Péter Bauer–József Kelemen

Estimation of uncertainty stemming from data revision of Hungarian GDP data

MNB Occasional Papers 129

2017

Péter Bauer–József Kelemen

Estimation of uncertainty stemming from data revision of Hungarian GDP data

MNB Occasional Papers 129

2017

The views expressed are those of the authors and do not necessarily reflect the official view of the central bank of Hungary (Magyar Nemzeti Bank).

MNB Occasional Papers 129

Estimation of uncertainty stemming from data revision OF Hungarian GDP data*

(Adatrevíziókból eredő bizonytalanság becslése a magyar GDP adatokon)

Written by Péter Bauer, József Kelemen

Budapest, July 2017

Published by the Magyar Nemzeti Bank

Publisher in charge: Eszter Hergár

H-1054 Budapest, Szabadság tér 9.

www.mnb.hu

ISSN 1585-5678 (online)

* We would like to express our gratitude for the valuable comments of Balázs Vonnák and the experts of the National Accounts Department of the Hungarian Central Statistical Office (HCSO). We also received valuable assistance from the HCSO for compiling data and useful information on the statistical office's revision practices.

Tartalom

Abstract

Due to the phenomenon of revision, published macroeconomic data can never be regarded as final, because they are subject to continuous change, although the size of the change decreases over time. Our methodology is able to give an estimate of expected future routine revisions, based on the time series properties of the revisions and the data. We intend to filter out the revisions stemming from methodological changes. In addition – in line with the practice used in the literature – we highlight the effect of turning points. Taking these into account, we can observe a more significant systematic bias. We applied this methodology to the time series of Hungarian GDP. Based on our results, GDP growth in 2016 is expected to be subsequently revised upwards in future publications by approximately 0.2 percentage points. Considerable uncertainty surrounds our estimate: the 90 per cent confidence interval surrounding the expected revision is approximately ±0.5 percentage points. Our results are also confirmed by two robustness analyses. In the first analysis, instead of using indicators in the model to capture the effect of the turning points, we define turning points based on the output gap. For the second estimate, we filtered out the effect of the latest major methodology change, i.e., the introduction of ESA 2010 from the revisions. In both cases we obtained results that barely differ from the baseline estimate for the expected revision.

JEL codes: C22, C53
Keywords: revision, GDP, state-space model, Kalman filter

Executive Summary

Due to the phenomenon of revision, published macroeconomic data can never be regarded as final as they are subject to continuous change, although the size of the change decreases over time. But decision-makers need to accurately know the macroeconomic variables, in order to gain a realistic overview of the economy for making decisions. For this reason, it may be useful to regard the published data as an estimate due to the fact that subsequent publications are very likely to be revised over time as updated information becomes available.

Building on the earlier findings of the literature, this study may serve as the basis for the forecasts of the MNB for quantifying the uncertainty of published macroeconomic data. Our methodology is able to give an estimate of expected future routine revisions, based on the time series properties of the revisions and the data. We intend to filter out the revisions stemming from methodological changes. In addition – in line with the practice used in the literature – we highlight the effect of turning points. Taking these into account, we can observe a more significant systematic bias. One of the possible ways of doing this is to include indicators in the model with the help of which we are able to capture the revision effect of turning points.

We applied this methodology to the time series of Hungarian GDP. We estimate the parameters of the model based on past revisions and the latest available data, and we use as an indicator the time series of the ESI backward-looking confidence index for services.

Based on our results, GDP growth in 2016 is expected to be subsequently revised upwards in future publications, by approximately 0.2 percentage points. Based on our model, this upward revision is mainly justified by two factors: the average upward bias of the GDP revision and the growth of the applied indicator in the second half of 2016. Our estimate is surrounded by considerable uncertainty: the 90 per cent confidence interval around the expected revision is approximately ±0.5 percentage points. Let us note here that based on our results, the average revision bias of the Hungarian GDP does not exceed the values obtained in the international literature.

The robustness of our estimates is supported by two other investigations. In the first one, instead of using indicators in the model to capture the effect of turning points, we define turning points based on the output gap. For the second estimate, we filtered out from the revisions the effect of the latest major methodology change, the introduction of ESA 2010. In both cases we obtained results barely different from the baseline estimate for the expected revision.

1 Introduction

By virtue of its nature, every measurement contains measurement error, and therefore so do macroeconomic data. Data providers face contradicting expectations regarding the compilation and publication of data series. On the one hand, the data requirement of decision-makers and analysts is gradually increasing as they want more data. On the other hand, users and even the general public expect published statistics to be sufficiently accurate, that is, the measurement error to be sufficiently small. Thirdly, the data need to be published as soon as possible to provide the most up-to-date input for decisions, analyses and forecasts. But as time goes by, the amount and precision of information grows and it becomes available from multiple sources. An additional key aspect is that the basic relationships among macroeconomic data must also be fulfilled, and this is particularly true for the system of national accounts.

As a result, data providers are unable to incorporate all information – provided that they are available at all – when releasing publications. So additional publications must follow the first publication of macroeconomic data. In the new data publications, more recent information that becomes available in the meantime is also used, any newfound errors are corrected and consistency with new data is achieved. As a result, statisticians revise the data. Accordingly, revision naturally accompanies the measurement of data and at the same time the revised data provide an increasingly reliable picture of the situation and functioning of the economy.

In addition to the above, the data may also be modified subsequently for methodological reasons. In this case, the definition of the given indicator or data series often changes, or the process of generating the data is modified. This type of revision is different in nature from the previously introduced routine revisions.

Due to revision, published data practically can never be considered final. If we disregard methodological revisions, it remains generally true that macroeconomic data can be revised after its first publication over a few additional quarters. The uncertainty of this can be captured by examining the properties of the revisions, and frequently an estimate can be given not only for the uncertainty, but also for the expected revision itself. The question of how revision can be predicted if statisticians use all available information in an optimal way may arise. But in practice, statisticians are bound by rules, and for good reason, as to what data they can work with and how they can apply them when compiling macro statistics. It may be the case that some relevant information exists that statisticians do not or cannot take into consideration because of these rules. For example, confidence indicators are not used in official statistics despite the fact that they are quickly available and often correlate closely with economic activity. Or the bias[1] of revisions also indicates that the use of information is not perfect. But statisticians cannot automatically adjust the statistics generated according to the defined rules. It is unlikely to accurately identify how the bias results from the methodology used for the statistics, but despite this, the revisions can be predicted by knowing the extent of the bias.

Building on the earlier findings of the literature and further fine-tuning the methodology and supplementing the database used, this study may serve as the basis for the MNB's forecasts for quantifying the uncertainty of published macroeconomic data.

The MNB has made efforts in the past to investigate the revision properties of GDP. Bauer et al. (2008) presented the average and the average absolute revision of the preliminary and detailed GDP data publications for the period between 2002 and 2007, but did not deal with the effect of methodological changes. Moreover, the

[1] On average for longer periods, data are revised either rather downward or upward.

