Rationality, asymmetry and predictability of GDP revisions. The case of France

Matteo Mogliani[1,*]

Banque de France - Business Conditions and Macroeconomic Forecasting Directorate

Thomas Ferrière

Banque de France - Business Conditions and Macroeconomic Forecasting Directorate

Abstract

We analyze French GDP revisions and we investigate the rationality of preliminary announcements of GDP. We consider nonlinearities, taking the form of business cycle asymmetry and time changes, and their effect on both revisions and the rationality of releases. We find that asymmetry represents a prevalent feature in French GDP releases. Our findings suggest that releases are overall unbiased but inefficient. This means that GDP estimates are not affected by a systematic prediction error, but do not seem to use efficiently information publicly available at the time of the release. Finally, we investigate the issue of forecastability of GDP revisions in real-time and we find out that revisions are overall predictable when asymmetry is taken into account.

Keywords: GDP Revisions, Real-time dataset, Efficiency, Unbiasedness, Forecasting
JEL: C22, C52, E37

[*]Corresponding author; Banque de France, 46-1376 DGEI-DCPM-SEMAP, 31 Rue Croix des Petits Champs, 75049 Paris CEDEX 01 (France). Phone: +33142929756. Fax: +33142924469.

Email addresses: `matteo.mogliani@banque-france.fr` (Matteo Mogliani), `thomas.ferriere@banque-france.fr` (Thomas Ferrière)

[1]We wish to thank participants in seminars at the Banque de France for helpful comments and discussions. The usual disclaimer applies. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Banque de France.

February 3, 2016

1. Introduction

The economic and econometric literature has been paying increasing attention to the statistical properties of preliminary announcements of key macroeconomic variables, such as GDP, and their subsequent revisions. For an exhaustive review of this literature, we refer to Mankiw et al. (1984), Mankiw and Shapiro (1986), Mork (1987), Faust et al. (2005), Swanson and Van Dijk (2006), Aruoba (2008), and references contained therein. The rationale of this research activity is substantially twofold. First, monetary and fiscal policy decisions usually depend on preliminary estimates of the state of the economy, represented by standard business cycle indicators. This implies that the more accurate the initial announcement for these indicators, the more effective the impact of policies. For this purpose, initial announcements must be "rational", *i.e.* unbiased and efficient predictors of final (or pseudo-final) data releases, with revisions behaving as the errors from rational predictions. The unbiasedness condition states that first estimates should not be affected by a systematic prediction error, while the efficiency condition implies that all the available information is efficiently employed in the estimation of preliminary values of farther released data, as suggested by the rational expectations theory. Second, understanding the statistical features of initial announcements is a central issue for empirical economists and forecasters. Indeed, when researchers investigate the effect of policy decisions through econometric models, the data chosen as input should be as close as possible to the data actually observed by policymakers at the time the policy decision was undertaken (Orphanides, 2001). In addition, the construction and evaluation of forecasting models should also account for the presence of data revisions, consistently with the target of the forecaster (Croushore, 2011). It follows that the analysis of macroeconomic data revisions and the implication for macro-econometric modeling and forecasting are the main issues this literature has recently dealt with.

With respect to GDP releases and revisions, recent empirical results have pointed to an interesting outcome: the unbiasedness and efficiency conditions do not hold overall on preliminary announcements for a large group of industrialized countries (see for instance Aruoba, 2008 and Faust et al., 2005). This means that early announcements of GDP cannot be considered as rational estimates of revised data. It follows that GDP revisions are not consistent with the definition of noise from rational forecasts of final values, but rather with a measurement error uncorrelated with final values and correlated with data available at the time of the announcement. In other words, revisions do not contain news, but reduce statistical noise (Croushore, 2011). However, the inefficiency of early estimates would also mean that revisions might be predictable, using for instance information available at the time of the initial announcement as predictor. This is a crucial point for policymakers, because even if preliminary GDP announcements are irrational, in practice subsequent revisions may be predicted by conditioning on some (neglected) information.

In this paper, we add to the previous literature on these points, by assessing the rationality of preliminary estimates of French GDP and by performing a forecasting analysis of revisions. For this purpose, we use a real-time database including complete vintage estimates of GDP spanning from Q1 1991 (released in Q2 1991) to Q4 2014 (released in Q1 2015), *i.e.* 96 preliminary estimates of quarterly GDP growth rates. Our analysis is performed on these data after controlling for the presence of annual and benchmark GDP revisions. The former arise from the release of national annual accounts in May of year T for year $T-1$, which may imply a substantial adjustment of quarterly GDP growth rates in order to match the annual rate. The latter is instead related to changes in the base-year and in the accounting methodology. As these revisions are unrelated with the issue that we are addressing here (unbiasedness and efficiency of preliminary announcements), we implement a simple approach to account for their effect on the data.

Our first goal is to investigate the main statistical features of French GDP revisions, especially the first two moments of their distribution. We find out that, as expected, revisions are unconditionally unbiased, *i.e* the average revision is not significantly different from zero, and that the volatility of revisions decreases with the order of releases. However, it has been shown in the literature that revision processes may be characterized by nonlinear features which my affect unconditional and conditional moments (Brodsky and Newbold, 1994; Rathjens and Robins, 1995). A reasonable candidate for the underlying nonlinear mechanism is given by the asymmetry of the business cycle (Swanson and Van Dijk, 2006). Indeed, the sign and the volatility of revisions may depend on the phases of the business cycle during which the estimate is released. For instance, the contraction of activity is unusually fast during recessions, such that the magnitude of the drop may not be correctly evaluated by National Statistical Agencies (NSAs hereafter) in their early announcements. In this example, early announcements may overstate the economic activity during recessions, but they may also understate it during expansions. In the present paper, we follow the literature and we perform the analysis under the assumption of both symmetric and asymmetric business cycle. We nevertheless additionally assume that alternative nonlinear mechanisms, based on looser definitions of the standard business cycle, may be at play. For instance, revisions may depend on the sign and/or the size of the growth rate of the released GDP. Further, revisions may be also correlated with the speed of expansion and contraction of the economy, rather than the (cycle or growth) phase of GDP. Finally, we investigate whether revisions of early announcements do evolve over-time, possibly according to changes in the accounting performance of NSAs. Globally, the results point to the presence of nonlinear features in both the mean and the volatility of revisions. We find out that preliminary announcements are affected by an overestimation bias during phases of contraction of GDP, which nevertheless disappears with later releases. We also find a positive correlation between revisions and the acceleration of GDP, suggesting that revisions become more positive when the economy is speeding up and less positive when the economy is slowing down. Further, results point to an increase in the volatility of early GDP revisions during phases of contraction, but also to a decrease in the volatility of revisions over time.

Our second goal is to provide an assessment of the rationality of GDP announcements. We do so by examining the entire revision history for GDP through a sequence of testing procedures based on a standard regression approach. The analysis is performed by including an information set (*e.g.*, past revisions, stock market returns, interest rate spreads, survey data) available at the time of GDP releases into linear, auxiliary, rationality regressions (see Faust et al., 2005, and Swanson and Van Dijk, 2006). We also assume that rationality may be affected by the business cycle asymmetry, by the acceleration of the activity, or by a structural change. For example, during recessions, data frequently used to compute early estimates may convey information on the severity of the contraction only with some lag. Results suggest that the efficiency hypothesis can be strongly rejected as a result of significant conditional bias in the contraction regime. However, it is worth noticing that inefficiency can be also detected, although less frequently, when the test is performed over the expansion regime. All in all, our results suggest that early GDP announcements are conditionally unbiased, but inefficient. The latter can be mainly attributes to the significant correlation between GDP revisions and external economic information.

Finally, our third goal is to study the forecastability of GDP revisions in real-time (Faust et al., 2005; Aruoba, 2008). This is particularly meaningful and appealing in the present context of overall rejection of the efficiency hypothesis. We perform the analysis by modeling GDP revisions as dynamic linear and nonlinear equations and by implementing a GEneral-To-Specific (GETS) approach to model selection, where the general unrestricted model includes the same set of predictors used to assess the rationality of releases. Evaluation of point and density forecasts suggests that while linear models fail to provide reasonable results, allowing for business cycle asymmetry in forecasting equations does improve significantly the overall predictive accuracy.

The paper is organized as follows. Section 2 describes the methodology followed in order to test for rationality of preliminary French GDP estimates, as well as for the presence of business cycle asymmetries and time variation in the revision process. Section 3 presents the real-time dataset used in the present study and the strategy followed to construct GDP revisions. Section 4 reports results on the main statistical features of revisions and the rationality of preliminary GDP estimates. In Section 5, we investigate the predictability of GDP revisions. Finally, Section 6 concludes.

2. The econometric methodology

2.1. Revisions and forecast rationality tests

Let us denote y_t^{t+k} the value of the quarterly growth rate of a variable of interest at quarter t, released at quarter $t+k$ (the vintage estimate). First-release GDP values (the preliminary announcements) are usually released about 30-50 days after the end of the reference quarter, *i.e.* with one quarter lag. Accordingly, the notation used for these values is y_t^{t+1}. In this case, k can be interpreted as the order of releases (first-release for $k = 1$, second-release for $k = 2$, and so on). Final data are denoted $y_t^{t+\ell}$, where ℓ is supposed to be a finite value large enough to exclude further revisions. We can now define a revision as the difference between two announcements released respectively at time $t+h$ and $t+k$:

$$ {}_{h}y_t^k = y_t^{t+h} - y_t^{t+k} \tag{1} $$

where $h \geq k+1$. From the general Equation (1), we can identify three useful expressions covering a wide space of possible revisions (Swanson and Van Dijk, 2006):

$$ {}_{k+1}y_t^k = y_t^{t+k+1} - y_t^{t+k} \tag{2a} $$
$$ {}_{k+1}y_t^1 = y_t^{t+k+1} - y_t^{t+1} \tag{2b} $$
$$ {}_{\ell}y_t^k = y_t^{t+\ell} - y_t^{t+k} \tag{2c} $$

namely, fixed-width revisions (FXW hereafter) in expression (2a), increasing-width revisions (INW) revisions in (2b), and remaining revisions (REM) in (2c). These expressions are used in Section 4.1 to provide a statistical assessment (mean and volatility) of the process governing the revisions of French GDP. It is worth noticing that INW revisions are equivalent to FXW revisions for $k=1$ and to REM revisions for $k=\ell-1$. Further, the expression ${}_{\ell}y_t^1 = y_t^{t+\ell} - y_t^{t+1}$ (*i.e.* INW revisions for $k=\ell-1$ and REM revisions for $k=1$) denotes the final revisions, computed as the difference between final (ℓ-th) releases and preliminary announcements. As noted by Aruoba (2008), final revisions are expected to have small variance compared to final releases, *i.e.* $\sigma_{{}_{\ell}y_t^1} < \sigma_{y_t^{t+\ell}}$.

The literature on testing for rationality of early announcements is broadly based on the framework provided by Mankiw and Shapiro (1986), where first-release data (y_t^{t+1}) are tested for the hypothesis of either rational forecasts or noisy estimate of final data ($y_t^{t+\ell}$). On the one hand, the rationality hypothesis implies that revisions are related to news unknown at the date of preliminary GDP estimates. In that case, revisions are uncorrelated with first-release data and we expect $\sigma_{y_t^{t+1}} < \sigma_{y_t^{t+\ell}}$, because efficient forecasts of later releases are expected to be smoother than exactly those later releases to be forecast. On the other hand, the measurement error hypothesis implies that provisional estimates are indeed an observation of later releases, but measured with an error (or noise) that is progressively reduced across subsequent revisions. Under this assumption, revisions are uncorrelated with final (fully revised) data and we expect $\sigma_{y_t^{t+1}} > \sigma_{y_t^{t+\ell}}$.

