DEUTSCHE BUNDESBANK
EUROSYSTEM

Evaluating the calibration of multi-step-ahead density forecasts using raw moments

Malte Knüppel

Discussion Paper
Series 1: Economic Studies
No 32/2011

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 978-3–86558–772–5 (Printversion)
ISBN 978-3–86558–773–2 (Internetversion)

Abstract:

The evaluation of multi-step-ahead density forecasts is complicated by the serial correlation of the corresponding probability integral transforms. In the literature, three testing approaches can be found which take this problem into account. However, these approaches can be computationally burdensome, ignore important information and therefore lack power, or suffer from size distortions even asymptotically. In this work, a fourth testing approach based on raw moments is proposed. It is easy to implement, uses standard critical values, can include all moments regarded as important, and has correct asymptotic size. It is found to have good size and power properties if it is based directly on the (standardized) probability integral transforms.

Keywords: Density forecast evaluation; normality tests

JEL-Classification: C12, C52, C53

Non-technical Summary

Today, predictions are often made in the form of density forecasts. An increasing number of central banks publishes density forecasts, which are displayed by fan charts. Compared to point forecasts, density forecasts contain additional information. From density forecasts for inflation, for example, it is possible to infer the probability of deflation or the probability of inflation being higher than the central bank's target.

Point forecasts can be evaluated according to properties like bias or efficiency. Similarly, density forecasts can also be evaluated. A forecast density should coincide with the true density of the variable under study. If this is the case, the density forecast is said to be correctly calibrated. However, if, for instance, over a certain period of time the realizations of the variable under study always occur within a very narrow interval around the means of the forecast densities, this would be a strong indication for incorrect calibration. The forecast densities probably have a too large width in this case.

If density forecasts for more than one period ahead are to be evaluated, this evaluation is complicated by the serial correlation of the outcomes with respect to the density forecasts. Suppose, for example, that one density forecast for inflation is made in January for July, and the next forecast is made in February for August. Then, if inflation in July turns out to be much higher than the mean of January's density forecast, it is very likely that inflation in August will also be considerably higher than the mean of February's density forecast.

One can distinguish three evaluation approaches that are used or suggested in the literature for these situations. However, each of them has certain disadvantages

with respect to the ability to detect incorrect calibration, to the possibility of falsely concluding that the density forecasts have incorrect calibration although it is actually correct, or to the ease of use. Therefore, an alternative evaluation approach, which does not suffer from any of these drawbacks, is suggested in this paper. In simulations, this new approach is found to yield good results and, thus, to be a viable alternative to the existing approaches.

Nicht-technische Zusammenfassung

Vorhersagen werden heutzutage oft in Form von Dichteprognosen gemacht. Auch Zentralbanken veröffentlichen in zunehmendem Maße Dichteprognosen, die als Fächerdiagramme (Fan Charts) dargestellt werden. Im Vergleich zu Punktprognosen enthalten Dichteprognosen zusätzliche Informationen. Aus Dichteprognosen für die Inflation ist zum Beispiel ersichtlich, wie hoch die Wahrscheinlichkeit für eine Deflation ist oder wie wahrscheinlich es ist, dass die Inflation über der Zielmarke der Zentralbank liegt.

Punktprognosen können bezüglich verschiedener Eigenschaften, wie Verzerrung oder Effizienz, beurteilt werden. In ähnlicher Weise ist auch die Beurteilung von Dichteprognosen möglich. Eine Prognosedichte sollte mit der wahren Dichte der untersuchten Variable übereinstimmen. Falls dies der Fall ist, spricht man von einer korrekt kalibrierten Dichteprognose. Wenn jedoch zum Beispiel die Realisationen der untersuchten Variable über einen längeren Zeitraum hinweg immer in einem sehr engen Intervall um die Mittelwerte der Prognosedichten liegen, so würde dies auf eine fehlerhafte Kalibrierung hindeuten. Die Prognosedichten würden in diesem Fall wahrscheinlich eine zu große Breite besitzen.

Falls Dichteprognosen für mehr als eine Periode im Voraus beurteilt werden sollen, so wird eine Beurteilung dadurch erschwert, dass die Realisationen in Bezug auf die Dichteprognosen autokorreliert sind. Man könnte beispielhaft den Fall von zwei Dichteprognosen für die Inflation betrachten, von denen eine im Januar für Juli und eine im Februar für August erstellt wird. Falls die Inflation im Juli deutlich über dem Mittelwert der Prognosedichte vom Januar liegt, dann ist es sehr wahrscheinlich, dass die Inflation im August ebenfalls beträchtlich über dem

Mittelwert der Prognosedichte vom Februar liegt.

Man kann im Wesentlichen drei Ansätze für die Beurteilung von Dichteprognosen in solchen Situationen unterscheiden, die in der Literatur verwendet oder vorgeschlagen werden. Allerdings besitzt jeder dieser Ansätze gewisse Nachteile in Bezug auf die Möglichkeit, fehlerhafte Kalibrierungen zu identifizieren, korrekt kalibrierte Prognosen fälschlicherweise als fehlerhaft zu klassifizieren oder in Bezug auf die Komplexität des Verfahrens. Daher wird in diesem Papier ein alternatives Bewertungsverfahren vorgeschlagen, das über keinen dieser Nachteile verfügt. In Simulationsstudien zeigt sich, dass der neue Ansatz gute Ergebnisse liefert und daher eine brauchbare Alternative zu den bestehenden Verfahren darstellt.

Contents

Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments[1]

1 Introduction

Today, predictions are often made in the form of density forecasts. Tay and Wallis (2000) give a survey of the use of density forecasts in macroeconomics and finance. In contrast to point forecasts, density forecasts contain information about the probability of outcomes. In general, optimal decision-making requires these probabilities in order to minimize expected losses.

Just like point forecasts, density forecasts should be evaluated in order to investigate whether they are correctly specified. Point forecasts, for example, can be tested for bias. Density forecasts, in general, are tested for correct calibration. Correct calibration means that the density forecast coincides with the true density of the predicted variable. If, for example, the observed realizations always occur within a range of only one standard error of the forecast density around the mean forecast, the forecast density is probably too dispersed. A test would be likely to reject the hypothesis of correct calibration in such a situation.

This work is concerned with the question, how an evaluation of density forecasts can be conducted if the percentiles of the realizations, calculated according to the forecast densities, are serially correlated. That is, the situation is studied where, for example, it is very likely that the next realization exceeds the median of the

[1]Author: Malte Knüppel, Deutsche Bundesbank, Research Centre, Wilhelm-Epstein-Straße 14, D-60431 Frankfurt am Main, Germany. E-mail address: malte.knueppel@bundesbank.de
The author would like to thank Jörg Breitung and Karl-Heinz Tödter for helpful comments and suggestions. This paper represents the author's personal opinion and does not necessarily reflect the views of the Deutsche Bundesbank.

forecast density for the next period if the current realization exceeds the median of the forecast density for the current period. Serial correlation of forecast errors, and, thus, of the percentiles of the realizations, is a typical feature of multi-step-ahead forecasts.[2]

The evaluation of density forecasts frequently rests on the transformation of the realizations to percentiles according to the forecast density, as described above. To be more precise, if the density forecasts are calibrated correctly, the probability integral transforms (henceforth PITs) of the realizations should yield the correct percentiles of the realizations, which then are uniformly distributed over the interval $(0, 1)$, as noted by Dawid (1984) and Diebold et al. (1998). The original idea for this evaluation approach dates back at least to Rosenblatt (1952). If the PITs are independent, they can be used directly for testing the calibration of density forecasts, employing, for example, the Kolmogorov–Smirnov test. Applying an inverse normal transformation to the PITs yields, in the case of correctly-calibrated density forecasts, a variable with standard normal distribution (henceforth the INTs, i.e. the inverse normal transforms). This second transformation is often employed, because "there are more tests available for normality, it is easier to test autocorrelation under normality than uniformity, and the normal likelihood can be used to construct likelihood ratio tests." (Wallis, 2007, p. 39).

For one-step-ahead forecasts, the PITs, in addition to uniformity, should display independence.[3] This implies that the PITs, and, consequently, the INTs

[2] The evaluation approaches investigated in this work can of course also be applied to one-step-ahead density forecasts. However, in these cases one often prefers to use tests which simultaneously check for correct calibration of the forecast densities and independence (and, thus, no serial correlation) of the percentiles.

[3] In the words of Mitchell and Wallis (2011), the density forecasts are *completely* calibrated if both conditions are fulfilled.

should not be autocorrelated. The likelihood ratio test proposed by Berkowitz (2001) can be applied to the INTs in order to test simultaneously for zero mean, unit variance, and zero autocorrelation based on a first-order autoregressive model (henceforth AR(1)-model) for the INTs. The approach of Berkowitz (2001), however, does not allow to test for departures from normality. Bao et al. (2007) consider an extension which can accomplish this task.

For multi-step-ahead forecasts, even optimal forecasts produce serially correlated forecast errors, and, thus, serially correlated PITs and INTs. The evaluation of multi-step-ahead forecasts found in the literature, mostly therefore, focuses on correct calibration only.[4] Basically, three approaches can be distinguished.

One approach, proposed by Corradi and Swanson (2006a), uses Kolmogorov-type tests that account for the serial correlation of the data. However, for these tests, critical values are data dependent and therefore, have to be determined individually for each sample under study employing a block bootstrap method.

Another approach rests on normality tests for the INTs which are valid in the presence of serial correlation. Mitchell and Wallis (2011) mention the skewness- and kurtosis-based normality tests proposed by Bai and Ng (2005). Corradi and Swanson (2006b) also suggest, inter alia, the tests proposed by Bai and Ng (2005), and related GMM type tests introduced by Bontemps and Meddahib (2005, 2007). Another test for the normality of time series was proposed by Lobato and Velasco (2004). The tests of Bai and Ng (2005) are, for example, employed by D'Agostino

[4] If an h-step-ahead density forecast is optimal ("Optimal" here means "completely calibrated". More on this follows in Section 3.2.), then its INT at time t are independent of the INTs at time $t \pm (h + i)$ with $i = 0, 1, \ldots$. While this independence property could in principle be tested as well, in practice this is apparently never done. As argued by Corradi and Swanson (2006a), it is important to know whether the density forecast is correctly calibrated, even if the forecasts are not optimal, thereby possibly causing dependence of the INTs.

et al. (forthcoming) for the evaluation of their density forecasts.[5]

Finally, in several applications like those by Clements (2004), Mitchell and Hall (2005), Jore et al. (2010), Bache et al. (2011), and Aastveit et al. (2011) one finds a variant of the test by Berkowitz (2001) adapted to the case of serially correlated INTs. Instead of testing for zero mean, unit variance and zero autocorrelation, only the first two hypotheses enter the test. Thus, no restriction is placed on the autoregressive coefficient of the AR(1)-model.[6]

Unfortunately, each of these approaches has certain disadvantages. The tests by Corradi and Swanson (2006a) are computationally burdensome. Therefore, in practice these tests are hardly applied. Concerning the normality tests proposed above, none of them was originally derived in order to evaluate density forecasts. Therefore, these tests are based on skewness and kurtosis, but ignore the information contained in first and second moments. Since the INTs have a *standard* normal distribution under the null hypothesis of correct calibration, large power gains can, of course, be achieved by considering those moments. Finally, the test by Berkowitz (2001) is based on the assumption of an AR(1)-process. If this assumption is incorrect, as would be expected in the case of, for example, optimal multi-step-ahead forecasts, the standard critical values are not valid, so that the test does not have the correct asymptotic size. Moreover, information from higher-order moments is ignored. It should be noted that, as in the case of the normality tests, the evaluation of multi-step-ahead forecasts is not the intended use of the

[5]To be more precise, D'Agostino et al. (forthcoming) use separate tests for zero skewness and zero excess kurtosis proposed by Bai and Ng (2005), instead of their normality test which uses both moments jointly.

[6]Interestingly, in several studies mentioned in this paragraph, one can also find p-values of additional tests which actually assume serial independence of the PITs and the INTs. Since the actual size of these tests is unknown in the presence of serial correlation, the information content of these p-values remains rather unclear.

test by Berkowitz (2001).

The tests that are proposed in this work do not suffer from any of the disadvantages mentioned, as they use standard critical values, can employ all moments regarded as important, and have correct asymptotic size. Actually, they are closely related to the normality tests mentioned above. Most likely, the skewness- and kurtosis-based tests by Bai and Ng (2005), Bontemps and Meddahib (2005), and Lobato and Velasco (2004) could easily be modified such that hypotheses about lower-order moments are included. However, it seems more obvious to directly consider the raw moments instead of standardized moments for several reasons.

Firstly, certain kinds of misspecifications are more likely to be discovered when tests for raw moments are used. For example, if the forecast density and the true density are normal, but the forecast density has an incorrect variance, this misspecification will show up in the fourth raw moment, but not in the kurtosis. Moreover, the tests based on raw moments are much simpler, because raw moments do not rely on estimates of mean and variance.[7] Finally, the estimators of skewness and kurtosis can be severely biased in small samples, even in the absence of serial correlation, whereas the estimators of raw moments are unbiased.

It should be noted that the effects of parameter estimation uncertainty for the parameters of the forecasting model on the evaluation of density forecasts is not addressed in this work. An excellent treatment of this issue can be found in Chen

[7] Since skewness and kurtosis use these estimates, for example, Bai and Ng (2005) have to estimate a four-dimensional long-run covariance matrix for their normality tests. Bontemps and Meddahib (2005) instead use transformations of the variable under study known as Hermite polynomials. Finally, Lobato and Velasco (2004) derive analytic formulas for the variance of skewness and kurtosis which take the estimation uncertainty for mean and variance into account. In contrast to that, with raw moments, the two-dimensional long-run covariance matrix of the third and fourth moment could be used directly.

(2011).[8]

2 Calibration Tests for Density Forecasts

Let the variable of interest be denoted by x_t and the forecast density for this variable in period t by $\hat{f}(x_t)$, where the forecast was made in period $t-h$, and h is a positive integer. The PIT proposed by Rosenblatt (1952) is given by

$$u_t = \hat{F}(x_t) = \int_{-\infty}^{x_t} \hat{f}(q)\, dq$$

where $\hat{F}(x_t)$ denotes the forecast distribution function associated with $\hat{f}(x_t)$. If the forecast density $\hat{f}(x_t)$ is equal to the true density $g(x_t)$,[9] then u_t is uniformly distributed over the interval $(0,1)$. The INT used by Berkowitz (2001) yields

$$z_t = \Phi^{-1}(u_t) = \Phi^{-1}\left(\hat{F}(x_t)\right)$$

where Φ^{-1} is the inverse of the standard normal distribution function. As stated in Berkowitz (2001), the density of z_t is given by

$$p(z_t) = \frac{g(x_t)}{\hat{f}(x_t)}\phi(z_t)$$

[8] Using the formulas (34) and (35) given in Chen (2011), which are based on West and McCracken (1998), it should be fairly easy to adapt the raw-moments tests proposed in what follows to the case where a forecasting model with estimated parameters is to be evaluated according to its out-of-sample density forecasts.

[9] Note that there might be more than one true density, depending on the conditioning information. Gneiting et al. (2007) give the example of a random variable m_t that equals $m_t = n_t + \varepsilon_t$. Let $N(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2. If n_t and ε_t are independently $N(0,1)$ distributed, then $N(0,2)$ and $N(n_t,1)$ are both true densities of m_t. $N(0,2)$ is the unconditional density and $N(n_t,1)$ the density conditional on n_t. So, both, $N(0,2)$ and $N(n_t,1)$, could serve as correctly-calibrated density forecasts for m_t.

where $\phi(\cdot)$ denotes the standard normal density function and $x_t = \hat{F}^{-1}(\Phi(z_t))$. In the following, I will describe the two tests typically used in empirical macroeconomic applications and an alternative approach based on raw moments.