December 2016 Inflation Report (MNB, 2016) presents an estimate of the expected revisions concerning GDP for 2016.

The international empirical literature also addresses this topic. Dynan and Elmendorf (2001) states that the preliminary GNP estimates do not capture the acceleration and deceleration of output, and hence neglect the economy's turning points.

Öller and Hansson (2004) use Swedish data to observe the quarterly revisions of GDP items on the expenditure side, based on the annual indices of the first detailed national accounts. Based on this, they found a 0.2 percentage point annual bias for GDP growth in the sample period. The authors also found that these data provide a good basis for economic cycles since the sign of the first data publication is rarely wrong. Finally, they identified an upward bias of GDP in international data.

Bishop et. al (2013) examined the change in revisions on Australian data. For one, they found that the average revision is done upwards, by around 0.1 percentage point for the last quarter in the case of quarterly growth and by 0.2-0.3 percentage point in the case of annual growth. The authors also found that when analysing the turning points in the case of peaks, upward revision was applied for the preceding period and downward revision was generally applied for the following period. An opposing revision trend can be observed in the troughs of the cycles.

Using preliminary French GDP data, Mogliani and Ferrière (2016) found that the revisions are unconditionally unbiased and that over time the volatility of the revisions decreases as the number of publications increases. Moreover, the sign and the volatility of the revisions may depend on the stage of the economic cycle and they correlate with confidence indicators, interest rate spreads and equity market yields.

In the theoretical literature, Mankiw and Shapiro (1986) attempt to answer the question whether revisions can be predicted. Analysing American GNP revisions in their study, they introduced the issue of news and noise. According to the latter interpretation, there is measurement error or noise in the revisions and consequently the more revisions are done the more the measurement error is reduced or eliminated. However, according to the former, preliminary data releases can be regarded as an efficient forecast that contain all information, and thus new publications reduce or eliminate prediction error by incorporating the new information. But this also means that in this case, revisions cannot be predicted.

Kapetanios and Yates (2004) point out that when examining revisions, the majority of researchers assume that the true value of a given variable can be approximated using its final or most recently published observation so as they can estimate the variance of the measurement error. The authors propose an alternative method where they model the activity of an imaginary statistical office where the published estimates of the past period are more uncertain. For this reason those models which underweight more recent information relative to earlier information can perform better.

In this study we present a model building on the methodology of Cunningham et al. (2007), which is capable of giving an estimate based on the time series properties of the revisions and of the data, also using indicators, for the expected revisions while also presenting the uncertainty of the macro data. Thereafter we apply the model to GDP and estimate the parameters of the model based on historic revisions of GDP and the latest available GDP data. Finally, we present the expected revision of the latest available GDP data, and the uncertainty of the published GDP data.

2 Methodology

Our methodology is partly based on our own considerations and partly on the paper of Cunningham et al. (2007). In line with the literature, we also treat the description of the revision model in a general way, enabling the method to be reproduced for any time series. For the sake of clarity, however, we focus on the one-variable case in addition to the lack of autocorrelation of the error term of the measurement equation, rendering practical implementation easier.

In the following we define the revision as the difference between the value published by the statistical office and the actual, unobserved variable to which the publications converge over time. On the one hand we have a notion about the latter based on economic considerations as to what process it may follow, and on the other hand we know that its value influences the estimate of the statistical office. These together define a state-space model which can be used to estimate the actual process with the help of observed data.

Let us assume in the state equation that the time series of the unobserved variable follows a stationary autoregressive process of order q with an expected value μ and an error term η_t. In this case, the value of the unobserved variable at time t, y_t can be expressed as a function of its own lagged variables with coefficients α_i:

$$y_t = \mu + \sum_{i=1}^{q} \alpha_i y_{t-1} + \eta_t \tag{1}$$

The value of the variable in question measured by the statistical office at time T $(T \geq t)$, y_t^T depends on the actual value of the variable, on the revision bias c^{T-t} and on the ε_t^T measurement error. In this case the measurement equation can be expressed as follows:

$$y_t^T = y_t + c^{T-t} + \varepsilon_t^T \tag{2}$$

The state-space model is defined by (1) and (2), the state equation and the measurement equation. But additional considerations are called for based on the literature. First of all we assume about the revision bias that as the statistical office revises an observation of a given date t more and more over time, that is, the observation becomes more mature and the bias of the revision is likely to be smaller thanks to the expansion of the information set.

$$c^{T-t} = c^0 \left(1+\lambda\right)^{T-t} \tag{3}$$

Upon the first publication of a given observation $(T = t)$ the bias is c^0. However, if the data for the observation are published several times $(T > t)$, the bias will be smaller and will exponentially tend to zero over time, where the convergence speed is determined by the parameter λ $(-1 \leq \lambda < 0)$.

Secondly we assume heteroscedasticity for the error term of the measurement equation. As revisions are received, the standard deviation of the published data ($\sigma^2_{\varepsilon^{T-t}}$) decreases, that is, as a data becomes more mature, the more precisely the statistical office can measure it. The error has an initial standard deviation ($\sigma^2_{\varepsilon^0}$) upon the first release, which then decays on the occasion of additional publications based on the following formula at a speed of δ $(-1 \leq \delta < 0)$

$$\sigma^2_{\varepsilon^{T-t}} = \sigma^2_{\varepsilon^0} \left(1+\delta\right)^{T-t} \tag{4}$$

2.1 TURNING POINTS

Turning points and the cyclical position of the economy play a key role in explaining the revisions. In expansions upward revisions, while in contractions downward revisions can be observed. One of the reasons may be if economic agents erroneously report their current position at turning points based on the previous periods. For example, during the crisis period, it was observable for industrial data that initially more optimistic numbers were reported.[2] But on the long-term average, these revisions stemming from expansions and contractions may neutralise one another. A prerequisite for this is that the length of the positive and negative cycles coincide, otherwise this may be shown as bias in the data. For example, if the negative cycles are shorter, then this may lead to the average upward revision of the macro variables. For the above reasons it is worthwhile to consider turning points when analysing the properties of the revision.

We try to capture the effect of turning points with two different methods to ensure the robustness of our results. According to one of the methods, we use explanatory variables in the state-space model that move closely together with turning points. In the other method, we apply the turning point effect in the bias of the measurement equation.

Based on the first approach we use non-revised data series (z_t) to improve the estimation of our unobserved variable. We suppose that the value of z_t depends on the unobserved variable with coefficient γ, on a constant c_z and a measurement error v_t.

$$z_t = c_z + \gamma y_t + v_t \tag{5}$$

In the other case, we identify the turning points in the bias of the measurement equation with the help of expansion and contraction periods so that

$$c^{T-t} = \begin{cases} c_f^0\left(1+\lambda_f\right)^{T-t}, & \text{in expansion} \\ c_l^0\left(1+\lambda_l\right)^{T-t}, & \text{in contraction} \end{cases} \tag{6}$$

But for this, we need to set up a rule to categorise the state of the economy to an expansion state or, on the contrary, to a contraction state. We can do this based on objective methods, but also based on subjective reasons. Naturally the definition selected will impact the results.

