

NBS Working paper
8/2023

MIDAS regression: a new horse in the race of filtering macroeconomic time series

Michal Benčík

© Národná banka Slovenska 2023

research@nbs.sk

This publication is available on the NBS website

www.nbs.sk/en/publications-issued-by-the-nbs/research-publications

The views and results presented in this paper are those of the authors and do not necessarily represent the official opinion of the National Bank of Slovakia.

Michal Benčík

Abstract

We propose a new method of dealing with the end point problem when filtering economic time series. The main idea is to replace filtered quarterly observations at the end of the sample with static forecasts from a MIDAS regression using higher frequency time series. This method is capable to improve stability of output gap estimates or other cyclical series, as we confirm by empirical analysis on selected CEE countries and the United States. We find that stability may still be violated due to structural breaks in business cycles, or by an excessive amount of short-term noise. While MIDAS regressions have the potential to improve output gap estimates compared to the HP filter approach, the country-specific circumstances play a considerable role and need to be considered.

JEL Codes: C22, E32

Keywords: end point problem, output gap, MIDAS regression

1 Introduction

Defining equilibrium level of macroeconomic variables is a key to any mean-reverting models. Still, vast majority of literature relies on simplistic extraction of trend, either using long term weighted averages or HP filters, although these are commonly understood to suffer from range of inaccuracies. The aim of this paper is to address this problem; suggest, test, and assess other options of filtering techniques that could potentially improve both consistency and the accuracy of equilibrium level of selected macroeconomic variables.

In this paper, we focus on filtering the gross domestic product (output) mainly for purposes of monetary policymaking. We will argue on one hand that existing simple methods have serious drawbacks that can lead to bias in monetary policy and on the other hand, that the multi-equation methods can be tricky to implement due to their complexity. We present a novel and relatively simple method of tying the output gap (the indicator of business cycle) to monthly variables. This new method stabilizes the estimates of output gap at the end of the sample (the most important part of the sample for policymaking).

To be more specific, let's consider the following example. While of the inflation target supposed to be reached in the long term is set by the monetary authority (central bank), the equilibrium level of output evolves in time and needs to be determined from the series of actual output (and possibly, other variables). There are qualitative definitions of potential output, for example, that it is the highest level of output that does not accelerate inflation, simpler assumptions are often used, the most popular is the premise that potential output evolves smoothly and coincides with actual output in the long run. Among the simplest and widely used methods is the Hodrick-Prescott (HP; 1997) filter, using the assumptions about smoothness and long run convergence to actual output. Its main advantage is ease of use, requiring only the choice of one parameter, defining the smoothness of filtered trend (potential output). There are various multivariate methods of computing output gap, but they use more complex relations with many parameters. These parameters often deviate from their expected values, in data for the studied economy is imperfect. The use of multivariate methods thus often opens its own group of problems and this fact decreases the reliability and convincing power of these methods, leading to use of HP filter or similar simpler methods.

The most important drawback of HP filter is the so-called end point problem. We think that from the perspective of policymakers in central bank, the end point problem is of greatest concern. It manifests itself firstly in instability of values of output gap at the end of the cycle and secondly, by the convergence of the series of potential output to actual output at the end of the sample. Note that the end of the cycle is typically the most important period for correct setting of monetary policy.

Gross domestic product is not the only variable linked to output gap, there are many monthly indicators, available with a smaller publication gap, that measure the slack in the economy. It thus makes sense to tie the output gap to these indicators. The simplest and parsimonious method to do so is Mixed data sampling or MIDAS (Ghysels et al., 2004). It is a generalization of linear regression in two aspects: it includes regressors with higher periodicity and it restricts the parameters of higher frequency regressors to be a certain function of lag. This allows for the number of variables to be greater than the number or estimated parameters.

In this paper, we design a method for stabilizing the output gap values at the end of the sample using forecasts from Midas regressions with monthly regressors. As we will demonstrate, this method lowers the variability of output gap estimates, when new observations are added to the

filtered output series, if the output gap is “well behaved”, i. e. it has a form of wave-like oscillations with the period typical for business cycle (3 to 8 years).

We apply our method to eight countries in Central Europe and the Baltics (Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovenia and Slovakia) and to the United States. Our method efficiently improves the stability of output gap estimates for all countries but Latvia and Slovakia.

We base our research on three strands of literature: (i) the papers about multivariate methods cited above show that the output gap is related to various macroeconomic variables, mainly surveys, (ii) Ostapenko (2022) and Feldkircher et al. (2015) show that the set of these covariates can be reduced to a couple of variables¹ and (iii) Ghysels et al. (2004). Armesto et al. (2010) present a novel kind of constrained regression with mixed frequency data (MIDAS) and Kuzin (2009) shows that MIDAS regressions can be superior to mixed frequency VARs in the short run.

This method can be used in practical policymaking in three ways:

- Using the corrected observations directly as output gap estimates
- When monthly data are available, but quarterly aren't yet, this method can be used to nowcasting the output gap,
- From the comparisons of uncorrected HP filter and MIDAS prediction it can be inferred, in which direction the HP filter estimates will be corrected, when new observations will be added to the sample.

We suppose that the corrected output gap estimates will be more useful in simpler models describing the short term impact of business cycle on other variables (for example Phillips curve) than in more sophisticated models, results of which depending upon their rich structure (for example DSGE models).

The rest of the paper is organized as follows: Part 2 summarizes drawbacks of HP filter and proposed solutions for the end point problem in the literature, Part 3 presents the nouvelle method, Part 4 contains empirical analysis, Part 5 presents the results and Part 6 concludes.

2 Drawbacks of HP filter and their solutions in the literature

Despite its intuitive and easy use, the HP filter is far from ideal. Hamilton (2017) identifies three issues: (i) HP filter produces series with spurious dynamic relationships, not based in the data generating process; (ii) estimates in the end of the sample are different from the observations from the center and (iii) value of smoothing parameter lambda is not supported by data. Mohr (2005) is more specific about the first point, how HP filter may impose properties of results in odds with the reality: (a) Objective function as a weighted sum of squares; (b) stochastic trend is a second order random walk and (c) cycle is proportional to the fourth difference of the trend.

¹ Latter paper deals with GDP growth rate. This variable, however, determines the majority of short-term variance of output gap, because the short term variance of potential output is smaller than the variance of actual output growth. We thus find the paper relevant to our research.

The end point problem is caused by the second point identified by Hamilton (2017). It manifests itself when output gap is computed repeatedly with gradually expanding sample. When a new observation is added to the sample and the output gap recomputed. Its values for the original observations often change. An example, how the weight matrix of HP filter changes, when new observations are added is in the Appendix 1. Sometimes, these changes arise from data revisions, but in most cases, the output gap changes ex post even if the original data remain the same. As we will demonstrate, these changes stem from changing weighting matrix as the new observation is included. The observations of HP filtered potential output from the middle of the sample and from the end of the sample are not created equal. While the last observation of potential product is a weighted average of the observations from the past only, an observation from the middle of potential output series is a linear function of both the past and future (leading) observations. As new observations are added, the former last observation will become a function of newer and newer observations, and this creates instability of output gap estimates.

There have been various suggestions for solving the problems with HP filter. The least radical are the end of sample correction. Bruchez (2003) suggests changing the parameter λ at the end of sample. This makes the weights and resulting output gap estimates somewhat more stable at the end of the sample.

Other authors propose to change the filtering process more radically. Wolf et al. (2020) propose one sided HP filter (using only the last observation from each data vintage) and stretching by linear transformation. Stretching makes the amplitude of the output gap from this filter comparable to that of original HP filter and the filter does not change ex post per definition. However, this approach might even backfire, because more weight is attributed to the observations most affected with the end point problem. Phillips and Z Shi (2021) propose so called boosted HP filter, e.g., apply the HP filter several times (replace actual output series with the cycle estimate) till the cycle is stationary. This procedure heavily depends on the character of the data and number iterations may change when observations are added. This is caused by the fact that stationarity tests for output gap are not robust and their outcome can change materially, when the added observation contains a shock. This, however, introduces abrupt changes into the resulting trend series, so that this method is unsuitable for repeated filtering, as done in the policymaking process. Finally, the whole minimalization problem of HP filter can be generalized and reformulated as in Mohr (2005), who uses explicit modeling of the cycle by means of rotation matrices and generalizes the trend model. By these extensions, the filtering becomes as complicated as using the multivariate methods.

Hamilton (2017) recently proposed an alternative to HP filter using regression of log output on its past values (lags 8 to 11 quarters). According to him, the new regression filter eliminates the shortcomings of HP filter introducing no major downside. Jönsson (2020) supports this conjecture by analysis of long time series for the US, where he demonstrates superior stability of the Hamilton filter over the HP filter mainly by documenting heavy revisions after only few more observations are added. As a response, Hodrick (2020) shows that this argument is valid for some series, as random walks, but not for series with non-stationary stochastic trend and longer cycles. Further support to validity of the HP filter comes from Franke and Kukacka (2023), who show that Hamilton filter is outperformed by HP filter on artificial data. They suggest using HP filter with much higher smoothing parameter than usual. Further to that, a study by Hall and Thomson (2020) concludes that (Hamilton filter) produces exaggerated volatilities and less credible trend movements during periods, when the business cycle is turning (2007 peak and 2009 through) . Apart from simulations with time series, the Hamilton filter basically decides whether a shock is a part of the trend or cycle based on its timing. For recent shocks (newer than two years), there is no way in standard Hamilton filter to account them as a part of trend, they are always a part of the cycle. This can be problematic, because (i)

as time goes by, part of the shock can be attributed to trend (there is a version of Hamilton filter where this always happens) or (ii) there can be consensus among economists, based on certain narrative, that certain shock affects supply and shall be viewed more as a part of the trend.

Last but not least, very popular way to eliminate the undesirable properties of simple methods of time series detrending are methods using multiple time series – for Slovakia for example Gylánik and Huček (2008), Ódor and Jurášeková-Kucserová (2014), Ostapenko (2022), Mičko et al. (2023) and Benčík (2019). Advanced time series models, like Mohr (2005) are of similar complexity, even if they use fewer time series and do not mimic the economic theory to the same extent as former methods. The use of these methods, as we already started, is usually not straightforward and the user has to take many relationships and parameters into account.

In order to avoid the aforementioned difficulties with multivariate filtering methods, we focused on methods using the log output series only. To replicate Jönsson (2020), we started our analysis by applying HP, Hamilton filter, asymmetric Christiano-Fitzgerald filter and boosted HP filters to eight transition countries and US, iteratively for all pseudo vintages of data, for which the specific filter was defined. From the resulting output gap estimates, we computed mean absolute revision, e. g. mean of absolute values of differences of output gap observations for the same period from two consecutive pseudo vintages². The empirical results show that Hamilton filter is not clearly superior to 2-sided HP filter (the standard one) neither for transition countries, nor for the US. The Christiano-Fitzgerald filter is less stable than HP filter for most countries as well. Thus, rather than using an alternative method of filtering, we therefore decided to look closer at the end point problem of conventional HP filter and to maximize its stability. The results are shown in Table 1.

