	0.58	NaN	NaN	1(*)
R	0.89	0.79	0.34	0.66	0.59	0.68	**0.93**	0.77	NaN	NaN	1(*)
Factor	**0.49**	0.34	0.01	0.38	0.24	0.41	0.39	0.32	NaN	NaN	0.7(*)
	Transitory and Filtered variables										
Y	-0.01	-0.01	-0.02	-0.02	-0.01	-0.03	-0.03	-0.02	-0.08	0.02	-0.03
C	0	0.01	-0.01	-0.04	0.02	-0.03	-0.02	-0.03	NaN	NaN	0
I	-0.03	0	-0.02	-0.02	-0.01	-0.01	-0.02	-0.03	NaN	NaN	0
H	-0.02	-0.01	-0.02	-0.02	-0.01	-0.02	-0.03	0	-0.02	-0.02	-0.02
RK	-0.04	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	0.02	NaN	NaN	-0.04
W	-0.02	0.01	-0.03	-0.06	0.01	-0.05	-0.04	0.02	NaN	NaN	-0.06
CapU	-0.04	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	0	NaN	NaN	-0.04
π	0.01	0.01	0	0	0.01	0.02	0.03	-0.01	NaN	NaN	-0.01
R	0.01	0.01	0	-0.01	0.01	-0.01	0.02	0.01	NaN	NaN	0
Factor	-0.04	-0.01	-0.01	-0.06	-0.01	-0.06	-0.04	-0.01	NaN	NaN	-0.04

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. Correlations are computed averaging over 100 data replications.In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table B.11: Average AR1 coefficient, DGP: SW, T=150

Variable	True	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Gaps and filtered variables											
Y	0.96	0.94	0.81	0.2	**0.95**	0.93	0.98	0.89	1.00	0.97	0.79	0.96(*)
C	0.96	0.95	0.81	0.21	**0.96**	0.93	0.98	0.88	1.00	NaN	NaN	0.97
I	0.98	0.96	0.88	0.5	0.97	0.93	**0.98**	0.91	0.99	NaN	NaN	0.97
H	0.96	0.9	0.76	0.11	0.92	0.92	**0.97**	0.88	1.00	0.93	0.93	0.93
RK	0.95	**0.95**	0.81	0.19	0.96	0.93	0.98	0.88	0.92	NaN	NaN	0.98
W	0.95	**0.96**	0.8	0.1	0.97	0.93	0.99	0.86	0.93	NaN	NaN	0.98
CapU	0.95	**0.95**	0.81	0.19	0.96	0.93	0.98	0.88	0.92	NaN	NaN	0.98
π	0.9	0.87	0.76	0.16	**0.88**	0.92	0.96	0.86	0.77	NaN	NaN	0.9 (*)
R	0.76	0.69	0.53	-0.21	**0.73**	0.91	0.94	0.72	0.29	NaN	NaN	0.78(*)
Factor	0.97	0.95	0.81	0.19	**0.96**	0.93	**0.98**	0.88	0.91	NaN	NaN	0.98(*)
	Transitory and filtered variables											
Y	0.95	0.93	0.81	0.2	**0.95**	0.93	0.98	0.89	1.00	0.95	0.86	0.94
C	0.96	**0.95**	0.83	0.24	**0.97**	0.93	0.98	0.88	1.00	NaN	NaN	0.96(*)
I	0.96	0.95	0.88	0.48	**0.96**	0.93	0.98	0.9	0.99	NaN	NaN	0.96(*)
H	0.94	0.89	0.76	0.09	0.92	0.92	0.97	0.87	0.99	**0.93**	**0.93**	0.91
RK	0.97	0.95	0.81	0.18	**0.96**	0.93	**0.98**	0.88	0.93	NaN	NaN	0.95
W	0.97	0.95	0.81	0.11	**0.97**	0.93	0.99	0.87	0.92	NaN	NaN	0.96
CapU	0.97	0.95	0.81	0.18	**0.96**	0.93	**0.98**	0.88	0.92	NaN	NaN	0.95
π	0.9	0.86	0.76	0.13	**0.89**	0.92	0.96	0.86	0.77	NaN	NaN	0.88
R	0.75	0.69	0.52	-0.21	**0.73**	0.91	0.94	0.72	0.29	NaN	NaN	0.73(*)
Factor	0.97	0.93	0.81	0.17	**0.96**	0.93	**0.98**	0.88	0.91	NaN	NaN	0.95

