DEUTSCHE BUNDESBANK
EUROSYSTEM

Evaluating macroeconomic risk forecasts

Malte Knüppel
Guido Schultefrankenfeld

Discussion Paper
Series 1: Economic Studies
No 14/2011

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0
Telex within Germany 41227, telex from abroad 414431

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978-3–86558–722–0 (Printversion)
ISBN 978-3–86558–723–7 (Internetversion)

Abstract:

Macroeconomic risk assessments play an important role in the forecasts of many institutions. A risk forecast is related to the potential asymmetry of the forecast density. In this work, we investigate how the optimality of such risk forecasts can be tested. We find that the Pearson mode skewness outperforms the standard third-moment-based skewness as a measure of asymmetry. We consider problems of the tests likely to be encountered in practice and try to offer remedies where possible. In general, tests for macroeconomic risk forecast optimality tend to have at best moderate power given the empirically available small sample sizes.

Keywords: Forecast evaluation; asymmetric densities, skewness

JEL-Classification: E37, C12, C53

Non-technical summary

Many central banks supplement their macroeconomic forecasts with an assessment of future risks, where a risk implies an asymmetry of the forecast density. Put differently, the presence of an upward risk commonly implies that outturns greater than the point forecast have a probability of more than 50%. Accordingly, the presence of a downward risk means that outturns less than the point forecast have a probability of more than 50%.

In this work, we investigate how macroeconomic risk forecasts can be evaluated. A typical problem of these evaluations is the small sample size of currently available risk forecasts. As a start, we find that an evaluation should not be based on the standard measure of asymmetry for probability density functions (the skewness), because an alternative measure, the Pearson mode skewness, allows a markedly more precise evaluation in small samples.

We also consider several problems to be encountered in practice and try to offer remedies where possible. For example, if the width of the forecast density, i.e. the dispersion of the forecast errors is systematically over- or underestimated, a certain test for the optimality of risk forecasts tends to yield misleading results. However, a modification of this test leads to correct results.

A general conclusion of the investigations presented is that tests for the optimality of macroeconomic risk forecasts cannot be expected to be very powerful given the mostly very small sample sizes which are currently available. So, it is not unlikely that no statistical evidence against the optimality of risk forecasts can be found, even if optimality is not present, i.e. if the asymmetry of the forecast density is not predicted correctly.

Nicht-technische Zusammenfassung

Viele Zentralbanken versehen ihre makroökonomischen Prognosen mit Einschätzungen über zukünftige Risiken, wobei ein Risiko eine Asymmetrie der Prognosedichte impliziert. Das heißt, dass das Vorliegen eines Aufwärtsrisikos das Auftreten von Werten oberhalb der Punktprognose mit einer Wahrscheinlichkeit von mehr als 50% bedeutet. Entsprechend liegt ein Abwärtsrisiko vor, wenn das Auftreten von Werten unterhalb der Punktprognose eine Wahrscheinlichkeit von mehr als 50% besitzt.

In der vorliegenden Arbeit wird geprüft, wie solche makroökonomischen Risikoprognosen evaluiert werden können. Ein typisches Problem solcher Evaluationen ist durch die kleinen Stichprobenumfänge der derzeit verfügbaren Risikoprognosen gegeben. Es zeigt sich zunächst, dass eine Evaluation von Risikoprognosen nicht auf Grundlage des Standardmaßes für die Asymmetrie einer Wahrscheinlichkeitsdichte (der Schiefe) erfolgen sollte, da ein alternatives Maß, die Pearsonsche Modusschiefe, in kleinen Stichproben eine deutlich genauere Evaluation ermöglicht.

Viele der in der Praxis auftretenden Probleme bei der Evaluation von Risikoprognosen werden untersucht und, soweit möglich, Lösungsvorschläge unterbreitet. So kann zum Beispiel das Problem beobachtet werden, dass ein bestimmter Test für die Optimalität von Risikoprognosen zu falschen Ergebnissen führt, falls die Breite der Prognosedichte, also die Streuung der zukünftigen Prognosefehler systematisch über- oder unterschätzt wird. Eine Modifikation des Tests sorgt jedoch dafür, dass er wieder korrekte Ergebnisse liefern kann.

Aus den durchgeführten Untersuchungen kann die Schlussfolgerung gezogen werden, dass von den Optimalitätstests für makroökonomische Risikoprognosen wegen der derzeit verfügbaren, häufig nur kleinen Stichprobenumfänge keine besonders hohe Güte erwartet werden kann. Das bedeutet, dass es nicht unwahrscheinlich ist, dass keine statistische Evidenz gegen die Optimalität einer Risikoprognose gefunden wird, selbst wenn diese Optimalität nicht vorliegt, wenn also die Asymmetrie der Prognosedichte nicht korrekt vorhergesagt wird.

Contents

Evaluating Macroeconomic Risk Forecasts[1]

1 Introduction

Many central banks and other institutions such as the Bank of England or the IMF complement their macroeconomic forecasts by assessments of risks. In general, these risk forecasts are related to possible asymmetries of forecast densities, as described in Knüppel & Schultefrankenfeld (2011). In Figure 1, two examples of forecast densities implying upside or downside risks to the forecast can be found.[2] Both densities are asymmetric, so that mean and mode do not coincide.

As noted by Leeper (2003), it would actually be important to verify whether macroeconomic risk assessments contain valuable information. Since the forecast of the central tendency is often a mode forecast[3], realizations above [below] the central tendency should be observed more often in the presence of an upward [downward] risk. If it turns out that this is not the case, the risk forecasts are not informative.

In this work, we attempt to identify methods which are suitable for the evalu-

[1] Authors: Malte Knüppel and Guido Schultefrankenfeld, Deutsche Bundesbank, Research Centre, Wilhelm-Epstein-Straße 14, D-60431 Frankfurt am Main, Germany. Corresponding author: malte.knueppel@bundesbank.de. The authors would like to thank Heinz Herrmann and Karl-Heinz Tödter for very helpful comments and discussions. This paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank.

[2] To be more precise, a risk is usually defined as an important event which, in case of its occurrence, would lead to a change of the central forecast. For example, the risk of an oil supply plunge would constitute an upward risk to the inflation forecasts. Of course, there might be several different risks to a forecast. If the upward [downward] risks dominate, it is often said that the balance of risks is tilted to the upside [downside]. Instead of using the expression "balance of risks", in this work we simply speak of upward or downward risks.

[3] This fact might be surprising, but many central banks and other institutions indeed explicitly state that their published point forecasts refer to the mode of the forecast density or, equivalently, to the most likely outcome. See Knüppel & Schultefrankenfeld (2011) for a survey.

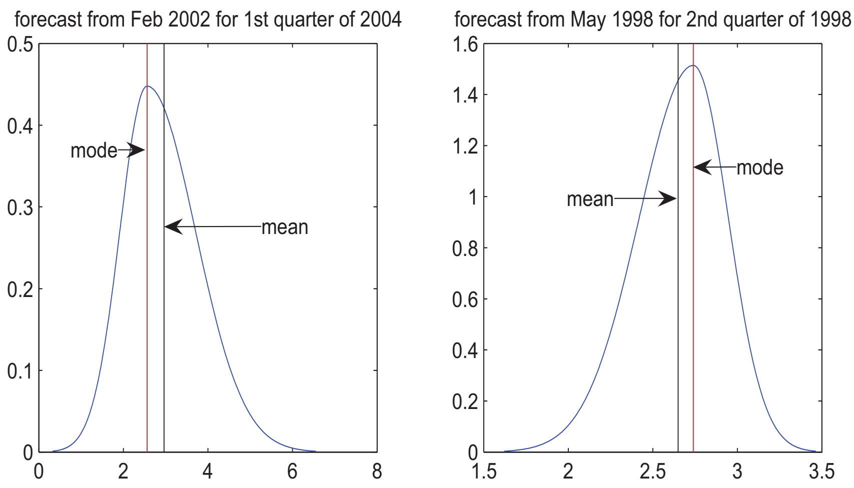

Figure 1: Two of the Bank of England's density forecasts for inflation. The forecast in the left panel implies an upward risk (mean > mode), the one in the right panel a downward risk (mean < mode).

ation of risk forecasts, focusing on tests for risk forecast optimality. Risk forecasts can be regarded as point forecasts, because the asymmetry of a density is usually summarized by a single parameter. While there is a vast literature concerning the evaluation of point forecasts, we are not aware of any work which focuses on the specific nature of macroeconomic risk forecast evaluation.