2.1 The Test of Berkowitz (2001)

Berkowitz (2001) suggests estimating the equation

$$z_t = \mu + \rho z_{t-1} + \varepsilon_t$$

with $t = 1, 2, \ldots, T$ and $\varepsilon_t \sim N(0, \sigma^2)$, so that the log-likelihood function is given by

$$\begin{aligned} \ln L &= c(T) - \frac{1}{2}\ln\left(\frac{\sigma^2}{1-\rho^2}\right) - \left(z_1 - \frac{\mu}{1-\rho}\right)^2 \left(\frac{1-\rho^2}{2\sigma^2}\right) \\ & -\frac{T-1}{2}\ln\left(\sigma^2\right) - \sum_{t=2}^{T}\left(\frac{\left(z_t - \mu - \rho z_{t-1}\right)^2}{2\sigma^2}\right). \end{aligned}$$ [10]

Denoting the maximum-likelihood estimates with a hat, a joint test of correct calibration and independence of the INTs can be conducted using the likelihood ratio test statistic

$$\hat{\beta}_{12}^{ind} = 2\left(\ln L\left(\hat{\mu}, \hat{\sigma}, \hat{\rho}\right) - \ln L\left(0, 1, 0\right)\right)$$

which converges to a $\chi^2(3)$-distribution under the null hypothesis. For multi-step-ahead forecasts, instead, the test statistic

$$\hat{\beta}_{12} = 2\left(\ln L\left(\hat{\mu}, \hat{\sigma}, \hat{\rho}\right) - \ln L\left(0, \sqrt{1-\hat{\rho}^2}, \hat{\rho}\right)\right)$$

and, correspondingly, a $\chi^2(2)$-distribution are used by practitioners in order to test for correct calibration.[11] As mentioned above, this test is frequently employed, probably not least because of its simplicity. Henceforth, we will refer to this test as the $\hat{\beta}_{12}$ test.

2.2 The Test of Bai and Ng (2005)

Bai and Ng (2005) actually propose two similar tests for the normality of time series which are based on the skewness and kurtosis of z_t. I will focus on the test that Bai and Ng (2005) appear to prefer because of higher power, and that, consequently, they use in their empirical application.[12] The test is based on the statistic

$$\hat{\mu}_{34} = \hat{\mathbf{a}}' \left(\hat{\mathbf{\Psi}} \hat{\mathbf{\Xi}} \hat{\mathbf{\Psi}}' \right)^{-1} \hat{\mathbf{a}}$$

where $\hat{\mathbf{a}}$ is the vector given by

$$\hat{\mathbf{a}} = \begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (z_t - \bar{z})^3 \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left((z_t - \bar{z})^4 - 3 \left(\hat{\sigma}^2 \right)^2 \right) \end{bmatrix},$$

[11] Of course, the values of the parameters will depend on the forecast horizon h under study.

[12] There is a typing error in Tables 4 and 5 in Bai and Ng (2005). While in their article, one reads that "The $\hat{\mu}_{34}$ test is generally more powerful than the $\hat{\pi}_{34}$ test" (Bai and Ng, 2005, p. 55), Table 4 apparently shows that the $\hat{\pi}_{34}$ test tends to reject more often than the $\hat{\mu}_{34}$ test, and for the empirical application in Table 5, only the $\hat{\pi}_{34}$ test is used. When replicating parts of Bai and Ng's (2005) Monte Carlo simulations and empirical applications, it turns out that the typing error occurred in the tables, not in the text. So as stated in their text, the $\hat{\mu}_{34}$ test rejects more often than the $\hat{\pi}_{34}$ test, and it is actually the $\hat{\mu}_{34}$ test which is used in the empirical application.

Although Bai and Ng (2005) prefer the $\hat{\mu}_{34}$ test because of power reasons, it should be noted that this test tends to overreject under the null hypothesis, whereas the $\hat{\pi}_{34}$ test tends to under-reject. So the $\hat{\pi}_{34}$ test could actually be superior in terms of size-adjusted power.

with $\bar{z}$ and $\hat{\sigma}^2$ being consistent estimates of mean and variance of z_t, respectively. $\hat{\mathbf{\Psi}}$ is given by

$$\hat{\mathbf{\Psi}} = \begin{bmatrix} -3\hat{\sigma}^2 & 0 & 1 & 0 \\ 0 & -6\hat{\sigma}^2 & 0 & 1 \end{bmatrix}$$

and $\hat{\mathbf{\Xi}}$ is the long-run covariance matrix of the vector series $\hat{\mathbf{b}}_t$ defined by

$$\hat{\mathbf{b}}_t = \begin{bmatrix} z_t - \bar{z} \\ (z_t - \bar{z})^2 - \hat{\sigma}^2 \\ (z_t - \bar{z})^3 \\ (z_t - \bar{z})^4 - 3\hat{\sigma}^4 \end{bmatrix}.$$

The long-run covariance matrix $\hat{\mathbf{\Xi}}$ can be consistently estimated using the approaches of Newey and West (1987) and Andrews (1991). $\hat{\mu}_{34}$ converges to a $\chi^2\,(2)$-distribution under the null hypothesis. In what follows, the test by Bai and Ng (2005) will be referred to as the $\hat{\mu}_{34}$ test.

2.3 Alternative Tests Based on Raw Moments

The major complications when testing higher-order moments as in the case of Bai and Ng (2005) arise from the fact that the lower-order moments are unknown. Therefore, a four-dimensional covariance matrix is needed for a joint test of only two moments, skewness and kurtosis. When testing for *standard* normality, however, also the lower-order moments are known under the null hypothesis. Therefore, one does not need to consider standardized moments like skewness and kurtosis. It is not even necessary to employ central moments like the variance. Instead, non-standardized, non-central moments, i.e. the raw moments can be used, so that tests can be constructed very easily.

Actually, the tests do not have to be based on the standard normal distribution, but any suitable transformation of the PITs can be used. Denote the transformed variables by

$$y_t = H\left(u_t\right)$$

where $H\left(u_t\right) = \Phi^{-1}\left(u_t\right)$ would yield standard normally distributed variables y_t. Let the $r-$th raw moment of y_t be denoted as

$$m_r = E\left[y_t^r\right]$$

and define the vector $\hat{\mathbf{D}}_{r_1 r_2 \ldots r_N}$ of the difference between the N empirical raw moments of interest $(\hat{m}_{r_1}, \hat{m}_{r_2}, \ldots, \hat{m}_{r_N})$ and the corresponding expected raw moments of $y_t = H\left(u_t\right)$ if u_t is uniformly distributed $(m_{r_1}, m_{r_2}, \ldots, m_{r_N})$ as

$$\hat{\mathbf{D}}_{r_1 r_2 \ldots r_N} = \begin{bmatrix} \hat{m}_{r_1} - m_{r_1} \\ \hat{m}_{r_2} - m_{r_2} \\ \vdots \\ \hat{m}_{r_N} - m_{r_N} \end{bmatrix},$$

where $\hat{m}_{r_i}$ is simply given by the sample mean $\hat{m}_{r_i} = \frac{1}{T}\sum_{t=1}^{T} y_t^{r_i}$ for $i = 1, 2, \ldots, N$. For convenience, I assume that the moments are ordered such that $r_1 < r_2 < \ldots < r_N$. Then a test for the distributional assumption for y_t can be based on the statistic

$$\hat{\alpha}_{r_1 r_2 \ldots r_N} = T\hat{\mathbf{D}}'_{r_1 r_2 \ldots r_N}\hat{\mathbf{\Omega}}^{-1}_{r_1 r_2 \ldots r_N}\hat{\mathbf{D}}_{r_1 r_2 \ldots r_N} \tag{1}$$

where $\hat{\mathbf{\Omega}}_{r_1 r_2 \ldots r_N}$ is the long-run covariance matrix of the vector series

$$\mathbf{d}_t = \begin{bmatrix} y_t^{r_1} - m_{r_1} \\ y_t^{r_2} - m_{r_2} \\ \vdots \\ y_t^{r_N} - m_{r_N} \end{bmatrix}.$$

Supposing that the central limit theorem holds for $\mathbf{d}_t$,[13] the test statistic $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ converges to a $\chi^2(N)$ distribution under the null.

If the transformation $y_t = \Phi^{-1}(u_t)$ is employed, under the null hypothesis this test is asymptotically equivalent to the $\hat{\mu}_{34}$ test if one chooses to use $r_1 = 3$ and $r_2 = r_N = 4$, so that $\mathbf{d}_t = \begin{bmatrix} y_t^3, & y_t^4 - 3 \end{bmatrix}'$ and the test statistic $\hat{\alpha}_{34}$ is obtained. However, important differences in power are to be expected. For example, if $y_t \sim N(0, \sigma^2)$ with $\sigma^2 \neq 1$, $\hat{\mu}_{34}$ converges to a $\chi^2(2)$-distribution while $\hat{\alpha}_{34}$ does not, because $E[y_t^4] = 3\sigma^4$ continues to hold, whereas $E[y_t^4] = 3$ does not. In addition, the common estimators for skewness and kurtosis as used for the $\hat{\mu}_{34}$ test can be strongly biased in small samples, whereas the r_i-th raw moment is estimated unbiasedly by $\hat{m}_{r_i}$. Note that, for the test based on $\hat{\alpha}_{34}$, only a two-dimensional covariance matrix $\hat{\mathbf{\Omega}}_{34}$ has to be estimated.

If the transformation $y_t = \Phi^{-1}(u_t)$ and $r_1 = 1$ and $r_2 = r_N = 2$ are employed, so that $\mathbf{d}_t = \begin{bmatrix} y_t, & y_t^2 - 1 \end{bmatrix}'$, the test is similar to the $\hat{\beta}_{12}$ test, because both tests are based on the first and second moment of the INTs. Yet, the $\hat{\beta}_{12}$ test assumes an AR(1)-process for z_t, whereas the $\hat{\alpha}_{12}$ test accounts for general forms of serial

[13] So, it is assumed that the moments of z_t are finite up to order $2r_N$ which is unproblematic if, for example, $H(u_t) = \Phi^{-1}(u_t)$ or $H(u_t) = u_t$. The asymptotic normality of the first four sample raw moments, if applied to linear processes with normal disturbances, is shown by Lomnicki (1961, Section 4).

correlation. The tests will, of course, have different power properties. Berkowitz (2001) mentions that the likelihood ratio test is the optimal test against one-sided alternatives in the case of independent observations, so that its power is known to be larger at least in certain situations.

In addition to allowing for general forms of serial correlation, another advantage of the tests proposed here is that the moments can be chosen flexibly according to the given circumstances. For example, if only a small sample of density forecasts is available, it is rather unlikely that the inclusion of higher-order moments is helpful, because they can increase size distortions and decrease power. In very small samples, one might actually just want to set $r_1 = r_N = 1$. One could also imagine situations where only certain moments are of interest, for example $r_1 = 1$ and $r_2 = 3$.[14] In larger samples, an obvious choice might be $r_i = i$ with $i = 1, 2, 3, 4$, although, of course, moments of even higher order could also be included.

If the transformed variables $y_t = H(u_t)$ have a density that is symmetric around 0, there is an alternative approach that, asymptotically, leads to the same results as the tests described above, but might behave differently in small samples. This approach is based on the fact that the long-run covariance of $y_t^{r_i} - m_{ri}$ and $y_t^{r_j} - m_{r_j}$ equals 0 if y_t is symmetrically distributed around 0 and if $r_i + r_j$ is odd. A proof of this property is given in Appendix A.

Based on this property, one can construct an alternative test statistic $\hat{\alpha}^0_{r_1 r_2 \dots r_N}$ as the sum of two test statistics

$$\hat{\alpha}^0_{r_1 r_2 \dots r_N} = \hat{\alpha}^{odd}_{r_1 r_2 \dots r_N} + \hat{\alpha}^{even}_{r_1 r_2 \dots r_N} \tag{2}$$

[14] This might, for example, be the case if the forecast densities are normal. In this case, a test using $r_1 = 1$ and $r_2 = 3$ could be employed to check whether the true densities are symmetric and have the same means as the forecast densities.

where $\hat{\alpha}^{odd}_{r_1 r_2 \ldots r_N}$ uses all odd raw moments and $\hat{\alpha}^{even}_{r_1 r_2 \ldots r_N}$ all even raw moments from the set of sample moments $\{\hat{m}_{r_i}\}$ which are considered for the test. $\hat{\alpha}^{odd}_{r_1 r_2 \ldots r_N}$ and $\hat{\alpha}^{even}_{r_1 r_2 \ldots r_N}$ are calculated in the same way as the test statistic $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ in (1), but only using the odd and even moments, respectively.[15] $\hat{\alpha}^{0}_{r_1 r_2 \ldots r_N}$ converges to a $\chi^2(N)$ distribution under the null.

So, if at least one even and at least one odd moment are involved in the test, two alternative test statistics can be used. If, for example, the first and second moment are to be used, so that $r_1 = 1$ and $r_2 = r_N = 2$, and these moments equal $E[y_t] = 0$ and $E[y_t^2] = 1$, the test statistics

$$\hat{\alpha}_{12} = T \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} y_t \\ \frac{1}{T}\sum_{t=1}^{T}(y_t^2 - 1) \end{pmatrix}' \begin{bmatrix} \hat{\sigma}^2_{y_t} & \hat{\sigma}_{y_t, y_t^2-1} \\ \hat{\sigma}_{y_t, y_t^2-1} & \hat{\sigma}^2_{y_t^2-1} \end{bmatrix}^{-1} \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} y_t \\ \frac{1}{T}\sum_{t=1}^{T}(y_t^2 - 1) \end{pmatrix}$$

and

$$\hat{\alpha}^0_{12} = \frac{T}{\hat{\sigma}^2_{y_t}} \left(\frac{1}{T} \sum_{t=1}^{T} y_t \right)^2 + \frac{T}{\hat{\sigma}^2_{y_t^2-1}} \left(\frac{1}{T} \sum_{t=1}^{T} (y_t^2 - 1) \right)^2$$

are obtained, where $\hat{\sigma}^2_x$ denotes the sample long-run variance of x, and $\hat{\sigma}_{x,y}$ denotes the sample long-run covariance of x and y. Both test statistics have the same asymptotic distribution (the $\chi^2(2)$-distribution) under the null, because $\hat{\sigma}_{y_t, y_t^2-1}$ converges to 0.[16]

[15] Alternatively, one could simply set the respective elements of $\hat{\mathbf{\Omega}}_{r_1 r_2 \ldots r_N}$ in (1) to zero if the truncation lag for the long-run covariance matrix is fixed a priori. However, with an automatic truncation lag selection procedure, the latter approach might be problematic.

[16] Actually, in Bai and Ng (2005), the test statistics $\hat{\pi}_{34}$ and $\hat{\mu}_{34}$ share a similar relation as $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ and $\hat{\alpha}_{r_1 r_2 \ldots r_N}$. In contrast to the calculation of $\hat{\mu}_{34}$, for the calculation of $\hat{\pi}_{34}$, information about asymptotic covariances equal to zero is used.

3 Monte Carlo Simulation Setup

3.1 The Densities

In order to assess the size and power properties of the tests presented, Monte Carlo simulations are used, where it is assumed that the density of the variable

$$x_t \sim N(0,1)$$

is to be predicted. The x_t's are identically, but not necessarily independently distributed, so that, in general, $\phi(x_t | x_{t-1}) \neq \phi(x_t)$ holds.

For the misspecified density forecasts, we consider normal, two-piece-normal, Student's t and normal mixture distributions. The normal distribution is employed to create correctly calibrated density forecasts, or forecasts whose expectation or variance differ from the true values of 0 and 1, respectively. The two-piece normal distribution is employed to construct density forecasts with correct expectation and variance, but with incorrect skewness and kurtosis.[17] In order to construct density forecasts with correct expectation, variance and skewness, but incorrect kurtosis, Student's t distribution is employed. Finally, the normal mixture distribution is set up such that its first four moments are identical to those of a standard normal distribution while the shapes of both densities differ markedly. Note that all densities except for the normal ones are standardized such that they have an expectation of 0 and a variance of 1. In what follows, the densities are described in more detail. Unless otherwise mentioned, the skewness of the densities presented equals 0 and their kurtosis equals 3.

[17] The two-piece-normal density is a relatively popular forecast density among central banks. For a survey, see Knüppel and Schultefrankenfeld (forthcoming).

The normal density forecast is given by

$$\hat{f}(x_t) = \frac{1}{\sigma}\phi\left(\frac{x_t - \mu}{\sigma}\right)$$

where μ is the forecast mean and σ the forecast standard deviation of x_t. Of, course, with $\mu = 0$ and $\sigma = 1$, the standard normal density is obtained.