2.2 ESTIMATION OF THE MODEL

The estimation of the parameters of the state-space model is performed in two steps. First, we determine one part of the parameters directly, with the use of a vintage database, and then the rest is estimated with the help of the maximum likelihood function of the state-space. The biggest advantage of this approach is that the entire vintage database is not needed to model the state-space. Thereafter, we can estimate the values of the unobserved time series with the help of the Kalman filter.

As the first step we introduce the notion of revision matrix. For this, we assume that after n $(n \geq 1)$ quarters have elapsed, the statistical office receives all the information and thereafter only methodology revisions affect the time series. Therefore we look at time series y_t^{t+n} as the actual value of the variable in question during the estimation of the parameters. At that point we can define a revision matrix, **W**, any element of which is as follows

$$w_{i,t} = y_t^{t+i} - y_t^{t+n} \tag{7}$$

[2] Based on the verbal information supplied by the experts of the HCSO.

Column *t* means the first *n* revisions of the observation pertaining to a given date and row *i* $(0 \le i \le n)$ of this column shows how much we deviate from the value of the *n*th publication considered to be the actual value at the *i*th publication. In case of an upward revision we obtain a negative number while in the case of a downward revision we obtain a positive number. This number times minus one means the extent of the revision remaining.[3]

With the help of revision matrix **W** we give an estimate for the extent of the bias and for the standard deviation of the error term of equation (2). To calculate the bias, we use the average of the matrix rows, that is, the revisions belonging to a given maturity. These indicate how much a given observation is revised on average at a given maturity level. We fit the following formula to it based on (3) so we can estimate the speed of decay and the initial error (t_0 is the starting date of the vintage time series).

$$\overline{w_{i,\cdot}} = \sum_{t=t_0}^{T-n} \frac{w_{i,t}}{T-n} = c^0 \left(1+\lambda\right)^i + \omega_i \tag{8}$$

We determine the uncertainty in the error term of equation (2) as in the previous case. When calculating this, we take the sample variance of the rows of matrix **W**, and then we do a regression based on equation (4).

$$var\left(w_{i,\cdot}\right) = \sigma^2_{\varepsilon^0} \left(1+\delta\right)^{T-t} + \theta_i \tag{9}$$

[3] We define the components of matrix **W** in this form to be consistent with the state-space model. Because based on (2) it is $y_t^T - c^{T-t} = y_t$ if there is no shock. This means that we add the revision to the observed variable which will, in this way, coincide with the unobserved variable.

3 Empirical results

In this section, we apply the methodology presented in the previous section for the Hungarian GDP time series. First, we present the data used, then we address the baseline model and the two modified versions, and finally we perform a robustness analysis.

3.1 DATA

From among the quarterly Hungarian GDP data, we use the year-on-year volume indices (the same period of the previous year = 100), adjusted for calendar effects. It needs some explanation as to why we examine the revision of indices adjusted for calendar effects instead of the seasonally adjusted data. In its forecasting practice, the MNB starts out from either the raw or the adjusted for calendar effects year-on-year indices and hence its GDP forecast is based on that. The analysts of the MNB redo the seasonal adjustment on the GDP data series supplemented with the forecast and reconcile the data obtained. The purpose of this practice is to reduce the endpoint problem of seasonal adjustment. However, in the revision of the seasonally adjusted and seasonally adjusted and reconciled GDP data, this endpoint problem also plays a role while it has no significance in the MNB forecast. We could also say that the MNB is in reality anticipating the revision of the seasonally adjusted data with the adjustment done in parallel with the forecast. This is why we decided to disregard the revision resulting from the seasonal adjustment when examining the revision.[4] Moreover, the use of the calendar effect-adjusted data instead of raw data is justified by their more favourable time series properties. But in this way we can only handle the seasonality found in the data by using the year-on-year volume indices.

From among the GDP data publications, we only deal with the detailed data releases; therefore, the revision of preliminary data is not part of our analysis. This is also justified by the forecasting practice of the MNB, as the forecast of the Inflation Report is based on detailed GDP data.

We use a so-called vintage database for processing the data which shows the GDP data pertaining to specific dates at various publication dates. This allows us to follow how the GDP data of a given quarter change over time in the new publications. The structure of the database is shown schematically in Table 1.

Table 1
Structure of the vintage database

Vintage/period	Q1 2001	Q2 2001	Q3 2001	…	Q3 2016	Q4 2016
Q1 2001	x					
Q2 2001	x	x				
Q3 2001	x	x	x			
…	…	…	…	…		
Q3 2016	x	x	x	…	x	
Q4 2016	x	x	x	…	x	x

Note: the "x"-s denote that there is data in the given cell. The latest publication is included in the last row of the table. The date of the vintage does not indicate the date of the publication but the quarter for which data was first released in the given publication.

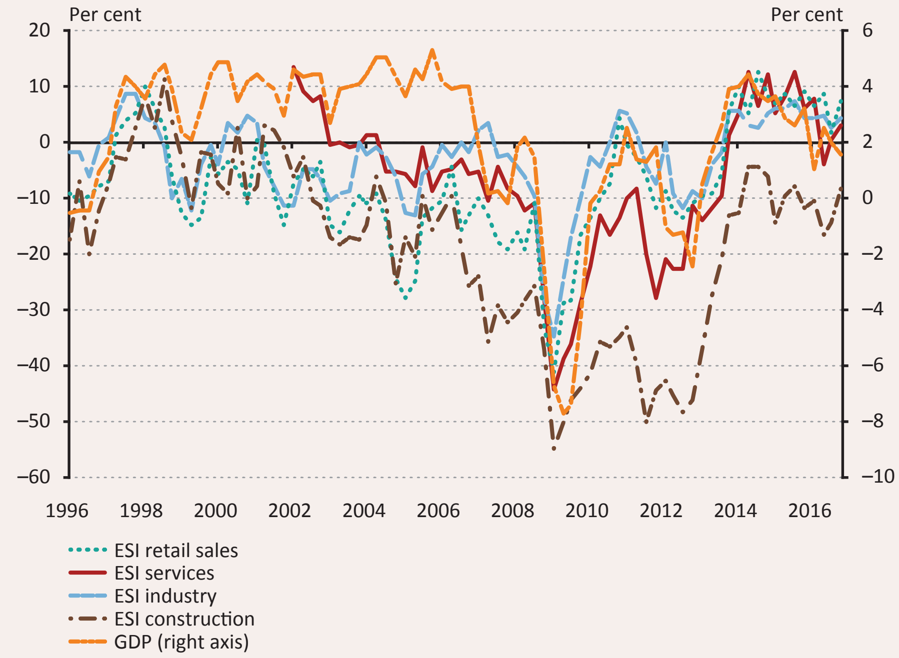
[4] We also performed the calculations for the seasonally adjusted year-on-year indices. Both the bias and the standard deviation of the revision proved to be larger than those of the calendar effect adjusted indices. The change in the standard deviation is as expected, but the causes of the larger bias are not clear.

