Universität Hamburg

Department Economics and Politics

Too Many Cooks? The German Joint Diagnosis and Its Production

Ulrich Fritsche
Ullrich Heilemann

DEP Discussion Papers
Macroeconomics and Finance Series
1/2010

Hamburg, 2010

Too Many Cooks? The German Joint Diagnosis and Its Production

by Ulrich Fritsche* and
Ullrich Heilemann**

Abstract:

The "Gemeinschaftsdiagnose" [Joint Diagnosis (JD)] is the most influential semi-annual macroeconomic forecast in Germany. Jointly produced by up to six institutes, its accuracy as well as the large number of involved participants is often criticised. This study examines the JD's growth and inflation forecasts from 1970 to 2007, including most of the contributions of the forecasts submitted by the five institutes at the start of the JD. Four questions are addressed: (i) Are these forecasts unbiased and efficient? (ii) How do results change if we presume an asymmetric loss function? (iii) Are any of the institutes more accurate than the JD? Are five/six institutes necessary and at what cost? (iv) Do the institutes make strategic forecasts to influence the JD forecast? Results show that there is no strong evidence of bias or inefficiency of the institutes' forecasts and no evidence of asymmetric loss functions. Five institutes are not necessary, but it is very hard to predict the redundant institutes; however, the loss of accuracy by employing only two is small.

Keywords: Forecast accuracy, joint forecasts, strategic forecast behaviour
JEL-classification: E 37, C 53, D 72
Acknowledgements: Thanks for critical comments go to Roland Döhrn and Herman Stekler. For able research support, we thank Andreas Strasser and Stefan Wappler.

* Faculty of Economics and Social Sciences, University of Hamburg, e-mail: ulrich.fritsche@wiso.uni-hamburg.de
** Faculty of Economics, University of Leipzig, e-mail: heilemann@wifa.uni-leipzig.de

1. Introduction

Twice a year for nearly six decades approximately 50 affiliates of five, and later six major German[1] economic research institutes have convened to produce the "*Gemeinschaftsdiagnose*" [Joint diagnosis (JD)]. This attracts considerable attention not only from the federal government, who commission the report, but also from the business community, business associations, unions, and the media. While it is within the forecast community a kind of "Star of Bethlehem," the JD is also under attack. Customers complain about the seemingly low accuracy (Döpke & Fritsche, 2009) of the forecast and also its short life. By presenting new forecasts of their own about four weeks after publication of the JD, the institutes degrade the results of their joint effort. The high cost of the report (almost 2 m € in 2007) is also criticised. Hence, it is argued that the size of this "cartel" should be reduced to one or two alternating institutes. Notwithstanding this criticism, customers defend the present structure by emphasizing the advantages of this unique accumulation of wide-ranging expertise as well as the JD's pluralistic policy view that in their eyes assures high forecasting accuracy.

Surprisingly, the latter argument has rarely been tested. The few exceptions so far include Klapper (2004) and Heilemann & Quaas (2006), which addressed the way in which the institutes' forecast proposals (see below) are combined. The present paper extends and broadens these studies by asking:

- How accurate are the institutes' forecast proposals at the start of the JD and at the final JD? Do they pass usual tests of unbiasedness and efficiency? How does this relate to the JD?
- Do all institutes contribute to the JD forecast and to its accuracy in an equal way? Do these contributions change over time and in which ways?
- Is there a role for strategic behaviour, personal elements, and reputation gains in joint production of a forecast like the JD?
- Would a smaller number of institutes be sufficient – or even better – and what would be its price in terms of accuracy gains or losses?

In short: Do too many cooks spoil the broth?

To answer these questions is more difficult than it may appear at first sight. Leaving different theories or models aside, differences in accuracy may be due to different assumptions and to

[1] As described in the next section, since spring 2007 things have changed somewhat.

differences in the data employed,[2] and hence, may not tell much about the capabilities of the forecaster. The importance of the assumptions, such as world trade, import prices, interest rates, government construction outlays, etc., is difficult for third parties to assess since they are not listed in the proposals. However, judging from the institute's official forecasts, these differences tend to be rather small, and their short run influences are limited; they should not count for much of the differences. Also for similar reasons the second argument can be dismissed. The institutes must submit their forecast at least four days before the start of the JD, and it can be safely assumed that all institutes share the same information and process it in a similar way. As with any comparison of the following type, the result does not permit statements about the quality of the hypotheses or the model employed.

The study had to be restricted in several ways. Firstly, we examine only the forecasts of growth and inflation as the most interesting and most fundamental variables. While both forecasts are in a material sense not independent of each other, we analyse them – as nearly all studies do – as separate forecasts. Secondly, we examine only the autumn forecasts with a forecast horizon of 1 ½ years. With respect to policy, these results should be more important than the spring forecasts for the current year. The sample period covers the forecasts for 1970 to 2007, the years 1969 to 2006. Finally, we assume that the production of the JD is not disturbed by new information that would make the institute's proposals obsolete, and if there were such information, it would not change their relative positions.

The next section (2) briefly describes the JD's main institutional and procedural details. Section 3 presents the data. In Section 4, we first examine the accuracy of the forecast proposals and compare them with that of the JD; second, we look for each institute's contribution to the accuracy of the joint forecast; third, we search for strategic and personal influences. The final section (5) summarizes the findings of this paper and offers some conclusions.

2 **The "Joint Diagnosis"**

From 1950 to 2007 five members (six since 1993) of the "*Arbeitsgemeinschaft Deutscher Wirtschaftswissenschaftlicher Forschungsinstitute*" jointly produce and publish, usually in April and in October, a detailed macroeconomic forecast for Germany (Antholz, 2006). The institutes involved include:

[2] The institute's data differences usually occur with respect to the current quarter. They tend to disappear when during the JD indicator results become available.

- The Deutsches Institut für Wirtschaftsforschung (DIW), Berlin;
- The Hamburgisches Welt-Wirtschafts-Archiv, (HWWA) after 2007: Hamburgisches Welt-Wirtschafts-Institut (HWWI), Hamburg (from 1953 to 2006);
- The Ifo-Institut, Munich;
- The Institut für Weltwirtschaft (IfW) Kiel (since 1970 as a full member);
- The Institut für Wirtschaftsforschung (IWH), Halle (since 1993);
- The Institut für Landwirtschaftliche Marktforschung (ILM), Braunschweig (until 1970), and
- The Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Essen.

The work was based on an annually prolonged contract with the federal government, and for most of that time the contract was with the Federal Ministry of Economics. In 2007 the customer changed the rules somewhat with the result being that among others, the JD is now produced by three consortia lead by Ifo, IWH, and RWI, which all entailed at least one other institute, not necessarily a German one, and by IfW. Furthermore, the procedure involves regular evaluation procedures and the contract will not be prolonged in a "quasi-automatic" way. This should lead to a somewhat higher level of competition in the future.[3]

A central part of the contract is a numerical macroeconomic forecast for Germany. Since 1969 this forecast has to be presented in quantitative terms within the NA framework (levels, rates of change against previous year) with a horizon of up to two years (spring) or 1 ½ year (autumn). While the forecasts are based on quarterly data, the presentation of the forecasts report only semi-annual and annual results.

Approximately six weeks before the proper beginning of the JD, a kick-off meeting of the persons leading the institutes' teams ("*Federführende*") and the heads of the international groups is held. In a subsequent meeting the customer specifies special questions he wants to see assessed in the report. Two weeks later, a consultative meeting with experts from the Deutsche Bundesbank, the European Central Bank, and from the federal ministry in charge takes place. After this, the JD starts with about 50 members from the participating institutes; the meeting continues for up to ten days and is held at one of the institutes. As previously mentioned, four days before the proper start, each institute is required to submit its detailed forecast. These estimates are summarized and averaged in a "synopsis" made available to all participants. These estimates may differ from the institutes' actual official forecast, but the

[3] Since the rules are quite new, it is too early to evaluate if the level of competition did increase.

institutes are prohibited, now in their contracts, to publish new forecasts four weeks before and after the JD publication.

The production of the JD is organized in a number of working groups, each analysing and forecasting one or more major macroeconomic aggregates (Stäglin, 1998). The forecasts are based on a set of assumptions agreed upon by the "*Federführenden*" and the members of the international and monetary policy group in the kick-off meeting. These assumptions typically refer to oil price developments, interest and exchange rates, and wage developments. After merging all individual forecasts and completing a detailed calculation and discussion process, the various groups report their forecasts to the "*Federführenden.*" The production of the forecast is structured according to the informal GDP model. The methods employed range from exogenous settings to econometric equations and macroeconometric models to check the overall results. The "*Federführenden*" put the group results in a national account framework, and then discuss the results among themselves and with the various groups. They usually demand revisions of the forecasts, recheck the new overall forecast for consistency, etc. The forecast is usually completed after two or three such rounds. The final forecast is discussed in a plenum session. The iteration steps, though a natural part of the informal GDP model, also reflect problems of coordinating different views in or between groups or between groups and the "*Federführenden.*" More substantial reasons may be unwelcome policy implications of a forecast or a significant contrast from an institute's own views. Of course, theory has its place in this process as well as new events, policies, new information from indicators, data revisions etc. So far, there is not much evidence that changes of the exogenous environment played a major role in this process, as may be deduced also from Table 1 below. Of course, the final forecast is not always just the result of an exclusively evidence-based discourse.

After about ten days, a final draft of the completed 60+ page report is discussed by the "*Federführenden*" and their deputies. If one or two institutes still find that they do not subscribe to the forecast (or a particular view expressed in the report), they have the opportunity to express the points of dissent (not necessarily a complete forecast) in a minority vote. Not surprisingly, there is much bargaining between the "*Federführenden*" to avoid such a situation because minority votes are seen as diminishing the JD's impact. The final report is transmitted to the federal ministry to enable them to comment on the report's press release the next day.