Table 1: Mean absolute revisions for output gap estimates from different filters.

Country	Czech Rep.	Estonia	Hungary	Lithuania	Latvia	Poland	Slovenia	Slovakua	U.S.
All									
2s HP	0.46	0.70	0.41	0.59	0.80	0.34	0.45	0.47	0.13
Hamilton	1.43	2.20	1.31	1.89	2.17	0.81	1.28	1.36	0.35
Christiano-F.	0.84	1.75	0.96	1.58	2.22	0.49	1.04	1.12	0.56
w/o 20 pseudo vintag									
2s HP	0.55	0.94	0.46	0.81	1.12	0.38	0.50	0.62	0.14
Hamilton	1.71	2.91	1.46	2.63	3.07	0.92	1.41	1.88	0.38
Christiano-F.	1.05	2.22	1.25	1.91	2.61	0.58	1.35	1.11	0.51
Last 20 pseudo vintag									
2s HP	0.28	0.21	0.31	0.14	0.17	0.25	0.37	0.17	0.06
Hamilton	0.96	0.98	1.06	0.63	0.64	0.64	1.08	0.48	0.10
Christiano-F.	0.43	0.83	0.39	0.93	1.44	0.30	0.42	1.14	0.91
boosted hp	1.25	0.87	0.94	0.33	0.43	0.25	1.09	0.18	0.06

Source: own computations.

Notes: In two-sided HP filter, $\lambda = 1600$. In asymmetric Christiano-Fitzgerald filter bounds 6 to 40 quarters were used, assuming order of integration 0 and removing constant and linear trend. In Hamilton filter, lags 8 to 11 were used. The boosted HP filter was applied according to Phillips and Shi (2019), as the computing procedure is more clearly described in the paper. All values were multiplied by 10^3 for better readability.

² Under pseudo vintages, we understand output gap series, where observations were removed at the end of the sample. For Table 1, the shortest pseudo vintage contains the minimum number of observations, for which the output gap estimate can be computed and depends on the method. The use of pseudo vintages in our novel method is described in Part 3.

3 Applying MIDAS regressions to forecasting of output gap

As we have seen in the previous chapter, HP filter is not defeated by newer single-series methods. On the other hand, methods using multiple equations require the estimation of large number of parameters. For transition countries, this is an important disadvantage, as the time series available are relatively short.

We propose a novel approach, where a regression for the output gap will be estimated and used. As Berger, Morley and Wong (2023) and Ostapenko (2022) show, output gap can be nowcasted with Bayesian mixed frequency VARs, demonstrating that output gap is linked to other higher frequency variables. According to Armesto et al. (2010), GDP can be nowcasted with MIDAS regression, using monthly employment data as a regressor. Feldkircher et al. (2015) identify the industrial production and rate of unemployment as the variables most closely related to output growth in post-socialist countries.

Mixed data sampling (MIDAS) is a special type of regression model, pioneered by Ghysels et al. (2004). Similarly, to mixed frequency VARs, this method combines data with different frequencies and can be used to forecast the variables with the lower frequency using data with the higher frequency. However, this approach uses much less parameters than mixed frequency VARs, because only one equation is estimated instead of whole VAR.

In this chapter, we describe the way we applied this method, utilizing the conclusions in the aforementioned literature to nowcast the output gaps for eight transition countries and United States. We show as well, how the resulting forecasts were evaluated.

3.1 Methodological apparatus

3.1.1 MIDAS regressions in general

MIDAS regression combines data with lower frequency (Y) and data with higher frequency (x), lags of Y and x have different lengths. The lag operator associated with dependent variable Y shifts the observation one quarter or three months back. The lag operator L_{HF} associated with monthly regressors shifts observations one month back. If the estimated equation contains more than three lags of monthly regressors, data from two different quarters are used. For example, consider a regression with a quarterly output gap as dependent variable and unemployment rate with five lags as a monthly regressors. Then, say the observation 2019Q1 of the output gap is determined by values of unemployment rate from March, February and January 2019 and December and November 2018. As defined by Armesto et al. (2010) we can shift the whole structure of the unemployment rate forward (introduce a lead) and perform intra-period forecasting. If the lead is one month, the value of output gap in 2019Q1 depends on April, March, February and January 2019 and December 2018. For model with quarterly and monthly data, new observations can be added once in three months. Adding one month would require estimating and use three models. MIDAS regressions contain three sets of parameters. Parameters β are unconstrained and correspond to quarterly variables. First tier parameters ϕ corresponding to monthly variables are constrained to have a certain shape (as a function of lag of corresponding variables), the shape of the function is determined by the tier parameters θ . Implemented groups of functions include Almon lags (generalized OLS), exponential and beta distribution. With Almon lags lag length can be optimized.

According to Armesto et al. (2010), MIDAS regressions have the general form:

$$Y_t = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \gamma \sum_{k=1}^m \Phi(k; \theta) L_{HF}^k X_t + \varepsilon_t \quad (1)$$

Where Y are quarterly variables, X are monthly variables, L is the quarterly lag operator, L_{HF} is the monthly lag operator, $\phi(k; \theta)$ is the function constraining monthly parameters and ε_t is a random noise.

It should be noted, that the parameters on monthly indicators X could be so scattered that they will not obey the restriction implied by the function $\phi(k; \theta)$. In that case, a variant of MIDAS models called MIDAS-U may be more appropriate. This version of MIDAS estimates the parameters of monthly variables directly, without assumption of their shape. However, from the practical point of view, we have found out that when the parameters in question are too scattered, the estimating procedure selects three lags using three parameters. The only restriction imposed in this case is the number of parameters and lags used, as the original regressors are merely linearly transformed by multiplying them with matrix H . If we used the variant MIDAS-U, we would be able to include more lags at the price of introducing more parameters. Such regressions would not be comparable with the rest, and we thus did not use it.

3.1.2 Almon lag model

As we stated in previous chapter, MIDAS regression imposes restrictions on the parameters of variables with higher periodicity³. These restrictions imply that the parameter follow a certain curve or shape. The restrictions can be implemented in such way, that the parameters follow a polynomial of pre-specified order (most often quadratic – a parabola) or the exponential or beta distribution. As the estimated parameters in our regression change signs for several countries, the methods using distributions are unsuitable for this project. Estimation of parameters of the distribution requires a non-linear optimization, adding further complexity and numerical instability. We thus use the simpler and more general Almon lag model, that allows the parameters to switch signs, as it is basically OLS with transformed variables.

Almon lag model was originally a method of reducing the number of parameters and was used originally in distributed lag models, a class of models unrelated to mixed frequency data. It is described in Batten and Thornton (1983) and the selection of optimal lag length for the function Φ is discussed as well. It restricts the first-tier parameters Φ to be polynomial in lag using second tier parameters θ . If the original model is $y=X\Phi$ then Φ obeys (we assume that parameter vector θ has p element and parameter vector Φ has q elements)

$$\phi_k = \theta_0 + \theta_1 k + \theta_2 k^2 + \dots + \theta_p k^p \quad (2a)$$

In matrix notation,

$$\Phi = H\theta \quad (2b)$$

with H defined as

³ There is a variant of MIDAS without restriction, so called U-MIDAS, but we found the restricted MIDAS have the advantage that the restriction led to inclusion of more observations of the monthly indicators for most countries. with fixed number of parameters. If the estimated parameters did not fit the quadratic parabola well, lag order 3 was selected and in effect, no restriction was imposed.

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 4 & 8 & \cdots & 2^p \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & q & q^2 & q^3 & \cdots & q^p \end{bmatrix} \quad (3)$$

The description of selection procedure for optimal lag length is beyond the scope of this paper and the reader is referred to Batten and Thornton, (1983), who present the procedure of Pagano and Hartley. In practice, the selection functioned in the way that it usually selected longer maximum lags, if the shape of the function Φ approximated the chosen polynomial (quadratic parabola, allowing for U-shaped and \cap -shaped histogram of parameters) well, but chose only three lags, if the first-tier parameters were scattered in a random pattern. As the parabola itself has three second tier parameters, their number was not changed. In effect, no restriction was implied, the parameters were merely transformed in this case. We used the polynomial function of second degree (with three parameters including the constant term, quadratic parabola), which is a often used setup with Almon models.

3.2 Stabilizing the output gap estimates by means of Midas regressions.

3.2.1 Assumptions and goals

In standard situation, actual output oscillates around the potential output in cycles with period up to eight years. The difference of these series, output gap, shall have the form of a wave with some random component. This implies that the output gap series is expected to be serially correlated. Besides, we assume that the output gap is linked to the monthly indicators (unemployment rate and industrial production), especially at the turning points. We will capture the inertia of the business cycle and impacts of monthly indicators in our regression. The oscillations of output gap can be asymmetric, the descent from peak to trough may be faster than the ascent from trough to peak.

The new method shall fulfil three criteria to be successful:

1. It shall reduce the volatility of output gap estimates at the end of the sample.
2. At least some parameters of monthly indicators in regressions shall be significant.
3. The properties of output gap series shall be altered by the method in least possible extent.

These criteria will ensure, that the improvements of output gap estimates will be underpinned by statistically significant linkages between variables and the results of our new method (which will be computed for the end of sample only) will be comparable with the rest of the series of output gap estimates, obtained by the HP filter.

3.2.2 Computing procedure

Transition countries are a group of smaller countries that share common history in many important aspects as a legacy of the past, they have relatively high share of industrial production in the economy and thus this indicator can be used to capture short-term changes of economic activity. Common history and relatively high importance of industrial production are the reason, why we applied our new method to a group of transition countries. Data for these countries starts around 1995, so the samples for individual transition countries are relatively short. In order to explore the behaviour of our method in longer samples, we applied

our method to United States as well. The data starts in 1947 there, so there are several hundreds of observations, and our new method will manifest its large sample properties for that country.

We have replicated this procedure for selected CEE countries and the U.S. (Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, Slovenia, and the United States). The sample varies due to availability of the data, but mostly begins around 2000 and ends in 2022.

We undertake the computing procedure in the following steps:

1. Define a pseudo vintage⁴ of log output series (1st iteration – observations 1 to T-20)
2. Apply two-sided HP filter,
3. Truncate the sample (cut first six and last nine observations; truncation at the end of the sample corresponds to Covid-19 pandemic)
4. Estimate a Midas regression:

$$g_t = c_0 + c_1 g_{t-1} + \sum_{k=0}^{p_1} \Phi_{1k}(\theta_1; k) \text{dlip}_{t-k} - 1 + \sum_{k=0}^{p_2} \Phi_{2k}(\theta_2; k) \text{ur}_{t-k+1} \quad (4)$$

with the output gap g_t as the dependent variable (and its lag g_{t-1} as the regressor), dlip is change in log industrial production (seasonally adjusted, where it was necessary and ur is unemployment rate. This regression is estimated for every country separately (country index is omitted for simplicity). Note that beside the lag structure imposed by the Almon model, industrial production enters the regression with one month lag and unemployment rate with one month lead. This is necessary, as the industrial production is a leading indicator for the countries in question and unemployment rate is a lagging indicator.

