

Discussion Paper No. 07-058

**The Accuracy and Efficiency of the
Consensus Forecasts:
A Further Application and Extension
of the Pooled Approach**

Philipp Ager, Marcus Kappler, and Steffen Osterloh

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 07-058

**The Accuracy and Efficiency of the
Consensus Forecasts:
A Further Application and Extension
of the Pooled Approach**

Philipp Ager, Marcus Kappler, and Steffen Osterloh

Download this ZEW Discussion Paper from our ftp server:

<ftp://ftp.zew.de/pub/zew-docs/dp/dp07058.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von
neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung
der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other
economists in order to encourage discussion and suggestions for revisions. The authors are solely
responsible for the contents which do not necessarily represent the opinion of the ZEW.

Non-technical summary

The last decade has seen marked economic fluctuations in the major industrial countries, which regularly present business cycle forecasters with a challenge. In this paper we are interested in how professional forecasters managed to predict GDP and price developments during the last decade. To this end, we explore the accuracy and evolution of the Consensus Forecast for twelve industrial countries for the years 1996 to 2006. This pooled forecast has the main advantage that it offers monthly publications of revised forecasts for the current and the next year, so that an explicit revision process of 24 forecasts for every target year can be observed.

The theoretical and econometric analysis is based on the framework of Davies and Lahiri (1995) and Clements et al. (2006). The latter employ a pooling procedure which permits the evaluation of all forecasts for each target variable over 24 horizons simultaneously. Adopting this methodology allows us to draw conclusions on evaluating systematic forecast bias and, by applying a test on the predictability of forecast revisions, on the efficient use of new information for all forecast horizons jointly. It is shown how the pooled approach needs to be adjusted in order to accommodate the forecasting scheme of the Consensus Forecasts. Furthermore, the pooled approach is extended by a sequential test with the purpose of detecting the critical horizon beyond which the forecast should be regarded as biased. Moreover, we show how the pooled tests for the predictability of forecast revisions can be improved by taking heteroscedasticity in the form of target year-specific variances of macroeconomic shocks into account.

In the empirical part we first present results in the form of analytical confidence intervals surrounding the horizon-specific bias estimates which allow intuitive and meaningful interpretations. The test for common bias reveals that several countries show biased forecasts, especially with forecasts covering more than 12 months. These results partially confirm the presumption that the macroeconomic forecasts for the past 10 years were severely affected by the pronounced shocks in that period. The fact that for individual countries systematic biases can be observed by applying the Consensus Forecasts reveals that in these countries the forecasting industry on the whole was not able to cope with the shocks specific to the past ten years. Revisions of past months for GDP growth forecast have significant explanatory power for current revisions for almost all countries. For inflation revisions we find significant past revisions for some countries. Overall, our results imply that a lack of information efficiency is more severe for GDP forecasts than for inflation forecasts.

The accuracy and efficiency of the Consensus Forecasts: A further application and extension of the pooled approach

October 2007

Philipp Ager

GSE, Universitat Pompeu Fabra, Barcelona

Marcus Kappler¹

Centre for European Economic Research (ZEW), Mannheim

Steffen Osterloh

Centre for European Economic Research (ZEW), Mannheim

Abstract: In this paper we analyze the macroeconomic forecasts of the Consensus Forecasts for 12 countries over the period from 1996 to 2006 regarding bias and information efficiency. A pooled approach is employed which permits the evaluation of all forecasts for each target variable over 24 horizons simultaneously. It is shown how the pooled approach needs to be adjusted in order to accommodate the forecasting scheme of the Consensus Forecasts. Furthermore, the pooled approach is extended by a sequential test with the purpose of detecting the critical horizon after which the forecast should be regarded as biased. Moreover, heteroscedasticity in the form of year-specific variances of macroeconomic shocks is taken into account. The results show that in the analyzed period which was characterized by pronounced macroeconomic shocks, several countries show biased forecasts, especially with forecasts covering more than 12 months. In addition, information efficiency has to be rejected in almost all cases.

Keywords: business cycle forecasting, forecast evaluation, Consensus Forecasts

JEL-Classification: C 52, E 32, E 37

¹ Corresponding author. P.O. Box 103443, D-68034, Mannheim, Germany, E-Mail: kappler@zew.de. For helpful feedback, we thank François Laisney and the audience at the International Workshop in Economics and Finance (IWEF 2007) at Tripolis, Greece, 16 June 2007. All remaining errors are the responsibility of the authors.

1 Introduction

The last decade has seen marked economic fluctuations in the major industrial countries, starting from the new economy boom beginning in the mid-nineties, followed by a lengthy downturn due to the burst of the dot-com bubble and accelerated by a number of events like 9/11 and concluding with the recent economic recovery. Such economic fluctuations regularly present business cycle forecasters with a challenge. In this paper we are interested in how professional forecasters managed to predict GDP and price developments during the last decade. To this end, we explore the biasedness and efficiency of the Consensus Forecast for twelve industrial countries for the years 1996 to 2006.

The Consensus Forecast is a pooled forecast based on a monthly survey among professional forecasters from the respective country. Due to its aggregation it is often regarded as the market's opinion on the future development of the economy. One main advantage the Consensus Forecasts offer to our analysis is that, because of the monthly publication of revised forecasts for the current and the next year, an explicit revision process of 24 forecasts for every target year can be observed. Moreover, due to international standardization regarding methodology and date of publication, these forecasts are highly relevant for international comparisons.

This kind of forecasts has received increasing attention in the forecast evaluation literature in the past years, starting with Loungani (2001) who compares the Consensus Forecasts for industrialized and developing countries in the years 1989-1998 regarding accuracy, biasedness and efficiency. In a recent work, Batchelor (2007) addresses similar questions by analyzing the bias of the Consensus Forecasts in the years 1990-2005. He finds evidence for overoptimistic GDP forecasts in Japan, Germany, France and Italy, and no evidence for a bias of the inflation forecasts. The biasedness of individual forecasters is a very popular object of investigation (see, for instance, Batchelor, 2001, for OECD and IMF forecasts; a survey on national studies can be found in Fildes & Stekler, 2002). The frequently found characteristic of systematically biased forecasts is usually attributed to error sources on the individual level such as model misspecifications, herding or political biases (see Stekler, 2007, for an overview). However, the studies applying the Consensus Forecasts use a pooled forecast, where these individual biases are typically expected to cancel out. Therefore, the biases found by Batchelor (2007), which are unequal in size between the countries, do not hint to irrationality on the individual level, but can be attributed to a common wrong dealing with

country-specific external influences such as declining growth trends, which were shared by all forecasters in a country.

As the Consensus Forecasts constitute a fixed-event forecast, with every target year forecasted separately from 24 horizons before, it complicates the empirical analysis. Batchelor (2007) applies a conventional approach, by testing single forecast horizons individually for biasedness. This may be accomplished by the Mincer-Zarnowitz test (Mincer & Zarnowitz, 1969), or by a more general t-test as introduced by Clements (2005, p.6). However, this kind of test has one decisive weak point for the high-frequency fixed-event forecasts of the Consensus Forecasts: Either all 24 horizons are tested individually, and the comparative results lose in explanatory power due to the complexity of 24 tests with potentially different results. Or, as done in Batchelor (2007), only selected horizons out of the 24 available are tested and compared, but then a loss of information occurs. To cope with this problem, in this paper a cutting edge pooling method proposed by Clements et al. (2007) and Davis & Lahiri (1995) is applied.

The second question of this paper, the efficiency of the forecasts, has received more attention in the literature regarding the application to fixed-event forecasts in the form of the Nordhaus test (Nordhaus, 1987) on the unpredictability of forecast revisions. Further applications of this methodology on the pooling of fixed-event forecasts over different target years can be found in Clements (1997) and Harvey et al. (2001). Isiklar et al. (2006) are the first who analyze Consensus Forecasts for 18 industrialized countries from 1989 to 2004 and who consider a pooling over several countries.

Again, a pooled approach based on the works of Clements et al. (2007) and Davis & Lahiri (1995) is chosen. However, this procedure demands several adjustments due to the characteristics of the data set: During the analyzed period, years with relatively few unforeseen events were followed by several other years which were dominated by macroeconomic shocks like the stock market crash, increasing oil prices due to the Iraq war and terrorist attacks, taking a combined as well as consecutive influence on the worldwide economic performance. Given such events, the assumption of homoscedastic macroeconomic shocks underlying the model does not seem reasonable. We show how the pooled tests for the predictability of forecast revisions can be improved by considering year-specific variances in the econometric estimations.

The paper proceeds as follows. Section 2 outlines the error model used for the forecast analyses and sets out the pooling approach for testing unbiasedness, predictability of forecast revisions and weak efficiency. Section 3 presents the empirical results for Belgium, Canada,

France, Germany, Italy, Japan, the Netherlands, Spain, Sweden, Switzerland, the United Kingdom and the US. The last section summarizes and concludes.

2 Pooling procedure and forecast analysis

To test for bias of the Consensus Forecasts and to analyse the forecast revisions, we directly build on a recently published approach by Clements et al. (2007). The authors propose to pool the forecasts for each variable across horizons. This approach enables one to employ more powerful econometric tests than in the traditional procedure which looks at forecasts separately for each horizon. In particular, the pooling procedure is ideally suited for our data set which comprises forecast over 24 horizons for each year, but covers only 11 target years. Conducting the conventional procedures to test for biasedness and efficiency is not sensible with very few observations available for each horizon.

In this section we briefly review the pooling approach and extend it to take some form of heteroscedasticity of macroeconomic shocks into account. In particular, we show that incorporating target year-specific shock variances improves the econometric tests to analyse forecast revisions. As we show in the empirical section, this source of variability is likely to be important for monthly errors of GDP growth and inflation forecasts. However, considering target year-specific variances in the pooled test of forecast unbiasedness does not advance the original test which assumes homoscedastic shocks.

Furthermore, we demonstrate how the pooling approach can be used to conduct a sequential test to detect forecast biasedness for increasing horizons and we offer some intuitive illustrations of the tests for separate bias for each horizon. By applying a battery of different tests which are all derived from a coherent error model framework, we are then able to present a comprehensive picture of the performance of the recent Consensus Forecasts.