At the start of the next JD, the "*Federführenden*" discuss, with different and varying passion, the last report and its forecast. Recently, the spring JD has a page or two analysing the accuracy of the *previous* spring forecast, but rarely goes beyond a simple forecast/actual comparison. There is no evidence that the institutes' forecast submissions also undergo a common *post mortem* analysis. One reason may be that with a multivariate, multi-period forecast not much can be learned from such an analysis. For the same reason it is difficult to pick a winner in a convincing way, and, as with all comparisons of forecast accuracy, it would be difficult for the group tied together by contract to draw any conclusions. Furthermore, for data reasons, the autumn forecast for the coming year can be properly evaluated only after about 18 months. Additionally, given the great amount of theories, assumptions, methods, and data employed by the institutes, higher accuracy is primarily seen as documenting good luck rather than greater skills. Finally and more generally, past success does not tell much about future accomplishments.

3 Data

The analysis concentrates on the annual forecast for the autumn year ahead including real GDP growth (rates of change against previous year) and inflation (deflator of private consumption, rates of change against previous year) submitted by DIW, HWWA/HWWI, Ifo, IfW, and RWI as parts of their NA forecasts; the IWH is not included because it joined the JD only in 1993. Actual data are the first-published official NA data as shown in the JD's spring forecast following the year covered by the forecast. The sample starts with the autumn forecasts for 1970 and ends with the forecasts for 2007. For some details, see Table A1a, b in the Appendix.

The submissions are not official forecasts and to protect their privacy we coded them as A through E. All data are available from the authors upon request; the decoding is available only to the institutes.

As to notation: $\hat{y}_{t|t-1}$ is the forecast of a variable for the year t with the information in $t-1$. y_t is its (first-published) realization. The forecast error reads as $\hat{e}_{t|t-1} = y_t - \hat{y}_{t|t-1}$.

4 Results

4.1 Accuracy, unbiasedness, efficiency: defining benchmarks

To get a first impression, we calculated the usual error statistics for the two variables. As Table 1 reveals, the errors are uncomfortably large but differ not much neither from German nor from international macroeconomic forecasting standards (Heilemann & Stekler, 2003). The differences between the institutes' forecasts as well as between them and the JD are rather small. As expected, the picture is more varied for sub-periods. The *decennium horribele,* the 1970s, witnesses not only considerable increases of the error statistics in general but also a tripling of the bias of growth forecasts. During the 1980s, accuracy increases and variance between the forecasters gets smaller. The 1990s see a further reduction of errors but a widening of the range of growth forecasts. The most recent period, because of its short length not quite comparable to the other results, experienced a new deterioration of the accuracy of the growth forecasts, while that of the inflation forecasts further improved.

The range of the institutes' proposals (minimum/maximum), the JD, and the actual development are confronted in Figure A 1a, b in the Appendix. Three features seem to be remarkable: First, the dispersion of the set of forecasts, at least in absolute terms, is generally declining, and most pronounced for inflation. The underlying factors might be the post-oil shocks better understanding of inflationary processes and a stability-oriented monetary system (Heilemann & Stekler, 2003). Second, the JD forecasts, with few exceptions, nest within the range of the proposals, suggesting that formally, the JD forecast is a weighted average of the proposals (Heilemann & Quaas, 2006). However, it is plausible that the weights were not constant over time and evolved according to the power and pervasiveness of different institution's arguments. Third, in most cases, the cross section distribution of all proposals misses the actual distribution. Even if the cross-section dispersion is sometimes referred to as an indicator of forecast uncertainty (Linden, 2003), this finding shows that the cross-section distribution is not a very simple proxy for the overall forecast uncertainty of the JD, a result that is in line with the findings of Bomberger (1996) and Rich & Tracey (2003).

About here Table 1

In addition to being of the highest accuracy, it is usually agreed that “rational” and non-intentional forecasts should be unbiased and should fulfil certain criteria of efficiency. In the present case, things may be more complicated. The proposals are not published and for a number of reasons the institutes may employ them to manoeuvre the JD forecast and the resulting policy advice in a particular direction. Such intentions cannot be ruled out, and, indeed, there is some anecdotic evidence for this, but as Table 1 reveals, the leeway for this is very small. Notwithstanding, we examine in detail whether forecasts are unbiased and efficient. We made use of a number of non-parametric tests which are summarized in Table 2:[4]

- As a test for bias, we applied the test of Dufour (1981) as used by Campbell & Ghysels, (1995) which has the advantage of avoiding restrictive assumptions on well-behaved residuals as in the case of most of the regression-based tests. Suppose that we compute the absolute value of the difference between each observation and the mean, and then rank these observations from high to low. Then the test is based on the idea that the sum of the ranks for the samples above and below the median should be similar. This is a signed rank test with the null hypothesis $H_0 : median\,(\hat{e}_{t|t-1}) = 0$. According to Campbell & Ghysels (1995), this test is preferable to parametric tests for small samples.
- As a test for efficiency, we again used a test proposed by Campbell and Ghysels (1995) to test for absence of serial correlation in the forecasts errors, i. e. that the median of the product of two consecutive forecast errors is centred on a median of zero, i.e. $H_0 : median\,(\hat{e}_{t|t-1} * \hat{e}_{t-1|t-2}) = 0$.

Overall, we have mixed evidence for deviations from unbiasedness or efficiency. On the one hand, the numbers in Table 1 showed quite remarkable error margins and point to at least a somewhat biased forecast of some institutions in certain periods. On the other hand, the non-parametric test procedure applied here (Table 2) does not point to significant anomalies given the few observations we have.[5]

About here Table 2

[4] For both tests, EViews reports a p-value for the asymptotic normal approximation to the Wilcoxon T-statistic. We performed the usual parametric tests as well; the results are available on request. One might argue that all these tests are of limited value since the forecasts are scientific forecasts, i.e. based on assumptions and are meaningful only if controlled for assumptions. This is of course true as Keereman (2003) has shown. However, due to limitations in our data set, we have to shift the analysis of this feature into the future.

[5] The results from parametric tests – using Newey-West standard errors – did also not show any evidence for bias or inefficiency.

Summing up the results from Table 2, the inflation forecasts show some signs of inefficiency in the 1970s and 1980s. As a consequence of the two oil crises and their aftermath, errors of inflation forecasts were large and serially correlated (Döpke & Fritsche, 2006a, Heilemann & Stekler, 2003). This problem diminished in the late 1980s and 1990s. All in all, there is evidence that inflation rate forecasts have to be seen as not very efficient during the early part of the sample – a problem which is obviously much less so at the present time. Given the results in the literature – namely Heilemann & Stekler (2003) – this is not astonishing.

4.2 Relative forecast performance: individual dimension

After we gathered information about the average accuracy of the institutes' proposals, we asked more precisely whether any of the institutions – according to the certain criteria explained below – "outperform" the JD, i.e. is significantly better in some features than the JD. From the huge battery of tests available and proposed in the literature (e.g., Diebold & Mariano, 1995), we decided to present only a few tests to make the point. First, we applied a standard test, namely the well-known Diebold-Mariano (1995) test with the null hypothesis of an average equality of expected forecast accuracy over the sample against the alternative of different forecast accuracies. We report the p-values for the (one-sided) test that the respective institution performs "better" than the JD forecast for the variable under investigation and vice versa (i.e. the JD outperforms the individual forecast). The test is reliable in larger samples only; therefore, only full sample results are reported (Table 3).

About here Table 3

The standard procedure shows evidence that some growth forecasts (B and D) differ significantly from the JD. In two out of five cases, the null of equal forecast accuracy is rejected at a 5 percent level; in four out of five cases, on a 10 percent level. The results, however, also indicate that the JD on an average outperforms the individual forecasts – a result in line with much of the forecast combination literature (Clemen, 1989; Makridakis & Hibon, 2000; Marcellino, 2004; Elliott & Timmermann, 2008).

Relative forecast performance on an average is, however, only part of the story. As Campbell & Murphy (2006) argue, the coverage rate of the individual forecasts taken together as a group before entering the merging procedure might be somewhat poor. Following Campbell & Murphy (2006) and based on the early work of Milton Friedman (1937, 1940), we apply another class of tests, not often used for forecast evaluation. These tests are based on the idea that the performance of each forecaster can at any point in time be ranked (in contrast to a ranking over the whole sample). If the group is homogenous, the forecasts cannot be distinguished; any forecaster has the same probability of achieving the same rank as any other. The average of the forecaster's rank should, therefore, not deviate from the mid-rank. Campbell & Murphy (2006) propose applying the tests to the "accuracy" and "style" of forecasts. Ranking by accuracy is understood as being based on the squared error (therefore independent if there is an over- or underestimation), by "style" as ranking by the size of the forecast (therefore aiming at detecting for some statistically significant differences in systematic over- or underestimation). (Table 4)

About here Table 4

When applying this test, one has to keep in mind that differences in the ranks must be pronounced to produce significant results in very small samples.[6] Nevertheless, even then there is some evidence that the growth forecasts in the 1990s differed significantly regarding their tendency to over- and underestimate. This is in line with the results of Antholz (2005) who showed that in this period the number of minority votes was the highest ever observed.

4.3 Relative forecast performance: JD versus a smaller group of forecasters

Since the differences between forecasters are so small, critics of the JD often question the usefulness of having a group of five/six institutes, as mentioned above. To examine this issue, we calculate the mean of the five proposals and compare the results with the JD outcome and the realizations (Table A.1 a, b in the Appendix).

The results again confirm Heilemann & Quaas (2006). The widespread perception by forecasters and the media is that the JD outcome can be pretty well guessed by averaging the proposals. In other words, looking only at JD's forecasting performance, the rewards of the joint efforts are very small.