The two-piece normal distribution, as described, for example, in Wallis (2004, p. 66), is defined by

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

with

$$A = \frac{2}{\sqrt{2\pi}\,(\sigma_1 + \sigma_2)}$$

and the forecast moments

$$\begin{aligned} E[x_t] &= \mu = m + \sqrt{\frac{2}{\pi}}\,(\sigma_2 - \sigma_1) \\ E\left[(x_t - \mu)^2\right] &= \mu_2 = \left(1 - \frac{2}{\pi}\right)(\sigma_2 - \sigma_1)^2 + \sigma_1\sigma_2. \end{aligned}$$

So, setting the mode m to

$$m = \sqrt{\frac{2}{\pi}}(\sigma_1 - \sigma_2)$$

guarantees that the forecast mean equals 0. Moreover, choosing

$$\begin{aligned}
\sigma_1 &= \sqrt{\left(1-\frac{3}{8}\pi\right)\gamma^2+1}-\gamma\sqrt{\frac{1}{8}\pi} \\
\sigma_2 &= \sqrt{\left(1-\frac{3}{8}\pi\right)\gamma^2+1}+\gamma\sqrt{\frac{1}{8}\pi}
\end{aligned}$$

with

$$\gamma = \mu - m$$

makes the forecast variance equal to 1, and the forecast density then only depends on γ.[18] The parameter γ controls the asymmetry of the density and represents the mean-mode difference.[19] Its possible values are restricted to the interval $\left[-\sqrt{\frac{2}{\pi-2}}, \sqrt{\frac{2}{\pi-2}}\right]$. A positive value of γ corresponds to a positively-skewed random variable x_t. Skewness and kurtosis of the standardized two-piece normal distribution are given by

$$\begin{aligned}
s &= E\left[x_t^3\right] = \left((3-\pi)\gamma^2+1\right)\gamma \\
k &= E\left[x_t^4\right] = ((22-3\pi)\pi-40)\frac{\gamma^4}{4}+(3\pi-8)\frac{\gamma^2}{2}+3.
\end{aligned}$$

With $\gamma = 0$, $\hat{f}(x_t)$ becomes the standard normal density.

Let $\tau(x_t, v)$ denote the density function of the t-distribution with v degrees of freedom, where I assume that $v > 4$ holds. Using this density, the forecast variance

[18] In order to see that the variance equals 1, simply insert the expressions for σ_1 and σ_2 into the expression for the variance $E\left[(x_t-\mu)^2\right]$.

[19] Since the variance equals $\mu_2 = 1$, γ is also equal to the Pearson mode skewness defined by $(\mu - m)/\sqrt{\mu_2}$.

μ_2 would equal

$$\mu_2 = \frac{v}{v-2}.$$

In order to obtain a forecast density with unit variance, the scaled forecast density given by

$$\hat{f}(x_t) = \left(\frac{v}{v-2}\right)^{-\frac{1}{2}} \tau\left(x_t \left(\frac{v}{v-2}\right)^{-\frac{1}{2}}, v\right)$$

is employed. The kurtosis of x_t equals

$$k = \frac{3v-6}{v-4}.$$

When v approaches infinity, $\hat{f}(x_t)$ converges to the standard normal density.

Finally, the normal mixture density considered is given by

$$\hat{f}(x_t) = \frac{1}{6\sigma}\phi\left(\frac{x_t+m}{\sigma}\right) + \frac{4}{6\sigma}\phi\left(\frac{x_t}{\sigma}\right) + \frac{1}{6\sigma}\phi\left(\frac{x_t-m}{\sigma}\right)$$

with

$$m = \sqrt{3\left(1-\sigma^2\right)},$$

and with $\sigma \in (0,1]$.[20] The standard normal density emerges if $\sigma = 1$. This density becomes trimodal if σ is sufficiently small.

It should be noted that differences between the moments of the true standard normal density and the forecast density, in general, do not translate one-to-one into differences between a standard normal density and the density of the INTs. For example, while the forecast density based on the t-distribution has a unit variance, the variance of the corresponding INTs can differ from 1. Yet, if the true density

[20] Actually, the condition on m only serves to obtain a unit variance. Even if m is chosen arbitrarily, the skewness and kurtosis of x_t continue to equal 0 and 3, respectively.

and the forecast density are symmetric and have the same expectation, the density of the INTs is also symmetric.

3.2 The Simulation Environment

It is well known that optimal h-step-ahead point forecasts lead to forecast errors following a moving-average process of order $h-1$ (henceforth MA($h-1$)-process).[21] Similarly, if an h-step-ahead density forecast for x_t is correctly calibrated and uses all the information relevant for the determination of x_t, i.e. if it is completely calibrated according to the terminology of Mitchell and Wallis (2011), it must hold that

$$f\left(y_t \left| y_{t-i}\right.\right) \begin{cases} \neq f\left(y_t\right) & i = 1, 2, \ldots, h-1 \\ = f\left(y_t\right) & i = h, h+1, \ldots \end{cases} . \tag{3}$$

where $f(\bullet)$ denotes the density of y_t. Formula (3) and the fact that y_t has a Wold representation suggest that, in the case of completely calibrated h-step-ahead density forecasts, the y_t's follow an MA($h-1$)-process. For example, in the case of completely calibrated h-step-ahead density forecasts for the linear and normal process

$$x_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

with $\varepsilon_t \sim N\left(0, \sigma^2\right)$, the INTs are actually described by the MA($h-1$)-process

$$y_t = z_t = \frac{1}{\sigma \sqrt{\sum_{i=0}^{h-1} b_i^2}} \sum_{i=0}^{h-1} b_i \varepsilon_{t-i}.^{22}$$

[21] See, for example, Diebold (1998, p. 341). Optimality here refers to the minimization of the mean squared forecast error.

[22] Yet, for example in the case of the PITs, it is not clear how the shocks of y_t are related to the shocks of the process for x_t.

Therefore, in what follows, an MA(1)-process is used in order to generate dependent standard normal variables x_t, so that x_t evolves according to

$$x_t = \varepsilon_t + \rho\varepsilon_{t-1}$$

with $\varepsilon_t \sim N\left(0, \left(1+\rho^2\right)^{-1}\right)$ for $t = 1, 2, \ldots T$. However, an AR(1)-process is also considered. In this case, x_t is determined by

$$x_t = \rho x_{t-1} + \varepsilon_t$$

with $\varepsilon_t \sim N\left(0, 1-\rho^2\right)$.

If the forecast density is standard normal, the MA(1)-process leads to y_t's which correspond to those of completely calibrated 2-step-ahead density forecasts, whereas the AR(1)-process produces y_t's which are closely related to completely calibrated h-step-ahead density forecasts only if h is sufficiently large and if the data-generating process is an AR(1)-process.

The tests considered are the $\hat{\beta}_{12}$ test, the $\hat{\mu}_{34}$ test, and various raw-moments tests based on $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ and $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$.[23] The parameters for the $\hat{\beta}_{12}$ test are estimated by maximum likelihood. For the $\hat{\mu}_{34}$ and the raw-moments tests, the covariance matrices $\hat{\boldsymbol{\Xi}}$ and $\hat{\boldsymbol{\Omega}}_{r_1 r_2 \ldots r_N}$ are estimated under the null hypothesis. That is, the covariances are determined without subtracting the estimated means of the vector series $\hat{\mathbf{b}}_t$ and $\mathbf{d}_t$, which both have an expectation of 0 under the null. With this approach we follow Bai and Ng (2005).[24] Subtracting the empirical mean

[23] It should be noted that the data-generating AR(1)-process corresponds to the assumption used by the $\hat{\beta}_{12}$ test, so that this test is expected to perform well in the corresponding simulation studies.

[24] This is not evident from the article itself, but becomes clear from the GAUSS codes provided

would tend to increase the size distortions of the tests, but also improve their size-adjusted power.

Concerning the raw-moments tests, the most parsimonious test is only based on the first moment. Tests with power against more types of density misspecification are obtained by consecutively adding higher moments. Wherever it is possible, both test statistics, $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ and $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$, are employed. The largest moment order considered is four. This yields the seven test statistics $\hat{\alpha}_1$, $\hat{\alpha}^0_{12}$, $\hat{\alpha}_{12}$, $\hat{\alpha}^0_{123}$, $\hat{\alpha}_{123}$, $\hat{\alpha}^0_{1234}$, and $\hat{\alpha}_{1234}$.

The raw-moments tests could be applied to any transformation of u_t yielding random variables with a distribution that is symmetric around 0. Natural candidates are the INTs and a standardized version of the PITs. The standardized PITs (henceforth S-PITs) are obtained as

$$y_t = \sqrt{12}\left(u_t - \frac{1}{2}\right).$$

In the case of correctly calibrated density forecasts, the S-PIT is a standard uniformly distributed random variable, i.e. a uniformly distributed variable with an expectation of 0 and a variance of 1. Moreover, its skewness and kurtosis continue to equal 0 and 1.8, respectively. Hence, the third and fourth raw moment also equal 0 and 1.8, respectively. The density of y_t is given by

$$f\left(y_t\right) = \begin{cases} \frac{1}{\sqrt{12}} & -\sqrt{3} \leq y_t \leq \sqrt{3} \\ 0 & \text{else} \end{cases}$$

in the case of correctly calibrated density forecasts. Otherwise, $f\left(y_t\right)$ will differ

by the authors.

from this functional form, but positive values of the density will, of course, continue to be restricted to the interval $-\sqrt{3} \leq y_t \leq \sqrt{3}$. Other potential transformations of u_t will be discussed in Section 5.

In order to facilitate meaningful comparisons between the test statistics, above all between the test statistics $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ and $\hat{\alpha}_{r_1 r_2 \ldots r_N}$, the *size-adjusted* power of the tests will be reported. This requires a reasonably precise estimation of their actual sizes. The 95% interval for the rate of the type I error $\hat{\delta}$ estimated from the simulations has an approximate width of $1.96\sqrt{\frac{\delta(1-\delta)}{N}}$, where N is the number of Monte Carlo simulations and δ is the true rate of the type I error. Setting $N = 200,000$ yields an accuracy that appears satisfactory for the given purpose. With a nominal significance level of $\delta^* = 5\%$, if the size distortions are not too severe, the 95% interval then has a width of about 0.001.[25] The critical value of the test statistics which is used for the power computations is determined by the 95% quantile of the $200,000$ test statistics computed under the null hypothesis. For the power computations, the number of Monte Carlo simulations is set to $10,000$ which corresponds to a maximal width of the 95% interval for the estimated rejection probability $\hat{\lambda}$ equal to 0.01.[26]

The sample sizes T considered are 50, 100, 200, 500, and 1000. The autoregressive and moving-average parameters ρ take on the values 0, 0.5, and 0.9. As suggested by Andrews (1991), the quadratic spectral kernel is used for the estimation of the long-run covariance matrix.[27] The truncation lag is also chosen

[25] The largest possible width equals 0.002. This value is attained if $\delta = 0.5$, i.e. in the rather extreme situation where, given a nominal significance level of $\delta^* = 5\%$, the test suffers from strong size distortions and rejects ten times more often than expected under the null.

[26] Of course, this value is attained if the true power equals $\lambda = 0.5$. In the calculation of the width, the uncertainty concerning the simulated critical values is ignored.

[27] Employing the Bartlett kernel as suggested by Newey and West (1987) instead of the quadratic spectral kernel tends to produce only slightly larger size distortions. Results are

according to Andrews (1991).[28]

The first misspecified normal forecast density considered has an expectation of $\mu = -0.5$ and unit variance. The next two misspecified normal forecast densities have expectations of 0, but their standard deviations $\sqrt{\mu_2} = \sigma$ equal $\frac{2}{3}$ and $\frac{3}{2}$, respectively. The mean-mode difference γ of the following standardized two-piece normal forecast density is equal to 0.8. The standardized density of the t-distribution has 5 degrees of freedom.[29] Finally, the standardized normal mixture density uses the parameter value $\sigma = 0.4$. All densities employed, their corresponding INTs, and standard normal densities are displayed in Figure 1. Plots of the S-PITs can be found in Figure 2. In the case of correctly-calibrated density forecasts, the density of the S-PITs would be flat and attain a value of about $0.3 \approx 1/\sqrt{12}$.

Assuming normality of the error terms and the (potential) non-normality of the forecast densities instead of the opposite (non-normal errors and normal forecast densities) has two advantages.[30] Firstly, the normality of the error terms has the convenient implication that the unconditional distribution of the data, i.e. of x_t, is always normal and does not change with the serial correlation. Secondly, computational problems are more likely if the forecast densities are normal. Such

available upon request.

[28] Prewhitening as suggested in Andrews and Monahan (1992) is not used due to the fact that the y_t's have a moving average representation of order $h-1$ in the case of complete calibration, whereas standard prewhitening procedures employ autoregressive processes.

[29] This value is chosen arbitrarily. It is the smallest integer value for which the fourth moment exists. Yet, the existence of moments is not required for the forecast density.

[30] Actually, the opposite case, i.e. normal forecast densities and non-normal error terms is probably more relevant from an empirical point of view. However, apart from the serial correlation the test results only depend on the densities of the INTs and SPITs. Hence, it does not matter whether, for example, the densities of the INTs and SPITs are the result of a two-piece normal forecast density and a normally distributed variable, or whether the same densities arise from a normal forecast density and a corresponding non-normal variable.

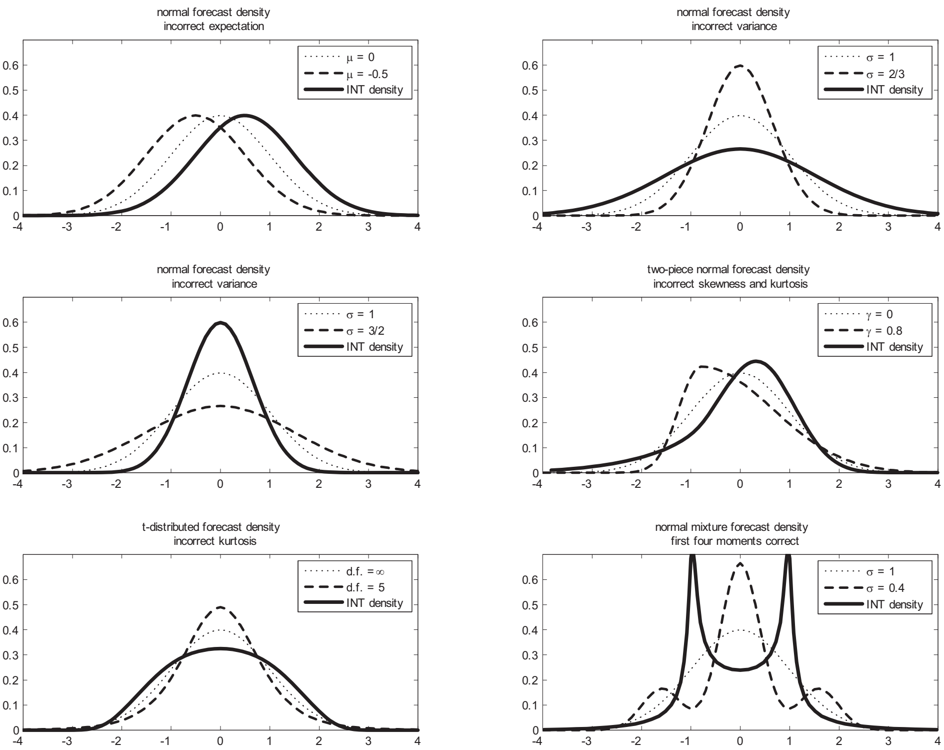

Figure 1: Misspecified forecast densities, the true standard normal forecast densities, and the corresponding INTs (i.e. the inverse normal transforms of the probability integral transforms)

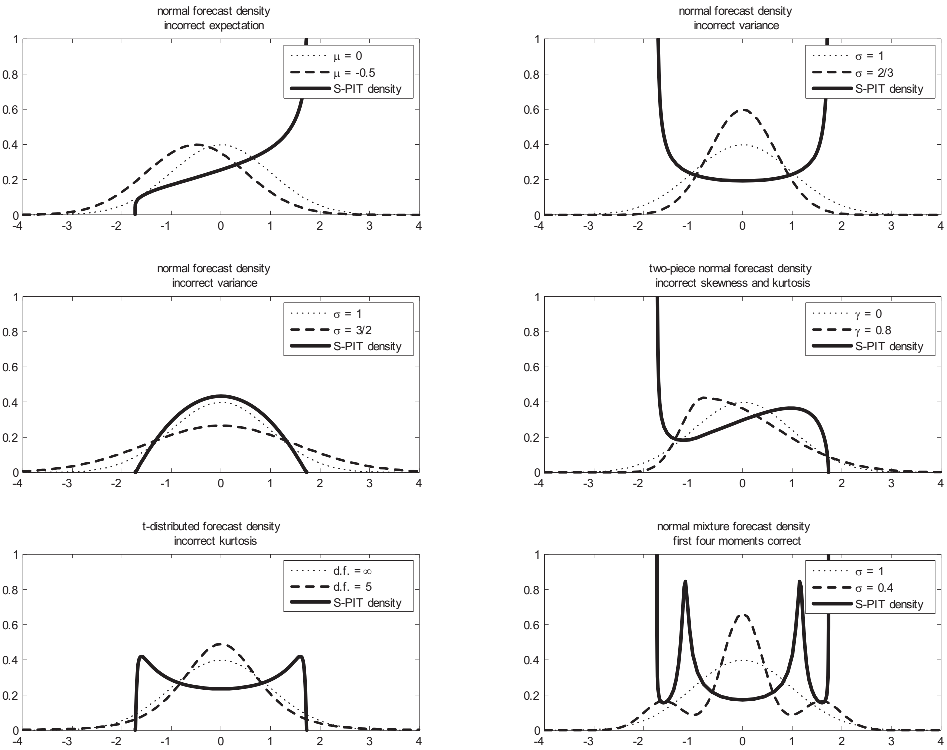

Figure 2: Misspecified forecast densities, the true standard normal forecast densities, and the corresponding S-PITs (i.e. the standardized probability integral transforms)

problems can arise if realizations occur in regions where the forecast density implies virtually zero probability. In the case of the the thin-tailed normal distribution as a forecast density, for example, the very small or very large random numbers coming from a t-distribution are likely to lead to values of y_t, i.e. of the PITs, which are so close to 0 or 1 that the computer rounds them to exactly 0 or 1, respectively. In these cases, a following inverse normal transformation is not feasible.[31]