Macroeconomic risk forecast evaluation has to deal with several issues. First, a choice has to be made concerning the parameter which is supposed to measure the asymmetry of the density forecast. Also, it has to be considered how the corresponding parameter reflecting the asymmetry of the *true* forecast density could be measured. Moreover, evaluations of macroeconomic risk forecasts typically have to deal with the small sample sizes of risk forecasts. In addition, the risks that are forecast are usually not very large, either. Finally, many institutions do not quantify the risk, i.e. the asymmetry of the forecast density, but only give the direction of the risk. That is, these institutions state that there is an upward or downward risk to their forecasts, or that the risks are balanced, but they do not reveal the magnitude of the forecast risks. Due to small samples and only moderate asymmetries of the forecast densities, the power of tests for risk forecast optimality can be expected to be rather low, amplifying the importance of power considerations. Therefore, this study focuses on a careful analysis of the power properties of these tests.

In what follows, we are concerned with the partial optimality of forecasts, where partial optimality is defined as in Diebold & Lopez (1996). As mentioned by Diebold & Lopez (1996), the original concept of partial optimality refers to optimality conditional on the information set being used by the forecaster and goes back to Brown & Maital (1981).[4] When we speak of optimality in this study, we always mean partial optimality with respect to the information set that is given by the independent variable(s) of a certain regression.

2 How to Measure the Forecast Risk?

In principle, the optimality of quantitative risk forecasts can be investigated in a similar way as the optimality of conditional mean forecasts. That is, in the regression

$$\tilde{r} = \alpha + \beta \hat{r} + \varepsilon, \tag{1}$$

where $\hat{r}$ is the risk forecast and $\tilde{r}$ is the realized risk, the joint null hypothesis $\alpha = 0, \beta = 1$ can be tested. If the risk forecasts are not optimal, the null hypothesis

[4] Therefore, it would actually seem more plausible to speak of conditional optimality instead of partial optimality.

should be rejected.

The major question here is how to measure risk. Since, in macroeconomic forecasting, the balance of risks is related to the skewness of the forecast density, measures of skewness are natural candidates for the measurement of risk. In what follows, we consider the third standardized moment (henceforth standard skewness) and the Pearson mode skewness of the density forecast. While the standard skewness measure is more familiar to statisticians, the Pearson mode skewness is more directly related to many macroeconomic forecasts by institutions which quantify risks in terms of the mean-mode difference. Interestingly, as described in Knüppel & Schultefrankenfeld (2011), these institution mostly focus on the mode as measure of the central tendency of their forecasts. Both measures of skewness are standardized. This is an advantage if the volatility of the forecast variable changes over time, or if the risk forecasts for different variables or different forecast horizons are to be analyzed simultaneously, for example in a panel study.

Suppose that the parameters of the forecast density are known. Denote the expectation of the variable of interest by μ and the corresponding mode by m. Then, using standard skewness as the measure of risk, the forecast risk $\hat{r}$ for the random variable Y, henceforth denoted by $\hat{\tau}$, is given by

$$\hat{r} = \hat{\tau} \equiv \frac{E\left[(Y-\mu)^3\right]}{\sigma^3},$$

where σ denotes the standard deviation of Y, the variable to be forecast. The realized risk $\tilde{r}$ of observation y[5], henceforth denoted by $\tilde{\tau}$, is then given by

$$\tilde{r} = \tilde{\tau} \equiv \frac{(y-\mu)^3}{\sigma^3}.$$

Employing the definition of the Pearson mode skewness gives the forecast risk $\hat{r}$, henceforth denoted by $\hat{\phi}$,

$$\hat{r} = \hat{\phi} \equiv \frac{E[Y] - m}{\sigma} = \frac{\mu - m}{\sigma},$$

and the corresponding realized risk of observation y, henceforth denoted by $\tilde{\phi}$,

$$\tilde{r} = \tilde{\phi} \equiv \frac{y - m}{\sigma}.$$

Both measures of realized risk depend on a location parameter, just as the usual measure of the realized forecast error which would be given by $(y - \mu)$. In addition, both measures depend on the standard deviation of the forecast variable.

A choice of the risk measure to be employed might be based on the power properties of tests for the optimality of risk forecasts in small samples. We attempt to find out which risk measure implies a larger power of the tests. To this end, we conduct a Monte Carlo study using the two-piece normal distribution (henceforth tpn distribution) as described, among others, by Wallis (2004). We use this type of distribution because it is employed by several institutions which publish macroeconomic risk forecasts, like the Bank of England or the International Monetary

[5] We apply the usual notation where upper case letters denote random variables and the corresponding lower case letters denote their realizations.

Fund.[6] The density of a tpn-distributed variable y is given by

$$f(y) = \begin{cases} A \exp\left(-\frac{(y-m)^2}{2\sigma_1^2}\right) & \text{if } y \leq m \\ A \exp\left(-\frac{(y-m)^2}{2\sigma_2^2}\right) & \text{if } y \geq m, \end{cases}$$

with $A = \frac{2}{\sqrt{2\pi}(\sigma_1+\sigma_2)}$. Its expectation μ, variance σ^2, and third central moment equal

$$\begin{aligned} E[y] &= \mu = m + \sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1) \\ E\left[(y-\mu)^2\right] &= \sigma^2 = \left(1 - \frac{2}{\pi}\right)(\sigma_2 - \sigma_1)^2 + \sigma_1\sigma_2 \\ E\left[(y-\mu)^3\right] &= \sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1)\left(\left(\frac{4}{\pi} - 1\right)(\sigma_2 - \sigma_1)^2 + \sigma_1\sigma_2\right). \end{aligned} \tag{2}$$

In what follows, without loss of generality, we will restrict the analyses to the case of a zero mode, i.e. to the case

$$m = 0.$$

We consider tpn-distributed variables which, in addition, fulfill the restrictions

$$\begin{aligned} \sigma^2 &= 1 \\ \sigma_2 &= \theta\sigma_1 \end{aligned} \tag{3}$$

where θ is varied from 1 to 4.[7] A standard normal distribution is obtained when θ equals 1, and the more θ differs from 1, the more asymmetric is the resulting

[6] The densities displayed in Figure 1 are densities of the two-piece normal distribution. Yet, there are of course many other types of distribution that could also be used. Classes of potentially asymmetric distributions can be found inter alia in Ramberg & Schmeiser (1974), Todd C. Headrick & Sheng (2008) or Knüppel & Tödter (2007).

[7] In principle, we could also consider values $0 \leq \theta \leq 1$.

density.

For each θ, we simulate N tpn-distributed variables and estimate the parameters β_τ and β_ϕ in the equations

$$\begin{aligned} \tilde{\tau}_i &= \beta_\tau \hat{\tau}_i + \varepsilon_{\tau,i} \\ \tilde{\phi}_i &= \beta_\phi \hat{\phi}_i + \varepsilon_{\phi,i} \end{aligned} \tag{4}$$

for $i = 1, 2, \ldots, N$ by OLS where $\varepsilon_{\tau,i}$ and $\varepsilon_{\phi,i}$ are iid and have an expectation of zero. So, in contrast to equation (1), we assume that α equals zero, which simplifies the following study.