In addition, we ranked, first, the institutes according to their performance over a certain decade (criteria: root mean squared error, RMSE); second, we selected the two best performing

[6] Comparing these results with those of Campbell / Murphy (2006) for Canada reveals that here, the differences are in general much less pronounced. One has to keep in mind, however, that the authors had a more detailed data set (demand components).

as well as the two worst performing institutes over the decade and calculated synthetic JD results by simply averaging the forecasts of the respective two institutions for the decade following the decade where the choice was made. In doing so, we assume a situation where all institutes' forecasts are evaluated in a way to serve as a base for a JD contract for the coming decade. Third, we compare the RMSE's of this synthetic JD's with the RMSE of the real JD (Table 5).

About here Table 5

The results can be interpreted in two ways. First, in most cases, the JD result slightly outperforms the synthetic JD's; however, the differences are very small. This is not astonishing as the results from the forecasting combination literature – a weighted average of a greater number of informative forecasts often outperforms an average based on a smaller number of forecasts – is confirmed. Much more astonishing here is that it makes no difference if we pick the best or the worst performers.

4.4 Relative forecast performance: cross-section information

It may be objected to that the synthetic JD is the result of a simple averaging of two forecasts while the actual JD is a joint product of five or six institutes. Table 5 gave no hint that this combination results in sizeable accuracy gains. Nevertheless, the question as to whether the submission of five or six forecasts to the JD forecast contains information not already covered by the average of the forecasts of some randomly selected institutes deserves a closer examination. We, therefore, had a look at the cross-section distribution of the forecasts. As indicated above, the cross-section distribution of forecast errors cannot serve as a very simple measure for individual forecast uncertainty due to theoretical reasons (Bomberger, 1996, Rich & Tracey, 2003); however, the range of the submissions might contain some information about the size of JD's future errors. The range between minimum and maximum may be seen as such an indicator or the size of the forecast errors to be expected – at least for the most prominent figure, the GDP/GNP growth rate.[7] The relationship is shown in Figure 1.

About here Figure 1

[7] The finding is confirmed by two simple regressions. In case of GDP/GNP growth, we receive $|e_t| = \underset{(2.18)}{0.48} + \underset{(3.11)}{0.56}\, Range_t, \bar{R}^2 = 0.18$. In the case of inflation, we receive $|e_t| = \underset{(3.16)}{0.57} + \underset{(1.00)}{0.24}\, Range_t, \bar{R}^2 = 0.02$ (t-statistics in parentheses).

The scatter plot suggests a positive relationship between the range of submissions and the JD's absolute forecast errors – at least for growth forecast errors. This is in line with previous findings for Germany (cf. Döpke & Fritsche 2006b) and many others countries (Dovern et al., 2009) which find an increasing disagreement across forecasters during times signalling upcoming recessions.[8] As a result, by publishing the institutes' proposals, the informational content of the JD could be increased. While the cost for doing this would be small, it might reduce the institutes' propensity to compromise and the JD's image of coherence might suffer. Similar conflicts arise with the selection of participating institutes and here the new arrangement of the JD has opted for a broader pool from which to choose contractors.

4.5 Strategic behaviour and changing influences

Since the JD in a sense competes with the participating institutes' (analysis) and forecasts, strategic behaviour at the JD for both institutional and personal reasons, cannot be excluded. Of particular interest is the question as to whether the forecasters' loss functions are symmetric, and, if they are not, this could mean that the proposals are intentional.

Indeed, some forecasters of the JD group have long since been assigned to the "Keynesian" or "Monetarist" camp (Döpke 2000). If this would hold also for their proposals, we would expect a tendency to underestimate growth and overestimate inflation. Therefore, we first test for the existence of asymmetric loss functions[9]. The above tests for bias and rationality implicitly assumed that forecasters have a symmetric loss function. However, this is not necessarily the case. There are many reasons why forecasters might try to avoid errors on the side of the distribution, in particular because of the personal cost in terms of reputation.

Second, we test for the existence of changing "dominance" in the JD. Given the JD practice of forecasting, it is obvious that not just theory, data and methods but also "men" play quite a role in the process of forecast production. Since we were not able to track the changes in the

---

[8] However, this result considers only the alpha error – neglecting the fact that a high range might give a wrong signal and does not indicate a future recession. In 2004, the range was high but did not signal a recession. This could indicate that a high range might be closer linked to turning points in general but not to slips into recessions alone. It may be added that in the period under study the autumn forecasts of the JD did never forecast a recession.

[9] Under present circumstances, with at least three parties involved – the customer, the JD, and the various institutes – the concept of loss function appears as very opaque. Each party might have its own trade-offs and be of different and varying influence.

personal structure of the JD and of the "*Federführenden,*" we test for changes in the weights of the institutes over time in the JD exercise of "forecast combination." Despite being well aware that such shifts may be attributable to a host of other factors, experience within the JD and outside suggest that these weights might well be seen as a function of "charisma" and "persuasive power."

4.4.1 Is there evidence for an asymmetric loss function?

A large body of literature has investigated the issue of accuracy, bias, and especially the various concepts of rationality and concluded that the full rationality of business cycle forecasts has to be rejected. (e.g. Döpke & Fritsche, 2006; Heilemann & Stekler, 2003, Kirchgässner & Savioz, 2001 Fildes & Stekler, 2002).

Batchelor & Peel (1998), Elliott et al. (2005) and Patton & Timmerman (2006) disputed this and argue that the observed forecast errors might well be in line with rational expectations, as long as the loss function of the respective forecasters is not symmetric, as it is assumed in the usual tests of the rationality hypothesis. Such an asymmetric loss function might also be quite relevant in "games" like the JD where intentional behaviour is not unlikely.

To test for asymmetries, we refer to the approach of Elliot et al., (2005) which starts from the general loss function:

$$\textbf{(1)}\quad L(p,\alpha,\Theta) = \left[\alpha + (1-2\alpha)\cdot \mathbf{I}(y_{t+1} - \hat{y}_{t+1} < 0)\left|y_{t+1} - \hat{y}_{t+1}\right|^{p}\right]$$

In this type of loss function, the parameter p represents the underlying general assumption about the shape of the function. In particular, $p = 1$ stands for a linear-linear (lin-lin) loss function, while in the case of $p = 2$ the calculations are based on a quadratic-quadratic (quad-quad) loss function. Furthermore, the loss function consists of a symmetry parameter α . It represents the degree of symmetry or asymmetry of the loss function. In particular, a value of $\alpha = 0.5$ yields a symmetric loss function, while $\alpha > 0.5$ represents the case of forecasters' incentives to issue optimistic forecasts. Finally, any value $\alpha < 0.5$ stands for cases of too-pessimistic forecasts. Using different parameters, we can test for a great number of loss functions. For example $L(1,\frac{1}{2},\Theta) = (y_{t+1} - \hat{y}_{t+1})$ yields a symmetric quadratic loss function (Elliot et al. 2005). The approach also to determine criteria of forecast optimality, which, in turn, can be interpreted as the moment condition(s) for an instrument variable estimator (here: GMM). By using the sequence of forecasts, the proposed GMM estimator yields the following expression:

$$\textbf{(2)}\quad \hat{\alpha}_T = \frac{\left[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} \upsilon_t \left|y_{t+1} - \hat{y}_{t+1}\right|^{p_0-1}\right]' \hat{S}^{-1} \left[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} \upsilon_t \mathbf{1}(y_{t+1} - \hat{y}_{t+1} < 0)\left|y_{t+1} - \hat{y}_{t+1}\right|^{p_0-1}\right]}{\left[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} \upsilon_t \left|y_{t+1} - \hat{y}_{t+1}\right|^{p_0-1}\right]' \hat{S}^{-1} \left[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} \upsilon_t \left|y_{t+1} - \hat{y}_{t+1}\right|^{p_0-1}\right]}$$

with $\hat{S}$ as a weighting matrix. Since $\hat{S}$ in turn depends on $\hat{\alpha}$, the estimation has to be performed iteratively, assuming $\hat{S} = \mathbf{I}$ in the first round and using a vector υ_t of instrument(s). Hence, the estimation is based on the considerations of Hansen (1982).[10] Elliot et al. (2005) show that the estimator of α_T is asymptotically normal and renders it possible to test for the hypothesis $\alpha = 0.5$ i.e. for loss function symmetry.

For the procedure to work appropriately, instruments are needed. Following Elliot et al. (2005), the instruments considered here are: i) a constant; ii) a constant and a lagged forecast error; iii) a constant and the lagged variable to be predicted (i.e. the growth and inflation rate, respectively); and iv) a constant, the lagged forecast error, and the lagged variable to be predicted. The estimation results for the data set under investigation are given in Table 6 and summarized in Figures 2 and 3.

About here Table 6, Figures 2 and 3

[10] See furthermore Hansen and West (2002) for a survey and a discussion of the relation of IV (GMM) estimation to macroeconomic applications.

In summary:

- First, the null of a symmetric loss function is almost never rejected if we look at the *full sample*. This holds for growth as well as inflation forecasts of all institutes and of the JD.
- Second, for the *growth forecasts* and by splitting the sample around the middle, we find that on an average they were slightly over-pessimistic in the first half of the sample and slightly over-optimistic in the second half of the sample, which confirms the results from the bias results. Using the J-test we can test for the significance of the bias. There is only one institute, where we find a significant deviation from symmetry in the loss function, Institute C in the second half of the sample. In all other cases, the deviations can be regarded as being too small to give an indication of asymmetric loss functions.
- In the case of *inflation forecasts,* we find a slight tendency of over-pessimism, irrespective of the assumed sample. In the second half of the sample this tendency became somewhat more pronounced, but is still far from being significant for any of the forecasters (again Institute C being the exception).