In this paper, we favor a testing approach based on the null hypothesis of forecast rationality. In this sense, revisions are "news", and hence expected to display zero mean and to be uncorrelated with the information available at time $t+k$.[1] We follow Faust et al. (2005) and Swanson and Van Dijk (2006), who propose a general testing approach of the forecast efficiency hypothesis based directly on the correlation between the revisions defined in Equation (2c) and the k-th data release.

[1]It is worth noticing that rejecting the hypothesis of "news" revisions in this framework does not necessarily imply accepting the hypothesis of "noise" revisions. This is because the two null hypotheses are mutually exclusive but not collectively exhaustive, meaning that both hypotheses may be simultaneously rejected (see Aruoba, 2008, and Jacobs and van Norden, 2011).

According to the notion of rational expectations, a data release is a rational forecast of the final release if and only if

$$y_t^{t+k} = \mathbb{E}\left[y_t^{t+\ell}|\mathbf{\Omega}_{t+k}\right],$$

where $\mathbf{\Omega}_{t+k}$ is the information available at time $t+k$. Following Swanson and Van Dijk (2006), forecast rationality can be tested by the means of the following regression model:

$$_{\ell}y_t^k = \alpha + y_t^{t+k}\beta + \mathbf{X}'_{t+k}\boldsymbol{\gamma} + \varepsilon_{t+k}, \tag{3}$$

where $\mathbf{X}_{t+k}$ is a vector of indicators representing the information set available at time $t+k$. These conditions can be verified through a Mincer-Zarnowitz-type rationality test (Mincer and Zarnowitz, 1969), by simply setting the joint null hypothesis $\alpha = \beta = 0$ and $\boldsymbol{\gamma} = 0$. Following Keane and Runkle (1990), the test of rationality in the context of Equation (3) can be broken down in an unbiasedness hypothesis ($\alpha = \beta = 0$, with the imposed restriction that $\boldsymbol{\gamma} = 0$) and an efficiency hypothesis ($\alpha = \beta = 0$ and $\boldsymbol{\gamma} = 0$, with no imposed restrictions). By the means of these two sub-hypothesis, we can readily ascertain whether the rejection of the hypothesis of data rationality depends on the presence of either a prediction bias or incomplete information when constructing preliminary announcements, or both.

A crucial issue in the analysis described above is the construction of robust test statistics for the relevant parameters (t-statistic) and the restrictions of the rational forecast hypothesis (F-statistic). Since GDP revisions often display an apparent autocorrelated structure as well as changing volatility (Aruoba, 2008), it seems reasonable to compute heteroskedasticity and autocorrelation consistent (HAC) estimates of the variance-covariance matrix (VCV). However, as pointed out by Andrews (1991), Andrews and Monahan (1992) and Den Haan and Levin (1997), the use of HAC estimators can lead the tests to substantial size distortions in finite samples. To attenuate this issue, we compute HAC estimates of the VCV following the approach suggested by Kiefer et al. (2000) and Kiefer and Vogelsang (2002a,b). This approach consists in computing HAC standard errors using a Bartlett kernel without truncation, with bandwidth equal to the sample size.[2] The resulting test statistics t^* and F^* are asymptotically invariant to nuisance parameters, but have non-standard distributions that depend on the number of restrictions q being tested:

$$t^* \Rightarrow W_1(1)\left[2\int_0^1 B_1(r)^2 \mathrm{d}r\right]^{-\frac{1}{2}} \qquad \text{and} \qquad F^* \Rightarrow W_q(1)'\left[2\int_0^1 B_q(r)B_q(r)'\mathrm{d}r\right]^{-1} W_q(1)/q,$$

[2] For a more general discussion on the underlying asymptotic theory (fixed-b asymptotics), see Kiefer and Vogelsang (2005).

where $B_q(r) = W_q(r) - rW_q(1)$ and $W_q(r)$ is a $q \times 1$ vector of independent Brownian motions. We use standard Monte Carlo techniques to simulate these distributions and compute critical and probability values.

Finally, we point out that, in the present linear regression framework, failures to reject the null hypothesis of rationality only imply the absence of *linear* correlation between revisions and information available at time $t + k$, because the Wald tests described above do not necessarily detect *nonlinear* dependence. Cognizant of this issue, we compare our results with those of alternative rationality tests consistent against generic nonlinearity by implementing the testing procedure described by Corradi et al. (2009). The proposed test statistic takes the form:

$$M_T = \sup_{\boldsymbol{\lambda} \in \Lambda} |m_T(\boldsymbol{\lambda})|,$$

where $m_T(\boldsymbol{\lambda}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} {}_{\ell}y_t^k \times w\left(\sum_{j=0}^{t-1} \boldsymbol{\lambda}_j' \Psi(\mathbf{Z}_{t+k})\right)$ and $\mathbf{Z}_{t+k} = \left(y_t^{t+k}, \mathbf{X}_{t+k}\right)'$. Following Corradi and Swanson (2002) and Corradi et al. (2009), we set $\lambda_{i,j} = \lambda_i(j+1)^{-2}$, $w(\cdot)$ as the exponential function, and $\Psi(\cdot)$ as the inverse tangent function. The limiting distribution of this statistic depends on the nuisance parameter $\boldsymbol{\lambda} \in \Lambda$, and thus standard critical values are not available. Corradi et al. (2009) suggest to implement a bootstrap procedure to obtain critical and probability values, leading to a bootstrap analog of M_T, namely $M_T^* = \sup_{\boldsymbol{\lambda} \in \Lambda} |m_T^*(\boldsymbol{\lambda})|$, where $m_T^*(\boldsymbol{\lambda}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} \left({}_{\ell}y_t^{k*} \times w^*(\boldsymbol{\lambda}) - {}_{\ell}y_t^k \times w(\boldsymbol{\lambda})\right)$, and ${}_{\ell}y_t^{k*}$ and $w^*(\boldsymbol{\lambda})$ are bootstrap resampled series.

2.2. Standard business cycle asymmetry

Business cycle fluctuations may affect the efficiency of NSAs, in the sense that the statistical properties of revisions and preliminary announcements may depend on the phase of the cycle. Indeed, changes in activity during recession may be exceptionally fast and thus hard to be estimated in real time. Furthermore, data related to the manufacturing sector (survey data and industrial production index) are more efficient to forecast first announcements of GDP than data related to other sectors of the economy (Mogliani et al., 2014). However, during recessions, the contraction is often sharper in the former than in the latter. For that reasons, first-release data may overstate the economic activity during contraction periods and understate it during expansion periods. We can test for business cycle asymmetry by the means of the following threshold model:

$${}_{\ell}y_t^k = \left(\alpha_1 + y_t^{t+k}\beta_1 + \mathbf{X}_{t+k}'\boldsymbol{\gamma}_1\right)\mathbb{I}(s_t) + \left(\alpha_2 + y_t^{t+k}\beta_2 + \mathbf{X}_{t+k}'\boldsymbol{\gamma}_2\right)[1 - \mathbb{I}(s_t)] + \varepsilon_{t+k}, \tag{4}$$

where $\mathbb{I}(s_t)$ is an indicator function, taking value of 1 if period t is part of a GDP expansion and 0 otherwise. Phases of expansion and contraction are here estimated through the Bry-Boschan algorithm (Harding and Pagan, 2002) on the latest available vintage (Q2 2015). Estimated turning points are quite stable across different vintages and are broadly equivalent to those provided by

CEPR and other authors (for instance, Cotis and Coppel, 2005).[3] Business cycle asymmetry in revision processes can be easily tested through a Wald test on parameters restrictions by setting the null hypothesis $\alpha_1 = \alpha_2$, with $\beta_1 = \beta_2 = 0$ and $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = 0$ imposed. When testing for data rationality, we proceed sequentially as follows: first, we test for business cycle asymmetry in the unbiasedness regression by setting the joint null hypothesis $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, with $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = 0$ imposed; second, if the asymmetry hypothesis cannot be rejected, we test again for unbiasedness and efficiency while imposing the non-linearity implied by Equation (4), *i.e.* we set respectively the joint null hypotheses $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$, with the imposed restriction that $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = 0$, and $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$ and $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = 0$, with no imposed restrictions. This procedure should allow us to ascertain whether the rejection of the hypothesis of data rationality depends on systematic predictive bias and/or inefficiency during a specific phase of the cycle. Two important remarks are nevertheless worth making. The definition of asymmetry here follows the usual peak and trough characterization of the business cycle (Burns and Mitchell, 1946), while alternative definitions involving a different number of GDP growth regimes and estimation methods (*e.g.* Markov switching models) are not considered for ease of analysis. More importantly, the results provided by the rationality tests performed on Equation (4) should be interpreted with care, because the dates of business cycle turning points (either estimated by an algorithm on a specific vintage or called by a Dating Committee) do not usually belong to the information set of agents until the end of a contraction period. Still, the sign and the magnitude of estimated coefficients should allow us to explore the importance of asymmetry in the data, as well as to assess whether NSAs do overstate or understate the activity in the business cycle.[4]

2.3. Threshold business cycle asymmetry

The main drawback of the business cycle analysis above is that the sample of expansion and contraction periods is usually unbalanced towards the former. Indeed, business cycle tends to move slowly, and so that only a very few number of observations refers to a contraction period. One way to address this issue is to explore the revisions process and test for data rationality conditionally on the speed, rather than the phase, of the business cycle. This analysis should shed light on whether first-release data overstate or understate the activity when GDP growth is positive or negative. More formally, we can check for the effect of the speed of business cycle by the means of the following

[3] Interestingly enough, the algorithm is not able to detect the recession in Q4 2002-Q2 2003, which disappeared from recent vintages of GDP. This is likely due to the combination of different factors, such as base-year changes and the introduction of major accounting innovations in 2014. To be consistent with a pseudo-real time detection of turning points, we set this period as a contraction phase.

[4] Roughly speaking, if β_1 is positive (negative) and β_2 is negative (positive), this means that preliminary announcements understate (overstate) GDP growth during expansion periods and conversely overstate (understate) GDP growth during contraction periods. Of course, preliminary announcements can also systematically understate (overstate) GDP if the sign of both β_1 and β_2 is positive (negative), but the size of these coefficients differs across the business cycle.

second threshold model:

$$
{}_{\ell}y_t^k = \left(\alpha_1 + y_t^{t+k}\beta_1 + \mathbf{X}'_{t+k}\boldsymbol{\gamma}_1\right)\mathbb{I}(y_t^{t+k};c) + \left(\alpha_2 + y_t^{t+k}\beta_2 + \mathbf{X}'_{t+k}\boldsymbol{\gamma}_2\right)\left[1 - \mathbb{I}(y_t^{t+k};c)\right] + \varepsilon_{t+k}, \quad (5)
$$

where $\mathbb{I}(y_t^{t+k};c)$ is an indicator function, taking value of 1 if $y_t^{t+k} > c$ and 0 otherwise, with $c = 0$ imposed.[5] Business cycle asymmetry is then tested as described is Section 2.2. It is worth noticing that the argument of the indicator function in (5) includes the k-th release of data. Indeed, the logic of rationality tests requires that the transition variable belongs to the information set of agents at the time the announcement is made. This restriction is unnecessary when testing for business cycle asymmetry in revision processes, and we set the transition variable as the vintage $y_t^{t+\ell}$.[6]

2.4. Business cycle acceleration

An interesting alternative way to account for business cycle features consists in estimating the correlation between revisions and a transformation of $y_t^{t+\ell}$ (an acceleration factor) supposed to capture the speed at which the economy is expanding or contracting (Dynan and Elmendorf, 2001; Bishop et al., 2013). Similarly to the analysis above, the sign and the magnitude of the estimated correlation coefficients should reveal whether first-release data either overstate or understate the activity, although here this behavior is conditional on phases during which the economy is speeding up or slowing down. More formally, we can check for the effect of the acceleration of business cycle on revision processes by the means of the following model:

$$
{}_{h}y_t^k = \alpha + \Phi(y_t^{t+\ell})\beta + \varepsilon_{t+k}, \quad (6)
$$

where $\Phi(y_t^{t+\ell})$ is a function defining the acceleration factor. We hence test whether β is statistically different from zero and positively (negatively) signed, which would suggest that revisions become more positive (negative) when the economy is speeding up and more negative (positive) when the economy is slowing down. Following Dynan and Elmendorf (2001), we can set this function as follows:

$$
\Phi(y_t^{t+\ell}) = y_t^{t+\ell} - \left(y_{t-1}^{t+\ell}\right)^{1/2}\left(y_{t-2}^{t+\ell}\right)^{1/2},
$$

[5] A generalization of Equation (5), accounting not only for the sign of GDP growth but also for its size, could be obtained by replacing the indicator function $\mathbb{I}(y_t^{t+k};c)$ with the (continuous) logistic function $\mathbb{G}(y_t^{t+k};\theta,c) = [1+\exp(-\theta(y_t^{t+k}-c)/\sigma_{y^{t+k}})]^{-1}$, where $\mathbb{G}(\cdot)$ is bounded between 0 and 1, θ is the smoothing parameter, $\sigma_{y^{t+k}}$ is the standard deviation of y_t^{t+k} (a scaling factor estimated through empirical moments), and $c = 0$ is the threshold. Preliminary results (not reported) suggest that for those revision processes for which standard linearity tests (Luukkonen et al., 1988; Van Dijk et al., 2002) cannot reject the alternative of logistic smooth-transition specifications, the estimated smoothing parameter is often quite sizeable, leading the logistic model to collapse into a threshold model.