4 Simulation Results

4.1 Size

If a standard normal forecast density is used, so that the densities are correctly calibrated, the actual size of the tests under study should be equal to the nominal size asymptotically. Yet, in small samples, both quantities can differ markedly. In Tables 1 and 2, the actual sizes of the $\hat{\beta}_{12}$ test, the $\hat{\mu}_{34}$ test, and the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ as well as the $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ tests can be found. Table 1 contains the results for the raw-moments tests being based on the INTs, whereas Table 2 contains the results if the S-PITs are used. The following statements concerning the size distortions refer to the absolute differences between the nominal and the actual size, unless

[31] Interestingly, the computer's floating point system for representing real numbers also implies a second form of potential problems. The computer's (absolute) accuracy decreases the more a number differs from 0. Thus, the PITs are more 'strongly' rounded near 1 than near 0. This has the effect that especially the INTs can appear asymmetric (to be more precise, skewed to the left), although forecast and true density are actually symmetric around their identical expected values. In the setup chosen here, this effect could be observed if the symmetric normal mixture density was used with σ being very small. For example, with $\sigma = 0.2$, the third raw moment of the INTs calculated by Monte Carlo simulations equals -0.17 instead of its true value of 0.

otherwise mentioned.[32]

The most notable observation concerning the actual sizes is given by the sometimes large size distortions of the raw-moments tests if they are based on the INTs. It is well-known that, for example, the distribution of the sample kurtosis estimator is far from normal even for relatively large sample sizes.[33] Table 1 shows that for many raw-moments tests, the size distortions can become relatively large. Even with the $\hat{\alpha}^0_{12}$ test, i.e. if only the first and second raw moment are considered, and the zero-long-run covariance property is used, the actual size can reach almost 10 percent if 200 observations are available and the persistence is strong (i.e. in the case of an AR(1)-process with $\rho = 0.9$). The $\hat{\alpha}^0_{123}$ test performs slightly better in most cases, but still has an actual size of almost 8 percent in the case of 500 observations and strong persistence.[34] If fourth moments are employed, the size distortions can become huge. If 200 observations are available, the actual sizes of the $\hat{\alpha}_{1234}$ test and the $\hat{\alpha}^0_{1234}$ test range from 13 to 35 percent, depending on the persistence parameter ρ.

The size distortions of the raw-moments tests based on the S-PITs, in contrast, are fairly contained according to Table 2. As in the case of the INTs, the size distortions are, in general, smaller if the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ tests are used instead of the $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ tests. In this case, the largest negative size distortions are observed for the case of 50 observations and strong persistence with actual sizes often being below 1 percent. The largest positive size distortion is again recorded for 200

[32] That is, for example, an actual size of 0.01 is considered a size distortion equal to 4 percentage points, thereby being larger than the size distortion associated with an actual size of 0.08, which is equal to 3 percentage points.

[33] Moreover, sample skewness and sample kurtosis of normal variables are uncorrelated, but not independent in small samples. For more details, see, for example, Doornik and Hansen (2008) and the references therein.

[34] The actual size slightly exceeds 8 percent, for example, if $\rho = 0.9$ and $250 \leq T \leq 450$.

observations and strong persistence, where the $\hat{\alpha}^0_{1234}$ test has an actual size of 7.3 percent.[35] Therefore, in what follows only the raw-moments tests based on the S-PITs will be considered.

For all tests considered, in general, the size distortions increase with ρ and decrease with T. There can be exceptions to this rule when tests, for example, are undersized in small samples and oversized in larger samples. In these cases, there appears to be an intermediate sample size where size distortions can be close to 0.

If the forecast variable follows an AR(1)-process with no or only moderate persistence, in general, the $\hat{\beta}_{12}$ test yields the smallest size distortions. In small samples with strong persistence, however, even this test has an actual size of more than 9 percent. Given an MA(1)-process, the $\hat{\beta}_{12}$ test suffers from size distortions which do not vanish asymptotically. With $\rho = 0.9$, its actual asymptotic size equals 2.3 percent, and its size distortions exceed those of all raw-moments tests for $T \geq 100$. The $\hat{\mu}_{34}$ test hardly ever rejects in the case of strong persistence unless the sample size is large. However, even with 1000 observations and without serial correlation, the $\hat{\mu}_{34}$ test can suffer from notable size distortions. In general, the smallest size distortions of the raw-moments test are obtained with the $\hat{\alpha}^0_{12}$ test.

Summing up, neither the raw-moments tests based on the S-PITs, nor the $\hat{\beta}_{12}$ test, nor the $\hat{\mu}_{34}$ test can guarantee small size distortions in all circumstances. However, the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ tests always perform well in the case of MA(1)-processes. In the case of AR(1)-processes, they are undersized in small samples with strong

[35] Additional simulations not reported in the tables suggest that this appears to be the largest positive size distortion among all sample sizes given the AR(1)-process with $\rho = 0.9$ and the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ tests considered. With $T = 150$, the actual size of the $\hat{\alpha}^0_{1234}$ test equals about 6.9 percent. With $T = 250$, it reaches 7.2 percent.

persistence, but one could argue that this is preferable to overrejecting as observed for the $\hat{\beta}_{12}$ test. The use of the $\hat{\mu}_{34}$ test and the $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ tests cannot be recommended.

The latter conclusions are uniquely motivated by size considerations. However, it should be noted that these considerations are of paramount importance for most practitioners, because they tend to rely on the critical values derived from the χ^2-distribution.

4.2 Size-Adjusted Power

The size-adjusted power (henceforth simply referred to as power) of the tests depends crucially on the sample moments of the S-PITs and INTs. Therefore, these moments are displayed in Table 3 for three sample sizes ($T = 50$, $T = 200$, $T = 1000$) and the case of no ($\rho = 0$) and strong ($\rho = 0.9$, AR(1)-process) persistence. Obviously, the expected sample raw moments do not depend on the sample size or persistence. Differences between the sample raw moments for a specific forecast density are only caused by the Monte Carlo error. The sample raw moments are only reported for the S-PITs, but all statements also apply to the sample raw moments of the INTs.

In contrast to the sample raw moments, the sample moment estimators for central and standardized moments can be severely biased. The sample variance, denoted by $\hat{\mu}_2$, is biased only if the data are serially correlated. The sample skewness estimator and, especially, the sample kurtosis estimator can be strongly biased even without serial correlation, unless the sample size is very large, as also found by Bai and Ng (2005).[36]

It is evident that misspecifications of the forecast density that lead to raw moments of the S-PITs (and INTs) different from those under the null of correct specification do not necessarily have the same effect on central or standardized moments. For example, all raw moments of the S-PITs corresponding to the normal forecast density with expectation $\mu = -0.5$ are different from their standard uniform counterparts, whereas variance, skewness and kurtosis of the the INTs are not. If the variance of the normal forecast density is misspecified, this misspecification causes the fourth raw moment of the S-PITs to differ from 1.8 (and the fourth raw of the INTs to differ from 3), whereas the kurtosis of the INTs continues to equal 3.

Several other observations are noteworthy as well. For example, the moments of INTs associated with the two-piece normal forecast density all deviate from those of a standard normal variable, but the sample mean is very close to 0.[37] The sample mean of the S-PITs is not too far from 0, either. In the same case, the second sample raw moment of the S-PITs only slightly differs from 1. Hence, it is not always obvious which moments of the transformed variables will indicate a certain type of misspecification. Concerning the sample kurtosis, values larger than 3, i.e. positive values of the sample excess kurtosis only appear in the case of the two-piece normal and the normal mixture forecast density. Interestingly, in the latter case, the sign of the sample excess kurtosis actually depends on the sample size and persistence. This property will be important for understanding

[36] The common sample skewness and sample kurtosis estimators, $\hat{s} = \frac{\frac{1}{T}\sum_{t=1}^{T}(z_t-\bar{z})^3}{\left(\frac{1}{T}\sum_{t=1}^{T}(z_t-\bar{z})^2\right)^{\frac{3}{2}}}$ and $\hat{k} = \frac{\frac{1}{T}\sum_{t=1}^{T}(z_t-\bar{z})^4}{\left(\frac{1}{T}\sum_{t=1}^{T}(z_t-\bar{z})^2\right)^{2}}$, are used. For alternative estimators see Joanes and Gill (1998).

[37] This property does not depend on the specific value of γ chosen here.

the results of the $\hat{\mu}_{34}$ test.

In what follows, the results for the individual forecast densities are presented. Remember that the size-adjusted power is reported, i.e. the critical values are simulated and can thus differ from the asymptotic critical values.

Normal forecast density with misspecified expectation ($\mu = -0.5$)

In the case of the normal forecast density with misspecified expectation, the results in Table 4 suggest that the most powerful tests are the $\hat{\beta}_{12}$ test and the $\hat{\alpha}_1$ test, where the $\hat{\beta}_{12}$ test works better in small samples with strong persistence. In general, the tests using zero-covariance properties, i.e. the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ tests have more power than the corresponding $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ tests. Therefore, and because of the better size properties, in the following we will only focus on the former.[38] The $\hat{\alpha}^0_{12}$ test, which is the raw-moment test corresponding most closely to the $\hat{\beta}_{12}$ test, does not have higher power than the latter in any of the settings considered. Additionally considering the third moment and fourth raw moment leads to power losses. The $\hat{\mu}_{34}$ test has power equal to size. Finally, in the setting with $T = 50$ and an AR(1)-process with $\rho = 0.9$, even the most powerful test, the $\hat{\beta}_{12}$ test, rejects in only 11% of the cases.

Normal forecast density with too small variance ($\sigma = 2/3$)

The results displayed in Table 5 show that the $\hat{\beta}_{12}$ test has the largest power in all settings. Interestingly, the tests based on $\hat{\alpha}^0_{12}$ and $\hat{\alpha}^0_{123}$ have relatively similar power properties, although the expectation of the third raw moment equals 0. The inclusion of the fourth raw moment causes more pronounced power losses. The power of the $\hat{\mu}_{34}$ test and the $\hat{\alpha}_1$ test, as expected, is about equal to size. In the

[38] The superior power of the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ tests was also observed for all subsequent misspecifications. Results for the $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ tests for the subsequent misspecifications are available upon request.

case of an AR(1)-process with $\rho = 0.9$ the tests based on $\hat{\alpha}^0_{12}$ and $\hat{\alpha}^0_{123}$ need more than 100 observations for their power to differ pronouncedly from size. For the $\hat{\alpha}^0_{1234}$ test, more than 200 observations are needed for this.

Normal forecast density with too large variance ($\sigma = 3/2$)

The results for the case that the forecast standard deviation exceeds the true value are shown in Table 6. In general, again the $\hat{\beta}_{12}$ test has the largest power. However, there are cases where the power of the $\hat{\alpha}^0_{12}$ and the $\hat{\alpha}^0_{123}$ test attain slightly higher values. So this is one of the situations which illustrate that the likelihood-ratio test is not the uniformly most powerful test here.[39] The other statements made for the case with $\sigma = 2/3$ continue to apply.

Two-piece normal forecast density ($\gamma = 0.8$)

The misspecifications implied by the two-piece normal forecast density are, in general, most successfully discovered by the $\hat{\mu}_{34}$ test and the $\hat{\alpha}^0_{123}$ test, as shown in Table 7. The $\hat{\beta}_{12}$ test attains a similar power only if $T = 50$. The power of the $\hat{\alpha}^0_{1234}$ test is comparable to that of the $\hat{\alpha}^0_{123}$ test. The $\hat{\alpha}_1$ test and the $\hat{\alpha}^0_{12}$ have rather low power, which does not seem surprising, because the mean of the S-PITs is close to 0, and the second raw moment is close to 1.[40]

***t*-distributed forecast density** (5 degrees of freedom)

As can be seen from Table 8, if the forecast density has a t-distribution with 5 degrees of freedom, the $\hat{\mu}_{34}$ test delivers the best power results. This result is noteworthy because Bai and Ng (2005) state that their normality tests derive hardly any power from the kurtosis component. However, here the INTs are symmetric,

[39] According to Berkowitz (2001), such a test does not exist for the null hypothesis $\mu = 0, \sigma^2 = 1$.

[40] However, such considerations can be misleading, as will be seen in the case of the normal mixture density.

so that the skewness equals 0. Thus, the test here, actually, derives all its power from the non-zero sample excess kurtosis. The fact that the $\hat{\mu}_{34}$ test has better properties than claimed by its developers is apparently related to the sign of the excess kurtosis. Bai and Ng (2005) almost exclusively study the power for random variables with positive excess kurtosis, and their kurtosis and normality tests indeed have very low power in these cases. However, in the case of negative excess kurtosis, their kurtosis and normality tests perform well in terms of size-adjusted power even in small samples.[41]

Concerning the tests based on raw moments, the $\hat{\alpha}^0_{12}$ test, the $\hat{\alpha}^0_{123}$ test and the $\hat{\alpha}^0_{1234}$ test again attain similar power which here clearly exceeds the power of the $\hat{\beta}_{12}$ test whenever power exceeds size.

Normal mixture forecast density ($\sigma = 0.4$)

The behavior of the $\hat{\mu}_{34}$ test observed here appears counterintuitive at first sight. Considering first the results for MA(1)-process, which are easier to explain, the power is around 0.25 for $T = 50$, but then decreases with the the sample size until $T = 1000$, where power is down to 0.03. These results are caused by the asymmetric power properties with respect to excess kurtosis, the bias of the sample kurtosis estimator, and the fact that the sample kurtosis estimator yields values around 3 in most settings.[42] In the small samples, the sample kurtosis estimator is strongly biased downwards, attaining values close to 3 or even below. For a given sample size, in every simulation with an estimate of the sample kurtosis lower than 3, the $\hat{\mu}_{34}$ test is much more likely to reject than with an estimate exceeding 3 by the same amount. So the sample size here has two effects on power. Firstly, a

[41] This result was also confirmed in additional Monte Carlo studies not reported here.
[42] See Table 3.

larger sample size implies a larger sample kurtosis, reducing power if the sample kurtosis is in the neighborhood of 3. Secondly, a larger sample size, of course, implies more precise estimates, increasing power if the kurtosis does not equal 3. For sample sizes up to $T = 1000$, here the first effect dominates. For very large sample sizes, of course, the second effect would dominate. With strong persistence, the issue is further complicated by the fact that persistence also has two opposite effects. Firstly, it amplifies the small sample bias, hence further reduces the sample kurtosis, thereby leading to more rejections. Secondly, it makes inference more difficult, resulting in lower power. All effects described give rise to the behavior observed for $\rho = 0.9$ in the case of the AR(1)-process, where the power increases from $T = 50$ to $T = 200$ and then decreases for $T = 500$ to $T = 1000$, attaining only 0.13 in the latter case. If even larger sample sizes were considered, the power, of course, would eventually increase again.

The highest power, in general, is attained by the $\hat{\alpha}^0_{1234}$ test. Only in the case of small samples and strong persistence, the $\hat{\mu}_{34}$ test delivers better results. The high power of the $\hat{\alpha}^0_{1234}$ test compared to all other raw-moments tests is surprising insofar as, according to Table 3, the fourth raw sample moment is virtually equal to 1.8, its value under the null hypothesis. Additional simulations not reported here reveal that, indeed, a test that only uses the fourth raw moment, i.e. the $\hat{\alpha}_4$ test, has power essentially equal to size. Further simulations show that, interestingly, the high power of the $\hat{\alpha}^0_{1234}$ test stems from the joint consideration of the second, third and fourth raw moment. If one of these moments does not enter the test, the power decreases considerably. Apparently, the joint distribution of these three sample moments is such that, usually, at least one of the moments is likely to signal departures from the standard uniform distribution.[43] Hence, this example shows that it can even be beneficial to include moments whose marginal distributions, at first sight, suggest that they cannot contribute to the power of the test.