In general, the case for any asymmetric loss function must therefore be seen as extremely weak. From that perspective, there is no indication for intentional forecasts. In contrast, in almost all cases, the forecasts are symmetric and not biased. Periods of over-optimism or pessimism seem to be relevant for the whole bulk of forecasters but not related with the "position" of certain institutions, if there ever was any one with regard to intentional forecast errors.[11]

4.4.2 Changing "dominance" over time

To assess the "dominance" problem, the influence of institutes on the JD and its change over time, we analyzed the stability of the institutes' weights in explaining the JD results by "forecast combination" regressions. More specifically, we regressed the JD forecast on the proposals of all institutes and restricted the sum of the coefficients to be one (Bates & Granger, 1969, Granger & Ramanathan, 1984) that is $JD_t = \sum_{i=A}^{E} \beta_i X_{it} + \varepsilon_t$ with X_{it} being the individual

[11] It may be argued that the institutes' theoretical positions are more closely reflected in the policy recommendations of the reports. This is confirmed by the fact that most of the "minority votes" address policy questions, depending on the degree in which the institutes had subscribed to Keynesian or Monetarist positions. A similar "bias" with respect to the institutes' forecast submissions, however, is hard to detect.

"proposals" and estimated under the restrictions, that all weights have to be positive and sum up to 1. To apply the restriction we used an exponential transformation and then applied non-linear least squares for estimation. For a better presentation, we multiplied the coefficients by the factor 100. Figure 4, Panels a and b, illustrate that for some periods some institutions seem to have been particularly influential. This influence also changed strongly.

About here Figure 4

Institute A, for example, was quite influential during the 1970s and 1980s, and for inflation also in the 1990s. After that time, other institutes took over this position. Institute D, for instance, had a minor influence in the 1980s but gained considerable influence in the 1990s, while Institute E seems to have dominated the new millennium so far. To which degree this reflects competence or factors like "charisma" or personal influence remains unclear, at least to third parties. However, we were not able to detect a clear pattern between forecast accuracy in the different decades (Table 1) and the forecast combination results.

However, with the possibility of "minority votes" the JD offers the institutes additional possibilities to let their voices be heard. The institutions use this option quite differently: only two of them made wide use of it, one in the 1980s and the other one in the 1990s; the others hardly used it at all. Of course, those institutes that entered the room with an "outsider" position and opted for a minority vote could not be detected as being particular influential in the result. Nevertheless, this does not mean that those institutes did not influence; nor should they be considered as bad forecasters by their client or by the public.

5. Summary and conclusions

This paper analysed a number of questions arising from the results of the "Joint Diagnosis" (JD), the most influential semi-annual macroeconomic forecast for Germany. While the forecast accuracy of the JD has often been studied, this is not the case with the forecast proposals of the participating institutes. Those forecast proposals have been employed to shed some light on the production process of this forecast. The analysis was restricted to the autumn forecasts of growth and inflation which have a forecast horizon of about 18 months. The period covered is from 1970 to 2007.

Accuracy and bias of the institutes' proposals for the two variables differ very little and the differences of the corresponding mean value with the JD are rarely larger than 0.5 percentage points. From a policy perspective, the accuracy of the growth forecasts is unsatisfactory low, but this is consistent with previous national and international findings. This is also true for the stability of the results: improvements of accuracy are mainly to be recorded for inflation forecasts. In general, the proposals passed the usual tests of unbiasedness and efficiency, but of course, this does not compensate for the disappointing low accuracy of the forecasts.

Since the JD forecasts compete with the institute's own forecasts, they could be expected to make strategic proposals to influence the outcome of the joint product, i.e. to have asymmetric loss functions. However, related tests failed; all five institutes seem to share standard loss functions. Of course, the institutes' hierarchy with respect to forecast accuracy changed over time and in some periods, and on an average some outperformed others and the JD. We found some evidence as to the dominance of some institutes and strategic behaviour, though this influence seems to have been weak.

Status, persuasive power, and bargaining power seemed to be disproportionally distributed among the group. This raises doubts as to the necessity of having such a large group of institutes involved in such a resource-intensive process. However, the selection of a set of forecasters based on past records did not warrant a better performance. In practice, however, with annual forecasts this is difficult to do, despite the fact that here, as in general, past forecasting success is no guarantee for future success. Therefore, it may be somewhat comforting that even the worst couple of forecasters to replace the JD would hardly make a noticeable difference.

References

Antholz, B. (2005). Zur Treffsicherheit von Wachstumsprognosen: prognostizierte und tatsächliche Wachstumsraten des Bruttoinlandsprodukts Deutschlands 1950-2004, Münster: EHA-Verlag.

Antholz, B. (2006). Geschichte der quantitativen Konjunkturprognose-Evaluation in Deutschland. *Vierteljahrshefte zur Wirtschaftsforschung*, 75(2): 12-33.

Batchelor, R. & D. Peel (1998). Rationality testing under asymmetric loss. *Economics letters* 61: 49-54.

Bates, J. M. & Granger C. W. J. (1969). The Combination of Forecasts. *Operations Research Quarterly* 20: 451-468.

Bomberger, W. A. (1996). Disagreement as a measure of uncertainty. *Journal of Money, Credit, and Banking*, 28: 318–392.

Campbell, B. & Ghysels, E. (1995). Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency. *The Review of Economics and Statistics*, 77 (1): 17-31.

Campbell, B. & Murphy S.(2006). The Recent Performance of the Canadian Forecasting Industry. *Canadian Public Policy – Analyse de Politques*, 32, No. 1, 2006.

Clemen, R.T. (1989). Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, 5: 559-581.

Diebold, F. & Mariano R. (1995). Comparing Predictive Accuracy. J*ournal of Business Economics and Statistics*, 13: 253-63.

Döpke, J. (2000). Haben Konjunkturprognosen einen politischen Bias? *Schmollers Jahrbuch - Zeitschrift für Wirtschafts- und Sozialwissenschaften*,120, 587-620.

Döpke, J. & Fritsche U. (2006a). Growth and inflation forecasts in Germany: a panel-based assessment of accuracy and dispersion. *Empirical Economics,* 31, 777-798.

Döpke, J. & Fritsche U.(2006b). When do Forecasters disagree? An Assessment of German Growth and Inflation Forecast Dispersion. *International Journal of Forecasting*,22, No. 1, 125-135.

Döpke, J. & U. Fritsche (2009). Zehn Gebote zum Umgang mit Konjunkturprognosen. *Zeitstchrift für Politikberatung*. 2 (1), 26-42.

Dovern, J. & U. Fritsche & J. Slacalek (2009), Disagreement among Forecasters in G7 countries. ECB Discussion Paper (forthcoming).

Dufour, J. (1981). Rank Tests for Serial Dependence. *Journal of Time Series Analysis*, 2 (1981). 117-128.

Elliot, G., Kommunjer, I., Timmermann A. (2004). Biases in macroeconomic forecasts: irrationality or asymmetric loss? http://repec.org/esNASM04/up.6440.1075609439.pdf

Elliot, G., Komunjer I., Timmermann A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies,* 72, 1107-1125.

Elliot, G. I. & A. Timmermann (2008). Economic Forecasting. *Journal of Economic Literature*, 46, 3-56

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation. *Econometrica*, 50, 987-1008.

Fildes, R. & Stekler, H.O. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics,* 24, 435-468.

Friedman, M.(1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32: 675-701.

Friedman, M. (1940). A Comparison of Alternative Tests of Significance for the Problem of m Rankings. *Annals of the Mathematical Statistics*, 11, 86-97.

Granger, C.W.J. & Ramanathan R. (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting* 3: 197-204.

Hansen, B. E. & West K. D. (2002). Generalized method of moments and macroeconomics. *Journal of Business & Economic Statistics*, Oct 2002, 20(4): 460-469.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistic*s, 23, 365–380.

Hansen, P.R. & Lunde, A. (2007). MULCOM 1.00 Econometric Toolkit for Multiple Comparisons. mimeo.

Heilemann U & Stekler H. O. (2003). Has the accuracy of German macroeconomic forecasts improved? *Discussion Paper of the German Research Council's research centre* 475. 31/03

Heilemann, U. & Quass G. (2006). Who dominates and who wins joint forecasts? Results for the German Joint Diagnosis. Paper presented at the 26th International Symposium on Forecasting (ISF), June 2006, Santander, Spain.

Keereman, F. (2003). External assumptions, the international environment and the track record of the Commission Forecasts. *European. Economy*, No. 189, 2003.

Kirchgässner, G. (1991). Sind die Prognosefehler der Konjunkturprognosen autokorreliert?,ifo-Studien, 37, 151-158.

Kirchgässner, G. & Savioz M. (2001). Monetary policy and forecasts for real GDP growth: an empirical investigation for the Federal Republic of Germany. Germany Economic Review 2: 339-365.

Klapper, Matthias (2004). Combining German Macroeconomic Forecasts Using Rank-Based Techniques, SFB 475 "Komplexitätsreduktion in multivariaten Datenstrukturen" http://hdl.handle.net/2003/4882

Makridakis, S. & Hibon M. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting* 16: 451-476.

Marcellino, M. (2004). Forecast pooling for short time series of macroeconomic variables. *Oxford Bulletin of Economic and Statistics,* 66, 91-112.

Patton, A.J. & A. Timmermann (2007). Testing Forecast Optimality under Unknown Loss. *Journal of American Statistical Association*, 102, 1172-1184

Rich, R., & Tracy, J. (2003). Modelling uncertainty: Predictive accuracy as a proxy for predictive confidence. *Federal Reserve Bank of New York Staff Reports*, 161.

