Kredit und Kapital, 38. Jahrgang, Heft 3
Seiten 377–399

Assessing Macroeconomic Forecast Uncertainty: An Application to the Risk of Deflation in Germany

By Dora Borbély and Carsten-Patrick Meier, Kiel

I. Introduction

Public perception treats macroeconomic forecasts as exact. It generally fails to acknowledge the uncertainty associated with them. The main reason is most likely that macroeconomic predictions are commonly given as point forecasts with no guidance of their likely accuracy. While point forecasts may sometimes be adequate, they should in general be supplemented by prediction intervals that represent the uncertainty of the forecast and allow taking into account alternative outcomes indicated by the interval (Chatfield 1993). In addition, the probabilities of certain events of particular interest to the user of the forecast may be stated explicitly, again to document the degree of imprecision of the prediction and to allow thinking in alternatives. As an example, a macroeconomic forecast may be supplemented by an explicit statement on the probability that a recession occurs over the forecast horizon (Fair 1993).

The uncertainty associated with a model-based forecast is partly inherent to the general uncertainty of future events, partly it arises from the estimated forecast model (Ericsson 2001). Clements and Hendry (1998, Chapter 7.3) distinguish five categories of model-based forecast errors: future changes in the underlying structure of the economy, misspecification of the model, mis-measurement of the data in the base period from which forecasting begins, inaccuracies in the estimates of the model's parameters, and the cumulation of future disturbances to the model. Although all five types of uncertainty are generally important, only the last two can be analysed quantitatively and are therefore themselves predictable. The first of these last two types captures the shocks that can occur to the economy given the model used for forecasting. The second is usually termed "parameter estimation uncertainty", resulting from estimating the model parameters on sample information instead of using the true population parameters.

However, current practice in macroeconomic forecasting suggests that "parameter estimation uncertainty" often entails more than the imprecision of the forecast arising from the actual estimation of the parameters. The broad selection of the forecast method (for instance univariate versus multivariate) and the variables used for forecasting (in a multivariate context) may well be guided by considerations that do not depend on the sample data at hand, e.g. by economic theory. Still, the "best-fitting" model that is ultimately employed for computing predictions is usually found by a data-based search procedure that possibly compares a large number of specifications. Given that the wrong model may be selected prior to parameter estimation and given that the costs, in terms of forecast accuracy, of model misspecification may be high, there is also model selection uncertainty in most real world forecasting problems (Chatfield 1996). Assessing the overall variability of a model-based macroeconomic forecast, thus, requires accounting for parameter estimation uncertainty as well as for the uncertainty that arises from data-based model selection.

The present paper works out a procedure that accounts for all three types of "predictable uncertainties" (Ericsson 2001) in macroeconomic forecasting. In the first step, a model selection procedure is defined that helps to choose the data-based specification of the macroeconomic model in a formalized and therefore replicable way. The procedure combines efficient use of information criteria to select subset models with additional tests for non-autocorrelation and for the presence of outliers to select the final model, thereby merging model selection approaches from multivariate time-series analysis, such as the general-to-specific principle, with procedures emphasized in univariate time-series literature such as the need to identify and model aberrant observations. Having defined a practical forecasting model guided by economic theory, this model selection procedure is used in the second step to specify the equations of model in detail. In the final step, point forecasts are generated from this model conditional on the last observations in the sample and a bootstrap is employed together with the formalized model selection procedure to estimate the associated conditional prediction densities that account for all three types of forecast uncertainties. In addition, event probabilities are estimated for selected outcomes of particular interest.

The macroeconomic question we apply our procedure to is that of estimating the risk of deflationary developments in Germany, a theme that figured prominently in public debate in Germany all over the year 2003.

German consumer price inflation has been lower than the euro area average ever since the start of the European Monetary Union. The reasons for these differences are mostly structural in nature, pertaining to a large extend to higher GDP growth in the rest of the euro area (Balassa-Samuelson effect). Over most of 2002, inflation largely came to a standstill while at the same time the real economy stagnated and was seen at the brink of recession by some commentators in early 2003. Against this background, there was widespread concern that with a weak economy and a monetary policy guided to a structurally higher euro area average inflation, there was a substantial risk of Germany drifting into a deflationary environment not unlikely to the situation Japan has been struggling to get out again since the late 1990s. As deflation may indeed be a self-enforcing process that is not easy to escape from once in, there is in general good reason for policy makers to avoid the economy slipping into such a situation. That is, he or she may want to minimize the risk of such a development occurring. Given this shape of the policy maker's 'loss-function', our model prediction of the risk of deflation can be the basis for deciding on the use of policy instruments to reduce that risk. Since there are various sensible definitions of deflation, pertaining to the strength and the length of the fall in the price level as well as to rest of the macroeconomic environment, we calculate different event probabilities for each of our definitions.

The remainder of the paper is organized as follows. Section 2 explains the model selection procedure we employ, outlines the general form of our forecasting model and gives estimation results. The bootstrap approach we use to estimate conditional forecast is portrayed, and the estimated predictions and prediction intervals are presented. Some conclusions are drawn in the final section.

II. The Empirical Approach

Our empirical approach to the problem of generating macroeconomic forecasts and forecast densities proceeds in three steps. First, we choose the economic variables of our forecast model on the basis of economic reasoning. Second, we employ a statistical model selection procedure to specify the model in detail, in particular its dynamic structure. Finally we generate point forecasts and forecast intervals using a simulation approach that accounts not only for the usual model uncertainty that stems from the model being only an approximation to reality but also for sampling uncertainty arising from parameter estimation and statistical model selection. Note that simulation of parameter estimation and model selection requires a complete formalization of all steps involved at this stage. The choice of the economic variables that enter the model, in contrast, is not formalized and simulated since it is not based on sample information.[1]

1. The Model Selection Procedure

Having specified the variables that enter the model, we have to select its empirical specification. Essentially this pertains to the question how the dynamics of the model's equations should be modelled. In addition, our procedure addresses the question of how to identify aberrant observations. As aberrant observations can distort model selection, but the identification of outliers also depends on having chosen the correct model, model selection and outlier detection proceed jointly in an iterative procedure.