4.3 Summary

From the Monte Carlo simulations conducted above, several conclusion can be drawn. Raw-moments tests should not be based on the INTs due to size distortions. Instead, the tests should be based on the S-PITs. Moreover, the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ tests yield better results than the $\hat{\alpha}_{r_1 r_2 \ldots r_N}$ tests in terms of size and power. Among the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ tests, the $\hat{\alpha}^0_{12}$ test often gives the smallest size distortions. However, the $\hat{\alpha}^0_{1234}$ test has power against more types of misspecification, while its size distortions are still fairly contained.

Concerning the choice among the $\hat{\beta}_{12}$ test, the $\hat{\mu}_{34}$ test and the $\hat{\alpha}^0_{r_1 r_2 \ldots r_N}$ tests, the $\hat{\mu}_{34}$ test often has the largest size distortions, it cannot detect misspecifications which affect first and second moments of the INTs only, and its power depends in complex ways on sample size and persistence. Therefore, this test does not appear to be well-suited for the evaluation of density forecasts. The $\hat{\beta}_{12}$ test has very good size properties if the underlying AR(1)-process assumption is correct, but otherwise suffers from size distortions which do not vanish asymptotically.[44] The size distortions are moderate in the setting chosen here, but could be larger in other

[43] That, is if, for example, the third and fourth sample moments are close to 0 and 1.8, respectively, the second sample moment is likely to differ significantly from 1. If the second and third sample moments are close to 0 and 1, respectively, the fourth is likely to differ significantly from 1.8. Finally, if the second and the fourth sample moments are close to 1 and 1.8, respectively, the third is likely to differ significantly from 0.

[44] If there is evidence against an AR(1)-process, one could try to employ a different process, but should be aware of the potential problems due to pre-testing and increased parameter estimation uncertainty.

situations. The $\hat{\beta}_{12}$ test is usually the best choice if the sample size is small, and the data is very persistent. Nevertheless, it should be noted that the test is also likely to overreject in these situations, even if the data follow an AR(1)-process.

If the misspecifications of the forecast density are restricted to third and fourth moments, the $\hat{\beta}_{12}$ test can nonetheless be useful for detecting miscalibration, because the mentioned misspecifications will usually translate into INTs with non-zero mean or at least a variance not equal to 1. However, tests considering higher moments, of course, tend to have larger power in these situations. If the sample is not too small, or if persistence is only moderate, as one would expect in the case of h being not too large, the $\hat{\alpha}^0_{1234}$ test appears to be a good option with power against many types of misspecification.

5 Extensions

In the previous simulations, raw-moments tests were investigated for only two types of transformations, the INTs and the S-PITs. However, any other symmetry-preserving transformation might be a candidate with potentially better size and power properties. The uniform distribution, for example, is a special case of the beta distribution with parameters $\alpha = \beta = 1$. If $\alpha = \beta$ holds, this distribution is symmetric, and choosing, for example $\alpha = \beta = \frac{1}{2}$ gives the U-shaped arcsine distribution function, whereas $\alpha = \beta = 2$ yields a concave distribution function. Just like the uniform distribution, these beta distributions can easily be transformed such that they are symmetric around 0 and have a variance of 1. However, Monte Carlo simulations suggest that none of these distributions consistently leads to better size or power properties than the standardized uniform distribution. Choosing

$\alpha = \beta = 2$ actually rather tends to lead to lower power and larger size distortions in the AR(1)-case, whereas with $\alpha = \beta = \frac{1}{2}$ the size distortions in the MA(1)-case slightly increase.[45]

Another possibility is given by the consideration of orthogonal functions instead of moments. That is, instead of the first four raw moments, one could use, for example, the first four Legendre or Chebyshev polynomials for the transformed variable $y_t = 2\left(u_t - \frac{1}{2}\right)$ with positive density over the interval $[-1, 1]$. In the case of the Legendre polynomials, for example, this would give the vector

$$\hat{\mathbf{D}}_{1234} = \begin{bmatrix} \frac{1}{T}\left(\sum_{t=1}^{T} y_t\right) \\ \frac{1}{T}\left(\sum_{t=1}^{T} \frac{1}{2}\left(3y_t^2 - 1\right)\right) \\ \frac{1}{T}\left(\sum_{t=1}^{T} \frac{1}{2}\left(5y_t^3 - 3y_t\right)\right) \\ \frac{1}{T}\left(\sum_{t=1}^{T} \frac{1}{8}\left(35y_t^4 - 30y_t^2 + 3\right)\right) \end{bmatrix}$$

and the corresponding vector series

$$\mathbf{d}_t = \begin{bmatrix} y_t \\ \frac{1}{2}\left(3y_t^2 - 1\right) \\ \frac{1}{2}\left(5y_t^3 - 3y_t\right) \\ \frac{1}{8}\left(35y_t^4 - 30y_t^2 + 3\right) \end{bmatrix}.$$

With this approach, all elements of $\mathbf{d}_t$ have contemporaneous covariances equal to 0, but this does not hold for the lagged covariances, and, hence, the long-run

[45] However, if a certain raw-moments test is chosen, the transformations considered could be used to limit size distortions. For example, the size of the $\hat{\alpha}_{1234}^{0}$ strongly depends on the convergence of the raw fourth sample moment to normality. If y_t has a uniform distribution with expectation 0, y_t^4 has a positively skewed distribution with mode at 0. The more peaked the distribution of y_t is at 0, the more skewed is y_t^4. With a U-shaped distribution of y_t, however, the skewness of y_t^4 decreases. Therefore, transforming u_t such that the resulting y_t has an arcsine distribution under the null tends to decrease the size distortions of the $\hat{\alpha}_{1234}^{0}$ test.

covariances. Only the long-run covariances of odd and even polynomials are equal to 0, as in the case of odd and even moments. It turns out that, again, neither Legendre nor Chebyshev polynomials yield consistently better results in terms of size and power than obtained with the approach based on raw moments and the standardized uniform distribution. Moreover, results based on these polynomials are probably more difficult to interpret.

Finally, one could modify the approach based on the INTs by using knowledge about the long-run covariance matrix under the null. According to Lomnicki (1961), the long-run covariance matrix of the raw moments of the INTs, $\mathbf{\Omega}_{1234}$, is given by

$$\mathbf{\Omega}_{1234} = \begin{bmatrix} \sum_{i=-\infty}^{\infty} R_i & 0 & 3R_0 \sum_{i=-\infty}^{\infty} R_i & 0 \\ & 2\sum_{i=-\infty}^{\infty} R_i^2 & 0 & 12R_0 \sum_{i=-\infty}^{\infty} R_i^2 \\ 3R_0 \sum_{i=-\infty}^{\infty} R_i & 0 & \sum_{i=-\infty}^{\infty} (9R_0^2 R_i + 6R_i^3) & 0 \\ 0 & 12R_0 \sum_{i=-\infty}^{\infty} R_i^2 & 0 & \sum_{i=-\infty}^{\infty} (72R_0^2 R_i^2 + 24R_i^4) \end{bmatrix},$$

where R_i denotes the autocovariance of y_t at lag i. Thus, the matrix $\hat{\mathbf{\Omega}}_{1234}$ can be constructed by choosing a cutoff value for i and estimating the sample autocovariances. However, with this approach, the size and power properties of the raw-moments tests based on the INTs improve only marginally, so that using the S-PITs continues to yield clearly better results.

In summary, none of the extensions considered appears to be clearly preferable to the approach based on the raw moments and the S-PITs.

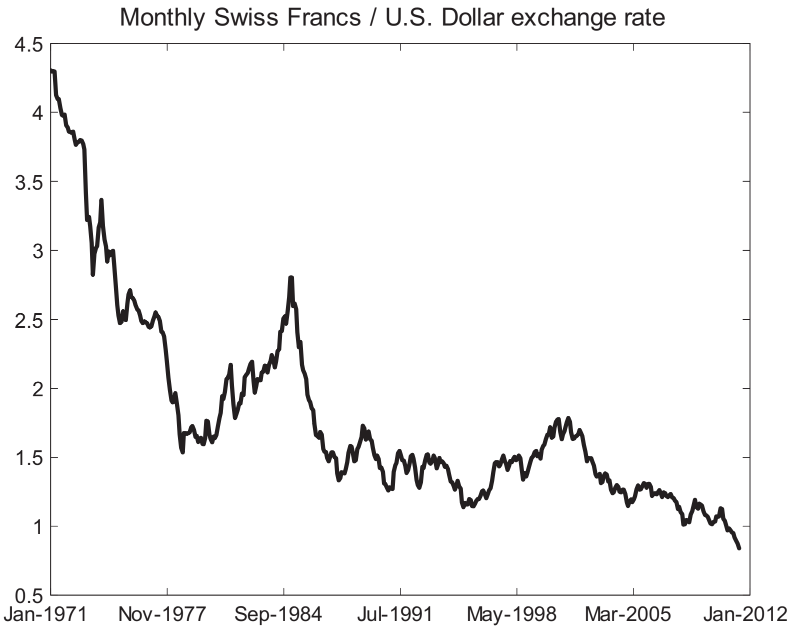

Figure 3: Monthly Swiss Francs / U.S. Dollar exchange

6 Empirical Application

In order to illustrate the usefulness of the tests based on raw moments, in what follows, out-of-sample density forecasts for the monthly Swiss Francs / U.S. Dollar exchange rate are investigated. The data cover the period from January 1971 to June 2011 and are displayed in Figure 3. I consider h-step-ahead forecasts with h ranging from 2 to 5 months. In light of the results of Meese and Rogoff (1983), the exchange rate is assumed to follow a random walk without drift, so that simple no-change forecasts are used. For each forecast horizon, the first 96 forecast errors available, corresponding to 8 years of data, are used to determine the type of forecast density and the initial estimates of the required parameters. This setup is used mainly because it is very easy to implement.

The $\hat{\mu}_{34}$ test does not reject the normality assumption of the forecast errors for any forecast horizon considered, so that the forecast densities are assumed to be normal.[46] Thus, the only parameter that needs to be estimated for the density forecasts is the variance. It is determined separately for each horizon, using a rolling window of 8 years. For the largest forecast horizon $h = 5$, this approach yields $T = 385$ density forecasts that can be evaluated. In order to obtain a balanced sample of density forecasts with $T = 385$ for every horizon h, for the smaller forecast horizons $h = 2, 3, 4$, the last $5 - h$ density forecasts are ignored.[47] The PITs of the resulting density forecast are shown as histograms in Figure 4. The dashed line indicates the expected height of the bars if the density forecasts were calibrated correctly. For the horizons $h = 2$ and $h = 3$, the most notable deviations from this line occur for the PITs in the interval $(0.9, 1.0)$, because pronouncedly less large positive forecast errors occur than expected.

Before evaluating the density forecasts, it is instructive to take a look at the autocorrelations and partial autocorrelations of the 385 INTs. These are displayed in Figure 5. Obviously, the assumption of an AR(1)-process would be rather problematic above all for the smaller forecast horizons. Actually, the dynamics of the INTs associated with the h-step-ahead density forecasts seem to be fairly well described by MA($h - 1$)-processes. Consequently, persistence increases with the forecast horizon. The autocorrelations and partial autocorrelations of the PITs shown in Figure 6 are very similar to those of the INTs, so that the same statements apply.

In order to check for miscalibration, the $\hat{\alpha}^0_{1234}$ test is employed. The $\hat{\beta}_{12}$ test

[46] The p-values range from 0.46 to 0.60. Nevertheless, according to the Monte Carlo results reported above, the $\hat{\mu}_{34}$ test is likely to be undersized and to have low power here, so that the normality assumption might not be without problems.

[47] So for the horizon $h = 4$, the density forecast for June 2011 is not used. For $h = 3$, the density forecasts for May and June 2011 are not used, etc.

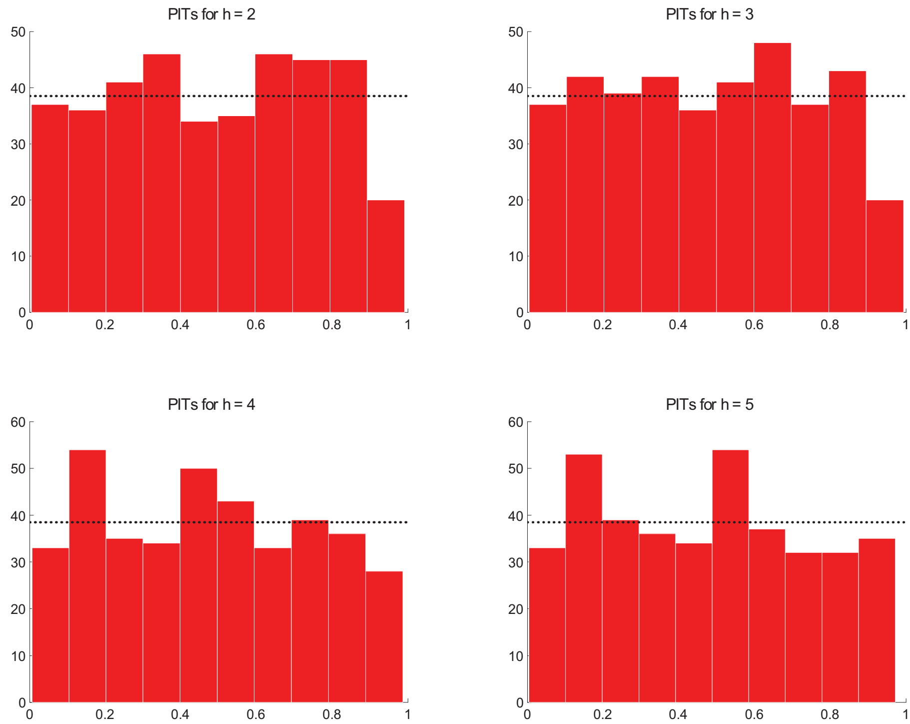

Figure 4: PITs of the density forecasts for the exchange rate at different forecast horizons h. The dashed line indicates the expected height of the bars if the density forecasts are calibrated correctly.

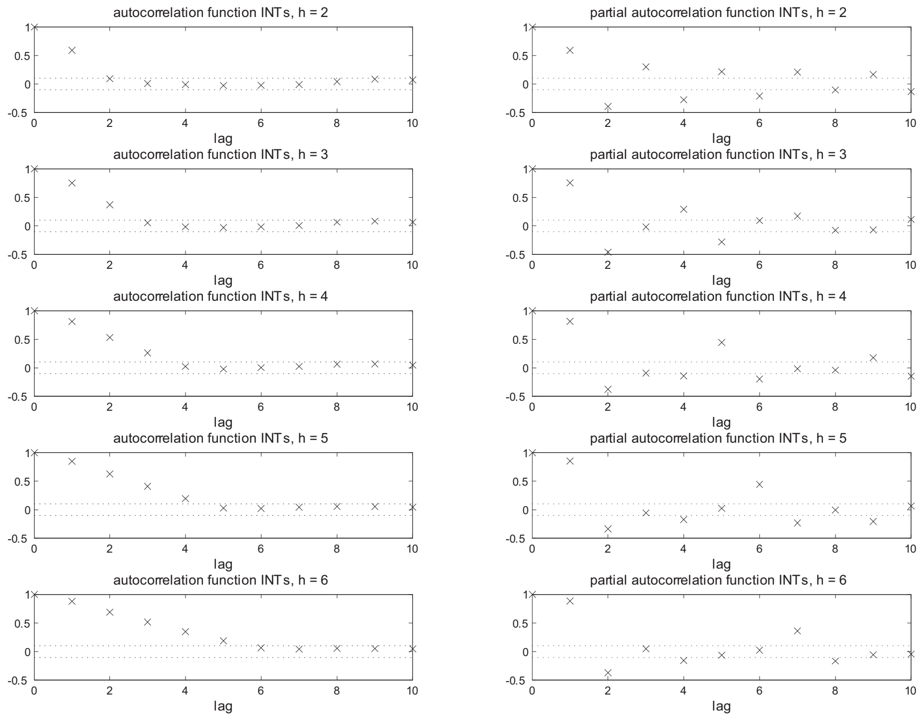

Figure 5: Autocorrelation functions of the INTs for the monthly Swiss Francs / U.S. Dollar exchange rate for different forecast horizons h based on no-change forecasts. Dashed lines indicate 95% confidence bounds, calculated as $\pm 2/\sqrt{T}$.

