Order Invariant Tests for Proper Calibration of Multivariate Density Forecasts*

Jonas Dovern[1] and Hans Manner[†2]

[1]Alfred-Weber-Institute for Economics, Heidelberg University

[2]Institute of Economics, University of Graz

March 8, 2018

Abstract

Established tests for proper calibration of multivariate density forecasts based on Rosenblatt probability integral transforms can be manipulated by changing the order of variables in the forecasting model. We derive order invariant tests. The new tests are applicable to densities of arbitrary dimensions and can deal with parameter estimation uncertainty and dynamic misspecification. Monte Carlo simulations show that they often have superior power relative to established approaches. We use the tests to evaluate GARCH-based multivariate density forecasts for a vector of stock market returns.

JEL Classification: C12, C32, C52, C53

Keywords: density calibration, goodness-of-fit test, predictive density, Rosenblatt transformation

---

*A previous draft of this paper has been circulated under the title "Order Invariant Evaluation of Multivariate Density Forecasts". We would like to thank Malte Knüppel, Fabian Krüger, Maik Wolters, and the participants of seminars at RWTH Aachen University, University of Cologne, University of Graz, Heidelberg University, University of Kiel, Maastricht University, and the University of Warwick for their helpful comments on earlier drafts of the paper.

†Corresponding author: Universitätsstr. 15/F4, 8010 Graz, Austria. hans.manner@uni-graz.at

1 Introduction

The use of density forecasts has recently become common in many scientific fields (Gneiting and Katzfuss, 2014) and, in particular, in many areas of economics. Density forecasts are increasingly used, for instance, in the fields of energy economics (Huurman et al., 2012), demand management (Taylor, 2012), finance (Shackleton et al., 2010; Kitsul and Wright, 2013; Hallam and Olmo, 2014; Ghosh and Bera, 2015), and macroeconomics (Clark, 2011; Herbst and Schorfheide, 2012; Aastveit et al., 2014; Wolters, 2015; Amisano and Geweke, 2017). Many tasks, such as the computation of Value-at-Risk measures for portfolios containing multiple assets or the planning of production for a firm that serves many markets from one central production facility, require the construction and evaluation of *multivariate* density forecasts. Beginning with Smith (1985) and Diebold et al. (1999), the literature has proposed several approaches for testing whether a sequence of multivariate density forecasts coincides with the corresponding true densities (e. g., Clements and Smith, 2000, 2002; Corradi and Swanson, 2006a; Bai and Chen, 2008; González-Rivera and Yoldas, 2012; Ko and Park, 2013a; Ziegel and Gneiting, 2014).

This strand of literature has neglected two important issues (Ziegel and Gneiting, 2014 being an exception). First, established tests depend on the order of variables in a multivariate model. As a consequence, researchers need to present a myriad of (sometimes inconclusive) results. The issue even offers room for data mining if a researcher decides to report only those results which correspond to one particular ("preferred") order.[1] Second, all empirical applications and many of the theoretical results focus on the bivariate case. However, many applications, especially in finance, require models of higher dimensionality to be useful. We address both issues in this paper.

Following Diebold et al. (1999), the most commonly used approach for testing the calibration of multivariate density forecasts is based on the Rosenblatt (1952) probability integral transform (PIT). Examples include Clements and Smith (2000), Clements and Smith (2002), Ko and Park (2013a), and Ko and Park (2013b). This approach relies on the factorization of the multivariate forecast distribution into conditional distributions

[1] Essentially, this is one form of "data snooping" as discussed in White (2000).

because these, in turn, can be used to form independent PITs which, for well-specified models, follow a uniform distribution.[2] Suitable transformations of these conditional PITs then lead to a reduction of the multivariate testing problem to a univariate one. How well a testing approach works depends crucially on the chosen transformation. The univariate tests can be implemented using any goodness-of-fit test (e. g., Neyman's smooth test, the Kolmogorov-Smirnov test, or the Anderson-Darling test).

Two aspects that are of utmost importance in practical applications that involve parametric forecasting models are parameter estimation uncertainty and dynamic misspecification. Corradi and Swanson (2006b) present a comprehensive overview about these aspects. Dynamic misspecification refers to the fact that a forecaster potentially uses only a subset of the relevant information to form a conditional density forecast. In most fields of economics and finance this is very likely. Parameter estimation uncertainty arises whenever a parametric forecast model is used to construct density forecasts whose parameters are estimated based on finite samples. Whether estimation uncertainty has to be dealt with when evaluating a sequence of predictive densities depends on the exact formulation of the null hypothesis one is interested in. One common approach is to test whether the forecast distribution belongs to a given parametric density family with parameters evaluated at their pseudo-true values. Thus one tests whether the *underlying model specification* is suitable to generate appropriate conditional density forecasts. An alternative view, proposed in Rossi and Sekhposyan (2016), is to test for the ability of a model to produce correct forecast distributions evaluated at the estimated parameter values. This means one tests whether the *estimated model* is suitable to generate appropriate conditional density forecasts. The tests that we propose are, in general, capable of handling both views.

A number of approaches have been suggested in the literature to address parameter estimation uncertainty. Bai (2003) (for the univariate case) and Bai and Chen (2008) (for multivariate densities) combine the Kolmogorov test with Khmaladze's martingale

[2]Henceforth, we use the term 'conditional distributions' in a way that includes the one marginal distribution that is needed for the factorization of the joint distribution. In addition, we will refer to the PITs of the conditional distributions as *conditional PITs*.

transformation to obtain a test which is distribution free in the presence of estimated parameters. Andrews (1997) solves this problem by using a parametric bootstrap. Duan (2004) uses a suitable sequence of transformations to obtain a parametric test that does not suffer from parameter estimation error. More recently, Chen (2011) adapts a number of tests from the parameter-free context to parameter-dependent density forecast evaluation, building on insights from Newey (1985) and Tauchen (1985) in the in-sample case and from West (1996) and West and McCracken (1998) in the out-of-sample case. This is the approach that we use in our paper.

Dynamic misspecification causes the PITs to be serially correlated.[3] A number of papers propose tests that are robust against dynamic misspecification, i. e., preserve this misspecification under the null hypothesis. Usually, at the same time, those papers also consider the effects of parameter estimation uncertainty. Pioneering work in this context has been made by Corradi and Swanson (2006a), who show that a block bootstrap can be used to adjust Kolmogorov-type tests under such conditions. Their parametric approach has the advantage of a higher rate of convergence relative to the non-parametric test proposed in Hong and Li (2005). Both papers assume a stationary data generating process. In contrast, Rossi and Sekhposyan (2013) relax this assumption and propose a test for correct specification of density forecasts that is robust, in addition, to structural breaks.

Other papers test jointly for uniformity and *i.i.d.* property of the PITs, thereby testing the null hypothesis of completely calibrated densities (Mitchell and Wallis, 2011). The first contributions in this context use simultaneous tests. Berkowitz (2001), for instance, develops a likelihood ratio test for this joint null hypothesis, allowing for serial correlation of different order under the alternative. Hong et al. (2007) (using a non-parametric kernel density estimate) and Ko and Park (2013b) (using a parametric density estimate) propose a test based on the joint distribution of consecutive PITs. Lin and Wu (2017), in contrast, propose a sequential procedure consisting of a data driven smooth

[3]Serial dependence of PITs is also an issue in the context of multi-step forecasts (Knüppel, 2015). PITs based on h-step-ahead density forecasts will generally follow a moving average process of order $h-1$.

Table 1: Classification of Testing Problems

Treatment of dynamic misspecification:	Known Parameters/ Forecasts as Primitives	Estimated Parameters
Ignored	Diebold et al. (1998), Diebold et al. (1999), Clements and Smith (2000, 2002), Ko and Park (2013a)	Andrews (1997), Bai (2003), Chen (2011)
Accounted for	Knüppel (2015)	Corradi and Swanson (2006a)
Tested for	Berkowitz (2001), Rossi and Sekhposyan (2016)	Hong and Li (2005), Hong et al. (2007), Ko and Park (2013b), Lin and Wu (2017), Gonzlez-Rivera et al. (2011), González-Rivera and Yoldas (2012), González-Rivera and Sun (2015)

Notes: This table contains a non-exhaustive collection of papers taking different views on how parameter estimation and dynamic misspecification should be treated when testing the calibration of (predictive) densities.

test for uniformity, preceded by a test of serial independence of the PITs. An alternative approach that relies on so-called (generalized) autocontours has recently been proposed by Gonzlez-Rivera et al. (2011) and González-Rivera and Sun (2015).

Thus, there is a wide range of views about how density forecasts should be tested which we summarize in Table 1. In practice, the exact formulation of the testing problem depends on several factors such as the type of application or whether predictive densities are model-based or obtained via a survey. The methods that we propose below are compatible with any combinations of views about how dynamic misspecification and parameter uncertainty should be treated.

In this paper, we contribute to the literature on the evaluation of multivariate density forecasts in the following way. We propose new transformations of the conditional PITs which can be combined with *any* goodness-of-fit test for univariate distributions. Our preferred transformations (called Z_t^{2*} and $Z_t^{2\dagger}$ below) are constructed as the sum of squares of inverse normal transformations of the conditional PITs corresponding to all possible orders of the variables. The new transformations have a number of advantages. First, they are *order invariant*, a concept we define below, meaning that test results do not depend on the order of variables in the forecasting model. We show that the distortions in rejection rates caused by a tendentious application of the established tests, which are not order invariant, can be very substantial. Second, the new tests are applicable to densities of arbitrary dimension. Third, they have better power (relative to established tests) against a wide range of alternatives. Furthermore, we show that our tests can also be used when dynamic misspecification and parameter uncertainty have to be taken into account. In an application, we show that the new tests are helpful for testing the appropriateness of density forecasts based on sophisticated multivariate models for vectors of financial returns. In particular, we show that the potential for data mining is immense when using the established tests in practice and that our order invariant tests are required to draw unambiguous conclusions.

The remainder of this paper is organized as follows. In Section 2, we describe the testing problem, generalize established tests, and derive new tests to evaluate multivariate densities. In Section 3, we assess the properties of different tests by means of Monte Carlo simulations. In Section 4, we demonstrate the usefulness of the newly proposed tests in an application to forecasting the distribution of a vector of stock returns. Section 5 concludes. The Appendix contains all proofs and technical derivations as well as additional simulation results.

2 Theory

2.1 Setup and Test Hypothesis

Let Y_t be a vector-valued continuous random variable with true (but unknown) conditional distribution function (CDF) $G_{Y_t}(Y_t|\mathfrak{I}_{t-1})$, where $\mathfrak{I}_{t-1}$ denotes the relevant information set available at time $t-1$. Furthermore, we consider the predictive CDF $F_{Y_t}(Y_t|\Omega_{t-1},\theta_0)$ with corresponding conditional probability density function (PDF) $f_{Y_t}(Y_t|\Omega_{t-1},\theta_0)$, where $\Omega_{t-1} \subseteq \mathfrak{I}_{t-1}$ is the information set available to the researcher and θ_0 denotes a parameter vector with compact and finite parameter space Θ. This framework takes into account that density forecasts are often constructed using parametric models and allows for dynamic misspecification as defined, for instance, by Corradi and Swanson (2006a). For the time being, we treat θ_0 as known. In practice, the parameters have to be estimated from the data. We discuss below how this affects the testing problem and how we can take estimation uncertainty into account.

Consider a sample $\{Y_t, \Omega_{t-1}\}_{t=1}^n$ of which the first R observations can potentially be used to estimate θ_0 and the remaining P observations are used to evaluate the predictive densities generated by $F_{Y_t}(Y_t|\Omega_{t-1}, \theta_0)$. We are interested in testing whether the model $F_{Y_t}(Y_t|\Omega_{t-1}, \theta_0)$ is correctly specified in the sense that

$$H_0 : F_{Y_t}(Y_t|\Omega_{t-1}, \theta_0) = G_{Y_t}(Y_t|\Im_{t-1}). \tag{2.1}$$

To specify the null exactly, assumptions need to be made about whether θ_0 has to be estimated and about whether dynamic misspecification can be ignored, should be controlled for, or should jointly be tested. In Table 1 above, we provide an overview about the assumption made in the literature.

In the univariate case, H_0 implies that the probability integral transform (PIT), given by $U_t = F_{Y_t}(Y_t)$, is uniformly distributed between 0 and 1 (see, e.g., Gneiting and Katzfuss, 2014). This fact can be used to test for proper density calibration (e. g., Dawid, 1984; Diebold et al., 1998). The uniformity can be checked either by graphical methods, such as QQ-plots and histograms, or by goodness-of-fit tests, such as the Kolmogorov-Smirnov test, the Anderson-Darling test, or Neyman's smooth test.

Unfortunately, matters are more complicated in the multivariate case because the distribution of the multivariate PITs of Y_t under the null is unknown, in general, for $d > 1$ (see, e. g., Genest and Rivest, 2001). In essence, the task then is to reduce the multivariate problem to a univariate one by using suitable transformations. One way to approach this problem, proposed in Ziegel and Gneiting (2014), is to work with the Kendall distribution function for $F_{Y_t}(Y_t|\Omega_{t-1}, \theta_0)$.[4] The more commonly used way to approach this problem is based on the factorization of the joint densities into the product of conditional densities. Let $F_{Y_i}(Y_{i,t}|\Omega_{t-1}, \theta_0)$ denote the marginal CDF for the i^{th} element of Y_t and denote by

[4]We experimented with this approach but results were much worse (in terms of power) than those based on alternative approaches presented below. Therefore, we do not report them in this paper.

$F_{Y_i|Y_{i-1},\dots,Y_1}(Y_{i,t}|Y_{i-1,t},\dots,Y_{1,t},\Omega_{t-1},\theta_0)$ the conditional distribution of $Y_{i,t}$ given $Y_{i-1,t},\dots,Y_{1,t}$. Suppressing the dependence on Ω_{t-1} and θ_0, we can then write

$$f_{Y_t}(Y_t) = f_{Y_d|Y_{d-1},\dots,Y_1}(Y_{d,t}) \times \dots \times f_{Y_2|Y_1}(Y_{2,t}) \times f_{Y_1}(Y_{1,t}). \tag{2.2}$$

Rosenblatt (1952) shows that the sequences of *conditional PITs* for the elements of Y_t

$$\begin{aligned} U_t^1 &= F_{Y_1}(Y_{1,t}), \\ U_t^{2|1} &= F_{Y_2|Y_1}(Y_{2,t}), \\ &\vdots \\ U_t^{d|1,\dots,d-1} &= U_t^{d|1:d-1} = F_{Y_d|Y_{d-1},\dots,Y_1}(Y_{d,t}) \end{aligned} \tag{2.3}$$

are independent of each other and distributed $\mathcal{U}(0,1)$. The next step is to obtain a univariate testing problem based on this vector of PITs. Diebold et al. (1999) achieve the reduction of dimension by stacking all conditional PITs. More formally, if we let

$$S_t = [U_t^{d|1:d-1},\dots,U_t^1]', \tag{2.4}$$

then $S = [S'_{R+1}, S'_{R+2},\dots,S'_n]'$ constitutes a vector of variables which are uniformly distributed under H_0.

Instead of stacking the conditional PITs, a commonly used alternative is to transform the vector-valued random variable Y_t into a scalar random variable and to compute PITs for this transformed random variable. This is also the approach that we use below when developing our new tests. To formalize the idea, consider the general transform function $g_t(\cdot) : \mathbb{R}^d \to \mathbb{R}$ and define the transformed series $W_t = g_t(Y_t)$ with distribution function F_{W_t}. The PIT of W_t is given by

$$U_t^W = F_{W_t}(W_t). \tag{2.5}$$

Testing H_0 then is equivalent to testing whether $U_t^W \sim \mathcal{U}(0,1)$. In the absence of dynamic misspecification, the PITs are also independently distributed across time under H_0, i. e.,

$U_t^W \overset{i.i.d.}{\sim} \mathcal{U}(0,1)$. Mitchell and Wallis (2011) call density forecasts which satisfy both features *completely calibrated*.