a) Selecting the Lag Structure

A possible approach to select the lag structure of a model is to choose the model that minimizes some information criterion such as BIC or AIC. However, the computational effort of this approach is substantial because when there are N potential coefficients in the model, $2N$ models have to be compared. Hansen (1999) proposes a short-cut to the full evaluation of $2N$ models, consisting of eliminating all coefficients in question on the basis of their empirical t-ratios until only 10 remain; among these last 10 coefficients it is computationally feasible to conduct a full search procedure that allows to select the model that is optimal according to some information criterion. However, even this final full search can be substituted by a computationally more efficient sequential elimination based on the lowest t-ratios since, as is shown by Brüggemann and Lütkepohl (2001), this is equivalent to sequentially eliminating coefficients based on a model selection criterion provided that at each step of the elimination process a suitable threshold t-value is used. The latter de-

[1] Clearly, if the choice of variables used is also based on statistical arguments, this would have to be accounted for. A way to assess the uncertainty arising from such a 'data mining' process via bootstrapping has recently been proposed by *White* (2000).

pends on the model selection criterion chosen, the sample size and the number of regressors in the model.

Using this result, we select the dynamic structure of the equations of our model on the following information criterion-based procedure. As regards the choice of the information criterion, we follow Hansen (1999) and favour BIC before AIC. We then choose some maximal lag order, estimate the model with all lags included and sequentially eliminate the coefficient with the lowest t-ratio until some upper limit for the t-ratio is reached (which we fix at 5.0) and record the BIC for each of the models that appeared in reduction process. In principle, the model with the minimum is our preferred model.

However, we find that the BIC at times selects models that are parameterized too parsimoniously to capture the full dynamics of the underlying process, as can be seen from the serial correlation of the estimated residuals. Since the bootstrap procedure we use for constructing confidence intervals requires serially uncorrelated residuals and, moreover, since serial correlation may cause parameter estimation bias and thus poor forecast performance, we augment our criterion-based model selection with a test for autocorrelation. That is, we select the specification that minimizes the BIC among all specifications that passed tests for non-autocorrelated residuals up to the first, the fourth and the eighth order.[2]

b) Identifying Aberrant Observations

In addition to selecting the dynamic structure of the model, a decision has to be made on how to address the problem of aberrant observations. Commonly referred to as outliers, such observations are quite often encountered in empirical research, often simply due to the approximate nature of the econometric model (Franses and Lucas 1998, Krasker et al. 1983). Their treatment is, however, still quite controversial in multivariate time series analysis. In univariate time-series analysis and in cross-section analysis, in contrast, there exists a large literature on the effects, the identification and the treatment of aberrant observations.[3]

[2] *Kilian* (2001) in a simulation study also finds the BIC to select models that are too tightly parameterized to account for higher order dynamics. He advocates using the AIC instead of the BIC. He does not consider combining information criteria and tests for autocorrelation, though.

[3] See *Krasker* and *Welsh* (1983) for a survey, and *Franses* (1998) and *Maddala* (1992) for textbook discussions.

This literature shows that if not accounted for in a suitable way, outliers may severely distort the model selection process and cause biases estimates of model parameters and confidence intervals. Practical time series applications that require good model identification and forecasting performance, such as procedures for seasonal adjustment, commonly employ routines to detect outliers to guard against their detrimental effects[4]. Consequently, we attempt to identify aberrant observations in our model selection process and neutralize their influence.

Procedures proposed in the literature to identify outliers commonly rely on assessing the influence of a particular observation by dropping it from the sample and checking whether the change in the model's fit at that observation is large. The first of the two approaches we use is to sequentially search the sample for a 'studentized residual', which is the empirical t-ratio of a dummy that takes the value 1 at the respective observation and 0 otherwise.[5] Following the strategy proposed by Chen and Lui (1993) for univariate models, we calculate the absolute studentized residual for all possible data points, choose the date where it reaches its maximum as the location of a potential outlier and use some critical value to decide on its statistical significance. In case the null hypothesis of no outlier is rejected, the respective observation is modelled by an impulse dummy and we repeat our search – until no more outlier is found. As critical value we chose to take 2.7, following the simulation evidence in Chen and Lui (1993).[6]

The iterative procedure just described will generally identify so-called innovation outliers in the dependent variable of an equation. To identify observations in the regressor set that are outside the majority of the observed data in the context of the model (Krasker et al. 1983, p. 661) a measure referred to as dfits has been proposed which like the studentized residual is a measure of the difference the fitted value of the dependent variable due to dropping the observation in question (see Belsley, Kuh and Welsch (1980) and Maddala (1992), Chapter 12). We follow Krasker et al. (1983) in using a critical value of $3\sqrt{p/T}$, where p is the number of parameters in the model and T the number of observations, to

[4] As an example, see *Findley* et al. (1998) for a describtion of the procedure used in the X-12-ARIMA routine of Bureau of Census.

[5] See *Franses* (1998, Chapter 6) and *Maddala* (1992, Chapter 12) for textbook expositions.

[6] *Chen* and *Lui* (1993) found for sample sizes up to $T = 100$ critical values between 2.5 and 3.0 to work well in terms of finding the correct number of aberrant observations.

identify a significant outlier. Again, we model an identified aberrant observation by an impulse dummy.

Since aberrant observations can distort model selection, but the identification of outliers also depends on having chosen the correct model, model selection and outlier detection have to proceed jointly in an iterative procedure (Chen and Lui 1993). We start our model selection process with an unrestricted equation containing all regressors up to the prespecified maximum lag order. We subject this equation to our two outlier detection tests and model identified outliers by impulse dummies. The resulting, possibly dummy-augmented, equation is simplified by eliminating coefficients on the basis of the lowest t-ratios and selecting the specification the one with the lowest BIC among all specifications with serially independent residuals. The resulting specification is then used for a second round of outlier detection tests. In case no more aberrant observations are encountered, the specification found in the previous step is the preferred specification, otherwise the equation is augmented by additional impulse dummies and the BIC-based simplification procedure starts again from the most general specification until the final specification is found. This specification is used for forecasting.