2.1 Uniform Bias for every forecast horizon

The core of the analytical framework is the forecast error model of Clements et al. (2007). According to this model, the error for a forecast with horizon h of a target variable in year t is given by

$$(1) \quad e_{th} = A_t - F_{th} = \alpha + v_{th}$$

where A_t is the actual value of the variable in period t , and F_{th} is the forecast made for the target variable in period t at time $t-h$ (i.e. $h = 1, \dots, H$ forecasts are made for variable A_t). The error term e_{th} is the sum of a (possible) bias α and the aggregate or common macroeconomic shocks v_{th} . The v_{th} are assumed to be common to all individuals and are the sum of all shocks u_{ij} that occurred between $t-h$ and t , labelled as u_{th} to u_{t1} :

$$(2) \quad v_{th} = \sum_{j=1}^h u_{tj}.$$

Note that in contrast to Clements et al. (2007), we do not consider including “idiosyncratic shocks” in the error model (1) from the beginning. Since we apply this error framework to analyse the aggregate Consensus Forecasts, it is plausible to assume that the idiosyncratic shocks cancel out due to the aggregation of the individual forecasts. In addition, omitting an idiosyncratic error component simplifies the subsequent derivation of test statistics considerably.

For the pooled estimation, the forecast errors are stacked into a vector with dimension TH that can be written as $\mathbf{e} = \mathbf{A} - \mathbf{F}$, where the vectors \mathbf{F} and \mathbf{A} are defined as $\mathbf{F}' = (F_{1H}, \dots, F_{11}, F_{2H}, \dots, F_{21}, \dots, F_{TH}, \dots, F_{T1})$ and $\mathbf{A} = \mathbf{A}^* \otimes \mathbf{i}_H$ with $\mathbf{A}^* = (A_1, A_2, \dots, A_T)'$, respectively. In matrix notation equation (1) takes the form:

$$(3) \quad \mathbf{e} = \mathbf{i}_{TH} \alpha + \mathbf{v}$$

where \mathbf{v} is the stacked vector of the error terms and \mathbf{i}_{TH} is a vector of ones with dimension $TH \times 1$. The null hypothesis of an unbiased forecast is $H_0 : \alpha = 0$. The test of bias is based on the OLS estimator of α in (3), which simply computes the mean forecast error over all horizons and periods

$$(4) \quad \hat{\alpha} = \frac{1}{TH} \sum_{t=1}^T \sum_{h=1}^H e_{th}$$

with the consistent variance estimator of $\hat{\alpha}$

$$(5) \quad \text{Var}(\hat{\alpha}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{(TH)^2} \mathbf{i}'_{TH} \mathbf{\Sigma} \mathbf{i}_{TH}$$

where $\mathbf{X} = \mathbf{i}_{TH}$ and $\mathbf{\Sigma}$ is the covariance matrix of the error terms \mathbf{v} under H_0 . We can assume that $E(\mathbf{v}) = \mathbf{0}$ for the first moment and $E(\mathbf{v}\mathbf{v}') = \mathbf{\Sigma}$ for the second moments. The covariance matrix comprises not only elements on the diagonal but also takes off-diagonal elements into account that represent the correlation structure for overlapping common macroeconomic shocks. The precise form of $\mathbf{\Sigma}$ depends on the forecast structure and forecast horizon of the variable of interest. Using the assumption about the error terms made as given by equation (2), it is convenient that the covariance matrix depends only on one single parameter which is $E(u_t^2) = \sigma^2$ assuming homoscedastic aggregate shocks. In addition, the variance of the aggregate shocks is directly proportional to h . Therefore, the covariance matrix $\mathbf{\Sigma}$ can be written as follows:

$$(6) \quad \mathbf{\Sigma} = \sigma^2 \mathbf{\Psi}$$

where $\mathbf{\Psi}$ is a $(TH \times TH)$ matrix that contains the specific correlation structure for the forecast variable \mathbf{F} . A detailed description of the relevant covariance matrices in our framework is discussed in section 3. Because $\mathbf{\Sigma}$ is unobservable it has to be replaced by the empirical counterpart $\hat{\mathbf{\Sigma}}$ to obtain feasible and consistent standard errors for OLS. Consequently, $\hat{\mathbf{\Sigma}}$ is computed as follows. In the first step the estimated $\hat{\mathbf{v}}$ is obtained by subtracting the estimated bias from the residuals.

$$(7) \quad \hat{\mathbf{v}} = \mathbf{e} - \hat{\alpha} \mathbf{i}_{TH}$$

As in Davies & Lahiri (1995) and Clements et al. (2007) we make use of the relationship $E(v_{it}^2) = h\sigma_u^2$ and thereby obtain in the second step the estimated variance $\hat{\sigma}^2$ by the following OLS regression:

$$(8) \quad \hat{\mathbf{v}} \odot \hat{\mathbf{v}} = \varphi \boldsymbol{\tau} + \boldsymbol{\omega}$$

The operator \odot refers to the element by element multiplication (Hadamard product). Furthermore, $\boldsymbol{\tau} = \mathbf{i}_T \otimes \boldsymbol{\tau}_H$ is a $(TH \times 1)$ horizon index, where $\boldsymbol{\tau}'_H = (H, H-1, \dots, 1)$ and OLS provides $\hat{\phi}$ which is an estimate of σ^2 , the variance of the common macro shocks. With this estimate of σ^2 at hand, the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ can be readily computed by following equation (6). Replacing $\boldsymbol{\Sigma}$ by $\hat{\boldsymbol{\Sigma}}$ in equation (5) we can test the unbiasedness of the forecasts by the usual t-statistic.

2.2 Separate bias for every forecast horizon and a sequential test

In section 2.1 we considered a common bias over all forecast horizons. Henceforth, we relax this restriction and allow for a separate bias for each horizon by using a parameter vector $\boldsymbol{\alpha}_H$ with dimension $(H \times 1)$. Horizon-specific alphas enable us to identify the evolution of systematic bias of forecasts over horizons. A graphical presentation of bias dispersion across forecast horizons, which provides an intuitive assessment of the forecast performance, will be given in the empirical section. A further advantage of separate alphas is that a variety of testing procedures can be employed for detecting systematic bias of forecast horizons. For instance, one can test not only if there is a bias in single forecast horizons, but also if there is a bias in several or all forecast horizons by imposing joint restrictions.

By allowing for separate biases, equation (3) takes the following form:

$$(9) \quad \mathbf{e} = (\mathbf{i}_T \otimes \mathbf{I}_H) \boldsymbol{\alpha}_H + \mathbf{v}$$

An estimate of the $(H \times 1)$ parameter vector $\boldsymbol{\alpha}_H$, $\hat{\boldsymbol{\alpha}}_H$, is obtained by running an OLS regression, where $(\mathbf{i}_T \otimes \mathbf{I}_H)$ is the matrix of the explanatory variables with dimension $(TH \times H)$. Moreover, the variance of $\hat{\boldsymbol{\alpha}}_H$, $\text{Var}(\hat{\boldsymbol{\alpha}}_H)$, is computed analogously to equation (5) by replacing $\hat{\boldsymbol{\alpha}}$ with $\hat{\boldsymbol{\alpha}}_H$, setting $\mathbf{X} = \mathbf{i}_T \otimes \mathbf{I}_H$ and using the fact that $(\mathbf{i}_T \otimes \mathbf{I}_H) \hat{\boldsymbol{\alpha}}_H$ is a $(TH \times 1)$ parameter vector.

2.2.1 A sequential test for forecast unbiasedness

A test statistic for the null hypothesis that there is no bias in the forecast up to horizon h can be obtained by computing the Wald statistic which uses only a subset $\hat{\boldsymbol{\alpha}}_h$ of $\hat{\boldsymbol{\alpha}}_H$ and places

zero restrictions on this vector of bias estimates. This allows one to conduct a sequential test in order to detect the critical horizon after which the forecast should be regarded as biased. As Clements et al. (2007) note, different biases for the individual horizons may result if there are systematic differences in the ways in which the forecasts of different horizons are computed. Therefore, a sequential test can help to assess up to which forecast horizon the model of a particular forecaster breaks down.

To formalize the idea, consider testing of the null hypothesis of zero bias up to horizon h , $H_0 : \hat{\alpha}_1 = \hat{\alpha}_2 = \dots = \hat{\alpha}_h = 0$, by computing the Wald statistic

$$(10) \quad W = (\hat{\mathbf{a}}_h)' [Var(\hat{\mathbf{a}}_h)]^{-1} (\hat{\mathbf{a}}_h)$$

in which the $(h \times 1)$ vector $\hat{\mathbf{a}}_h = (\alpha_1, \alpha_2, \dots, \alpha_h)'$ contains the bias estimates in ascending order and $Var(\hat{\mathbf{a}}_h)$ is the $(h \times h)$ symmetric upper left block of the full covariance matrix $Var(\hat{\mathbf{a}}_H)$. This Wald statistic is χ^2 -distributed with h degrees of freedom. The corresponding F statistic can be obtained by dividing W by h , which has better properties in small samples.² In the empirical part we present the results of such a sequential test in more detail.

2.3 Are forecasts revisions predictable? – A test for efficiency

If forecasts are efficient in the sense that they employ all relevant information about the future development of the target variable, then the forecast revision process should behave like a random walk (Nordhaus, 1987). Define the forecast revisions between $h+1$ and h as $r_{t;h,h+1} \equiv F_{t,h} - F_{t,h+1}$. The forecast revision $r_{t;h,h+1}$ should be uncorrelated with the information known when $F_{t,h+1}$ was made. In particular, it should be uncorrelated with past revisions. Therefore, a standard procedure is to test forecast efficiency by testing whether the past revisions $\{r_{t;h+1,h+2}, r_{t;h+2,h+3}, \dots\}$ for target A_t help to predict the current revision $r_{t;h,h+1}$. In the literature this test is typically referred to as a test of weak efficiency.