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. The AR1 coefficient is computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table B.12: Average variability, DGP: SW, T=150

Variable	True	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Gaps and filtered variables											
Y	15.65	**14.85**	3.95	1.56	36.28	3.21	12.1	13.7	192.03	425.21	24.07	20.88
C	17.59	**9.23**	2.06	0.81	24.54	1.67	8.11	7.98	190.93	NaN	NaN	18.35(*)
I	44.67	87.39	**25.58**	6.85	205.24	22.25	70.02	80.12	242.9	NaN	NaN	93.09
H	4.66	3.4	1.34	0.65	7.72	1.08	2.58	3.58	186.35	**5.41**	**5.41**	3.54
RK	1.9	**1.92**	0.38	0.15	4.87	0.3	1.61	1.53	0.23	NaN	NaN	2.65
W	11.4	5.54	0.86	0.38	**14.81**	0.64	4.81	3.75	0.51	NaN	NaN	9.36
CapU	6.02	**6.07**	1.2	0.48	15.44	0.94	5.12	4.85	29.82	NaN	NaN	8.41
π	0.42	0.27	0.14	0.07	**0.54**	0.11	0.18	0.3	0.08	NaN	NaN	0.26
R	0.53	**0.38**	0.23	0.22	0.75	0.15	0.22	0.41	0.06	NaN	NaN	0.32
Factor	5.36	1.92	0.49	0.27	**4.82**	0.4	1.56	1.7	0.26	NaN	NaN	2.8
	Transitory and filtered variables											
Y	18.21	**13.18**	3.93	1.57	33.83	3.1	11.28	13.14	204.37	652.08	25.43	19.44(*)
C	17.43	**10.23**	2.28	0.87	27.95	1.75	9.1	8.86	207.63	NaN	NaN	15.46(*)
I	113.84	**70.9**	24.77	6.64	168.26	21.36	58.08	68.56	237.58	NaN	NaN	102.97(*)
H	5.68	3.39	1.31	0.65	**7.98**	1.04	2.61	3.55	169.43	6.93	6.93	4.72(*)
RK	3.95	1.62	0.36	0.14	**4.02**	0.29	1.32	1.31	0.23	NaN	NaN	2.2
W	10.97	5.44	0.89	0.38	**15.03**	0.66	4.85	3.83	0.45	NaN	NaN	8.79(*)
CapU	12.52	5.15	1.14	0.46	**12.75**	0.91	4.18	4.14	22.5	NaN	NaN	6.99
pi	0.38	0.25	0.13	0.06	**0.53**	0.1	0.17	0.27	0.08	NaN	NaN	0.32(*)
R	0.5	0.36	0.22	0.21	**0.71**	0.14	0.21	0.4	0.06	NaN	NaN	0.45(*)
Factor	4.53	1.82	0.54	0.27	**4.58**	0.43	1.51	1.72	4.38	NaN	NaN	2.64

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component component, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output, C consumption, I investment, H hours, RK the real rate, W the real wage, CapU capacity utilization, R the nominal rate, and π the inflation rate. Factor is the first principal component of the nine series. The variability is computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table B.13: Average number of turning points, durations, and amplitudes, DGP: SW, T=150