In the vintage database, the first published time series taken into account contain data only for Q1 2001 while last published time series contain data only for the period between Q1 2001 and Q4 2016. The length of the vintage time series increases by one additional quarter after each quarter.[5] Consequently, we examine the properties of the revision starting from Q1 2001. This may be favourable because the previous period still included the effects of an economic transition, and therefore the properties of the data revisions applied in that period may have been different from the revisions characterising the later stages of the economy.

We perform the estimation of certain parameters of the model presented in the methodology section based on the vintage database and the details thereof are discussed in that section. We estimate the other parameters of the model and the expected revision as well as the uncertainty of the GDP data based on the latest available time series released, which lasts from Q1 1996 until Q4 2016.

In addition to GDP, we also used some indicators for the estimation for which we expect that they move closely together with the GDP, that they are quickly available and that they should not be revised, or if revised, only to a negligible extent. We selected four confidence indices that meet the above conditions: the ESI backward-looking confidence indices for 3 months for retail trade, for industry, for construction and for services. For every quarter we took the last month's value of the given quarter for every one of these sectors. Moreover we used their level because based on experiences their level moves closely together with the annual GDP index and not their year-on-year index (Chart 1). The indicators are available from Q1 1996 until Q4 2016, except for the indicator regarding services which only starts from Q1 2002. It is not practical to use all 4 indices at the same time because they strongly correlate with one another. Therefore we performed principal component analysis and found that the ESI index of services has the strongest correlation with the first principal component, and

Chart 1
GDP and ESI confidence indices
(Year-on-year change in per cent, adjusted for calendar effects and the value of the last month of the quarter, respectively)



Source: HCSO, European Commission.

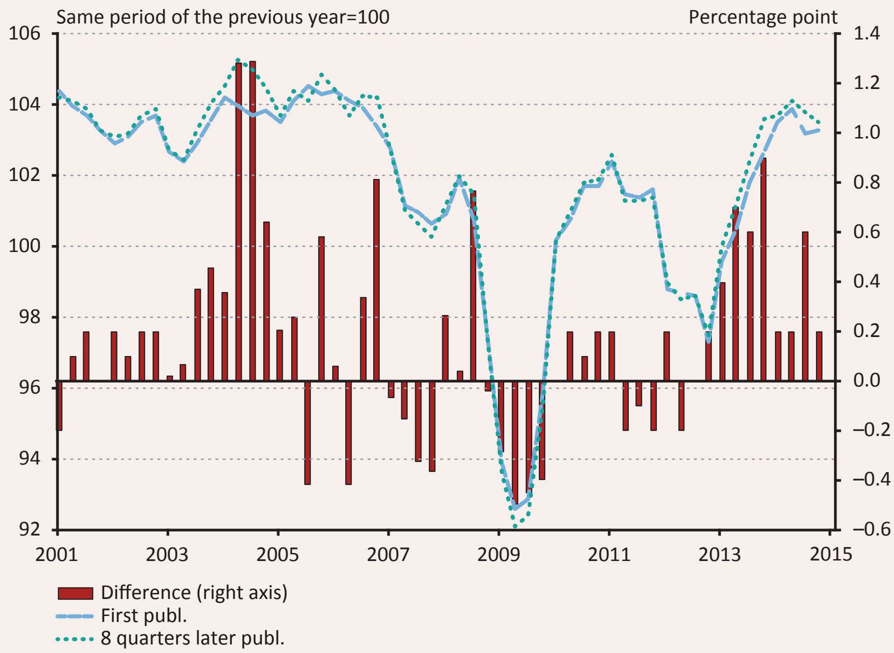
[5] It should be noted that the HCSO did not publish any data adjusted for calendar effects in its publications prior to 2005, therefore we substituted them with raw data. To our knowledge, the reason for this is that during that period, in the course of the seasonal adjustment of GDP, the HCSO did not include any calendar effects. As a result of this solution, for Q1 2004, the impact of the leap day causes a more significant, approximately 0.5 percentage point revision, but this only distorts the average revision upwards to a negligible extent.

moreover, this correlation is very high (greater than 0.9). So as baseline, we selected the ESI index for services as indicator.[6]

In addition to the above time series we also use the time series of the output gap estimated by the MNB for one of the robustness analysis, which is available from Q1 2002 until Q3 2016.

When calculating the revisions we rely on the assumption that the eighth quarter following the first publication (that is, the ninth publication) closely approximates the final "true" GDP data. This is because the routine revision of the GDP affects the two years preceding the reference period based on the HCSO's practice. Revisions beyond two years may take place due to methodological changes and therefore revisions beyond two years should not be taken into account. Thanks to this solution, we can filter out most of the impact caused by changes in methodology on revisions. But some of their effects may still persist, since the HCSO may also revise some data of less than two years when applying methodological changes. The difference between the first detailed GDP publication and the one published 8 quarters later is shown in Chart 2.

Chart 2
First detailed GDP publication and the one published 8 quarters later, and the difference between them
(data adjusted for calendar effects, same period of the previous year = 100, in percentage points, respectively)



Source: HCSO, own calculation.

We calculate the revision in an unusual way, not relative to the first publication, but as the difference between the given publication and the publication considered to be final. Accordingly, the definition of the revision matrix already introduced under the model estimation is the following in the case of GDP data. We generate the various revision time series belonging to the various maturity levels (which data publication is concerned) from the GDP data that we place in the rows of matrix **W**. The revision time series belonging to maturity level *i*: $w_{it} = y_t^{t+i} - y_t^{t+8}$ where the superscripts show when the data pertaining to date *t* was published. In the formula, *i = 0* denotes the first publication but we do not indicate the publication time lag. So according to the definition, if w_{it} is positive, it means that the data considered to be final is lower than the published data, therefore the HCSO will revise the data downward. If w_{it} is negative, the opposite is true: the data considered to be final is higher than the published data, therefore the HCSO will revise the data upward.

[6] In principle we are not allowed to use the principal component itself as an indicator as it is revised upon the receipt of new data. Despite this, as a robustness analysis, we also estimated the model with the first principal component used as an indicator. The results were basically no different from the baseline.

In this way matrix **W** starts from Q1 2001 and lasts until Q4 2014 based on the vintage database, because the ninth publication is last available for this data. Accordingly, in our case **W** is an 8 by 56 matrix the structure of which can be seen in Table 2.