As mentioned before, a substantial contribution to the revision analysis that will be presented in the next section is that we allow for target year-specific shock variances to accommodate an important source of data variability.

2.3.1 Test on past revisions controlling for heteroscedastic common shocks

From equation (1) it readily follows that the forecast revision between $h+1$ and h is³

$$(11) \quad r_{t,h,h+1} = u_{t,h+1} + \alpha_{h+1} - \alpha_h.$$

Once again, instead of looking at forecast revisions separately for each horizon, these are stacked as in Clements et al. (2007) into a $((H-2)T \times 1)$ vector $\mathbf{r} = [\mathbf{r}_{H-2,H-1}, \dots, \mathbf{r}_{1,2}]'$. The vector \mathbf{r} has length $(H-2)$ since two forecast horizons get “lost” due to computing revisions and using previous month’s revisions, i.e. lagged revisions, in the test regression. The revisions for the individual horizon pairs (according to the notational convention adopted above) are stacked over time: $\mathbf{r}_{H-2,H-1} = (r_{t,H-2,H-1} \dots r_{T,H-2,H-1})'$, \dots , $\mathbf{r}_{1,2} = (r_{t,1,2} \dots r_{T,1,2})'$. In the next step a pooled regression is run:

$$(12) \quad \mathbf{r} = \gamma \mathbf{r}_{-1} + \boldsymbol{\delta}_h + \boldsymbol{\omega}$$

where $\mathbf{r}_{-1} = (r_{t,H-1,H} \dots r_{T,H-1,H}, \dots, r_{t,2,3} \dots r_{T,2,3})'$ and both \mathbf{r} and \mathbf{r}_{-1} are $T(H-2) \times 1$ vectors. The matrix $\boldsymbol{\delta}_h = (\mathbf{I}_{H-2} \otimes \mathbf{i}_T) \boldsymbol{\delta}$ with $\boldsymbol{\delta} = (\delta_{H-2}, \dots, \delta_1)'$ is included to control for horizon-specific bias. Furthermore, we can assume the following property $E(\boldsymbol{\omega}) = 0$ for the first moment and $E(\boldsymbol{\omega}\boldsymbol{\omega}') = \boldsymbol{\Omega}$ for the second moments. If $\hat{\gamma}$ is not significantly different from zero, current forecast revisions are generally uncorrelated with the past information set (past revisions). For the moment we assume that the common errors u_{ij} are homoscedastic. Thus, the covariance matrix of $\boldsymbol{\omega}$ takes the following form under the null:

$$(13) \quad \hat{\boldsymbol{\Omega}} = \hat{\sigma}^2 \boldsymbol{\Lambda}$$

where $\hat{\boldsymbol{\Omega}}$ and $\boldsymbol{\Lambda}$ are $((H-2)T \times (H-2)T)$ matrices. The covariance matrix $\hat{\boldsymbol{\Omega}}$ comprises not only elements on the diagonal but also takes off-diagonal elements into account that represent an additional correlation structure in the forecast revision context. The precise form of $\boldsymbol{\Omega}$ depends on $\boldsymbol{\Lambda}$, which contains the specific correlation structure for the forecast revision \mathbf{r} . A

² Cf. Greene (2003), chapter 6, for an illustration of the relation between the Wald test and the F statistic.

³ We slightly change the notion of equation (1) and allow for separate horizon specific bias.

detailed description of $\mathbf{\Lambda}$ in our framework is discussed in section 3. Moreover, the estimated variance $\hat{\sigma}^2$ can be obtained from the OLS residuals of the regression of \mathbf{r} on $\mathbf{\delta}_h$.

However, assuming homoscedastic shocks may not describe the variability in the data reasonably well and we therefore relax this assumption in the following. The Consensus data set of monthly GDP forecasts enables us to cope with the problem of heteroscedasticity by considering target year-specific error variances. Particularly in periods where several macroeconomic shocks such as the stock market crash, increasing oil prices due to the Iraq war and terrorist attacks took a combined as well as consecutive influence on the worldwide economic performance, there are no reasons to maintain the assumption of homoscedastic macroeconomic shocks. Moreover, we find evidence for heteroscedastic shocks in the last decade by inspecting the squared residuals, $\hat{\mathbf{v}}^2$, of twelve countries over time (see section 3).

Estimating target year-specific variances is straightforward. In a first step, the vector of stacked revisions \mathbf{r} is regressed on the matrix of horizon-specific $\mathbf{\delta}_h$ dummy variables. The residuals of this regression provide estimates of bias-removed revisions $\hat{r}_{t,h,h+1} = r_{t,h,h+1} - \hat{\delta}_h$ with $\hat{\delta}_h$ as an estimate for $\alpha_{h+1} - \alpha_h$. Assuming target year-specific variances and using $E(u_{t,h}) = \sigma_t^2$, $t=1, \dots, T$, consistent estimates of the T variances can be obtained by regressing $\hat{\mathbf{r}} \odot \hat{\mathbf{r}}$ on a matrix of year dummies, whereas the vector $\hat{\mathbf{r}}$ is constructed analogous to \mathbf{r} . Denote the estimated variances as $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_T^2$. The following formula will insert these variances into the proper positions of the covariance matrix of the revision errors:

$$(14) \quad \hat{\mathbf{\Omega}} = \mathbf{I}_{H-2} \otimes \text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_T^2) + \mathbf{C} \otimes \text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_{T-1}^2; -1)$$

in which $\text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_{T-1}^2; -1)$ is a $(T \times T)$ matrix that contains the variance terms $\hat{\sigma}_1^2, \dots, \hat{\sigma}_{T-1}^2$ on the first diagonal below the main diagonal. \mathbf{I}_{H-2} is a $(H-2) \times (H-2)$ identity matrix and \mathbf{C} a matrix of the same dimension that takes the special correlation structure of $\mathbf{\Lambda}$ into account (see section 3 for details):

$$\mathbf{C} = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & 0 & 1 & \ddots & \vdots \\ 0 & & & & \ddots & \ddots & 0 \\ 1 & 0 & & & & & 1 \\ 0 & 1 & \ddots & & & & 0 \\ \vdots & \ddots & \ddots & & & & \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix}$$

in which the ones are put on the $(H-2)/2$ th diagonal above and below the main diagonal. The test $\gamma = 0$ in the regression (12) can now be conducted by employing robust standard errors.

It is worth mentioning at this point that target year-specific variances help to obtain robust standard errors in the revision regressions, but they do not change the standard errors of the bias test as outlined above compared to the constant variance assumption. The reason that the covariance matrix that builds on target year-specific variances and the covariance-matrix that builds on a constant error variances yield numerically equivalent standard errors of $\boldsymbol{\alpha}_H$ is that in the test for unbiasedness (recall equation (3) and (5)) the forecast errors are regressed only on a fixed constant.⁴

3 Data and Empirical results

For the empirical analysis we explore the evolution and accuracy of the Consensus Forecast for twelve industrial countries for the years 1996 to 2006. The data set from Consensus Economics inc. contains - among other things - the pooled forecasts of the annual real GDP growth and inflation, measured as the change of consumer prices. The Consensus Forecast is a pooled forecast, based on a monthly survey among professional forecasters from the respective country, usually at a number of between 10 and 30. Due to its aggregation, it is often regarded as the market's opinion of the future development of the economy. As mentioned before, one main advantage the Consensus Forecasts offer to our analysis is that due to the monthly publication of revised forecasts for the current and the next year, an explicit revision process of 24 forecasts for every target year can be observed.

⁴ Cf. Davidson & MacKinnon (1993), Chapter 16.

3.1 Results for the unbiasedness test

One objection which is often raised against the application of the unbiasedness test to Consensus Forecasts is the so-called “aggregation bias” (see e.g. Bonham & Cohen, 2001). Keane & Runkle (1990) state two different sources of errors connected with this issue: (1) Individual forecasts differ because they are based on individual-specific information sets which implies that they are not rational on any other information set. (2) Aggregation masks individual biases, as averaging makes opposite biases to cancel out. However, albeit these objections are valid analyzing the behavior of individual forecasters, this is not in the focus of our tests. Here, the question is not whether the forecasts underlying the Consensus Forecast are biased and therefore imply non-rationality at the individual level, but whether the Consensus itself is biased. Therefore, the treatment of the Consensus Forecast as a self-contained forecast while disregarding the underlying individual forecasts is reasonable.

Appropriately testing the null hypothesis of an unbiased forecast requires a derivation of the correct covariance matrix of the error terms. Overlapping forecast targets induce autocorrelation between forecast horizons within and between adjacent forecast periods. Consequently, the correlation structure in our study has to reflect the monthly frequency in which forecasts for the current and the following year are published. The time interval in figure 1 illustrates the overlapping structure of the monthly forecasts for two adjacent periods:

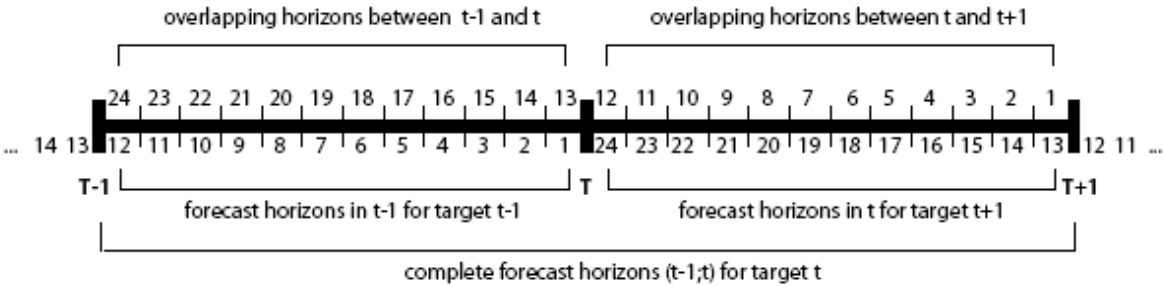


Figure 1: Structure of the Consensus Forecasts

To understand the serial correlation structure of the monthly Consensus Forecast more clearly, consider the following generic forecast made in January of the year t for the target year $t+1$. The realized value of that target is known only 24 months later and therefore comprises 24 monthly aggregate shocks which are not known to the forecaster in January of year t . However, the next forecast for the same target year $t+1$ is published one month later, in February of year t , and comprises 23 monthly aggregate shocks that are not yet known to the producer of the forecast. Since these 23 monthly shocks are not part of the available

information set of the January *and* February forecasts, both forecasts share these 23 shocks as a common component which induces a moving average process of order 23, denoted $MA(23)$. In an analogous manner the forecast produced in March of year t is correlated with the forecasts of January and February of year t for the target in year $t+1$ and so on.