5. Compute forecasts (static in the preferred version) for last nine observations using the estimated regression, store results in matrices,
6. Add one quarterly observation to the pseudo vintage, go to Step 1. In Step 4, three corresponding monthly observations are added to the series of monthly regressors X . repeat till all observations are used in Step 2.

Revisions (differences between estimates for the same period from two consecutive pseudo vintages), mean compensated revisions (arithmetic mean on non-missing revisions), mean absolute revisions (arithmetic mean of absolute values of non-missing revisions)⁵ and median revisions are computed from the stored forecasts for every model and country. Country specific mean absolute revisions are averaged. The reduction of mean absolute revision is the most important criterium for success of our method, as stated in Part 3.2.1.

⁴ Pseudo vintages are created from the last vintage on data series (newest published data) by omitting certain number of observations at the end of the sample. They do not take revisions of output series into account. We have chosen this way, because we are primarily concerned with instability of output gap estimates caused by adding observations, rather than by revisions of past data.

⁵ Let $g_{HP,t,i}$ be output gap estimate for period t and pseudo vintage i (ranges of t and i are implied by the computing procedure) and $g_{HP,t,i-1}$ be the estimate for the same period, but for previous pseudo vintage. mean absolute revision is defined as

$$MAR_{HP} = \frac{1}{N} \sum_{t,i} |g_{HP,t,i} - g_{HP,t,i-1}|,$$

Where the sum goes over all non-missing values of the difference in absolute values, N is the number of non-missing values. Country index was omitted for clarity. Mean absolute revisions for forecasts from Midas regressions were computed analogously. For every pseudo vintage, eight observations of revisions were computed, as the forecast interval shifts one observation forwards in every iteration.

In search of optimal model specification, we used several alternative specifications, with variety of available variables (i.e., exports, major purchases, economic sentiment indicator, order books or confidence indicators) and alternative functional forms apart from the AR term. From among these alternatives, in terms of the three conditions stipulated in section 3.2.1, specifications with the unemployment rate and industrial production produced the best results. Some other specifications were more efficient in reducing the mean absolute revision of output gap estimates but failed due to insignificant parameters. The cause of that is non-linearity and/or instability of the relationship between output gap and these variables.

4 Results

4.1 Main results

We report computed mean absolute revision for all countries. The mean absolute revisions are averages of absolute values of differences of output gaps for the same time period using two consecutive pseudo vintages. We find this measure more suitable than a quadratic measure similar to root mean square error, because mean absolute revision is linear and the trade-off between rise in one revision and drop in another is constant and equal to unity. Mean absolute revisions are used to evaluate the stability of output gap estimates by Hulej and Grabek (2015)⁶. The lower the mean absolute revision, the less the output gap estimates change after adding new observations. In our settings, reduction of mean absolute revision thus indicates improvement in the stability of output gap estimates at the end of the sample. It is the indicator showing how successful or unsuccessful our new method is, so we report it for all countries. We present them for all countries in Table 2, both for 2-sided HP filter (second column) and the preferred MIDAS forecast model version (as specified in Step 4 of computing procedure; third column). In the last column of Table 2, we present the reduction of mean absolute revision⁷. Comparison of mean absolute revisions for HP filter and MIDAS forecasts shows that the latter estimates are more stable most of the time, but the efficiency of Midas projection varies strongly among countries. It should be noted that for Latvia, the preferred version of MIDAS projection yielded higher mean absolute revision than the HP filter and for Slovakia, the mean absolute error are practically the same. The cause is that the data for these countries did not obey the assumptions of our method. This topic is discussed in Part 4.2 in detail.

Table 2: Reductions of mean absolute revisions⁸.

Method\Country	Czech Rep.	Estonia	Hungary	Lithuania	Latvia	Poland	Slovenia	Slovakia	U.S.
HP 2s	2.53	1.85	2.86	1.34	1.56	2.28	3.3	1.65	1.58
Midas	2.06	1.28	2.48	0.94	1.83	1.61	2.63	1.62	1.18
Reduction_%	18.5	31.3	13.3	29.5	-17.1	29.3	20.5	1.7	25.2

Source: own computations

Note: Values of mean average revisions were multiplied by 10^3 for better readability.

⁶ They call the indicator average absolute revision.

⁷ The relative reduction for country X is defined as

$$reduction_ \%_X = 100 \times (MAR_{X,HP} - MAR_{X,MIDAS}) / MAR_{X,HP}$$

Where $MAR(\cdot)$ is mean average revision.

⁸ The mean absolute revisions for HP filter are computed only for those observations, where their counterparts from Midas forecasts exist. They are thus different from the results from Table 1, which were computed over all available observations.

In the following graphs 1a-b to 9a-b, we illustrate the end point sample estimates for HP filter and the preferred Midas projection for all countries covered in this paper.

It is evident from the graphs, that the output gap estimates from HP filter move in a wider band than those from MIDAS projection. Especially observations for the year 2019 from HP filter change rapidly depending upon the inclusion of the observations corresponding to Covid-19 pandemic.

It is also evident from the graphs that the estimates from MIDAS regressions change less for most countries. Thus, the MIDAS regressions partly alleviate the end point problem, to the extent given in the last row of the Table 2.

. In order to find out, how the stability of the output gaps estimates from the HP filter and Midas forecasts change in time, we computed the mean absolute revision for observations that were gradually added to the sample in the computing procedure. The mean is computed across pseudo vintages. For most countries, mean absolute revision for MIDAS forecast is lower for all observations, but for Latvia and Slovakia there are quarters, where it is lower for HP filter. The series of the mean absolute revisions are depicted in Figures A2 to A10 in Appendix 4. The spike in variability at the end of 2019 is often visible, but in general, the relative efficiency of Midas forecast neither worsens nor improves during the Covid-19 pandemic.

However, there are three problems with the estimates from MIDAS regressions:

1. The monthly indicators contribute with additional noise in the estimates. This noise does not vanish ex post, by adding new observations,
2. The inclusion of the AR terms in Midas regressions may cause that troughs (and peaks) of business cycle occur with a small lag in the estimates from MIDAS regression, compared to HP filter. Despite this problem, we find the inclusion of the AR term in estimated equations necessary, we would have an omitted variable problem, if we dropped them.

The amplitude of output gap fluctuations from MIDAS regressions can be different from the amplitude from HP filter. However, the strong shocks affecting economies in question are highly atypical and country-specific and may be connected to this problem.

Figures 1 – 9: End of sample estimates

Figure 1a: HP filter, Czech Republic

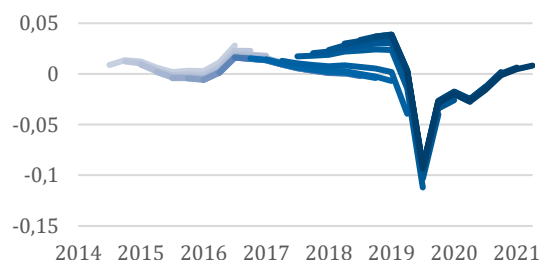
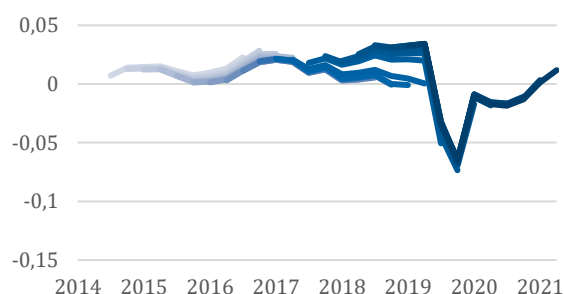


Figure 1b: MIDAS forecast, Czech Republic



Source: own computations.

Figure 2a: HP filter, Hungary

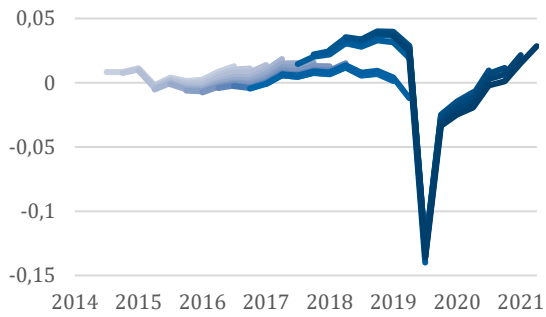
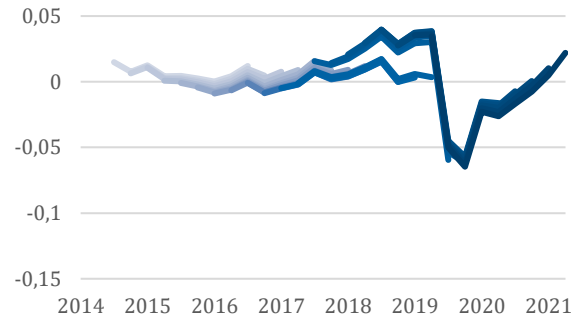


Figure 2b: MIDAS forecast, Hungary



Source: own computations.

Figure 3a: HP filter, Slovakia

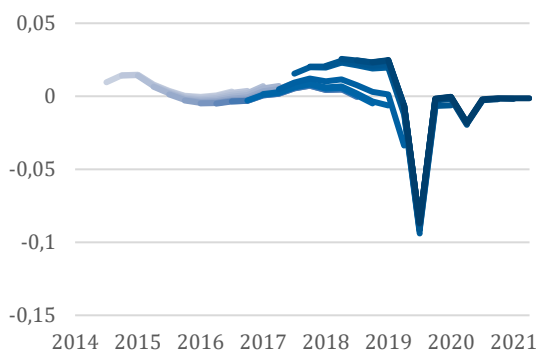
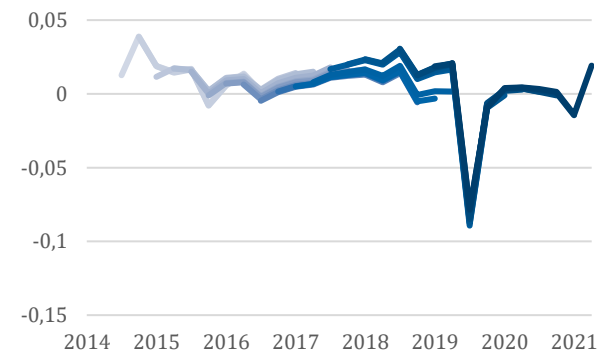


Figure 3b: MIDAS forecast, Slovakia



Source: own computations.