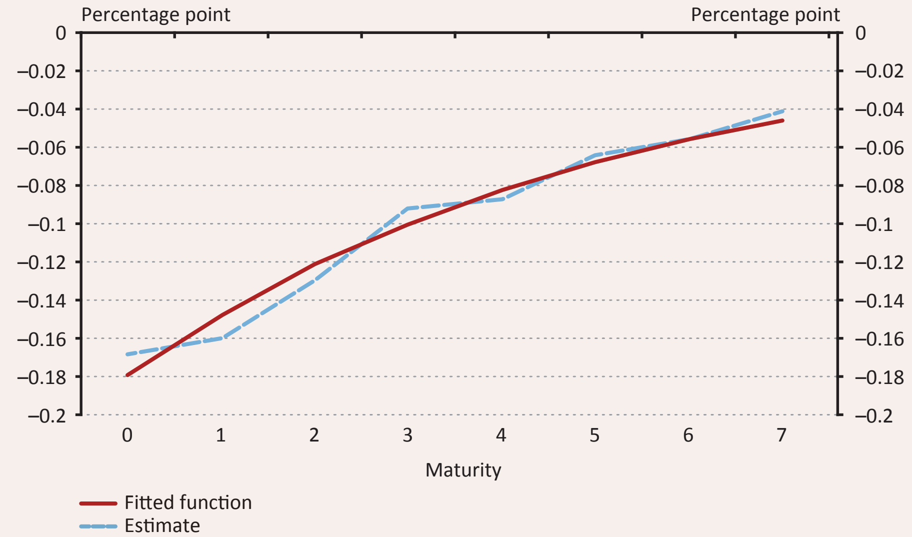
Table 2
Structure of the revision time series matrix (*W*)

Maturity/date	Q1 2001	Q2 2001		Q3 2014	Q4 2014
0	$y_{01Q1}^{01Q1}-y_{01Q1}^{03Q1}$	$y_{01Q2}^{01Q2}-y_{01Q2}^{03Q2}$	$\cdots$	$y_{14Q3}^{14Q3}-y_{14Q3}^{16Q3}$	$y_{14Q4}^{14Q4}-y_{14Q4}^{16Q4}$
1	$y_{01Q1}^{01Q2}-y_{01Q1}^{03Q1}$	$y_{01Q2}^{01Q3}-y_{01Q2}^{03Q2}$	$\cdots$	$y_{14Q3}^{14Q4}-y_{14Q3}^{16Q3}$	$y_{14Q4}^{15Q1}-y_{14Q4}^{16Q4}$
2	$y_{01Q1}^{01Q3}-y_{01Q1}^{03Q1}$	$y_{01Q2}^{01Q4}-y_{01Q2}^{03Q2}$	$\cdots$	$y_{14Q3}^{15Q1}-y_{14Q3}^{16Q3}$	$y_{14Q4}^{15Q2}-y_{14Q4}^{16Q4}$
3	$y_{01Q1}^{01Q4}-y_{01Q1}^{03Q1}$	$y_{01Q2}^{02Q1}-y_{01Q2}^{03Q2}$	$\cdots$	$y_{14Q3}^{15Q2}-y_{14Q3}^{16Q3}$	$y_{14Q4}^{15Q3}-y_{14Q4}^{16Q4}$
4	$y_{01Q1}^{02Q1}-y_{01Q1}^{03Q1}$	$y_{01Q2}^{02Q2}-y_{01Q2}^{03Q2}$	$\cdots$	$y_{14Q3}^{15Q3}-y_{14Q3}^{16Q3}$	$y_{14Q4}^{15Q4}-y_{14Q4}^{16Q4}$
5	$y_{01Q1}^{02Q2}-y_{01Q1}^{03Q1}$	$y_{01Q2}^{02Q3}-y_{01Q2}^{03Q2}$	$\cdots$	$y_{14Q3}^{15Q4}-y_{14Q3}^{16Q3}$	$y_{14Q4}^{16Q1}-y_{14Q4}^{16Q4}$
6	$y_{01Q1}^{02Q3}-y_{01Q1}^{03Q1}$	$y_{01Q2}^{02Q4}-y_{01Q2}^{03Q2}$	$\cdots$	$y_{14Q3}^{16Q1}-y_{14Q3}^{16Q3}$	$y_{14Q4}^{16Q2}-y_{14Q4}^{16Q4}$
7	$y_{01Q1}^{02Q4}-y_{01Q1}^{03Q1}$	$y_{01Q2}^{03Q1}-y_{01Q2}^{03Q2}$	$\cdots$	$y_{14Q3}^{16Q2}-y_{14Q3}^{16Q3}$	$y_{14Q4}^{16Q3}-y_{14Q4}^{16Q4}$

Note: y_t^s denotes the GDP data published in s+1 quarter pertaining to date t

Based on equation (8) and (9) we can calculate the bias and the variance of the error of the measurement equation. The exponential function fits well on the average revisions (Chart 3).

Chart 3
Estimates of average revision belonging to the given maturity level and the fitted exponential function
(Compared to maturity level 8, percentage point)



Note: Negative values mean upward revisions
Source: own calculation.

3.2 RESULTS OF THE BASELINE MODEL

In the baseline model we use as an indicator the ESI backward-looking confidence index for services, which helps in estimating the actual, unobserved GDP data. In order to motivate the use of indicators, we examined whether the utilised confidence index correlates with the revisions, that is, whether the revisions can be predicted when we know the indicator. According to our results (for more details, see Appendix, Table 7) the confidence index has a significant correlation with the revisions: the higher the index value the more upward the GDP data will be revised by the HCSO in the future, at least in case of the early publications.

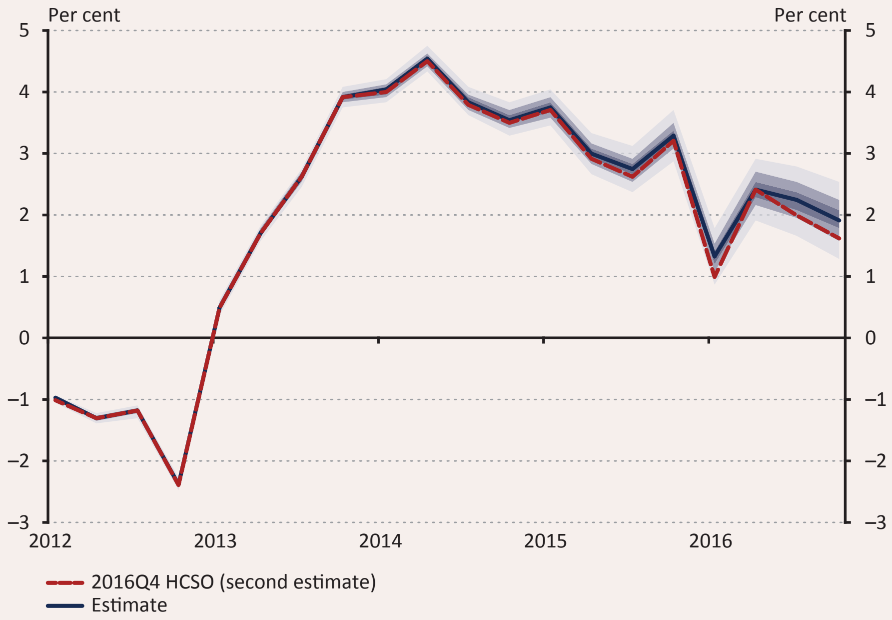
We present the estimated values of the parameters in the Appendix (Table 8 and Table 9); the main text includes only the estimates of the "actual" unobservable value of the GDP. We compare the model estimates for the quarters of 2016 with the latest data published by the HCSO on 7 March 2017 in a table format and we present the fan chart showing the uncertainty of the estimation for the time series (Table 3 and Chart 4).