Variable	True	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
		Output Gap and filtered output										
Number of TP	19.97	20.44	20.56	17.9	**19.66**	20.36	21.63	19.47	20.07	22.12	17.18	21.08
DurE	5.94	**5.75**	**5.75**	6.91	6.19	5.64	5.34	6	5.64	5.33	6.81	5.61
DurR	5.75	5.61	5.57	6.35	5.66	5.51	5.24	**5.83**	5.64	5.31	6.9	5.38
AmpE	-2.47	-3.58	-3.55	-4.7	-5.12	-1.41	**-1.48**	-5.29	-1.83	-11.88	-5.9	-2.66(*)
AmpR	2.46	3.59	3.55	4.72	5.1	1.4	**1.47**	5.27	1.81	11.86	5.92	2.67(*)
		factor Gap and filtered factor										
Number of TP	19.37	19.08	**19.23**	17.73	19.71	20.23	18.5	21.48	21.68	NaN	NaN	19.89
DurE	6.06	6.01	**6.02**	6.77	6.15	5.66	6.32	5.37	5.24			5.8
DurR	5.92	6.13	**5.94**	6.55	5.79	5.54	6.26	5.25	5.69	NaN	NaN	5.84
AmpE	-1.53	-1.21	-1.19	**-1.55**	-1.68	-0.46	-1.82	-0.48	-0.46	NaN	NaN	-0.89
AmpR	1.53	1.21	1.19	**1.55**	1.68	0.46	1.83	0.48	0.46	NaN	NaN	0.89
		Output Transitory and filtered output										
Number of TP	21.05	19.99	20.11	17.67	19.6	20.47	22.04	19.05	**20.44**	22.77	17.87	19.85
DurE	5.47	5.66	5.63	6.85	5.98	**5.48**	5.08	6.34	5.82	4.7	6.73	5.72
DurR	5.57	6.01	5.84	6.38	5.85	**5.61**	5.18	6.16	**5.61**	5.63	6.8	6.06
AmpE	-3.42	-3.6	**-3.56**	-4.75	-5.18	-1.42	-1.47	-5.25	-1.84	-15.51	-5.46	-3.61
AmpR	3.42	3.61	**3.57**	4.74	5.17	1.42	1.46	5.24	1.81	15.46	5.48	3.61
		Factor transitory and filtered factor										
Number of TP	19.97	19.07	19.1	17.85	**19.59**	20.33	19.02	21.68	21.84	NaN	NaN	18.98
DurE	5.91	6.14	6.14	6.62	**6.04**	5.58	6.05	5.28	5.24	NaN	NaN	6.2
DurR	5.73	6.02	5.97	6.44	**5.71**	5.61	6.26	5.26	5.39	NaN	NaN	6.03
AmpE	-1.34	**-1.27**	-1.26	-1.66	-1.8	-0.49	-1.87	-0.5	-1.41	NaN	NaN	-1.28(*)
AmpR	1.34	**1.27**	1.26	1.67	1.81	0.5	1.88	0.5	1.41	NaN	NaN	1.28(*)

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. Y is output and Factor is the first principal component of the nine series. DurE and DurR are the durations of expansions and recessions; AmpE and AmpR the amplitude of expansions and recessions. Statistics are computed averaging over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

Table B.14: Average Phillips curve and Okun law predictions, DGP: SW, T=150

Step ahead	POLY	HP	FOD	LD	BP	Wa	Ham	UC	BN	BQ	BW
	Phillips curve prediction: Output Gap										
1	0.46	0.6	0.43	0.48	0.64	0.44	0.49	**0.36**	0.42	0.52	0.35(*)
4	2.49	2.76	2.06	2.75	3.92	2.5	2.55	**2.08**	2.8	2.93	1.65(*)
	Phillips curve prediction: Transitory Output										
1	0.58	0.64	0.53	**0.45**	0.53	0.5	0.55	0.61	0.49	0.52	0.57
4	2.52	2.33	2.22	**2.16**	3.26	2.39	2.24	2.6	2.28	2.21	2.45
	Okun law prediction: Output Gap										
1	0.79	1.21	0.75	0.71	3.26	**0.68**	0.69	0.7	0.84	0.69	0.84
4	12.92	26.59	**7.96**	8.98	39.17	8.92	8.25	8.31	14.3	11.56	9.36
	Okun law prediction: transitory Output										
1	1.46	1.44	1.53	1.39	1.45	**1.38**	1.53	1.45	1.51	1.45	1.48
4	13.94	13.69	**12.89**	13.54	15.62	15.03	15.63	16.69	15.78	13.5	15.48

Notes: POLY is polynomial detrending, HP is Hodrick and Prescott filtering, FOD is first order differencing, LD is 24th order differencing, UC is unobservable component filtering, BP is band pass filtering, Ham is local projection detrending, BN and BQ are bivariate Beveridge and Nelson and Blanchard and Quah decompositions, BW is Butterworth filtering. The Phillips curve and the Okun law predictions are regression of the form ; $x_{t+m} = \alpha_0 + \alpha_1 x_t + \sum_{j=1}^{3} \beta_j y_{t-j}$ where y_{t-j} is the true gap (transitory) or the estimated one, $x_t = \pi_t$ (inflation) or H_t (hours) and m=1,4. Reported the difference in variance of the prediction error between each procedure and the true prediction error, averaged over 100 data replications. In bold the best approach. A (*) in the last column indicates that the BW filter improves over or is comparable to the best approach.