Figure 4a: HP filter, United States

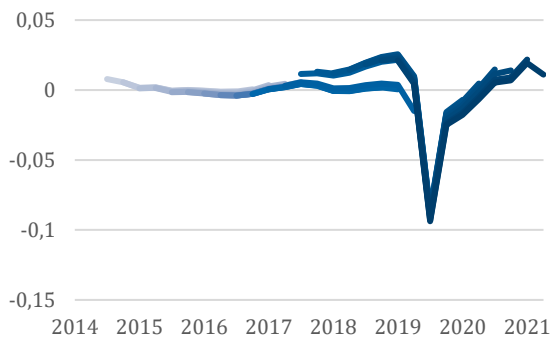
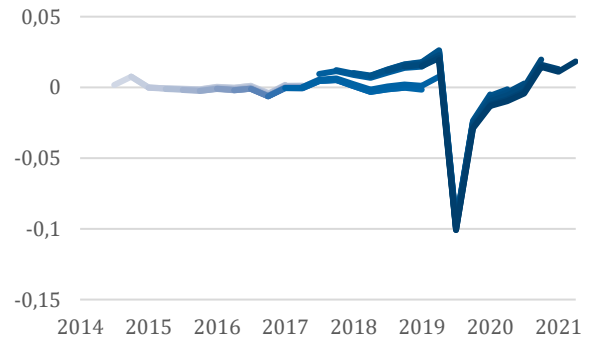


Figure 4b: MIDAS forecast, United States



Source: own computations.

Figure 5a: HP filter, Lithuania

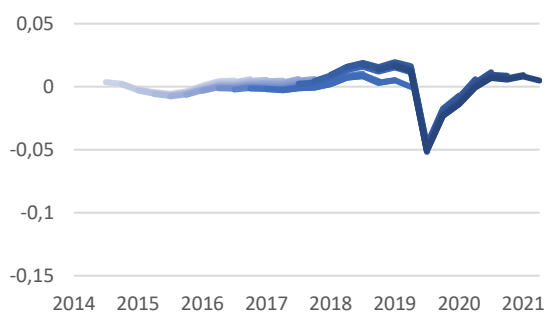
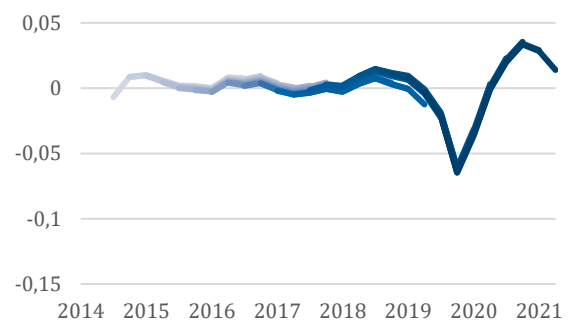


Figure 5b: MIDAS forecast, Lithuania



Source: own computations.

Figure 6a: HP filter, Latvia

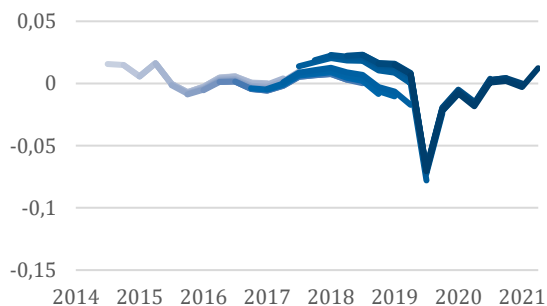
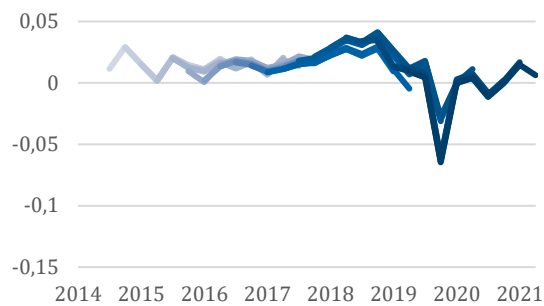


Figure 6b: MIDAS forecast, Latvia



Source: own computations.

Figure 7a: HP filter, Slovenia

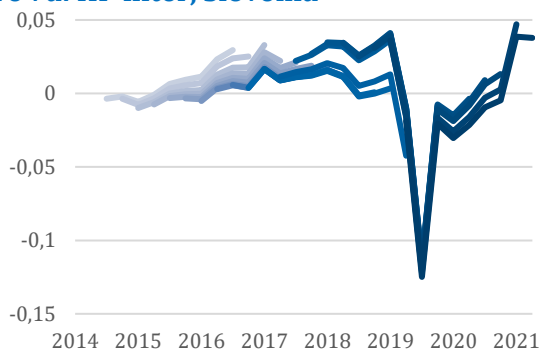


Figure 7b: MIDAS forecast, Slovenia

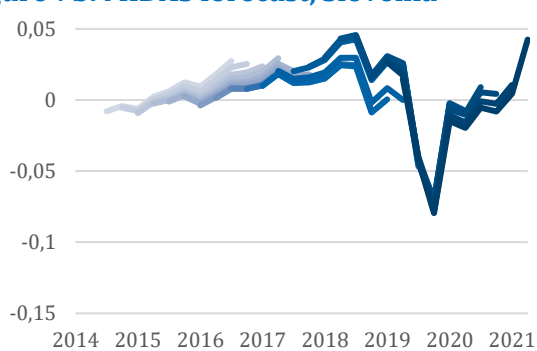


Figure 8a: HP filter, Estonia

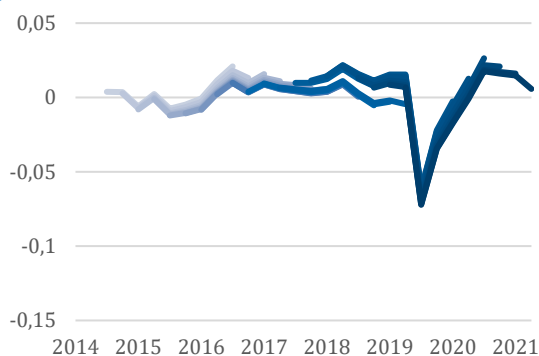
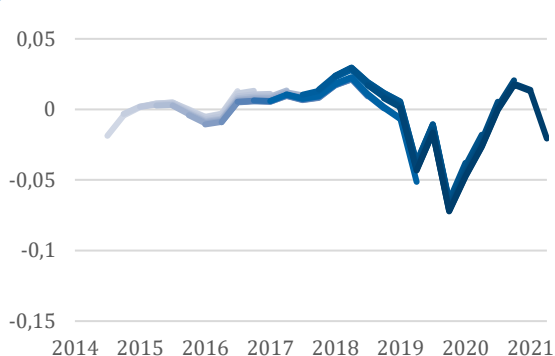


Figure 8b: MIDAS forecast, Estonia



Source: own computations.

Figure 9a: HP filter, Poland

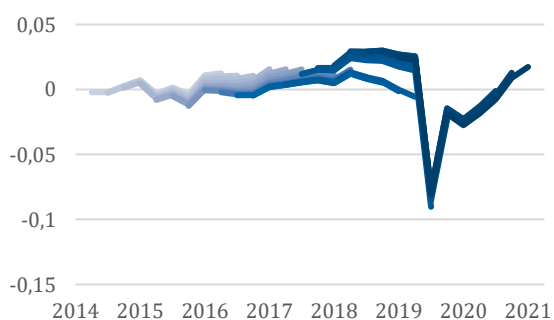
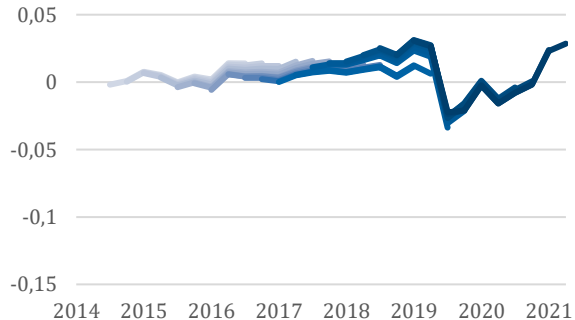


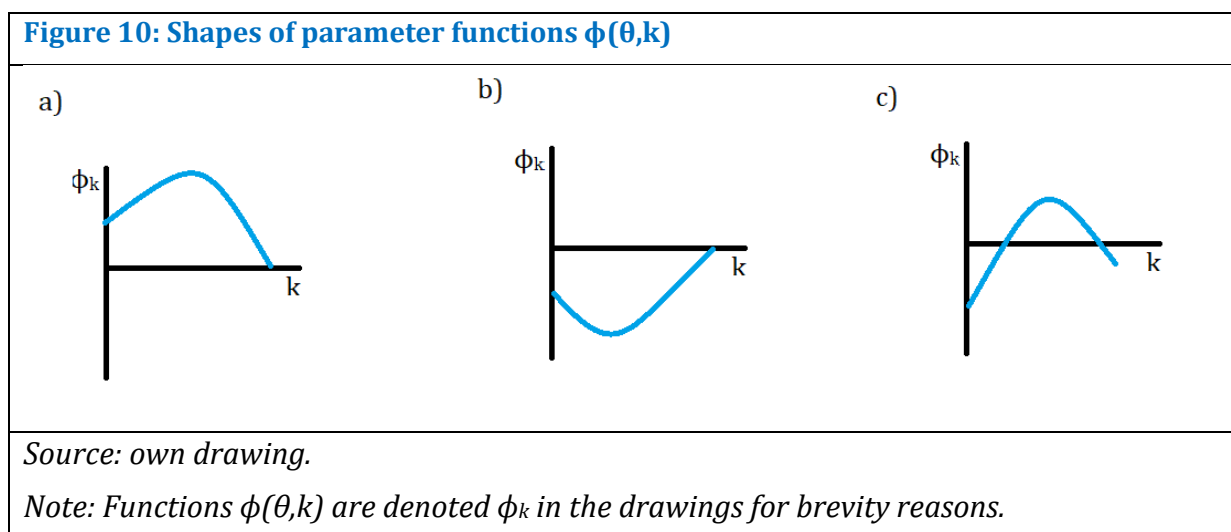
Figure 9b: MIDAS forecast, Poland



Source: own computations.

The statistical parameters for the underlying MIDAS regressions are in the Appendix 2. These are the estimates up to in 2019Q4 (observations from 2020 and later are removed to avoid the contamination of the sample by the Covid-19 period). To explore the fit of the equations with shorter samples, we saved the coefficients of determination from all estimators for every country. We present the minimum, maximum and final coefficient of determination for all countries in Appendix 5. The values do not change much for all countries except for Poland. The atypical nature of the business cycle in Poland is the most likely reason. But in general, the coefficients of determination show that the estimates for shorter samples are not significantly worse than the final estimates presented in Appendix 2.

The parameters ϕ_k measure the unit effect of specific monthly variable on the output gap value. The lag polynomials for industrial production contain three to six such terms. As we use MIDAS with restrictions, these parameters are generated by the function $\phi(\theta, k)$, where θ are the estimated parameters and k is lag index, integer going from 0 to p , $3 \leq p \leq 6$. Formulae for these functions are given in Appendix 2. The lag order p is country and variable specific. This way, they are constrained to follow certain shape. These restrictions are implemented by Almon lag model that restricts the parameters ϕ_k to the shape of quadratic parabola. Three shapes of the function $\phi(\theta, k)$ that were identified most often by our estimation are shown in Figure 10.