2.2 The Order of Variables

So far, we have implicitly assumed that there exists a natural order of variables from 1 to d. This, of course, is not true as already mentioned in most papers on the topic (Diebold et al., 1999; Clements and Smith, 2002; Hong and Li, 2005; Ishida, 2005). Ordering the elements in Y_t in a different way will generally lead to different results because the Rosenblatt transform in (2.3) clearly depends on the order of the variables. Consequently, the outcome of a hypothesis test will depend on the selected order. This is an undesirable property for a test since a researcher who is interested in supporting or discrediting a certain model may perform the hypothesis test for all distinct orders and only report the outcome with the largest or smallest p-value.[5] While it is certainly true that for low-dimensional cases results for all possible permutations can be presented and discussed, this becomes quickly impossible for larger d. In addition, even when multiple test statistics are presented, it is unclear how an overall decision should be made.

We use the following notation for different permutations of the variables. Let π_k, for $k = 1, \ldots, d!$, be the set of all possible permutations of the data. Furthermore, let $\pi_k(i)$ denote the index (or "position") of variable i in the k^{th} permutation. Then, the conditional PITs under permutation π_k are given by

$$\begin{aligned}
U_t^{\pi_k(1)} &= F_{Y_{\pi_k(1)}}(Y_{\pi_k(1),t}) \\
U_t^{\pi_k(2)|\pi_k(1)} &= F_{Y_{\pi_k(2)}|Y_{\pi_k(1)}}(Y_{\pi_k(2),t}) \\
&\vdots \\
U_t^{\pi_k(d)|\pi_k(1):\pi_k(d-1)} &= F_{Y_{\pi_k(d)}|Y_{\pi_k(d-1),t},\ldots,Y_{\pi_k(1),t}}(Y_{\pi_k(d),t}).
\end{aligned} \tag{2.6}$$

The following definition formalizes the concept that the initial order of the data is not

[5] Note that in a different context, Clements and Hendry (1993) show that for commonly used selection criteria the choice of an optimal forecast model depends on the data transformation (e.g., levels or first differences) that is chosen, providing opportunity for a similar type of "data mining".

relevant for the test outcome.

Definition 1. *Let $T(\pi_k)$ be a test statistic based on $\{Y_t\}_{t=1}^n$ under permutation π_k. We call a test statistic $T(\pi_k)$ order invariant if $T(\pi_k) = T(\pi_j)$, $\forall\, k \neq j$.*

In the next section, we show that established transformations are order invariant only under very restrictive conditions and we derive new transformations that are always order invariant.

2.3 Established Transformations

Different transformations $g_t(\cdot)$ have been considered in the literature. Clements and Smith (2000) propose to evaluate density forecasts based on the product of the conditional PITs corresponding to one particular permutation of the variables.[6] In this case, the transformation function $g_t(\cdot)$ is given by

$$CS_{t,d} = g(Y_t) = \prod_{i=1}^{d} U_t^{i|1:i-1}, \tag{2.7}$$

where we define $U_t^{1|1:0} = U_t^1$. Ko and Park (2013a) explain why tests based on $CS_{t,d}$ have good power only against correlations lower than the hypothesized value. They suggest a location-adjusted version which does not suffer from this asymmetry and is given by

$$KP_{t,d} = g(Y_t) = \prod_{i=1}^{d} (U_t^{i|1:i-1} - 0.5). \tag{2.8}$$

Tests based on the transformations suggested by Diebold et al. (1999), Clements and Smith (2000), and Ko and Park (2013a) are not, in general, insensitive to the choice of the permutation.

Proposition 1. *Test statistics $T(\pi_k)$ based on $\{CS_{t,d}\}_{t=1}^n$, $\{KP_{t,d}\}_{t=1}^n$ and on the stacked transformation $\{S_{t,d}\}_{t=1}^n$ are order invariant if and only if under H_0 the variables $Y_{1,t}, \ldots, Y_{d,t}$ are independent, i. e., when $f_{Y_t}(Y_t) = f_{Y_1}(Y_{1,t}) \times \ldots \times f_{Y_d}(Y_{d,t})$.*

---

[6] Clements and Smith (2002) consider the ratio of the conditional PITs as an alternative. We do not discuss this transformation here since it is not obvious how to extend this approach to higher dimensions.

2.4 New Transformations

The first transformation that we propose leads to order-invariant test statistics under less restrictive conditions and forms the basis for additional transformations that always lead to order invariant tests. Consider the following transformation which is based on the squares of inverse normal transforms of the PITs:

$$Z_{t,d}^2 = \sum_{i=1}^{d} \left(\Phi^{-1} \left(U_t^{i|1:i-1} \right) \right)^2, \tag{2.9}$$

where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution.

Proposition 2. *Test statistics $T(\pi_k)$ based on $\{Z_t^2\}_{t=1}^n$ are order invariant if under H_0 $Y_t \sim \mathcal{N}(\mu, \Sigma)$, i. e., when Y_t follows a multivariate normal distribution with mean vector μ and covariance matrix Σ.*

Of course, Z_t^2 can also be used to test non-Gaussian densities. In this case, however, the corresponding test statistics are not generally order invariant, except for the obvious case of independence. The proof of Proposition 2 in the appendix shows that under the null hypothesis of normality it holds that $Z_t^2 = (Y_t - \mu)'\Sigma^{-1}(Y_t - \mu)$, which is the transformation proposed by Ishida (2005).

Ideally, however, we would like to obtain a transformation that is order invariant in general. A transformation that fulfils this criterion is similar in structure to Z_t^2 but considers the sum over all distinct conditional PITs. Consider all possible permutations π_k for $k = 1, \ldots, d!$ and the corresponding sequences of conditional PITs defined by (2.6). This yields a total of $d \times d!$ terms. However, the number of distinct PITs is only $d \times \sum_{k=0}^{d-1} \binom{d-1}{k} = d \times 2^{d-1}$.[7] To formalize, let γ_i^k, for $k = 1, \ldots, 2^{d-1}$, be the set of all sets of conditioning variables (including the empty set) corresponding to all distinct

[7] To see why, note that there are d variables that can each be ordered first to last. When a particular variable is ordered second, there are $\binom{d-1}{1} = d - 1$ possible conditioning variables. When this variable is ordered third, there are $\binom{d-1}{2}$ distinct sets of conditioning variables (each containing two of the other variables), and so on.

conditional PITs for $Y_{i,t}$. Then the suggested transformation has the form

$$Z_t^{2*} = \sum_{i=1}^{d} \sum_{k=1}^{2^{d-1}} \left(\Phi^{-1} \left(U_t^{i|\gamma_i^k} \right) \right)^2 . \tag{2.10}$$

Since all distinct conditional PITs enter into this transformation and, thus, the initial order of the variables in Y_t is irrelevant, order invariance is clearly ensured for any test statistic based on Z_t^{2*}.

When d increases, the number of terms entering Z_t^{2*} can become prohibitively large.[8] In this case, it appears sensible to use a transformation for which the number of terms grows only linearly with d. Such transformation that, at the same time, is always order invariant can be obtained by picking a suitable subset from the set of all distinct conditional PITs on which Z_t^{2*} is based. The transformation that we propose considers only such conditional PITs corresponding to each $Y_{i,t}$ which are conditional on all other variables. Denoting those conditional PITs by $U_t^{i|-i}$, the transformation is given by

$$Z_t^{2\dagger} = \sum_{i=1}^{d} \left(\Phi^{-1} \left(U_t^{i|-i} \right) \right)^2 . \tag{2.11}$$

We think that this particular choice has some merits since the considered subset of conditional PITs contains rich information about the dependence structure of the elements of Y_t.

2.5 Distribution of Transformations

To test H_0 based on the transformations, we need to know their distribution under the null hypothesis as indicated by (2.5). The distributions of the established transformations are as follows. S_t is simply a vector of independent uniformly distributed random variables under H_0. Clements and Smith (2000) derive the distribution of CS_t for $d = 2, 3$. In the appendix, we show that for arbitrary d its distribution under H_0 is described by the

[8] With $d = 10$, for instance, the number of terms equals 5,120. **[Can we delete this footnote?]**

following PDF and CDF:[9]

$$f_{CS_d}(CS_{t,d}) = \frac{(-1)^{d-1}}{(d-1)!}\log^{d-1}(CS_{t,d})$$
$$F_{CS_d}(CS_{t,d}) = CS_{t,d}\sum_{i=0}^{d-1} f_{CS_{d-i}}(CS_{t,d})$$

Note that for $d = 2, 3$ the densities derived in Clements and Smith (2000) is recovered.

Ko and Park (2013a) provide the distribution of KP_t for $d = 2$.[10] In the appendix, we show that for arbitrary d its distribution under H_0 is described by the following PDF and CDF:

$$f_{KP_d}(KP_{t,d}) = \frac{2^{d-1}}{(d-1)!}\log^{d-1}\left|\frac{1}{2^d KP_{t,d}}\right|$$
$$F_{KP_d}(KP_{t,d}) = KP_{t,d}2^{d-1}\sum_{i=1}^{d}\frac{1}{(d-i)!}\log^{d-i}\left|\frac{1}{2^d KP_{t,d}}\right| + \frac{1}{2}$$

In the next two subsections, we derive the distributions of the new transformations. We distinguish two cases. In the first case, we do not make any assumptions about the distribution of Y_t, except that it is continuous. The corresponding results include, for instance, cases in which H_0 implies non-Gaussian parametric distributions of Y_t or its distribution is not available analytically so that the conditional PITs have to be calculated numerically. In the second case, we show that if Y_t is normally distributed the distributions of Z^{2*} and $Z^{2\dagger}$ become much more tractable.

2.5.1 Distributions of New Transformations: General Case

As shown in Section 2.1, the different conditional PITs for one particular permutation are independent. Therefore, H_0 implies that $Z^2_{t,d} \sim \chi^2_d$, where χ^2_d denotes the chi-squared distribution with d degrees of freedom. Denoting by $F_{\chi^2_d}$ the CDF of this distribution, $U_t^{Z^2} = F_{\chi^2_d}(Z^2_{t,d})$ is distributed $\mathcal{U}(0, 1)$ under H_0.

[9] Note that Clements and Smith (2000) suggest a sequential algorithm based on which the PDFs and CDFs for $d > 3$ could also be obtained.

[10] Note that the density given in the appendix of Ko and Park (2013a) needs to be multiplied by a factor of 2.

${Z^2_{t,d}}^*$ is similar to $Z^2_{t,d}$. However, due to the fact that the summands in (2.10) are not independent in general, Z^{2*}_t no longer follows a χ^2 distribution under H_0. The same argument apply in the case of $Z^{2\dagger}_t$. However, we can straightforwardly obtain the distributions of the transformations by Monte Carlo simulation as long as it is possible to generate random draws from the density model under H_0.

The following algorithm describes how the distributions of $Z^\bullet_t = \{Z^{2*}_t, Z^{2\dagger}_t\}$ can be approximated numerically to compute $U^{Z^\bullet}_t$:[11]

1. Generate B' conditional forecasts, $y^{(b)}_t$, based on the model under H_0, i.e., draw repeatedly from the conditional predictive densities $f_{Y_t}(\cdot)$.
2. Given $f_{Y_t}(\cdot)$, construct $U^{i|\gamma^k_i}_{t,(b)}$, $\forall i,k$, for $b = 1,\ldots,B'$ along the lines described in Section 2.4.
3. Compute the corresponding inverse PITs as $\Phi^{-1}\left(U^{i|\gamma^k_i}_{t,(b)}\right)$.
4. Based on the set of $\Phi^{-1}\left(U^{i|\gamma^k_i}_{t,(b)}\right)$, compute $Z^\bullet_{t,(b)}$ using (2.10) or (2.11), respectively.
5. Compute $U^{Z^\bullet}_t = Pr\left(Z^\bullet_t < Z^\bullet_{t,(b)}\right)$ by simply counting how often the transformed statistic based on the actual realizations is smaller than the transformed statistics based on conditional forecasts that are generated under H_0.

If H_0 holds, $U^{Z^\bullet}_t$ is distributed $\mathcal{U}(0,1)$.

2.5.2 Distributions of New Transformations: Gaussian Case

Under the assumption that Y_t is normally distributed, the distributions of Z^{2*}_t and $Z^{2\dagger}_t$ are available analytically and do not need to be simulated. In this case, the terms $\Phi^{-1}\left(U^{i|\gamma^k_i}_t\right)$ jointly follow a multivariate normal distribution. However, since their marginal distributions are not independent, the transformations do not follow a chi-squared distribution but a mixture of chi-squared distributions, where the weights depend on the dependence structure of the $\Phi^{-1}\left(U^{i|\gamma^k_i}_t\right)$. For Z^{2*}_t, we obtain the following result:

[11] Note that this algorithm is exclusively used to approximate this distribution for a given parameter value that can be either θ_0 or an estimate $\hat{\theta}$. We show below how parameter uncertainty is accounted for in the latter case at another step of the testing procedure.

Proposition 3. *Let* $Y_t \sim \mathcal{N}(\mu, \Sigma)$*. Then* Z_t^{2*} *is distributed as* $\sum_{i=1}^{d} \lambda_i Z_i^2$*, for independent* $\mathcal{N}(0,1)$ *variables* $Z_1, \ldots, Z_d$ *and* $\lambda_1, \ldots, \lambda_d$ *the non-zero eigenvalues of the rank* d *matrix* R_{Z^*}*, which is the correlation matrix of all distinct terms* $\Phi^{-1}\left(U_t^{i|\gamma_i^k}\right) \forall\, i, k$ *entering* Z_t^{2*}*. A typical entry of* R_{Z^*} *is given by*

$$Corr\left(\Phi^{-1}\left(U_t^{i|\gamma_i^k}\right), \Phi^{-1}\left(U_t^{j|\gamma_j^l}\right)\right) = (\Sigma_{i,i} - \Sigma_{i,\gamma_i^k}\Sigma^{-1}_{\gamma_i^k,\gamma_i^k}\Sigma_{\gamma_i^k,i})^{-1/2}(\Sigma_{j,j} - \Sigma_{j,\gamma_j^l}\Sigma^{-1}_{\gamma_j^l,\gamma_j^l}\Sigma_{\gamma_j^l,j})^{-1/2}\times$$
$$(\Sigma_{i,j} - \Sigma_{j,\gamma_j^l}\Sigma^{-1}_{\gamma_j^l,\gamma_j^l}\Sigma_{\gamma_j^l,i} - \Sigma_{i,\gamma_i^k}\Sigma^{-1}_{\gamma_i^k,\gamma_i^k}\Sigma_{\gamma_i^k,j} + \Sigma_{i,\gamma_i^k}\Sigma^{-1}_{\gamma_i^k,\gamma_i^k}\Sigma_{\gamma_i^k,\gamma_j^l}\Sigma^{-1}_{\gamma_j^l,\gamma_j^l}\Sigma_{\gamma_j^l,j}),$$

where the $\Sigma_{r,c}$ $(r, c \in \{i, \gamma_i^k\})$ *are scalars, vectors, and matrices containing those elements of* Σ *that are defined by the row(s) corresponding to the variable(s) defined by* r *and the column(s) corresponding to the variable(s) defined by* c*.*

The distribution of $Z_t^{2\dagger}$ in the Gaussian case is given by the following corollary:

Corollary 1. *Let* $Y_t \sim \mathcal{N}(\mu, \Sigma)$*. Then* $Z_t^{2\dagger}$ *is distributed as* $\sum_{i=1}^{d} \lambda_i Z_i^2$*, for independent* $\mathcal{N}(0,1)$ *variables* $Z_1, \ldots, Z_d$ *and* $\lambda_1, \ldots, \lambda_d$ *the eigenvalues of the matrix* $R_{Z^\dagger}$*, which is the correlation matrix of all terms* $\Phi^{-1}\left(U_t^{i|-i}\right)$ *for* $i = 1, \ldots, d$ *entering* $Z_t^{2\dagger}$*. A typical entry of* $R_{Z^\dagger}$ *is given by*

$$Corr\left(\Phi^{-1}\left(U_t^{i|-i}\right), \Phi^{-1}\left(U_t^{j|-j}\right)\right) = (\Sigma_{i,i} - \Sigma_{i,-i}\Sigma^{-1}_{-i,-i}\Sigma_{-i,i})^{-1/2}(\Sigma_{j,j} - \Sigma_{j,-j}\Sigma^{-1}_{-j,-j}\Sigma_{-j,j})^{-1/2}\times$$
$$(\Sigma_{i,j} - \Sigma_{j,-j}\Sigma^{-1}_{-j,-j}\Sigma_{-j,i} - \Sigma_{i,-i}\Sigma^{-1}_{-i,-i}\Sigma_{-i,j} + \Sigma_{i,-i}\Sigma^{-1}_{\gamma_i^k,-i}\Sigma_{-i,-j}\Sigma^{-1}_{-j,-j}\Sigma_{-j,j}),$$

where the index $-i$ *denotes all rows/columns of* Σ *except for the* i^{th} *one.*

Note that, of course, $Z_{t,d}^2 \sim \chi_d^2$ continues to hold under H_0 in the Gaussian case.