While θ is varied from 1 to 4 with increments of 0.25, we assume that the risk forecaster forecasts $\theta = 2$ for all N variables. Using (2) and (3) it can be shown that $\theta = 2$ implies

$$\begin{aligned} \hat{\tau}_i &= \frac{\sqrt{2}}{(3\pi - 2)^{\frac{3}{2}}} (\pi + 4) \approx 0.50 \\ \hat{\phi}_i &= \frac{\sqrt{2}}{\sqrt{3\pi - 2}} \approx 0.52 \end{aligned}$$

for $i = 1, 2, \ldots, N$. For the calculation of $\tilde{\tau}_i$ and $\tilde{\phi}_i$, we use, for each θ, the true values of the mode $m = 0$, the mean μ, and the variance σ^2, so that $\tilde{\tau}_i$ and $\tilde{\phi}_i$ are determined by

$$\begin{aligned} \tilde{\tau}_i &= \frac{(y_i - \mu)^3}{\sigma} = (y_i - \mu)^3 \\ \tilde{\phi}_i &= \frac{y_i - m}{\sigma} = y_i \end{aligned}$$

θ	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00
$\tau_i = E[\tilde{\tau}_i]$	0.00	0.18	0.31	0.42	0.50	0.56	0.62	0.66	0.69	0.72	0.75	0.77	0.79
$\phi_i = E\left[\tilde{\phi}_i\right]$	0.00	0.18	0.32	0.43	0.52	0.59	0.66	0.71	0.76	0.80	0.83	0.86	0.89

Table 1: Expected standard skewness and Pearson mode skewness depending on θ.

for $i = 1, 2, \ldots, N$. Thus, for each θ, the equations to be estimated become

$$\begin{aligned} (y_i - \mu)^3 &= \beta_\tau \cdot 0.50 + \varepsilon_{\tau,i} \\ y_i &= \beta_\phi \cdot 0.52 + \varepsilon_{\phi,i}. \end{aligned}$$

The null hypotheses for optimal forecasts are given by $\beta_\tau = 1$ and $\beta_\phi = 1$. The forecasts are optimal only if $\theta = 2$. We use a t-test with a nominal size of 5%. Of course, the estimators $\hat{\beta}_\tau$ and $\hat{\beta}_\phi$ are not t-distributed, because $\varepsilon_{\tau,i}$ and $\varepsilon_{\phi,i}$ as well as y_i are not normally distributed, but the t-test should work well at least asymptotically. The expected values of $\tilde{\tau}_i$ and $\tilde{\phi}_i$, τ_i and ϕ_i, are shown in Table 1. Interestingly, both measures of skewness are quite similar for given values of θ unless θ becomes large.

The power curves of the optimality tests are plotted in Figure 2.[8] Obviously, the optimality test based on standard skewness performs extremely badly for small sample sizes. The power curve for $N = 20$ is essentially flat, equalling about 0.15 for all values of θ considered. For $N = 40$ and $N = 100$, the curves are downward sloping also in the range $2 \geq \theta \geq 4$. Even for $N = 1000$, the test appears to reject slightly more often at $\theta = 2.25$ than at $\theta = 2$ and only reaches a rejection probability of about 40% for $\theta = 4$. Only for $N = 5000$ the test seems to be

[8] They are based on simulations with 10,000 observations for each point of each power curve.

unbiased, although this conclusion might change if a finer grid for θ was used.

In contrast to that, the optimality test based on the Pearson mode skewness works reasonably well in small samples. For $N = 20$, its power is not very large, reaching up to about 60% if the density is standard normal, i.e. if $\theta = 1$. However, its size is close to 5%, although $\varepsilon_{\phi,i}$ is not normally distributed. For $N = 40$, the power at $\theta = 1$ increases to about 90%, and for $N = 100$ to 100%. In both cases the test has approximately the correct size. For $N = 1000$ and $N = 5000$, the power is large for all $\theta \neq 2$ considered and the empirical size is close to 5%.

These results show that tests for the optimality of risk forecasts based on the standard skewness are not useful in small samples. Optimality tests based on the Pearson mode skewness, however, seem to yield fairly satisfactory results even in small samples. In what follows, skewness will refer to the Pearson mode skewness, not to the standard measure of skewness based on third moments.

3 Testing the Optimality of Macroeconomic Risk Forecasts in Practice

The examples given above show that the Pearson mode skewness is a more promising measure of asymmetry than the standard skewness if risk forecast optimality is to be tested. However, the examples in the previous section are highly stylized for several reasons. For example, θ and hence the value of the Pearson mode skewness $\hat{\phi}$ does not vary within the sample. Moreover, mode and standard deviation are known to the forecaster. In addition, the risk forecasts are quantitative, whereas in practice, many institutions only publish the direction of the forecast risk. There-

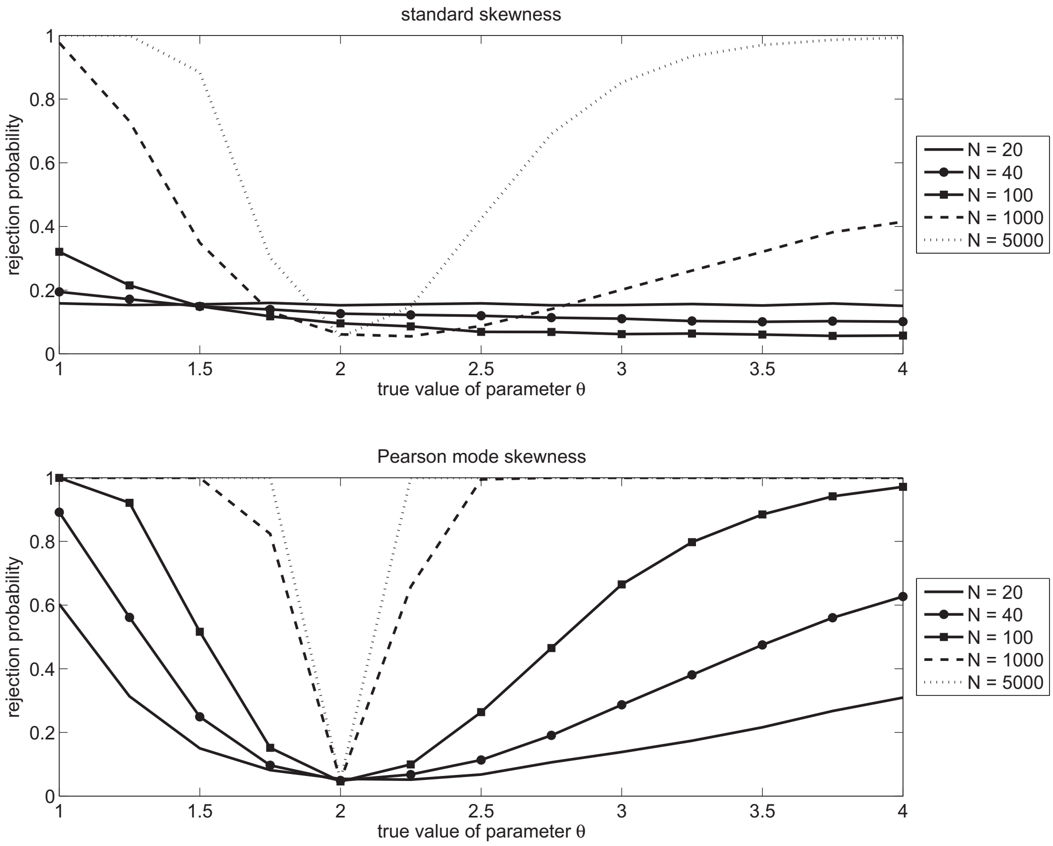

Figure 2: Rejection probabilities of two tests for risk forecast optimality with nominal size equal to 5%. The null hypothesis is $\theta = 2$.

fore, in what follows we will study the size and power of the optimality tests under more realistic conditions.

Another problem likely to be encountered in practice is given by the serial correlation of realized risks at least for forecast horizons larger than one. This problem can be addressed by using Newey & West (1987) standard errors. As this approach is well-known, we will not elaborate on this issue.

3.1 Varying Asymmetry

Instead of assuming that the true skewness ϕ as well as the forecast skewness $\hat{\phi}$ are fixed for all $i = 1, 2, ..., N$, we will now assume that they are uniformly distributed over the interval $\left[-\bar{\phi}, \bar{\phi}\right]$. Given a tpn-distribution, the largest possible value for $\bar{\phi}$ equals 1.32.[9] However, such a large value is rather unlikely to be encountered in practice, since it would imply a folded normal distribution, meaning that the realization is going to be larger than the mode with certainty. The absolute values for $\hat{\phi}$ found in the sample of risk forecasts by central banks studied in Knüppel & Schultefrankenfeld (2011) do not exceed 0.5. Moreover, the number of non-zero risk forecasts in that sample hardly exceeds 30 per forecast horizon, so that we will focus on samples of that size.