[6] Using $y_t^{t+\ell}$ is more appropriate than using y_t^{t+k} to compute business cycle features, under the assumption that final data can measure these features more accurately than preliminary announcements.

where the second term in $\Phi(\cdot)$ is a weighted geometric average intended to smooth past growth rates of GDP.[7] Since the transformation applied to $y_t^{t+\ell}$ is not consistent with a proper rationality analysis, we do not consider tests for unbiasedness and efficiency accounting for business cycle acceleration features.

2.5. Testing for changes in NSAs performance

Revision processes and rationality features may evolve over-time according to changes in the accounting performance of NSAs. For instance, statistical agencies can improve their data collection and processing, with a significant reduction of the impact of future revisions on preliminary announcements. Examples of these improvements may be found in the refining of seasonal adjustment procedures, the inclusion of new early indicators conveying information about current activity in specific sectors, and the reshaping of accounting methodologies. Changes can be either abrupt or smooth, depending on whether they take place, for instance, in the context of a benchmark revision or progressively over-time. The former can be compared to a break in the revision process and in the forecast rationality, while the latter takes the form of a time-varying process. The presence of breaks in our data can be explored by the means of a test for structural changes on the parameters of the forecast rationality regressions. Here we follow Swanson and Van Dijk (2006) and we implement the *AveF* Wald test proposed by Andrews (1993) and Andrews and Ploberger (1994), with approximate asymptotic p-values computed using the method of Hansen (1997, 2000). This is a test of the null hypothesis of parameters constancy against the alternative of one structural change at break date τ, the latter being endogenously estimated. More formally, the testing equation takes the following form:

$$ {}_{\ell}y_t^k = \left(\alpha_1 + y_t^{t+k}\beta_1 + \mathbf{X}'_{t+k}\boldsymbol{\gamma}_1\right)\mathbb{I}(\tau) + \left(\alpha_2 + y_t^{t+k}\beta_2 + \mathbf{X}'_{t+k}\boldsymbol{\gamma}_2\right)[1-\mathbb{I}(\tau)] + \varepsilon_{t+k}, \tag{7} $$

where $\mathbb{I}(\tau)$ is an indicator function, taking value of 1 if $t < \tau$ and 0 otherwise, and τ is the date at which the SSR of the testing regression is minimized. If the null hypothesis is rejected, Equation (7) is re-estimated with a structural change imposed. It is worth noticing that the test is valid against the alternative of one-time change in parameters. While restrictive, this assumption is not too strong in presence of relatively small samples, because parameters are not expected to change significantly too often. Tests for the hypothesis of multiple breaks are provided in the literature (Bai and Perron, 1998, 2003), but not implemented in this paper.

Smooth changes may take place when the performance of NSAs follows a slow, but continuous, improvement process. This can be consistent with a progressive introduction of new accounting procedures and methods, rather than through regular shocks operated by benchmark revisions. As a result, efficiency may gradually increase over-time. This hypothesis is explored here by the means

[7] The use of the alternative functions suggested by Dynan and Elmendorf (2001), namely $\Phi(y_t^{t+\ell}) = y_t^{t+\ell} - y_{t-1}^{t+\ell}$ and $\Phi(y_t^{t+\ell}) = y_t^{t+\ell} - (y_{t-1}^{t+\ell})^{4/7}(y_{t-2}^{t+\ell})^{2/7}(y_{t-3}^{t+\ell})^{1/7}$, provided qualitatively similar results and they are hence not reported.

of the joint test for parameter instability in Equation (3) against the alternative of a martingale process (the L_C test statistic; see Nyblom, 1989, and Hansen, 1992). We then take the result of this test as a rough evidence of the presence of parameters instability.

3. The data

Seasonally-adjusted vintages of quarterly French GDP growth rates are collected in a real-time dataset spanning from Q1 1991 (vintage Q2 1991) to Q1 2015 (vintage Q2 2015), such that the data are contained in an upper-triangular matrix, where the main diagonal counts 97 observations (the preliminary announcements). The subsequent analysis is then carried out on releases and revisions extracted from this dataset. It is worth noticing that the French NSA (INSEE hereafter) releases two preliminary announcements of GDP within the same quarter: the first one (named the "preliminary estimate") is released about 45 days after the end of the quarter and it provides an early estimate of transactions on goods and services, while the second one (named the "detailed estimate") is released about 85 days after the end of the quarter and it provides an early estimate of agent accounts.[8] Since GDP values are usually only slightly revised between these two preliminary releases, in what follows we refer exclusively to the former and compute our real-time dataset based on vintages of "preliminary estimates" only.

Vintage GDP estimates are periodically affected by a few structural revisions, the so-called "benchmark revisions". Unlike regular revisions, which add news or reduce noise, benchmark revisions usually involve base-year changes, weighting changes, seasonal adjustment refining, introduction of new indicators for the calibration of quarterly measures, and general reshaping of the accounting methodology. These revisions are unpredictable by construction and they do not represent any news or noise component, so that they should be offset by the econometrician. The solution proposed in this paper is to set the value of benchmark revisions to the unconditional mean of regular revisions by proceeding as follows: first, we regress the vector of the k-th revisions on a constant and a set of dummy variables, taking value 1 at the quarter known to involve a benchmark revision and 0 elsewhere; second, we select through an automatic algorithm (*Autometrics*; see Doornik, 2009) statistically significant benchmark revisions (up to 1% nominal level); finally, we remove from the original series the effect of these revisions on the unconditional mean (*i.e.* the estimated coefficient of selected dummy variables). A similar procedure is implemented in order to offset "irregular revisions", *i.e.* those revisions that are strictly related to rare growth episodes, such as the *Great Recession*, and that can be considered as outliers.[9]

[8] Compared to INSEE, both the BEA (US) and the ONS (UK) release three preliminary announcements of GDP.

[9] In practice, we consider the following vintages involving non-regular revisions: Q2 1999, Q2 2005, Q2 2011, Q2 2014 (benchmark revisions; base-change to 1995, 2000, 2005 and 2010); Q2 2008, Q3 2008, Q4 2008, Q1 2009 (irregular revisions due to the *Great Recession*); all the Q2 vintages (benchmarking of quarterly accounts on annual accounts).

In the literature, the series of final releases ($y_t^{t+\ell}$) is usually defined as either the series corresponding to a specific vintage or the latest available vintage (where $\ell = \infty$), or alternatively the series constructed in such a way that there are no further revisions (excluding benchmark revisions) after a finite number of periods. While the first option is not necessarily optimal, due precisely to benchmark revisions, the second one is more appealing because ℓ can be set as to replicate the schedule for revisions followed by NSAs (Aruoba, 2008).[10] For this reason, it is often reasonable to set a value for ℓ ranging between 12 and 14. An interesting but problematic issue with French GDP vintages is that this definition of final release, *i.e.* a vintage immune from further (regular) revisions, does not seem to apply to actual data. This means that values released in old vintages keep being revised substantially even in more recent vintages. Causes for such a behavior of data can be found, for instance, in the continuous updating of the seasonal-adjustment procedure. To address this issue, we compute the vintage of final release data by following the official regular revision schedule employed by INSEE: GDP values released in year T can be considered as fully revised when the preliminary announcement of the first quarter of year $T+3$ is released. However, this strategy comes to a price, because it follows from above that the number of revisions between a preliminary announcement at time T and its final release at time $T+3$ varies across quarters: 12 for Q1, 11 for Q2, 10 for Q3 and 9 for Q4. In the subsequent analysis we thus set ℓ consistently with this schedule, also meaning that the last observation available for the sample of final releases $y_t^{t+\ell}$ is set to Q4 2012.

Figure 1 reports preliminary announcements of GDP and the last vintage series (Q2 2015), as well as the range defined by the historical maximum and minimum values of GDP releases. A visual inspection of the series suggests that, except for a few observations (mostly concentrated at the beginning of the sample and related to the general strike episode at the end of 1995, as well during as the *Great Recession*), preliminary announcements and last vintage data are relatively close. However, the range of historical releases appears quite large and points to a pattern of substantial variation across vintages. Figure 2 reports final GDP revisions (${}_\ell y_t^1$) computed as the difference between the preliminary announcements and either the last available vintage series ($\ell = \infty$) or the series constructed using the revision schedule described above (both with benchmark and irregular revisions not offset). These two series have overall similar moments (tests reject the hypotheses of different mean and variance), but display sometimes noticeable differences in terms of both size and sign, whether revisions are computed for recent or past releases.

[Figures 1 and 2 about here]

The path of the main k-th revisions considered alongside the present contribution ($k = 1, 2, 3, 4, 8, \ell - 1$) can be roughly explored in Figure 3, which reports boxplots of the distribution of FXW, INW and REM revisions, respectively, described by Equations (2a), (2b) and (2c). From the boxplots, we

[10] For instance, the BEA follows a schedule of 13 revisions, so that a GDP value becomes "final" after 14 releases.

can observe that the mean and the median of revisions are usually close to zero, while the variance tends to decrease or increase consistently with the order and the nature of revisions. For instance, the size of FXW revisions is quite small (ranging between ± 0.2) and the variance seems to decrease progressively, since the accuracy of k-th releases with respect to final releases is expected to increase overtime. An intriguing exception is represented by the last revision, for which the dispersion seems to increase compared to the previous steps. Preliminary results (not reported) revealed that this finding is mainly due to our choice of ℓ, which is set to change across quarters. Fixing ℓ to, say, 13 is sufficient to attenuate this issue. On the other hand, as expected the variance of INW revisions tends to increase (see Figure 3b), while that of REM revisions tends to decrease (see Figure 3c). Further, the empirical distribution of ${}_{\ell}y_t^{t+1}$ indicates that the final revision ranges in the interval ± 0.5, which is quite large compared to the average of preliminary announcements (about 0.4%). Finally, it is quite clear-cut from Figures 3a and 3c that ${}_{\ell}y_t^{\ell-1} \neq 0$, meaning that the last revision does not reveal the true value of GDP overall.