2.6 Tests for Proper Calibration

In this section, we describe how we can construct tests of H_0 based on the transformations derived in the previous section. Depending on the formulation of the null hypothesis, this involves either testing that $U_t^W \sim \mathcal{U}(0,1)$ or jointly testing that $U_t^W \sim \mathcal{U}(0,1)$ and independent across time. In the former case, we use Neyman's (1937) smooth test and distinguish two cases. First, we do not account for potential autocorrelation in U_t^W and assume that we know the parameters of the model that is used to generate the densities.

Then, we construct tests that are robust against the existence of autocorrelation in U_t^W and assume that we have to estimate the parameters of the density model. To test the joint hypothesis in the second case, we resort to the concept of Generalized Autocontour (G-ACR) suggested by González-Rivera and Sun (2015).

2.6.1 Known Parameters and no Dynamic Misspecification

The first case that we consider ignores the issues of parameter uncertainty and dynamic misspecification. We assume θ_0 is known and $\Omega_{t-1} = \Im_{t-1}$. The null hypothesis is $U_t^W \sim \mathcal{U}(0,1)$.[12] Many tests can be used in this context. We follow, Bera and Ghosh (2002) and De Gooijer (2007) who advocate testing uniformity with Neyman's (1937) smooth test.

Let's take a closer look at Neyman's smooth test. Consider the alternative family of smooth distributions

$$s(u) = b_0 \exp\left(\sum_{i=1}^{k} b_i \psi_i(u)\right), \quad u \in [0,1], \tag{2.12}$$

with b_0 a normalization constant and ψ_i the orthonormal Legendre polynomials. Assuming that F_{W_t} is a member of the family of distributions defined by (2.12), testing uniformity—and hence H_0—boils down to testing $b_i = 0$ for all $i = 1, \ldots, k$. Here we consider the first four Legendre polynomials, but in principle one could also determine the number of polynomials in a data-driven fashion as suggested by Ledwina (1994) and applied, for instance, by Lin and Wu (2017).[13]

A score test is easily computed as follows. Denoting the vector of (log-)scores by $\xi_t = [\psi_1(U_t^W), \ldots, \psi_4(U_t^W)]'$ it follows that under H_0

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \xi_t \xrightarrow{d} N(0, I_4), \tag{2.13}$$

[12] This approach can also be used if autocorrelation is not of concern and the tested densities are not model based (for instance, because they are obtained from a survey).

[13] The first four Legendre polynomials are given by $\psi_1(u) = \sqrt{12}(u-1/2)$, $\psi_2(u) = \sqrt{5}(6(u-1/2)^2 - 1/2)$, $\psi_3(u) = \sqrt{7}\left(20(u-1/2)^3 - 3(u-1/2)\right)$, and $\psi_4(u) = 210(u-1/4)^4 - 45(u-1/2)^2 + 9/8$. **[Can we delete this footnote?]**

where I_4 is the 4×4 identity matrix. The Neyman smooth test statistic is then given by

$$NST = P^{-1} \left[\sum_{t=R+1}^{n} \xi_t \right]' \left[\sum_{t=R+1}^{n} \xi_t \right], \tag{2.14}$$

which follows a χ^2_k distribution under H_0. This result, however, only holds when the model parameters are assumed to be known and when there is no dynamic misspecification.

2.6.2 Estimated Parameters and Accounting for Dynamic Misspecification

Parameter uncertainty and dynamic misspecification are often relevant in practice when parametric forecast models are used and the DGP of the variables to be forecast (including the true values of the relevant parameters) is unknown to the forecaster. Ignoring both issues will, in general, lead to oversized tests in an out-of-sample evaluation. Thus, we now assume estimates of the parameters, $\hat{\theta}$, are obtained using a $\sqrt{T}$-consistent estimator and $\Omega_{t-1} \subset \Im_{t-1}$. We test again if $U_t^W \sim \mathcal{U}(0,1)$.

We adjust Neyman's smooth test by relying on results in West (1996) and West and McCracken (1998) to derive suitable tests in the presence of parameter uncertainty *and* potential dynamic misspecification. Recall that we split our n observations into R in-sample observations, which we use to estimate the parameters, and P out-of-sample observations, which we use to evaluate the forecast model. Let $\hat{\xi}_t = [\psi_1(\hat{U}_t^W), \ldots, \psi_4(\hat{U}_t^W)]'$ denote the Legendre polynomials in the *estimated* PITs of the (univariate) transformed series W_t, where $\hat{U}_t^W$ has been computed using the in-sample parameter estimates. It follows that under H_0 the elements of $\hat{\xi}_t$ are no longer independently distributed with unit variance as in (2.13), but

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \hat{\xi}_t \xrightarrow{d} N(0, \Sigma), \tag{2.15}$$

where (using the notation in Chen, 2011)

$$\Sigma = S^* - \eta_1 \left(D^* A^{-1} C' + C A^{-1} D^{*\prime} \right) + \eta_2 \left(C A^{-1} B^* A^{-1} C' \right). \tag{2.16}$$

Given the score function $s_t = \frac{\partial}{\partial \theta_0} \ln f_t(Y_t | \Omega_{t-1}, \theta_0)$, the elements of Σ are given by $A = E(\frac{\partial}{\partial \theta_0} s_t)$, $B = E(s_t s_t')$, $C = E(\frac{\partial}{\partial \theta_0} \xi_t)$, $D = E(\xi_t s_t')$, $S^* = \sum_{k=-\infty}^{\infty} E(\xi_t \xi_{t-k}')$, $B^* = \sum_{k=-\infty}^{\infty} E(s_t s_{t-k}')$, and $D^* = \sum_{k=-\infty}^{\infty} E(\xi_t s_{t-k}')$. The constants η_1 and η_2 are determined by the sampling scheme (fixed, rolling, or recursive) used to estimate the parameters and the limiting value of the ratio of in-sample and out-of-sample observations $\lambda = \lim_{n\to\infty} P/R$; see Chen (2011) for the precise formulas.

In order to avoid the evaluation of the matrices A and C (the latter of which may be particularly tedious to obtain), we use the fact that the equalities $A + B = 0$ and $C + D = 0$ continue to hold even under dynamic misspecification, see White (1994).[14] Thus, we can rewrite (2.16) as

$$\Sigma = S^* - \eta_1 \left(D^* B^{-1} D' + D B^{-1} D^{*\prime} \right) + \eta_2 \left(D B^{-1} B^* B^{-1} D' \right). \tag{2.17}$$

The matrices B and D can be estimated straightforwardly by their sample counterparts. In contrast, S^*, B^*, and D^* need to be estimated by an appropriate estimator that is autocorrelation consistent. While in principle the widely used HAC estimator by Newey and West (1987) could be used, we found results in finite samples to be better (in terms of size) if we use the quadratic spectral estimator proposed by Andrews (1991). Neyman's smooth test statistic is then given by

$$NST = P^{-1} \left[\sum_{t=R+1}^{n} \hat{\xi}_t \right]' \hat{\Sigma}^{-1} \left[\sum_{t=R+1}^{n} \hat{\xi}_t \right], \tag{2.18}$$

which follows a χ_k^2 (χ_4^2 in our case) distribution under H_0. In addition, we can consider two intermediate cases. In the absence of dynamic misspecification, it holds that $B^* = B$, $D^* = D$ and $S^* = I_4$; so (2.16) simplifies to $\Sigma = I_4 + (\eta_2 - 2\eta_1) D B^{-1} D'$. In the absence of parameter uncertainty, we obtain $\Sigma = S^*$.

[14] Note, however, that in this case they cannot be interpreted as (generalized) information matrix equalities. **[Can we delete this footnote?]**

2.6.3 Estimated Parameters and Testing for Independence

To jointly test for proper calibration and temporal independence of PITs, we have to test the null hypothesis $U_t^W \overset{i.i.d.}{\sim} \mathcal{U}(0,1)$. The G-ACR proposed in González-Rivera and Sun (2015) can be used to test this hypothesis and seems to work well in practice. Therefore, we adapt this approach to our testing problem. The idea is the following: Consider the pair of PITs $[U_t^W, U_{t-k}^W] \subset R^2$. Under the null hypothesis, the G-ACR for a level of $\alpha \in (0,1)$ is given by the set of points B such that

$$\begin{aligned} \text{G-ACR}_{\alpha,k} = &\{B(U_t^W, U_{t-k}^W) \subset R^2 || \; 0 \leq U_t^W \leq \sqrt{\alpha} \\ &\text{and } 0 \leq U_{t-k}^W \leq \sqrt{\alpha}, \text{ s.t.: } U_t^W \times U_{t-k}^W \leq \alpha\}. \end{aligned}$$

Note that we apply the G-ACR method to the PITs of the transformed variable(s) which implies that we stay in a univariate framework.[15] A test is then based on the comparison of the empirical frequency with which the (current and lagged) PITs fall into the square defined by the G-ACR to its distribution under the null hypothesis (potentially involving results for a range of values for α and k). **[Schau mal, ob Du meinst, dass das reicht. Sonst kannst Du ja vlt. noch was ergnzen ...]**

3 Monte Carlo Simulations

We use Monte Carlo simulations to analyze how severe the size and power distortions caused by data mining can be in the case of the order-dependent approaches and how the size and power of the tests based on the different transformations compare.

[15]We would loose the order invariance of our results if we directly compute the G-ACRs for the multivariate densities because the multivariate version proposed by González-Rivera and Sun (2015) also relies on the Rosenblatt transform and, thus, one particular order of the variables.

3.1 Simulation Setup

3.1.1 Known Parameters and no Dynamic Misspecification

For the simulations *without* dynamic misspecifications and parameter uncertainty, we consider multivariate normal and t distributions under the null. In the first case, we assume that the data generating process (DGP) under the null hypothesis is given by

$$y_t = \varepsilon_t, \qquad \text{with } \varepsilon \sim \mathcal{N}(0, \Sigma). \tag{3.1}$$

The $d \times d$ covariance matrix Σ is constructed such that all elements of y_t have unit variances ($\sigma_i^2 = 1$ for $i = 1, \dots, d$) and the correlation between any two elements of y_t is equal to 0.5 ($\rho_{ij} = 0.5$ for all $i \neq j$). We consider different dimensions up to $d = 50$ and sample sizes of $P = \{50, 100, 200\}$. **Throughout the paper, we use 10,000 iterations for our Monte Carlo simulations.** We consider seven alternative DGPs which imply different (combinations of) deviations from H_0:

- **Alternative 1 (H_1):** The innovations are generated from a multivariate normal distribution with $\sigma_i = 1.1$ and $\rho_{ij} = 0.5$.
- **Alternative 2 (H_2):** The innovations are generated from a multivariate normal distribution with $\rho_{ij} = 0.5$, $\sigma_i = 1.1$ for $i = 1, \dots, \lfloor d/3 \rfloor$, and $\sigma_i^2 = 1.0$ for $i = \lfloor d/3 \rfloor + 1, \dots, d$.
- **Alternative 3 (H_3):** The innovations are generated from a multivariate normal distribution with $\sigma_i = 1.0$ and $\rho_{ij} = 0.4$.
- **Alternative 4 (H_4):** The innovations are generated from a multivariate normal distribution with $\sigma_i = 1.1$ and $\rho_{ij} = 0.4$.
- **Alternative 5 (H_5):** The innovations are generated from a multivariate t distribution with 8 degrees of freedom with $\sigma_i = 1.0$ and $\rho_{ij} = 0.5$.
- **Alternative 6 (H_6):** The innovations are generated from a multivariate t distribution with 8 degrees of freedom with $\sigma_i = 1.1$ and $\rho_{ij} = 0.4$.

- **Alternative 7 (H_7):** The innovations are generated from a multivariate Gaussian CCC-GARCH(1,1) model which we parameterize such that the unconditional covariance matrix is equal to that under H_0.

In the second case, we assume that y_t follows a multivariate t distribution with 5 degrees of freedom under the null and in the case of alternatives 1–4, keeping variances and correlations as in the simulations with the Gaussian null model. Likewise, we change alternatives 5 and 6 to be based on multivariate normal distributions. We use Neyman's smooth test as described in Section 2.6.1 for this set of simulations.

3.1.2 Estimated Parameters and Accounting for Dynamic Misspecification

We consider two types of simulations *with* dynamic misspecification and estimated parameters. First, we generate data by the following homoskedastic dynamic model:[16]

$$y_t = \alpha I_{d\times d} y_{t-1} + \varepsilon_t, \qquad \text{with } \varepsilon \sim \mathcal{N}(0, \Sigma) \tag{3.2}$$

If not noted otherwise, we set $\alpha = 0.5$. To simulate dynamic misspecification, we generate predictive densities using the static model without the autoregressive term (assuming $\alpha = 0$) under the null. Since parameters are now estimated, we consider only alternative 5 from the above list. Rejecting the null hypothesis might be very hard in small samples if the alternative is a t distribution with 8 degrees of freedom. Therefore, we consider an additional alternative with a more substantial deviation from H_0:

- **Alternative 8 (H_8):** The innovations are generated from a multivariate t distribution with 4 degrees of freedom with $\sigma_i = 1.0$ and $\rho_{ij} = 0.5$.

In addition, we run two alternatives where we generate data with GARCH models. The first assumes no dynamic misspecification in the mean equation (by setting $\alpha = 0$ when simulating the data), even though we control for it when testing. The second involves GARCH effects and dynamic misspecification:

[16]Results assuming a dynamic moving average structure, $y_t = 0.8 I_{d\times d}\varepsilon_{t-1} + \varepsilon_t$, are very similar and not reported below.

- **Alternative 9 (H_9):** The innovations are generated from a multivariate Gaussian CCC-GARCH(1,1) model which we parameterize such that the unconditional covariance matrix is equal to that under H_0. In addition, we set $\alpha = 0$ to "switch off" dynamic misspecification.

- **Alternative 10 (H_{10}):** The innovations are generated from a multivariate Gaussian CCC-GARCH(1,1) model which we parameterize such that the unconditional covariance matrix is equal to that under H_0.

We use a fixed estimation scheme and set $\lambda = 1/4$ to determine the size of the estimation sample $R = P/\lambda$.

Alternatively, we assume that the innovations under the null follow a Gaussian CCC-GARCH(1,1) process while they are generated by CCC-t-GARCH(1,1) models with 8 and 4 degrees of freedom under the alternatives:

- **Alternative 11 (H_{11}):** The innovations are generated from a multivariate CCC-t-GARCH(1,1) model with 8 degrees of freedom.