2. *Estimating Prediction Intervals*

Given the paramter estimates of our simple five-variable system of equations, it is straightforward to calculate point forecasts for the variables of interest such as GDP, the output gap and, notably, the rate of inflation. These point forecasts are, however, uncertain. As regards the case of deflation this means that even though the model's point forecast may be that there is no fall in the price level in the forecast period, we may not be able to rule out this possibility completely since the model's forecasts may simply not precise enough to allow this.

a) The Bootstrap

To assess the uncertainty associated with the forecasts of our model, we estimate prediction intervalls using a bootstrap technique. The general idea of this simulation method is to measure the variability of an estimate obtained from some statistical procedure by applying the procedure repeatedly to an artifical data set that is constructed by resampling from the original observations (Efron 1979). In times series analysis, the resampling is implemented by randomly reordering the time series of estimated residuals of a regression equation, resulting in artifical data with the same stochastic structure as the original sample.[7] In case the statistical procedure to be evaluated is a forecasting model, evaluation of the variability of the generated forecast commonly proceeds as follows[8]: estimate the model on the basis of artifical data, then generate forecasts using the estimated model in conjunction with artificial 'future' disturbances resampled from the original residuals. Repeat the procedure a number of times and use the resulting distribution of forecasts to approximate the distribution of the real forecast.

The specific procedure we employ differs from the method just explained only in that we not only re-estimate given specifications of the model's equations on the artificial data at each replication, but apply our complete model selection process to the data prior to estimation, such that it is possible that in each replication a different specification is found and used for parameter estimation and forecasting. This way our procedure accounts for the forecast variability arising from selecting the model specification from sample information. Failure to take model selection uncertainty into account would result in estimated confidence bands that were too narrow and therefore underestimated the true risks associated with a forecast.

b) Applying the Bootstrap

Starting point of our procedure is the vector of the estimated residuals of the k equations of the model, $\hat{\varepsilon}_t = \{\hat{\varepsilon}_{1t}, \ldots, \hat{\varepsilon}_{kt}\}$, $t = 1, \ldots, T$, which by virtue of our model selection procedure is independently and identically distributed. We center the residuals and follow the convention to rescale all residuals as proposed by Stine (1987). To generate a bootstrap replicate sample of our data set, we randomly draw (with replacement) T times from $\hat{\varepsilon}_t$, giving us the vector of artificial residuals $\hat{\varepsilon}_t^*, t = 1, \ldots, T$ and then calculate the equations recursively by substituting the empirical residuals $\hat{\varepsilon}_t$ by their bootstrap counterparts $\hat{\varepsilon}_t^*$, using some original sample data as starting values. Note that since we draw the residuals $\hat{\varepsilon}_t$

7 This residuals-based procedure is termed a nonparametric bootstrap. See *Horrowitz* (2001) for a recent general survey on bootstrap methods and *Berkowitz* and *Kilian* (2000) for a survey on the time-series aspects. A non-technical introduction is provided by *Brownstone* and *Valetta* (2001).

8 See *Clements* and *Taylor* (2001) for a recent survey on generating forecast densities using bootstrap methods.

in tandem, the contemporaneous correlation between them is preseved in the artifical data set. Next, the dynamics of model's equations are specified using our model selecting procedure and the parameters are estimated on the artificial data, giving a new set of parameter estimates whose variability over the various repetitions accounts for model specification and parameter estimation uncertainty. Finally, these estimated parameters are used to generate the forecast in conjunction with a set new set of 'future' disturbances drawn from the empirical residuals.

More formally, our bootstrap procedure works as follows[9]. Let the AR(p) process $z_t = \varphi_1 z_{t-1} + \ldots + \varphi_p z_{t-p} + \varepsilon_t$, where ε is a sequence of iid random disturbances, represent one of the equations of our model. We estimate the equation on the original sample data giving us the vector of coefficient estimates $\hat{\varphi}$ and the empirical residuals $\hat{\varepsilon}_t$. We then generate a bootstrap sample series z_t^* by drawing T artifical disturbances ε_t^* from the dummy-adjusted, centered and rescaled empirical residuals and recursively calculating $z_t^* = \hat{\varphi}_1 z_{t-1}^* + \ldots + \hat{\varphi}_p z_{t-p}^* + \varepsilon_t^*$ for $t = p+1, \ldots, T$ using the first p original sample observation as starting values. On this artificial series z_t^* we apply our model selection procedure which identifies some autoregressive model with parameter vector $\hat{\varphi}^* = \left\{ \hat{\varphi}_1^*, \hat{\varphi}_2^*, \ldots, \hat{\varphi}_{p^*}^* \right\}$, where p^* may or may not coincide with p. To generate a forecast for the next h periods, we draw from $\hat{\varepsilon}_t$ h artificial future disturbances ε_t^* for $t = T+1, \ldots, T+h$ and using this we recursively calculate $z_{t+h}^* = \hat{\varphi}_1^* z_{t+h-1}^* + \ldots + \hat{\varphi}_1^* z_{t+h-p}^* + \varepsilon_t^*$. Repeating this procedure B times gives an empirical distribution for z_{t+h}^* which is the bootstrap approximation of the unknown forecast distribution. The quantiles of this distribution define the upper and lower confidence band around the point forecast. For instance, for $B = 1000$, the upper and lower values of a 95% confidence band are found by taking the 975^{th} and 25^{th} element of the vector of the decreasingly ordered realisations of z_{t+h}^*.