[Figure 3 about here]

4. Empirical results

4.1. GDP Revisions: main statistical features

We start our empirical analysis by investigating the main statistical properties of French GDP revisions. For $k = 1, 2, 3, 4, 8, \ell - 1$, Table 1 reports descriptive statistics for FXW revisions (Panel A), INW revisions (Panel B), and REM revisions (Panel C). Average revisions (μ) appear overall small and statistically not different from zero, which means that revisions are broadly unconditionally unbiased. As expected, volatility (σ) and mean absolute revisions ($|\mu|$) are a decreasing function of k for FXW and REM revisions, but an increasing function for INW revisions (see also Figure 3). These findings are consistent with the increasing rate of accuracy of later releases, as new available data convey progressively less information for the refining of GDP estimates. According to the Jarque-Bera test, revisions appear overall normally distributed.

[Table 1 about here]

Since the mean and the standard deviation of final revisions may not be very informative *per se* of the size of revisions, we also report ratios of mean revisions to mean releases as well as noise-to-signal ratios, the latter being defined as the standard deviation of revisions divided by the standard deviation of releases. We propose these statistics to shed light on the relative size of revisions (in terms of first two moments) compared to the underlying original data (Aruoba, 2008). As expected, results point to negligible size for mean revisions compared to mean releases. However, results for noise-to-signal ratios appear more substantial, ranging mostly between 0.13 and 0.60 depending on the nature and the order of revisions.[11] Hence, findings suggest that revisions may be sizeable compared to releases, meaning that one of the desirable properties of revisions (σ small) is broadly not supported by our data.

Table 2 reports results on the correlation of revisions with the business cycle. According to the Wald test for the hypothesis of equal mean revisions across standard business cycle phases (Business Cycle I in the table), asymmetry can be detected at $k = 1, 3, 8$ for FXW revisions, for all k for INW revisions, and for $k = 1, 2$ for REM revisions. When the null hypothesis is rejected, results point overall to zero mean revisions in phases of expansion and to negative mean revisions in phases of contraction, the latter suggesting an overestimate of GDP during recessions. It is worth noticing that, for the case of INW revisions, the size of mean revisions during contraction periods is an increasing function of k, suggesting that the asymmetry increases with the number of revisions. The opposite holds for REM revisions. Turning to a threshold definition of the business cycle (Business Cycle II in the table), the hypothesis of equal mean revisions across positive and negative growth phases can be rejected at $k = 8$ for FXW revisions, $k = 8, \ell - 1$ for INW revisions, and $k = 1, 2, 3, 4$ for REM revisions. Again, results point to sizeable overestimates of GDP during phases of negative growth rates, while the bias appears overall negligible when GDP growth rates are positive. All in all, these findings suggest that preliminary announcements suffer from an overestimation bias during phases of contraction of GDP compared to their final releases, but this bias progressively disappears, such that last revisions are not affected by any business cycle asymmetry. Finally, there is some evidence of significant correlation between the acceleration of GDP and FXW revisions at $k = 3, 4$ and INW revisions at $k = 4, 8, \ell - 1$. With respect to REM revisions, significant correlation between the acceleration of GDP and revisions can be found at $k = 1, 2, 3$. Estimated correlation coefficients are positive signed, suggesting that revisions become more positive when the economy is speeding up and less positive when the economy is slowing down.

[Table 2 about here]

Table 3 reports results on the correlation of the volatility of revisions with the business cycle. According to Wald tests, we have some evidence, although often weak, of standard business cycle asymmetry at $k = 2$ for FXW revisions and $k = 2, 3, 4, 8$ for INW revisions. When a threshold definition of the business cycle is considered, asymmetry can be detected at $k = 1, 2$ for FXW revisions and $k = 1$ for INW revisions. Conversely, the volatility of INW revisions does not seem to be affected by any business cycle asymmetry. Results point overall to an increase in the volatility of early GDP revisions during phases of contraction or negative GDP growth, but this asymmetry

[11] Since $\sigma_{{}_h y_t^k} / \sigma_{y_t^{t+h}} = \sqrt{1 - \left[\sigma^2_{y_t^{t+k}} \left(1 - 2\rho\sigma^{-1}_{y_t^{t+k}} \sigma_{{}_h y_t^k}\right) / \sigma^2_{y_t^{t+h}}\right]}$, the noise-to-signal ratio is bounded below by zero, but not necessarily bounded above by unity due to the sign of the coefficient ρ, that denotes the correlation between y_t^{t+k} and ${}_h y_t^k$.

progressively disappears. There is also evidence of significant correlation between the acceleration of GDP and FXW revisions (and hence also for INW revisions) at $k = 1$. The estimated correlation coefficient is positive signed, suggesting that volatility of the first revision increases when the economy is speeding up and decreases when the economy is slowing down. With respect to REM revisions, no significant correlation can be found.

[Table 3 about here]

Results for the instability in mean and volatility of revisions are reported in Table 4. According to structural change test results, the hypothesis of stability can be only seldom rejected for mean revisions: at $k = 2$ for FXW revisions, $k = 2, 4$ for INW revisions and $k = 3, 4$ for REM revisions. Significant break dates are usually estimated around the end of the 90s and the begininning of the 2000s. Conversely, evidence of structural change is substantially stronger for the volatility of revisions, for which the hypothesis of stability can be quite frequently rejected across revisions. Turning to time-varying parameters tests, the results are broadly consistent with the findings reported above: the hypothesis of stability can be often rejected for the volatility of revisions, while mean revisions appear statistically stable. All in all, results suggest that only the second moment of revisions has changed over-time, and additional investigation on the sign of these changes (not reported) reveals that volatility has overall decreased in the last decade. These findings would support the idea of a recent improvement in the accuracy of data releases.

[Table 4 about here]

4.2. Data rationality

As discussed in Section 2.1, we break down the hypothesis of forecast rationality into a conditional unbiasedness hypothesis and an efficiency hypothesis. We start with the test of conditional unbiasedness of revisions, by estimating Equation (3) with the imposed restriction $\boldsymbol{\gamma} = 0$ and setting the joint null hypothesis $\alpha = \beta = 0$. Wald test results, reported in Table 5, suggest that revisions are overall conditionally unbiased. This finding is due to the absence of both substantial systematic prediction errors (α) and significant correlation between revisions and their respective k-th releases (β). Business cycle asymmetry is tested as described in Sections 2.2 and 2.3. The null hypothesis of standard business cycle asymmetry, $F^*(s)$, cannot be rejected at $k = 1, 2$. When the asymmetry is imposed, unbiasedness can be rejected at $k = 1, 2$ as a result of significant conditional bias in the contraction regime. When a threshold definition of the business cycle is considered, asymmetry and unbiasedness can be instead detected at $k = 2, 3$. Again, these findings arise from the contraction regime, for which unbiasedness can be nevertheless rejected only at $k = 2$. Structural change tests provide no evidence of parameters instability. However, time-varying tests reject the null of stability at $k = 1, 3, 4$, mainly due to changes in the volatility of regression residuals.

[Table 5 about here]

Testing for efficiency implies relaxing the restriction on $\boldsymbol{\gamma} = 0$. For ease of analysis, we split the vector $\mathbf{X}_{t+k}$, representing the information available at time $t+k$, into two groups of indicators. The first one includes past features of the revision process, such as its first lag $({}_{\ell}y_{t-1}^{k})$ and the revision accounted between the first and the k-th release (*i.e.* the k-th INW revision ${}_{k}y_{t}^{1}$, for $k > 1$), with the aim of conditioning each revision sequence on its own history. The second group $(\mathbf{x}_{t+k})$ includes standard external indicators, known at the time of the release of y_t^{t+k}, such as the spread between yields on long- and short-term Government bonds, the quarterly return on stock market index (SBF 250) and the quarterly change of crude oil (Brent) prices (see, for instance, Faust et al., 2005, and Swanson and Van Dijk, 2006). Recent contributions to the literature on nowcasting French GDP pointed out the predictive content of survey data on services and construction sectors when the target variable is implicitly the final GDP growth, while they appear far less useful than survey data on manufacturing sector when the target variable is explicitly the first-release GDP growth (see Mogliani et al., 2014, and references cited therein for a discussion). It follows that these indicators might carry valuable information on the revision process. We hence test this assumption by including business indicators on services and construction sectors released by INSEE in the information set $\mathbf{X}_{t+k}$. The data described above are sampled at monthly frequency, with small publication lags (usually, a few days). In order to enter the information set at time $t+k$, these variables are measured at the end of the quarter $t+k-1$, by taking their quarterly average. Since the INSEE surveys are designed to collect opinions over the past and following three months, we only consider the third-month value for each quarter. Finally, we include in the efficiency regressions a set of centered seasonal dummies $\sum_{s=1}^{3} \delta_s D_{s,t}^{*}$, where $D_{s,t}^{*} = D_{s,t} - D_{4,t}$, with $D_{s,t} = 1$ if time period t corresponds to quarter s and 0 otherwise, and $\delta_4 = -\sum_{s=1}^{3} \delta_s$. We follow Swanson and Van Dijk (2006) and we report $\delta^{*} = \sqrt{\sum_{s=1}^{4} \delta_s^2}$ as a measure of the importance of seasonal effects.

Results based on efficiency regressions including information on the revisions history, $\mathbf{X}_{t+k} = \left({}_{k}y_{t}^{1}, {}_{\ell}y_{t-1}^{k}\right)'$, are reported in Table 6. Wald test results suggest that revisions are overall efficient: the null hypothesis of $\alpha = \beta = \boldsymbol{\gamma} = 0$ cannot be rejected for all k. This result is due to both the unbiasedness of revisions ($\alpha = \beta = 0$) and the absence of significant correlation with the revision history ($\boldsymbol{\gamma} = 0$). On the other hand, the null hypothesis of standard business cycle asymmetry cannot be rejected for all k, excluding $k = \ell - 1$. When the asymmetry is imposed, efficiency can be now rejected for all k with the exception of $k = \ell - 1$, as a result of significant correlation of revisions with $\mathbf{X}_{t+k}$ in the contraction regime. When a threshold definition of the business cycle is considered, the picture is quite different: asymmetry can be rejected only at $k = \ell - 1$, while efficiency can be rejected at $k = 2$ when asymmetry is imposed. Again, these findings arise from the contraction regime. Structural change tests provide some evidence of parameters instability at $k = 2, 4$, with break dates detected in 1997 and 2002. Time-varying tests can reject the joint hypothesis of stability at $k = 1, 3$, with time-varying volatility detected at $k = 1, 2, 3, 4$.

[Table 6 about here]

Results based on efficiency regressions including information on external indicators, $\mathbf{X}_{t+k} = \mathbf{x}_{t+k}$, are reported in Table 7. Unlike the findings described above, test results suggest that the inclusion of additional indicators leads to strong rejections of the efficiency hypothesis. Indeed, the null hypothesis of $\alpha = \beta = \boldsymbol{\gamma} = 0$ can be strongly rejected for all k, except at $k = \ell - 1$, because of the presence of significant correlation between revisions and the selected indicators. Detailed results (not reported) reveal that failure to reject inefficiency can be mainly attributed to the interest rate spread and the survey indicator on services. Furthermore, revisions do not seem to be affected by seasonal patterns, as δ^* is not sizeable compared to the conditional mean. The null hypothesis of standard business cycle asymmetry cannot be rejected for all k. When the asymmetry is imposed, efficiency can be strongly rejected and we observe significant correlation of revisions with $\mathbf{X}_{t+k}$ in the contraction regime. This finding appears robust to the definition of the the business cycle: when a threshold business cycle asymmetry is imposed, the efficiency hypothesis can be rejected for all k, but also at $k = 1, 2, 4, 8$ when the test is performed over the expansion regime. Conversely, both structural change tests and time-varying tests provide only moderate evidence of parameters instability, arising mainly from the volatility of regression residuals.

[Table 7 about here]

These findings are broadly consistent with those provided by the testing approach proposed by Corradi et al. (2009). Results (reported in Table 8) suggest that the null hypothesis of generic nonlinear efficiency cannot be rejected when the testing regressions include only information on the revisions history, but it can (at 10% level) when external indicators are considered, except for $k = 4$.