- **Alternative 12 (H_{12}):** The innovations are generated from a multivariate CCC-t-GARCH(1,1) model with 4 degrees of freedom.

Here, we consider one set of simulations where dynamic misspecification is present and one set of simulations where it is not. Since we need more data to reliably estimate the GARCH models and want to hold $\lambda = 1/4$ fix, we consider also $P = 500$ in these simulations. Again, we use a fixed estimation scheme and the modified smooth test.

3.1.3 Estimated Parameters and Testing for Independence

Here, we use similar DGPs as in the previous section. What is different is the formulation of the null hypothesis (compare Section 2.6.3) and that we rely on G-ACRs to implement a joint test of the combined null hypothesis. The null model is the multivariate normal distribution given by (3.1). We consider the following alternatives:

- **Alternative 13 (H_{13}):** We induce dynamic misspecification by generating data according to (3.2) with $\alpha = 0.5$.

- **Alternative 14 (H_{14}):** Equal to H_5 without dynamic misspecification ($\alpha = 0$).
- **Alternative 15 (H_{15}):** Equal to H_5 with dynamic misspecification ($\alpha = 0.5$).
- **Alternative 16 (H_{16}):** Equal to H_8 with dynamic misspecification ($\alpha = 0.5$).

Like above, we consider evaluation samples of size P={50,100,200} and set $\lambda = 1/4$. We show results in Table B.3 in the appendix. All tests are properly sized. If only distributional deviations from the null model but no dynamic misspecification is present (H_{14}) the new transformations clearly outperform the established ones also with the G-ACR approach. When both features are present under the alternative, the new test and tests based on S clearly outperform the other two tests. The ranking of the former depend on whether the distributional deviation from the null is moderate (H_{15}, power is highest for S) or strong (H_{16}, power is highest for the new tests). The new tests are clearly dominated by tests based on S and interestingly also CS when the only deviation from the null hypothesis is given by dynamic misspecification (H_{13}); hence in this—rather unlikely—case there is apparently a price that one has to pay to make tests order invariant.

3.2 Potential for Data Mining

In this section, we present results that show whether considering different permutations of the data can have a serious impact on the outcomes of the tests that are not order invariant. The idea is the following: a researcher who wants to discredit (support) the hypothesis that a particular model produces good density forecasts could, in principle, search across all permutations and select the one which yields the highest (lowest) test statistic. We present results for H_0 and H_5 based on $P = 100$ and the assumption that parameters are known; results are similar for other settings and available upon request.

The left part of Figure 1 shows how severe data mining can be under the null hypothesis. The solid line indicates the nominal size of 5 % which, as we show below, is obtained when tests are applied properly (meaning that the order of variables is chosen randomly). The other lines refer to the rejection frequencies that we obtain for the tests based on S,

(a) Under H_0 (b) Under H_5

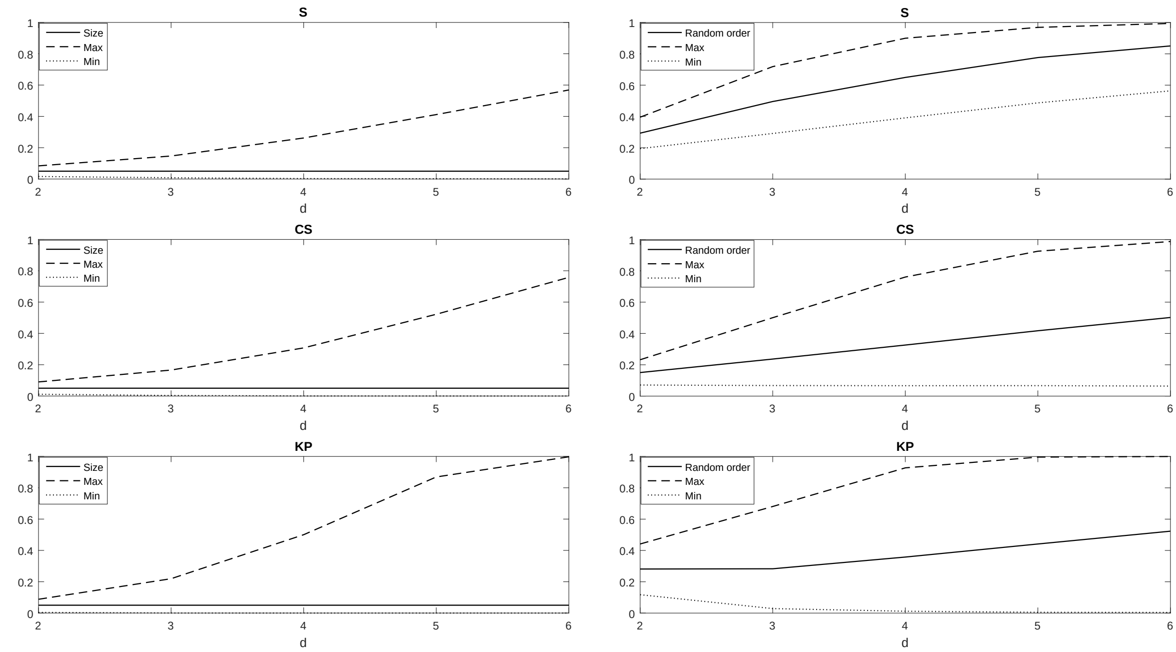

Figure 1: Scope for data mining.

CS, and KP, respectively, when we always choose the permutation for which we obtain the highest (lowest) test statistic. At the lower end of obtainable rejection rates, it is clearly possible to virtually never reject the null hypothesis for any dimension. On the other hand, the null hypothesis can be rejected much too frequently if one chooses those permutations that yield high test statistics. For $d = 2$ the scope for data mining is rather limited, with obtainable rejection rates being around 10 %. However, once the dimension (and consequently the number of possible permutations) increases, obtainable rejection rates increase quickly. They lie above 50% for $d = 6$ for all transformations considered and reach virtually 100 % for the test based on KP.

In the right part of Figure 1, the solid lines indicate the power that is obtained when the tests are applied properly. The upper (lower) lines show the rejection rates that one obtains when always selecting the highest (lowest) test statistic across all possible permutations. The range of obtainable rejection rates is considerable in all cases. For tests based on CS and KP, the lower line is very close to 0. This means that even though the data are generated from a different DGP, a researcher would be able to purposely select permutations in such a way that H_0 is almost never rejected.

3.3 Size and Power

3.3.1 Known Parameters and no Dynamic Misspecification

Table 2 shows the Monte Carlo results concerning the size and power for the different transformations under the assumption of known parameters and no dynamic misspecification. Focusing on the upper panel of the table, we see that none of the approaches suffers from notable size distortions. In all cases, the obtained actual sizes are very close to the nominal size of 5 %.

The second panel of the table reveals that tests based on our three new transformations and on S have the best power when the alternative implies deviations of the variances (H_1). Tests based on our new transformations outperform the test based on S for large dimensions, as also documented by the performance against H_2, the more challenging alternative in which we change only a third of the variances. Results for H_3 show that the three new approaches consistently outperform the tests based on established transformations when deviations from H_0 are specified in terms of the correlation structure of the multivariate density. In the case of simultaneously misspecified variances *and* correlations (H_4), the new approaches consistently outperform the tests based on CS or KP. For small samples ($P = 50$), they also outperform the test based on S; for larger samples tests based on either S or the new transformations quickly approach a power of 1.

Turning to the power of the different tests for detecting misspecification of the kurtosis (H_5), we see that the new approaches outperform all established tests by a wide margin. Especially for $P = 50$ the results are stunning: the power of the new approaches exceeds that of even the best-performing established approach by a factor of more than two in many cases. Adding wrongly calibrated variances and correlations to the misspecification of the distribution in H_6 leads to a decrease of this outperformance because there is little room for the power of the new approaches to improve while, at the same time, tests based on the old transformations gain a lot of power through these additional deviations from H_0. However, our new approaches still clearly outperform the established approaches. Finally, the new transformations have also better power against GARCH effects (H_7). In light of our previous results, this is expected given that this alternative leads to innovations that are unconditionally distributed with excess kurtosis. The second set of simulations which assumes a multivariate t distribution as the null model yields very similar results with respect to the relative performance of the different transformations; results can be found in the appendix.

3.3.2 Estimated Parameters and Accounting for Dynamic Misspecification

Now, we turn to the case where the model parameters have to be estimated from in-sample data and we account for dynamic misspecification. In general, the results indicate that tests based on all transformations are substantially oversized if one does not adjust Neyman's smooth test (Table 3). Using the adjusted version described in Section 2.6.2, yields correctly sized tests for $P \geq 100$. Furthermore, the results indicate that having to deal with estimated parameters and dynamic misspecification results in a considerable loss of power against H_5. Large evaluation samples seem to be necessary to detect this deviation from the null hypothesis reasonably well (especially for low dimensional densities). Power increases considerably for sample sizes of $P \geq 100$ in the case of H_8 which implies a much stronger deviation from H_0. At the same time, the ranking of the competing tests remains largely unaffected under both alternatives so that the new tests proposed in this paper continue to perform substantially better than established tests. Interestingly, when GARCH effects are present, the test based on S outperforms other tests both when dynamic misspecification is present (H_{10}) and when it is not (H_9).

In further simulations, we show that the new tests have the highest power against CCC-t-GARCH models when the null hypothesis is a Gaussian CCC-GARCH model if no dynamic misspecification is present. If, on the other hand, dynamic misspecification is present, tests based on S have the highest power for small P and comparable power to the new tests for larger evaluation samples; results can be found in the appendix. Furthermore, we show that tests based on the idea of G-ACRs (discussed in Section 2.6.3)

Table 2: Size and power - known parameters and no dynamic misspecification

Size	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.047	0.051	0.047	0.051	0.050	0.052	0.051	0.050	0.052	0.045	0.054	0.055	0.053	0.051	0.050	0.051	0.052	0.053
$d=3$	0.049	0.047	0.048	0.047	0.053	0.052	0.050	0.047	0.050	0.047	0.049	0.047	0.052	0.047	0.047	0.053	0.052	0.055
$d=4$	0.049	0.050	0.048	0.050	0.048	0.051	0.053	0.052	0.050	0.045	0.049	0.047	0.051	0.050	0.051	0.045	0.051	0.049
$d=5$	0.051	0.049	0.051	0.049	0.052	0.054	0.051	0.047	0.048	0.049	0.048	0.047	0.050	0.050	0.049	0.049	0.053	0.052
$d=6$	0.047	0.051	0.049	0.048	0.047	0.048	0.049	0.054	0.049	0.047	0.049	0.048	0.052	0.049	0.049	0.052	0.053	0.051
$d=10$																		
$d=20$																		
$d=50$																		
Power against H_1	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.197	0.137	0.139	0.198	0.199	0.164	0.328	0.210	0.211	0.336	0.336	0.273	0.556	0.338	0.358	0.596	0.583	0.484
$d=3$	0.253	0.150	0.162	0.273	0.272	0.227	0.435	0.223	0.240	0.464	0.472	0.394	0.739	0.385	0.411	0.777	0.766	0.673
$d=4$	0.323	0.164	0.184	0.338	0.335	0.289	0.557	0.247	0.280	0.603	0.597	0.516	0.859	0.435	0.487	0.893	0.882	0.822
$d=5$	0.385	0.175	0.205	0.418	0.407	0.360	0.654	0.279	0.315	0.698	0.693	0.635	0.925	0.472	0.539	0.948	0.940	0.909
$d=6$	0.449	0.193	0.219	0.482	0.475	0.434	0.741	0.301	0.361	0.783	0.763	0.721	0.961	0.527	0.600	0.977	0.971	0.954
$d=20$	0.925	0.384	0.497	0.953	-	0.941	0.999	0.650	0.785	1.000	-	0.999	1.000	0.915	0.978	1.000	-	1.000
Power against H_2	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=3$	0.076	0.056	0.067	0.081	0.081	0.075	0.092	0.061	0.079	0.097	0.099	0.094	0.132	0.068	0.091	0.147	0.153	0.134
$d=6$	0.089	0.057	0.080	0.102	0.101	0.098	0.129	0.063	0.089	0.149	0.149	0.143	0.217	0.076	0.120	0.256	0.255	0.231
$d=10$	0.104	0.056	0.075	0.122	-	0.115	0.154	0.068	0.095	0.189	-	0.184	0.281	0.096	0.132	0.338	-	0.324
$d=20$	0.157	0.062	0.098	0.189	-	0.187	0.276	0.086	0.127	0.330	-	0.323	0.520	0.142	0.204	0.610	-	0.593
$d=50$	0.386	0.097	0.172	0.464	-	0.457	0.700	0.174	0.275	0.783	-	0.774	0.953	0.346	0.465	0.982	-	0.979
Power against H_3	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.066	0.046	0.100	0.067	0.072	0.106	0.077	0.052	0.144	0.081	0.080	0.145	0.100	0.063	0.244	0.106	0.105	0.235
$d=3$	0.090	0.052	0.083	0.098	0.111	0.168	0.136	0.063	0.103	0.146	0.166	0.274	0.219	0.086	0.136	0.247	0.283	0.492
$d=4$	0.135	0.060	0.106	0.149	0.175	0.238	0.217	0.076	0.139	0.241	0.290	0.406	0.377	0.114	0.199	0.429	0.513	0.691
$d=5$	0.174	0.065	0.121	0.195	0.247	0.308	0.306	0.095	0.154	0.350	0.436	0.538	0.546	0.145	0.259	0.612	0.730	0.836
$d=6$	0.225	0.075	0.138	0.252	0.324	0.373	0.409	0.115	0.200	0.462	0.570	0.643	0.706	0.187	0.327	0.762	0.856	0.915
Power against H_4	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.305	0.135	0.256	0.326	0.325	0.378	0.510	0.203	0.423	0.549	0.549	0.627	0.810	0.341	0.696	0.844	0.848	0.904
$d=3$	0.504	0.180	0.292	0.550	0.559	0.619	0.788	0.290	0.473	0.834	0.849	0.894	0.980	0.493	0.757	0.987	0.991	0.996
$d=4$	0.664	0.227	0.350	0.724	0.762	0.791	0.929	0.387	0.587	0.953	0.966	0.976	0.999	0.660	0.881	0.999	1.000	1.000
$d=5$	0.801	0.290	0.436	0.850	0.875	0.890	0.980	0.489	0.690	0.988	0.993	0.995	1.000	0.792	0.940	1.000	1.000	1.000
$d=6$	0.892	0.346	0.508	0.923	0.941	0.946	0.997	0.589	0.787	0.998	0.999	0.999	1.000	0.881	0.976	1.000	1.000	1.000
Power against H_5	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.107	0.077	0.080	0.183	0.188	0.156	0.160	0.105	0.123	0.302	0.302	0.241	0.299	0.172	0.218	0.544	0.545	0.439
$d=3$	0.142	0.085	0.095	0.322	0.314	0.257	0.241	0.122	0.171	0.545	0.538	0.431	0.437	0.207	0.313	0.843	0.837	0.729
$d=4$	0.177	0.091	0.125	0.481	0.472	0.391	0.311	0.143	0.214	0.763	0.750	0.652	0.563	0.247	0.413	0.970	0.970	0.925
$d=5$	0.231	0.109	0.149	0.620	0.622	0.551	0.399	0.165	0.269	0.889	0.883	0.822	0.677	0.291	0.519	0.996	0.995	0.987
$d=6$	0.264	0.114	0.173	0.747	0.736	0.670	0.456	0.186	0.327	0.961	0.955	0.924	0.752	0.344	0.619	1.000	1.000	0.998
Power against H_6	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.186	0.108	0.181	0.299	0.300	0.320	0.293	0.155	0.275	0.486	0.478	0.514	0.504	0.248	0.478	0.756	0.762	0.795
$d=3$	0.298	0.155	0.185	0.507	0.525	0.524	0.493	0.233	0.276	0.762	0.773	0.778	0.772	0.402	0.465	0.958	0.963	0.968
$d=4$	0.406	0.199	0.232	0.666	0.689	0.688	0.656	0.323	0.364	0.904	0.913	0.911	0.905	0.563	0.597	0.994	0.997	0.996
$d=5$	0.521	0.260	0.283	0.793	0.802	0.791	0.781	0.430	0.446	0.965	0.971	0.968	0.965	0.691	0.711	1.000	1.000	1.000
$d=6$	0.588	0.304	0.326	0.859	0.872	0.865	0.847	0.505	0.514	0.988	0.991	0.989	0.987	0.784	0.799	1.000	1.000	1.000
Power against H_7	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.290	0.204	0.252	0.319	0.320	0.285	0.362	0.280	0.310	0.411	0.410	0.356	0.413	0.314	0.376	0.477	0.477	0.423
$d=3$	0.300	0.206	0.162	0.341	0.342	0.289	0.383	0.260	0.212	0.453	0.452	0.383	0.467	0.327	0.280	0.558	0.559	0.487
$d=4$	0.353	0.216	0.169	0.406	0.402	0.350	0.433	0.272	0.242	0.510	0.506	0.438	0.491	0.316	0.285	0.623	0.619	0.545
$d=5$	0.373	0.208	0.171	0.427	0.420	0.373	0.444	0.262	0.239	0.533	0.527	0.467	0.517	0.320	0.291	0.668	0.656	0.586
$d=6$	0.388	0.222	0.194	0.457	0.443	0.402	0.462	0.270	0.246	0.569	0.559	0.510	0.547	0.335	0.304	0.712	0.699	0.641

Notes: Rejection frequencies of Neyman's smooth test based on the transformations introduced in Sections 2.3 and 2.4 for the null hypothesis of multivariate normality with $\sigma_i = 1$ for $i = 1, \ldots, d$ and $\rho_{ij} = 0.5$ for all $i \neq j$. All Monte Carlo simulations are based on 10,000 iterations. The alternative models deviate from the null in terms of wrong variances (H_1), partly wrong variances (H_2), wrong correlations (H_3), wrong variances and wrong correlations (H_4), fat tails (H_5), fat tails, wrong variances, and wrong correlations (H_6), and GARCH effects (H_7). The exact hypotheses are defined in Section 3.1.