The matrix Ψ , which was introduced in section 2.1, incorporates this special correlation structure for the monthly Consensus Forecasts and has the following form:

$$\Psi = \begin{pmatrix} \mathbf{A} & \mathbf{B} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{B}' & \mathbf{A} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \mathbf{B} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{B}' & \mathbf{A} \end{pmatrix},$$

where the component matrices, each of dimension $(H \times H)$, are given by:

$$\mathbf{A} = \begin{pmatrix} 24 & 23 & 22 & 21 & 20 & 19 & \dots & 3 & 2 & 1 \\ 23 & 23 & 22 & 21 & 20 & 19 & \dots & 3 & 2 & 1 \\ 22 & 22 & 22 & 21 & 20 & 19 & \dots & 3 & 2 & 1 \\ \vdots & & & \ddots & & & & \vdots & & \\ 3 & 3 & 3 & 3 & 3 & 3 & \dots & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & \dots & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 12 & 11 & \dots & 2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 12 & 11 & \dots & 2 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & & & & & & \vdots & & \\ 12 & 11 & \dots & 2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 11 & 10 & \dots & 2 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & & & & & & \vdots & & \\ 1 & 1 & \dots & 1 & 1 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}.$$

and $\mathbf{0}$ is a $(H \times H)$ matrix of zeros.

We test the unbiasedness of the real GDP and inflation forecasts each pooled over horizons. As actual values the realized values published in the June issue of the subsequent year were chosen.⁵

First of all we assume a common bias for all horizons to test if forecasts exhibit systematic bias on average. Table 1 shows the common bias ($\hat{\alpha}$) and standard errors for each country for GDP and inflation.

⁵ As discussed in Batchelor (2001), this choice follows the conventional view. Moreover, we redid our estimations with the actual values published in March of the subsequent year, and achieved qualitatively the same results.

Table 1: Common Bias of GDP growth and inflation forecasts

Country/Bias	GDP growth	Inflation
Germany	-0.4777* (-1.66)	-0.0333 (-0.18)
Belgium	-0.0636 (-0.22)	0.0231 (0.12)
Canada	0.0189 (0.06)	0.0519 (0.37)
France	-0.2663 (-1.15)	0.0205 (0.13)
Italy	-0.6333*** (-2.57)	0.0273 (0.14)
Japan	-0.0492 (-0.10)	-0.0545 (-0.47)
Netherlands	-0.1739 (-0.44)	0.0595 (0.36)
Spain	0.1155 (0.66)	0.1879 (0.94)
Sweden	-0.0992 (-0.38)	-0.5970** (-1.98)
Switzerland	-0.3947 (-1.13)	-0.3398* (-1.84)
UK	-0.0117 (-0.07)	-0.1087 (-0.91)
US	0.3568 (1.14)	0.0511 (0.27)

Note: Each column contains the estimated α and the t-statistic in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level.

The Consensus Forecasts of two countries show significant biases for the GDP growth rate and also two for the inflation rate. In particular, the Consensus participants in Germany and Italy predicted on average GDP growth rates which have been significantly biased upwards as indicated by the negative sign of the α estimate. In contrast, no significant common bias can be found for the inflation rate forecasts of these countries. Instead, a significant and systematic deviation from actual inflation is found for the Consensus Forecasts for Sweden and Switzerland. The former bias estimate is significant at the 5%-level, whereas the latter shows a significance level of 10%.

However, if there are different biases for different horizons, a test for common bias may wrongly fail to reject the null hypothesis of unbiasedness. Therefore, in a second step, we estimated the pooled forecast error model by assuming that there are h different bias terms. The results of these estimations are shown in figure 2 and figure 3 which display these estimates for ascending forecast horizons for GDP growth and inflation along with two times the robust standard errors. These standard error bands approximate the 95% interval of significance. Obviously, a horizon-specific bias that is significant at the 5% level will fall

outside these confidence bands. The width of the confidence intervals reflects the volatility of the country-specific macroeconomic shocks and the use of a coherent scaling of the ordinate of each graph permits cross-country comparisons. Accordingly, the forecast uncertainty was comparatively high for GDP growth in Japan, the Netherlands and Switzerland. In contrast, the narrow confidence intervals for the United Kingdom and Spain reflect relative low variances of the underlying monthly macroeconomic shocks u_{it} .

For GDP growth, the hypothesis of zero bias over one to 24 horizons cannot be rejected for most countries, but the figures also suggest that constant bias may not be an appropriate assumption for the forecasts of several countries. For Italy, all individual bias estimates for forecasts which exceed horizons of seven month are negative and significantly different from zero. Similar results are found for the GDP growth forecast for Germany. Here, all forecasts which are released 15 months or earlier before the end of the target year was reached are significantly biased and suggest systematic over-prediction. For the inflation forecast, again for Sweden and Switzerland evidence is found for the presence of systematic biases. For Sweden, forecasts beyond a horizon of 10 months significantly deviate from actual inflation while for Switzerland the cut-off point beyond which the hypothesis of zero bias can no longer be maintained is 12 months.

Altogether, the analysis of the horizon-specific forecast biases confirms the results of the common bias tests.