The parameters associated with industrial production are defined by functions that have the shape shown in Figure 10a). All parameters are positive, they rise for small k (k I lag index), peak and finally decrease for larger k . This structure is similar to using the average of monthly values in regression. As the industrial production series contains a great share of short-term noise, this structure is of advantage, as the one-month-shocks average out.

The parameters for unemployment rate are defined by functions that have either the shape or pattern shown in Figure 10b or 10c. The case shown in Figure 10b is analogous to Figure 10a, only with opposite sign. This case corresponds to the theoretical assumption that output gap is a function of the level of unemployment rate. The case shown in Figure 10c corresponds to the situation, where the output gap is a function of second derivative, curvature, or acceleration of unemployment rate (see Appendix 3). This interpretation is supported by the fact that the long run impact of unemployment rate of output gap (the sum of corresponding parameters) is near zero.

4.2 Behaviour of Midas forecasts when the assumptions of the model are violated.

We have previously stated that the assumptions of our approach are that

- a) the output series contains the business cycle in form of slow persistent oscillations,
- b) the (monthly rate of change of) industrial production and rate of unemployment are determined predominantly by business cycle factors.

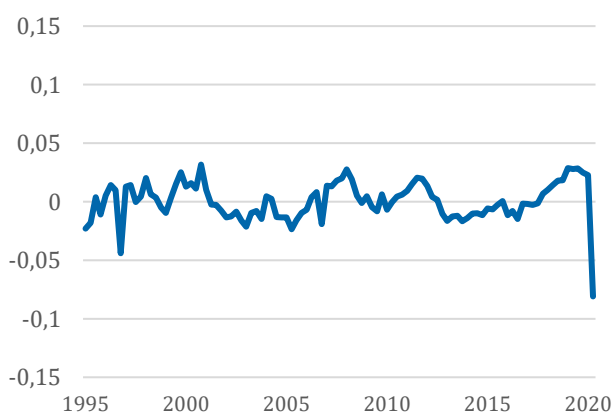
In this part, we discuss what happens if one of these assumptions is violated.

Business cycle contains massive breaks or mostly noise: In these situations, the AR term in the MIDAS regression propagates the shocks and causes error in the forecasts. Consequently, the forecasts of the MIDAS regressions do not improve the quality of HP filter, as indicated by mean absolute revision. The case with missing business cycle frequencies can be illustrated with the case of Poland and the case with massive breaks can be illustrated with the case of Latvia.

The economic development in Poland started from a low basis in the beginning of 1990s and output rose without interruption till the Covid pandemic. This implies that the output gap estimates using HP filters contains a low share of business cycle frequencies (and the cycles are very irregular) and a high share of short-term noise. This leads to insignificant parameters of Midas regressions for this country, but the out of sample forecasts still have lower variability (and uncertainty) than the HP filter.

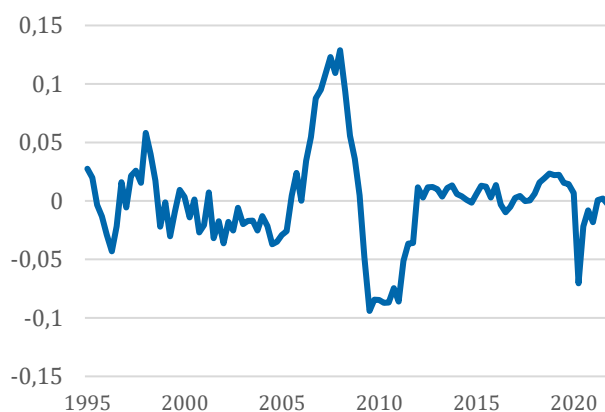
Latvia experienced a strong boom and bust cycle in the late 2000s, followed by an output drop of several tens of percent during the Global Financial Crisis. On the other hand, in several years at the end of the sample, the actual output is mostly at the level of potential. Thus, the series of output gap contains structural breaks, leading to poor performance of forecasts that fail to reduce the volatility of output gap estimates, even when the parameters in Midas regression are significant.

Figure 11a: Output gap estimates Poland



Source: own computations.

Figure 11b: Output gap estimates Latvia



Monthly indicators are materially affected by non-cyclical factors: The level and overall character of monthly indicators changes sometimes due to supply side factors that have a non-cyclical character. In this case, the variable in question is insignificant in MIDAS regressions and the method has decreased efficiency. This happened in Slovakia, where the level of the unemployment rate trended downwards despite temporary swings due to the business cycle.

This was caused mainly by atypical demographic developments. Until 2010 the number of persons in productive age increased steadily, but it decreased since then. We used the unemployment series in its original form in the model for better comparability with other countries, but may adjust the unemployment rate for the demographic decline and estimate the MIDAS model with adjusted series, more closely tied to the business cycle, in the future.

Theoretically, Industrial production can be affected as well. It is not case for Slovakia, nor Czech Republic, but if the structure of the economy in question⁹ changes permanently, for example in such a way, that the share of industrial production decreases and the share of services increases. The changing structure of economy makes the relation between industrial production and output gap unstable in time. On the other hand, our method can cope well with one-off spikes in industrial production, because all parameters associated with this variable are positive and the spikes get averaged out.

4.3 Discussion

The conclusion of computations is that the HP filtered output gap at the end of the cycle can be improved by replacing its values with static forecasts from MIDAS models. This requires, however, that the business cycle takes the form of slow persistent oscillations and that industrial production and the rate of unemployment are determined predominantly by business cycle factors, Tying the output gap to other variables by means of single equation stabilizes the output gap estimates. Thus, despite minor drawbacks, we find using the forecasts from Midas models useful, because they reduce the effects of changing weights in the HP filter (when new observations are added) and show the direction, where the HP filter estimates are likely to converge in the future.

This method can be compared with two other methods of dealing with the end point problem of HP filter: multivariate methods incorporating monthly variables (for example Ostapenko 2022) and using forecasts for output gap as input data in HP filter. Compared with the former, our method has less parameters to estimate and is thus likely to work better with shorter time series. Regarding the latter alternative, as Berger, Morley and Wong (2023) state, using output predictions for alleviating the end point problem works, but these predictions need to be known before filtering and this introduces a new open question into output gap computation. Our method has the advantage that it replaces these unknown predictions with the known monthly indicators.

5 Summary and conclusion

We designed a novel method for reducing the end point problem of HP filter and applied it to eight transition countries and United States. We have chosen transition countries as they share a great deal of common economic history and have relatively high share of industrial production in GDP. Thus, we could expect that the monthly changes in industrial production will be a good explanatory variable for the output gap. We used the U.S. data to picture the procedure using long data history. After computing the HP estimates of output gap, we suggest removing the some observations from the end of the sample (in our case, number of removed observations is chosen so that data from Covid-19 pandemics do not contaminate the

⁹ We computed the analogous MIDAS regression for UK (unreported in the paper), where the industrial production was insignificant all together.

estimation sample, but it can be chosen so that it comprises observations deemed biased by the end point problem by the researcher), fitting a MIDAS regression to the rest and replacing the removed observations with static forecasts from the MIDAS regression. The MIDAS regression contains a naïve forecast (extrapolation) of output gap and two monthly indicators in several lags as regressors. For our preferred specification, those monthly indicators are monthly growth rate of industrial production and unemployment rate. The sample was truncated at the end and static forecasts were computed. Replacing the observations from the end of the sample with the forecasts from this regression reduces the mean absolute revisions by tens of percentage points for the majority of countries. We found that under the assumptions, that the output series contains a business cycle in form of slow persistent oscillations and that industrial production and rate of unemployment are determined predominantly by business cycle factors (they are not contaminated by non-cyclical shocks), the HP filtered output gap at the end of the cycle can be improved by replacing its values with static forecasts from MIDAS models. Beside reducing the variability of output gap estimates, we made sure that our preferred model does not alter the character of output gap series in material extent. Thus, tying the output gap to monthly indicators, that are available with lower lag than the quarterly data, helps to reduce the uncertainty of output gap estimates in the time period most relevant for adequate setting of monetary policy.

This method can be used in three purposes:

- For computing estimates of output gap estimates less affected by the end point problem than conventional HP filter,
- nowcasting in absence of quarterly data, or finally,
- predicting the future course of HP filter estimates when new observations are added.

Further research will be needed to augment this method for the cases when the assumptions of the model are violated. Most notably, non-cyclical component of unemployment rate needs to be addressed. This, however, is better done case-by case for affected countries.

Literature

- [1] Armesto, M. T., Engemann, K. M., and Owyang, M. T. (2010): Forecasting with Mixed Frequencies, *The Federal Reserve Bank of St. Louis Review*, Volume 92, (Nov.), pp. 521-536.
- [2] Batten D. S. And Thornton, D. L. (1983): Polynomial Distributed Lags and the Estimation of the St. Louis Equation, *The Federal Reserve Bank of St. Louis Review*, Volume 65, (Apr.), pp. 13-25.
- [3] Benčík, M. (2019): Construction of a Survey-based Measure of Output Gap, [Working Paper 3/2019], NBS, Bratislava
- [4] Berger, T., Morley, J., Wong, B. (2023): Nowcasting the output gap, *Journal of Econometrics*, Volume 232, Issue 1, 18-34.
- [5] Bruchez, P.-A. (2003, August). A modification of the HP filter aiming at reducing the end-point bias. Working paper, Swiss Federal Finance Administration, Bundesgasse 3, CH-3003 Bern.; Bruchez PA (2003): A Modification of the HP Filter Aiming at Reducing the End-Point Bias. Working Paper Swiss Federal Finance Administration, ÖT/2003/3.)
- [6] Feldkircher, Martin & Huber, Florian & Schreiner, Josef & Tirpák, Marcel & Tóth, Peter & Wörz, Julia. (2015). Bridging the information gap: small-scale nowcasting models of GDP growth for selected CESEE countries. *Focus on european economic integration*. 56-75.
- [7] Franke, R., Kukacka, J. and Sacht, S. (2023):, Is the Hamilton Regression Filter Really Superior to Hodrick-Prescott Detrending? (2023).
- [8] Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004): The MIDAS Touch: Mixed Data Sampling Regression Models, [Cirano Working paper 2004s-20], CIRANO, Montreal.
- [9] Gylánik, M. and Huček, J. (2009): Odhad NAIRU v slovenskej ekonomike, *Biatec*, 17, No 12, 13-18.
- [10] Hall, V.B. and Thomson P. (2021): Does Hamilton's OLS regression provide a "better alternative" to the Hodrick-Prescott filter? A New Zealand Business Cycle Perspective, *Journal of Business Cycle Research*, Vol. 17, No 2, pp. 151-183, November.
- [11] Hamilton, J. D. (2018): Why You Should Never Use the Hodrick-Prescott Filter, *The Review of Economics and Statistics*, Vol. 100, No 5, pp. 831-843.
- [12] Hodrick, R. J. (2020): An Exploration of Trend-Cycle Decomposition Methodologies In Simulated Data, [Working Paper 26750], NBER, Cambridge.
- [13] Hodrick, R. J., and E. C. Prescott (1997). "Postwar US business cycles: an empirical investigation," *Journal of Money, Credit, and Banking*, Vol. 29, issue 1, 1-16
- [14] Hulej, M. and Grabek, G. (2015): Output Gap Measure Based on Survey Data, [NBP Working Papers No 200], NBP, Economic Institute, Warsaw.
- [15] Jönsson, K. (2020): Real-time US GDP gap properties using Hamilton's regression-based filter, *Empirical Economics*, Vol, 59, No 1, pp.307-314.
- [16] Kuzin, V., Marcellino, M. And Schumacher, Ch. (2011): MIDAS vs. Mixed-Frequency VAR:
- [17] Nowcasting GDP in the Euro Area, *International Journal of Forecasting*, Elsevier, Vol. 27, No 2, pp. 529-542.
- [18] Mičko, T., Karšay, A., Múčka, Z., Šrámková, L. (2023): Closer to Finding Yeti (Rethinking the output gap methodology through the lens of fiscal policy stance plausibility), Working Paper, CBR Bratislava