Table 3
Results of the baseline model for 2016
(Adjusted for calendar effects, year-on-year, %)

			Confidence interval (90%)	
	HCSO	Estimate	Low value	High value
Q1 2016	1.0	1.3	0.9	1.8
Q2 2016	2.4	2.4	1.9	2.9
Q3 2016	2.0	2.2	1.7	2.8
Q4 2016	1.6	1.9	1.3	2.5
Average	**1.75**	**1.97**	**1.43**	**2.51**

Source: own calculation, HCSO.

Chart 4
Fan chart of baseline model results
(Adjusted for calendar effects, year-on-year, %)



Note: The bands show the confidence intervals of the estimation (30%, 60% and 90%).
Source: own calculation, HCSO.

Our results show that the expected revision of the 2016 GDP data is approximately 0.2 percentage points. The expected revision of 2015 is significantly lower, and practically zero during the preceding years. In parallel with the extent of the expected revision, the uncertainty of the GDP data release also decreases. The upward revision in 2016 is basically justified by the historic upward bias of the revisions and the growth of the indicator in the second half of 2016.[7]

3.3 ROBUSTNESS ANALYSIS

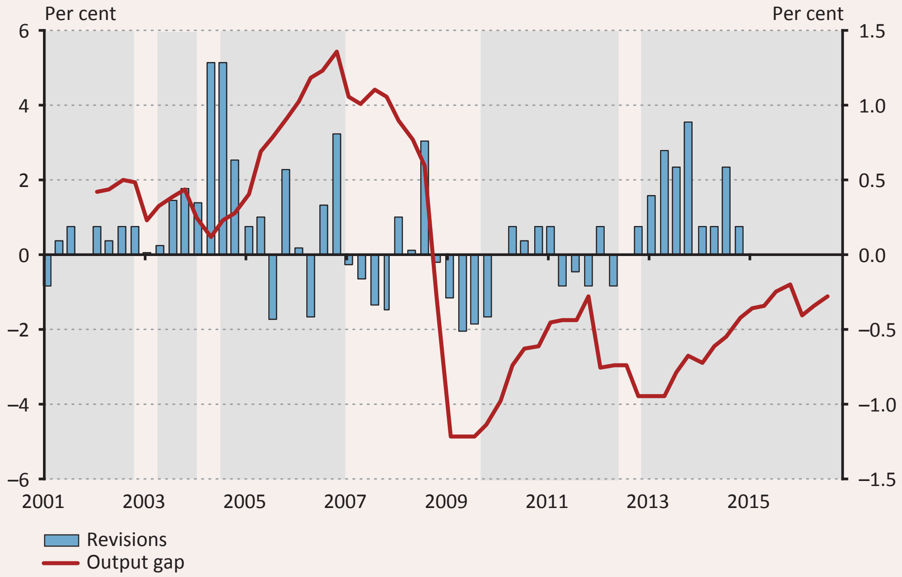
Below, we also present two alternative estimations in addition to the baseline model.

3.3.1 Identification of turning points based on output gap

We follow the method already outlined, where we define the bias separately for the expansion and contraction periods. We identify the various periods based on the output gap estimated by the MNB. According to our definition, the economy is in an expansion if the value of the output gap increases in at least two consecutive quarters. The economy is in a contraction if the value of the output gap decreases in at least two consecutive quarters. According to our definition, if the output gap moves in a direction contrary to the previous one in only one quarter, the state of the economy does not change. Let us note here that the above definition may be considered somewhat ad-hoc, and one that does not necessarily capture the turning points in an accurate way. Our objective was to support the robustness of our baseline model with another type of model.

The last public output gap data is available only until Q3 2016 in the Inflation Report of December 2016 when the economy was in an expansion. Moreover, we assume based on the detailed macroeconomic forecast of the Inflation Report that in Q1 2017 the value of the output gap will increase (decrease in absolute value). It follows that the economy remained in expansion during Q4 2016. Chart 5 shows the output gap and the two states of the economy and how the revision of the GDP evolved during these two states. The figure confirms that GDP is revised upward much more strongly during expansions than during contractions.

Chart 5
Relationship between GDP revisions and the output gap and periods of expansion and contraction
(Difference between the first detailed GDP publication and the one published 8 quarters later and the deviation from the potential GDP in per cent)



Note: the expansions are indicated by the grey areas
Source: MNB, own calculation.

[7] The indicator represents substantial additional information for the model. Without using the indicator, the revision expected for 2016 would be lower, approximately 0.15 percentage points, and its uncertainty would be slightly higher.

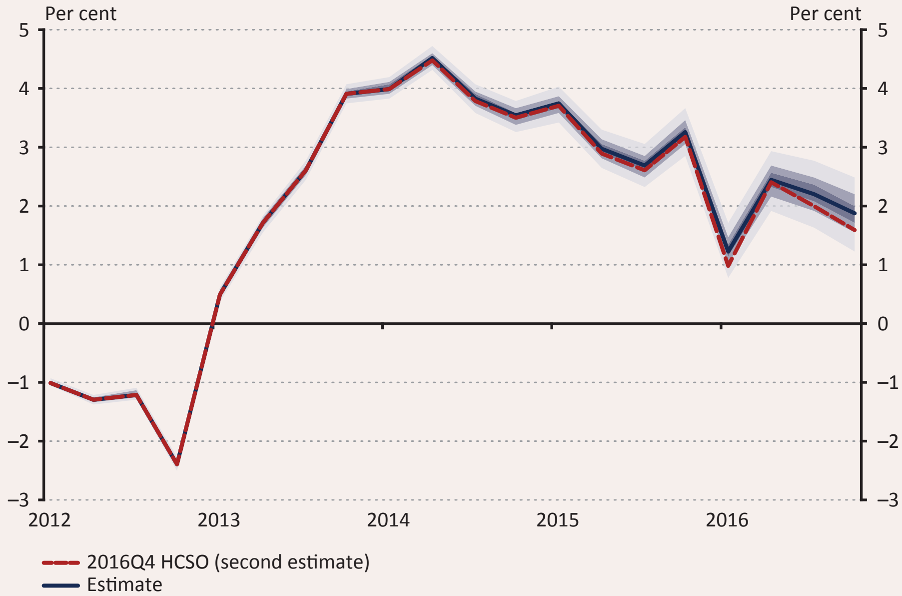
Thereafter we apply our model for the "actual", unobservable GDP estimation and as previously we compare the results with the last publication, also presenting the uncertainty of the estimates (Table 4 and Chart 6).