In the following Monte Carlo simulations, we consider two cases. In the first case, the true risks ϕ_i are independent of the risk forecasts $\hat{\phi}_i$, so that the risk forecasts have no information content. In the second case, the risk forecasts are optimal, i.e. $\phi_i = \hat{\phi}_i$ holds. For both cases, we simulate N tpn-distributed vari-

[9] The exact number is given by $\max\left(\bar{\phi}\right) = \sqrt{\frac{2/\pi}{1-2/\pi}}$.

ables[10] and estimate the parameter β in the equation

$$\tilde{\phi}_i = \beta\hat{\phi}_i + \varepsilon_{\phi,i} \tag{5}$$

with $i = 1, 2, \ldots, N$ as above in equation (4). However, in contrast to the simulations above, here in general $\hat{\phi}_i \neq \hat{\phi}_j$ holds for $i \neq j$. If the risk forecasts are optimal, $\tilde{\phi}_i$ is drawn from a tpn-distribution with skewness equal to $\hat{\phi}_i$ where the parameter $\hat{\phi}_i$ is drawn from a uniform distribution over the interval $\left[-\bar{\phi}, \bar{\phi}\right]$. If the risk forecasts have no information content, $\tilde{\phi}_i$ is drawn from a tpn-distribution with skewness ϕ_i, where ϕ_i is drawn from a uniform distribution over the interval $\left[-\bar{\phi}, \bar{\phi}\right]$, and $\hat{\phi}_i$ is drawn independently from the same distribution.[11]

We test the hypotheses of no information content ($\beta = 0$) and of optimal risk forecasts ($\beta = 1$) at a level of significance of 5% and for various values of N, but focusing on $N = 30$. We do so for five distinct values of $\bar{\phi}$, ranging from 0.25 to 1.32. The rejection probabilities, based on 10,000 simulations[12], are displayed in Table 2. Apparently, the optimality tests have the correct size. Their power, however, can be quite low if the largest possible value of the true and the forecast risks is low. When $N = 30$, with $\hat{\phi}_i$ coming from the interval $[-0.25, 0.25]$, (and in the case $\beta = 0$ with ϕ_i coming from the same interval,) the wrong null is rejected only in 12% of all simulations. In case that $\bar{\phi}$ equals 0.5, this number increases to slightly more than 30%. Only if the interval includes more extreme values of

[10] Given a value for ϕ_i and $\sigma = 1$, the corresponding values of σ_1 and σ_2 can be calculated as $\sigma_1 = \sqrt{\left(1 - \frac{3}{8}\pi\right)\phi_i^2 + 1} - \sqrt{\frac{1}{8}\pi}\phi_i$ and $\sigma_2 = \sqrt{\left(1 - \frac{3}{8}\pi\right)\phi_i^2 + 1} + \sqrt{\frac{1}{8}\pi}\phi_i$

[11] Alternatively, in case of no information content, ϕ_i could be drawn from a symmetric distribution. Using the standard normal distribution, we find results which are similar to those reported below.

[12] This holds for all following Monte Carlo simulations as well.

$\bar{\phi}$	N	$\beta = 0$		$\beta = 1$	
		H_0: $\beta = 0$	H_0: $\beta = 1$	H_0: $\beta = 0$	H_0: $\beta = 1$
0.25	30	0.05	0.12	0.12	0.05
0.50	30	0.05	0.31	0.32	0.05
0.75	30	0.05	0.56	0.63	0.05
1.00	30	0.05	0.74	0.88	0.05
1.32	30	0.05	0.88	0.99	0.05
0.50	50	0.05	0.49	0.52	0.06
0.50	100	0.05	0.78	0.81	0.05
0.50	200	0.05	0.97	0.98	0.05

Table 2: Size and power of tests for optimality of risk forecasts. Skewness is uniformly distributed over interval $\left[-\bar{\phi}, \bar{\phi}\right]$. The nominal size of the tests is 5%. Rejection probabilities are based on 10,000 simulations.

skewness the power of the tests increases to more satisfactory levels, reaching almost 100% if the forecasts are optimal and $\bar{\phi}$ equals 1.32. Given $\bar{\phi} = 0.5$, one needs about 200 observations to reach a comparable power.

3.2 Unknown Mode and Standard Deviation

In practice, additional complications arise from the fact that the mode m as well the standard deviation σ of the forecast density are potentially time-varying and unknown. Therefore m_i and σ_i have to be predicted themselves. So a risk forecast actually consists of three elements: the forecast mode, the difference between the forecast mean and forecast mode, and the forecast standard deviation.

So the question arises how problems with forecasting the mode or the standard deviation affect the results of tests for risk forecast optimality. Suppose, for example, that the forecaster optimally forecasts the difference between the mode and the mean as well as the standard deviation, but the mode forecasts are biased. If, in this case, a test for risk forecast optimality would reject the null of optimality due to the biased mode forecast, its result would be misleading.

Denote the mode forecast by $\hat{m}_i$, the mean forecast by $\hat{\mu}_i$ and the standard deviation forecast by $\hat{\sigma}_i$. The equation used in order to test for risk forecast optimality is then given by

$$\frac{y_i - \hat{m}_i}{\hat{\sigma}_i} = \alpha + \beta \frac{\hat{\mu}_i - \hat{m}_i}{\hat{\sigma}_i} + \varepsilon_i. \tag{6}$$

Assume that the forecast mode and the forecast standard deviation are related to the true mode m_i and standard deviation σ_i according to

$$\begin{aligned} \hat{\sigma}_i &= a_\sigma + \sigma_i + u_{\sigma,i} \\ \hat{m}_i &= a_m + m_i + u_{m,i} \end{aligned} \tag{7}$$

where $u_{\sigma,i}$ and $u_{m,i}$ are iid zero-mean error terms with finite variances σ_{u_σ} and σ_{u_m}.[13] If a_σ and a_m equal zero, the forecasts of mode and standard deviation are unbiased.

Suppose that the forecast standard deviation $\hat{\sigma}_i$ is equal to the true standard deviation σ_i and substitute for $\hat{m}_i$ on the left-hand side of equation (6). This yields

$$\frac{y_i - m_i}{\sigma_i} - \frac{a_m}{\sigma_i} = \alpha + \beta \frac{\hat{\mu}_i - \hat{m}_i}{\sigma_i} + \varepsilon_i + \frac{u_{m,i}}{\sigma_i}.$$

Note that we do not need to substitute for $\hat{m}_i$ on the right-hand side of equation (6), because $(\hat{\mu}_i - \hat{m}_i)/\sigma_i = (\hat{\mu}_i - \hat{m}_i)/\hat{\sigma}_i$ correctly measures the risk forecast. Assuming, for the sake of simplicity, that $\sigma_i = \sigma$ holds, this equation implies that

[13] In addition, the support of $u_{\sigma,i}$ must be bounded from below such that $\min(\hat{\sigma}_i) > 0$ holds.

the estimators for α and β in equation (6) converge to

$$\begin{aligned} plim\left(\hat{\alpha}\right) &= -\frac{a_m}{\sigma}+\alpha \\ plim\left(\hat{\beta}\right) &= \beta. \end{aligned}$$

Now suppose that the forecast mode $\hat{m}_i$ is equal to the true mode m_i and substitute for $\hat{\sigma}_i$ on the left-hand side of (6). This gives

$$\frac{y_i-m_i}{a_\sigma+\sigma_i+u_{\sigma,i}}=\alpha+\beta\frac{\hat{\mu}_i-m_i}{\hat{\sigma}_i}+\varepsilon_i.$$

Again, we we do not need to substitute for $\hat{\sigma}_i$ on the right-hand side of (6), because $\left(\hat{\mu}_i-m_i\right)/\hat{\sigma}_i=\left(\hat{\mu}_i-\hat{m}_i\right)/\hat{\sigma}_i$ correctly measures the risk forecast. Using the linear approximation

$$\frac{y_i-m_i}{a_\sigma+\sigma_i+u_{\sigma,i}}\approx\frac{y_i-m_i}{\sigma_i}\left(1-\frac{a_\sigma}{\sigma_i}-\frac{u_{\sigma,i}}{\sigma_i}\right) \tag{8}$$

and assuming that $\sigma_i=\sigma$ holds, it follows that

$$\begin{aligned} plim\left(\hat{\alpha}\right) &\approx \alpha\left(1-\frac{a_\sigma}{\sigma}\right) \\ plim\left(\hat{\beta}\right) &\approx \beta\left(1-\frac{a_\sigma}{\sigma}\right). \end{aligned} \tag{9}$$

So if $a_\sigma>0$, both estimates will be biased towards zero, and if $a_\sigma<0$, both estimates will be biased away from zero. Yet, if α is close to zero, the bias will tend to be negligible for the estimate $\hat{\alpha}$.