- [19] Mohr, M. (2005): A Trend-Cycle(-Season) Filter, [Working Paper Series 499], European Central Bank, Frankfurt am Main.
- [20] Ódor, L. and Jurášeková Kucserová, J. (2014): Finding Yeti: More robust estimates of output gap in Slovakia, [Working Paper No. 1/2014], NBS, Bratislava.
- [21] Ostapenko, N. (2022): Do output gap estimates improve inflation forecasts in Slovakia? [Working Paper 4/2022], NBS, Bratislava.
- [22] Phillips, P. C. B. and Shi, Z. (2019): Boosting the Hodrick-Prescott Filter, [Cowles Foundation Discussion Paper No. 2192], Yale University, New Haven.
- [23] Phillips, P. C. B. and Shi, Z. (2021): Boosting: Why You Can Use the HP Filter, *International Economic Review* 62/2, pp. 521-570
- [24] Sess Adiabouah, A. (2020): Revisiting the Hodrick-Prescott Filter: A New Perception of the Minimization Problem, *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, Volume 64, No 1, pp 179-186.
- [25] Wolf, E., Mokinski, F. and Schüller, Y. (2020): On adjusting the one-sided Hodrick-Prescott Filter, [Discussion Paper No 11/2020], Deutsche Bundesbank, Frankfurt am Main.

Appendix 1 An example of genesis of end point problem

HP filter is a linear filter, every observation of filtered series is a linear combination of all observations of the actual series. Filtered trend f is a product of weight matrix W and actual series y :

$$f = Wy \tag{A1}$$

The matrix W is defined as

$$W = (I + \lambda KK')^{-1} \tag{A2}$$

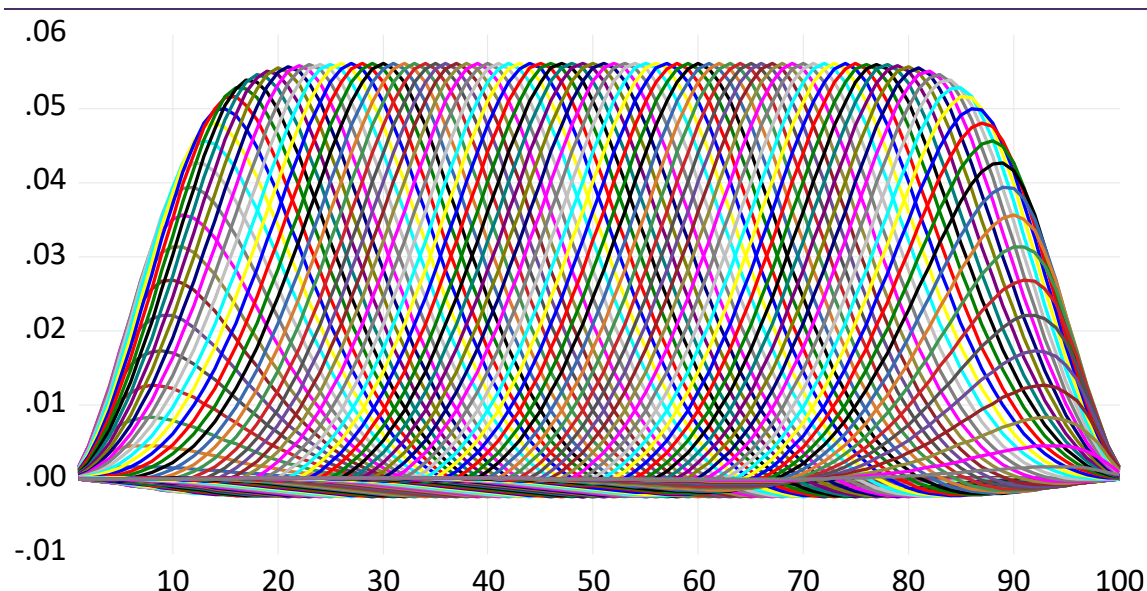
K is an $n \times n+2$ difference matrix (n is the number of observations), its elements are defined as follows: $k_{ij}=1$ for $j=i$ and $j=i+2$, $k_{ij}=-2$ for $j=i+1$ and $k_{ij}=0$ otherwise.

Thus, it has the structure:

$$K = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix} \tag{A3}$$

As an example, graph of a complete weight matrix for $n=100$ and $\lambda=1600$ is in Figure A1. From the graph two problematic features of HP filter are evident: first that the weight matrix implies different character of observations at the beginning and end of the sample on one hand and in the middle of the sample at the other hand, as stated in Hamilton (2017); and second that there are negative weights as well, as stated in Bruchez (2003).

Figure A1: Graph of complete weight matrix for HP filter for $n=100$ and $\lambda=1600$



Source: own calculations.

Note: Lines represent columns, horizontal axis labels are rows

Note that the weight matrix contains negative weights as well. Wolf et al. (2020) list several papers, where weights of HP filter or their approximations are analytically derived. Due to Equation 1, the elements of matrix W can be interpreted as derivatives of the observation of

filtered series with the row index i with respect to observation of actual series with column index j :

$$w_{ij} = \frac{\partial f_i}{\partial y_j} \quad (A4)$$

From the point of view of practical policymaking, it is important that the elements matrix W change when new observations are added. In order to illustrate this, parts of the weight matrix for varying number of observations and $\lambda = 1600$ are in Table A1.

Table A1: Evolution of weights with increasing number of observations

n	i	j=80	j=81	j=82	J=83
80	80	0.0005	0	0	0
81	80	0.0021	0.0009	0	0
81	81	0.0009	0.0005	0	0
82	80	0.0049	0.0030	0.0012	0
82	81	0.0030	0.0021	0.0009	0
82	82	0.0012	0.0009	0.0005	0
84	80	0.0133	0.0105	0.0072	0.0041
84	81	0.0105	0.0087	0.0062	0.0036
84	82	0.0072	0.0062	0.0049	0.0030
84	83	0.0041	0.0036	0.0030	0.0021
90	80	0.0416	0.0391	0.0356	0.0312
90	81	0.0391	0.0378	0.0351	0.0313
90	82	0.0356	0.0351	0.0335	0.0305
90	83	0.0312	0.0313	0.0305	0.0287
100	80	0.0557	0.0549	0.0530	0.0502
100	81	0.0549	0.0554	0.0546	0.0526
100	82	0.0530	0.0546	0.0551	0.0542
100	83	0.0502	0.0526	0.0542	0.0546

Source: own calculations.

Note: n is number of observations (dimension of the weight matrix is $n \times n$), i is the row index and j is the column index.

Table 2 simulates a case realistic in many transition countries. We begin with 20 years of data, corresponding to 80 observations. Thus, we begin with 80×80 weight matrix and observe the weight of the last actual observation in the last filtered observation. It is given in the row of the Table 2 for $n=80$ and $j=80$. Next, we add an observation to the sample, so the weight matrix is now 81×81 . We see now that the weight of 80-th actual observation in 80-th filtered observation changed from 0,0005 to 0.002078 and a weight for the 81-th actual observation equal to 0.000888. There are 81 filtered observations now instead of 80, and the weights for last two actual observations in the last filtered observation are now 0.000888 and 0.0005 respectively, as given by the row of Table 2 corresponding to $n=81$ and $i=81$. Analogously, we extend the weight matrix, so that it has 82, 84, 90 and 100 rows (corresponding to a filtering problem for a series with the same number of observations). We present the weights of the actual observations 80-83 in the filtered observations 80-83.

It is evident from the table, that the weights for older actual observation change to a great extent, so the impact of older shocks in the filtered series can change, although often the filtered series is just getting smoother. Furthermore, weights for the same observations vary to a great extent, when new observations are added to the sample. These weights do not depend on actual data (as it is shown by Equation 2). Thus, when the added observations contain shocks, these shocks will be inevitably introduced into older observations of the filtered series and the value of those observations will change. We find these ex-post changes in the weight matrix, as well as step-by-step inclusion of new observations and accompanying changes in the weight matrix W to be the most important cause of instability of two-sided HP filter.

Appendix 2 Parameters of MIDAS regressions

Table A2: MIDAS regression parameters, Czech Republic

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Dependent Variable: Output_gap				
Sample (adjusted): 2000Q3 2019Q4				
Included observations: 78 after adjustments				
Quarterly parameters β :				
C	0.0067	0.0026	2.550	0.013
Output_gap(-1)	0.8949	0.0388	23.056	0.000
Monthly parameters θ				
Unemp_rate(+1)/100				
$\theta(1,1)$	1.04278	1.03188	1.01056	0.31570
$\theta(1,2)$	-1.87957	1.17313	-1.60218	0.11360
$\theta(1,3)$	0.48349	0.25657	1.88443	0.06370
Indust_prod(-1)				
$\theta(2,1)$	-0.01645	0.06205	-0.26509	0.79170
$\theta(2,2)$	0.09040	0.04390	2.05938	0.04320
$\theta(2,3)$	-0.01401	0.00630	-2.22449	0.02930
Monthly Parameters $\varphi(k)$				
Unemp_rate(+1)/100	Lag	Coefficient		
Lags: 4	0	-0.3533		
	1	-0.7824		
	2	-0.2445		
	3	1.2604		
Indust_prod(-1)	Lag	Coefficient		
Lags: 6	0	0.0600		
	1	0.1083		
	2	0.1287		
	3	0.1211		
	4	0.0854		
	5	0.0218		
R-squared	0.9536	S. E. of R.	0.00453	
Durbin-Watson stat	1.8270	Sum u ²	0.00144	

Note:

Source: own calculations.