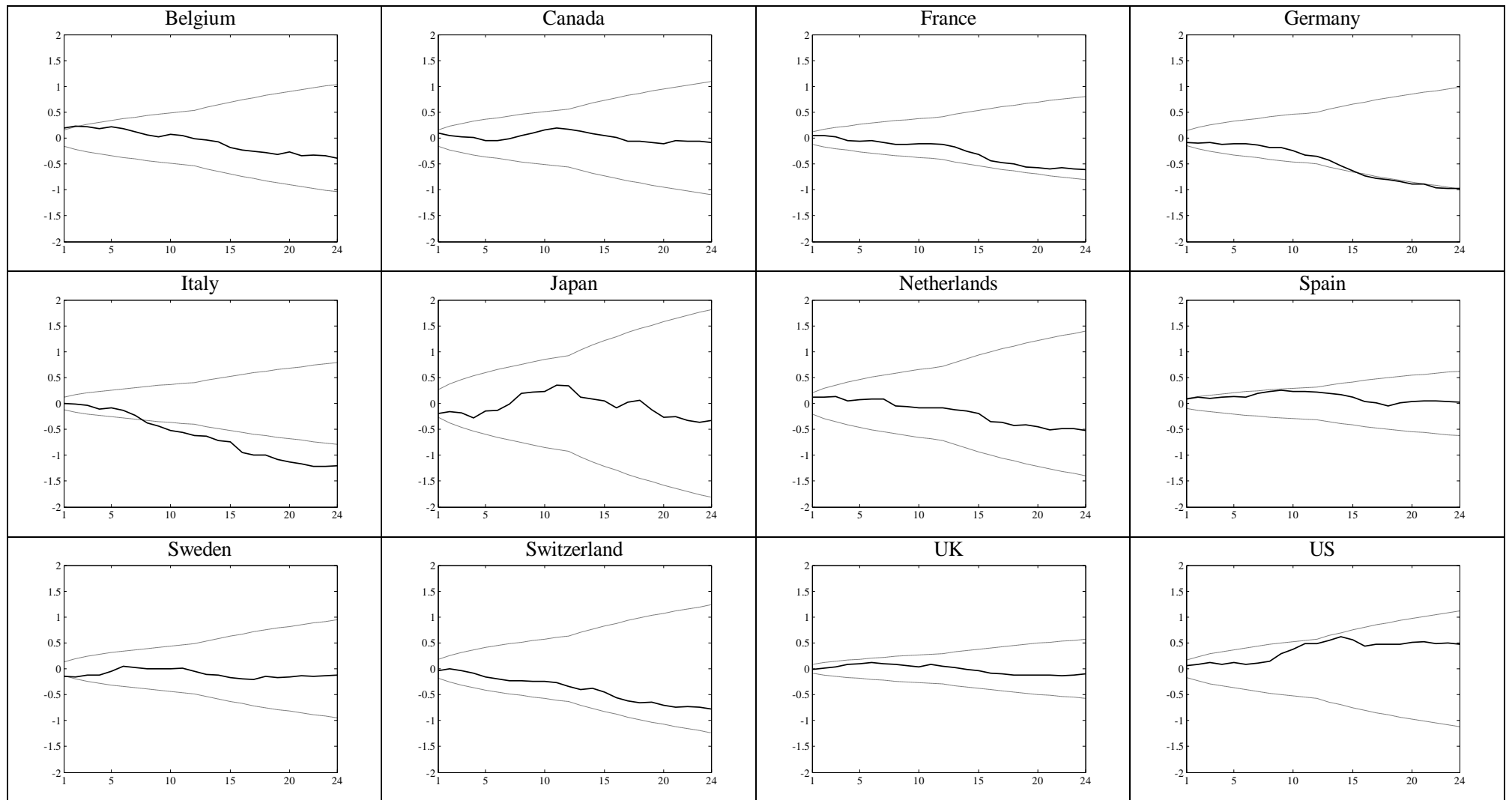


Figure 2: Horizon-specific bias estimates of GDP forecasts

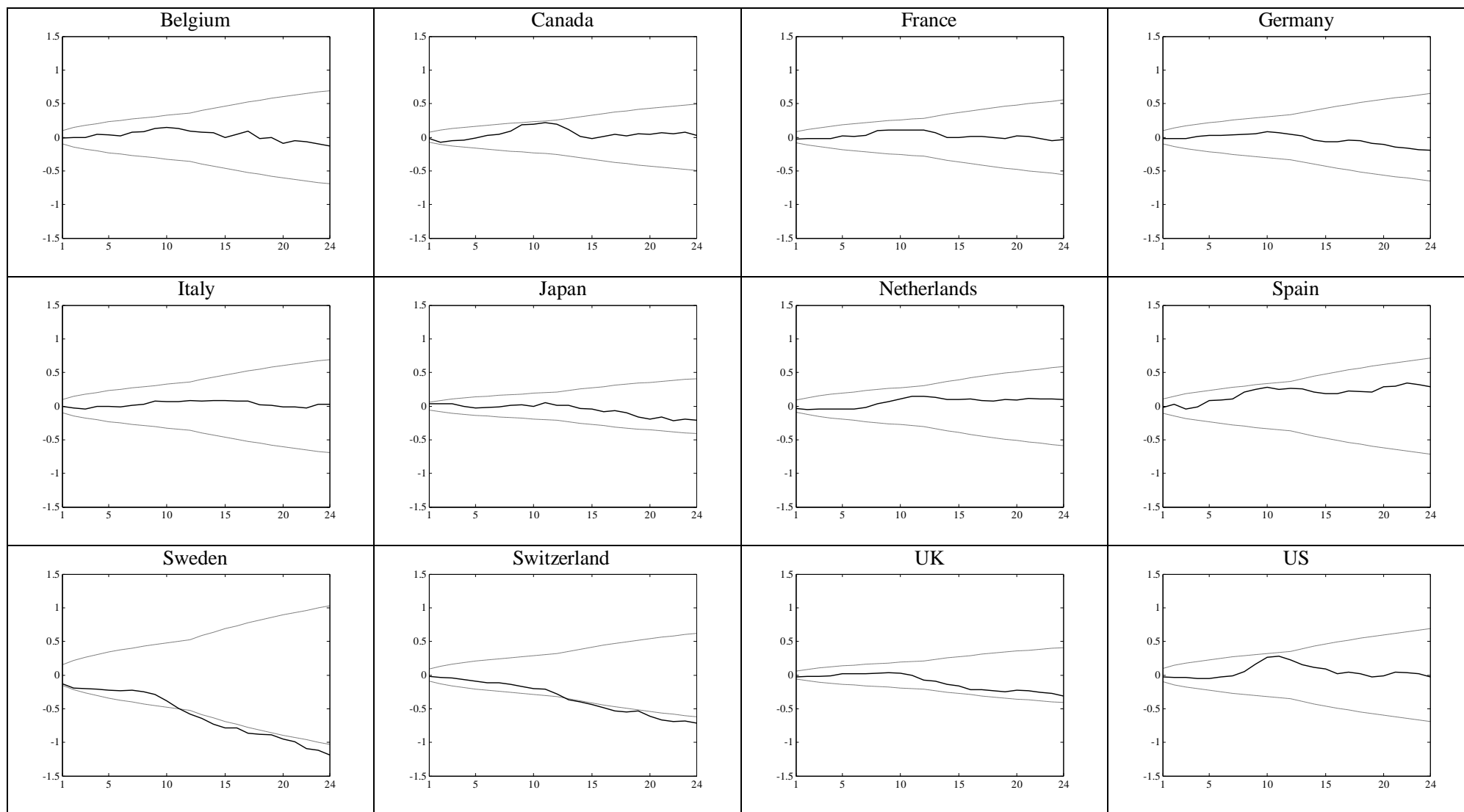


Figure 3: Horizon-specific bias estimates of inflation forecasts

Next, we tested the null hypothesis of unbiased forecasts by a sequence of F tests that allow for horizon-specific biases. The test procedure can be described as follows. The test runs as a sequence, that is, we first test whether the error from the forecasts with horizon h are zero, $H_0: \alpha_h = 0$, and then we continue with testing whether the errors from forecasts with horizon one and two are jointly zero until we test whether the errors from all forecast horizons are jointly zero, $H_0: \alpha_1 = 0, \dots, \alpha_H = 0$, and therefore unbiased. The corresponding results for the GDP and the inflation forecasts are shown in table 2 and table 3.

A general finding is that for most countries and for both target variables, we fail to reject the hypothesis that all individual bias estimates up to a horizon of at least 12 months are zero. Exceptions are the results for the GDP growth forecasts for Belgium, where the first two horizons are jointly significant different from zero, and for Sweden and Spain for which we fail to reject the hypothesis that the biases for the horizon of one month is zero. At the same time we obtain significant F statistics for many countries if we consider more than 12 months in the joint test for zero bias. Recall the forecasting scheme of the Consensus in which each month the participants are asked to release their forecasts for the current year and for the following year. Starting from January to December, each forecast with a horizon of 1 to 12 months refers to the current year whereas forecasts for the following year have horizons from 13 to 24 months. Given these relations, the findings reported in table 2 and 3 imply that the Consensus Forecasts generally provide an unbiased view on the GDP and price developments of the current year, but fail to do so when predictions for the following year are also taken into consideration. Furthermore, we only find insignificant F statistics throughout for the Netherlands and Switzerland (see table 2), but it has to be noted that the horizon-specific biases are not very precisely estimated for these countries. Therefore, although the magnitude of the horizon-specific bias estimates for the Netherlands and Switzerland can be quite substantial (recall figure 2), the standard errors for these estimates are also relatively large due to a high estimated volatility of the monthly aggregate shocks. In contrast, the bias estimates for the UK are very precise according to the standard error bands and therefore lead to a rejection of the null hypothesis of zero bias for all forecast which exceed horizons of 15 months although the individual estimates appear to be small in magnitude (<0.13).

Table 2: Testing the null hypothesis of unbiased GDP forecasts (horizon-specific bias)

	Germany	Belgium	Canada	France	Italy	Japan
h	F-Stat	F-Stat	F-Stat	F-Stat	F-Stat	F-Stat
1	1.5179	5.9488**	1.4770	0.8146	0.0000	2.1337
2	0.7665	3.0823**	0.9582	0.4186	0.0469	1.1021
3	0.5161	2.0594	0.6551	0.3093	0.0626	0.7406
4	0.4478	1.5985	0.4944	0.6902	0.4224	0.6888
5	0.3704	1.3220	0.5517	0.5567	0.3802	0.7495
6	0.3086	1.1376	0.4618	0.4677	0.4576	0.6253
7	0.2841	1.0695	0.4115	0.4526	0.7979	0.6594
8	0.2960	0.9965	0.4577	0.4215	1.4492	0.8685
9	0.2631	0.9098	0.4408	0.3746	1.3820	0.7739
10	0.3112	0.8525	0.4565	0.3394	1.4339	0.6970
11	0.3946	0.7861	0.4250	0.3086	1.3377	0.7122
12	0.3731	0.7756	0.3937	0.2904	1.3221	0.6532
13	0.3524	3.3755***	1.4238	1.3643	1.2620	0.6342
14	1.0957	3.4582***	1.3231	1.8765**	1.6700*	0.7242
15	1.7165**	3.8650***	1.2824	1.9848**	1.5713*	0.6840
16	1.9056**	3.6506***	1.2442	2.0499**	3.3061***	0.6810
17	2.1080***	3.5811***	1.1732	2.0026**	3.7175***	0.6415
18	2.0103***	3.3839***	1.1113	2.0020**	3.7344***	0.6194
19	1.9120**	3.2286***	1.1355	1.9743**	3.5450***	2.0187***
20	1.8173**	3.4920***	1.4498	1.8950**	3.8615***	3.7430***
21	1.7349**	3.3675***	1.3879	1.860/**	3.7164***	3.5648***
22	1.6576**	3.2589***	1.4981*	1.8020**	3.5767***	3.4944***
23	1.7309**	3.1190***	1.4557*	1.7747**	3.4990***	3.6515***
24	1.6859**	2.9950***	1.3952	1.7044**	3.6587***	3.5248***
	Netherlands	Spain	Sweden	Switzerland	UK	US
h	F-Stat	F-Stat	F-Stat	F-Stat	F-Stat	F-Stat
1	1.4626	3.7125*	4.1843**	0.1526	0.0450	0.5657
2	0.7350	2.0233	2.1003	0.1526	0.1126	0.3348
3	0.5000	1.3984	1.4874	0.1526	0.2102	0.2848
4	0.5261	1.1323	1.1155	0.2003	0.4392	0.2598
5	0.4268	0.9133	1.1016	0.2538	0.3874	0.2286
6	0.3569	0.7672	1.2476	0.2369	0.3904	0.2078
7	0.3060	0.9971	1.0904	0.2249	0.3925	0.1847
8	0.4505	0.9467	0.9725	0.1968	0.3660	0.1847
9	0.4038	0.8580	0.8645	0.1791	0.3453	0.4926
10	0.3664	0.8056	0.7780	0.1612	0.3513	0.5588
11	0.3331	0.7324	0.7088	0.1500	0.4218	0.6350
12	0.3060	0.6744	0.7164	0.1884	0.4204	0.5830
13	1.1005	2.6466***	1.1904	0.2218	0.4754	0.5388
14	1.0353	2.9587***	1.1054	0.2075	1.0690	0.6325
15	1.1415	3.0109***	1.6147*	0.2991	1.5652*	1.0021
16	1.2791	4.6613***	1.5668*	0.4276	2.9454***	1.3860
17	1.2240	4.5794***	1.8537**	0.4029	2.9948***	1.3108
18	1.3054	4.6330***	1.7978**	0.3809	3.3013***	1.2655
19	1.2390	4.4074***	1.7036**	0.4230	3.3174***	1.2104
20	1.3804	4.1914***	1.6838**	0.5366	3.1602***	1.1504
21	1.3697	3.9921***	1.6440**	0.5235	3.1078***	1.5941*
22	1.3359	3.8830***	1.5736*	0.5097	2.9728***	2.0398***
23	1.2778	3.7513***	1.5052*	0.4876	2.9167***	2.1854***
24	1.2324	3.5951***	1.6552**	0.4918	3.0368***	2.1367***

Note: The asterisks *, ** and *** denote significance at the 10%, 5% and 1% level.