Table 3: Size and power - estimated parameters and dynamic misspecification

Size (original test)	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.252	0.236	0.154	0.124	0.124	0.127	0.248	0.229	0.157	0.122	0.123	0.124	0.240	0.221	0.153	0.123	0.124	0.122
$d = 3$	0.262	0.226	0.098	0.133	0.132	0.143	0.253	0.228	0.102	0.134	0.133	0.133	0.259	0.237	0.102	0.130	0.130	0.132
$d = 4$	0.270	0.229	0.101	0.153	0.152	0.174	0.261	0.233	0.089	0.136	0.135	0.146	0.260	0.235	0.086	0.132	0.132	0.136
$d = 5$	0.291	0.235	0.094	0.170	0.165	0.206	0.260	0.225	0.092	0.140	0.140	0.163	0.254	0.224	0.082	0.129	0.130	0.136
$d = 6$	0.295	0.239	0.102	0.182	0.182	0.252	0.270	0.235	0.090	0.157	0.150	0.189	0.261	0.223	0.085	0.138	0.137	0.157
Size (adjusted test)	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.039	0.020	0.038	0.039	0.041	0.037	0.052	0.055	0.064	0.057	0.059	0.059	0.055	0.063	0.066	0.059	0.061	0.057
$d = 3$	0.036	0.013	0.030	0.031	0.035	0.031	0.049	0.053	0.052	0.049	0.052	0.052	0.052	0.060	0.056	0.055	0.056	0.056
$d = 4$	0.026	0.015	0.024	0.025	0.029	0.028	0.042	0.051	0.041	0.047	0.050	0.046	0.048	0.054	0.053	0.056	0.059	0.055
$d = 5$	0.018	0.010	0.015	0.019	0.020	0.018	0.035	0.045	0.035	0.047	0.048	0.046	0.045	0.058	0.047	0.052	0.054	0.050
$d = 6$	0.010	0.006	0.010	0.014	0.018	0.014	0.027	0.041	0.033	0.039	0.041	0.040	0.043	0.058	0.041	0.047	0.049	0.050
Power against H_5	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.050	0.019	0.043	0.022	0.024	0.022	0.084	0.071	0.091	0.056	0.056	0.051	0.121	0.096	0.122	0.129	0.130	0.106
$d = 3$	0.046	0.016	0.036	0.013	0.014	0.015	0.094	0.064	0.073	0.059	0.058	0.049	0.140	0.102	0.123	0.229	0.225	0.161
$d = 4$	0.038	0.013	0.024	0.007	0.007	0.009	0.091	0.058	0.073	0.071	0.070	0.059	0.170	0.092	0.136	0.360	0.351	0.262
$d = 5$	0.026	0.007	0.018	0.003	0.004	0.005	0.091	0.053	0.064	0.093	0.090	0.066	0.209	0.099	0.153	0.514	0.504	0.397
$d = 6$	0.022	0.005	0.009	0.003	0.003	0.003	0.085	0.044	0.055	0.116	0.106	0.090	0.223	0.095	0.166	0.648	0.629	0.538
Power against H_8	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.074	0.016	0.042	0.015	0.016	0.013	0.235	0.128	0.199	0.191	0.185	0.142	0.471	0.266	0.424	0.654	0.649	0.541
$d = 3$	0.085	0.013	0.034	0.010	0.009	0.009	0.305	0.112	0.192	0.289	0.283	0.209	0.649	0.299	0.507	0.891	0.888	0.784
$d = 4$	0.088	0.010	0.023	0.007	0.007	0.006	0.368	0.108	0.204	0.402	0.392	0.294	0.758	0.320	0.613	0.979	0.978	0.942
$d = 5$	0.082	0.008	0.015	0.006	0.006	0.003	0.411	0.090	0.207	0.515	0.501	0.397	0.834	0.334	0.688	0.997	0.996	0.985
$d = 6$	0.077	0.007	0.010	0.004	0.004	0.002	0.454	0.081	0.211	0.600	0.572	0.470	0.883	0.348	0.757	0.999	0.999	0.997
Power against H_9	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.141	0.075	0.092	0.081	0.098	0.075	0.248	0.204	0.260	0.216	0.230	0.197	0.327	0.267	0.336	0.315	0.321	0.280
$d = 3$	0.157	0.070	0.052	0.064	0.075	0.057	0.265	0.181	0.174	0.197	0.211	0.184	0.356	0.250	0.246	0.330	0.339	0.298
$d = 4$	0.169	0.065	0.045	0.049	0.065	0.051	0.281	0.168	0.167	0.201	0.224	0.195	0.392	0.248	0.254	0.373	0.363	0.323
$d = 5$	0.153	0.050	0.034	0.036	0.047	0.041	0.287	0.165	0.174	0.197	0.206	0.189	0.429	0.258	0.267	0.412	0.405	0.351
$d = 6$	0.152	0.045	0.029	0.030	0.042	0.036	0.312	0.154	0.164	0.188	0.196	0.183	0.448	0.241	0.258	0.427	0.412	0.370
Power against H_{10}	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.069	0.019	0.047	0.045	0.049	0.043	0.184	0.143	0.216	0.138	0.149	0.126	0.239	0.200	0.295	0.244	0.248	0.217
$d = 3$	0.075	0.016	0.037	0.025	0.033	0.030	0.173	0.119	0.156	0.117	0.126	0.109	0.261	0.182	0.233	0.252	0.250	0.209
$d = 4$	0.066	0.011	0.025	0.020	0.020	0.018	0.172	0.096	0.134	0.097	0.100	0.098	0.297	0.161	0.243	0.285	0.279	0.221
$d = 5$	0.055	0.008	0.018	0.011	0.014	0.013	0.176	0.082	0.125	0.087	0.093	0.089	0.321	0.163	0.240	0.303	0.285	0.232
$d = 6$	0.044	0.006	0.011	0.005	0.007	0.008	0.184	0.073	0.119	0.077	0.074	0.070	0.347	0.148	0.233	0.319	0.295	0.247

Notes: Rejection frequencies of Neyman's smooth test based on the transformations introduced in Sections 2.3 and 2.4 for the null hypothesis of multivariate normality with $\sigma_i = 1$ for $i = 1, \ldots, d$ and $\rho_{ij} = 0.5$ for all $i \neq j$. All Monte Carlo simulations are based on 10,000 iterations. The null model is a Gaussian VAR(1) model. The alternative models are VAR(1) models with innovations following multivariate t distributions with 8 degrees of freedom (H_5) and 4 degrees of freedom (H_8), a Gaussian GARCH(1,1) without dynamic misspcification of the mean equation (H_9), and one with dynamic misspecification in the mean equation (H_{10}). The exact hypotheses are defined in Section 3.1.

are properly sized and that tests of this kind based on S have the best power against dynamic misspecification while tests based on our new transformations have the best power against distributional misspecifications. Again, results can be found in the appendix.

4 Predicting the Distribution of Stock Market Returns

In this section, we provide an application of the tests discussed above. The application shows that using tests that are not order invariant offers room for data mining in many situations. We consider the problem of forecasting the joint distribution of five international stock market indices. Our data consist of weekly returns of the MSCI indices for the US, Japan (JA), UK, Australia (AU), and Germany (GE) which we obtained from Datastream. The sample spans the period from January 1971 until October 2013 for a total of 2,232 weekly returns. We consider eight different time periods of four years for which we evaluate density forecasts. These (out-of-sample) evaluation periods are 1981-1984, 1985-1988, 1989-1992, 1993-1996, 1997-2000, 2001-2004, 2005-2008, and 2009-2013. In addition, we evaluate the forecast models over the entire sample from 1981-2013. For each period, the previous ten years are considered as in-sample data to estimate the models of interest. The models are re-estimated for each week using a recursive scheme.

Three competing models of increasing complexity are considered: (i) a Gaussian DCC-GARCH model (Engle, 2002), (ii) a time-varying t-copula with t-GARCH margins,[17] and (iii) a time-varying t-copula with skewed-t-GJR-GARCH margins. For the DCC-GARCH model the marginal models for $i = 1, \ldots, d$ are given by

$$
\begin{aligned}
Y_{i,t} &= \mu_i + \varepsilon_{i,t} \\
\varepsilon_{i,t} &= \sqrt{h_{i,t}} z_{i,t} \\
h_{i,t} &= \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1}
\end{aligned}
$$

with $z_{i,t} \sim \mathcal{N}(0,1)$, $\omega_i, \alpha_i, \beta_i \geq 0$ and $\alpha_i + \beta_i < 1$. The correlation matrix R_t of the

[17] The time-varying correlation matrix of the copula is driven by DCC-type dynamics as described in the text, see also Manner and Reznikova (2012). **[Can we simply cite your paper in brackets and delete the footnote?]**

innovations $z_t = [z_{1,t}, \ldots, z_{d,t}]$ is given by

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}, \tag{4.1}$$

where

$$Q_t = (1 - \alpha_c - \beta_c)\bar{Q} + \alpha_c z'_{t-1} z_{t-1} + \beta_c Q_{t-1}, \tag{4.2}$$

with $\alpha_c, \beta_c \geq 0$, $\alpha_c + \beta_c \leq 1$, and $\bar{Q} = E(z'_t z_t)$, which, in practice, is estimated using the sample covariance matrix of z_t.

For the second model, the marginal models are the same as above, with the difference that the innovations $z_{i,t}$ follow a t distribution with ν_i degrees of freedom. The dependence between the t-distributed GARCH innovations z_t is given by a t-copula with degrees of freedom ν_c and correlation matrix R_t. For details and properties of the t-copula see, e. g., Joe (2014). The evolution of the correlation matrix is given by (4.1) and (4.2), but with $z_{i,t}$ replaced by $T_{\nu_c}^{-1}(U_{i,t})\sqrt{\frac{\nu_c - 2}{\nu_c}}$. Note that this model is slightly more flexible than a DCC-GARCH model based on a multivariate t distribution since the copula approach allows all marginal series to have distinct degrees of freedom. The estimation of the copula-based model is naturally done in two steps, ensuring numerical stability at the price of a small loss in statistical efficiency (Joe, 2005).

The third model is even more flexible by assuming that the GARCH innovations $z_{i,t}$ follow the skewed t distribution of Hansen (1994) and by relying on the GJR-GARCH model of Glosten et al. (1993), for which the conditional variance follows

$$h_{i,t} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1} + \gamma_i \varepsilon_{i,t-1}^2 I(\varepsilon_{i,t-1} < 0).$$

The dependence is again given by the DCC-t-copula model.

For each model and each time period, we compute the Rosenblatt PITs and apply the established and new transformations described above. Recall that for non-Gaussian models the distribution of Z_t^{2*} and $Z_t^{2\dagger}$ is not known but can be computed numerically as explained in Section 2.5.1. The null hypothesis of correctly predicted densities is then tested with Neyman's smooth test (Neyman, 1937), accounting for parameter estimation and potential dynamic misspecification as explained in Section 2.6.2. We estimate the long-run covariance matrices using a quadratic spectral kernel and the automatic lag selection as proposed in Andrews (1991). In Table 4, we report the p-values based on the different transformations. For those tests which are not order invariant, we consider all $5! = 120$ permutations of the data. We report the p-value of a random permutation of the variables (based on the arbitrary order in which we downloaded the data: US, JP, UK, AU, GE) and, in brackets, the lowest and highest p-values across all permutations.