Table 4
Results of the turning point model for 2016
(Adjusted for calendar effects, year-on-year, %)

			Confidence interval (90%)	
	HCSO	**Estimate**	**Low value**	**High value**
Q1 2016	1.0	1.2	0.8	1.7
Q2 2016	2.4	2.4	1.9	3.0
Q3 2016	2.0	2.2	1.6	2.8
Q4 2016	1.6	1.9	1.2	2.5
Average	**1.75**	**1.94**	**1.40**	**2.48**

Source: own calculation, HCSO

Chart 6
Fan chart of turning point model results for 2016
(Adjusted for calendar effects, year-on-year, %)



Note: The bands show the confidence intervals of the estimation (30%, 60% and 90%).
Source: own calculation, HCSO.

Our results show that the expected revision of the 2016 GDP data evolves around 0.2 percentage points according to this alternative approach as well. This confirms the results of the baseline model. In this case, we estimated an expected revision of a similar extent than in the baseline model despite the omission of the indicator. It comes from the fact that we observed higher bias in the past during expansions of the economy.

3.3.2 Filtering out the ESA 2010 methodology change

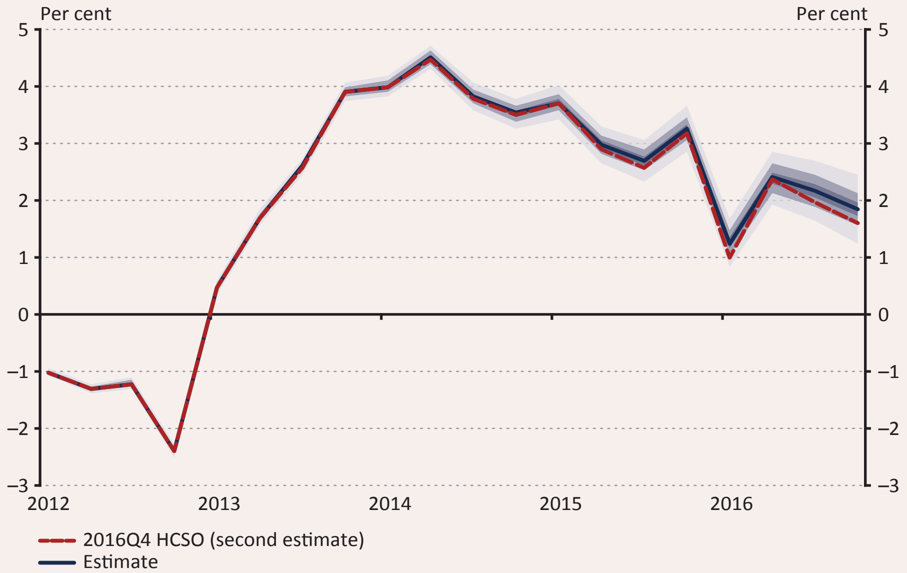
As already mentioned on several occasions, we consider the ninth publication as final data, thereby reducing the effect of methodology revisions on our estimate. Nevertheless, methodology revisions may still have an impact on the results despite this as we do not filter out the effects of the revisions within the 8 quarters. The latest methodology change with a substantial impact in the national accounts was the introduction of the ESA 2010 methodology in September 2014.[8] As a robustness analysis, we tried to filter out its effect from the revisions and we re-estimate our model in the obtained database. For this, we consider all of the revisions performed between the publication in September 2014 and the previous detailed data publication of June 2014 as the effect of methodology change and we apply this change retrospectively on all the previously published GDP data. But this method too can only be regarded as an approximation since it is not able to take into account the routine revisions performed in September 2014. This is the reason for not being able to filter out the effect of all methodology changes from the GDP revision because by doing so we would also eliminate a significant part of the routine revisions.[9] The estimates for 2016 and the related fan chart are shown in Table 5 and Chart 7.

Table 5
Results of the model filtering out ESA 2010 for 2016
(Adjusted for calendar effects, year-on-year, %)

			Confidence interval (90%)	
	HCSO	**Estimate**	**Low value**	**High value**
Q1 2016	1.0	1.3	0.8	1.7
Q2 2016	2.4	2.4	1.9	2.9
Q3 2016	2.0	2.2	1.7	2.7
Q4 2016	1.6	1.9	1.3	2.4
Average	**1.75**	**1.93**	**1.42**	**2.44**

Source: own calculation, HCSO.

Chart 7
Fan chart of the results of the model filtering out ESA 2010
(Adjusted for calendar effects, year-on-year, %)



Note: The bands show the confidence intervals of the estimation (30%, 60% and 90%).
Source: own calculation, HCSO.

Our results show in this case that the extent of the revision expected for 2016 is less, but continues to evolve close to 0.2 per cent. This means that although the effect of methodology revisions may influence the results of the baseline model, but filtering out the latest major methodology changes barely modifies our conclusions.

[8] In terms of quarterly data, it was performed with the data release of December 2014.

[9] The HCSO gives a detailed account of the methodology changes simultaneously with the annual GDP data publication, but it only shows its effects for the annual current price data. This information is insufficient for filtering out the methodology change from the quarterly constant price GDP revision.

4 Conclusion

In the first part of the study we presented a methodological framework allowing the estimation of the expected revision of macroeconomic variables and the uncertainty of the variable stemming from the revisions with the help of a state-space model. In the methodology used, we attempted to quantify the phenomenon also observed in the literature, the revision effect of turning points. For this we proposed two different methods to ensure the robustness of the results. One was the introduction of macroeconomic indicators, and the other one was distinguishing expansion and contraction periods based on the output gap.

In the second part of the study we gave examples of the practical use of the methodology using the most recent Hungarian GDP time series. The models estimated a revision of around 0.2 percentage points for the average of 2016. Our estimate is surrounded by considerable uncertainty: the 90 per cent confidence interval around the expected revision is approximately ±0.5 percentage points. It should be noted that based on our results, the average revision bias of Hungarian GDP does not exceed the values obtained in the international literature. Our results can be attributed to several factors: the historic bias of the revisions and the effect related to the behaviour of the economic cycle. In the baseline model we used the ESI confidence index for services. In addition, we also gave two other estimates that confirmed the robustness of the results. In the first one, instead of the indicators, we identified the turning points of economic activity based on the output gap. The results essentially coincided with the ones obtained in the baseline case. In the second case, we filtered out the effect of the introduction of the ESA 2010 methodology from the revisions. In this case, the estimation of the revision bias is somewhat lower than in the baseline case, thus the revision expected for 2016 is slightly smaller than the one calculated in the baseline case. The summary of the estimates of the baseline and the two robustness analyses are shown in Table 6.