Sachs, Lothar (2002). Angewandte Statistik, 11th ed., Berlin et al: Springer.

Stäglin, R. Beiträge der Wirtschaftsforschungsinstitute zur kurzfristigen Wirtschaftsbeobachtung. *Allgemeines Statistisches Archiv*, 82, 66-80.

Zarnowitz, V. (1992). *Business Cycles: Theory, History, Indicators, and Forecasting.* Chicago: University of Chicago Press.

Zarnowitz V (1999). Has the business cycle been abolished? *NBER Working Paper*, 6367.

Table 1: Forecasts of growth and inflation, accuracy[1] of the Institutes' proposals and of the JD, 1970 to 2007

		1970-2007					
		JD	A	B	C	D	E
GDP	Bias	-0.20	-0.10	-0.30	-0.40	-0.30	-0.30
	MAE	1.30	1.30	1.20	1.20	1.20	1.30
	RMSE	1.80	1.70	1.70	1.60	1.70	1.80
	RMSPE	292.00	301.00	373.00	347.00	369.00	330.00
	RMSPE/M	88.40	86.20	83.90	80.60	85.30	93.20
	Theils U	0.80	0.90	0.90	0.90	0.90	0.90
Inflation	Bias	0.00	0.10	0.10	0.10	0.00	-0.10
	MAE	0.70	0.70	0.80	0.80	0.70	0.70
	RMSE	0.90	0.90	1.00	1.00	0.90	1.00
	RMSPE	109.00	115.00	113.00	106.00	96.10	99.80
	RMSPE/M	31.50	31.70	33.50	34.90	30.80	31.70
	Theils U	0.80	1.00	0.80	0.90	1.00	0.80
		1970-1979					
GDP	Bias	-0.30	-0.70	-0.50	-0.60	-0.90	-0.50
	MAE	2.30	2.10	2.00	1.60	2.00	2.20
	RMSE	2.80	2.70	2.50	2.10	2.70	2.90
	RMSPE	213.00	280.00	255.00	148.00	280.00	287.00
	RMSPE/M	143.00	134.00	127.00	108.00	134.00	146.00
	Theils U	1.40	1.00	1.00	1.00	1.00	1.00
Inflation	Bias	0.40	0.50	0.60	0.70	0.20	0.40
	MAE	1.20	1.10	1.30	1.40	1.10	1.10
	RMSE	1.30	1.20	1.40	1.50	1.20	1.30
	RMSPE	30.80	28.80	33.60	32.40	29.70	32.10
	RMSPE/M	42.90	40.60	47.80	50.60	41.00	42.40
	Theils U	1.00	1.00	1.10	1.10	0.90	1.10
		1980-1989					
GDP	Bias	0.10	0.30	0.20	0.20	0.40	-0.10
	MAE	1.00	1.00	1.00	1.20	0.90	1.10
	RMSE	1.20	0.50	1.20	1.40	1.10	1.30
	RMSPE	82.80	70.50	74.50	127.00	82.00	125.00
	RMSPE/M	60.30	25.20	63.00	68.60	56.20	65.90
	Theils U	1.70	1.70	1.70	1.70	0.50	2.10

Table 1: cont'd

		1980-1989					
Inflation	Bias	-0.10	-0.10	-0.10	-0.10	0.00	-0.10
	MAE	1.00	1.00	1.00	1.10	1.00	1.00
	RMSE	1.10	1.00	1.10	1.20	0.90	1.20
	RMSPE	205.00	216.00	212.00	198.00	179.00	182.00
	RMSPE/M	36.70	33.30	36.70	40.00	30.00	40.00
	Theils U	0.80	1.00	0.80	0.90	1.10	0.80
		1990-1999					
GDP	Bias	-0.40	0.10	-0.40	-0.70	-0.40	-0.50
	MAE	0.80	0.90	0.90	1.00	0.70	0.90
	RMSE	1.00	1.20	1.20	1.30	1.00	1.20
	RMSPE	55.70	58.00	61.10	74.70	57.80	66.60
	RMSPE/M	51.80	61.60	58.20	68.00	52.70	61.80
	Theils U	0.60	0.60	0.60	0.80	0.60	0.70
Inflation	Bias	-0.30	-0.30	-0.30	-0.30	-0.30	-0.40
	MAE	0.40	0.40	0.40	0.40	0.50	0.50
	RMSE	0.50	0.50	0.50	0.40	0.50	0.60
	RMSPE	44.00	42.50	48.00	42.90	44.30	49.70
	RMSPE/M	16.30	17.40	16.30	14.90	18.20	19.90
	Theils U	1.00	0.90	1.10	0.90	1.10	1.20
		2000-2007					
GDP	Bias	-0.30	-0.30	-0.40	-0.60	-0.30	-0.30
	MAE	1.00	1.00	1.10	1.10	1.20	1.10
	RMSE	1.20	1.20	1.20	1.30	1.40	1.30
	RMSPE	579.00	567.00	753.00	719.00	733.00	624.00
	RMSPE/M	59.80	59.20	62.80	67.70	69.30	65.30
	Theils U	0.43	0.43	0.570	0.53	0.55	0.47
Inflation	Bias	-0.10	0.20	0.00	0.00	0.10	-0.10
	MAE	0.20	0.40	0.30	0.30	0.20	0.20
	RMSE	0.30	0.50	0.30	0.30	0.20	0.30
	RMSPE	24.40	37.40	20.50	26.20	16.00	21.40
	RMSPE/M	9.30	17.60	9.10	11.20	7.60	8.30
	Theils U	0.52	0.96	0.50	0.67	0.52	0.50

Authors' computations. Data from the five institutes participating at the Joint Diagnosis, the Joint Diagnosis and the Federal Statistical Office. – [1] Errors: actual ./. forecast. – MAE: Mean absolute error; RMSE: Root mean square error; RMSPE: Root mean square percentage error; RMSPE/M: Root mean square percentage error/Mean; Bias: Sum of absolute errors; Theils U: RMSE of forecast in relation to RMSE under naïve forecast (here: previous year's realization).

Table 2: Bias and Efficiency Tests of the Institutes' proposals and of the JD, 1970 to 2007[1]

		A	B	C	D	E	JD
		1970-2007					
Growth	Bias	0.919	0.759	0.507	0.444	0.409	0.553
	Efficiency	0.343	0.236	0.553	0.477	0.529	0.369
Inflation	Bias	0.667	0.788	0.988	0.689	0.446	0.706
	Efficiency	0.011	0.016	0.077	0.035	0.019	0.018
		1970-1979					
Growth	Bias	0.407	0.554	0.919	0.760	0.507	0.444
	Efficiency	0.529	0.363	0.343	0.236	0.636	0.477
Inflation	Bias	0.407	0.308	0.202	0.185	0.221	0.721
	Efficiency	0.107	0.044	0.044	0.058	0.236	0.044
		1980-1989					
Growth	Bias	0.541	0.683	0.838	0.192	0.918	0.838
	Efficiency	0.838	0.919	0.610	0.441	0.610	0.610
Inflation	Bias	0.878	0.683	0.759	0.919	0.919	0.838
	Efficiency	0.041	0.025	0.036	0.221	0.041	0.041
		1990-1999					
Growth	Bias	0.605	0.343	0.114	0.236	0.192	0.236
	Efficiency	0.610	0.234	0.760	0.834	0.441	0.441
Inflation	Bias	0.065	0.059	0.024	0.041	0.032	0.066
	Efficiency	0.838	0.944	0.726	0.683	0.888	0.838
		2000-2007					
Growth	Bias	0.441	0.352	0.363	0.527	0.624	0.447
	Efficiency	0.726	0.529	0.528	0.441	0.726	0.401
Inflation	Bias	0.326	0.944	1.000	0.526	0.551	0.944
	Efficiency	0.944	0.674	0.944	0.833	1.000	0.293

Authors' computations. Sources: See Table 1. For the tests see text. – [1] p-values.

Table 3: Diebold-Mariano Test Results
(p-values)

		Comparison with JD	
		H_0: JD = "Institute i" vs. JD better than "Institute i"	H_0: JD = "Institute i" vs. JD better than "Institute i"
		p-value	p-value
Growth	Institute A	0.054	0.946
	Institute B	0.036	0.964
	Institute C	0.365	0.635
	Institute D	0.044	0.956
	Institute E	0.061	0.939
Inflation	Institute A	0.443	0.557
	Institute B	0.119	0.881
	Institute C	0.119	0.881
	Institute D	0.630	0.370
	Institute E	0.275	0.725

Authors' computations. Sources: Table 1.