The results are mixed and depend on the time period under study. However, a few things clearly stand out. First, the Gaussian DCC model is rejected by almost all tests for all time periods except the 1997-2000 period. Second, model specifications (ii) and (iii) perform much better, but are still rejected for some periods. Notably, most tests reject these models for the aforementioned 1997-2000 period. This suggests that during that period returns had much lighter tails than during other periods. A shorter in-sample period may be appropriate to reflect such non-stationarities. Third, the most flexible specification (iii) does not consistently outperform specification (ii), confirming the known fact that model complexity may yield superior in-sample fit, but not necessarily a better forecasting performance. Fourth, when we use the entire sample, all models are rejected. However, given the very large number of observations and the fact that none of the models features time-varying parameters that might help dealing with structural breaks, this is not too surprising.[18]

Finally, the potential for data mining using the tests based on S, CS, and KP, respectively, is immense. For the wide majority of periods one can find permutations that reject and permutations that do not reject the null hypothesis of properly calibrated density forecasts for any of the models. Note, however, that in line with our results from Section 3.2, the range of p-values for tests based on S is, on average, smaller than for the ones based on CS and KP. Finally, turning to the results for Z^2 which are not

[18] **Not least due to the assumption of frequent structural breaks, applications in finance (e.g. for VaR calculations) are usually based on much shorter evaluation samples.**

Table 4: Density forecast evaluation for stock market returns

Gaussian DCC	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
1981-1984	0.013 [0.004, 0.195]	0.006 [0.002, 0.067]	0.021 [0.008, 0.955]	0.003	0.006	0.015
1985-1988	0.000 [0.000, 0.000]	0.002 [0.000, 0.005]	0.618 [0.001, 0.913]	0.012	0.018	0.010
1989-1992	0.001 [0.000, 0.002]	0.145 [0.002, 0.418]	0.014 [0.000, 0.019]	0.000	0.000	0.000
1993-1996	0.000 [0.000, 0.000]	0.013 [0.000, 0.024]	0.020 [0.000, 0.020]	0.000	0.000	0.000
1997-2000	0.158 [0.009, 0.734]	0.022 [0.002, 0.406]	0.702 [0.039, 0.978]	0.052	0.207	0.237
2001-2004	0.000 [0.000, 0.000]	0.008 [0.006, 0.169]	0.037 [0.000, 0.086]	0.002	0.001	0.004
2005-2008	0.000 [0.000, 0.255]	0.000 [0.000, 0.019]	0.020 [0.000, 0.673]	0.001	0.001	0.004
2009-2013	0.000 [0.000, 0.000]	0.024 [0.000, 0.153]	0.022 [0.000, 0.648]	0.000	0.000	0.001
1981-2013	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.000	0.000	0.000
t-GARCH-tDCC-Cop	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
1981-1984	0.249 [0.043, 0.593]	0.013 [0.006, 0.119]	0.300 [0.066, 0.997]	0.415 [0.244, 0.424]	0.254	0.046
1985-1988	0.000 [0.000, 0.000]	0.004 [0.000, 0.043]	0.543 [0.009, 0.993]	0.221 [0.069, 0.307]	0.180	0.510
1989-1992	0.370 [0.007, 0.403]	0.809 [0.070, 0.901]	0.942 [0.001, 0.942]	0.107 [0.063, 0.220]	0.119	0.096
1993-1996	0.001 [0.000, 0.034]	0.062 [0.002, 0.071]	0.647 [0.000, 0.698]	0.000 [0.000, 0.000]	0.000	0.007
1997-2000	0.054 [0.001, 0.068]	0.090 [0.008, 0.750]	0.591 [0.003, 0.955]	0.000 [0.000, 0.001]	0.001	0.022
2001-2004	0.000 [0.000, 0.000]	0.030 [0.006, 0.412]	0.189 [0.000, 0.443]	0.027 [0.026, 0.034]	0.018	0.001
2005-2008	0.000 [0.000, 0.436]	0.001 [0.000, 0.050]	0.059 [0.000, 0.954]	0.091 [0.073, 0.250]	0.306	0.220
2009-2013	0.000 [0.000, 0.006]	0.204 [0.003, 0.529]	0.016 [0.000, 0.971]	0.033 [0.021, 0.056]	0.027	0.044
1981-2013	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.011 [0.000, 0.702]	0.007 [0.001, 0.051]	0.003	0.000
st-GJR-tDCC-Cop	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
1981-1984	0.297 [0.036, 0.556]	0.017 [0.007, 0.105]	0.165 [0.056, 0.998]	0.582 [0.358, 0.638]	0.457	0.143
1985-1988	0.000 [0.000, 0.001]	0.008 [0.000, 0.085]	0.653 [0.015, 0.993]	0.126 [0.047, 0.147]	0.090	0.102
1989-1992	0.048 [0.000, 0.065]	0.582 [0.060, 0.912]	0.616 [0.000, 0.945]	0.154 [0.096, 0.277]	0.142	0.156
1993-1996	0.000 [0.000, 0.009]	0.026 [0.000, 0.026]	0.373 [0.000, 0.543]	0.000 [0.000, 0.000]	0.000	0.003
1997-2000	0.052 [0.001, 0.244]	0.258 [0.016, 0.929]	0.215 [0.002, 0.978]	0.001 [0.001, 0.001]	0.002	0.036
2001-2004	0.000 [0.000, 0.000]	0.097 [0.011, 0.318]	0.024 [0.000, 0.044]	0.001 [0.001, 0.002]	0.001	0.000
2005-2008	0.000 [0.000, 0.245]	0.010 [0.001, 0.481]	0.003 [0.000, 0.592]	0.072 [0.070, 0.100]	0.075	0.057
2009-2013	0.245 [0.005, 0.659]	0.672 [0.075, 0.794]	0.150 [0.002, 0.962]	0.767 [0.727, 0.788]	0.626	0.262
1981-2013	0.000 [0.000, 0.000]	0.000 [0.000, 0.000]	0.013 [0.000, 0.779]	0.022 [0.005, 0.104]	0.007	0.001

Notes: The table shows p-values corresponding to the different transformations introduced in Sections 2.3 and 2.4 using the adjusted version of Neyman's smooth test (Neyman, 1937) that accounts for parameter estimation and potential dynamic misspecification as explained in Section 2.6.2. The data are weekly MSCI stock index returns for the US, Japan, UK, Australia and Germany. Forecasts are evaluated for the stated periods and the previous 10 years of data are used as the in-sample period. For transformations which are not order invariant, the numbers in brackets show the lowest and highest obtained p-values across all permutations of the variables; for these transformations, the first p-value is for an arbitrarily selected permutation.

order invariant for the non-Gaussian models, one can see that the range of the p-values is very limited and that there is only moderate scope for data mining based on this transformation.

In summary, we recommend evaluating the density forecasts based on Z^{2*} and $Z^{2\dagger}$, and possibly based on Z^2. The results based on the other transformations are not reliable as different permutations can lead to substantially different conclusions regarding the performance of the models. Furthermore, our Monte Carlo simulations show that the new tests are superior in terms of power. Using a 1% significance level, specifications (ii) and (iii) are rejected by the test based on Z^{2*} ($Z^{2\dagger}$) for only 2 (2) and 3 (2) sub-samples, respectively. When using a Bonferroni correction to address the fact that this is a case of multiple testing, specification (ii) is only rejected for the 1993-1996 period (based on both Z^{2*} and $Z^{2\dagger}$) and for the 1997-2000 period (based on Z^{2*}).[19] Thus, overall, the

[19]Since we apply the tests to eight different sub-samples, a test at the 5% significance level should reject when the p-value is smaller than $0.05/8 = 0.0063$.

t-GARCH model with a time-varying t-copula can be recommended for modeling and predicting the joint density of weekly stock market returns.

5 Conclusion

In this paper we derive order-invariant tests for proper calibration of multivariate densities of arbitrary dimension. We demonstrate that distortions in rejection rates can be very large when established tests, which are not order invariant, are used for data mining. Furthermore, we show that the new tests have very good power against a wide range of deviations from the null hypothesis; this holds true, in particular, when the data exhibit fat tails that are not taken into account by the null model. We do not find that one of our new tests dominates the others in terms of power regardless of the alternatives. Our Monte Carlo simulations indicate that tests based on Z^{2^*} have slightly better power against misspecification of the variance and in the presence of fat tails while tests based on $Z^{2^\dagger}$ have slightly better power against misspecification of the correlation structure. Therefore, we recommend using simultaneously the tests based on Z^{2^*} and $Z^{2^\dagger}$ whenever there is no strong prior about the nature of potential deviations from the specified null model.

We want to stress again that our approach, which essentially relies on transforming the multivariate problem to a univariate one, is compatible with any method for testing univariate distributions. We recommend using the powerful Neyman smooth test and show how it can be adjusted to account for parameter uncertainty and dynamic misspecification. If the aim is to test the joint hypothesis of completely calibrated distributions, G-ACR-based tests seem to work well.

We believe there is a wide range of other applications in various fields. First, the proposed methods are useful whenever properly calibrated density forecasts are crucial to form well-informed decisions (about production, investment, pricing, etc.) and will foster the use of multivariate density forecasts in situations in which decisions are based on loss functions that take more than one variable as input arguments. Our tests could, for instance, be used to assess the overall forecast performance of macroeconomic DSGE models used at central banks. Second, the proposed methods are useful to improve the specification of multivariate models taking higher moments into account; obvious applications of this kind are common in financial econometrics, e. g., for estimating the Value-at-Risk of a portfolio, but it can be expected that the modeling of the dependence structure of higher moments of multivariate data becomes more common also for demand management or in macroeconomics.

Our study leaves room for future research along several dimensions. First, especially for financial applications, it would be interesting to extend the analytical results of our paper which are limited to the case of multivariate Gaussian processes under the null hypothesis to more general settings. Second, we believe it may be possible to develop tests with even better power for very high-dimensional densities; this could be achieved by selecting the terms entering the $Z^{*\dagger}$ transformation in a data-driven way or by assigning weights to the conditional PITs entering the transformations that are based on the relevant loss function. Finally, it might be worthwhile to investigate whether powerful order invariant tests can be constructed that are not based on the Rosenblatt transformation.

References

Aastveit, K. A., Gerdrup, K. R., Jore, A. S., and Thorsrud, L. A. (2014). Nowcasting GDP in real time: a density combination approach. *Journal of Business & Economic Statistics*, 32(1):48–68.

Amisano, G. and Geweke, J. (2017). Prediction using several macroeconomic models. *Review of Economics and Statistics*, 99(5):912–925.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Andrews, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica*, 65(5):1097–1128.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics*, 85(3):531–549.

Bai, J. and Chen, Z. (2008). Testing multivariate distributions in GARCH models. *Journal of Econometrics*, 143(1):19–36.

Bera, A. K. and Ghosh, A. (2002). Neyman's smooth test and its applications in econometrics. In Ullah, A., Wan, A. T. K., and Chaturvedi, A., editors, *Handbook of Applied Econometrics and Statistical Inference*, pages 177–230. Marcel Dekker, New York.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–74.

Chen, Y.-T. (2011). Moment tests for density forecast evaluation in the presence of parameter estimation uncertainty. *Journal of Forecasting*, 30:409–450.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3):327–341.

Clements, M. P. and Hendry, D. F. (1993). On the limitations of comparing mean square forecast errors. *Journal of Forecasting*, 12:617–637.

Clements, M. P. and Smith, J. (2000). Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment. *Journal of Forecasting*, 19:144–165.

Clements, M. P. and Smith, J. (2002). Evaluating multivariate forecast densities: a comparison of two approaches. *International Journal of Forecasting*, 18(3):397–407.

Corradi, V. and Swanson, N. R. (2006a). Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics*, 133(2):779–806.

Corradi, V. and Swanson, N. R. (2006b). *Predictive Density Evaluation*, volume 1 of *Handbook of Economic Forecasting*, chapter 5, pages 197–284. Elsevier.

Dawid, A. P. (1984). Statistical theory: the prequential approach. *J. Roy. Statist. Soc. Ser. A*, 147(2):278–292.

De Gooijer, J. G. (2007). Power of the Neyman smooth test for evaluating multivariate forecast densities. *Journal of Applied Statistics*, 34(4):371–381.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–83.

Diebold, F. X., Hahn, J., and Tay, A. S. (1999). Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange. *The Review of Economics and Statistics*, 81(4):661–673.

Duan, J.-C. (2004). A specification test for time series models by a normality. Econometric Society 2004 North American Winter Meetings 467, Econometric Society.

Engle, R. F. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20:339–350.

Genest, C. and Rivest, L.-P. (2001). On the multivariate probability integral transformation. *Statistics & Probability Letters*, 53(4):391–399.

Ghosh, A. and Bera, A. K. (2015). Density forecast evaluation for dependent financial data: Theory and applications. Working Paper 1-2015, Research Collection Lee Kong Chian School Of Business.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Applications*, 1(1):125–151.

González-Rivera, G. and Sun, Y. (2015). Generalized autocontours: Evaluation of multivariate density models. *International Journal of Forecasting*, 31(3):799–814.

González-Rivera, G. and Yoldas, E. (2012). Autocontour-based evaluation of multivariate predictive densities. *International Journal of Forecasting*, 28(2):328–342.

Gonzlez-Rivera, G., Senyuz, Z., and Yoldas, E. (2011). Autocontours: Dynamic specification testing. *Journal of Business & Economic Statistics*, 29(1):186–200.

Hallam, M. and Olmo, J. (2014). Semiparametric density forecasts of daily financial returns from intraday data. *Journal of Financial Econometrics*, 12(2):408–432.

Hansen, E. B. (1994). Autoregressive density estimation. *International Economic Review*, 35:705–730.

Herbst, E. and Schorfheide, F. (2012). Evaluating DSGE model forecasts of comovements. *Journal of Econometrics*, 171(2):152–166.

Hong, Y. and Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies*, 18(1):37–84.

Hong, Y., Li, H., and Zhao, F. (2007). Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates. *Journal of Econometrics*, 141(2):736–776.

Huurman, C., Ravazzolo, F., and Zhou, C. (2012). The power of weather. *Computational Statistics & Data Analysis*, 56(11):3793–3807.

Ishida, I. (2005). Scanning multivariate conditional densities with probability integral transforms. CARF F-Series CARF-F-045, Center for Advanced Research in Finance, Faculty of Economics, The University of Tokyo.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401–419.

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Kitsul, Y. and Wright, J. H. (2013). The economics of options-implied inflation probability density functions. *Journal of Financial Economics*, 110(3):696–711.

Knüppel, M. (2015). Evaluating the calibration of multi-step-ahead density forecasts using raw moments. *Journal of Business & Economic Statistics*, 33(2):270–281.

Ko, S. I. M. and Park, S. Y. (2013a). Multivariate density forecast evaluation: a modified approach. *International Journal of Forecasting*, 29(3):431–441.

Ko, S. I. M. and Park, S. Y. (2013b). Multivariate density forecast evaluation: Smooth test approach. Working paper, Chinese University of Hong Kong.

Ledwina, T. (1994). Data driven version of the Neyman smooth test of fit. *Journal of the American Statistical Association*, 89(427):1000–1005.

Lin, J. and Wu, X. (2017). A sequential test for the specification of predictive densities. *Econometrics Journal*, 20(2):190–220.

Manner, H. and Reznikova, O. (2012). A survey on time-varying copulas: Specification, simulations and application. *Econometric Reviews*, 31(6):654–687.

Mitchell, J. and Wallis, K. F. (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, 26(6):1023–1040.

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, 53(5):1047–1070.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–08.

Neyman, J. (1937). Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift*, 20:150–199.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3):470–472.

Rossi, B. and Sekhposyan, T. (2013). Conditional predictive density evaluation in the presence of instabilities. *Journal of Econometrics*, 177(2):199–212.

Rossi, B. and Sekhposyan, T. (2016). Alternative tests for correct specification of conditional predictive densities. Economics working papers, Department of Economics and Business, Universitat Pompeu Fabra.

Shackleton, M. B., Taylor, S. J., and Yu, P. (2010). A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices. *Journal of Banking & Finance*, 34(11):2678–2693.

Smith, J. Q. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting*, 4:283–291.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics*, 30(1-2):415–443.

Tay, A. (2015). A brief survey of density forecasting in macroeconomics. Macroeconomic Review October 2015, Monetary Authority of Singapore.

Taylor, J. W. (2012). Density forecasting of intraday call center arrivals using models based on exponential smoothing. *Management Science*, 58(3):534–549.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

West, K. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64:1067–1087.

West, K. and McCracken, M. (1998). Regression based tests of predictive ability. *International Economic Review*, 39:817–840.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

Wolters, M. H. (2015). Evaluating point and density forecasts of DSGE models. *Journal of Applied Econometrics*, 30(1):74–96.

Ziegel, J. F. and Gneiting, T. (2014). Copula calibration. *Electronic Journal of Statistics*, 8(2):2619–38.

APPENDIX

Appendix A Proofs and Derivations

Proof of Proposition 1. Under independence, we have $U_t^{i|1:i-1} = U_t^i$, i. e., the conditional CDF is equal to the marginal CDF. In this case, the product transformation reduces to $P_{t,d} = \prod_{i=1}^{d} U_t^i$. This is clearly robust to permutations. The same argument can be made for the location-adjusted version $P_{t,d}^*$. The stacked transformation then becomes $S_t = [U_t^1, \ldots, U_t^d]'$, which again is obviously order invariant.

Now consider the following two permutations: $\pi_1 = (1, 2, 3, \ldots, d)$ and $\pi_2 = (2, 1, 3, \ldots, d)$. For these permutations, the product transformations only differ in their first two components. So w.l.g., we only check that independence is needed for $U_t^1 \cdot U_t^{2|1} = U_t^2 \cdot U_t^{1|2}$ to hold. The latter equality is equivalent to $\frac{U_t^1}{U_t^2} = \frac{U_t^{1|2}}{U_t^{2|1}}$ for all t, which does not hold in general, unless we have independence.