Table 3: Testing the null hypothesis of unbiased Inflation forecasts (horizon-specific bias)

	Germany	Belgium	Canada	France	Italy	Japan
h	F-Stat	F-Stat	F-Stat	F-Stat	F-Stat	F-Stat
1	0.1386	0.0302	0.2365	0.4274	0.0000	1.3908
2	0.0693	0.0302	1.1827	0.2375	0.1358	0.6954
3	0.0462	0.0201	0.8673	0.1583	0.1308	0.4636
4	0.1126	0.2035	0.6653	0.1187	0.2867	0.6954
5	0.1178	0.1689	0.7215	0.2470	0.2293	0.7128
6	0.0982	0.1608	0.7589	0.2137	0.1961	0.6085
7	0.0891	0.2929	0.6843	0.2103	0.1854	0.5340
8	0.0823	0.2601	0.7836	0.5640	0.1773	0.5107
9	0.0770	0.3150	1.3536	0.5066	0.2414	0.4636
10	0.1005	0.2955	1.2241	0.4559	0.2203	0.4520
11	0.1039	0.2796	1.1612	0.4145	0.2002	0.6954
12	0.1068	0.2965	1.1088	0.3799	0.1936	0.8186
13	0.1200	0.2896	2.0251**	0.4595	0.1921	1.2650
14	0.8101	0.3142	2.9232***	1.6493*	0.3509	2.0691**
15	0.8093	0.8608	3.2557***	1.5394*	0.3658	1.9645**
16	0.8515	0.8093	3.1453***	1.4603	0.6771	1.8459**
17	0.8018	1.1185	2.9630***	1.5872*	0.6373	2.3957***
18	0.7684	2.0571***	3.6320***	1.4992*	0.8522	2.9332***
19	0.9811	2.0580***	3.5262***	1.6303**	0.8801	4.4278***
20	1.0165	2.9890***	3.8739***	1.6953**	0.9628	4.7878***
21	1.1972	2.8491***	4.5052***	1.6621**	1.0880	4.6640***
22	1.3455	2.7823***	4.3569***	1.6985**	1.0479	4.8502***
23	1.2873	2.6973***	4.1753***	1.8150**	1.2745	4.7705***
24	1.2393	2.5878***	4.0747***	1.7850**	1.2454	4.7069***
	Netherlands	Spain	Sweden	Switzerland	UK	US
h	F-Stat	F-Stat	F-Stat	F-Stat	F-Stat	F-Stat
1	0.6724	0.1146	2.6966	0.1507	0.7807	0.2766
2	0.4203	0.4153	1.6854	0.1507	0.4337	0.1537
3	0.2942	0.8879	1.1282	0.1130	0.2892	0.1024
4	0.2206	0.7805	0.8496	0.1224	0.2386	0.1076
5	0.1765	1.1973	0.6907	0.1658	0.3470	0.0861
6	0.1471	1.0025	0.5778	0.1946	0.2892	0.1178
7	0.1801	0.8756	0.4973	0.1668	0.2479	0.1185
8	0.3467	1.1994	0.4420	0.1648	0.2277	0.2920
9	0.3502	1.1171	0.4311	0.2135	0.2121	0.7513
10	0.4203	1.0512	0.5256	0.2260	0.1995	1.0480
11	0.4432	0.9973	0.6579	0.2089	0.2524	0.9639
12	0.4062	0.9237	0.7177	0.3924	0.6940	0.9758
13	0.4249	0.8612	0.8625	1.2235	0.6610	1.2794
14	0.4187	1.8737**	0.8403	1.2099	1.8604**	1.3089
15	0.4041	1.9497**	0.9212	1.2675	2.0363**	1.3279
16	0.3940	1.9642**	0.8677	1.3256	3.4003***	1.6678*
17	0.4990	2.0111**	1.0572	1.3844	3.4190***	1.6635*
18	0.4847	1.9328**	1.0037	1.3434	3.3401***	1.8938**
19	0.4602	1.9001**	0.9661	1.2842	3.1906***	2.2274***
20	0.9416	1.8133**	1.0242	1.6347**	3.0608***	2.2846***
21	0.8976	1.7816**	0.9766	1.6146**	3.0018***	2.3790***
22	1.1953	1.7146**	0.9453	1.5420*	2.8924***	3.2449***
23	1.2825	1.6429**	1.1198	1.5093*	2.7781***	3.2160***
24	1.2392	1.7341**	1.0771	1.5433*	2.8855***	3.0835***

Note: See table 2.

For the inflation forecast, we generally estimate the monthly shocks with higher precision than the shocks for the GDP growth rates. Only for Sweden fairly substantial standard errors for the horizon-specific bias estimates appear (see figure 3). As a consequence, all 24 sequential F tests reported in table 3 fail to reject the hypothesis of zero bias estimates.

Overall, the outcomes of the common bias test as well as the findings of the sequential test provide a detailed picture of the accuracy of the Consensus Forecasts but also demonstrate that care must be taken in interpreting individual test results.

3.2 Results for the forecast revisions

In this section we test the efficient use of information for forecast revisions across all horizons in a pooled regression. The test is constructed in order to check whether there is a significant impact of past revisions (the proxy for the past information set) on current revisions.

This test is conducted for both the first and second lag of revisions. The use of the second lag in the revision regressions for the evaluation of Consensus Forecasts goes back to Isiklar (2005). He shows that due to a problem associated with individual-specific information sets, the Consensus regressions using two lags of revisions prevent the danger of inconsistent estimations and even increase the chance of detecting forecast inefficiency as opposed to tests with the individual forecasts.⁶

As in section 3.1, in order to obtain robust standard errors of regression coefficients, we need to derive the covariance matrix of the error term. Again, this covariance matrix comprises in addition to the elements on the diagonal off-diagonal elements that represent the correlation structure in the forecast revision context. The following time interval illustrates the overlapping structure of monthly forecast revisions of two adjacent periods:

⁶ When considering two lags of revisions, the pooled regression to test for weak efficiency is

$$\mathbf{r} = \gamma \mathbf{r}_{-2} + \boldsymbol{\delta}_h + \boldsymbol{\omega} \text{ with } \mathbf{r} = \left[\mathbf{r}_{H-3,H-2}, \dots, \mathbf{r}_{1,2} \right]' \text{ and } \mathbf{r}_{-2} = (r_{t,H-1,H} \dots r_{T,H-1,H}, \dots, r_{t,2,3} \dots r_{T,3,4})'$$

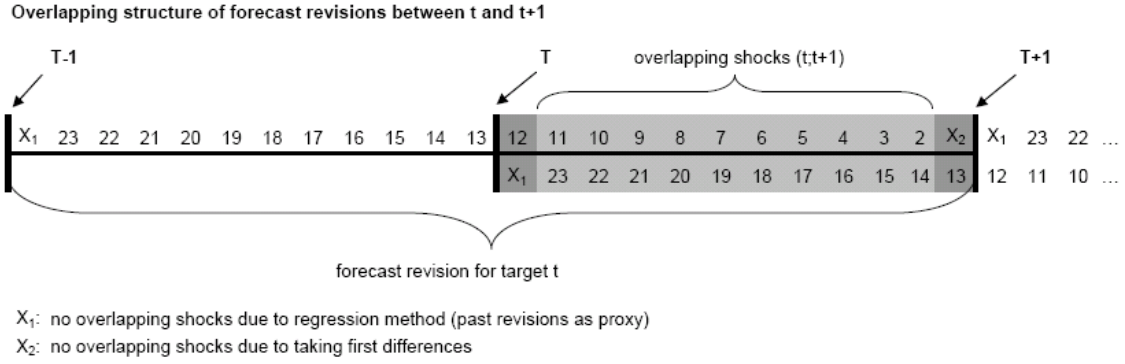


Figure 4: Structure of monthly forecast revisions

As already mentioned in section 2.3, the overlapping interval is determined by the method of estimating the forecast revisions. The numbers above and below the bold line stand for the forecast revisions between horizon $h+1$ and h for target t (numbers above) and $t+1$ (numbers below). For example, the first number above the bold line, 23, refers to the remaining error term, $u_{t,23}$, of the forecast revision between horizons 23 and 22 for target t . The following example should illustrate the overlapping structure of the forecast revisions for the Consensus data set. A revision between horizon 11 and 10 for target t induces autocorrelation between forecast revisions of adjacent periods, because the revision between horizon 23 and 22 for target $t+1$ occurs at the same time. This kind of correlation structure proceeds until the revision between horizon 2 and 1 for target t and simultaneously between horizon 14 and 13 for target $t+1$ is made⁷. As a consequence, the correlation structure described by matrix $\mathbf{\Lambda}$ of equation (13) takes the following form: unity elements on the diagonal as well as for off-diagonal elements that represent the autocorrelation of forecast revisions between adjacent forecast periods and zeros elsewhere.

More precisely,

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{A} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{B} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{B} & & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & & & \ddots & \vdots \\ \mathbf{0} & & \ddots & & & & & \mathbf{B} \\ \mathbf{B} & \mathbf{0} & & & & & & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & & & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & & \ddots & \ddots & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{B} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{A} \end{pmatrix}$$

⁷ The first difference method rules out the possibility of autocorrelation between forecast horizons within a forecast period.

where $\mathbf{A} = \mathbf{I}_T$ is an identity matrix of dimension T and \mathbf{B} is a $(T \times T)$ matrix which has ones on the first diagonal below the main diagonal:

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \ddots & \vdots \\ 0 & 1 & \ddots & \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

$\mathbf{0}$ is a $(T \times T)$ matrix of zeros which is placed $(H-2)/2$ times between the A and B matrices in each row of $\mathbf{\Lambda}$. This structure and displacement of the individual elements of $\mathbf{\Lambda}$ is owed to the scheme of the monthly forecast revisions as illustrated above and may change from data set to data set.⁸

We tested the predictability of forecast revisions of the real GDP and Inflation forecasts between 1996 and 2006 for the twelve industrial countries, each with the pooled method described in section 2.3. Table 4 and table 5 record the results for the real GDP and the Inflation forecasts. Column (1) and (2) compare the outcomes of the forecast revision test for the homoscedastic (1) and heteroscedastic (2) assumptions of the variance structure of the error terms, using one lag of past revisions as regressor. In addition, each specification controls for horizon-specific bias. Furthermore, we are interested whether forecast revisions are still predictable if we use two lags of past revisions as explanatory variable. The corresponding results are reported in columns 3 and 4 and we again test in the homoscedastic (3) and heteroscedastic (4) case.