[Table 8 about here]

All in all, the results reported above suggest that GDP releases are conditionally unbiased, but inefficient when revisions are conditioned on external indicators. Results provide evidence of strong correlation of revisions with macroeconomic and financial indicators, as well as of the asymmetry with respect to phases of the business cycle. The failure to reject inefficiency means that revisions do not contain news but reduce statistical noise, which also means that they are technically predictable. In the next section, we investigate this issue by building forecasting models of GDP revisions and performing an evaluation analysis of point and density forecasts.

5. Forecasting GDP revisions in real-time

5.1. Design of the forecasting experiment

The out-of-sample analysis is based on recursive regressions over a hold-out sample spanning from T_1=Q1 2004 to T_f=Q4 2012. We compute one-step ahead forecasts of GDP revisions using the following generalized forecast rationality specification:

$$ {}_{\ell}y_{t+h}^{k} = \alpha + \sum_{j=1}^{4} \varphi_j \, {}_{\ell}y_{t-j}^{k} + \sum_{j=0}^{4} \beta_j y_{t-j}^{t+k} + \sum_{j=0}^{4} \mathbf{X}_{t-j+k}' \boldsymbol{\gamma}_j + \mathbf{D}_t^{*\prime} \boldsymbol{\delta} + \varepsilon_{t+k}, \tag{8} $$

where $h = 0$ for ease of notation $\mathbf{X}_{t+k}$ is the full set of indicators. This ARDL specification has the advantage of providing a more flexible representation of GDP revisions than the rationality regression described in Section 2.1. However, it has the drawback of being heavily parameterized (large number of parameters to be estimated relative to the number of observations), leading to in-sample overfit and out-of-sample poor performance compared to more parsimonious models. To address this issue, we implement a shrinkage method to model selection based on the GEneral-To-Specific (GETS) approach popularized by Krolzig and Hendry (2001) and implemented here through the *Autometrics* algorithm (Doornik, 2009). This model reduction algorithm can deal with the 2^N paths generated by Equation (8), where N is the number regressors (including the deterministic terms), focusing efficiently on a subselection of paths only. Model selection is thus obtained through simplification of the general model by sequentially eliminating those variables failing a battery of forward and backward tests of inclusion in the parsimonious final model. A training sample spanning from Q1 1991 to Q4 2003 is used to select the best model specifications. However, to avoid misspecification issues related to this initial model selection, the specifications are allowed to evolve overtime through a sequential update taking place every four quarters.[12]

Point and density forecasts are evaluated relatively to the performance of a benchmark ARMA(p,q) model, where the autoregressive and moving-average orders are set by optimizing the BIC criterion.[13] Let us denote T_h the number of out-of-sample observations, $\tilde{e}_{t+h+k} = {}_{\ell}y_{t+h}^{k} - {}_{\ell}\tilde{y}_{t+h}^{k}$ the forecast errors of the benchmark ARMA model, and $\hat{e}_{t+h+k} = {}_{\ell}y_{t+h}^{k} - {}_{\ell}\hat{y}_{t+h}^{k}$ the forecast errors of the selected competing model. The out-of-sample predictive accuracy is then measured in terms of relative root mean squared forecast error (RMSFE):

$$ \text{RMSFE}_k = \sqrt{\sum_{t=1}^{T_h} \left(\hat{e}_{t+h+k}\right)^2 \Big/ \sum_{t=1}^{T_h} \left(\tilde{e}_{t+h+k}\right)^2}, \tag{9} $$

where lower values suggest that the selected model outperforms the benchmark model. Mean squared error differences are evaluated through the Diebold and Mariano (1995) and West (1996) test (DMW hereafter) for unconditional equal predictive accuracy, with a small-sample adjustment to the consistent estimate of the variance proposed by Harvey et al. (1997). It is also interesting to

[12] Stability and reliability of model estimates are ensured by including selected impulse dummies in Equation (8), following the indicator saturation approach developed by Santos et al. (2008). See also Johansen and Nielsen (2009) and Castle et al. (2012).

[13] We set the maximum AR and MA orders to 4. Further, following the design of GETS approach described in Section 5.1, the selected specifications are updated every four quarters.

investigate the ability of our models to predict the sign of GDP revisions, other than their values. For this purpose we implement the Pesaran and Timmermann (1992, 2009) sign test (PT hereafter).

Density forecasts $g^k_{t+h}\left({}_\ell y^k_{t+h}\right)$ are computed under the assumption of normality of the probability distribution of predictions. From the sequence of predictive densities, we compute the mean log-predictive score (MLPS) as the realization of the variable evaluated at the out-turn of the probability densities, averaged over T_h:

$$\text{MLPS}_{y,k} = T_h^{-1}\sum_{t=1}^{T_h}\log S\left({}_\ell y^k_{t+h}\right),$$

where $\log S\left({}_\ell y^k_{t+h}\right) = \log g^k_{t+h}\left({}_\ell y^k_{t+h}\right)$. For our set of predictions ${}_\ell\hat{y}^k_{t+h}$ and ${}_\ell\tilde{y}^k_{t+h}$, we report the relative density performance, $\Delta\text{MLPS}_{y,k} = \text{MLPS}_{\hat{y},k} - \text{MLPS}_{\tilde{y},k}$, such that positive values suggest that the competing model outperforms the benchmark. Further, equal density forecast accuracy is here investigated by the means of a DMW-type test, in the lines of Mitchell and Hall (2005) and Amisano and Giacomini (2007). Density forecasts are also evaluated individually over their probability integral transform (PIT) values:

$$z^k_{t+h} = \int_{-\infty}^{{}_\ell y^k_{t+h}} g^k_{t+h}(u)\mathrm{d}u.$$

If g^k_{t+h} coincides with the sequence of true densities $\bar{g}^k_{t+h}$, then the sequence of PITs z^k_{t+h} is uniformally distributed (Diebold et al., 1998). In this case, it can be shown that the inverse normal transformation of PITs, given by $z^{k,*}_{t+h} = \Phi^{-1}(z^k_{t+h})$, where $\Phi^{-1}(\cdot)$ is the standard normal distribution function, is normally distributed. We implement the Doornik and Hansen (2008) test (DH hereafter) to check for normality of the CDF of $z^{k,*}_{t+h}$, while we test for the joint hypothesis of independence and normality of $z^{k,*}_{t+h}$ through the Berkowitz (2001) likelihood-ratio test (Bk hereafter) against a first-order autoregressive alternative (see also Mitchell and Wallis, 2011).

5.2. Results

Consistently with the analysis reported in the previous sections, results for $k = 1, 2, \ldots, \ell - 1$ are presented in Table 9. It is nevertheless worth noticing that, from the point of view of the policy-maker, results for the case $k = 1$ are certainly more relevant for the conduct of monetary and/or fiscal policy than for $k > 1$. Hence, in what follows we mainly focus our discussion on the former case.

[Table 9 about here]

Forecast results for the model presented in Equation (8) are reported in Panel A of Table 9. With respect to point forecasts, we find that the selected specification at $k = 1$ does not outperform the ARMA benchmark, given a slightly positive RMSFE. However, the PT test can reject at 10%

significance level the hypothesis of distributional independence between (the binary transformation of) ${}_\ell y^1_{t+1}$ and ${}_\ell \hat{y}^1_{t+1}$, so that the model displays some predictive power for the sign of the final revision of the preliminary announcement of GDP. Slightly better results in terms of RMSFE are obtained for the specifications at $k = 2$ and $k = 3$, but the DMW test does not reject the hypothesis of equal predictive accuracy with the benchmark. Similarly to the case $k = 1$, PT test results point to a predictive ability of the sign of revisions for these specifications. For the case $k > 3$, results are overall disappointing on the side of both point and sign accuracy. This means that, in spite of the efficiency results reported in the first panel of Table 7, the indicators selected among the vector $\mathbf{X}_{t+k}$ have low predictive power for later revisions. This is also suggestive of the presence of a large share of news revisions in late releases of French GDP. With respect to density forecasts, results for relative accuracy point to a very low outperformance of the specification at $k = 1$, which is nevertheless not statistically significant. Predictive densities appear also poorly fitted, according to both DH and Bk tests, which reject their null at 5% level. Better results can be found at $k = 2$ and $k = 4$, for which the relative density performance of the selected specifications is statistically different from that of the benchmark. Further, goodness-of-fit of predictive densities cannot be rejected at any conventional significance level.

All in all, the findings reported above suggest that even though releases appear overall inefficient, revisions can be only poorly predicted using a simple linear model. However, from the results reported in Section 4.2, forecast inefficiency can be strongly detected when business cycle asymmetry is imposed to the rationality testing regression. We can hence exploit this feature and extend the forecasting model in (8) to a non-linear formulation, which takes the form of a threshold regression. We then impose the threshold business cycle asymmetry described in Section 2.3 to the forecasting model, we implement the selection algorithm, and we compute the forecasts. Results are reported in Panel B of Table 9. Compared to our previous findings, the selected specification at $k = 1$ does now substantially outperform the ARMA benchmark: the predictive gain is about 22% according to RMSFE and statistically significant according to the DMW test. Further, the sign test strongly rejects the null of independence. Hence, we have clear-cut evidence of predictive power of the selected indicators (mainly survey data) when business cycle asymmetry is allowed to the forecasting equation. We find similar results for the specifications at $k = 3$ and $k = 4$, for which the predictive gains are positive (about 14% and 12%, respectively) and statistically different from zero, and the PT test rejects the null at 5% level. For the specification at $k = 2$, we cannot find any predictive improvement compared to the benchmark in terms of point forecasts, but we have strong evidence of sign accuracy. Finally, the specifications at $k = 4$ and $k = \ell - 1$ perform similarly to the symmetric model. With respect to density forecasts, we have strong evidence of relative density outperformance of the benchmark at $k = 1, 2, 4$. Further, goodness-of-fit tests cannot reject the hypothesis on normality and independence of predictive densities overall.

As stated in Section 4.2, rejection of the efficiency hypothesis under business cycle asymmetry can be mainly attributed to the significant correlation of revisions with external information in the contraction regime. Hence, one can question whether our previous findings are actually driven by forecasts of revisions falling in quarters displaying a negative GDP growth. To shed light on this issue, we perform an out-of-sample evaluation of a sub-set of predictions covering the expansion regime only. Results, reported in in Panel C of Table 9, are broadly in line with the findings reported above: the specifications at $k = 1, 3, 4$ outperform the benchmark, with a statistically significant (except for $k = 3$) predictive gain ranging between 10-16%. We also observe an interesting improvement of predictive accuracy at $k = 2$, although the gain does not appear statistically different from zero. With respect to density forecasts, results reported in Panel C are not substantially different from those reported in Panel B, suggesting a significant outperformance of the benchmark at $k = 1, 2, 4$ and an overall goodness-of-fit of the estimated densities.

To summarize, French GDP revisions are statistically inefficient and hence possibly predictable. While linear models fail to provide a reasonable accuracy compared to a simple benchmark, we nevertheless show that better results can be obtained by allowing for business cycle asymmetry in the forecasting equations.

6. Concluding remarks

In this paper, we performed a statistical analysis of French GDP revisions, based on quarterly data spanning from 1991 to 2014. The results pointed to a bunch of interesting features of revisions. First, mean revisions are not statistically different from zero, meaning that GDP releases are overall unbiased. Second, the volatility of revisions decreases with the order of releases, which is consistent with the idea that an increasing rate of accuracy of later releases may be driven by either additional news or reduced measurement errors, or even both, conveyed by revisions. Third, we allowed for a possible asymmetry in revisions due to the business cycles and the hypothesis was strongly accepted. When asymmetry is imposed, we found out that the statistical features of revisions are substantially different across business cycle regimes, notably during phases of GDP contraction or slowdown. In that case, preliminary GDP announcements are overestimated and the volatility of the first revisions is significantly high. We then performed a further analysis in order to assess the rationality of GDP releases. Following Faust et al. (2005) and Swanson and Van Dijk (2006), the hypothesis of rationality was broken down into two sub-hypotheses: unbiasedness (revisions are mean-zero and uncorrelated with GDP announcements) and efficiency (revisions are uncorrelated with information available at the time the release is made). Results suggest that both unbiasedness and efficiency hold overall. However, after taking into account business cycle asymmetry, both hypotheses can be strongly rejected. This result can be mainly attributed to the specific features of revisions during phases of GDP contraction or slowdown.