For these two permutations order invariance in S_t is given only if $[U_t^1 \; U_t^{2|1}]'$ is equal to $[U_t^2 \; U_t^{1|2}]'$ for all t, which again only holds under independence. □

Proof of Proposition 2. W.l.g. let $\mu = 0$, which can be achieved by demeaning the original data. Rewrite Y_t as

$$
\begin{aligned}
Y_{1,t} &= Z_{1,t} \\
Y_{2,t} &= \beta_{2,1} Y_{1,t} + Z_{2t} \\
&\vdots \\
Y_{d,t} &= \beta_{d,1} Y_{1,t} + \beta_{d,2} Y_{2,t} + \ldots + \beta_{d,d-1} Y_{d-1,t} + Z_{d,t},
\end{aligned}
$$

with $Z_{i,t}$ normally distributed. Writing this more compactly we obtain

$$BY_t = Z_t,$$

where $Z_t = (Z_{1,t}, \ldots, Z_{d,t})'$, with

$$\mathbb{E}(Z_t Z_t') = D = \text{diag}\begin{pmatrix} \sigma_1^2 \\ \sigma_{2|1}^2 \\ \vdots \\ \sigma_{d|1:d-1}^2 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 1 & 0 & 0 & \ldots \\ -\beta_{2,1} & 1 & 0 & \ldots \\ -\beta_{3,1} & -\beta_{3,2} & 1 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is the matrix of population regression coefficients, whose precise form in terms of the covariance matrix is directly available using standard results on conditional normal random variables. It holds that $U_t^1 = \Phi(Z_{1,t}/\sigma_1), U_t^{2|1} = \Phi(Z_{2,t}/\sigma_{2|1}), \ldots, U_t^{d|1:d-1} = \Phi(Z_{d,t}/\sigma_{d|1:d-1})$. Furthermore, note that

$$\text{Cov}(Y_t) = \mathbb{E}(Y_t Y_t') = \Sigma = B^{-1} D B^{-1'}.$$

Consequently,

$$Z_t^2 = (Z_t' D^{-1/2})(D^{-1/2} Z_t) = Y_t' B' D^{-1} B Y_t = Y_t' \Sigma^{-1} Y_t.$$

The last term is clearly invariant to the order of the variables. □

***Derivation of the distribution of* $\mathbf{CS_{t,d}}$ *for arbitrary* $\mathbf{d}$.** The proof is done by induction. To simplify notations we drop the time subscript and replace the conditional PITs by a sequence of independent $\mathcal{U}(0,1)$ random variables $U_1, U_2, \ldots$.

Step 1 ($d = 2$): For $d = 2$ the density is given by

$$f_{CS_2}(CS_2) = \frac{(-1)^1}{1!} \log(CS_2) = -\log(CS_2),$$

which is equal to the density derived in Clements and Smith (2000). Note that we could also start at $d = 1$, for which the density is equal to 1, corresponding to the uniform distribution.

Step 2 ($d \to d+1$): Consider the change of variables

$$CS_{d+1} = CS_d U_{d+1}$$

The determinant of the Jacobian for the inverse transformation is

$$J = \det \frac{\partial(CS_d, U_{d+1})}{\partial(CS_{d+1}, U_{d+1})} = \begin{vmatrix} \frac{1}{U_{d+1}} & -\frac{CS_{d+1}}{U_{d+1}^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{U_{d+1}}.$$

The joint density of CS_{d+1} and U_{d+1} is

$$f_{CS_{d+1},U_{d+1}}(CS_{d+1}, U_{d+1}) = f_{CS_d}\left(\frac{CS_{d+1}}{U_{d+1}}\right) \cdot \frac{1}{U_{d+1}} = \frac{(-1)^{d-1}}{(d-1)!} \log^{d-1}\left(\frac{CS_{d+1}}{U_{d+1}}\right) \cdot \frac{1}{U_{d+1}},$$

with $0 < CS_{d+1} < U_{d+1} < 1$. Therefore, the marginal PDF of CS_{d+1} is

$$f_{CS_{d+1}}(CS_{d+1}) = \int_{CS_{d+1}}^{1} f_{CS_d}\left(\frac{CS_{d+1}}{U_{d+1}}\right) \cdot \frac{1}{U_{d+1}} d \cdot U_{d+1} = \frac{(-1)^d}{d!} \log^d(CS_{d+1}) = f_{d+1}(CS_{d+1}).$$

To show that the CDF is correct first note that

$$f'_{CS_d}(CS_d) = \frac{(-1)^{d-1}}{(d-2)!} \log^{d-2}(CS_d) \cdot \frac{1}{CS_d} = -1 \cdot f_{CS_{d-1}}(CS_d) \cdot \frac{1}{CS_d}.$$

It follows that

$$F'(CS_d) = \sum_{i=0}^{d-1} f_{CS_{d-i}}(CS_d) - \sum_{i=1}^{d-1} f_{CS_{d-i}}(CS_d) = f_{CS_d}(CS_d).$$

□

Derivation of the distribution of $\mathbf{KP_{t,d}}$ ***for arbitrary*** $\mathbf{d}$. Again, the proof is done by induction and again for simplicity we consider a sequence of independent $\mathcal{U}(0,1)$ random variables $U_1, U_2, \ldots$.

Step 1 ($d = 2$): Consider the change of variables

$$KP_2 = (U_1 - 0.5)(U_2 - 0.5) = U_1^* U_2^*.$$

The determinant of the Jacobian is

$$J = \frac{1}{U_2^*},$$

so the joint density of KP_2 and U_2^* is given by

$$f_{KP_2, U_2^*} = \left| \frac{1}{U_2^*} \right|.$$

Integrating out U_2^* gives

$$f_{KP_2}(KP_2) = \int_{-1/2}^{1/2} \left| \frac{1}{U_2^*} \right| 2 \cdot U_2^* = 2 \cdot \int_{|2KP_2|}^{1/2} \frac{1}{U_2^*} = 2 \log(U_2^*) \Big|_{|2KP_2|}^{1/2} = 2 \log \left| \frac{1}{4KP_2} \right|,$$

where the second equality follows from the symmetry around 0 and the fact that $|2KP_2| < |U_2^*| < 1/2$.

Step 2 ($d \to d+1$): Consider the following change of variables

$$KP_{d+1} = KP_d(U_{d+1} - 0.5) = KP_d U_{d+1}^*.$$

The determinant of the Jacobian is

$$J = \frac{1}{U_{d+1}^*},$$

and therefore the joint density of KP_{d+1} and U_{d+1}^* is

$$f_{KP_{d+1}, U_{d+1}^*} = f_{KP_d}\left(\frac{KP_{d+1}}{U_{d+1}^*}\right) \left| \frac{1}{U_{d+1}^*} \right|.$$

The PDF of KP_{d+1} then is

$$
\begin{aligned}
f_{KP_{d+1}}(KP_{d+1}) &= \int_{-1/2}^{1/2} f_{KP_d}\left(\frac{KP_{d+1}}{U^*_{d+1}}\right)\left|\frac{1}{U^*_{d+1}}\right| d\cdot U^*_{d+1} \\
&= 2\cdot\int_{|2^dKP_{d+1}|}^{1/2} \frac{2^{d-1}}{(d-1)!}\log^{d-1}\left(\frac{U^*_{d+1}}{2^d\,|KP_{d+1}|}\right)\frac{1}{U_{d+1}} d\cdot U_{d+1} \\
&= 2\cdot\frac{2^{d-1}}{(d-1)!}\frac{1}{d}\log^{d}\left(\frac{U^*_{d+1}}{2^d\,|KP_{d+1}|}\right)\Bigg|_{|2^dKP_2|}^{1/2} = \frac{2^d}{(d)!}\log^d\left|\frac{1}{2^{d+1}KP_{d+1}}\right|.
\end{aligned}
$$

Again the symmetry around 0 and the fact that $|2^dKP_{d+1}| < |U^*_{d+1}| < 1/2$ was used.

Now consider the CDF. Note that

$$
\frac{f'_{KPd}(KP_d)}{2^{d-1}} = \frac{1}{(d-2)!}\log\left|\frac{1}{2^dKP_d}\right|(-1)\frac{1}{P_d}.
$$

Then using the product rule

$$
\begin{aligned}
F'_{KP_d}(KP_d) &= 2^{d-1}\sum_{i=1}^{d}\frac{1}{(d-i)!}\log^{d-i}\left|\frac{1}{2^dKP_d}\right|KP_d - KP_d\sum_{i=2}^{d}\frac{1}{(d-i)!}\log^{d-i}\left|\frac{1}{2^dKP_d}\right|\frac{1}{KPd} \\
&= \frac{2^{d-1}}{(d-1)!}\log^{d-1}\left|\frac{1}{2^dKP_d}\right| = f_{KP_d}(KP_d).
\end{aligned}
$$

The addition of 1/2 (see Proposition **??**) ensures that the CDF lies between 0 and 1. □

Proof of Proposition 3. Consider the generic term $\Phi^{-1}\left(U_t^{i|\gamma_i^k}\right) \sim \mathcal{N}(0,1)$, where γ_i^k stands for a set of indices representing the conditioning variables. Under normality, these terms are also jointly normally distributed. Then the fact that Z_t^{*2} has a mixture of independent χ^2_1 random variables follows directly from Lemma 17.1 in van der Vaart (1998). The weights of the mixture are given by the eigenvalues of the covariance matrix of the terms $\Phi^{-1}\left(U_t^{i|\gamma_i^k}\right)$ for all $i = 1,\dots,d$ and $k = 1,\dots,2^{d-1}$. This matrix is actually a correlation matrix due to the unit variance of the inverse normal transformation.

To compute this correlation matrix, we start with the covariance between $Y^t_{i|\gamma_i^k}$ and

$Y^t_{j|\gamma^l_j}$. Then, dropping the time index, Y_i conditional on the vector $Y_{\gamma^k_i}$ is

$$Y_i|Y_{\gamma^k_i} = Y_i - \Sigma_{i,\gamma^k_i}\Sigma^{-1}_{\gamma^k_i,\gamma^k_i}Y_{\gamma^k_i},$$

which has variance equal to $\Sigma_{ii} - \Sigma_{i,\gamma^k_i}R^{-1}_{\gamma^k_i,\gamma^k_i}\Sigma_{\gamma^k_i,i}$. Consequently,

$$\Phi^{-1}\left(U_t^{i|\gamma^k_i}\right) = \frac{Y_i - \Sigma_{i,\gamma^k_i}\Sigma^{-1}_{\gamma^k_i,\gamma^k_i}Y_{\gamma^k_i}}{(\Sigma_{ii} - \Sigma_{i,\gamma^k_i}\Sigma^{-1}_{\gamma^k_i,\gamma^k_i}\Sigma_{\gamma^k_i,i})^{1/2}}$$

and analogously for $\Phi^{-1}\left(U_t^{j|\gamma^l_j}\right)$. Then the computation of the covariance/correlation is straightforward. The reduced rank of R_{Z^*} follows from the fact that all conditional variables $Y^t_{i|\gamma^k_i}$ are a linear combination of the original d variables. □

Appendix B Further Simulation Results

Table B.1: Size and power when the null is a multivariate t distribution with known parameters

Size	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.047	0.044	0.050	0.049	0.052	0.054	0.047	0.050	0.047	0.049	0.046	0.050	0.051	0.052	0.051	0.047	0.050	0.048
$d=3$	0.050	0.052	0.053	0.053	0.053	0.059	0.053	0.049	0.051	0.052	0.058	0.052	0.054	0.048	0.050	0.051	0.051	0.047
$d=4$	0.055	0.048	0.049	0.053	0.051	0.049	0.053	0.053	0.046	0.047	0.051	0.054	0.054	0.051	0.052	0.050	0.049	0.050
$d=5$	0.052	0.049	0.048	0.052	0.052	0.053	0.050	0.050	0.051	0.053	0.052	0.052	0.045	0.049	0.047	0.049	0.050	0.051
$d=6$	0.051	0.045	0.050	0.048	0.049	0.047	0.048	0.050	0.052	0.051	0.055	0.051	0.055	0.054	0.051	0.054	0.062	0.060
$d=10$																		
$d=20$	0.054	0.051	0.048	0.050	-	0.048	0.048	0.048	0.043	0.049	-	0.048	0.048	0.051	0.047	0.048	-	0.053
$d=50$	0.048	0.048	0.051	0.052	-	0.054	0.052	0.050	0.052	0.049	-	0.044	0.050	0.049	0.052	0.050	-	0.062

Power against H_1	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.130	0.100	0.102	0.121	0.100	0.093	0.196	0.134	0.149	0.185	0.182	0.159	0.344	0.212	0.227	0.345	0.304	0.241
$d=3$	0.151	0.102	0.111	0.135	0.113	0.099	0.234	0.133	0.150	0.231	0.262	0.195	0.442	0.208	0.240	0.447	0.395	0.302
$d=4$	0.165	0.101	0.115	0.155	0.160	0.126	0.270	0.127	0.152	0.264	0.326	0.248	0.495	0.205	0.252	0.511	0.487	0.383
$d=5$	0.171	0.096	0.114	0.158	0.170	0.137	0.290	0.127	0.161	0.289	0.277	0.227	0.542	0.199	0.258	0.559	0.566	0.458
$d=6$	0.177	0.098	0.111	0.169	0.167	0.138	0.312	0.130	0.162	0.319	0.279	0.223	0.581	0.193	0.268	0.609	0.552	0.437
$d=20$	0.195	0.086	0.117	0.192	-	0.196	0.327	0.111	0.159	0.393	-	0.395	0.682	0.179	0.280	-	0.711	

Power against H_2	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=3$	0.146	0.099	0.107	0.127	0.141	0.115	0.232	0.128	0.141	0.225	0.213	0.151	0.436	0.204	0.237	0.444	0.482	0.387
$d=6$	0.180	0.093	0.112	0.168	0.158	0.141	0.310	0.126	0.158	0.320	0.321	0.245	0.577	0.192	0.270	0.603	0.553	0.436
$d=10$	0.185	0.091	0.113	0.180	0.173	0.161	0.350	0.122	0.167	0.362	0.357	0.307	0.641	0.190	0.275	0.684	0.723	0.673
$d=20$	0.197	0.088	0.112	0.193	-	0.195	0.362	0.112	0.170	0.393	-	0.389	0.680	0.186	0.287	0.724	-	0.668
$d=50$	0.166	0.069	0.111	0.167	-	0.226	0.327	0.098	0.149	0.338	-	0.437	0.613	0.154	0.238	0.669	-	0.835

Power against H_3	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.060	0.046	0.093	0.062	0.058	0.087	0.073	0.052	0.127	0.065	0.070	0.136	0.081	0.058	0.203	0.083	0.078	0.146
$d=3$	0.079	0.051	0.078	0.074	0.078	0.122	0.096	0.058	0.088	0.098	0.094	0.187	0.145	0.083	0.111	0.147	0.153	0.319
$d=4$	0.091	0.054	0.085	0.091	0.099	0.149	0.136	0.064	0.106	0.140	0.159	0.252	0.223	0.094	0.139	0.231	0.309	0.468
$d=5$	0.108	0.053	0.091	0.104	0.139	0.189	0.169	0.075	0.120	0.174	0.240	0.332	0.304	0.115	0.171	0.320	0.454	0.616
$d=6$	0.122	0.059	0.101	0.120	0.205	0.249	0.204	0.079	0.128	0.216	0.329	0.388	0.369	0.123	0.186	0.392	0.474	0.624

Power against H_4	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.195	0.091	0.192	0.185	0.195	0.253	0.326	0.118	0.314	0.318	0.392	0.481	0.578	0.181	0.547	0.587	0.631	0.708
$d=3$	0.284	0.102	0.192	0.278	0.301	0.360	0.515	0.136	0.300	0.519	0.522	0.608	0.826	0.233	0.539	0.833	0.900	0.946
$d=4$	0.372	0.111	0.207	0.378	0.384	0.437	0.633	0.158	0.348	0.659	0.744	0.800	0.929	0.280	0.599	0.938	0.948	0.969
$d=5$	0.428	0.116	0.231	0.435	0.534	0.571	0.747	0.181	0.398	0.761	0.827	0.844	0.969	0.338	0.671	0.977	0.991	0.993
$d=6$	0.482	0.123	0.245	0.494	0.560	0.569	0.800	0.197	0.417	0.821	0.899	0.892	0.984	0.369	0.707	0.990	0.996	0.997