The assumption of target year-specific variances is highly plausible if we look at the squared forecast errors of the target variables under consideration. Figure 5 and figure 6 in the Appendix show these squared errors \hat{v}_{it}^2 for the GDP growth forecast and the inflation forecasts which are a measure of the variance of the cumulated macroeconomic shocks (see equation (6)). For both target variables, it is fairly obvious that the constant variance assumption does not describe the error variability appropriately. In all countries and for both variables, periods which are associated with high volatility of aggregate shocks alternate with periods in which the volatility of shocks is relatively moderate or low. Therefore, taking this

⁸ When regressing current revisions on second lags, the matrix $\mathbf{\Lambda}$ has dimension $((H-3)T \times (H-3)T)$.

heteroscedasticity of forecast errors into account should result in more robust test statistics compared to the standard procedure which assumes homoscedastic errors.

Table 4: Results of the forecast Revisions regressions: growth rate of GDP ($H_0: \gamma = 0$):

Country	(1) L=1, HOM	(2) L=1, HET	(3) L=2, HOM	(4) L=2, HET
Germany	0.4585*** (5.74)	0.4585*** (5.18)	0.3811*** (4.71)	0.3811*** (4.24)
Belgium	0.2531*** (3.15)	0.2531*** (2.98)	0.2772*** (3.43)	0.2772*** (3.24)
Canada	0.4899*** (6.23)	0.4899*** (5.05)	0.2570*** (3.20)	0.2570*** (2.58)
France	0.3418*** (4.22)	0.3418*** (4.15)	0.2120** (2.57)	0.2120** (2.53)
Italy	0.3193*** (4.01)	0.3193*** (3.64)	0.1871** (2.30)	0.1871** (2.07)
Japan	0.3180*** (4.08)	0.3180*** (3.89)	0.2706*** (3.43)	0.2706*** (3.26)
Netherlands	0.3801*** (4.58)	0.3801*** (4.38)	0.3944*** (4.66)	0.3944*** (4.46)
Spain	0.1955** (2.39)	0.1955** (2.32)	0.2004** (2.47)	0.2004** (2.38)
Sweden	0.3298*** (3.90)	0.3298*** (3.63)	0.3527*** (4.11)	0.3527*** (3.82)
Switzerland	0.3288*** (3.90)	0.3288*** (3.65)	0.3049*** (3.56)	0.3049*** (3.32)
UK	0.2797*** (3.88)	0.2797*** (3.36)	0.3191*** (4.23)	0.3191*** (3.67)
US	0.3328*** (4.15)	0.3328*** (3.04)	0.2255*** (2.75)	0.2255** (2.01)

Note: Each column contains the estimate of γ and the t-statistic in parentheses. *, ** and *** respectively denote significance at the 10%, 5% and 1% level. All estimations include horizon-specific dummies. (1): Testing predictability of forecast revisions using one lag of revisions as regressor and homoscedastic shocks. (2): Testing predictability of forecast revisions using one lag of revisions as regressor and heteroscedastic shocks. (3): Testing predictability of forecast revisions using two lags of revisions as regressor and homoscedastic shocks. (4): Testing predictability of forecast revisions using two lags of revisions as regressor and heteroscedastic shocks.

The results of the revision regressions for the GDP growth forecasts are clear-cut. None of the entries in table 4 is insignificant which implies that we reject the hypothesis that the forecast revisions of the Consensus are unpredictable, throughout and for every country. It makes no difference whether one lag or two lags of revisions are considered and the assumption on the error variance (homoscedasticity or heteroscedasticity) is also irrelevant for the test decision. However, the heteroscedastic specifications generally lead to smaller standard errors.

In contrast, the outcomes for the inflation revisions are much more diverse (see table 5). A significant coefficient of the first lag of revisions is found for the Consensus Forecasts in Germany, Belgium, Canada, Italy, Sweden and the US. The second lag of revisions is significant in most countries; exceptions are the regression results for Belgium and the UK. The significance of the regression coefficient for Belgium drops from the 1% level to the 10%

level when switching from homoscedastic to heteroscedastic errors. In the cross-country comparison of table 5, the UK is the only country for which neither specification of the revision regression results in a rejection of the predictability and efficiency hypothesis.

Table 5: Results of the forecast Revisions regressions: inflation($H_0: \gamma = 0$):

Country/Bias	(1) L=1, HOM	(2) L=1, HET	(3) L=2, HOM	(4) L=2, HET
Germany	0.3313*** (4.10)	0.3313*** (4.02)	0.2662*** (3.26)	0.2662*** (3.19)
Belgium	-0.1971*** (-2.71)	-0.1971* (-1.87)	0.0616 (0.83)	0.0616 (0.56)
Canada	0.1878** (2.43)	0.1878** (2.42)	0.1647** (2.10)	0.1647** (2.09)
France	0.0999 (1.23)	0.0999 (1.27)	0.1959** (2.43)	0.1959** (2.50)
Italy	0.2795*** (3.62)	0.2795*** (3.04)	0.3325*** (4.24)	0.3325*** (3.59)
Japan	-0.0788 (-1.07)	-0.0788 (-1.09)	0.1742** (2.34)	0.1742** (2.39)
Netherlands	0.0993 (1.28)	0.0993 (1.20)	0.1866** (2.34)	0.1866** (2.19)
Spain	-0.0805 (-1.04)	-0.0805 (-0.85)	0.3914*** (4.77)	0.3914*** (3.95)
Sweden	0.3799*** (4.69)	0.3799*** (4.57)	0.2829*** (3.44)	0.2829*** (3.34)
Switzerland	0.0500 (0.61)	0.0500 (0.57)	0.2169*** (2.59)	0.2169** (2.40)
UK	0.0183 (0.23)	0.0183 (0.22)	0.1112 (1.38)	0.1112 (1.29)
US	0.2801*** (3.56)	0.2801*** (3.52)	0.2529*** (3.19)	0.2529*** (3.15)

Note: see table 4.

4 Summary and Conclusions

In this paper we asked whether the fixed-target forecasts of the Consensus Forecasts for 12 countries delivered a reliable view about the actual economic development over the period from 1996 to 2006. For this purpose we employed the pooled approach of Clements et al. (2007) which permits the evaluation of all forecasts for each target variable over 24 horizons simultaneously. We showed how the pooled approach needs to be adjusted in order to accommodate the forecasting scheme of the Consensus Forecasts in the analysis of bias, predictability of forecast revisions, and weak efficiency. Furthermore, we extended the pooled approach by taking heteroscedasticity in the form of year-specific variances of macroeconomic shocks for the test of predictability of forecast revisions and weak efficiency into account. Year-specific shock variances are a marked feature of the data. Extending the

pooled approach in these ways allows us to run several powerful tests in terms of efficiency within a coherent error model framework.

Bias was tested by assuming a common bias over all horizons and by permitting biases to vary over forecast horizons. A common bias for the GDP growth forecasts was detected only for Germany and Italy. In the case of inflation, a common bias was found for Sweden and Switzerland. A salient result is that for both target variables the variances associated with the macroeconomic shocks varied greatly between countries over the last decade. The uncertainty for the GDP growth was particularly high in Japan, the Netherlands, Switzerland and the US, while the same holds for the inflation forecasts for Sweden, Italy and Spain. These results partially confirm the presumption that the macroeconomic forecasts for the past 10 years were severely affected by the pronounced shocks in that period. The fact that for individual countries systematic biases can be observed applying the Consensus Forecasts reveals that these biases cannot be attributed to error sources at the individual level. Instead, this reflects that in these countries the forecasting industry on the whole was not able to cope with the shocks specific to the past ten years.

However, when considering different bias estimates for different horizons, more differentiated conclusions can be drawn. The outcomes of a sequence of F tests for testing the horizon-specific bias estimates for joint unbiasedness implied that the Consensus Forecast generally provided an unbiased view on the GDP and price developments of the current year. However, unbiasedness did not hold when predictions for the following year were also taken into consideration.

Applying the second lags of revisions following the approach by Isiklar (2005), the GDP growth predictions were devastating in terms of efficiency. For no single country we found evidence in favour of weak efficiency. For inflation, the outcomes of the efficiency tests were somewhat more encouraging. Here, we were not able to reject the hypothesis that forecasters are weakly efficient in information processing in Belgium and the United Kingdom.

The application of the first lags of revisions shows similar results. In addition to lacking information efficiency at the individual level such as the “forecast smoothing”, as proposed by Nordhaus (1987), this may also be explained by the construction of the Consensus Forecasts. As the forecasters do not send back their questionnaires simultaneously but over a space of several days, it may be assumed that news appearing during this space of time are not entirely incorporated in the Consensus Forecast of that month.

A further factor contributing to the correlation of the first lags of revision might be an imitation behaviour of the forecasters. As it is shown by Gallo et al. (2002), individual

forecasters are attracted by the mean forecast, i.e. the Consensus Forecast. Hence, news is not incorporated by all forecasters at the same time and not reflected by a large revision of the Consensus Forecast, but spreads via smaller revisions over several months, leading to a positive correlation of the revisions. Therefore, it can be assumed that this herding behaviour adds to the inefficiencies caused by smoothing at the individual level.