Table 4 **:** Results of the Friedman test, 1970 to 2007

Variable	criterion	Full Sample	1970-79	1980-89	1990-99	2000-07
Growth	Accuracy	0.74	1.39	1.55	2.51	1.51
	Style	3.41	6.05	3.95	9.71**	5.91
Inflation	Accuracy	0.70	2.16	2.21	4.22	3.91
	Style	1.04	3.71	2.34	2.73	2.98

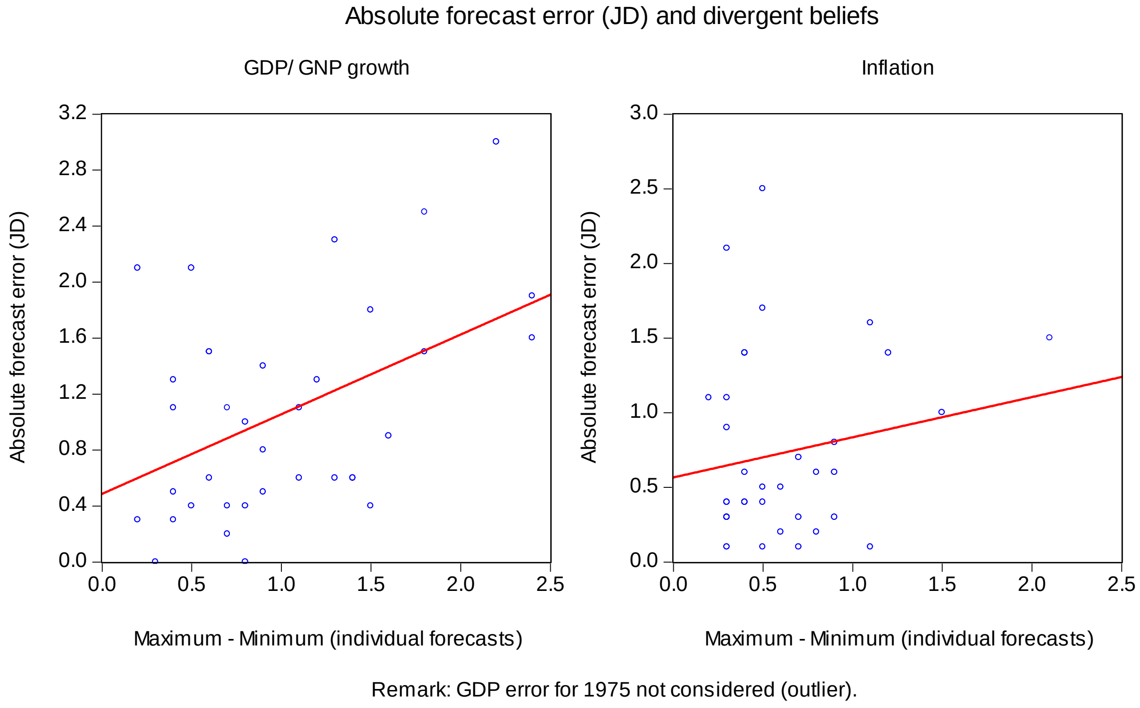
Authors' computations. Sources: See Table 1. ***,**,* indicate significance at 10, 5, 1 per cent level. Critical values for small samples are taken from Sachs (2002).

Table 5 **:** Comparison of JD with two synthetic JD's [1]

	Growth			Inflation		
	Best 2	Worst 2	JD	Best 2	Worst 2	JD
1980-1989	1.2	1.2	1.2	1.2	1.2	1.1
1990-1999	1.1	1.2	1	0.6	0.5	0.5
2000-2007	1.3	1.3	1.2	0.3	0.2	0.3

Authors' computations. Sources: See Table 1. „Synthetic JDs" as defined in the text. – [1] RMSE.

Figure 1: Forecast errors and disagreement range (1970-2007)



Authors' computations. Sources: See Table 1.

Table 6: Estimation results for asymmetry parameter α

		Full Sample 1972-2007											
		Estimate of α (using k instruments)											
		α (k=1)	s.e.	p-value	α (k=2)	s.e.	p-value	α (k=3)	s.e.	p-value	α (k=4)	s.e.	p-value
Growth	Institute A	0.472	0.083	0.738	0.472	0.083	0.732	0.469	0.083	0.713	0.467	0.083	0.691
	Institute B	0.472	0.083	0.738	0.469	0.083	0.713	0.471	0.083	0.731	0.467	0.083	0.693
	Institute C	0.611	0.081	0.171	0.615	0.081	0.158	0.620	0.081	0.138	0.627	0.081	0.116
	Institute D	0.528	0.083	0.738	0.528	0.083	0.738	0.531	0.083	0.713	0.531	0.083	0.705
	Institute E	0.556	0.083	0.502	0.563	0.083	0.449	0.557	0.083	0.493	0.566	0.083	0.423
	JD	0.528	0.083	0.738	0.530	0.083	0.717	0.529	0.083	0.730	0.532	0.083	0.701
Inflation	Institute A	0.556	0.083	0.502	0.588	0.082	0.282	0.556	0.083	0.500	0.591	0.082	0.265
	Institute B	0.583	0.082	0.310	0.619	0.081	0.143	0.583	0.082	0.310	0.625	0.081	0.121
	Institute C	0.611	0.081	0.171	0.630	0.080	0.105	0.624	0.081	0.126	0.639	0.080	0.083
	Institute D	0.556	0.083	0.502	0.573	0.082	0.377	0.556	0.083	0.500	0.573	0.082	0.375
	Institute E	0.583	0.082	0.310	0.611	0.081	0.173	0.584	0.082	0.308	0.612	0.081	0.170
	JD	0.611	0.081	0.171	0.643	0.080	0.073	0.611	0.081	0.171	0.650	0.079	0.058

Table 6: cont'd

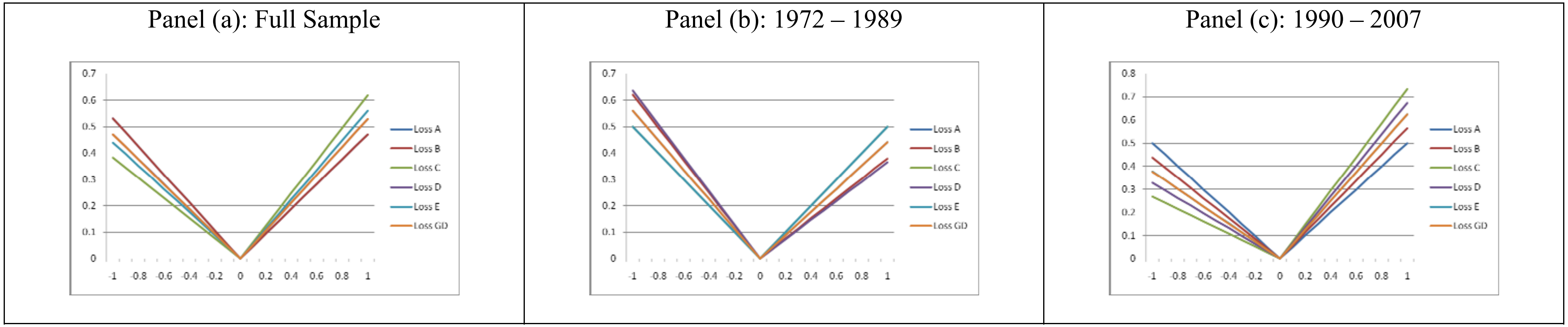
		Sample: 1972-1989											
		Estimate of α (using k instruments)											
		α (k=1)	s.e.	p-value	α (k=2)	s.e.	p-value	α (k=3)	s.e.	p-value	α (k=4)	s.e.	p-value
Growth	Institute A	0.444	0.117	0.635	0.442	0.117	0.620	0.440	0.117	0.609	0.435	0.117	0.576
	Institute B	0.389	0.115	0.334	0.383	0.115	0.308	0.378	0.114	0.286	0.367	0.114	0.242
	Institute C	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000
	Institute D	0.389	0.115	0.334	0.388	0.115	0.331	0.353	0.113	0.192	0.327	0.111	0.118
	Institute E	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000
	JD	0.444	0.117	0.635	0.440	0.117	0.606	0.441	0.117	0.617	0.435	0.117	0.579
Inflation	Institute A	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000
	Institute B	0.556	0.117	0.635	0.652	0.112	0.176	0.557	0.117	0.625	0.688	0.109	0.085
	Institute C	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000
	Institute D	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000
	Institute E	0.556	0.117	0.635	0.633	0.114	0.242	0.556	0.117	0.631	0.665	0.111	0.137
	JD	0.556	0.117	0.635	0.618	0.115	0.301	0.556	0.117	0.635	0.642	0.113	0.209

Table 6: cont'd

		Sample: 1990-2007											
		Estimate of α (using k instruments)											
		α (k=1)	s.e.	p-value	α (k=2)	s.e.	p-value	α (k=3)	s.e.	p-value	α (k=4)	s.e.	p-value
Growth	Institute A	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000	0.500	0.118	1.000
	Institute B	0.556	0.117	0.635	0.571	0.117	0.544	0.557	0.117	0.629	0.573	0.117	0.534
	Institute C	0.722	0.106	0.035	0.743	0.103	0.019	0.723	0.105	0.034	0.745	0.103	0.017
	Institute D	0.667	0.111	0.134	0.667	0.111	0.134	0.679	0.110	0.104	0.679	0.110	0.103
	Institute E	0.611	0.115	0.334	0.625	0.114	0.273	0.621	0.114	0.291	0.638	0.113	0.224
	JD	0.611	0.115	0.334	0.626	0.114	0.271	0.624	0.114	0.276	0.637	0.113	0.228
Inflation	Institute A	0.611	0.115	0.334	0.617	0.115	0.308	0.637	0.113	0.226	0.637	0.113	0.226
	Institute B	0.611	0.115	0.334	0.611	0.115	0.333	0.611	0.115	0.333	0.611	0.115	0.333
	Institute C	0.722	0.106	0.035	0.722	0.106	0.035	0.726	0.105	0.031	0.727	0.105	0.031
	Institute D	0.611	0.115	0.334	0.622	0.114	0.286	0.612	0.115	0.332	0.624	0.114	0.279
	Institute E	0.611	0.115	0.334	0.613	0.115	0.326	0.616	0.115	0.311	0.617	0.115	0.309
	JD	0.667	0.111	0.134	0.670	0.111	0.124	0.678	0.110	0.107	0.684	0.110	0.093

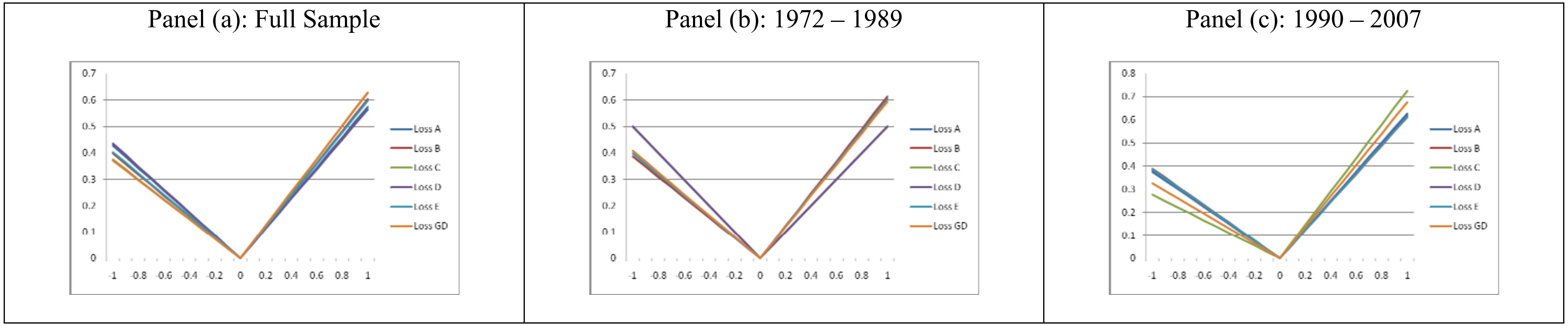
Authors' computations. Sources: See Table 1.