These results suggest that preliminary GDP announcements are overall inefficient, and that revisions can then be predicted. A noticeable finding provided in this paper is the presence of a systematic bias in preliminary GDP announcements during phases of contraction. Hence, further research should investigate how to attenuate this bias at the early stages of GDP measurement. However, we show that even though the information set considered here seems to convey some predictive content to GDP revisions, obtaining a reasonable forecasting performance requires a great deal of flexibility in data handling and model selection. This means that the efficient use of available information in the early stages of GDP measurement appears a rather hard task that goes largerly beyond the scope of data producers such as NSAs. It follows that the quality of measurements provided by INSEE when releasing quarterly National Accounts is therefore far from being questioned here.

References

Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. Journal of Business & Economic Satistics 25 (2), 177–190.

Andrews, D. W. K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.

Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61 (4), 821–856.

Andrews, D. W. K., Monahan, J. C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60 (4), 953–966.

Andrews, D. W. K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62 (6), 1383–1414.

Aruoba, S. B., 2008. Data revisions are not well behaved. Journal of Money, Credit and Banking 40 (2-3), 319–340.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66 (1), 47–78.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18 (3), 1–22.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business & Economic Statistics 19 (4), 465–474.

Bishop, J., Gill, T., Lancaster, D., March 2013. GDP revisions: Measurement and implications. RBA Bulletin - Reserve Bank of Australia (March), 11–22.

Brodsky, N., Newbold, P., 1994. Late forecasts and early revisions of United States GNP. International Journal of Forecasting 10 (3), 455–460.

Burns, A. F., Mitchell, W. C., 1946. Measuring Business Cycles. New York (US): NBER.

Castle, J. L., Doornik, J. A., Hendry, D. F., 2012. Model selection when there are multiple breaks. Journal of Econometrics 169 (2), 239–246.

Corradi, V., Fernandez, A., Swanson, N. R., 2009. Information in the revision process of real-time datasets. Journal of Business and Economic Statistics 27 (4), 455–467.

Corradi, V., Swanson, N. R., 2002. A consistent test for non-linear out-of-sample predictive accuracy. Journal of Econometrics 110 (2), 353–381.

Cotis, J. P., Coppel, J. G., 2005. Business cycle dynamics in OECD countries: Evidence, causes and policy implications. In: Kent, C., David, N. (Eds.), The Changing Nature of the Business Cycle. Sidney (Australia): RBA Annual Conference Volume, Reserve Bank of Australia, pp. 8–57.

Croushore, D., 2011. Frontiers of real-time data analysis. Journal of Economic Literature 49 (1), 72–100.

Den Haan, W. J., Levin, A., 1997. A preactionner's guide to robust covariance matrix estimation. In: Maddala, G. S., Rao, C. R. (Eds.), Handbook of Statistics: Robust Inference. New York (US): Elsevier, pp. 299–342.

Diebold, F. X., Gunther, A., Tay, K., 1998. Evaluating density forecasts with application to financial risk management. International Economic Review 39 (4), 863–883.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13 (3), 253–263.

Doornik, J. A., 2009. Autometrics. In: Castle, J. L., Shephard, N. (Eds.), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Oxford (UK): Oxford University Press, pp. 88–121.

Doornik, J. A., Hansen, H., 2008. An omnibus test for univariate and multivariate normality. Oxford Bulletin of Economics and Statistics 70 (S1), 927–939.

Dynan, K. E., Elmendorf, D. W., 2001. Do provisional estimates of output miss economic turning points? Finance and Economics Discussion Series 2001-52, Board of Governors of the Federal Reserve System.

Faust, J., Rogers, J. H., Wright, J. H., 2005. News and noise in G-7 GDP announcements. Journal of Money, Credit and Banking 37 (3), 403–419.

Hansen, B. E., 1992. Testing for parameter instability in linear models. Journal of Policy Modeling 14 (4), 517–533.

Hansen, B. E., 1997. Approximate asymptotic p-values for structural-change tests. Journal of Business & Economic Statistics 15 (1), 60–67.

Hansen, B. E., 2000. Testing for structural change in conditional models. Journal of Econometrics 97 (1), 93–115.

Harding, D., Pagan, A. R., 2002. Dissecting the cycle: A methodological investigation. Journal of Monetary Economics 49 (2), 365–381.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13 (2), 281–291.

Jacobs, J. P. A. M., van Norden, S., 2011. Modeling data revisions: Measurement error and dynamics of "true" values. Journal of Econometrics 161 (2), 101–109.

Johansen, S., Nielsen, B., 2009. An analysis of the indicator saturation estimator as a robust regression estimator. In: Castle, J. L., Shephard, N. (Eds.), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Oxford (UK): Oxford University Press, pp. 1–36.

Keane, M. P., Runkle, D. E., 1990. Testing the rationality of price forecasts: New evidence from panel data. American Economic Review 80 (4), 714–735.

Kiefer, N. M., Vogelsang, T. J., 2002a. Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. Econometrica 70 (5), 2093–2095.

Kiefer, N. M., Vogelsang, T. J., 2002b. Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. Econometric Theory 18 (6), 1350–1366.

Kiefer, N. M., Vogelsang, T. J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory 21 (6), 1130–1164.

Kiefer, N. M., Vogelsang, T. J., Bunzel, H., 2000. Simple robust testing of regression hypotheses. Econometrica 68 (3), 695–714.

Krolzig, H. M., Hendry, D. F., 2001. Computer automation of general-to-specific model selection procedures. Journal of Economic Dynamics and Control 25 (6-7), 831–866.

Luukkonen, R., Saikkonen, P., Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75 (3), 491–499.

Mankiw, N. G., Runkle, D. E., Shapiro, M., 1984. Are preliminary announcements of the money stock rational forecasts? Journal of Monetary Economics 14 (1), 15–27.

Mankiw, N. G., Shapiro, M., 1986. News or noise: An analysis of GNP revisions. Survey of Current Business 66 (5), 20–25.

Mincer, J., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J. (Ed.), Economic Forecasts and Expectations. New York (US): NBER, pp. 1–46.

Mitchell, J., Hall, S. G., 2005. Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation. Oxford Bulletin of Economics and Statistics 67 (S1), 995–1033.

Mitchell, J., Wallis, K. F., 2011. Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. Journal of Applied Econometrics 26 (6), 1023–1040.

Mogliani, M., Brunhes-Lesage, V., Darné, O., Pluyaud, B., 2014. New estimate of the MIBA forecasting model. Modeling first-release GDP using the Banque de France's monthly business survey and the "blocking" approach. Working Paper Series N. 473, Banque de France.

Mork, K. A., 1987. Ain't behavin': Forecast errors and measurement errors in early GNP estimates. Journal of Business & Economic Statistics 5 (2), 165–175.

Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84 (405), 223–230.

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

Pesaran, M. H., Timmermann, A., 1992. A simple nonparametric test of predictive performance. Journal of Business and Economic Statistics 10 (4), 461–465.

Pesaran, M. H., Timmermann, A., 2009. Testing dependence among serially correlated multicategory variables. Journal of the American Statistical Association 104 (485), 325–337.

Rathjens, P., Robins, R. P., 1995. Do government agencies use public data? The case of GNP. Review of Economics and Statistics 77 (1), 170–172.

Santos, C., Hendry, D. F., Johansen, S., 2008. Automatic selection of indicators in a fully saturated regression. Computational Statistics 23 (2), 317–335.

Swanson, N. R., Van Dijk, D., 2006. Are statistical reporting agencies getting it right? Data rationality and business cycle asymmetry. Journal of Business & Economic Statistics 24 (1), 24–42.

Van Dijk, D., Teräsvirta, T., Franses, P. H., 2002. Smooth transition autoregressive models - a survey of recent developments. Econometric Reviews 21 (1), 1–47.

West, K. D., 1996. Asymptotic inference about predictive ability. Econometrica 64 (5), 1067–1084.

Figure 1: Preliminary announcements, last vintage release, and minimum-maximum releases of French GDP

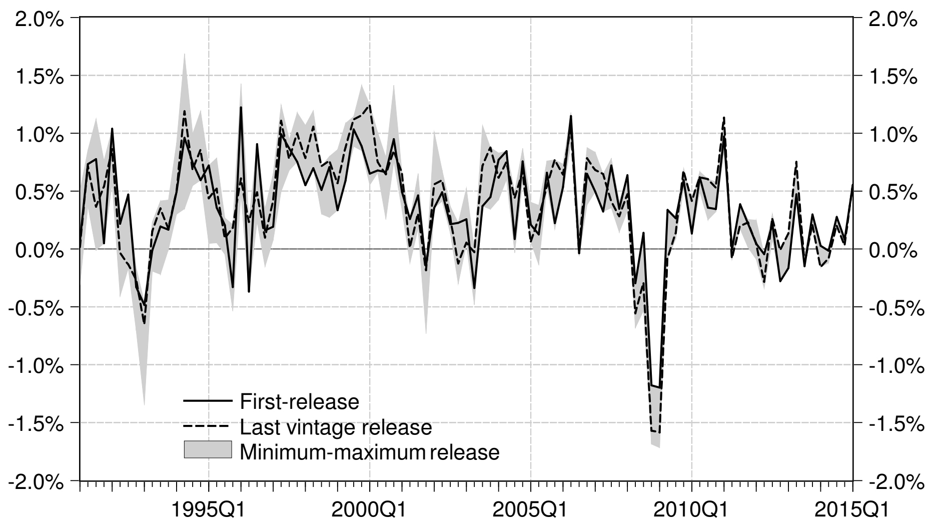


Figure 2: Final revisions of French GDP (${}_{\ell}y_t^1$): latest vintage ($\ell = \infty$) and scheduled revisions

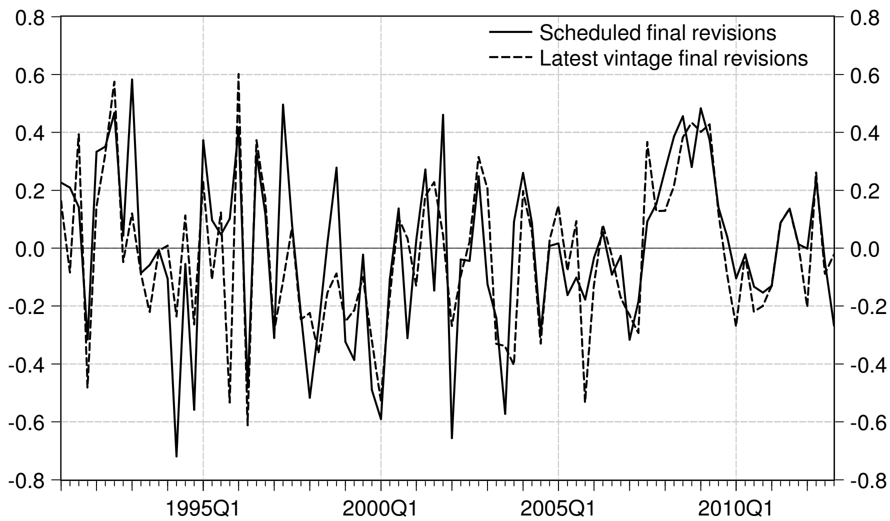
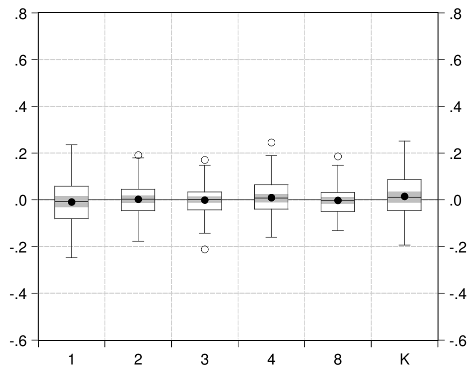
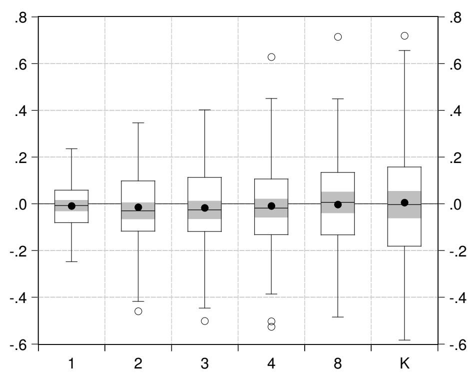