Power against H_5	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.087	0.080	0.087	0.116	0.121	0.095	0.136	0.104	0.120	0.242	0.275	0.196	0.256	0.156	0.203	0.495	0.534	0.322
$d=3$	0.104	0.074	0.095	0.197	0.176	0.104	0.175	0.103	0.146	0.442	0.447	0.250	0.349	0.172	0.259	0.808	0.795	0.544
$d=4$	0.119	0.079	0.106	0.300	0.242	0.142	0.215	0.108	0.177	0.652	0.646	0.435	0.427	0.177	0.311	0.953	0.966	0.822
$d=5$	0.128	0.080	0.118	0.408	0.379	0.257	0.251	0.119	0.194	0.803	0.745	0.499	0.508	0.186	0.361	0.993	0.994	0.929
$d=6$	0.138	0.079	0.123	0.519	0.461	0.304	0.280	0.115	0.221	0.907	0.871	0.704	0.578	0.200	0.417	0.999	0.996	0.957

Power against H_6	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.456	0.188	0.371	0.454	0.496	0.528	0.767	0.310	0.635	0.786	0.827	0.841	0.975	0.567	0.907	0.983	0.987	0.987
$d=3$	0.691	0.209	0.417	0.721	0.734	0.745	0.951	0.362	0.699	0.969	0.969	0.965	1.000	0.646	0.950	1.000	1.000	1.000
$d=4$	0.830	0.241	0.493	0.878	0.858	0.850	0.994	0.437	0.796	0.998	0.998	0.999	1.000	0.744	0.982	1.000	1.000	1.000
$d=5$	0.917	0.275	0.553	0.960	0.969	0.961	0.999	0.495	0.853	1.000	1.000	1.000	1.000	0.823	0.992	1.000	1.000	1.000
$d=6$	0.961	0.303	0.600	0.987	0.990	0.983	1.000	0.558	0.897	1.000	1.000	1.000	1.000	0.878	0.997	1.000	1.000	1.000

Power against H_7	$P=50$						$P=100$						$P=200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d=2$	0.479	0.306	0.376	0.504	0.501	0.445	0.644	0.487	0.567	0.676	0.674	0.614	0.796	0.660	0.749	0.826	0.842	0.806
$d=3$	0.554	0.300	0.349	0.586	0.564	0.487	0.701	0.468	0.520	0.737	0.728	0.677	0.848	0.650	0.720	0.882	0.867	0.823
$d=4$	0.589	0.296	0.366	0.624	0.633	0.567	0.743	0.462	0.560	0.787	0.780	0.722	0.895	0.643	0.765	0.926	0.923	0.891
$d=5$	0.632	0.299	0.392	0.671	0.649	0.596	0.780	0.462	0.583	0.813	0.803	0.754	0.912	0.649	0.792	0.942	0.944	0.924
$d=6$	0.664	0.306	0.406	0.698	0.686	0.646	0.806	0.458	0.615	0.840	0.849	0.806	0.934	0.651	0.822	0.955	0.942	0.922

Notes: Rejection frequencies of Neyman's smooth test based on the transformations introduced in Sections 2.3 and 2.4 for the null hypothesis of a multivariate t distribution with $\sigma_i = 1$ for $i = 1, \ldots, d$ and $\rho_{ij} = 0.5$ for all $i \neq j$. All Monte Carlo simulations are based on 10,000 iterations. The alternative models deviate from the null in terms of wrong variances (H_1), partly wrong variances (H_2), wrong correlations (H_3), wrong variances and wrong correlations (H_4), normal distribution (H_5), normal distribution, wrong variances, and wrong correlations (H_6), and GARCH effects (H_7). The exact hypotheses are defined in Section 3.1.

Table B.2: Size and power – CCC-GARCH vs. CCC-t-GARCH

	$P=50$						$P=200$						$P=500$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
Only estimation uncertainty																		
Size																		
$d=2$	0.097	0.068	0.076	0.090	0.084	0.081	0.079	0.055	0.048	0.061	0.068	0.055	0.056	0.047	0.051	0.043	0.041	0.031
$d=3$	0.132	0.081	0.070	0.107	0.098	0.092	0.066	0.054	0.044	0.068	0.065	0.058	0.065	0.051	0.041	0.051	0.061	0.043
$d=4$	0.142	0.064	0.057	0.122	0.102	0.099	0.084	0.048	0.059	0.073	0.068	0.060	0.066	0.046	0.045	0.049	0.053	0.048
$d=5$	0.146	0.078	0.057	0.121	0.104	0.094	0.089	0.051	0.048	0.069	0.068	0.075	0.072	0.042	0.042	0.070	0.056	0.048
$d=6$	0.180	0.069	0.043	0.127	0.094	0.118	0.081	0.054	0.036	0.065	0.062	0.066	0.062	0.050	0.052	0.065	0.047	0.063
Power against H_{11}																		
$d=2$	0.204	0.097	0.141	0.229	0.216	0.194	0.300	0.155	0.244	0.532	0.521	0.408	0.657	0.376	0.517	0.905	0.886	0.816
$d=3$	0.239	0.096	0.102	0.332	0.315	0.253	0.459	0.223	0.310	0.847	0.819	0.693	0.815	0.449	0.683	1.000	0.997	0.973
$d=4$	0.284	0.104	0.110	0.452	0.404	0.331	0.549	0.239	0.342	0.965	0.934	0.869	0.926	0.546	0.824	1.000	1.000	0.999
$d=5$	0.276	0.100	0.110	0.523	0.473	0.379	0.656	0.286	0.507	0.993	0.986	0.969	0.956	0.642	0.898	1.000	1.000	1.000
$d=6$	0.346	0.105	0.112	0.626	0.577	0.487	0.702	0.296	0.558	1.000	0.997	0.984	0.983	0.735	0.942	1.000	1.000	1.000
Power against H_{12}																		
$d=2$	0.434	0.227	0.321	0.587	0.567	0.493	0.937	0.681	0.864	0.997	0.993	0.983	1.000	0.842	0.999	1.000	1.000	1.000
$d=3$	0.574	0.242	0.351	0.790	0.755	0.686	0.992	0.733	0.957	1.000	1.000	1.000	1.000	0.774	1.000	1.000	1.000	1.000
$d=4$	0.655	0.248	0.347	0.891	0.874	0.801	0.998	0.769	0.985	1.000	1.000	1.000	1.000	0.767	1.000	1.000	1.000	1.000
$d=5$	0.730	0.245	0.440	0.924	0.909	0.859	1.000	0.772	0.992	1.000	1.000	1.000	1.000	0.700	1.000	1.000	1.000	1.000
$d=6$	0.758	0.243	0.447	0.930	0.922	0.897	1.000	0.774	0.994	1.000	1.000	1.000	1.000	0.664	1.000	1.000	1.000	1.000
Estimation uncertainty and dynamic misspecification																		
Size																		
$d=2$	0.042	0.011	0.050	0.053	0.063	0.046	0.118	0.071	0.111	0.121	0.117	0.090	0.131	0.096	0.103	0.109	0.102	0.079
$d=3$	0.070	0.012	0.031	0.046	0.056	0.032	0.110	0.060	0.060	0.116	0.103	0.091	0.114	0.073	0.079	0.101	0.096	0.082
$d=4$	0.086	0.018	0.031	0.041	0.037	0.024	0.114	0.061	0.060	0.109	0.095	0.081	0.113	0.079	0.060	0.088	0.080	0.078
$d=5$	0.095	0.013	0.017	0.023	0.024	0.016	0.105	0.058	0.053	0.112	0.078	0.066	0.130	0.074	0.075	0.119	0.084	0.078
$d=6$	0.097	0.008	0.009	0.022	0.021	0.013	0.151	0.081	0.068	0.105	0.078	0.070	0.127	0.081	0.055	0.131	0.097	0.097
Power against H_{11}																		
$d=2$	0.069	0.014	0.044	0.025	0.027	0.034	0.171	0.103	0.136	0.228	0.198	0.151	0.316	0.157	0.234	0.532	0.400	0.380
$d=3$	0.108	0.011	0.034	0.020	0.023	0.020	0.245	0.127	0.145	0.383	0.273	0.235	0.396	0.181	0.287	0.873	0.733	0.650
$d=4$	0.113	0.017	0.028	0.020	0.019	0.015	0.276	0.111	0.165	0.588	0.451	0.396	0.545	0.211	0.385	0.974	0.918	0.873
$d=5$	0.157	0.009	0.032	0.009	0.014	0.006	0.373	0.124	0.193	0.752	0.597	0.512	0.646	0.229	0.501	0.995	0.977	0.962
$d=6$	0.160	0.010	0.015	0.014	0.010	0.007	0.402	0.095	0.218	0.859	0.721	0.642	0.709	0.280	0.583	1.000	0.999	0.997
Power against H_{12}																		
$d=2$	0.120	0.026	0.051	0.041	0.041	0.033	0.516	0.293	0.457	0.792	0.688	0.615	0.896	0.584	0.838	0.999	0.988	0.983
$d=3$	0.207	0.023	0.055	0.041	0.023	0.019	0.686	0.334	0.522	0.961	0.901	0.834	0.975	0.631	0.946	1.000	1.000	1.000
$d=4$	0.274	0.030	0.041	0.036	0.024	0.017	0.793	0.296	0.628	0.992	0.972	0.947	0.997	0.653	0.992	1.000	1.000	1.000
$d=5$	0.326	0.018	0.024	0.046	0.027	0.017	0.881	0.348	0.728	1.000	0.998	0.984	1.000	0.667	0.998	1.000	1.000	1.000
$d=6$	0.373	0.015	0.021	0.046	0.023	0.015	0.938	0.394	0.826	1.000	0.999	0.998	1.000	0.681	0.998	1.000	1.000	1.000

Notes: Rejection frequencies of Neyman's smooth test based on the transformations introduced in Sections 2.3 and 2.4 for the null hypothesis of a Gaussian CCC-GARCH(1,1). The alternatives are CCC-t-GARCH(1,1) models with 8 (H_{11}) and 4 (H_{11}) degrees of freedom, respectively. All Monte Carlo simulations are based on 10,000 iterations. The exact hypotheses are defined in Section 3.1.

Table B.3: Size and power – Tests using the G-ACR approach

Size	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.051	0.048	0.052	0.050	0.050	0.054	0.051	0.051	0.049	0.049	0.051	0.052	0.053	0.051	0.050	0.048	0.048	0.049
$d = 3$	0.049	0.051	0.045	0.050	0.051	0.053	0.051	0.048	0.051	0.049	0.052	0.048	0.050	0.053	0.049	0.050	0.047	0.051
$d = 4$	0.049	0.050	0.051	0.051	0.050	0.051	0.051	0.051	0.051	0.049	0.046	0.050	0.049	0.047	0.051	0.049	0.052	0.052
$d = 5$	0.051	0.050	0.051	0.050	0.048	0.050	0.053	0.052	0.047	0.049	0.052	0.050	0.050	0.051	0.048	0.052	0.051	0.048
$d = 6$	0.051	0.053	0.052	0.048	0.049	0.052	0.050	0.048	0.049	0.048	0.048	0.048	0.049	0.044	0.048	0.044	0.049	0.048

Power against H_{13}	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.394	0.265	0.144	0.078	0.070	0.061	0.591	0.379	0.189	0.088	0.093	0.095	0.799	0.523	0.237	0.102	0.097	0.114
$d = 3$	0.485	0.278	0.104	0.072	0.067	0.073	0.725	0.402	0.123	0.078	0.078	0.080	0.923	0.518	0.131	0.127	0.110	0.115
$d = 4$	0.593	0.287	0.089	0.099	0.087	0.084	0.808	0.409	0.092	0.107	0.102	0.102	0.977	0.549	0.102	0.101	0.121	0.092
$d = 5$	0.659	0.294	0.072	0.096	0.088	0.104	0.883	0.388	0.078	0.090	0.094	0.082	0.984	0.527	0.081	0.126	0.117	0.112
$d = 6$	0.710	0.281	0.068	0.111	0.098	0.113	0.914	0.390	0.071	0.094	0.091	0.097	0.994	0.547	0.068	0.118	0.115	0.133

Power against H_{14}	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.104	0.044	0.065	0.217	0.232	0.181	0.118	0.065	0.087	0.352	0.367	0.289	0.189	0.072	0.108	0.564	0.549	0.415
$d = 3$	0.101	0.041	0.084	0.350	0.349	0.301	0.143	0.047	0.100	0.562	0.535	0.429	0.296	0.075	0.147	0.811	0.806	0.665
$d = 4$	0.111	0.036	0.080	0.459	0.458	0.365	0.234	0.054	0.110	0.734	0.713	0.608	0.387	0.057	0.210	0.926	0.925	0.842
$d = 5$	0.149	0.029	0.083	0.564	0.546	0.464	0.242	0.047	0.126	0.831	0.820	0.727	0.497	0.060	0.259	0.984	0.980	0.942
$d = 6$	0.161	0.044	0.099	0.680	0.666	0.556	0.281	0.045	0.166	0.895	0.879	0.823	0.566	0.074	0.295	0.996	0.989	0.984

Power against H_{15}	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.376	0.254	0.162	0.262	0.245	0.216	0.585	0.355	0.202	0.373	0.354	0.291	0.794	0.472	0.288	0.547	0.538	0.450
$d = 3$	0.506	0.237	0.115	0.322	0.326	0.282	0.702	0.330	0.157	0.477	0.487	0.405	0.901	0.426	0.171	0.755	0.748	0.616
$d = 4$	0.584	0.260	0.097	0.426	0.410	0.353	0.782	0.277	0.119	0.618	0.609	0.536	0.967	0.463	0.159	0.868	0.859	0.793
$d = 5$	0.650	0.253	0.103	0.475	0.469	0.420	0.861	0.335	0.114	0.736	0.756	0.642	0.987	0.446	0.154	0.939	0.935	0.867
$d = 6$	0.734	0.275	0.120	0.580	0.564	0.491	0.933	0.337	0.134	0.826	0.800	0.759	0.993	0.461	0.191	0.978	0.966	0.948

Power against H_{16}	$P = 50$						$P = 100$						$P = 200$					
	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$d = 2$	0.416	0.218	0.191	0.561	0.554	0.469	0.630	0.321	0.274	0.789	0.785	0.692	0.866	0.473	0.434	0.967	0.959	0.918
$d = 3$	0.539	0.224	0.183	0.735	0.720	0.640	0.780	0.273	0.282	0.931	0.934	0.878	0.959	0.400	0.461	0.999	0.996	0.991
$d = 4$	0.633	0.235	0.176	0.822	0.816	0.720	0.845	0.277	0.291	0.972	0.971	0.948	0.991	0.352	0.514	1.000	0.999	0.999
$d = 5$	0.719	0.224	0.193	0.897	0.901	0.820	0.942	0.280	0.333	0.998	0.995	0.969	0.999	0.389	0.655	1.000	1.000	1.000
$d = 6$	0.790	0.218	0.228	0.936	0.921	0.862	0.959	0.284	0.380	0.998	0.996	0.993	0.999	0.397	0.696	1.000	1.000	1.000

Notes: Rejection frequencies for joint tests of uniform distribution and independence of U_t^W using the G-ACR approach described in Section 2.6.3. All Monte Carlo simulations are based on 10,000 iterations. H_{13} involves only dynamic misspecification. H_{14} involves only a misspecification of the distribution. H_{15} and H_{16} involve dynamic misspecification plus a deviation from the null distribution. The exact hypotheses are defined in Section 3.1.