Complications to this standard procedure arise when impulse dummies are used in the original equations to model aberrant observations. The impulse dummies are used to reduce the biases in model selection and parameter and confidence band estimation associated with aberrant observations. However, since the estimated residual at the observations modelled by a dummy is zero, resampling from that residual series will in general underestimate the true uncertainty of the forecast.[10] To ensure

---

9 See also *Clements* and *Taylor* (2001).

10 The exception is when the use of the dummy variable is motivated by structural economic information rather than statistical testing and the event that trig-

the residual series has the variability it would have without the dummies, we replace all estimated zero residuals by the coefficient estimates of the associated dummy variables and, in turn, exclude the impulse dummies from the process used to generate the artifical data. The modified residual series is then centered and rescaled and the procedure works as explained above.

c) Conditional Forecast Intervals

So far, the procedure estimates prediction intervals conditional on the estimated parameters and the artificial data used to estimate these parameters. The intervalls are therefore close to being unconditional. We are interested, however, in a forecast interval that is like our point forecast conditioned on the last (p) original observations of our sample. Thombs and Schucany (1990) propose obtaining conditional forecast intervals using parameter estimates based on artifical data from $t = 1, \ldots T - p$ that is constructed by backcasting taking the last p original observations of the sample as starting values.

However, Pascual et al. (2001) show that simply conditioning on the past p observations of the original sample instead of the artificial data in the procedure explained above gives asymptotically the same results. The idea is that for large B, the convergence of the bootstrap parameter estimate $\hat{\varphi}^*$ to the true parameter φ is independent of the artificial data being conditioned on the last p observations, see Clements and Taylor (2001). Therefore, the parameters can be estimated on a separate data set, which makes backcasting unnecessary. We therefore rely on this approach to obtain conditional forecast intervals.

III. Assessing the Risk of Deflation in Germany

According to a common defintion, deflation is a process of falling prices. This may harm economic growth mainly via two channels (Newman et al. 1992). First, it can be due to a 'Fisher effect': If nominal interest rates do not fall sufficiently to make up for the (expected) fall in prices, the real interest rate rises and deters investment. Second, if nom-

gered the observation to deviate from the rest of the sample can be excluded to occur again in the forecast period. An example for such a case in our model is German unification, which causes a break in our real GDP series in 1991 that is modelled with an impulse dummy.

inal wages are downwardly sticky while deflation occurs, real wages will increase and cause employment to fall. In both cases, the problem is that once the economy is in a deflationary situation, the following fall in demand causes prices to fall even further, so a reinforcing process, often termed a deflationary spiral, may unfold. The greatest cyclical deflation occured in the United States between 1929 and 1933, where the fall in the price level reached eight percent per annum over a four year period. Conventional economic wisdom has it that deflation was a major cause of the Great Depression that occured at that time. More recently, Japan has been experiencing deflation for a number of years, coupled with poor performance of the real economy. There is therefore good reason for policy makers to be concerned about the risk of deflation.

From economic grounds, it may, however, be debatable whether a small negative "dip" of the consumer price index or even a somewhat more sustained fall in the price level at low rates (say 0.5 or 1.0 percent) will already trigger a deflationary spiral. Also, a fall in the price level caused by external influences such as a sharp drop in the world price of crude oil would generally not be associated with a self-enforcing deflationary environment since it would not be expected to be permanent and would moreover increase rather than decrease firm's profit expectations. In our empirical application we therefore estimate event probabilities for alternative definitions of deflationary developments and differentiate between changes in consumer prices caused by oil price fluctuations and other changes.

1. *The Empirical Model*

We are now ready to start specifying our macroeconomic forecasting model. In the simplest case, such a model this could be a set of autoregressive equations, more ambitious are vector autoregressive (VAR) or a dynamic simultaneous equations models. Our model belongs to the latter class, despite its simple recursive structure. In setting up the model, we try to strike a balance between good forecast performance and economic interpretability. The forecasts of the model should not only be accurate but also reasonable from an economic point of view. This desired feature guided the choice of variables included in the model and also precluded relying completely on vector autoregressions. We therefore decide to model the inflation process in terms of a modified Phillips curve relationship and to give survey-based economic sentiment indicators a prominent role in forecasting economic activity.

a) Theoretical Considerations

The Phillips Curve is widely regarded as a central tool for forecasting inflation. The conventional Phillips Curve specifies inflation as negatively dependent on cyclical unemployment expressed as the deviation of unemployment from its natural rate. In this paper we use an alternative specification, which describes the positive relationship between inflation rate and the degree of capacity utilization in an economy. We use the output gap – measured by the deviation of real GDP from its potential value – as a dimension of the degree of capital utilisation. This specifiaction is commonly used, since the output gap and the unemployment gap show directly opposed movements in an economy. We find, that the output gap-based Phillips curve provides better forecasts with smaller mean squared errors than the unemployment-based.[11] The Phillips curve specification used in this model is

$$\pi_t = \alpha(y - \bar{y}) + \varepsilon$$

where π denotes the percentage change of consumer prices and $(y - \bar{y})$ denotes the output gap. Alternative ways of forecasting inflation would be to use e.g. interest rate differentials or long-run growth rate of monetary aggregates in the Phillips curve.

To make the Phillips curve operational for forecasting inflation, we need an estimate of the parameter α and a forecast of the output gap $(y - \bar{y})_t$ over the forecast horizon. The parameter α can be estimated from a dynamic model for π_t which includes as regressors lags of $(y - \bar{y})_t$ and lags of π_t. The particular procedure we use to select the lag structure of the model is explained in the previous section in greater detail. We augment the empirial specification of this equation by the contemporanous lagged rates of the rate change of the euro price of crude oil (UK Brent), because much of the short-run dynamics of inflation in Germany are associated with changes in energy prices. The euro price of crude oil itself is modelled as a simple autoregressive process.