Parameters ϕ_k and θ_{ij} are related by following equations: (defining quadratic parabola and restricting parameters ϕ_k):

$$\text{For unemployment rate: } \phi(\theta, k) = \phi_k = \theta_{1,1} + \theta_{1,2} * (k + 1) + \theta_{1,3} * (k + 1)^2 \quad (\text{A5})$$

$$\text{For industrial production: } \phi(\theta, k) = \phi_k = \theta_{2,1} + \theta_{2,2} * (k + 1) + \theta_{2,3} * (k + 1)^2 \quad (\text{A6})$$

These equations contain a shift compared to Chapter 3.1.2, but it does not change the functioning of Almon lag model significantly.

Table A3: MIDAS regression parameters, Estonia

Dependent Variable: Output_gap
Sample (adjusted): 2000Q3 2019Q4
Included observations: 75 after adjustments
Quarterly parameters β :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0125	0.0060	2.102	0.0393
Output_gap(-1)	0.7941	0.0675	11.768	0.000

Monthly parameters θ

Unemp_rate(+1)/100				
$\theta(1,1)$	-17.39750	5.26226	-3.30608	0.00150
$\theta(1,2)$	16.73151	5.27752	3.17034	0.00230
$\theta(1,3)$	-3.26727	1.05821	-3.08754	0.00290
Indust_prod(-1)				
$\theta(2,1)$	0.03624	0.11920	0.30400	0.76210
$\theta(2,2)$	0.22055	0.12984	1.69865	0.09400
$\theta(2,3)$	-0.07347	0.03101	-2.36901	0.02070

Monthly Parameters $\varphi(k)$

Unemp_rate(+1)/100		
Lags: 4	Lag	Coefficient
	0	-3.9332
	1	2.9965
	2	3.3916
	3	-2.7477
Indust_prod(-1)		
Lags: 3	Lag	Coefficient
	0	0.1833
	1	0.1835
	2	0.0367

R-squared	0.9212	S. E. of R.	0.01232
Durbin-Watson stat	2.1934	Sum u ²	0.01232

Source: own calculations.

Table A4: MIDAS regression parameters, Hungary

Dependent Variable: Output_gap
Sample (adjusted): 2000Q3 2019Q4
Included observations: 78 after adjustments
Quarterly parameters β :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0001	0.0025	0.057	0.9546
Output_gap(-1)	0.9571	0.0451	21.223	0.000
Monthly parameters θ				
Unemp_rate(+1)/100				
$\theta(1,1)$	1.07322	1.30586	0.82185	0.41400
$\theta(1,2)$	-0.90558	1.09093	-0.83010	0.40930
$\theta(1,3)$	0.14896	0.17954	0.82968	0.40950
Indust_prod(-1)				
$\theta(2,1)$	0.00959	0.07299	0.13140	0.89580
$\theta(2,2)$	0.18821	0.06614	2.84587	0.00580
$\theta(2,3)$	-0.04318	0.01312	-3.29080	0.00160
Monthly Pamameters $\varphi(k)$				
Unemp_rate(+1)/100		Lag	Coefficient	
Lags: 5		0	0.3166	
		1	-0.1421	
		2	-0.3028	
		3	-0.1657	
		4	0.2694	
Indust_prod(-1)		Lag	Coefficient	
Lags: 4		0	0.1546	
		1	0.2133	
		2	0.1856	
		3	0.0716	
R-squared	0.9034		S. E. of R.	0.00569
Durbin-Watson stat	1.9451		Sum u ²	0.00227

Source: own calculations.

Table A5: MIDAS regression parameters, Lithuania

Dependent Variable: Output_gap
Sample (adjusted): 1998Q3 2019Q4
Included observations: 86 after adjustments
Quarterly parameters β :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0121	0.0061	1.984	0.0507
Output_gap(-1)	0.8015	0.0597	13.421	0.000
Monthly parameters θ				
Unemp_rate(+1)/100				
$\theta(1,1)$	3.25174	1.54813	2.10043	0.03890
$\theta(1,2)$	-4.21808	1.55797	-2.70742	0.00830
$\theta(1,3)$	0.96753	0.31531	3.06850	0.00300
Indust_prod(-1)				
$\theta(2,1)$	0.00393	0.05433	0.07233	0.94250
$\theta(2,2)$	0.07166	0.04242	1.68926	0.09520
$\theta(2,3)$	-0.01130	0.00581	-1.94378	0.05550
Monthly Pamameters $\varphi(k)$				
Unemp_rate(+1)/100		Lag	Coefficient	
Lags: 4		0	0.0012	
		1	-1.3143	
		2	-0.6948	
		3	1.8598	
Indust_prod(-1)		Lag	Coefficient	
Lags: 6		0	0.0643	
		1	0.1021	
		2	0.1173	
		3	0.1099	
		4	0.0799	
		5	0.0273	
R-squared	0.8723	S. E. of R.		0.01353
Durbin-Watson stat	2.0584	Sum u ²		0.01427

Source: own calculations.

Table A6: MIDAS regression parameters, Latvia

Dependent Variable: Output_gap				
Sample (adjusted): 2000Q3 2019Q4				
Included observations: 78 after adjustments				
Quarterly parameters β :				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0186	0.0085	2.178	0.0328
Output_gap(-1)	0.8952	0.0661	13.550	0.000
Monthly parameters θ				
Unemp_rate(+1)/100				
$\theta(1,1)$	-11.14430	7.58471	-1.46931	0.14620
$\theta(1,2)$	12.60128	9.31960	1.35213	0.18070
$\theta(1,3)$	-3.02532	2.37123	-1.27585	0.20620
Indust_prod(-1)				
$\theta(2,1)$	0.19155	0.16479	1.16240	0.24900
$\theta(2,2)$	-0.07949	0.11087	-0.71695	0.47580
$\theta(2,3)$	0.01295	0.01541	0.84040	0.40350
Monthly Pamameters $\varphi(k)$				
Unemp_rate(+1)/100				
Lags: 3	Lag	Coefficient		
	0	-1.5684		
	1	1.9570		
	2	-0.5684		
Indust_prod(-1)				
Lags: 6	Lag	Coefficient		
	0	0.1250		
	1	0.0844		
	2	0.0697		
	3	0.0808		
	4	0.1179		
	5	0.1809		
R-squared	0.9095		S. E. of R.	0.01458
Durbin-Watson stat	2.1474		Sum u ²	0.01487

Source: own calculations.

Table A7.: MIDAS regression parameters, Poland

Dependent Variable: Output_gap
Sample (adjusted): 2001Q1 2019Q4
Included observations: 76 after adjustments
Quarterly parameters β :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0060	0.0030	1.992	0.0504
Output_gap(-1)	0.7324	0.0712	10.285	0.000
Monthly parameters θ				
Unemp_rate(+1)/100				
$\theta(1,1)$	0.89982	0.75093	1.19828	0.23500
$\theta(1,2)$	-0.74072	0.55794	-1.32761	0.18870
$\theta(1,3)$	0.11086	0.07956	1.39343	0.16800
Indust_prod(-1)				
$\theta(2,1)$	0.05589	0.13812	0.40467	0.68700
$\theta(2,2)$	0.13243	0.14082	0.94043	0.35030
$\theta(2,3)$	-0.03193	0.02767	-1.15405	0.25250
Monthly Pamameters $\varphi(k)$				
Unemp_rate(+1)/100		Lag	Coefficient	
Lags: 6		0	0.2700	
		1	-0.1382	
		2	-0.3246	
		3	-0.2893	
		4	-0.0323	
		5	0.4464	
Indust_prod(-1)		Lag	Coefficient	
Lags: 4		0	0.1564	
		1	0.1931	
		2	0.1658	
		3	0.0748	
R-squared	0.7558		S. E. of R.	0.00696
Durbin-Watson stat	2.3608		Sum u ²	0.00329

Source: own calculations.

Table A8.: MIDAS regression parameters, Slovenia

Dependent Variable: Output_gap				
Sample (adjusted): 2001Q2 2019Q4				
Included observations: 75 after adjustments				
Quarterly parameters β :				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0086	0.0068	1.263	0.2109
Output_gap(-1)	0.8808	0.0564	15.616	0.000
Monthly parameters θ				
Unemp_rate(+1)/100				
$\theta(1,1)$	-6.21562	3.52633	-1.76263	0.08250
$\theta(1,2)$	5.66441	3.52546	1.60672	0.11280
$\theta(1,3)$	-1.06260	0.70771	-1.50146	0.13790
Indust_prod(-1)				
$\theta(2,1)$	0.05341	0.04401	1.21359	0.22920
$\theta(2,2)$	0.08409	0.02970	2.83156	0.00610
$\theta(2,3)$	-0.01681	0.00433	-3.88126	0.00020
Monthly Pamameters $\varphi(k)$				
Unemp_rate(+1)/100		Lag	Coefficient	
Lags: 4		0	-1.6138	
		1	0.8628	
		2	1.2143	
		3	-0.5595	
Indust_prod(-1)		Lag	Coefficient	
Lags: 6		0	0.1207	
		1	0.1544	
		2	0.1544	
		3	0.1209	
		4	0.0537	
		5	-0.0471	
R-squared	0.9156		S. E. of R.	0.00704
Durbin-Watson stat	1.9398		Sum u ²	0.00332

Source: own calculations.

Table A9.: MIDAS regression parameters, Slovakia

Dependent Variable: Output_gap
Sample (adjusted): 2000Q3 2019Q4
Included observations: 78 after adjustments
Quarterly parameters β :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0084	0.0058	1.445	0.1528
Output_gap(-1)	0.7890	0.0868	9.085	0.000
Monthly parameters θ				
Unemp_rate(+1)/100				
$\theta(1,1)$	-0.49135	1.12646	-0.43619	0.66400
$\theta(1,2)$	0.07212	0.95576	0.07546	0.94010
$\theta(1,3)$	0.02350	0.15980	0.14703	0.88350
Indust_prod(-1)				
$\theta(2,1)$	-0.22163	0.08561	-2.58877	0.01170
$\theta(2,2)$	0.25480	0.07245	3.51695	0.00080
$\theta(2,3)$	-0.03947	0.01205	-3.27536	0.00160
Monthly Pamameters $\varphi(k)$				
Unemp_rate(+1)/100				
Lags: 5	Lag	Coefficient		
	0	-0.3957		
	1	-0.2531		
	2	-0.0635		
	3	0.1731		
	4	0.4566		
Indust_prod(-1)				
Lags: 5	Lag	Coefficient		
	0	-0.0063		
	1	0.1301		
	2	0.1876		
	3	0.1661		
	4	0.0657		
R-squared	0.7368	S. E. of R.	0.01222	
Durbin-Watson stat	2.3833	Sum u ²	0.01046	

Source: own calculations.