Figure 3: Boxplot of French GDP regular revisions

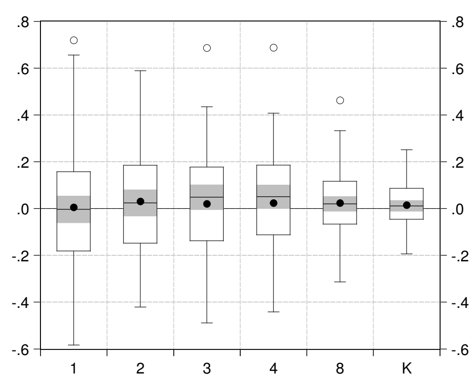
(a) Fixed-width revisions



(b) Increasing-width revisions



(c) Remaining revisions



Notes: The box denotes the hinges (first and third quartile). The staples denote the last data point within (or equal to) each inner fence (first quartile plus/minus 1.5 times the interquartile range, *i.e.* the difference between the hinges). Full dots denote the mean, empty dots and stars denote the outliers (observations outside the inner fences). The x-axis denotes the k-th revision, where $K = \ell - 1$.

Table 1: Main descriptive statistics: Mean (μ) and volatility (σ) of GDP revisions

	$k=1$	$k=2$	$k=3$	$k=4$	$k=8$	$k=\ell-1$
A. FXW revisions						
T	96	95	94	93	89	88
μ	-0.01 (0.01)	0.00 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.02 (0.01)
σ	0.11	0.07	0.07	0.08	0.06	0.10
JB	0.81	0.74	0.61	0.80	0.07	0.62
$\lvert\mu\rvert$	0.09	0.06	0.05	0.06	0.04	0.08
$\mu/\mu_{y_t^{t+k+1}}$	-0.02	0.01	0.00	0.03	0.00	0.04
$\sigma/\sigma_{y_t^{t+k+1}}$	0.28	0.18	0.17	0.18	0.13	0.22
B. INW revisions						
T	96	95	94	93	89	88
μ	-0.01 (0.01)	-0.01 (0.02)	-0.02 (0.02)	-0.01 (0.02)	0.00 (0.01)	0.01 (0.02)
σ	0.11	0.16	0.18	0.20	0.23	0.29
JB	0.81	0.80	0.83	0.17	0.82	0.52
$\lvert\mu\rvert$	0.09	0.13	0.14	0.15	0.18	0.22
$\mu/\mu_{y_t^{t+k+1}}$	-0.02	-0.04	-0.04	-0.02	-0.01	0.01
$\sigma/\sigma_{y_t^{t+k+1}}$	0.28	0.39	0.43	0.44	0.49	0.60
C. REM revisions						
T	88	88	88	88	88	88
μ	0.01 (0.02)	0.03 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.01)	0.02 (0.01)
σ	0.29	0.24	0.23	0.22	0.15	0.10
JB	0.52	0.50	0.72	0.37	0.46	0.62
$\lvert\mu\rvert$	0.22	0.19	0.19	0.18	0.11	0.08
$\mu/\mu_{y_t^{t+\ell}}$	0.01	0.07	0.05	0.06	0.06	0.04
$\sigma/\sigma_{y_t^{t+\ell}}$	0.60	0.51	0.49	0.47	0.31	0.22

Notes: JB denotes p-values for the Jarque-Bera test of normality. HAC standard errors in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10% levels, respectively, for the t^* statistics.

Table 2: Mean GDP revisions and the business cycle

	$k=1$	$k=2$	$k=3$	$k=4$	$k=8$	$k=\ell-1$
A. FXW revisions						
Business Cycle Asymmetry I						
μ_1	0.00 (0.01)	0.00 (0.01)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)
μ_2	-0.06*** (0.01)	-0.01 (0.01)	-0.05* (0.01)	-0.02 (0.02)	-0.04** (0.01)	0.04 (0.02)
$F^*(\mu_1=\mu_2)$	0.08	0.57	0.06	0.30	0.06	0.58
Business Cycle Asymmetry II						
μ_1	-0.01 (0.01)	0.01 (0.01)	0.01 (0.00)	0.02 (0.01)	0.00 (0.00)	0.01 (0.01)
μ_2	-0.02 (0.01)	0.00 (0.01)	-0.02 (0.01)	-0.01 (0.01)	-0.02 (0.01)	0.04 (0.02)
$F^*(\mu_1=\mu_2)$	0.69	0.57	0.23	0.03	0.12	0.29
Business Cycle Acceleration						
β_Φ	0.04 (0.03)	0.03 (0.01)	0.04** (0.01)	0.05* (0.01)	0.00 (0.01)	0.03 (0.02)
B. INW revisions						
Business Cycle Asymmetry I						
μ_1	0.00 (0.01)	0.01 (0.02)	0.02 (0.02)	0.03 (0.02)	0.04* (0.01)	0.05* (0.01)
μ_2	-0.06*** (0.01)	-0.15*** (0.01)	-0.21*** (0.03)	-0.24*** (0.01)	-0.25*** (0.03)	-0.28** (0.04)
$F^*(\mu_1=\mu_2)$	0.08	0.02	0.04	0.00	0.00	0.00
Business Cycle Asymmetry II						
μ_1	-0.01 (0.01)	0.00 (0.02)	0.01 (0.02)	0.03 (0.03)	0.03 (0.01)	0.05 (0.02)
μ_2	-0.02 (0.01)	-0.06 (0.02)	-0.10 (0.04)	-0.13 (0.04)	-0.16** (0.03)	-0.22*** (0.03)
$F^*(\mu_1=\mu_2)$	0.69	0.40	0.31	0.25	0.03	0.01
Business Cycle Acceleration						
β_Φ	0.04 (0.03)	0.09 (0.04)	0.14 (0.06)	0.19* (0.05)	0.16* (0.05)	0.25** (0.04)
C. REM revisions						
Business Cycle Asymmetry I						
μ_1	0.05* (0.01)	0.06** (0.01)	0.04 (0.02)	0.03 (0.02)	0.03 (0.01)	0.01 (0.00)
μ_2	-0.28** (0.04)	-0.17 (0.05)	-0.11 (0.05)	-0.03 (0.04)	0.00 (0.04)	0.04 (0.02)
$F^*(\mu_1=\mu_2)$	0.00	0.04	0.15	0.46	0.74	0.58
Business Cycle Asymmetry II						
μ_1	0.05 (0.02)	0.07* (0.02)	0.05 (0.02)	0.04 (0.02)	0.03* (0.01)	0.01 (0.01)
μ_2	-0.22*** (0.03)	-0.16** (0.03)	-0.13 (0.04)	-0.07 (0.03)	-0.02 (0.02)	0.04 (0.02)
$F^*(\mu_1=\mu_2)$	0.01	0.00	0.01	0.02	0.16	0.29
Business Cycle Acceleration						
β_Φ	0.25** (0.04)	0.21*** (0.02)	0.16** (0.02)	0.11 (0.03)	0.08** (0.01)	0.03 (0.02)

Notes: HAC standard errors in parentheses. F^* denotes p-values of the indicated null hypothesis. ***, ** and * denote statistical significance at the 1, 5 and 10% levels, respectively, for the t^* statistics.

Table 3: Volatility of GDP revisions and the business cycle

	$k=1$	$k=2$	$k=3$	$k=4$	$k=8$	$k=\ell-1$
A. FXW revisions						
Business Cycle Asymmetry I						
σ_1	0.11	0.08	0.07	0.08	0.06	0.10
σ_2	0.09	0.06	0.09	0.09	0.05	0.13
$F^*(\sigma_1=\sigma_2)$	0.13	0.10	0.32	0.31	0.62	0.14
Business Cycle Asymmetry II						
σ_1	0.12	0.08	0.07	0.08	0.06	0.10
σ_2	0.08	0.06	0.08	0.09	0.06	0.12
$F^*(\sigma_1=\sigma_2)$	0.00	0.02	0.70	0.35	0.68	0.20
Business Cycle Acceleration						
β_Φ	0.03***	0.01	0.00	-0.01	0.01	0.00
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
B. INW revisions						
Business Cycle Asymmetry I						
σ_1	0.11	0.15	0.17	0.18	0.21	0.27
σ_2	0.09	0.21	0.26	0.28	0.33	0.37
$F^*(\sigma_1=\sigma_2)$	0.13	0.07	0.08	0.01	0.09	0.14
Business Cycle Asymmetry II						
σ_1	0.12	0.16	0.17	0.19	0.22	0.27
σ_2	0.08	0.16	0.21	0.24	0.26	0.33
$F^*(\sigma_1=\sigma_2)$	0.00	0.91	0.58	0.23	0.59	0.28
Business Cycle Acceleration						
β_Φ	0.03***	0.03	0.3	0.04	0.03	0.05
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
C. REM revisions						
Business Cycle Asymmetry I						
σ_1	0.27	0.22	0.22	0.22	0.14	0.10
σ_2	0.37	0.32	0.29	0.22	0.18	0.13
$F^*(\sigma_1=\sigma_2)$	0.14	0.11	0.19	0.94	0.25	0.14
Business Cycle Asymmetry II						
σ_1	0.27	0.23	0.22	0.22	0.14	0.10
σ_2	0.33	0.28	0.27	0.21	0.17	0.12
$F^*(\sigma_1=\sigma_2)$	0.28	0.35	0.33	0.73	0.21	0.20
Business Cycle Acceleration						
β_Φ	0.05	0.04	0.03	0.05	-0.01	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)

Notes: HAC standard errors in parentheses. F^* denotes p-values of the indicated null hypothesis. ***, ** and * denote statistical significance at the 1, 5 and 10% levels, respectively, for the t^* statistics.