Next, need to endogenize $(y - \bar{y})_t$. To do so, we need forecasts for real GDP and for potential output. To forecast GDP we employ survey-based sentiment indicators. The advantage of these indicators is that they have a close correlation with GDP, generally with a lead of one or two quarters, and therefore a proven ability to forecast relatively accurately. We

11 Our findings correspond to those found by *Stock* and *Watson* (1999).

specify a vector autoregressive (VAR) model for the trend-adjusted logarithm of real GDP, the ifo-business climate – a survey-based index of business expectations and sentiments which is the most reliable business cycle indicator in Germany –, and the ISM index, which is the most important indicator for industrial production in the United States. To obtain a forecast for potential output, $\bar{y}_t$, we apply the Hodrick-Prescott filter (Hodrick and Prescott 1980) on our VAR-forecast for real GDP[12].

b) The Equations

In the following we present the results of employing the model selection procedure outlined above to the five equations we specified on theoretic grounds. We employ quartely data for Germany ranging from 1970 : 1 to 2002 : 3 to select the specification and to obtain estimates of the parameters of the equations. The data refers to West-Germany before 1991 and to unified Germany thereafter. In constructing the German data series breaks due to unification have been avoided by using growth rates. In the following we present the main equations of the model. All variables are estimated by OLS, all equations are free from autocorrelation up to the eights order. We start with our oil-price augmented Phillips curve equation, were application of the above procedure yields (t-values for H_0: $\beta_k = 0$ given in parantheses)[13]

$$\begin{aligned}\Delta p_t = {} & \underset{(3.35)}{0.07(y-\bar{y})_{t-1}} + \underset{(6.52)}{0.005 s_{1t}} - \underset{(-3.15)}{0.002 s_{3t}} + \underset{(5.00)}{0.28\Delta p_{t-1}} \\ & + \underset{(3.37)}{0.18\Delta p_{t-2}} + \underset{(6.61)}{0.04\Delta p_{t-4}} + \underset{(7.53)}{0.01\Delta p_t^{oil}} + \hat{\varepsilon}_{1t}\end{aligned} \tag{1}$$

$$\bar{R}^2 = 0.78 \quad T = 114 \quad JB : 0.72$$

where p stands for logarithm of the German index of consumer prices, $(y-\bar{y})$ for the output gap estimated using the Hodrick-Prescott Filter, p^{oil} is the price of cruide oil (UK Brent), denominated in euros. Δ denotes first differences and s_{1t} and s_{3t} are seasonal dummy variables. In

12 To reduce the instability of the filter at the end of the forecast horizon, the Hodrick-Prescott-Filter is calculated on a sample of forecasts for real GDP that reaches 12 quarters beyond our forecast horizon for the output gap, following *Baxter* and *King* (1995).

13 All estimations and simulations were performed using RATS 5.1.

addition, our procedure identified three aberrant observations that are modeled by 0/1-Dummy-Variables in the equation for 74 : 1, 91 : 3 and 97 : 3, the coefficient estimates of which have been suppressed when presenting the equation. Adjusted R^2 indicates quite good fit for a regression in changes, while the marginal significance level of the Jarque-Bera test (JB) shows that the residuals of the equation are normally distributed.[14]

Our Phillips curve forecasts inflation conditional on values for the output gap and the change in oil prices. To obtain a forecast for the output gap, we need to specify an equation for real GDP. This raises the question whether real GDP should be modelled as trend-stationary or difference stationary. Applying the ADF-test on a time-trend augmented autoregressive equation selected by our model selection procedure, we can reject the null hypothesis of non-stationarity in favor of the trend-stationary model. In the following, we therefore model real GDP as trend-stationary. Next, we include a variable measuring the so-called Business Climate which is a survey based measure of the assessment of the current economic situtation and the prospects six months ahead produced monthly by the ifo Institute in Munich. Since the Business Climate is clearly stationary, it may enter the equation for real GDP both levels and in first differences, giving the following error-correction model:

$$
\begin{aligned}
\Delta y_t = {} & \underset{(2.44)}{0.37} - \underset{(-2.84)}{0.07 y_{t-1}} + \underset{(2.40)}{0.001 t} + \underset{(7.02)}{0.27 I^G_{t-1}} - \underset{(-3.69)}{0.26 \Delta y_{t-1}} \\
& + \underset{(2.54)}{0.16 \Delta y_{t-4}} + \underset{(7.29)}{0.14 \Delta I^G_t} - \underset{(-2.97)}{0.05 \Delta I^G_{t-4}} + \hat{\varepsilon}_{2t}
\end{aligned} \tag{2}
$$

$$\bar{R}^2 = 0.64 \quad T = 107 \quad JB : 0.89$$

where I^G stands for the Business Climate indicator. It has a significant influence on real GDP both in levels and in differences. Besides we use some dummies for the following quarters: 74:4, 76:4, 79:2, 84:3; 87:1, 89:1, 91:1, 91.3 and 92:1.

To forecast, in turn, the Business Climate, we specify a bivariate autoregressiv equation for the German business climate by including a vari-

14 Note that the Phillips curve implies that consumer prices are best modeled by a unit root process. This assumption is important when simulating the prediction intervals, since only under this assumption bootstrapping the residuals from equation (1) gives the correct precdiction variances.

able measuring the US business climate, which is the ISM index provided by the US national association of purchasing managers. In employing this additional variable, we account for the interdependencies between the German and the US business cycles.

$$\begin{aligned} \Delta I_t^G = {} & \underset{(5.41)}{13.45} - \underset{(-5.66)}{0.15} I_{t-1}^G + \underset{(6.67)}{2.51} s_{1t} + \underset{(9.74)}{0.57} \Delta I_{t-1}^G + \underset{(4.28)}{0.25} \Delta I_{t-3}^G \\ & + \underset{(2.51)}{0.09} \Delta I_{t-8}^{US} + \hat{\varepsilon}_{3t} \end{aligned} \tag{3}$$

$$\bar{R}^2 = 0.70 \quad T = 104 \quad JB: 0.10$$

As expected the German business climate indicator enters into the regression with a negative sign in levels and a positiv sign in differences. This underlines the typical business cylce movements. The positive sign of the US business climate indicator is also obvious. It indicates a positive correlation, a comovement, between the business cycles in the two respective countries. Included 0/1-dummy variables are: 73 : 3, 73 : 4, 74 : 1, 82 : 3, 83 : 1, 84 : 2, 92 : 4, 95 : 1 and 99 : 3.