Table A10: MIDAS regression parameters, United States

Dependent Variable: Output_gap
Sample (adjusted): 1948Q3 2019Q4
Included observations: 286 after adjustments
Quarterly parameters β :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.0002	0.0015	-0.145	0.8851
Output_gap(-1)	0.8841	0.0256	34.497	0.000
Monthly parameters θ				
Unemp_rate(+1)/100				
$\theta(1,1)$	0.97216	0.42976	2.26211	0.02450
$\theta(1,2)$	-1.14789	0.36819	-3.11768	0.00200
$\theta(1,3)$	0.22435	0.06214	3.61044	0.00040
Indust_prod(-1)				
$\theta(2,1)$	-0.18443	0.08310	-2.21945	0.02730
$\theta(2,2)$	0.30325	0.05898	5.14144	0.00000
$\theta(2,3)$	-0.05591	0.00957	-5.84414	0.00000
Monthly Pamameters $\varphi(k)$				
Unemp_rate(+1)/100		Lag	Coefficient	
Lags: 5		0	0.0486	
		1	-0.4262	
		2	-0.4524	
		3	-0.0298	
		4	0.8414	
Indust_prod(-1)		Lag	Coefficient	
Lags: 5		0	0.0629	
		1	0.1984	
		2	0.2221	
		3	0.1340	
		4	-0.0660	
R-squared	0.8944		S. E. of R.	0.00528
Durbin-Watson stat	2.1319		Sum u ²	0.00776

Source: own calculations.

Appendix 3 Illustration of similarity of sign patterns of estimated parameters to expressions with second difference

Assume for simplicity that output gap g_t is a linear function of unemployment rate UR_t only. The theory predicts that the corresponding parameter should be negative because output gap shall drop in response to rising unemployment rate. Assume for example, that a rise of unemployment rate by 1% causes the output gap drop by 0.6 % (output gap is computed from log output so that this corresponds to the linear relationship). The equation for output gap has then the form

$$g_t = -0.6UR_t$$

Now, assume that the output gap depends on unemployment rate difference $UR_t - UR_{t-1}$. The equation for output gap has then the form

$$g_t = -0.6(UR_t - UR_{t-1}) = -0.6UR_t + 0.6UR_{t-1}$$

Note that by inserting the first difference into the equation, we made the output gap to rise with rising lagged value of unemployment rate. In empirical work, we can encounter a situation, when the impact of unemployment rate on the output gap is strongest during periods, when the business cycle turns and the unemployment rate reaches either a bottom or peak. In this situation, it makes sense for output gap to be function of the second difference of unemployment rate, since the curvature (measured by the second difference) of the path of the variable is greatest during bottoms and peaks. The second difference for the unemployment rate has the form $(UR_t - UR_{t-1}) - (UR_{t-1} - UR_{t-2})$ or after removing the brackets $UR_t - 2UR_{t-1} + UR_{t-2}$.

$$g_t = -0.6(UR_t - 2UR_{t-1} + UR_{t-2}) = -0.6UR_t + 1.2UR_{t-1} - 0.6UR_{t-2}$$

If we generalize the second difference and use $UR_t - UR_{t-2}$ instead of $UR_t - UR_{t-1}$ in the first differencing, we get $(UR_t - UR_{t-2}) - (UR_{t-1} - UR_{t-3})$ or $UR_t - UR_{t-1} - UR_{t-2} + UR_{t-3}$. After plugging this into our equation for output gap we get:

$$g_t = -0.6(UR_t - UR_{t-1} - UR_{t-2} + UR_{t-3}) = -0.6UR_t + 0.6UR_{t-1} + 0.6UR_{t-2} - 0.6UR_{t-3}$$

The pattern of signs of parameters in this equation (negative for lag 0, positive for lags 1 and 2 and negative for last lag 3) is important, because it matches the pattern implied by the function $\phi(\theta, k)$, shown in Figure 5c and shown in estimation results for some countries in Appendix 2. We can thus interpret the estimated parameters as describing the dependence of the output gap generalized second difference of unemployment rate.

Appendix 4 Series of mean absolute revisions across pseudo vintages

Figure A2: Series of mean absolute revisions, Czech Republic

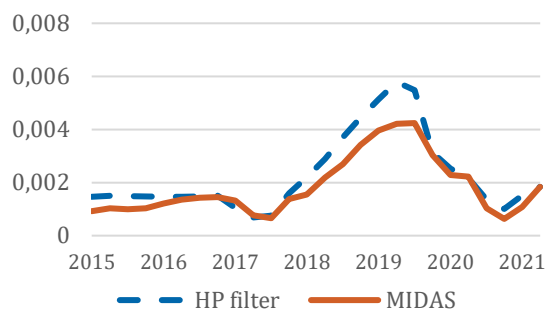
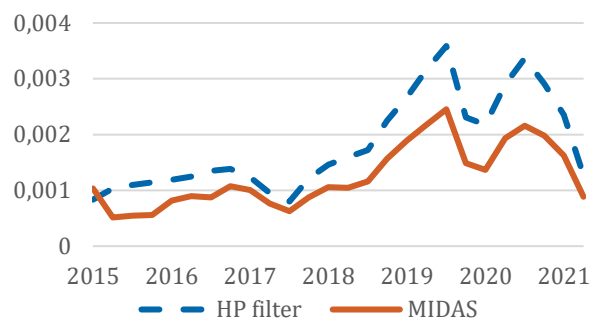


Figure A3: Series of mean absolute revisions, Estonia



Source: own computations.

Figure A4: Series of mean absolute revisions, Hungary

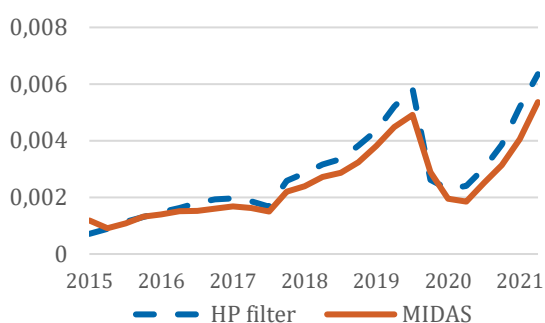
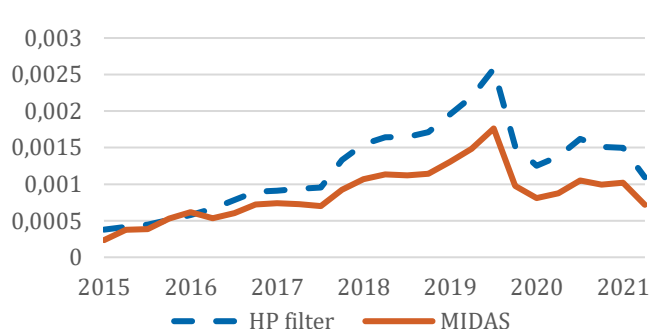


Figure A5: Series of mean absolute revisions, Lithuania



Source: own computations.

Figure A6: Series of mean absolute revisions, Latvia

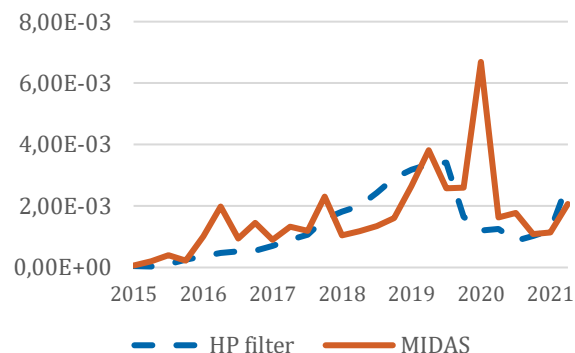
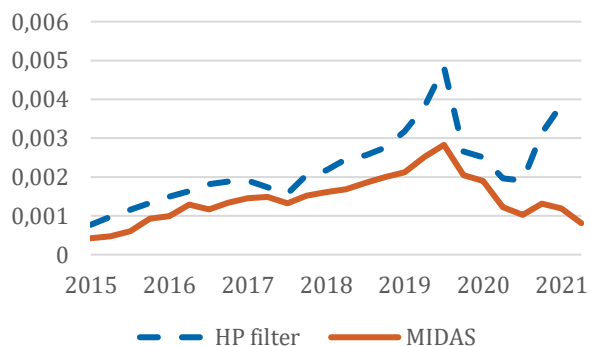
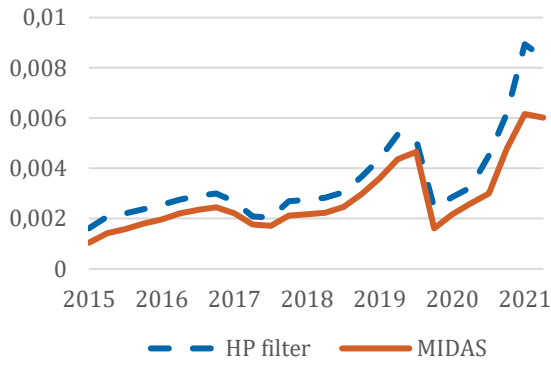


Figure A7: Series of mean absolute revisions, Poland



Source: own computations.

Figure A8: Series of mean absolute revisions, Slovenia



Source: own computations.

Figure A9: Series of mean absolute revisions, Slovakia

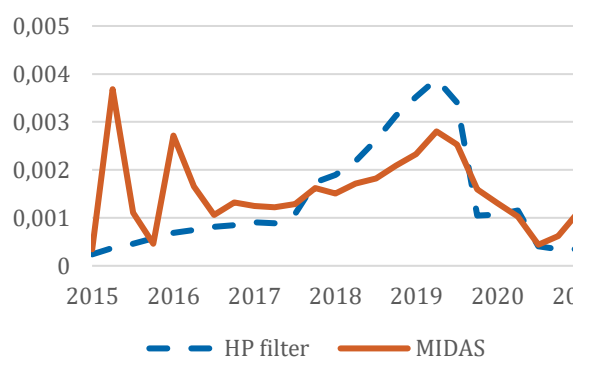
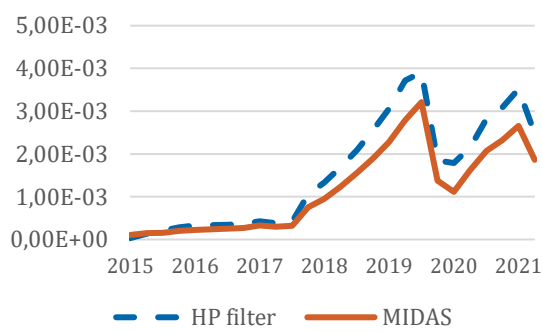


Figure A10: Series of mean absolute revisions, United States



Source: own computations.

Appendix 5 Minimum, maximum and final coefficient of determination of MIDAS regression

Table A11.: Coefficients of determination in MIDAS regressions

Source: own calculations

Country	Czech Rep.	Estonia	Hungary	Lithuania	Latvia	Poland	Slovenia	Slovakia	US
RSQmin	0.942	0.921	0.886	0.872	0.909	0.648	0.916	0.709	0.894
RSQmax	0.955	0.931	0.903	0.878	0.922	0.756	0.929	0.737	0.895
RSQfin	0.954	0.921	0.903	0.872	0.91	0.756	0.916	0.737	0.894

Note: RSQ min-minimal value, RSQ max-maximal value, RSQ_{fin} - final value