Given these results, a careful analysis of risk forecasts should firstly check for bias in the forecasts of mode and standard deviation. Unfortunately, without further assumptions it is unfeasible to evaluate the bias of mode forecasts. One might assume that the risks are balanced on average, so that $E\left[\frac{\mu_i - m_i}{\sigma}\right] = 0$ holds, implying $E[\mu_i] = E[m_i]$. In this case, one could test for the bias of mode forecasts in the same way as for the bias of mean forecasts.[14] If one is unwilling to make the assumption of balanced risks on average, one might just focus on the bias of the standard deviation and concentrate on tests of hypotheses about β.[15] Of course, a potential bias of the risk forecasts given by $\alpha \neq 0$ will remain undetected in this case.

If the forecasts of mode and standard deviation are unbiased, the main difference with respect to the results of the Monte Carlo simulations conducted above is mostly a loss of power. Results of Monte Carlo simulations where mode and standard deviation are unknown are presented in Table 3. The parameter $\bar{u}_\sigma$ denotes the limits of the interval $(-\bar{u}_\sigma, \bar{u}_\sigma)$ of the uniform distribution from which u_σ is drawn. u_m is drawn from a normal distribution. The true standard deviation is given by $\sigma_i = 1$ and the true mode by $m_i = 0$. Moreover, as in most simulations above, $N = 30$ and $\bar{\phi} = 0.5$.

The results confirm the considerations from above. Tests about β can suffer

[14] If the standard deviation can vary, one would need the additional assumption that $(\mu_i - m_i)$ and σ_i are independent, because then $E\left[\frac{\mu_i - m_i}{\sigma_i}\right] = \frac{1}{E[\sigma_i]} E[\mu_i - m_i]$ holds, so that $E\left[\frac{\mu_i - m_i}{\sigma_i}\right] = 0$ continues to imply $E[\mu_i - m_i] = 0$.
Otherwise, $E\left[\frac{\mu_i - m_i}{\sigma_i}\right] = 0$ would still imply $E\left[\frac{m_i}{\sigma_i}\right] = E\left[\frac{\mu_i}{\sigma_i}\right]$. So one could test the hypothesis $E\left[\frac{m_i}{\sigma_i}\right] = 0$. However, note that the standard test for the bias of a mean forecast should actually be based on the regression $\frac{y_i}{\sigma_i} - \frac{\mu_i}{\sigma_i} = \alpha + \varepsilon_i$ instead of $y_i - \mu_i = \alpha + \varepsilon_i$. Yet, it is always the latter equation that is employed. This approach could be justified if σ_i hardly varies or if heteroskedasticity-consistent standard errors are used. In any case, it does not seem too problematic to use the regression $y_i - m_i = \alpha + \varepsilon_i$ as well for testing the bias of mode forecasts in case of balanced risks.

[15] Another possible but even more restrictive assumption would be given by the constancy of the mode, i.e. by $m = m_i$. In this case, kernel density estimation would be needed to estimate m.

				$\alpha=0,\beta=0$				$\alpha=0,\beta=1$			
						H_0 is				H_0 is	
a_m	σ_{u_m}	a_σ	$\bar{u}_\sigma$	$\hat{\alpha}$	$\hat{\beta}$	$\beta=0$	$\beta=1$	$\hat{\alpha}$	$\hat{\beta}$	$\beta=0$	$\beta=1$
0	0	0	0	0.00	−0.01	0.05	0.31	0.00	1.00	0.31	0.04
0.3	0	0	0	−0.30	0.00	0.05	0.31	−0.30	1.01	0.32	0.05
0	0.5	0	0	0.00	0.01	0.05	0.25	0.00	0.99	0.26	0.05
0	0	0.3	0	0.00	0.00	0.05	0.47	0.00	0.77	0.31	0.08
0	0	0	0.25	0.00	−0.01	0.05	0.29	0.00	1.02	0.30	0.05
0.3	0.5	0	0.25	−0.31	0.01	0.05	0.25	−0.31	1.01	0.25	0.05

Table 3: Size and power of test for optimality of risk forecasts if mode and standard deviation are unknown. Skewness is uniformly distributed over interval $[-0.5, 0.5]$. Sample size N equals 30. The nominal size of the tests is 5%. Rejection probabilities are based on 10,000 simulations.

from size distortions if $a_\sigma \neq 0$, because then $plim\left(\hat{\beta}\right) \neq \beta$.[16]

3.3 Direction-of-Risk Forecasts

If risk forecasts are only qualitative, i.e. only the direction of the risk is forecast but not its magnitude, it is nevertheless possible to conduct tests for forecast optimality. In this case, the realized and the forecast risks are categorical variables. Even if the risk forecasts are quantitative, a transformation of quantitative risks (interval variables) to qualitative risks (categorical variables) might be interesting if the focus rather lies on the direction of the risks.

Actually, in the case of qualitative risk forecasts, there are only two possible outcomes: success (direction of forecast risk equals direction of realized risk) and failure (direction of forecast risk differs from direction of realized risk). Therefore, one could use tests based on the binomial distribution. However, in empirical applications one is most likely going to be confronted with the problem of serial

[16] Note that $\hat{\beta}$ on average equals 0.77, differing slightly from the value 0.70 implied by equation (9), because the latter result is based on an approximation.

correlation. This problem can be addressed more easily if the categorical variables are analyzed in a regression context.[17]

Recode the realized and forecast risks according to

$$\tilde{\omega}_i = \begin{cases} 1 & \text{if } \tilde{\phi}_i > 0 \\ 0 & \text{if } \tilde{\phi}_i \leq 0 \end{cases}, \tag{10}$$

and

$$\hat{\omega}_i = \begin{cases} 1 & \text{if } \hat{\phi}_i > 0 \\ 0 & \text{if } \hat{\phi}_i \leq 0 \end{cases}. \tag{11}$$

Then tests for forecast optimality can be based on the regression

$$\tilde{\omega}_i = \alpha + \beta_\omega \hat{\omega}_i + \varepsilon_{\omega,i}. \tag{12}$$

Obviously, if the risk forecasts have no information content, β_ω equals zero. If the risk forecasts are optimal, β_ω is larger than zero but, in contrast to β, it is not equal to one in general. This is due to the fact that, even if the forecast density is strongly skewed, there is still a positive probability mass on each side of the mode.[18] However, necessary (though not sufficient) conditions for risk forecast optimality could be tested using one-sided tests. If the risk forecasts are optimal, the null hypothesis $\beta_\omega \leq 0$ should be rejected, whereas the null hypothesis $\beta_\omega \geq 0$ should not. Note that optimal forecasts do not imply $\alpha = 0$.

We use the same Monte Carlo simulation design as for the results in Table 2 in order to obtain information about the rejection probabilities. That is, we simulate

[17] See Pesaran & Timmermann (2009).

[18] Only in the special case of a folded normal distribution, i.e. the case $\phi = \pm\sqrt{\frac{2/\pi}{1-2/\pi}}$, optimal risk forecasts would always correctly predict the direction of the realized risk.

$\tilde{\phi}_i$ and $\hat{\phi}_i$ as in Section 3.1, and we construct the variables $\tilde{\omega}_i$ and $\hat{\omega}_i$ according to equations (10) and (11). Then we estimate equation (12) and test hypotheses about β_ω. The results are displayed in Table 4. Here, the sample size N always equals 30, and the nominal size of the tests is again 5%.

If the risk forecasts have no information content, β_ω equals zero and the null hypotheses about β_ω ($\beta_\omega = 0$, $\beta_\omega \leq 0$, $\beta_\omega \geq 0$) have the correct size. If the risk forecasts are optimal, β_ω is positive, but can be close to zero if $\bar{\phi}$ is small. Consequently, the null hypothesis $\beta_\omega = 0$ can hardly be rejected if $\bar{\phi}$ is small. The power of the test only marginally exceeds its nominal size if $\bar{\phi} = 0.25$.