The US business climate indicator also has to be endogenized for the model. The regression yields the following.

$$\begin{aligned} \Delta I_t^{US} = {} & \underset{(3.51)}{8.17} - \underset{(-3.58)}{0.15} I_{t-1}^{US} + \underset{(7.22)}{0.45} \Delta I_{t-1}^{US} - \underset{(-4.88)}{0.32} \Delta I_{t-4}^{US} \\ & - \underset{(-3.80)}{0.23} \Delta I_{t-8}^{US} - \underset{(-2.38)}{0.13} \Delta I_{t-13}^{US} + \hat{\varepsilon}_{4t} \end{aligned} \tag{4}$$

$$\bar{R}^2 = 0.67 \quad T = 102 \quad JB: 0.98$$

Included 0/1-dummy variables are: 73:4, 74:4, 80:2, 80:3, 81:1, 83:1, 91:4, and 2002:1.

Finally, we specify an equation for the oil price, which is simply an autoregressive model specified in first differences:

$$\Delta p_t^{oil} = \underset{(3.51)}{0.22} \Delta p_{t-1}^{oil} - \underset{(-3.58)}{0.14} \Delta p_{t-5}^{oil} + \hat{\varepsilon}_{5t} \tag{5}$$

$$\bar{R}^2 = 0.08 \quad T = 110 \quad JB: 0.00$$

According to the expectations, the fit of the equation is quite poor.

2. Results

In the following we present the results of applying our empirical approach to estimate forecast uncertainty to the question of assessing the risk of deflation in Germany as of late 2002. We estimate the above model with data running up to 2002 : 3 and generate conditional forecasts for the rate of change of the consumer price index starting in 2002 : 4. We then apply our bootstrap approach, using 5000 replications in each simulation, to estimate the corresponding conditional density of the forecast. The result is a probability statement on the development of consumer prices over the forecast horizon that implies a quantification of the risk of the consumer price level falling by a certain extend.

a) Forecast Intervals

The main findings of our analysis are summarized in Figure 1. Reading from left to right the figure presents the model forecasts for the business climate indicator, the output gap, the percentage annual rate of growth of real GDP $\left((y_t - y_{t-4}) \cdot 100\right)$ and the percentage annual change in consumer prices $\left((p_t - p_{t-4}) \cdot 100\right)$ over a horizon of seventeen quarters. Evidently, the model predicts a rise in business confidence and a subsequent increase of real GDP growth that peaks in the middle of 2004. The output gap, which is negative at the start of our forecasting period, will be closed by then. Given this forecast, for real activity, it is obvious that the mean forecast of the model does not imply the change in consumer prices to fall below the zero line in the forecast period.

This forecast is, however, uncertain, as indicated by the prediction interval estimated for the forecast. The forecast intervals become quite wide after very few forecast steps, indicating the limitations of our model for predicting future events. For instance, as regards the forecast for GDP growth, the upper prediction interval for the first forecast step is already as high as 1.9 percentage points and the lower interval is –1.6 percentage points. After four forecast steps, these figures have increased (in absolute value) to 3.2 and –3.4 percentage points, respectively. For forecasting the rate of inflation, the intervals are somewhat lower, lying at 0.9/–0.8 percentage points for the first step and 2.4/–2.3 for the fourth. The lower bound of the 95 percent prediction interval for the inflation forecast, thus, very quickly slides below the zero line. At the 5 % significance level, deflation can not be ruled out for the year of 2003 and the following years.

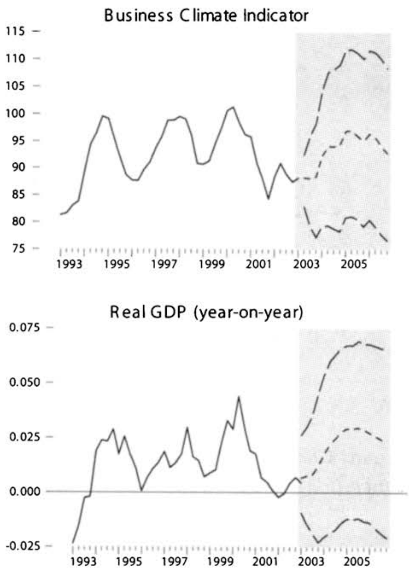

Figure 1: Model Forecasts with 95% Prediction Intervals 2003–2006

To gain an insight into the probability distribution of the forecast change in consumer prices, Figure 2 displayes the prediction intervals for inflation associated with different levels of significance. The outermost interval shows the results with allowing for 5% probability of error, followed by 10%, 25% and the most tight interval corresponds to a 50% probability of error. The 5% and 10% intervals reach the aera with negative growth rate before the end of 2003. At the 25% level, deflation can nearly be ruled out over the forecast horizon; the associated interval falls slightly below the zero by the start of 2004 and remains quite close to it until the end of the forecast horizon. Deflation can completly be expelled only for the 50% level of significance.

b) Event Probabilities

As indicated in the theoretical motivation above, not every fall in the price level may be damaging for real economic activity. We therefore need to clarify the definition of what is regarded as a harmful deflationary development. In order not have to rely on one specific definition, we use various alternatives and observe them under varying circumstances.

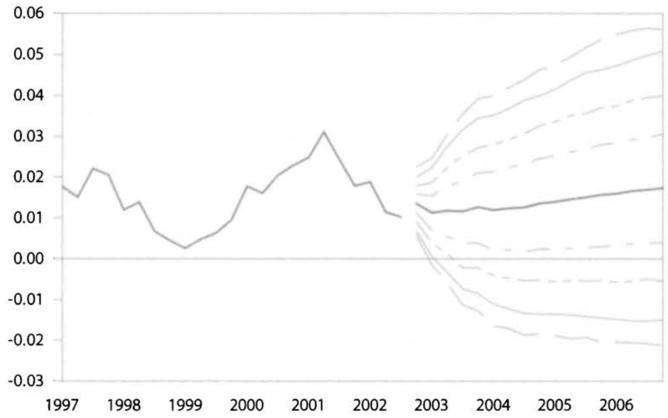

Figure 2: Prediction Intervals for the Annual Change in Consumer Prices for Alternative Levels of Confidence (50, 75, 90 and 95%)

First, it may be argued that a very small fall of the consumer price level may not cause much damage, at least not more than a small increase. Since there may be different views on what is small in this context, we use three definitions according to which the economy is in a deflation when the change in consumer prices lies either below 0.0, below -0.5 percent or below -1 percent. Second, a deflation may be a period of sustained decreases of the price level. Thus, we require negative growth of consumer price to hold on for at least 2 or 3 quarters in a row, respectively. In addition, we look at three forecast horizons: until the end of the year 2003, 2004 or 2005.