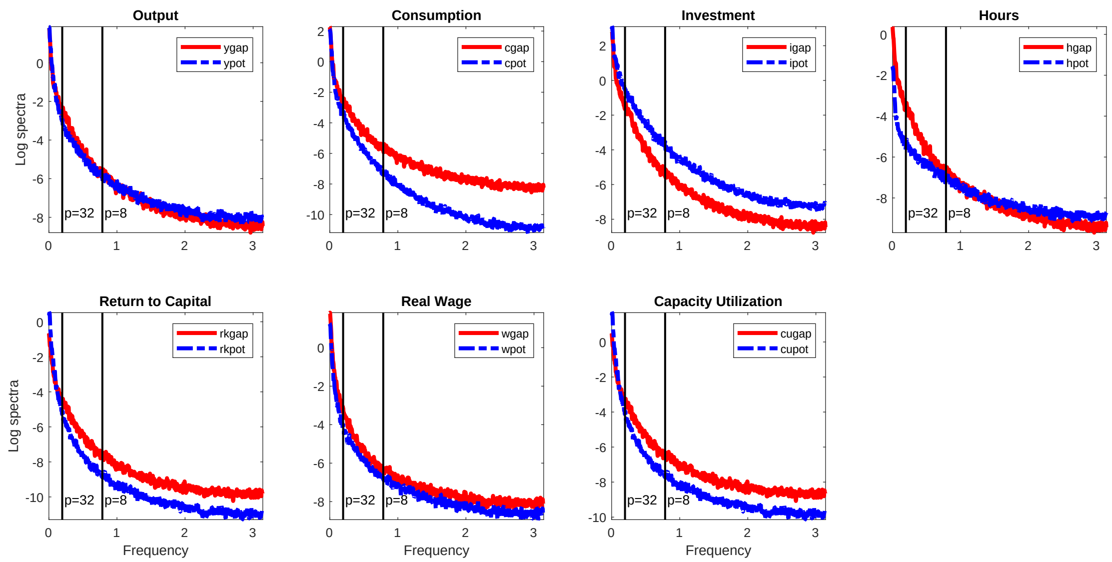

Figure C.1: Log spectral densities of gaps and potentials: Stationary SW model.

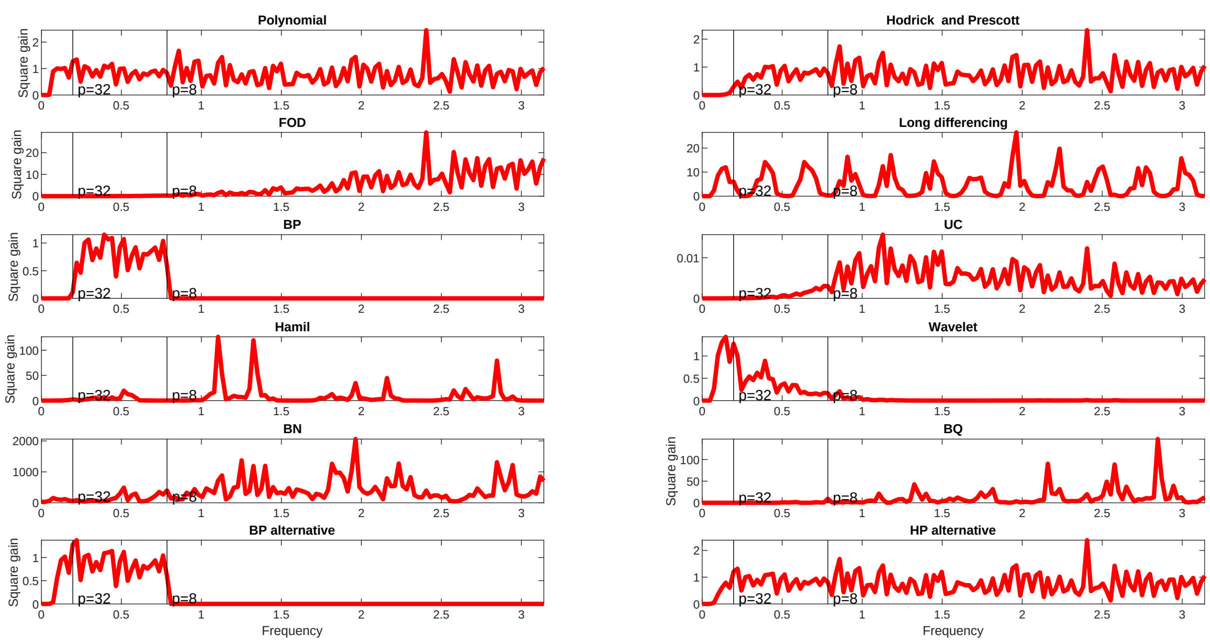

Figure C.2: Estimated squared gain functions for output; SW unit root in TFP.