Table 6
The most recent publication of the HCSO, the baseline estimates and the two robustness analysis estimates
(Adjusted for calendar effects, year-on-year, %)

	KSH	Baseline	Turning point-based	ESA 2010 filtered out
Q1 2016	1.0	1.3	1.2	1.3
Q2 2016	2.4	2.4	2.4	2.4
Q3 2016	2.0	2.2	2.2	2.2
Q4 2016	1.6	1.9	1.9	1.9
Average	1.75	1.97	1.94	1.93
Deviation from the HCSO		0.22	0.19	0.18

Source: own calculation, HCSO.

A future area of research may be the analysis of methodological revisions. The basic objective of the modeller when predicting revisions is to determine routine revisions by filtering out all technical effects. If this technical or methodological effect is unbiased, then the estimation of routine revisions also remains unbiased, but the contrary may also be the case. If, for example, statisticians make mostly changes on the occasion of methodological changes that increase GDP growth, then the effect of methodological revisions will be positive.[10] To determine this, we need empirical methods for which the current price revisions may serve as a basis, because we have data only for this in case of methodological changes. By quantifying these effects the results can be made more precise.

[10] For example, during methodological changes concerning GDP, the scope of activities considered often expands, but this only entails the increase of the GDP's level, but not necessarily the increase of the growth rate.

5 References

Bauer, P., Eppich, Gy., Nobilis, B. and Virág, B. (2008): *Macroeconomic statistics with the eyes of an analyst*. (only in Hungarian: *Makrogazdasági statisztikák elemzői szemmel)*. Statisztikai Szemle, Volume 86, issue 7–8.

Bishop, J., Gill, T. and Lancaster, D. (2013): *GDP Revisions: Measurement and Implications*. Reserve Bank of Australia Bulletin.

Cunningham, A., Eklund, J., Jeffery, C., Kapetanios, G. and Labhard, V. (2007): *A state-space approach to extracting the signal from uncertain data*. Bank of England Working Paper, No. 336.

Cunningham, A. and Jeffery, C. (2007): *Extracting a better signal from uncertain data*. Bank of England Quarterly Bulletin.

Dynan, K.E. and Elmendorf, D. W. (2001): *Do provisional estimates of output miss economic turning points?* FRB FEDS Discussion Paper, No. 2001-52.

Garrat, A., Lee, K., Mise, E. and Shields, K. (2005): *Real time representations of the output gap*. University of Leicester Discussion Paper, No. 130.

Kapetanios, G. and Yates, T. (2004): *Estimating time-variation in measurement error from data revisions; an application to forecasting in dynamic models*. Bank of England Working Paper, No. 238.

Mankiw, N. G. and Shapiro, M. (1986): *News or noise: An analysis of GNP revisions*. Survey of Current Business, 66 (5).

Kapetanios, G. and Yates, T. (2005): *Monetary policy and data uncertainty*. Bank of England Working Paper, No. 281.

Mogliani, M. and Ferrière, T. (2016): *Rationality of announcements, business cycle asymmetry, and predictability of revisions. The case of French GDP*. Banque de France Working Paper Series, No. 600.

MNB (2016): *Inflation Report.* Magyar Nemzeti Bank, December 2016.

Öller, L.-E. and Hansson, K.-G. (2004): *Revision of National Accounts: Swedish Expenditure Accounts and GDP*. Journal of Business Cycle Measurement and Analysis, 2004 (3).

6 Appendix

6.1 CORRELATION OF ESI SERVICE INDICATORS WITH REVISIONS

We analysed the correlation between the confidence indicator and the revisions with the help of the following 8 regressions, where *i = 0,…,7* shows which level of maturity (that is, publication *i+1*) is concerned.

$$w_{it} = \alpha + \beta z_t + \varepsilon_t$$

That is, we explained the future revision of the value published for time *t* with the help of the indicator. We obtained the following results which support that the indicator is able to predict the revisions, and a strongly significant correlation can be observed until the sixth publication.

Table 7
Correlation between the confidence indicator and the revisions
(regression coefficients)

Publication	constant	ESI serv.
1.	–0.295***	–0.013***
2.	–0.27***	–0.012***
3.	–0.229***	–0.011***
4.	–0.168***	–0.008***
5.	–0.158**	–0.008***
6.	–0.122**	–0.006***
7.	–0.094**	–0.004*
8.	–0.058**	–0.002

*Note: *,**,*** denotes the coefficients significant at 10, 5 and 1 per cent. We used Newey-West standard errors.*
Source: own calculation.

6.2 MODEL

6.2.1 Estimated parameters

Parameters estimated directly from the vintage database:

Table 8
Parameters estimated from the vintage database

	Baseline	Turning point	ESA 2010
c^0	–0.179	–	–0.14
λ	–0.177	–	–0.162
$\sigma^2_{\varepsilon^0}$	0.167	0.163	0.146
δ	–0.205	–0.203	–0.198
c^0_f	–	–0.236	–
λ_f	–	–0.196	–
c^0_l	–	–0.057	–
λ_l	–	–0.3	–

Source: own calculation.

6.2.2 State-space model estimated in Eviews

State equation:

$$y_t = c(1) + c(2) y_{t-1} + c(3) y_{t-2} + c(4) y_{t-3} + c(5) y_{t-4} + c(6) y_{t-5} + \eta_t$$

$$Var(\eta_t) = \exp(c(10))$$

Measurement equation:

$$y_t^T = y_t + c_t^{T-t} + \varepsilon_t^T$$

$$Var(\varepsilon_t) = \sigma^2_{\varepsilon_t}$$

A measurement equation also applies to the indicator in the baseline model and in the method filtering out the ESA 2010 effects:

$$z_t = c(21) + c(22) y_t + \upsilon_t$$

$$Var(\upsilon_t) = \exp(c(23))$$

Parameters estimated with maximum likelihood:

Table 9
Parameters estimated from the state-space model using the Maximum likelihood estimation

	Baseline	Turning point	ESA 2010
C(1)	0.062	0.03	0.062
C(2)	1.188***	1.192***	1.186***
C(3)	–0.181	–0.179	–0.179
C(4)	–0.136	–0.134	–0.136
C(5)	–0.236	–0.242	–0.237
C(6)	0.288***	0.296***	0.289***
C(10)	0.212	0.223	0.213
C(21)	–13.613***	–	–13.596***
C(22)	3.557***	–	3.554***
C(23)	4.013***	–	4.017***

*Note: *,**,*** denote the coefficients significant at 10, 5 and 1 per cent.*
Source: own calculation.

MNB OCCASIONAL PAPERS 129

ESTIMATION OF UNCERTAINTY STEMMING FROM DATA REVISION OF HUNGARIAN GDP DATA

July 2017

Print: Prospektus–SPL consortium

6 Tartu u., Veszprém H-8200