Combining those definitons results in eighteen different dimensions of what may be regarded as a deflationary development. To assess the likelihood of these definitions to appear over the forecast period, we estimate event probabilities in the sense of Fair (1993). That is, when running the replications, we record in each draw whether or not the specified event has occured. The probalitiy of the event is then simply the number of times it occured devided by the number of replications. Table 1 provides an overview on how often deflation falling in one of the defined categories has been predicted by the model, using again 5000 replications of our bootstrap procedure.

The probability of a deflation until the end of 2003 defined as two quarters of negative change of consumer prices accounts for 14.9 per-

Table 1

Probability of a Fall in the Consumer Price Index

	< 0.0 %	< –0.5 %	< –1.0 %
	until the end of 2003		
2 quarters in a row	14.9	7.5	3.4
3 quarters in a row	10.0	4.2	1.6
	until the end of 2004		
2 quarters in a row	25.7	15.5	8.9
3 quarters in a row	20.8	11.8	6.6
	until the end of 2005		
2 quarters in a row	33.1	21.4	13.4
3 quarters in a row	27.5	17.4	10.3

cent, whereas it only accounts for 1.6 percent if you define deflation as the percentage change of consumer prices lying below –1.0 percent three quarters in a row. The probability of deflation increases with a longer time horizont, since uncertainty of the model prediction increases accordingly. Considering the time horizont as almost two years, the probability of negative inflation rates over two periods rise up to 25.7 percent. Even for the most "lax" definition of deflation, the probability that such a development occurs until the end of 2005 still lies over 10 percent, whereas in one third of the replications the change of consumer prices lies below zero in two subsequent periods.

Still, these results may not be useful for assessing the economic risks associated with deflation. Recall, that the circumstances which cause consumer prices to fall determine whether it is harmful or not. If for example deflation is due to high productivity gains in production, as it has been experienced by the information technology sector in the last decades, falling prices are a normal reaction of the market, which are in this case even advantageous. Lower production costs translate into lower prices which stimulate demand for IT-products. Alternatively, negative price changes might be traced back to falling raw material costs, e.g. oil prices, which again is not a harmful but a natural reaction of the

Table 2

Probabilities Assuming Constant Oil Prices

	< 0.0 %	< –0.5 %	< –1.0 %
	until the end of 2003		
2 quarters in a row	11.3	3.9	1.2
3 quarters in a row	7.0	1.9	0.4
	until the end of 2004		
2 quarters in a row	20.6	9.9	4.7
3 quarters in a row	15.9	7.2	3.0
	until the end of 2005		
2 quarters in a row	26.5	14.4	7.3
3 quarters in a row	21.7	11.4	5.6

market. Since oil prices are extremely volatile, it is useful to find out the probability of deflation excluding the changes in consumer prices caused by oil price changes. To do so, we run our simulation procedure under the assumption that the oil price remains constant over the forecast period. Our model easily allows for the modification since oil price changes appear directly in the equation determing the inflation rate. In the following scenario, consumer prices are, thus, no longer influenced by the volatility of the oil price and forecast intervals should become narrower. This should tend to result in a decrease of the probability of deflation.

According to table 2 the percentage share of those simulations which forecast deflation is indeed lower than in the previous cases. Until the end of 2003 e. g. the probability of deflation ranges now only between 0.4 and 11.3 percent instead of 1.6 to 14.9 percent in the case of including the changes of the oil price. Controlling for the volatility of the oil price means that any deflationary or inflationary tendency might be attributed to business cycle movements.

IV. Conclusions

This paper has proposed a method for assessing the risks associated with model-based macroeconomic forecasts. It was argued, that estimated forecast intervals should account not only for the uncertainty arising from the model being an approximation to reality and from the model's parameters being estimated. Instead, they should also account for the uncertainty arising from selecting the very specification of the model from the sample data. To allow for model selection uncertainty to be considered systematically, we formalize a model selection procedure that specifies the lag structure of a model and also accounts for aberrant observations. The procedure can be used to bootstrap the complete model selection process when estimating forcast intervals.

In our application, we estimate the risk of deflationary developments occuring in Germany over a specified forecast horizon. The forecast intervals estimated using the outlined procedure implied the risk of a fall in the price level to be nonnegligible. We then examined alternative economically sensible definitions of a deflationionary development, bearing in mind that deflation may not always be harmful to the real economy. Among other things, we find the risk of a deflation occuring together with a recession, a development that reminds at the Great Depression in the United States or the recent Japanese experience, is very small in the period analysed.

References

Baxter, M. and *King,* R. (1995): Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. National Bureau of Economic Research Working Paper No. 5022. Cambridge Mass. – *Belsley,* D. A., *Kuh,* E. and *Welsch,* R. E. (1980): Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, New York. – *Berkowitz,* J. and *Kilian,* L. (2000): Recent Developments in Bootstrapping Time Series. Econometric Reviews 19, 1–48. – *Brownstone,* D. and *Valetta,* R. (2001): The Boostrap and Multiple Imputations: Harnessing Increased Computing Power for Improved Statistical Tests. Journal of Economic Perspectives 15 (4), 129–141. – *Brüggemann,* R. and *Lütkepohl,* H. (2001): Lag Selection in Subset VAR Models with an Application to a U.S. Monetary System. In: Econometric Studies: A Festschrift for Joachim Frohn, Münster, pp. 107–128. – *Chen,* C. and *Lui,* L.-M. (1993): Joint Estimation of Model Parameters and Outlier Effects in Time Series. Journal of the American Statistical Association 88, 284–297. – *Chatfield,* C. (1996): Model Uncertainty and Forecast Accuracy. Journal of Forecasting 15, 495–508. – *Chatfield,* C. (1993): Calculation Interval Forecasts. Journal of Business & Economic Statistics, Vol. 11, No. 2, 121–144. – *Clements,* M.