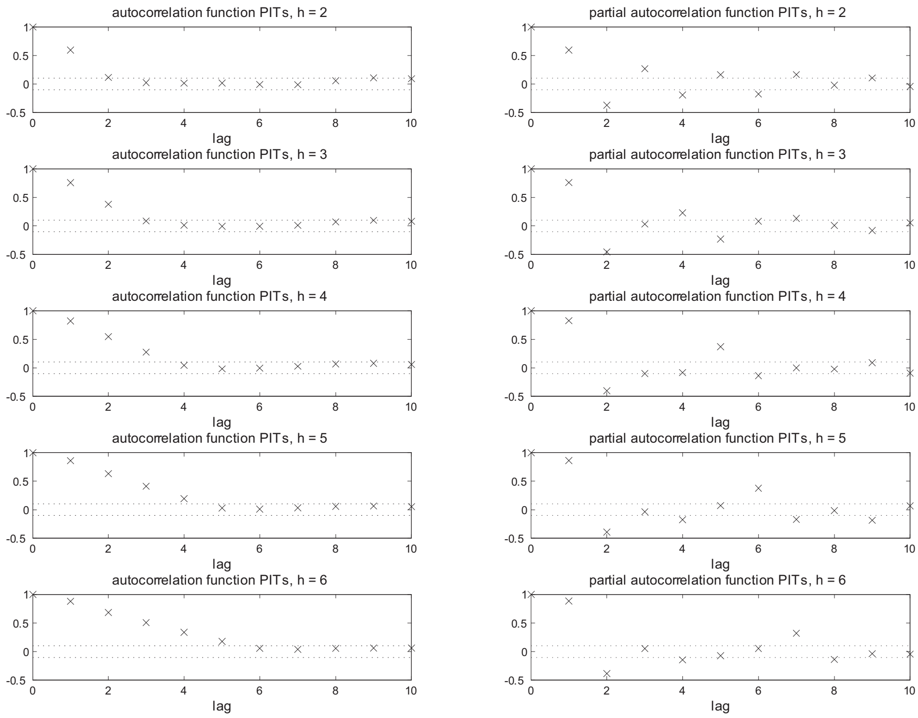

Figure 6: Autocorrelation functions of the PITs for the monthly Swiss Francs / U.S. Dollar exchange rate for different forecast horizons h based on no-change forecasts. Dashed lines indicate 95% confidence bounds, calculated as $\pm 2/\sqrt{T}$.

and the $\hat{\alpha}^0_{12}$ test are also used for the sake of comparisons. In Table 10, in addition to the test results, the first four sample raw moments of the S-PITs as well as the sample mean and sample variance $\hat{\mu}_2$ of the INTs are shown. For all forecast horizons, the sample means of the S-PITs and INTs are negative, and the second raw moments of the S-PITs and the variance of the INTs are smaller than 1. These results suggest that the density forecasts assign too much probability to positive forecast errors and to forecast errors that are large in absolute value. The fact that all third raw moments of the S-PITs are negative, and that all fourth raw moments of the S-PITs are smaller than 1.8 tend to support these conclusions.

Based on the raw moments of the S-PITs, the $\hat{\alpha}^0_{1234}$ test rejects the null hypothesis of correct calibration for the forecast horizons $h = 2$ and $h = 3$ at the conventional significance level of 5%. For $h = 4$ and $h = 5$, no rejection occurs. In contrast to these results, the $\hat{\beta}_{12}$ test and the $\hat{\alpha}^0_{12}$ test do not reject the null hypothesis for any horizon. Since the $\hat{\beta}_{12}$ test and the $\hat{\alpha}^0_{12}$ test do not make use of higher-order moments, it appears likely that the differences in the test results are caused by the differences in the moments considered.[48] The fact that the $\hat{\alpha}^0_{1234}$ test does not reject for the larger horizons could be caused by the power losses due to stronger persistence.

7 Conclusion and Outlook

In this work, two existing tests for the calibration of multi-step-ahead density forecasts are compared to new tests based on raw moments. The existing tests use

[48]Moreover, the $\hat{\beta}_{12}$ test is undersized in the presence of an MA(1)-process. This property could, of course, also hold for higher-order MA-processes of the forms suggested by Figure 5. This would be another reason why the $\hat{\beta}_{12}$ test is less likely to reject than the $\hat{\alpha}^0_{1234}$ test.

the inverse normal transforms (INTs) of the probability integral transforms (PITs). The raw-moments tests can, in principle, be based on any transformation which yields a symmetric zero-mean distribution function under the null hypothesis of correct calibration. In the present study, the raw-moments tests are based on the INTs and standardized PITs (S-PITs). Despite of the autocorrelation of the INTs and S-PITs, all tests considered here rely on standard critical values and, therefore, are attractive for practitioners in the first place. The third existing test for the calibration of multi-step-ahead density forecasts proposed by Corradi and Swanson (2006a) is not included in this study because its critical values are data dependent, their derivation is burdensome, and the test appears to be hardly applied.

We find that one of the existing tests, the $\hat{\mu}_{34}$ test of Bai and Ng (2005), cannot be recommended for the evaluation of density forecasts due to potentially large size distortions, complicated power properties, and, more importantly, not using important information contained in the first and second moments of the INTs. The second existing test, the $\hat{\beta}_{12}$ test, due to its relatively large power especially in small samples with strong persistence, can be very useful for the evaluation of density forecasts if the dynamics of the INTs correspond to the assumption used in the test. Otherwise, however, size distortions occur which do not vanish asymptotically. Moreover, the test does not use information contained in higher-order moments. It should be noted that both tests mentioned were not designed for the evaluation of multi-step-ahead density forecasts, but are applied or recommended for this purpose in the literature.

The raw-moments tests presented do not suffer from the drawbacks of the other tests mentioned above. Tests based on the S-PITs are found to have good size and power properties, and can therefore, and because of their simplicity, be a very useful tool for the evaluation of density forecasts. In contrast, raw-moments tests based on the INTs are subject to large size distortions. Raw-moments tests which use the fact that under the null, odd and even moments are uncorrelated perform better in terms of size and power than their counterparts which do not employ the zero-correlation property.

The $\hat{\alpha}^0_{12}$ test based on the S-PITs has very good size properties in all settings investigated in this study. The size distortions of the $\hat{\alpha}^0_{1234}$ test, in general, are slightly larger, but still fairly contained, and the test has considerable power against many alternatives. Therefore, the latter test, which uses the first four raw moments of the S-PITs appears to be the most recommendable raw-moments test. However, it should be noted that this test tends to have low power in small samples with strong persistence, so that the $\hat{\beta}_{12}$ test can be a better choice in such situations.

In an empirical application, the $\hat{\alpha}^0_{1234}$ test is applied to density forecasts of the monthly Swiss Francs / U.S. Dollar exchange rate, where the density forecasts are based on a random-walk model and the assumption of normally distributed forecast errors. The null hypothesis of correct calibration is rejected for the 2- and 3- month-ahead forecasts, but not for the 4- and 5- month-ahead forecasts.

The testing approach presented here can easily be extended in order to test other hypotheses of interest. For example, instead of only regressing $y_t^{r_i} - m_{r_i}$ on a constant, one could include $y_{t-h}^{r_i} - m_{r_i}$ as an additional regressor. Based on this setup, it would be possible to test for complete calibration by including the hypothesis that all coefficients equal zero. Moreover, one could easily test for the correct calibration for several forecast horizons jointly by considering the respective elements $y_t^{r_i} - m_{r_i}$ of the distinct forecast horizons in $\mathbf{d}_t$. Of course, the size and power properties of these approaches remain to be investigated.

Finally, it might be interesting to note that the serial correlation of the PITs is not only a feature of horizon-specific multi-step-ahead density forecasts. It also emerges if path density forecasts are evaluated. Path density forecasts are given by the forecasts of the joint density for several forecast horizons as considered in Jorda and Marcellino (2010).

References

Aastveit, Knut Are, Karsten R. Gerdrup, Anne Sofie Jore, and Leif Anders Thorsrud (September 2011). 'Nowcasting GDP in real-time: A density combination approach.' Working Paper 2011/11, Norges Bank

Andrews, Donald W. K. (May 1991). 'Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation.' *Econometrica*, 59(3), 817–858

Andrews, Donald W. K. and J. Christopher Monahan (1992). 'An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator.' *Econometrica*, 60(4), 953–66

Bache, Ida Wolden, Anne Sofie Jore, James Mitchell, and Shaun P. Vahey (October 2011). 'Combining VAR and DSGE forecast densities.' *Journal of Economic Dynamics and Control*, 35(10), 1659–1670

Bai, Jushan and Serena Ng (2005). 'Tests of Skewness, Kurtosis, and Normality for Time Series Data.' *Journal of Business and Economic Statistics*, 23(1), 49–60

Bao, Yong, Tae-Hwy Lee, and Burak Saltoglu (2007). 'Comparing Density Forecast Models.' *Journal of Forecasting*, 26, 203–225

Berkowitz, Jeremy (2001). 'Testing Density Forecasts, With Applications to Risk Management.' *Journal of Business & Economic Statistics*, 19(4), 465–474

Bontemps, Christian and Nour Meddahib (2005). 'Testing normality: a GMM approach.' *Journal of Econometrics*, 124, 149–186

— (2007). 'Testing Distributional Assumptions: A GMM Approach.' *IDEI Working Paper*, 486

Chen, Yi-Ting (2011). 'Moment Tests for Density Forecast Evaluation in the Presence of Parameter Estimation Uncertainty.' *Journal of Forecasting*, 30, 409–450

Clements, Michael P. (2004). 'Evaluating the Bank of England Density Forecasts of Inflation.' *The Economic Journal*, 114, 844–866

Corradi, Valentina and Norman R. Swanson (2006a). 'Bootstrap conditional distribution tests in the presence of dynamic misspecification.' *Journal of Econometrics*, 133, 779–806

— (2006b). ‘Predictive Density Evaluation.’ In Graham Elliott, Clive W. J. Granger, and Allan Timmermann, editors, ‘Handbook of Economic Forecasting,’ volume 1, chapter 5, pp. 197–284. North Holland

D’Agostino, Antonello, Luca Gambetti, and Domenico Giannone (forthcoming). ‘Macroeconomic Forecasting and Structural Change.’ *Journal of Applied Econometrics*

Dawid, A. P. (1984). ‘Statistical Theory: The Prequential Approach.’ *Journal of the Royal Statistical Society A*, 147(2), 278–292

Diebold, Francis X. (1998). *Elements of Forecasting.* South-Western College Publishing, Cincinnati

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay (1998). ‘Evaluating Density Forecasts with Applications to Financial Risk Management.’ *International Economic Review*, 39(4), 863–883

Doornik, Jurgen A. and Henrik Hansen (December 2008). ‘An Omnibus Test for Univariate and Multivariate Normality.’ *Oxford Bulletin of Economics and Statistics*, 70(s1), 927–939

Gneiting, Tilmann, Fadoua Balabdaoui, and Adrian Raftery (April 2007). ‘Probabilistic forecasts, calibration and sharpness.’ *Journal of the Royal Statistical Society B*, 69(2), 243–268

Joanes, D. N. and C. A. Gill (1998). ‘Comparing measures of sample skewness and kurtosis.’ *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 183–189. doi:10.1111/1467-9884.00122

Jorda, Oscar and Massimiliano Marcellino (June/July 2010). ‘Path Forecast Evaluation.’ *Journal of Applied Econometrics*, 25(4), 635–662

Jore, Anne Sofie, James Mitchell, and Shaun P. Vahey (2010). ‘Combining Forecast Densities from VARs with Uncertain Instabilities.’ *Journal of Applied Econometrics*, 25(4), 621–634

Knüppel, Malte and Guido Schultefrankenfeld (forthcoming). ‘How Informative Are Central Bank Assessments of Macroeconomic Risks?’ *International Journal of Central Banking*

Lobato, Ignacio N. and Carlos Velasco (August 2004). ‘A Simple Test Of Normality For Time Series.’ *Econometric Theory*, 20(04), 671–689

Lomnicki, Z. A. (1961). ‘Tests for departure from normality in the case of linear stochastic processes.’ *Metrika*, 4(1), 37–62

Meese, Richard A. and Kenneth S. Rogoff (1983). ‘Empirical exchange rate models of the seventies : Do they fit out of sample?’ *Journal of International Economics*, 14, 3–24

Mitchell, James and Stephen G. Hall (2005). ‘Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR Fan Charts of Inflation.’ *Oxford Bulletin of Economics and Statistics*, 67, 995–1033

Mitchell, James and Kenneth F. Wallis (2011). ‘Evaluating Density Forecasts: Forecast Combinations, Model Mixtures, Calibration and Sharpness.’ *Journal of Applied Econometrics*, 26(3)

Newey, Whitney K. and Kenneth D. West (May 1987). ‘A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.’ *Econometrica*, 55(3), 703–708

Rosenblatt, Murray (1952). ‘Remarks on a Multivariate Transformation.’ *Annals of Mathematical Statistics*, 23(3), 470–472

Tay, Anthony S. and Kenneth F. Wallis (2000). ‘Density Forecasting: A Survey.’ *Journal of Forecasting*, 19, 235–254

Wallis, Kenneth F. (2004). ‘An Assessment of Bank of England and National Institute Inflation Forecast Uncertainties.’ *National Institute Economic Review*, 189, 64–71

— (2008). ‘Forecast Uncertainty, Its Representation and Evaluation.’ In Roberto S. Mariano and Yiu-Kuen Tse, editors, ‘Econometric Forecasting and High-Frequency Data Analysis,’ volume 13. World Scientific Publishing Company

West, Kenneth D and Michael W McCracken (November 1998). ‘Regression-Based Tests of Predictive Ability.’ *International Economic Review*, 39(4), 817–40

A Appendix: Proof

The following proof shows that the long-run covariance of $y_t^{r_i} - m_{ri}$ and $y_t^{r_j} - m_{r_j}$ equals 0 if y_t is symmetrically distributed around 0 and if $r_i + r_j$ is odd. As a starting point, consider the inverse normal transform $z_t = \Phi^{-1}(u_t)$ which yields zero-mean variables, and denote other symmetry-preserving transformations by $y_t = S(z_t)$. Symmetry of the random variable y_t is obtained if $S(z_t)$ is an odd function, that is, if

$$S(-z_t) = -S(z_t)$$

holds. A simple example is given by the function $S(z_t) = z_t$. The symmetric density of y_t will be denoted by $f(y_t)$.

Suppose that r_i is odd and r_j is even. Then, for the contemporaneous covariance of $y_t^{r_i}$ and $y_t^{r_j}$, we have that

$$E\left[y_t^{r_i}\left(y_t^{r_j} - m_{r_j}^y\right)\right] = E\left[y_t^{r_i+r_j}\right] - E\left[y_t^{r_i}\right] m_{r_j}^y$$

where $m_{r_j}^y$ denotes the expectation $E\left[y_t^{r_j}\right]$. In order to show that the contemporaneous covariance of $y_t^{r_i}$ and $y_t^{r_j}$ is 0, it is thus enough to show that $E\left[y_t^r\right]$ equals 0 if r is odd. For the normally distributed zero-mean variable z_t, the expectation $E\left[z_t^r\right]$ equals

$$E\left[z_t^r\right] = \int_{-\infty}^{0} z_t^r \phi(z_t)\, dz_t + \int_{0}^{\infty} z_t^r \phi(z_t)\, dz_t = 0,$$

because the facts that the normal density $\phi(z_t)$ is an even function and that z_t^r is an odd function imply that the product $z_t^r \phi(z_t)$ is an odd function and, hence, that

$$\int_{-\infty}^{0} z_t^r \phi\left(z_t\right) dz_t = -\int_{0}^{\infty} z_t^r \phi\left(z_t\right) dz_t.$$

holds. Thus, the expectation $E\left[y_t^r\right]$ equals

$$\begin{aligned} E\left[y_t^r\right] &= \int_{-\infty}^{0} y_t^r f\left(y_t\right) dy_t + \int_{0}^{\infty} y_t^r f\left(y_t\right) dy_t \\ &= 0 \end{aligned}$$

because, like z_t^r, y_t^r is an odd function, and, like $\phi\left(z_t\right)$, $f\left(y_t\right)$ is an even function.