Figure 2: Implied loss functions for GDP/ GNP growth forecasts (1972-2007, 1972-1989, 1990-2007)



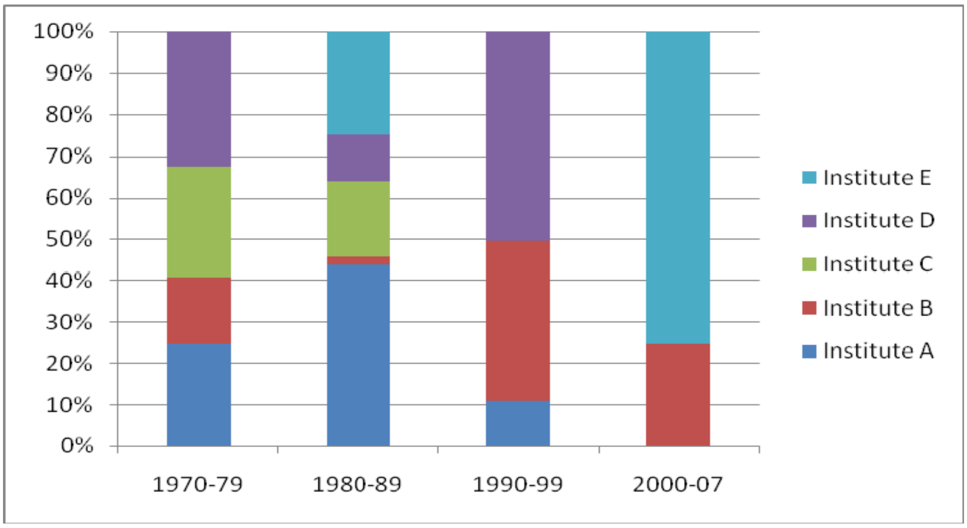
Authors' computations. Sources: See Table 1.

Figure 3: Implied loss functions inflation forecasts (1972-2007, 1972-1989, 1990-2007)

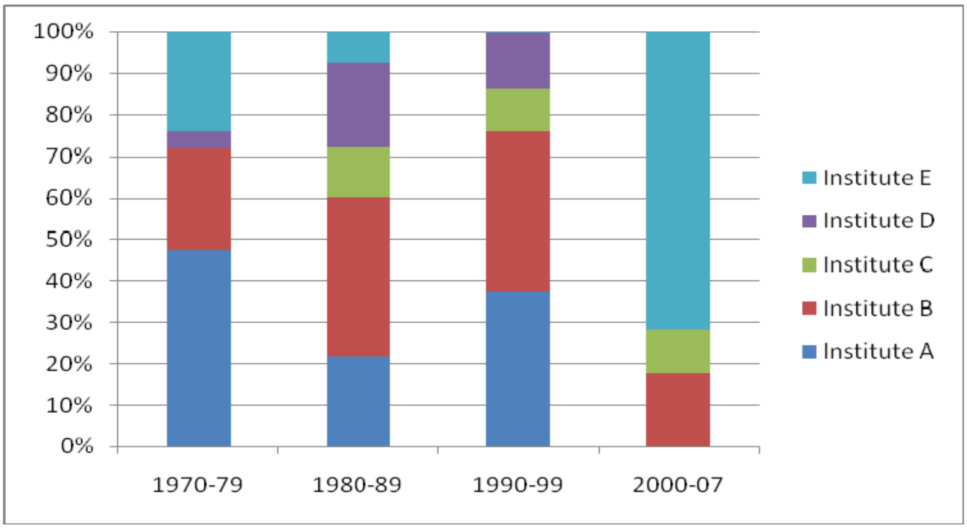


Authors' computations. Sources: See Table 1.

Figure 4: Forecast encompassing weights (sum scaled to 100 per cent), 1970 to 2007



Panel (a) Growth forecasts



Panel (b) Inflation forecasts

Authors' computations. Sources: See Table 1.

Appendix:

Table A 1a: Range of autumn growth forecasts, median JD outcome and actuals, 1970 to 2007

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year	Minimum	Maximum	Mean	JD	Reference	Abs. error mean	Abs. error JD	Difference (7) - (6)
1970	3.50	4.20	3.80		4.90	1.10		
1971	3.50	4.40	3.82	4.20	2.80	1.02	1.40	0.38
1972	0.70	3.10	1.38	1.30	2.90	1.52	1.60	0.08
1973	4.30	5.60	5.08	4.70	5.30	0.22	0.60	0.38
1974	2.10	3.90	3.42	2.90	0.40	3.02	2.50	-0.52
1975	2.20	2.60	2.42	2.30	-3.40	5.82	5.70	-0.12
1976	3.00	4.50	3.98	3.80	5.60	1.62	1.80	0.18
1977	3.90	6.10	5.28	5.40	2.40	2.88	3.00	0.12
1978	2.20	3.60	2.98	2.80	3.40	0.42	0.60	0.18
1979	2.90	4.30	3.60	3.80	4.40	0.80	0.60	-0.20
1980	1.60	2.70	2.30	2.40	1.80	0.50	0.60	0.10
1981	-0.90	0.60	-0.08	0.10	-0.30	0.22	0.40	0.18
1982	-1.10	1.30	0.56	0.80	-1.10	1.66	1.90	0.24
1983	-0.90	0.30	0.00	0.00	1.30	1.30	1.30	0.00
1984	1.60	2.50	1.94	2.10	2.60	0.66	0.50	-0.16
1985	1.50	2.30	1.90	2.00	2.40	0.50	0.40	-0.10
1986	2.70	3.20	2.86	2.80	2.40	0.46	0.40	-0.06
1987	2.40	3.50	2.88	2.80	1.70	1.18	1.10	-0.08
1988	1.70	2.30	2.02	1.90	3.40	1.38	1.50	0.12
1989	1.70	1.90	1.80	1.90	4.00	2.20	2.10	-0.10
1990	2.70	3.40	3.04	3.00	2.60	0.44	0.40	-0.04
1991	2.50	2.90	2.72	2.60	3.10	0.38	0.50	0.12
1992	1.50	1.90	1.70	1.80	1.50	0.20	0.30	0.10
1993	0.10	1.40	0.84	0.40	-1.90	2.74	2.30	-0.44
1994	-0.70	1.10	0.54	0.80	2.30	1.76	1.50	-0.26
1995	1.10	2.70	2.22	2.50	1.60	0.62	0.90	0.28
1996	1.90	2.70	2.34	2.40	1.40	0.94	1.00	0.06
1997	1.70	2.40	2.16	2.40	2.20	0.04	0.20	0.16
1998	2.30	3.10	2.80	2.80	2.80	0.00	0.00	0.00
1999	2.00	2.90	2.42	2.30	1.50	0.92	0.80	-0.12
2000	2.40	2.60	2.52	2.70	3.00	0.48	0.30	-0.18
2001	2.50	3.00	2.70	2.70	0.60	2.10	2.10	0.00
2002	1.50	1.90	1.64	1.30	0.20	1.44	1.10	-0.34
2003	1.30	1.90	1.64	1.40	-0.10	1.74	1.50	-0.24
2004	1.50	1.80	1.68	1.70	1.70	0.02	0.00	-0.02
2005	1.10	1.70	1.44	1.50	0.90	0.54	0.60	0.06
2006	1.00	1.40	1.20	1.20	2.50	1.30	1.30	0.00
2007	1.00	1.70	1.38	1.40	2.50	1.12	1.10	-0.02