P. and *Taylor,* N. (2001): Bootstrapping Prediction Intervals For Autoregressive Models. International Journal of Forecasting 17, 247–267. – *Clements,* M. and *Hendry,* D. F. (1998): Forecasting Economic Time Series. Cambridge U. K. – *Efron,* B. (1979): Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics 7, 1–26. – *Ericsson,* N. R. (2001): Forecast Uncertainty in Economic Modelling. Board of Governors of the Federal Reserve System. International Finance Discussion Papers No. 697. Washington D. C. – *Fair,* R. C. (1993): Estimation Event Probabilities From Macroeconomic Models Using Stochastic Simulation. In: Stock, J. H. (ed.) Business Cycles, Indicators, and Forecasting. 157–176. – *Findley,* D. F., *Monsell,* B. C., *Bell,* W. R., *Otto,* M. C. and *Chen,* B.-C. (1998): New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. Journal of Business and Economic Statistics 16, 127–152. – *Franses,* P. H. and *Lucas,* A. (1998): Outlier Analysis in Cointegration Analysis. Journal of Business and Economic Statistics 16, 459–468. – *Franses,* P. H. (1998): Time Series Models for Business and Economic Forecasting. Cambridge U.K. – *Hansen,* B. E. (1999): Discussion of 'Data Mining Reconsidered', Econometrics Journal 2, 192–201. – *Hodrick,* R. and *Prescott,* E. (1980): Post-War U.S. Business Cycles: An Empirical Investigation. Carnegie-Mellon University Working Paper. – *Horrowitz,* J. L. (2001): The Bootstrap. In: E. E. Leamer and J. J. Heckman (eds.), Handbook of Econometrics, Vol. V, Amsterdam, Chapter 52, 3159–3228. – *Kilian,* L. (2001): Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order. Journal of Forecasting 20, 161–179. – *Krasker,* W. S., *Kuh,* E., *Welsch,* R. (1983): Estimation for Dirty Data and Flawed Models. In: Z. Griliches and M. D. Intriligator (eds.), Handbook of Econometrics, Vol. I, Amsterdam, 652–698. – *Maddala,* G. S. (1992): Introduction to Econometrics. Second Edition, New York. – *Newman,* P., *Milgate,* M. and *Eatwell,* J. (ed.) (1992): The New Palgarave Dictionary of Money & Finance. The Macmillan Press Limited, London and Basingstoke. – *Pascual,* L., *Romo,* J. and *Ruiz,* E. (2001): Effects of Parameter Estimation on Prediction Densities: A Bootstrap Approach. International Journal of Forecasting 17, 83–103. – *Stine,* R. (1987): Estimating Properties of Autoregressive Forecasts. Journal of the American Statistical Association 82, 1072–1078. – *Stock,* H. James and *Watson,* W. Mark (1999). Forecasting Inflation. Journal of Monetary Economics 44, 293–335. – *Thombs,* L. A. and *Schucany,* W. R. (1990): Bootstrap Prediction Intervals for Autoregression. Journal of American Statistical Association 85, No.410, 486–492. *White,* H. (2000): A Reality Check for Data Snooping. Econometrica 68, No. 5, 1097–1126.

Summary

Assessing Macroeconomic Forecast Uncertainty: An Application to the Risk of Deflation in Germany

This paper proposes an approach for estimating the uncertainty associated with model-based macroeconomic forecasts. We argue that estimated forecast intervals should account for the uncertainty arising from selecting the specification of an empirical forecasting model from the sample data. To allow this uncertainty to be considered systematically, we formalize a model selection procedure that specifies the lag structure of a model and accounts for aberrant observations. The procedure can be used to bootstrap the complete model selection process when estimating forecast intervals. We apply the procedure to generating forecasts and forecast intervals for the change in the consumer price index in Germany, with special emphasis on assessing the risk of deflationary developments. (JEL C5, E0, E5)

Zusammenfassung

Zum Abschätzen der Unsicherheit gesamtwirtschaftlicher Prognosen: Eine Anwendung auf das Risiko einer Deflation in Deutschland

Der vorliegende Beitrag schlägt einen Ansatz zum Schätzen der Unsicherheit vor, die mit modellbasierten gesamtwirtschaftlichen Prognosen verbunden ist. Unserer Ansicht nach sollte dabei die Unsicherheit berücksichtigt werden, die sich aus der Wahl der Modellspezifikation anhand der Stichprobendaten ergibt. Um diese Unsicherheit systematisch zu berücksichtigen, formalisieren wir eine Modellauswahlroutine, die die Struktur der Verzögerungen eines empirischen Modells bestimmt und etwaige Ausreißer modelliert. Wir verwenden diese, um den gesamten empirischen Modellauswahlprozess im Rahmen eines Bootstrap-Ansatzes stochastisch zu modellieren. Als Anwendung dienen uns Prognosen für den Preisindex der Lebenshaltung in Deutschland, unter spezieller Berücksichtigung der Frage, ob es im Prognosezeitraum zu deflationären Tendenzen kommen wird.

Résumé

Evaluation de l'incertitude des prévisions macroéconomiques: Une application sur le risque d'une déflation en Allemagne

Les auteurs proposent ici une approche en vue d'estimer l'incertitude associée aux prédictions macroéconomiques basées sur des modèles. Selon eux, il faut prendre en compte l'incertitude qui provient du choix de la spécification d'un modèle de prévision à partir de données d'échantillon. Pour prendre en compte cette incertitude de manière systématique, les auteurs formalisent une procédure de sélection de modèle qui spécifie la structure des retards d'un modèle empirique et tient compte d'observations aberrantes. Les auteurs utilisent cette procédure pour modéliser tout le processus empirique de sélection du modèle dans le cadre d'une approche de Bootstrap. Ils appliquent cette procédure pour prédire le changement de l'indice des prix à la consommation en Allemagne, en évaluant tout particulièrement le risque de développements déflationnistes.