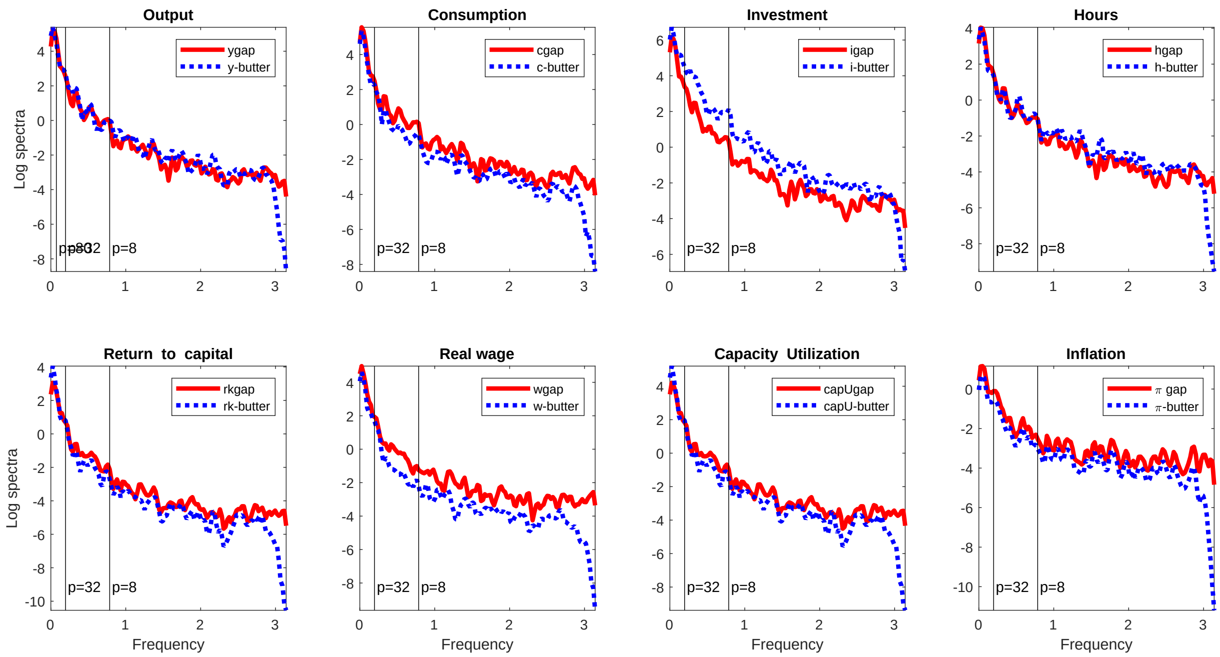

Figure C.3: Log spectra of gaps and of Butterworth estimates.