References

- Batchelor, R. A. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting*, 23, 189–203.
- Batchelor, R. A. (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus. *Applied Economics*, 33, 225–235.
- Bonham, C. S. & Cohen, R. H. (2001). To Aggregate, Pool, or Neither: Testing the Rational-Expectations Hypothesis Using Survey Data. *Journal of Business & Economic Statistics*, 19, 278–291.
- Clements, M. P. (2005). *Evaluating Econometric Forecasts of Economic and Financial Variables*. New York: Palgrave Macmillan.
- Clements, M. P. (1997). Evaluating the Rationality of Fixed-event Forecasts. *Journal of Forecasting*, 16, 225–239.
- Clements, M. P., Joutz, F. & Stekler, H. O. (2007). An Evaluation of the Forecasts of the Federal Reserve: A Pooled Approach. *Journal of Applied Econometrics*, 22, 121–136.
- Davidson, R. & MacKinnon, J.G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.
- Davies, A. & Lahiri, K. (1995). A New Framework for Analyzing Survey Forecasts Using Three-Dimensional Panel Data. *Journal of Econometrics*, 68, 205–227.
- Fildes, R. A. & Stekler, H. O. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics*, 24, 435–468.
- Gallo, G. M., Granger, C. W. J. & Jeon, Y. (2002). Copycats and Common Swings: the Impact of the Use of Forecasts in Information Sets. *IMF Staff Papers*, 49, 4–21.
- Greene, W.H. (2003); *Econometric Analysis*, 5th ed., Prentice Hall.
- Harvey, D. I., Leybourne, S. J. & Newbold, P. (2001). Analysis of a panel of UK macroeconomic forecasts. *Econometrics Journal*, 4, 37–55.
- Isiklar, G. (2005). On aggregation bias in fixed-event forecast efficiency tests. *Economics Letters*, 89, 312–316.
- Isiklar, G., Lahiri, K. & Loungani, P. (2006). How quickly do forecasters incorporate news? Evidence from cross-country surveys. *Journal of Applied Econometrics*, 21, 703–725.
- Keane, M. P. & Runkle, D. E. (1990). Testing the Rationality of Price Forecasts: New Evidence from Panel Data. *American Economic Review*, 80, 714–735.

- Loungani, P. (2001). How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth. *International Journal of Forecasting*, 17, 419–432.
- Mincer, J. & Zarnowitz, V. (1969). The Evaluation of Economic Forecasts. In: Mincer, J. (Ed.) *Economic Forecasts and Expectations*. New York: Columbia University Press, pp.3–46.
- Nordhaus, W. D. (1987). Forecasting Efficiency: Concepts and Applications. *The Review of Economics and Statistics*, 69, 667–674.
- Stekler, H. O. (2007). The future of macroeconomic forecasting: Understanding the forecasting process. *International Journal of Forecasting*, 23, 237–248.

Appendix

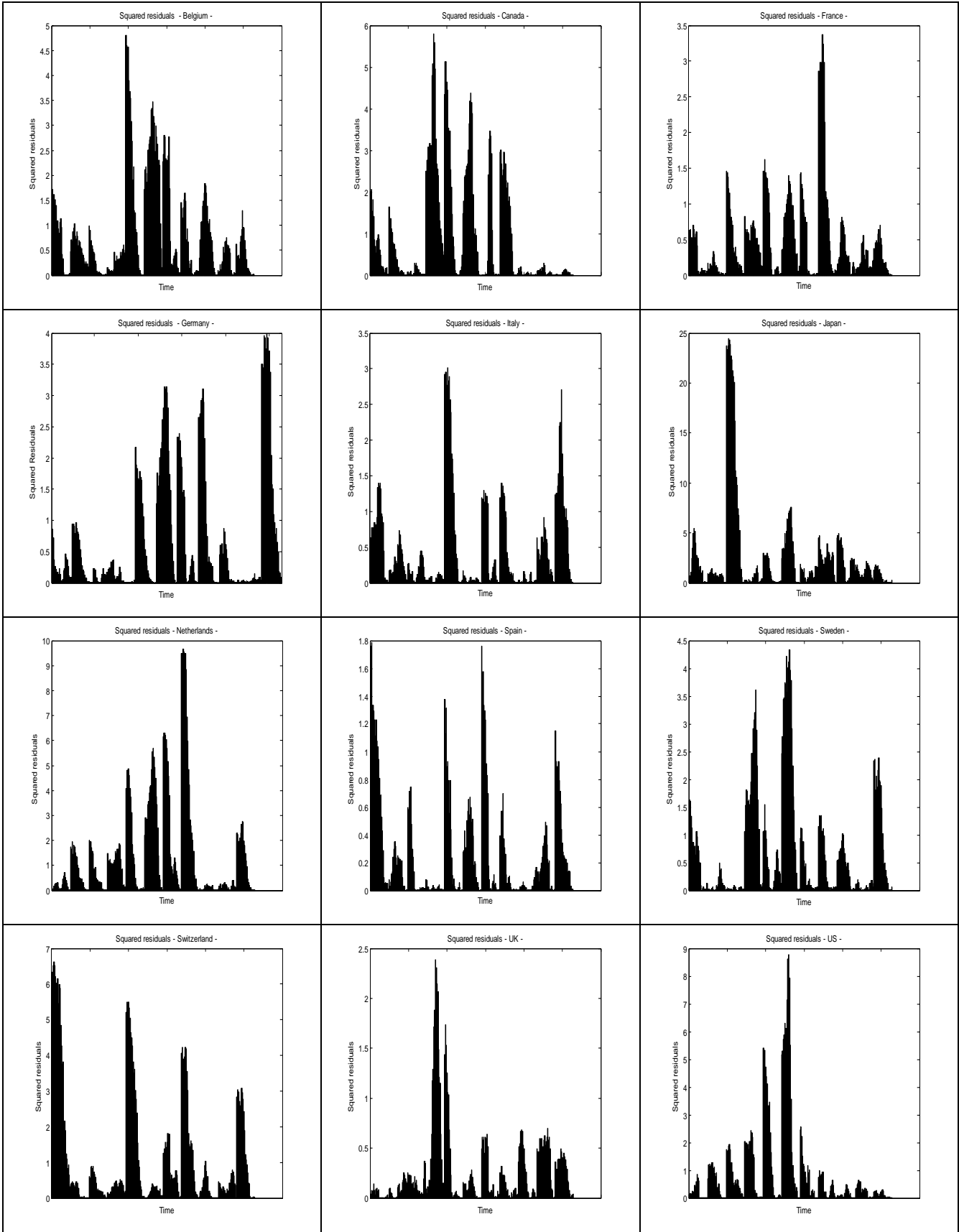


Figure 5: Squared forecast errors for real GDP

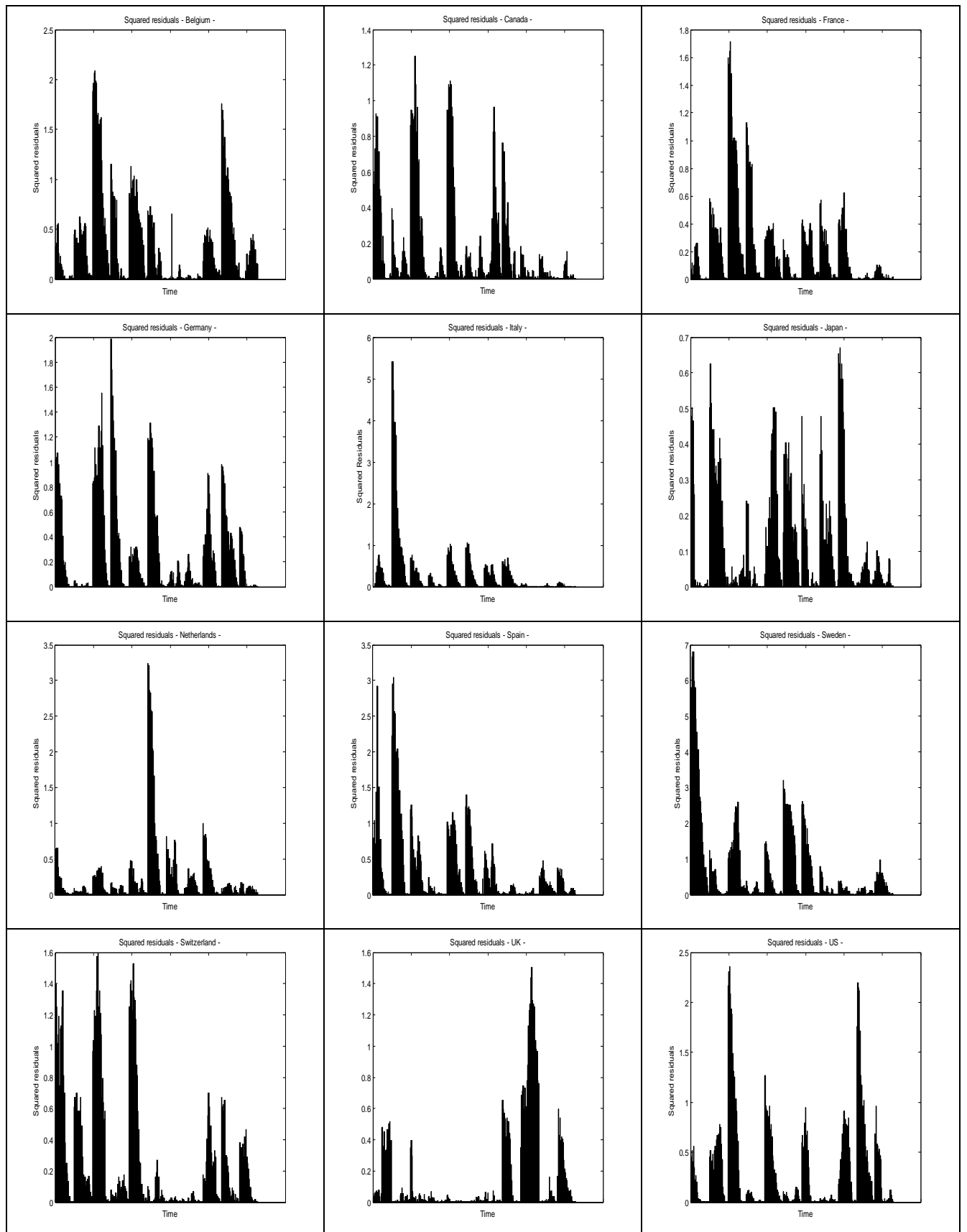


Figure 6: Squared forecast errors for inflation