The results show that it is preferable to conduct tests for risk forecast optimality based on quantitative risk forecasts if available. For example, according to Table 2 the power of the test for $\beta = 0$ if $\beta = 1$ is 33% if $\bar{\phi} = 0.5$. This is considerably larger than the 13% attained with qualitative risk forecasts and the test for $\beta_\omega = 0$.

If $\beta = 1$, the one-sided test for $\beta_\omega \leq 0$ of course rejects more often than the test for $\beta_\omega = 0$, but its power is not very large either. For example, it rejects in 22% of the simulations if $\bar{\phi} = 0.5$. Given $\beta = 1$, the one-sided test for $\beta_\omega \geq 0$ hardly ever rejects even if $\bar{\phi}$ is small.

The 95% confidence intervals for the estimate of β_ω interestingly include negative values unless $\bar{\phi}$ is very large. This implies that, even if the risk forecasts are optimal, it is not unlikely to observe more failures than successes of the qualitative risk forecasts in small samples.

One might expect that the fact that mode and standard deviation are unknown further reduces the power of the tests if qualitative risk forecasts are analyzed. This, however, is not always the case. If the realized risks are transformed to

$\bar{\phi}$	$\beta = 0$				$\beta = 1$				
		H_0 is				H_0 is			
	β_ω	$\beta_\omega = 0$	$\beta_\omega \leq 0$	$\beta_\omega \geq 0$	β_ω	$\beta_\omega = 0$	$\beta_\omega \leq 0$	$\beta_\omega \geq 0$	95% CI for $\hat{\beta}_\omega$
0.25	0	0.05	0.05	0.05	0.08	0.06	0.11	0.02	$(-0.28, 0.43)$
0.50	0	0.05	0.05	0.05	0.16	0.13	0.22	0.01	$(-0.20, 0.50)$
0.75	0	0.05	0.05	0.05	0.24	0.25	0.38	0.00	$(-0.13, 0.59)$
1.00	0	0.05	0.05	0.05	0.33	0.43	0.56	0.00	$(-0.02, 0.66)$
1.32	0	0.05	0.05	0.05	0.45	0.72	0.82	0.00	$(0.11, 0.75)$

Table 4: Size and power of tests for optimality of risk forecasts. Skewness is uniformly distributed over interval $\left[-\bar{\phi}, \bar{\phi}\right]$. Sample size N equals 30. The nominal size of all tests is 5%. CI stands for confidence interval.

categorical variables according to

$$\tilde{\omega}_i = \begin{cases} 1 & \text{if } \frac{y_i - \hat{m}_i}{\hat{\sigma}_i} > 0 \\ 0 & \text{if } \frac{y_i - \hat{m}_i}{\hat{\sigma}_i} \leq 0 \end{cases}$$

and the forecast risks according to equation (11), the fact that the standard deviation is unknown and thus has to be forecast is not problematic. This is because in the case of categorical variables, only the sign of $(y_i - \hat{m}_i) / \hat{\sigma}_i$ matters. This sign does obviously not depend on the forecast of $\hat{\sigma}_i$. The mode forecast, however, can affect the size and the power of the tests for forecast optimality.

Several results of Monte Carlo simulations can be found in Table 5. Here u_m is drawn from a normal distribution. The true standard deviation is given by $\sigma_i = 1$ and the true mode by $m_i = 0$. Moreover, $N = 30$ and $\bar{\phi} = 0.5$. The effects of mode forecast errors on $\hat{\beta}_\omega$ are non-linear in case of categorical variables.[19] In general, it seems that only large deviations from $a_m = 0$ and $\sigma_{u_m} = 0$ have a noticeable

[19] For example, with known mode, i.e. $a_m = 0$ and $\sigma_{u_m} = 0$, β_ω in equation (12) equals 0.16. With $a_m = 0.6$, the estimate of β_ω hardly changes, decreasing to 0.15. With $a_m = 1.2$, the estimate of β_ω drops to 0.11 and with $a_m = 1.8$ to 0.05.

		$\alpha=0,\beta=0$		$\alpha=0,\beta=1$		
		$\beta_\omega=0$	H_0 is	$\beta_\omega=0.16$	H_0 is	
a_m	σ_{u_m}	$\hat{\beta}_\omega$	$\beta_\omega=0$	$\hat{\beta}_\omega$	$\beta_\omega=0$	95% CI for $\hat{\beta}_\omega$
0	0	0.00	0.05	0.16	0.13	$(-0.20, 0.52)$
0.3	0	0.00	0.05	0.15	0.14	$(-0.20, 0.50)$
0.6	0	0.00	0.05	0.15	0.14	$(-0.17, 0.47)$
1.2	0	0.00	0.03	0.11	0.11	$(-0.12, 0.35)$
1.8	0	0.00	0.01	0.05	0.02	$(-0.08, 0.21)$
0	0.5	0.00	0.05	0.15	0.13	$(-0.21, 0.50)$
0	2.0	0.00	0.05	0.08	0.07	$(-0.28, 0.45)$
0.3	0.5	0.00	0.05	0.15	0.12	$(-0.20, 0.48)$

Table 5: Size and power of test for optimality of categorical risk forecasts if modes and standard deviations are unknown. Skewness is uniformly distributed over the interval $[-0.5, 0.5]$. Sample size N equals 30. The nominal size of all tests is 5%. CI stands for confidence interval.

effect on the size and power of tests for risk forecast optimality if qualitative risk forecasts are analyzed.[20]

3.4 An Alternative Test for Risk Forecast Optimality?

One might think that an alternative way of testing the optimality of risk forecasts would be given by using a standard test for mean forecast optimality, i.e. testing the null hypothesis $\alpha = 0$ and $\beta = 1$ in the regression $y_i = \alpha + \beta\hat{\mu}_i + \varepsilon_i$, where ε_i is an error term. This alternative test could work, because, if the null hypothesis is true, this implies that $(y_i - m_i)/\sigma_i = (\hat{\mu}_i - m_i)/\sigma_i + \varepsilon_i/\sigma_i$ holds, independently of the properties of the forecasts $\hat{m}_i$ and $\hat{\sigma}_i$. So this test might seem preferable if one wants to circumvent problems arising from forecasting m_i and σ_i. However,

[20] In the most extreme case, a_m can deviate so strongly from zero that the dependent variable becomes a constant, because one would always observe a downward risk if a_m is very large or an upward risk if a_m is very small.

it turns out that strong deviations from risk forecast optimality can easily remain undetected by standard tests of forecast optimality.

Consider the following case: Suppose that $\hat{m}_i = m_i$ and that $\hat{\sigma}_i = \sigma_i$. Suppose further that the risk forecasts are given by $\hat{\mu}_i - \hat{m}_i = \delta(\mu_i - m_i)$. So if $\delta > 1$, the magnitude of risk is overestimated, whereas if $0 \leq \delta < 1$, the magnitude of risk is underestimated. This implies that the mean forecast, given by $\hat{\mu}_i = \hat{m}_i + (\hat{\mu}_i - \hat{m}_i)$, equals

$$\hat{\mu}_i = \hat{m}_i + \delta(\mu_i - m_i) = (1-\delta)\, m_i + \delta\mu_i.$$

For the sake of simplicity, also assume that $E[m_i] = E[\mu_i] = 0$. The condition $E[m_i] = E[\mu_i]$ implies that the (unscaled) risks are balanced on average.

Even if δ strongly differs from 1, i.e if the risk forecasts are strongly non-optimal, it might be very difficult to detect that $\hat{\mu}_i \neq \mu_i$. Firstly, since $E[m_i] = E[\mu_i]$ holds, $E[\hat{\mu}_i] = E[\mu_i]$ holds as well. Secondly, the correlation between $\hat{\mu}_i$ and μ_i can be very large even if δ strongly differs from 1. Assuming that $Cov[m_i, \mu_i - m_i] = 0$, the correlation equals

$$\frac{E[\hat{\mu}_i\mu_i]}{\sqrt{E[\hat{\mu}_i^2]\, E[\mu_i^2]}} = \frac{\sigma_m^2 + \delta\sigma_{\mu-m}^2}{\sqrt{\left(\delta^2\sigma_{\mu-m}^2 + \sigma_m^2\right)\left(\sigma_{\mu-m}^2 + \sigma_m^2\right)}} \tag{13}$$

where σ_m^2 denotes the variance of the mode and $\sigma_{\mu-m}^2$ denotes the variance of the unscaled risk, i.e. the variance of the difference between mode and mean.[21] If $\sigma_{\mu-m}^2$ is small compared to σ_m^2, this correlation will be close to 1 even if δ strongly differs from 1.