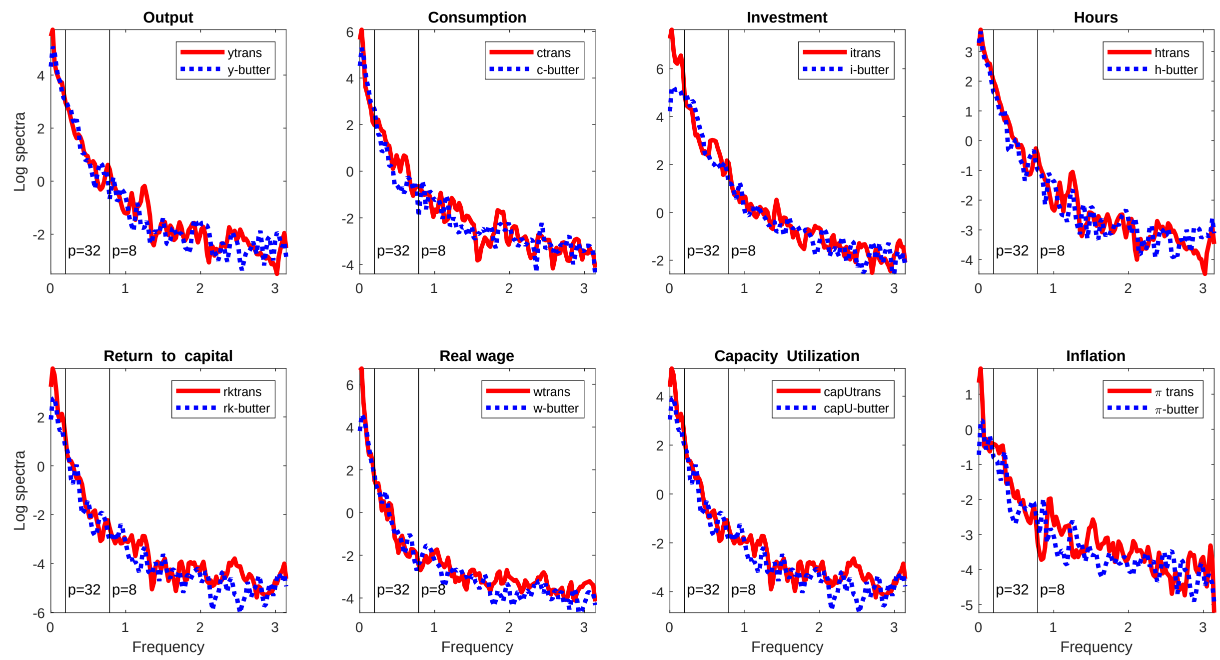

Figure C.4: Log spectra of transitory components and of Butterworth estimates.

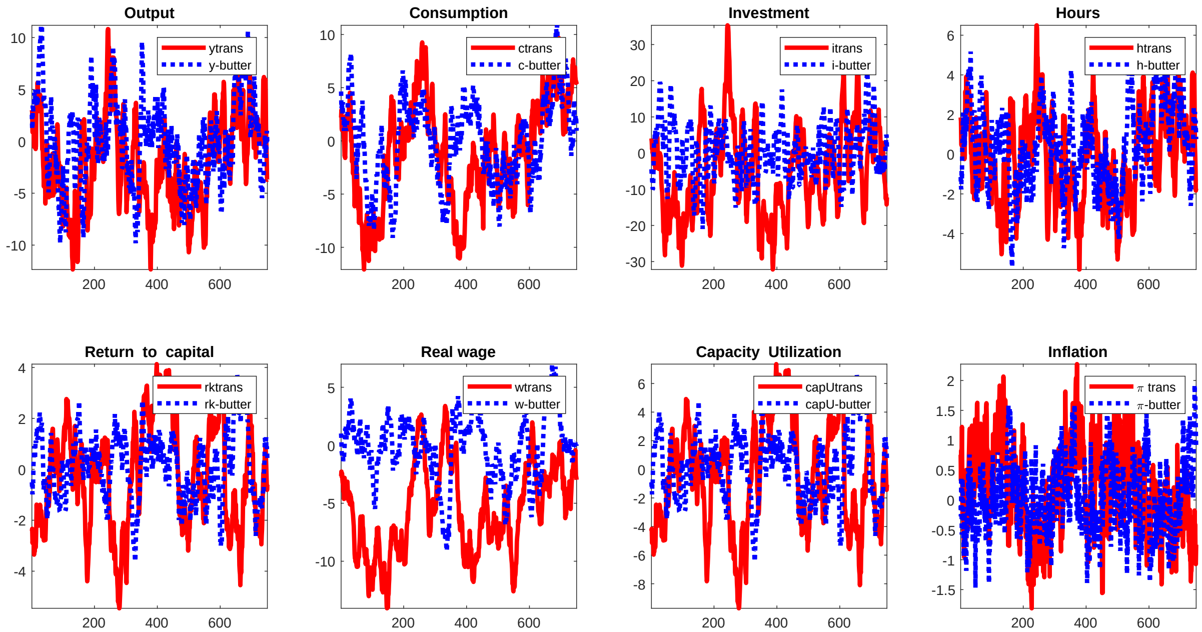

Figure C.5: Time series of transitory components and of Butterworth estimates.