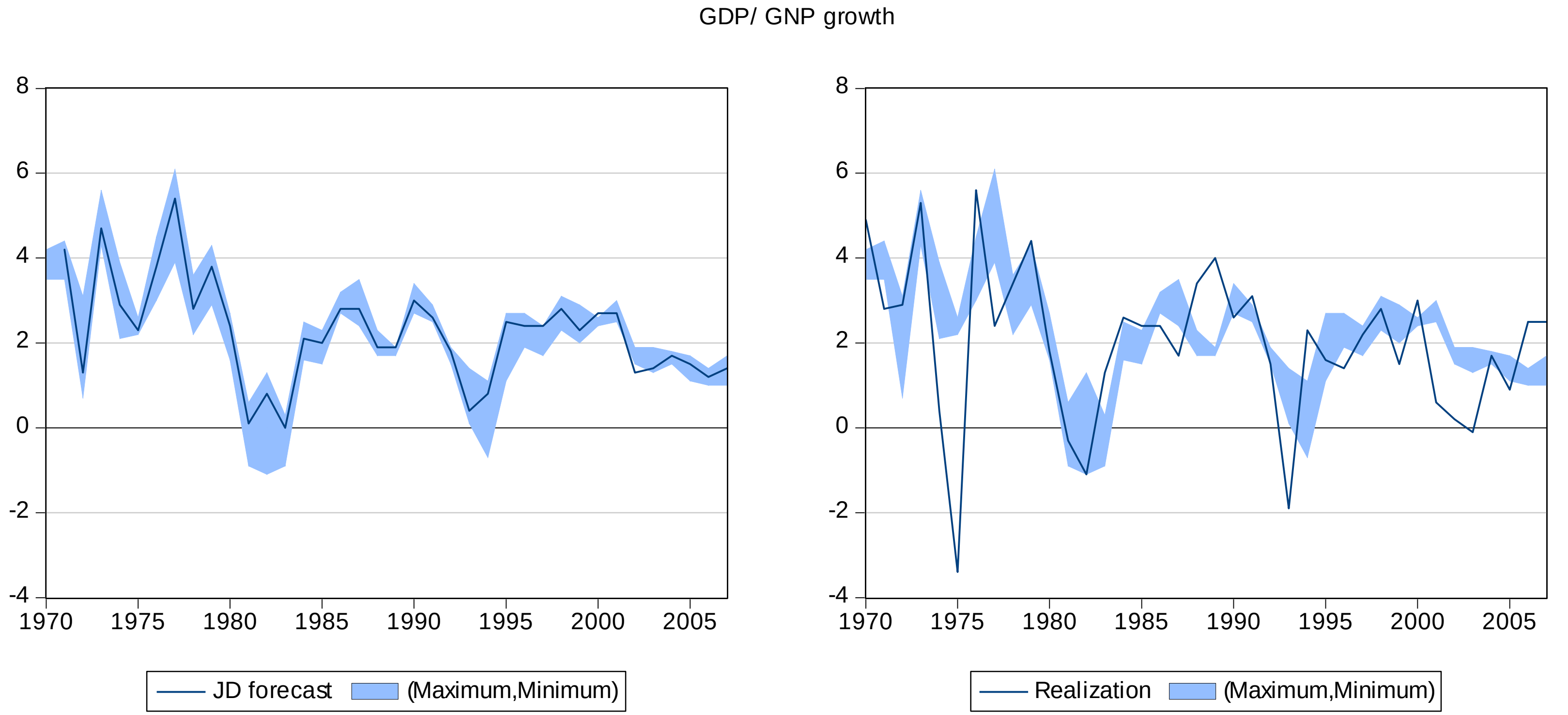
Authors' computations. Sources: See Table 1.

Table A 1b: Range of autumn inflation forecasts, median JD outcome and actuals, 1970 to 2007

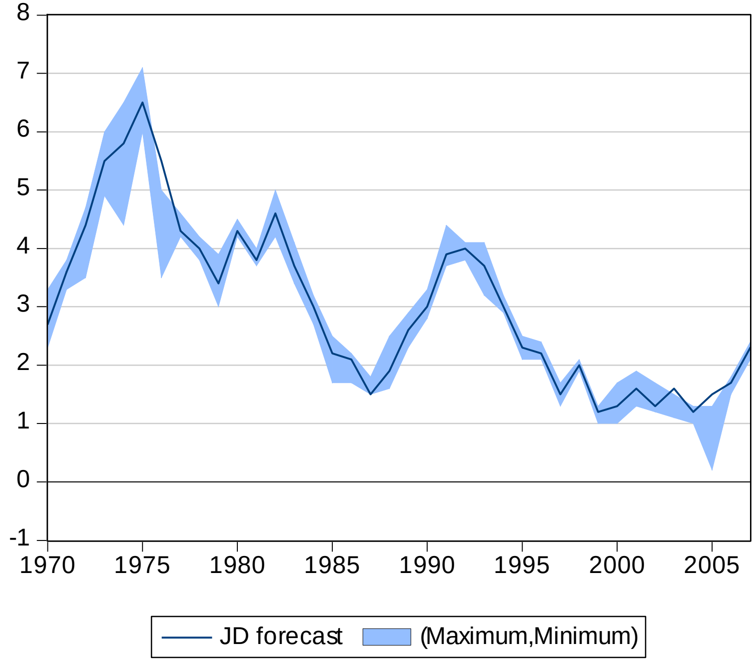
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year	Minimum	Maximum	Mean	JD	Reference	Abs. error mean	Abs. error JD	Difference (7) - (6)
1970	2.30	3.30	2.95	2.70	3.80	0.85	1.10	0.25
1971	3.30	3.80	3.56	3.60	5.30	1.74	1.70	-0.04
1972	3.50	4.70	4.20	4.40	5.80	1.60	1.40	-0.20
1973	4.90	6.00	5.40	5.50	7.10	1.70	1.60	-0.10
1974	4.40	6.50	5.78	5.80	7.30	1.52	1.50	-0.02
1975	6.00	7.10	6.60	6.50	6.10	0.50	0.40	-0.10
1976	3.50	5.00	4.62	5.50	4.50	0.12	1.00	0.88
1977	4.20	4.60	4.36	4.30	2.90	1.46	1.40	-0.06
1978	3.80	4.20	3.98	4.00	2.60	1.38	1.40	0.02
1979	3.00	3.90	3.40	3.40	4.20	0.80	0.80	0.00
1980	4.20	4.50	4.32	4.30	5.40	1.08	1.10	0.02
1981	3.70	4.00	3.80	3.80	5.90	2.10	2.10	0.00
1982	4.20	5.00	4.58	4.60	5.20	0.62	0.60	-0.02
1983	3.40	4.10	3.76	3.70	3.00	0.76	0.70	-0.06
1984	2.70	3.20	3.02	3.00	2.60	0.42	0.40	-0.02
1985	1.70	2.50	2.24	2.20	2.00	0.24	0.20	-0.04
1986	1.70	2.20	1.98	2.10	-0.40	2.38	2.50	0.12
1987	1.50	1.80	1.60	1.50	0.60	1.00	0.90	-0.10
1988	1.60	2.50	1.94	1.90	1.30	0.64	0.60	-0.04
1989	2.30	2.90	2.54	2.60	3.10	0.56	0.50	-0.06
1990	2.80	3.30	3.04	3.00	2.50	0.54	0.50	-0.04
1991	3.70	4.40	3.96	3.90	3.60	0.36	0.30	-0.06
1992	3.80	4.10	3.98	4.00	4.10	0.12	0.10	-0.02
1993	3.20	4.10	3.54	3.70	3.40	0.14	0.30	0.16
1994	2.90	3.20	3.00	3.00	2.60	0.40	0.40	0.00
1995	2.10	2.50	2.24	2.30	1.90	0.34	0.40	0.06
1996	2.10	2.40	2.24	2.20	1.90	0.34	0.30	-0.04
1997	1.30	1.70	1.54	1.50	1.90	0.36	0.40	0.04
1998	1.90	2.10	2.02	2.00	0.90	1.12	1.10	-0.02
1999	1.00	1.30	1.22	1.20	0.80	0.42	0.40	-0.02
2000	1.00	1.70	1.42	1.30	1.40	0.02	0.10	0.08
2001	1.30	1.90	1.64	1.60	1.80	0.16	0.20	0.04
2002	1.20	1.70	1.44	1.30	1.40	0.04	0.10	0.06
2003	1.10	1.50	1.34	1.60	1.00	0.34	0.60	0.26
2004	1.00	1.30	1.16	1.20	1.50	0.34	0.30	-0.04
2005	0.20	1.30	0.92	1.50	1.40	0.48	0.10	-0.38
2006	1.50	1.80	1.68	1.70	1.40	0.28	0.30	0.02
2007	2.10	2.40	2.24	2.30	2.20	0.04	0.10	0.06

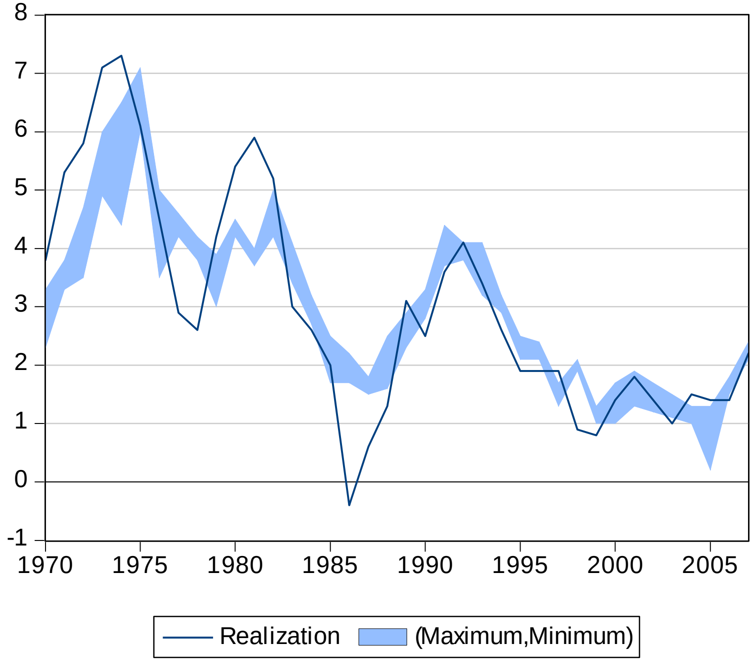
Authors' computations. Sources: See Table 1.

Figure A1: Range of Forecasts and JD (left), Range of Forecasts and Realization (Right), 1979 to 2007



Inflation





Authors' computations. Sources: See Table 1.