As an example, consider the inflation nowcasts of the Bank of England stud-

[21] See Appendix A for details.

ied in Knüppel & Schultefrankenfeld (2011).[22] The variance of the unscaled risk forecasts, i.e. of the mean-mode difference equals 0.0022, whereas the variance of the mode forecasts equals 0.4704. Assuming that these values are good estimates of the true variances of mode and unscaled risk, the variance of the mode is more than 200 times larger than the variance of the unscaled risk. Even if the risk forecasts contain no information, so that $\delta = 0$, the correlation between between $\hat{\mu}_i$ and μ_i would equal 0.9975. When testing the null hypothesis $\alpha = 0$ and $\beta = 1$ in the regression $y_i = \alpha + \beta\hat{\mu}_i + u_i$ in such a case, one will need an extremely large number of observations to obtain an acceptable power of the test. With $N = 30$, the power of the test is virtually identical to its size.[23]

4 Conclusion

In this work, we have investigated several issues arising when evaluating macroeconomic risk forecasts, focusing on tests for the optimality of these forecasts. Such tests are confronted with the problems of small samples, at best moderate risks (i.e. asymmetries) and potentially only categorical instead of quantitative risk forecasts. Although all simulation results presented rest on certain assumptions like the tpn-distribution of the densities and the uniform distribution for the risks, they nevertheless clearly show that one should not expect too much power of the optimality tests for risk forecasts. However, they also give guidance in order for the tests to have the correct size and as much power as possible under the given circumstances.

[22]For these nowcasts, the assumption $\hat{m}_i = m_i$ might be a good approximation to the more realistic relation $\hat{m}_i = m_i + u_{m,i}$, because the variance of the mode forecasts is much larger than the variance of the forecast density. The latter can be thought of as a proxy for $E\left[u_{m,i}^2\right]$.

[23]Results of Monte Carlo simulations are available on request.

We find that the asymmetry of the risk forecasts should not be measured by the standard coefficient of skewness, because the power of the test would be extremely low in this case. Instead, the Pearson mode skewness is preferable yielding much larger power in small samples. The asymmetry of the true density then has to be measured using the forecasts of mode and standard deviation. This can cause problems in terms of power loss if these forecasts are imprecise and in terms of size distortions if these forecasts are biased.

In order to evaluate the bias of the mode forecasts, potentially problematic assumptions are required. If these assumptions are to be avoided, tests for risk forecast optimality can still be conducted if the null hypothesis of the test is modified. That is, from the composite null hypothesis about the constant and the slope parameter, one can exclude the hypothesis about the constant and concentrate on the slope parameter only, because the estimate of the slope parameter remains unaffected by a potential bias of the mode forecasts.

If the risk forecasts are available on a categorical basis only, the tests suffer from a further loss in power. However, they are not affected by the potential impreciseness of the forecast standard deviation.

Considering the relation of standard tests for mean forecast optimality and risk forecast optimality, one might think that the results of these standard tests contain information about risk forecast optimality. While this is true in principle, we find that in practice the results of these standard tests should not be interpreted in terms of risk forecast optimality.

In order to tackle the problem of potentially low power, risk forecasts could possibly better be evaluated in a panel setup. This may be feasible if the risk forecasts for a certain variable are, for example, available for several forecast horizons.

References

Brown, B. W. & Maital, S. (1981), 'What do economists know? an empirical study of experts' expectations', *Econometrica* **49**, 491–504.

Diebold, F. X. & Lopez, J. A. (1996), Forecast evaluation and combination, *in* G. S. M. und C. R. Rao, ed., 'Handbook of Statistics', Vol. 14, Elsevier, chapter 8, pp. 241–268.

Knüppel, M. & Schultefrankenfeld, G. (2011), 'How informative are central bank assessments of macroeconomic risks?', mimeo.

Knüppel, M. & Tödter, K.-H. (2007), 'Quantifying risk and uncertainty in macroeconomic forecasts', *Deutsche Bundesbank Discussion Paper Series 1: Economic Studies* **25**.

Leeper, E. M. (2003), 'An inflation reports report', *NBER Working Paper* **10089**.

Newey, W. K. & West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

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

Ramberg, J. S. & Schmeiser, B. W. (1974), 'An approximate method for generating asymmetric random variables', *Communications of the ACM* **Volume 17**(2), 78–82.

Todd C. Headrick, R. K. K. & Sheng, Y. (2008), 'Parametric probability densities and distribution functions for tukey g-and-h transformations and their use for fitting data', *Applied Mathematical Sciences* **2**(9), 449 – 462.

Wallis, K. F. (2004), 'An assessment of bank of england and national institute inflation forecast uncertainties', *National Institute Economic Review* **189**, 64–71.

A Appendix: Correlation of $\hat{\mu}_i$ and μ_i

As stated in the text, suppose that $\hat{m}_i = m_i$ and that $\hat{\sigma}_i = \sigma_i$, and that the risk forecasts are given by $\frac{\hat{\mu}_i - \hat{m}_i}{\sigma_i} = \delta \frac{\mu_i - m_i}{\sigma_i}$. This implies that the mean forecast, given by $\hat{\mu}_i = \hat{m}_i + \sigma_i \frac{\hat{\mu}_i - \hat{m}_i}{\sigma_i}$, equals

$$\hat{\mu}_i = \hat{m}_i + \delta \left(\mu_i - m_i \right).$$

Using the definition $\phi_i = \frac{\mu_i - m_i}{\sigma_i}$, the true mean is given by

$$\mu_i = m_i + \sigma_i \phi_i.$$

Because of the non-restrictive assumption $E[\hat{\mu}_i] = E[\mu_i] = 0$, the correlation of $\hat{\mu}_i$ and μ_i equals

$$\frac{E[\hat{\mu}_i \mu_i]}{\sqrt{E[\hat{\mu}_i^2] E[\mu_i^2]}} = \frac{E[((1-\delta) m_i + \delta \mu_i) \mu_i]}{\sqrt{E\left[((1-\delta) m_i + \delta \mu_i)^2\right] E[\mu_i^2]}}.$$

In addition, because of $E[m_i] = E[\mu_i]$ and using the additional assumption that the (unscaled) risk forecasts are uncorrelated with the mode forecasts, i.e. $Cov[m_i, \sigma_i \phi_i] = Cov[m_i, \mu_i - m_i] = 0$, one obtains the expression

$$\frac{E[\hat{\mu}_i \mu_i]}{\sqrt{E[\hat{\mu}_i^2] E[\mu_i^2]}} = \frac{E\left[m_i^2 + \delta (\mu_i - m_i)^2\right]}{\sqrt{E\left[\delta^2 (\mu_i - m_i)^2 + m_i^2\right] E\left[(\mu_i - m_i)^2 + m_i^2\right]}}.$$

Using the definitions $\sigma_m^2 = E[m_i^2]$ and $\sigma_{\mu-m}^2 = E\left[(\mu_i - m_i)^2\right]$ gives equation (13).

The assumption $Cov[m_i, \mu_i - m_i] = 0$ could be problematic if, for example, downwards risks tend to prevail when the mode is above its average. However, according to the Bank of England data used in Knüppel & Schultefrankenfeld (2011), at least the correlation of the forecast modes and unscaled risks $Corr\left[\hat{m}_i, \hat{\mu}_i - \hat{m}_i\right]$ is close to zero, ranging between -0.18 and 0.15, depending on the forecast horizon. Moreover, as long as the variance ratio $\sigma_m^2/\sigma_{\mu-m}^2$ is large, the effect of a non-zero covariance $Cov\left[m_i, \mu_i - m_i\right]$ on the correlation of $\hat{\mu}_i$ and μ_i is very limited.