Table 4: GDP revisions and structural changes

	$k=1$	$k=2$	$k=3$	$k=4$	$k=8$	$k=\ell-1$
A. FXW revisions						
Structural change tests						
$AveF(\mu)$	0.23 [08Q1]	0.08 [01Q2]	0.52 [04Q2]	0.11 [06Q2]	0.64 [01Q1]	0.31 [07Q3]
$AveF(\sigma)$	0.00 [05Q2]	0.24 [00Q4]	0.02 [94Q4]	0.00 [06Q2]	0.43 [08Q2]	0.03 [99Q2]
Time-varying tests						
$L_C(\mu)$	0.24	0.09	0.42	0.12	0.67	0.36
$L_C(\sigma)$	0.00	0.21	0.04	0.17	0.17	0.25
L_C	0.01	0.12	0.06	0.10	0.22	0.34
B. INW revisions						
Structural change tests						
$AveF(\mu)$	0.23 [08Q1]	0.05 [00Q4]	0.12 [00Q4]	0.06 [04Q1]	0.43 [02Q4]	0.26 [07Q3]
$AveF(\sigma)$	0.00 [05Q2]	0.00 [01Q3]	0.01 [95Q3]	0.04 [03Q4]	0.00 [02Q2]	0.00 [03Q4]
Time-varying tests						
$L_C(\mu)$	0.24	0.06	0.12	0.06	0.32	0.24
$L_C(\sigma)$	0.00	0.01	0.01	0.12	0.02	0.02
L_C	0.01	0.00	0.00	0.07	0.04	0.04
C. REM revisions						
Structural change tests						
$AveF(\mu)$	0.26 [07Q3]	0.17 [97Q1]	0.07 [97Q4]	0.09 [97Q4]	0.30 [99Q2]	0.31 [07Q3]
$AveF(\sigma)$	0.00 [03Q4]	0.00 [04Q4]	0.00 [97Q3]	0.00 [97Q3]	0.01 [99Q2]	0.03 [99Q2]
Time-varying tests						
$L_C(\mu)$	0.24	0.16	0.06	0.09	0.31	0.36
$L_C(\sigma)$	0.02	0.14	0.03	0.03	0.32	0.25
L_C	0.04	0.12	0.02	0.02	0.38	0.34

Notes: $AveF$ denotes approximate asymptotic p-values (Hansen, 1997) for the Andrews (1993) average F test of the null hypothesis of one unknown breakpoint (estimated break dates are in brackets). L_C denotes p-values for the Nyblom (1989) and Hansen (1992) test of individual (mean and variance) and joint parameters instability.

Table 5: Forecast rationality: unbiasedness

	$k=1$	$k=2$	$k=3$	$k=4$	$k=8$	$k=\ell-1$
			Unbiasedness Regressions			
T	88	88	88	88	88	88
α	0.00 (0.03)	0.00 (0.02)	0.02 (0.02)	0.04 (0.02)	0.02 (0.01)	0.01 (0.01)
β	0.02 (0.04)	0.08 (0.03)	0.01 (0.03)	-0.03 (0.03)	0.02 (0.01)	0.02 (0.02)
$F^*(\alpha, \beta = 0)$	0.94	0.29	0.91	0.67	0.40	0.31
			Business Cycle Asymmetry I			
$F^*(s)$	0.00	0.00	0.17	0.52	0.91	0.34
$F^*(\alpha, \beta = 0)$	0.00	0.01	0.13	0.35	0.73	0.51
$F^*(\alpha_1, \beta_1 = 0)$	0.25	0.12	0.04	0.11	0.43	0.26
$F^*(\alpha_2, \beta_2 = 0)$	0.01	0.01	0.31	0.85	0.96	0.80
			Business Cycle Asymmetry II			
$F^*(s)$	0.46	0.16	0.09	0.09	0.51	0.06
$F^*(\alpha, \beta = 0)$	0.64	0.09	0.05	0.15	0.35	0.30
$F^*(\alpha_1, \beta_1 = 0)$	0.88	0.40	0.38	0.46	0.39	0.30
$F^*(\alpha_2, \beta_2 = 0)$	0.51	0.05	0.41	0.03	0.50	0.53
			Structural change tests			
$AveF$	0.58 [07Q3]	0.36 [95Q1]	0.17 [97Q4]	0.11 [97Q4]	0.55 [99Q2]	0.20 [99Q2]
			Time-varying tests			
$L_C(\sigma)$	0.02	0.12	0.03	0.03	0.30	0.33
L_C	0.06	0.19	0.05	0.05	0.51	0.26

Notes: HAC standard errors in parentheses. F^* denotes p-values for the indicated null hypothesis. ***, ** and * denote statistical significance at the 1, 5 and 10% levels, respectively, for the t^* statistic. $AveF$ denotes heteroskedasticity-corrected bootstrap p-values (Hansen, 2000) for the Andrews (1993) average F test of the null hypothesis of one unknown breakpoint (estimated break dates are in brackets). L_C denotes p-values for the Nyblom (1989) and Hansen (1992) test of individual (variance) and joint parameters instability.

Table 6: Forecast rationality: efficiency - $\mathbf{X}_{t+k} = \left({}_{k}y_t^1, {}_{\ell}y_{t-1}^k\right)'$

	$k=1$	$k=2$	$k=3$	$k=4$	$k=8$	$k=\ell-1$
			Efficiency Regressions			
T	87	87	87	87	87	87
α	0.02 (0.02)	0.02 (0.02)	0.03 (0.02)	0.05 (0.02)	0.03 (0.01)	-0.01 (0.01)
β	-0.02 (0.04)	0.04 (0.03)	-0.01 (0.04)	-0.04 (0.05)	-0.01 (0.01)	0.04 (0.01)
$\gamma_{{}_{\ell}y_{t-1}^k}$	0.16 (0.10)	0.13 (0.07)	0.11 (0.08)	0.03 (0.09)	0.10 (0.04)	0.19 (0.07)
$\gamma_{{}_{k}y_t^1}$		0.02 (0.03)	0.01 (0.01)	0.01 (0.01)	0.02* (0.00)	-0.02 (0.01)
$F^*(\alpha, \beta = 0)$	0.91	0.57	0.79	0.53	0.49	0.34
$F^*(\boldsymbol{\gamma} = 0)$	0.43	0.36	0.52	0.67	0.20	0.43
$F^*(\alpha, \beta, \boldsymbol{\gamma} = 0)$	0.79	0.07	0.89	0.79	0.51	0.64
			Business Cycle Asymmetry I			
$F^*(s)$	0.00	0.00	0.00	0.00	0.04	0.29
$F^*(\alpha, \beta, \boldsymbol{\gamma} = 0)$	0.00	0.00	0.00	0.00	0.03	0.31
$F^*(\alpha_1, \beta_1, \boldsymbol{\gamma}_1 = 0)$	0.44	0.16	0.12	0.15	0.15	0.19
$F^*(\alpha_2, \beta_2, \boldsymbol{\gamma}_2 = 0)$	0.00	0.00	0.00	0.00	0.00	0.54
			Business Cycle Asymmetry II			
$F^*(s)$	0.55	0.31	0.24	0.16	0.66	0.02
$F^*(\alpha, \beta, \boldsymbol{\gamma} = 0)$	0.18	0.04	0.12	0.58	0.56	0.13
$F^*(\alpha_1, \beta_1, \boldsymbol{\gamma}_1 = 0)$	0.06	0.22	0.44	0.68	0.50	0.22
$F^*(\alpha_2, \beta_2, \boldsymbol{\gamma}_2 = 0)$	0.62	0.00	0.28	0.12	0.27	0.08
			Structural change tests			
$AveF$	0.16 [07Q4]	0.10 [02Q3]	0.13 [97Q3]	0.01 [97Q3]	0.94 [02Q3]	0.62 [97Q3]
			Time-varying tests			
$L_C(\sigma)$	0.01	0.10	0.03	0.04	0.27	0.49
L_C	0.04	0.11	0.06	0.11	0.85	0.62

Notes: See Table 5.

Table 7: Forecast rationality: efficiency - $\mathbf{X}_{t+k} = \mathbf{x}_{t+k}$

	$k=1$	$k=2$	$k=3$	$k=4$	$k=8$	$k=\ell-1$
			Efficiency Regressions			
T	87	87	87	87	87	87
α	0.06	0.05	0.06	0.07 *	0.04 *	0.00
	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
β	- 0.14	- 0.05	- 0.10	- 0.13	- 0.04	0.02
	(0.04)	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)
δ^{*}	0.04	0.06	0.08	0.05	0.02	0.04
$F^{*}(\alpha, \beta = 0)$	0.24	0.39	0.35	0.25	0.16	0.32
$F^{*}(\boldsymbol{\gamma} = 0)$	0.00	0.00	0.01	0.02	0.01	0.90
$F^{*}(\alpha, \beta, \boldsymbol{\gamma} = 0)$	0.00	0.00	0.00	0.01	0.01	0.46
			Business Cycle Asymmetry I			
$F^{*}(s)$	0.00	0.02	0.00	0.00	0.00	0.04
$F^{*}(\alpha, \beta, \boldsymbol{\gamma} = 0)$	0.00	0.00	0.00	0.00	0.00	0.00
$F^{*}(\alpha_1, \beta_1, \boldsymbol{\gamma}_1 = 0)$	0.43	0.17	0.36	0.15	0.04	0.37
$F^{*}(\alpha_2, \beta_2, \boldsymbol{\gamma}_2 = 0)$	0.00	0.00	0.00	0.00	0.00	0.00
			Business Cycle Asymmetry II			
$F^{*}(s)$	0.15	0.25	0.04	0.00	0.04	0.33
$F^{*}(\alpha, \beta, \boldsymbol{\gamma} = 0)$	0.00	0.04	0.00	0.00	0.00	0.14
$F^{*}(\alpha_1, \beta_1, \boldsymbol{\gamma}_1 = 0)$	0.00	0.01	0.11	0.09	0.03	0.26
$F^{*}(\alpha_2, \beta_2, \boldsymbol{\gamma}_2 = 0)$	0.00	0.08	0.00	0.00	0.00	0.02
			Structural change tests			
$AveF$	0.21	0.17	0.25	0.11	0.16	0.05
	[95Q1]	[95Q2]	[97Q1]	[97Q1]	[02Q2]	[99Q1]
			Time-varying tests			
$L_C(\sigma)$	0.03	0.13	0.03	0.05	0.33	0.23
L_C	0.26	0.42	0.33	0.23	0.20	0.19

Notes: See Table 5.

Table 8: Corradi et al. (2009) forecast rationality tests

	$k=1$	$k=2$	$k=3$	$k=4$	$k=8$	$k=\ell-1$
			$\mathbf{X}_{t+k} = \left({}_{k}y_t^1, {}_{\ell}y_{t-1}^k\right)'$			
M_T^*	0.17	0.13	0.27	0.15	0.46	0.14
			$\mathbf{X}_{t+k} = \mathbf{x}_{t+k}$			
M_T^*	0.08	0.09	0.08	0.12	0.06	0.10

Notes: M_T^* denotes bootstrap p-values obtained after 999 replications.

Table 9: Forecast results

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 8$	$k = \ell - 1$
			A. Symmetric model			
RMSFE	1.02	0.91	0.96	1.00	1.07	1.02
PT	0.08	0.02	0.05	0.81	0.89	0.77
ΔMPLS	0.01	**0.11**	0.11	**0.13**	-0.03	0.00
DH	0.05	0.42	0.06	0.28	0.78	0.75
Bk	0.03	0.18	0.63	0.33	0.24	0.17
			B. Business Cycle Asymmetric model			
RMSFE	**0.78**	1.00	**0.86**	**0.88**	1.07	1.03
PT	0.00	0.01	0.05	0.01	0.82	0.99
ΔMPLS	**0.23**	0.03	**0.16**	**0.26**	-0.03	-0.01
DH	0.76	0.19	0.34	0.87	0.92	0.75
Bk	0.07	0.20	0.49	0.35	0.33	0.07
			C. Business Cycle Asymmetric model - expansion regime			
RMSFE	**0.80**	0.89	0.81	**0.84**	1.04	1.00
PT	0.01	0.03	0.14	0.00	0.99	0.99
ΔMPLS	**0.18**	0.03	**0.12**	**0.29**	-0.02	0.02
DH	0.60	0.66	0.22	0.91	0.97	0.66
Bk	0.07	0.05	0.17	0.29	0.40	0.57

Notes: RMSFE denotes relative root mean squared forecast error (<1 means outperformance of the benchmark model). ΔMPLS denotes relative mean log predictive score (>0 means outperformance of the benchmark model). Bold values denote rejection of the null hypothesis of equal predictive accuracy at 10% level according to the one-sided t-statistic version of the DMW test. PT, DH, and Bk denote p-values for the Pesaran and Timmermann (1992, 2009), Doornik and Hansen (2008), and Berkowitz (2001) tests, respectively.