For the non-contemporaneous covariance of $y_t^{r_i}$ and $y_{t-v}^{r_j}$ with $v \in \mathbb{Z}$ one obtains

$$\begin{aligned} E\left[y_t^{r_i}\left(y_{t-v}^{r_j} - m_{r_j}^y\right)\right] &= E\left[y_t^{r_i} y_{t-v}^{r_j}\right] - E\left[y_t^{r_i}\right] m_{r_j}^y \\ &= E\left[y_t^{r_i} y_{t-v}^{r_j}\right]. \end{aligned}$$

Again starting with the normally distributed zero-mean variables, the latter expectation can be rewritten as

$$\begin{aligned} E\left[z_t^{r_i} z_{t-v}^{r_j}\right] &= \int_{0}^{\infty}\int_{0}^{\infty} z_t^{r_i} z_{t-v}^{r_j} \phi\left(z_t, z_{t-v}\right) dz_t dz_{t-v} \\ &\quad + \int_{0}^{\infty}\int_{-\infty}^{0} z_t^{r_i} z_{t-v}^{r_j} \phi\left(z_t, z_{t-v}\right) dz_t dz_{t-v} \\ &\quad + \int_{-\infty}^{0}\int_{0}^{\infty} z_t^{r_i} z_{t-v}^{r_j} \phi\left(z_t, z_{t-v}\right) dz_t dz_{t-v} \\ &\quad + \int_{-\infty}^{0}\int_{-\infty}^{0} z_t^{r_i} z_{t-v}^{r_j} \phi\left(z_t, z_{t-v}\right) dz_t dz_{t-v} \end{aligned}$$

where $\phi\left(z_t, z_{t-v}\right)$ denotes the joint normal density of z_t and z_{t-v}. Here it holds that

$$\int_0^\infty \int_0^\infty z_t^{r_i} z_{t-v}^{r_j} \phi\left(z_t, z_{t-v}\right) dz_t dz_{t-v}$$
$$= -\int_{-\infty}^0 \int_{-\infty}^0 z_t^{r_i} z_{t-v}^{r_j} \phi\left(z_t, z_{t-v}\right) dz_t dz_{t-v}$$

because $z_{t-v}^{r_j}$ is an even function, $z_t^{r_i}$ is an odd function and $\phi\left(z_t, z_{t-v}\right) = \phi\left(-z_t, -z_{t-v}\right)$ holds. Moreover, since $\phi\left(z_t, -z_{t-v}\right) = \phi\left(-z_t, z_{t-v}\right)$, we have that

$$\int_0^\infty \int_{-\infty}^0 z_t^{r_i} z_{t-v}^{r_j} \phi\left(z_t, z_{t-v}\right) dz_t dz_{t-v}$$
$$= -\int_{-\infty}^0 \int_0^\infty z_t^{r_i} z_{t-v}^{r_j} \phi\left(z_t, z_{t-v}\right) dz_t dz_{t-v}$$

holds, so that

$$E\left[z_t^{r_i} z_{t-v}^{r_j}\right] = 0.$$

As above, considering $y_t^{r_i}$ and $y_t^{r_j}$ instead of the odd function $z_t^{r_i}$ and the even function $z_t^{r_j}$ leads to the same result, because, firstly, $y_t^{r_i}$ also is an odd function and $y_t^{r_j}$ also is an even function, and secondly, $f\left(y_t, y_{t-v}\right) = f\left(-y_t, -y_{t-v}\right)$ and $f\left(y_t, -y_{t-v}\right) = f\left(-y_t, y_{t-v}\right)$ must hold because $y_t = S\left(z_t\right)$ is a symmetry-preserving transformation. Therefore,

$$E\left[y_t^{r_i} y_{t-v}^{r_j}\right] = 0$$

holds for all $v \in \mathbb{Z}$, implying that the long-run covariance of an odd and an even raw moment equals 0 if $y_t = H\left(u_t\right)$ follows a distribution which is symmetric around 0.

T	ρ	$\hat{\beta}_{12}$	$\hat{\mu}_{34}$	$\hat{\alpha}_{1}$	$\hat{\alpha}^{0}_{12}$	$\hat{\alpha}_{12}$	$\hat{\alpha}^{0}_{123}$	$\hat{\alpha}_{123}$	$\hat{\alpha}^{0}_{1234}$	$\hat{\alpha}_{1234}$
					MA(1)-process					
50	0.0	0.049	0.024	0.039	0.052	0.071	0.038	0.132	0.169	0.321
50	0.5	0.035	0.016	0.032	0.055	0.072	0.040	0.111	0.161	0.272
50	0.9	0.025	0.012	0.026	0.051	0.052	0.035	0.079	0.147	0.213
100	0.0	0.051	0.059	0.046	0.054	0.067	0.045	0.113	0.162	0.296
100	0.5	0.033	0.039	0.042	0.060	0.077	0.049	0.124	0.178	0.323
100	0.9	0.024	0.032	0.040	0.059	0.073	0.047	0.120	0.183	0.328
200	0.0	0.051	0.089	0.047	0.053	0.061	0.047	0.089	0.128	0.221
200	0.5	0.033	0.070	0.047	0.059	0.069	0.051	0.106	0.142	0.254
200	0.9	0.023	0.064	0.045	0.057	0.067	0.048	0.106	0.147	0.264
500	0.0	0.051	0.095	0.050	0.052	0.056	0.049	0.071	0.095	0.149
500	0.5	0.032	0.088	0.050	0.055	0.059	0.050	0.079	0.104	0.170
500	0.9	0.024	0.086	0.049	0.055	0.060	0.050	0.082	0.107	0.179
1000	0.0	0.050	0.085	0.049	0.051	0.053	0.049	0.061	0.078	0.111
1000	0.5	0.031	0.084	0.049	0.053	0.055	0.050	0.067	0.084	0.126
1000	0.9	0.023	0.085	0.049	0.052	0.055	0.050	0.068	0.087	0.133
					AR(1)-process					
50	0.0	0.050	0.024	0.039	0.053	0.072	0.040	0.132	0.169	0.319
50	0.5	0.058	0.012	0.035	0.069	0.091	0.052	0.127	0.167	0.265
50	0.9	0.094	0.002	0.001	0.017	0.000	0.003	0.000	0.012	0.000
100	0.0	0.051	0.059	0.045	0.055	0.067	0.045	0.113	0.163	0.297
100	0.5	0.054	0.029	0.050	0.074	0.106	0.063	0.162	0.201	0.369
100	0.9	0.075	0.001	0.006	0.072	0.012	0.041	0.002	0.143	0.011
200	0.0	0.050	0.090	0.047	0.053	0.060	0.047	0.090	0.128	0.222
200	0.5	0.052	0.058	0.055	0.069	0.091	0.062	0.137	0.162	0.299
200	0.9	0.064	0.006	0.027	0.098	0.094	0.079	0.140	0.292	0.353
500	0.0	0.050	0.095	0.049	0.052	0.055	0.049	0.070	0.095	0.149
500	0.5	0.051	0.082	0.055	0.063	0.073	0.059	0.102	0.119	0.206
500	0.9	0.055	0.018	0.053	0.088	0.128	0.079	0.211	0.248	0.449
1000	0.0	0.050	0.084	0.050	0.051	0.053	0.049	0.061	0.077	0.109
1000	0.5	0.052	0.083	0.055	0.060	0.066	0.058	0.083	0.096	0.154
1000	0.9	0.052	0.041	0.057	0.080	0.106	0.074	0.172	0.198	0.363

Note: Actual sizes when the nominal size equals 0.05.

Table 1: Size distortions of tests, raw-moments tests based on INTs

T	ρ	$\hat{\beta}_{12}$	$\hat{\mu}_{34}$	$\hat{\alpha}_{1}$	$\hat{\alpha}^{0}_{12}$	$\hat{\alpha}_{12}$	$\hat{\alpha}^{0}_{123}$	$\hat{\alpha}_{123}$	$\hat{\alpha}^{0}_{1234}$	$\hat{\alpha}_{1234}$
					MA(1)-process					
50	0.0	0.051	0.023	0.040	0.036	0.039	0.030	0.041	0.034	0.048
50	0.5	0.035	0.015	0.034	0.033	0.034	0.021	0.023	0.030	0.024
50	0.9	0.024	0.013	0.027	0.029	0.022	0.017	0.011	0.026	0.010
100	0.0	0.051	0.060	0.045	0.043	0.046	0.040	0.048	0.044	0.054
100	0.5	0.034	0.039	0.043	0.044	0.046	0.034	0.043	0.041	0.049
100	0.9	0.024	0.033	0.040	0.040	0.040	0.030	0.035	0.038	0.040
200	0.0	0.050	0.090	0.048	0.046	0.048	0.045	0.049	0.046	0.052
200	0.5	0.032	0.071	0.047	0.048	0.048	0.043	0.048	0.046	0.052
200	0.9	0.023	0.064	0.046	0.047	0.046	0.040	0.045	0.044	0.049
500	0.0	0.050	0.095	0.049	0.048	0.049	0.048	0.050	0.049	0.051
500	0.5	0.032	0.088	0.050	0.050	0.050	0.048	0.050	0.049	0.051
500	0.9	0.024	0.087	0.049	0.050	0.048	0.047	0.048	0.048	0.050
1000	0.0	0.050	0.084	0.049	0.049	0.049	0.049	0.049	0.048	0.050
1000	0.5	0.031	0.085	0.049	0.050	0.050	0.049	0.049	0.049	0.050
1000	0.9	0.023	0.085	0.050	0.051	0.050	0.048	0.049	0.050	0.051
					AR(1)-process					
50	0.0	0.050	0.023	0.040	0.036	0.039	0.029	0.041	0.034	0.048
50	0.5	0.057	0.012	0.039	0.040	0.046	0.024	0.026	0.034	0.025
50	0.9	0.094	0.002	0.001	0.018	0.000	0.006	0.000	0.004	0.000
100	0.0	0.051	0.059	0.045	0.043	0.046	0.040	0.048	0.043	0.054
100	0.5	0.054	0.029	0.052	0.052	0.063	0.038	0.057	0.047	0.065
100	0.9	0.075	0.002	0.007	0.045	0.014	0.026	0.002	0.044	0.000
200	0.0	0.051	0.091	0.048	0.047	0.048	0.045	0.049	0.047	0.053
200	0.5	0.052	0.057	0.056	0.056	0.061	0.047	0.061	0.051	0.068
200	0.9	0.063	0.006	0.033	0.055	0.053	0.037	0.031	0.073	0.032
500	0.0	0.050	0.095	0.049	0.049	0.050	0.048	0.050	0.049	0.051
500	0.5	0.051	0.083	0.056	0.057	0.058	0.052	0.058	0.053	0.062
500	0.9	0.056	0.018	0.057	0.064	0.081	0.047	0.090	0.068	0.116
1000	0.0	0.051	0.084	0.050	0.050	0.050	0.049	0.050	0.050	0.051
1000	0.5	0.050	0.082	0.055	0.056	0.057	0.054	0.056	0.054	0.058
1000	0.9	0.052	0.041	0.059	0.065	0.073	0.054	0.084	0.065	0.104

Note: Actual sizes when the nominal size equals 0.05.

Table 2: Size distortions of tests, raw-moments tests based on S-PITs

		S-PITs				INTs			
T	ρ	$\hat{m}_1$	$\hat{m}_2$	$\hat{m}_3$	$\hat{m}_4$	$\hat{m}_1$	$\hat{\mu}_2$	$\hat{s}$	$\hat{k}$
				standard normal forecast density					
∞		0.00	1.00	0.00	1.80	0.00	1.00	0.00	3.00
				normal forecast density, $\mu = -0.5$					
50	0.0	0.48	1.13	0.96	2.19	0.50	1.00	0.00	2.88
50	0.9	0.48	1.14	0.95	2.20	0.49	0.71	0.00	2.52
200	0.0	0.48	1.13	0.96	2.19	0.50	1.00	0.00	2.97
200	0.9	0.47	1.13	0.94	2.17	0.50	0.92	0.00	2.78
1000	0.0	0.48	1.13	0.96	2.19	0.50	1.00	0.00	3.00
1000	0.9	0.48	1.13	0.96	2.19	0.50	0.98	0.00	2.95
				normal forecast density, $\sigma = 2/3$					
50	0.0	0.00	1.46	0.00	3.26	0.00	2.26	0.00	2.88
50	0.9	0.00	1.46	0.00	3.25	−0.01	1.59	−0.01	2.53
200	0.0	0.00	1.46	0.00	3.25	0.00	2.25	0.00	2.97
200	0.9	0.00	1.46	−0.01	3.27	0.00	2.06	0.00	2.79
1000	0.0	0.00	1.46	0.00	3.25	0.00	2.25	0.00	2.99
1000	0.9	0.00	1.46	0.00	3.25	0.00	2.21	0.00	2.95
				normal forecast density, $\sigma = 3/2$					
50	0.0	0.00	0.60	0.00	0.76	0.00	0.44	0.00	2.88
50	0.9	−0.01	0.60	0.00	0.76	0.00	0.31	0.00	2.53
200	0.0	0.00	0.60	0.00	0.76	0.00	0.44	0.00	2.97
200	0.9	0.00	0.60	0.00	0.76	0.00	0.41	0.00	2.78
1000	0.0	0.00	0.60	0.00	0.76	0.00	0.44	0.00	3.00
1000	0.9	0.00	0.60	0.00	0.76	0.00	0.44	0.00	2.95
				two-piece normal forecast density, $\gamma = 0.8$					
50	0.0	0.06	1.01	−0.09	1.90	−0.03	1.30	−0.93	4.33
50	0.9	0.06	1.02	−0.10	1.91	−0.02	0.93	−0.52	3.07
200	0.0	0.06	1.01	−0.09	1.91	−0.03	1.29	−1.05	4.91
200	0.9	0.07	1.02	−0.09	1.91	−0.03	1.19	−0.84	4.05
1000	0.0	0.07	1.02	−0.09	1.91	−0.03	1.30	−1.09	5.12
1000	0.9	0.06	1.02	−0.09	1.91	−0.03	1.28	−1.03	4.82
				t-distributed forecast density, 5 degrees of freedom					
50	0.0	0.00	1.14	0.00	2.13	0.00	1.11	0.00	2.29
50	0.9	−0.01	1.14	−0.01	2.13	0.02	0.78	0.00	2.34
200	0.0	0.00	1.14	0.00	2.12	0.00	1.11	0.00	2.29
200	0.9	0.01	1.14	0.01	2.13	0.00	1.02	0.00	2.31
1000	0.0	0.00	1.14	0.00	2.13	0.00	1.11	0.00	2.29
1000	0.9	0.00	1.14	0.00	2.13	0.00	1.09	0.00	2.29
				normal mixture forecast density, $\sigma = 0.4$					
50	0.0	0.00	1.10	0.00	1.79	0.00	1.09	0.00	3.22
50	0.9	0.01	1.10	0.01	1.79	0.00	0.78	−0.01	2.63
200	0.0	0.00	1.10	0.00	1.80	0.00	1.09	0.00	3.72
200	0.9	0.01	1.10	0.01	1.80	0.00	1.00	−0.01	3.01
1000	0.0	0.00	1.10	0.00	1.80	0.00	1.09	0.00	3.89
1000	0.9	0.00	1.10	0.00	1.80	0.00	1.08	0.00	3.66

Note: $\hat{m}_i$ denotes i-th raw moment, $\hat{\mu}_2$ variance, $\hat{s}$ skewness, $\hat{k}$ kurtosis.

Table 3: Raw sample moments of S-PITs and sample moments of INTs for all forecast densities

T	ρ	$\hat{\beta}_{12}$	$\hat{\mu}_{34}$	$\hat{\alpha}_1$	$\hat{\alpha}_{12}^0$	$\hat{\alpha}_{12}$	$\hat{\alpha}_{123}^0$	$\hat{\alpha}_{123}$	$\hat{\alpha}_{1234}^0$	$\hat{\alpha}_{1234}$
				MA(1)-process						
50	0.0	0.87	0.05	0.90	0.81	0.76	0.71	0.60	0.58	0.43
50	0.5	0.57	0.05	0.60	0.42	0.25	0.30	0.13	0.17	0.07
50	0.9	0.51	0.05	0.52	0.34	0.19	0.23	0.10	0.13	0.06
100	0.0	0.99	0.05	1.00	0.99	0.99	0.98	0.98	0.97	0.95
100	0.5	0.89	0.05	0.93	0.85	0.79	0.77	0.65	0.64	0.49
100	0.9	0.85	0.05	0.89	0.79	0.71	0.68	0.54	0.52	0.34
200	0.0	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
200	0.5	1.00	0.05	1.00	1.00	0.99	0.99	0.99	0.98	0.97
200	0.9	0.99	0.06	1.00	0.99	0.99	0.98	0.96	0.96	0.93
500	0.0	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.5	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.9	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.0	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.5	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.9	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
				AR(1)-process						
50	0.0	0.86	0.05	0.90	0.81	0.75	0.71	0.59	0.59	0.42
50	0.5	0.41	0.05	0.33	0.24	0.10	0.17	0.06	0.09	0.04
50	0.9	0.11	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04
100	0.0	1.00	0.05	1.00	0.99	0.99	0.99	0.98	0.97	0.95
100	0.5	0.72	0.05	0.74	0.61	0.45	0.50	0.31	0.35	0.18
100	0.9	0.15	0.05	0.07	0.03	0.03	0.03	0.03	0.04	0.03
200	0.0	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
200	0.5	0.96	0.05	0.97	0.94	0.90	0.90	0.83	0.85	0.75
200	0.9	0.28	0.05	0.17	0.08	0.04	0.06	0.03	0.03	0.03
500	0.0	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.5	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.9	0.63	0.06	0.63	0.48	0.29	0.41	0.19	0.19	0.10
1000	0.0	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.5	1.00	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	0.9	0.91	0.05	0.93	0.87	0.78	0.82	0.66	0.71	0.53

Note: Raw-moments tests are based on S-PITs.