The following Discussion Papers have been published since 2010:

Series 1: Economic Studies

01	2010	Optimal monetary policy in a small open economy with financial frictions	Rossana Merola
02	2010	Price, wage and employment response to shocks: evidence from the WDN survey	Bertola, Dabusinskas Hoeberichts, Izquierdo, Kwapil Montornès, Radowski
03	2010	Exports versus FDI revisited: Does finance matter?	C. M. Buch, I. Kesternich A. Lipponer, M. Schnitzer
04	2010	Heterogeneity in money holdings across euro area countries: the role of housing	Ralph Setzer Paul van den Noord Guntram Wolff
05	2010	Loan supply in Germany during the financial crises	U. Busch M. Scharnagl, J. Scheithauer
06	2010	Empirical simultaneous confidence regions for path-forecasts	Òscar Jordà, Malte Knüppel Massimiliano Marcellino
07	2010	Monetary policy, housing booms and financial (im)balances	Sandra Eickmeier Boris Hofmann
08	2010	On the nonlinear influence of Reserve Bank of Australia interventions on exchange rates	Stefan Reitz Jan C. Ruelke Mark P. Taylor
09	2010	Banking and sovereign risk in the euro area	S. Gerlach A. Schulz, G. B. Wolff
10	2010	Trend and cycle features in German residential investment before and after reunification	Thomas A. Knetsch

11	2010	What can EMU countries' sovereign bond spreads tell us about market perceptions of default probabilities during the recent financial crisis?	Niko Dötz Christoph Fischer
12	2010	User costs of housing when households face a credit constraint – evidence for Germany	Tobias Dümmler Stephan Kienle
13	2010	Extraordinary measures in extraordinary times – public measures in support of the financial sector in the EU and the United States	Stéphanie Marie Stolz Michael Wedow
14	2010	The discontinuous integration of Western Europe's heterogeneous market for corporate control from 1995 to 2007	Rainer Frey
15	2010	Bubbles and incentives: a post-mortem of the Neuer Markt in Germany	Ulf von Kalckreuth Leonid Silbermann
16	2010	Rapid demographic change and the allocation of public education resources: evidence from East Germany	Gerhard Kempkes
17	2010	The determinants of cross-border bank flows to emerging markets – new empirical evidence on the spread of financial crisis	Sabine Herrmann Dubravko Mihaljek
18	2010	Government expenditures and unemployment: a DSGE perspective	Eric Mayer, Stéphane Moyen Nikolai Stähler
19	2010	NAIRU estimates for Germany: new evidence on the inflation-unemployment trade-off	Florian Kajuth
20	2010	Macroeconomic factors and micro-level bank risk	Claudia M. Buch Sandra Eickmeier, Esteban Prieto

21	2010	How useful is the carry-over effect for short-term economic forecasting?	 Karl-Heinz Tödter
22	2010	Deep habits and the macroeconomic effects of government debt	 Rym Aloui
23	2010	Price-level targeting when there is price-level drift	C. Gerberding R. Gerke, F. Hammermann
24	2010	The home bias in equities and distribution costs	P. Harms M. Hoffmann, C. Ortseifer
25	2010	Instability and indeterminacy in a simple search and matching model	Michael Krause Thomas Lubik
26	2010	Toward a Taylor rule for fiscal policy	M. Kliem, A. Kriwoluzky
27	2010	Forecast uncertainty and the Bank of England interest rate decisions	 Guido Schultefrankenfeld
01	2011	Long-run growth expectations and "global imbalances"	M. Hoffmann M. Krause, T. Laubach
02	2011	Robust monetary policy in a New Keynesian model with imperfect interest rate pass-through	 Rafael Gerke Felix Hammermann
03	2011	The impact of fiscal policy on economic activity over the business cycle – evidence from a threshold VAR analysis	 Anja Baum Gerrit B. Koester
04	2011	Classical time-varying FAVAR models – estimation, forecasting and structural analysis	S. Eickmeier W. Lemke, M. Marcellino

05	2011	The changing international transmission of financial shocks: evidence from a classical time-varying FAVAR	Sandra Eickmeier Wolfgang Lemke Massimiliano Marcellino
06	2011	FiMod – a DSGE model for fiscal policy simulations	Nikolai Stähler Carlos Thomas
07	2011	Portfolio holdings in the euro area – home bias and the role of international, domestic and sector-specific factors	Axel Jochem Ute Volz
08	2011	Seasonality in house prices	F. Kajuth, T. Schmidt
09	2011	The third pillar in Europe: institutional factors and individual decisions	Julia Le Blanc
10	2011	In search for yield? Survey-based evidence on bank risk taking	C. M. Buch S. Eickmeier, E. Prieto
11	2011	Fatigue in payment diaries – empirical evidence from Germany	Tobias Schmidt
12	2011	Currency blocs in the 21st century	Christoph Fischer
13	2011	How informative are central bank assessments of macroeconomic risks?	Malte Knüppel Guido Schultefrankenfeld
14	2011	Evaluating macroeconomic risk forecasts	Malte Knüppel Guido Schultefrankenfeld

Series 2: Banking and Financial Studies

01	2010	Deriving the term structure of banking crisis risk with a compound option approach: the case of Kazakhstan	Stefan Eichler Alexander Karmann Dominik Maltritz
02	2010	Recovery determinants of distressed banks: Regulators, market discipline, or the environment?	Thomas Kick Michael Koetter Tigran Poghosyan
03	2010	Purchase and redemption decisions of mutual fund investors and the role of fund families	Stephan Jank Michael Wedow
04	2010	What drives portfolio investments of German banks in emerging capital markets?	Christian Wildmann
05	2010	Bank liquidity creation and risk taking during distress	Berger, Bouwman Kick, Schaeck
06	2010	Performance and regulatory effects of non-compliant loans in German synthetic mortgage-backed securities transactions	Gaby Trinkaus
07	2010	Banks' exposure to interest rate risk, their earnings from term transformation, and the dynamics of the term structure	Christoph Memmel
08	2010	Completeness, interconnectedness and distribution of interbank exposures – a parameterized analysis of the stability of financial networks	Angelika Sachs
09	2010	Do banks benefit from internationalization? Revisiting the market power-risk nexus	C. M. Buch C. Tahmee Koch, M. Koetter

10	2010	Do specialization benefits outweigh concentration risks in credit portfolios of German banks?	Rolf Böve Klaus Düllmann Andreas Pfingsten
11	2010	Are there disadvantaged clienteles in mutual funds?	Stephan Jank
12	2010	Interbank tiering and money center banks	Ben Craig, Goetz von Peter
13	2010	Are banks using hidden reserves to beat earnings benchmarks? Evidence from Germany	Sven Bornemann, Thomas Kick Christoph Memmel Andreas Pfingsten
14	2010	How correlated are changes in banks' net interest income and in their present value?	Christoph Memmel
01	2011	Contingent capital to strengthen the private safety net for financial institutions: Cocos to the rescue?	George M. von Furstenberg
02	2011	Gauging the impact of a low-interest rate environment on German life insurers	Anke Kablau Michael Wedow
03	2011	Do capital buffers mitigate volatility of bank lending? A simulation study	Frank Heid Ulrich Krüger
04	2011	The price impact of lending relationships	Ingrid Stein
05	2011	Does modeling framework matter? A comparative study of structural and reduced-form models	Yalin Gündüz Marliese Uhrig-Homburg
06	2011	Contagion at the interbank market with stochastic LGD	Christoph Memmel Angelika Sachs, Ingrid Stein

07	2011	The two-sided effect of financial globalization on output volatility	Barbara Meller
08	2011	Systemic risk contributions: a credit portfolio approach	Klaus Düllmann Natalia Puzanova
09	2011	The importance of qualitative risk assessment in banking supervision before and during the crisis	Thomas Kick Andreas Pfingsten

Visiting researcher at the Deutsche Bundesbank

The Deutsche Bundesbank in Frankfurt is looking for a visiting researcher. Among others under certain conditions visiting researchers have access to a wide range of data in the Bundesbank. They include micro data on firms and banks not available in the public. Visitors should prepare a research project during their stay at the Bundesbank. Candidates must hold a PhD and be engaged in the field of either macroeconomics and monetary economics, financial markets or international economics. Proposed research projects should be from these fields. The visiting term will be from 3 to 6 months. Salary is commensurate with experience.

Applicants are requested to send a CV, copies of recent papers, letters of reference and a proposal for a research project to:

Deutsche Bundesbank
Personalabteilung
Wilhelm-Epstein-Str. 14

60431 Frankfurt
GERMANY