Table 4: Size-adjusted power, normal forecast density with $\mu = -0.5$

T	ρ	$\hat{\beta}_{12}$	$\hat{\mu}_{34}$	$\hat{\alpha}_1$	$\hat{\alpha}^0_{12}$	$\hat{\alpha}^0_{123}$	$\hat{\alpha}^0_{1234}$
			MA(1)-process				
50	0.0	0.96	0.05	0.05	0.67	0.57	0.43
50	0.5	0.93	0.05	0.05	0.58	0.53	0.32
50	0.9	0.91	0.05	0.05	0.53	0.49	0.26
100	0.0	1.00	0.05	0.05	0.97	0.94	0.92
100	0.5	1.00	0.05	0.05	0.93	0.90	0.82
100	0.9	1.00	0.05	0.05	0.90	0.87	0.74
200	0.0	1.00	0.05	0.05	1.00	1.00	1.00
200	0.5	1.00	0.05	0.05	1.00	1.00	1.00
200	0.9	1.00	0.05	0.05	1.00	1.00	0.99
500	0.0	1.00	0.05	0.05	1.00	1.00	1.00
500	0.5	1.00	0.05	0.05	1.00	1.00	1.00
500	0.9	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.0	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.5	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.9	1.00	0.05	0.05	1.00	1.00	1.00
			AR(1)-process				
50	0.0	0.96	0.05	0.05	0.68	0.58	0.43
50	0.5	0.88	0.05	0.05	0.46	0.43	0.21
50	0.9	0.33	0.07	0.06	0.03	0.02	0.02
100	0.0	1.00	0.05	0.05	0.97	0.94	0.92
100	0.5	0.99	0.05	0.05	0.87	0.83	0.67
100	0.9	0.55	0.06	0.06	0.07	0.06	0.01
200	0.0	1.00	0.05	0.05	1.00	1.00	1.00
200	0.5	1.00	0.05	0.05	1.00	1.00	0.99
200	0.9	0.83	0.06	0.06	0.34	0.32	0.04
500	0.0	1.00	0.05	0.05	1.00	1.00	1.00
500	0.5	1.00	0.05	0.05	1.00	1.00	1.00
500	0.9	0.99	0.06	0.05	0.92	0.90	0.62
1000	0.0	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.5	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.9	1.00	0.06	0.05	1.00	1.00	0.99

Note: Raw-moments tests are based on S-PITs.

Table 5: Size-adjusted power, normal forecast density with $\sigma = 2/3$

T	ρ	$\hat{\beta}_{12}$	$\hat{\mu}_{34}$	$\hat{\alpha}_1$	$\hat{\alpha}^0_{12}$	$\hat{\alpha}^0_{123}$	$\hat{\alpha}^0_{1234}$
			MA(1)-process				
50	0.0	0.93	0.05	0.05	0.88	0.77	0.51
50	0.5	0.73	0.05	0.05	0.73	0.65	0.34
50	0.9	0.65	0.04	0.05	0.64	0.58	0.27
100	0.0	1.00	0.05	0.05	1.00	1.00	0.98
100	0.5	0.99	0.05	0.05	0.99	0.98	0.92
100	0.9	0.97	0.04	0.05	0.98	0.96	0.85
200	0.0	1.00	0.05	0.05	1.00	1.00	1.00
200	0.5	1.00	0.05	0.05	1.00	1.00	1.00
200	0.9	1.00	0.04	0.06	1.00	1.00	1.00
500	0.0	1.00	0.05	0.05	1.00	1.00	1.00
500	0.5	1.00	0.05	0.05	1.00	1.00	1.00
500	0.9	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.0	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.5	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.9	1.00	0.05	0.05	1.00	1.00	1.00
			AR(1)-process				
50	0.0	0.94	0.05	0.06	0.88	0.79	0.52
50	0.5	0.51	0.05	0.05	0.56	0.53	0.24
50	0.9	0.09	0.03	0.05	0.03	0.01	0.00
100	0.0	1.00	0.05	0.05	1.00	1.00	0.98
100	0.5	0.89	0.05	0.05	0.94	0.92	0.79
100	0.9	0.16	0.03	0.05	0.10	0.09	0.01
200	0.0	1.00	0.05	0.05	1.00	1.00	1.00
200	0.5	1.00	0.04	0.05	1.00	1.00	1.00
200	0.9	0.33	0.04	0.04	0.40	0.38	0.07
500	0.0	1.00	0.05	0.05	1.00	1.00	1.00
500	0.5	1.00	0.04	0.05	1.00	1.00	1.00
500	0.9	0.81	0.04	0.05	0.90	0.88	0.79
1000	0.0	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.5	1.00	0.05	0.05	1.00	1.00	1.00
1000	0.9	0.99	0.04	0.05	1.00	1.00	0.99

Note: Raw-moments tests are based on S-PITs.

Table 6: Size-adjusted power of tests, normal forecast density with $\sigma = 3/2$

T	ρ	$\hat{\beta}_{12}$	$\hat{\mu}_{34}$	$\hat{\alpha}_1$	$\hat{\alpha}^0_{12}$	$\hat{\alpha}^0_{123}$	$\hat{\alpha}^0_{1234}$
			MA(1)-process				
50	0.0	0.27	0.24	0.07	0.07	0.26	0.23
50	0.5	0.25	0.24	0.07	0.06	0.19	0.14
50	0.9	0.26	0.23	0.07	0.06	0.16	0.12
100	0.0	0.41	0.59	0.10	0.08	0.56	0.52
100	0.5	0.37	0.56	0.08	0.06	0.45	0.40
100	0.9	0.36	0.50	0.08	0.07	0.40	0.34
200	0.0	0.58	0.94	0.15	0.12	0.89	0.88
200	0.5	0.55	0.91	0.11	0.09	0.82	0.81
200	0.9	0.52	0.87	0.10	0.08	0.79	0.77
500	0.0	0.89	1.00	0.30	0.24	1.00	1.00
500	0.5	0.84	1.00	0.19	0.15	1.00	1.00
500	0.9	0.81	1.00	0.17	0.14	1.00	1.00
1000	0.0	0.99	1.00	0.53	0.46	1.00	1.00
1000	0.5	0.98	1.00	0.34	0.28	1.00	1.00
1000	0.9	0.97	1.00	0.30	0.24	1.00	1.00
			AR(1)-process				
50	0.0	0.27	0.24	0.08	0.07	0.27	0.24
50	0.5	0.22	0.24	0.06	0.05	0.13	0.10
50	0.9	0.14	0.13	0.05	0.05	0.06	0.05
100	0.0	0.39	0.58	0.10	0.09	0.55	0.52
100	0.5	0.32	0.54	0.06	0.05	0.37	0.31
100	0.9	0.17	0.20	0.06	0.05	0.07	0.06
200	0.0	0.58	0.94	0.15	0.12	0.89	0.88
200	0.5	0.46	0.88	0.09	0.07	0.77	0.75
200	0.9	0.22	0.37	0.06	0.05	0.11	0.07
500	0.0	0.89	1.00	0.31	0.24	1.00	1.00
500	0.5	0.76	1.00	0.14	0.11	1.00	1.00
500	0.9	0.30	0.73	0.06	0.06	0.43	0.30
1000	0.0	0.99	1.00	0.53	0.45	1.00	1.00
1000	0.5	0.95	1.00	0.23	0.18	1.00	1.00
1000	0.9	0.45	0.95	0.08	0.06	0.87	0.85

Note: Raw-moments tests are based on S-PITs.

Table 7: Size-adjusted power, two-piece normal forecast density with $\gamma = 0.8$

T	ρ	$\hat{\beta}_{12}$	$\hat{\mu}_{34}$	$\hat{\alpha}_1$	$\hat{\alpha}^0_{12}$	$\hat{\alpha}^0_{123}$	$\hat{\alpha}^0_{1234}$
			MA(1)-process				
50	0.0	0.04	0.22	0.05	0.13	0.11	0.10
50	0.5	0.04	0.17	0.05	0.10	0.09	0.08
50	0.9	0.04	0.15	0.06	0.09	0.10	0.08
100	0.0	0.06	0.47	0.05	0.23	0.20	0.18
100	0.5	0.05	0.39	0.05	0.18	0.17	0.16
100	0.9	0.05	0.35	0.05	0.17	0.15	0.15
200	0.0	0.10	0.86	0.05	0.48	0.41	0.40
200	0.5	0.09	0.80	0.05	0.39	0.34	0.34
200	0.9	0.08	0.75	0.05	0.35	0.30	0.32
500	0.0	0.26	1.00	0.05	0.89	0.85	0.85
500	0.5	0.21	1.00	0.05	0.81	0.76	0.79
500	0.9	0.20	1.00	0.05	0.76	0.70	0.75
1000	0.0	0.54	1.00	0.05	1.00	0.99	0.99
1000	0.5	0.47	1.00	0.05	0.99	0.97	0.99
1000	0.9	0.42	1.00	0.05	0.98	0.96	0.98
			AR(1)-process				
50	0.0	0.04	0.22	0.05	0.12	0.11	0.10
50	0.5	0.04	0.16	0.05	0.08	0.08	0.07
50	0.9	0.05	0.06	0.05	0.03	0.03	0.06
100	0.0	0.06	0.48	0.05	0.23	0.20	0.19
100	0.5	0.04	0.36	0.05	0.14	0.13	0.13
100	0.9	0.04	0.08	0.06	0.03	0.03	0.05
200	0.0	0.09	0.85	0.05	0.46	0.40	0.39
200	0.5	0.07	0.75	0.05	0.31	0.27	0.28
200	0.9	0.04	0.13	0.06	0.04	0.04	0.05
500	0.0	0.25	1.00	0.05	0.89	0.84	0.84
500	0.5	0.15	1.00	0.05	0.72	0.65	0.72
500	0.9	0.04	0.44	0.05	0.12	0.12	0.16
1000	0.0	0.53	1.00	0.05	1.00	0.99	0.99
1000	0.5	0.34	1.00	0.06	0.97	0.94	0.97
1000	0.9	0.06	0.84	0.05	0.28	0.25	0.42

Note: Raw-moments tests are based on S-PITs.

Table 8: Size adjusted power, t-distributed forecast density with 5 degrees of freedom

T	ρ	$\hat{\beta}_{12}$	$\hat{\mu}_{34}$	$\hat{\alpha}_1$	$\hat{\alpha}^0_{12}$	$\hat{\alpha}^0_{123}$	$\hat{\alpha}^0_{1234}$
			MA(1)-process				
50	0.0	0.10	0.27	0.05	0.09	0.08	0.48
50	0.5	0.10	0.25	0.05	0.08	0.07	0.46
50	0.9	0.10	0.24	0.06	0.07	0.07	0.44
100	0.0	0.12	0.21	0.05	0.16	0.14	0.82
100	0.5	0.12	0.21	0.05	0.13	0.12	0.80
100	0.9	0.12	0.22	0.05	0.12	0.11	0.77
200	0.0	0.16	0.12	0.05	0.32	0.27	0.99
200	0.5	0.15	0.12	0.05	0.25	0.22	0.99
200	0.9	0.15	0.14	0.05	0.23	0.20	0.98
500	0.0	0.26	0.04	0.05	0.71	0.64	1.00
500	0.5	0.24	0.04	0.05	0.61	0.55	1.00
500	0.9	0.24	0.05	0.05	0.54	0.48	1.00
1000	0.0	0.43	0.03	0.05	0.96	0.93	1.00
1000	0.5	0.38	0.03	0.05	0.91	0.87	1.00
1000	0.9	0.35	0.03	0.05	0.87	0.82	1.00
			AR(1)-process				
50	0.0	0.10	0.25	0.05	0.09	0.08	0.47
50	0.5	0.09	0.26	0.05	0.06	0.06	0.41
50	0.9	0.10	0.15	0.07	0.05	0.04	0.10
100	0.0	0.12	0.21	0.05	0.16	0.14	0.82
100	0.5	0.11	0.22	0.05	0.10	0.09	0.75
100	0.9	0.08	0.22	0.07	0.03	0.03	0.16
200	0.0	0.16	0.11	0.05	0.32	0.27	0.99
200	0.5	0.14	0.14	0.05	0.20	0.18	0.98
200	0.9	0.10	0.27	0.06	0.04	0.03	0.32
500	0.0	0.26	0.04	0.05	0.71	0.64	1.00
500	0.5	0.20	0.05	0.05	0.50	0.44	1.00
500	0.9	0.12	0.21	0.05	0.08	0.07	0.89
1000	0.0	0.42	0.03	0.05	0.96	0.93	1.00
1000	0.5	0.31	0.03	0.05	0.84	0.78	1.00
1000	0.9	0.15	0.13	0.05	0.17	0.15	1.00

Note: Raw-moments tests are based on S-PITs.

Table 9: Size-adjusted power, normal mixture forecast density with $\sigma = 0.4$

	moments						p-values		
	S-PITs				INTs				
	$\hat{m}_1$	$\hat{m}_2$	$\hat{m}_3$	$\hat{m}_4$	$\hat{m}_1$	$\hat{\mu}_2$	$\hat{\alpha}^0_{1234}$	$\hat{\alpha}^0_{12}$	$\hat{\beta}_{12}$
$h = 2$	−0.04	0.88	−0.15	1.42	−0.06	0.78	0.027	0.073	0.085
$h = 3$	−0.07	0.88	−0.17	1.42	−0.07	0.77	0.039	0.175	0.191
$h = 4$	−0.09	0.89	−0.18	1.45	−0.09	0.77	0.197	0.233	0.286
$h = 5$	−0.10	0.91	−0.19	1.47	−0.10	0.77	0.166	0.351	0.380

Note: Raw-moments tests are based on S-PITs. Sample sizes equal $T = 385$. $\hat{m}_i$ denotes the i-th raw moment, $\hat{\mu}_2$ the variance.

Table 10: Moments of SPITs and INTs and test results for calibration of density forecasts for monthly Swiss Francs / U.S. Dollar exchange rate

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
15	2011	Crises, rescues, and policy transmission through international banks	Claudia M. Buch Cathérine Tahmee Koch Michael Koetter
16	2011	Substitution between net and gross settlement systems – A concern for financial stability?	Ben Craig Falko Fecht

17	2011	Recent developments in quantitative models of sovereign default	Nikolai Stähler
18	2011	Exchange rate dynamics, expectations, and monetary policy	Qianying Chen
19	2011	An information economics perspective on main bank relationships and firm R&D	D. Hoewer T. Schmidt, W. Sofka
20	2011	Foreign demand for euro banknotes issued in Germany: estimation using direct approaches	Nikolaus Bartzsch Gerhard Rösl Franz Seitz
21	2011	Foreign demand for euro banknotes issued in Germany: estimation using indirect approaches	Nikolaus Bartzsch Gerhard Rösl Franz Seitz
22	2011	Using cash to monitor liquidity – implications for payments, currency demand and withdrawal behavior	Ulf von Kalckreuth Tobias Schmidt Helmut Stix
23	2011	Home-field advantage or a matter of ambiguity aversion? Local bias among German individual investors	Markus Baltzer Oscar Stolper Andreas Walter
24	2011	Monetary transmission right from the start: on the information content of the eurosystem's main refinancing operations	Puriya Abbassi Dieter Nautz
25	2011	Output sensitivity of inflation in the euro area: indirect evidence from disaggregated consumer prices	Annette Fröhling Kirsten Lommatzsch
26	2011	Detecting multiple breaks in long memory: the case of U.S. inflation	Uwe Hassler Barbara Meller

27	2011	How do credit supply shocks propagate internationally? A GVAR approach	Sandra Eickmeier Tim Ng
28	2011	Reforming the labor market and improving competitiveness: an analysis for Spain using FiMod	Tim Schwarzmüller Nikolai Stähler
29	2011	Cross-border bank lending, risk aversion and the financial crisis	Cornelia Düwel, Rainer Frey Alexander Lipponer
30	2011	The use of tax havens in exemption regimes	Anna Gumpert James R. Hines, Jr. Monika Schnitzer
31	2011	Bank-related loan supply factors during the crisis: an analysis based on the German bank lending survey	Barno Blaes
32	2011	Evaluating the calibration of multi-step-ahead density forecasts using raw moments	Malte Knüppel

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
10	2011	Bank bailouts, interventions, and moral hazard	Lammertjan Dam Michael Koetter
11	2011	Improvements in rating models for the German corporate sector	Till Förstemann
12	2011	The effect of the interbank network structure on contagion and common shocks	Co-Pierre Georg
13	2011	Banks' management of the net interest margin: evidence from Germany	Christoph Memmel Andrea Schertler
14	2011	A hierarchical Archimedean copula for portfolio credit risk modelling	Natalia Puzanova

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