Order Invariant Evaluation of Multivariate Density Forecasts

Jonas Dovern & Hans Manner
Heidelberg University & University of Cologne
jonas.dovern@awi.uni-heidelberg.de — +49 (6221) 54 2958

Abstract
We derive new tests for proper calibration of multivariate density forecasts based on Rosenblatt probability integral transforms. These tests have the advantage that they do not depend on the ordering of variables in the forecasting model, ii) are applicable to densities of arbitrary dimensions, and iii) have superior power relative to existing approaches. We furthermore devise additional tests that allow for estimated parameters and, consequently, can be used as in-sample specification tests. We demonstrate the problems of existing tests and how our new approaches can overcome those using Monte Carlo simulations as well as two applications based on multivariate GARCH-based models for stock market returns and on a macroeconomic Bayesian vector autoregressive model.

Motivation

• Only few existing tests for proper forecast calibration in multivariate setup:
  – Most based on multivariate version of probability integral transforms (PITs).
  – Existing tests suffer from two main shortcomings:
    – Most tests available only for d ≤ 2 or d ≤ 3.
    – Sensitive to the ordering of variables (”Prove to cheating”).

• Issue of dependence of test statistic on ordering of variables not addressed in literature.

Research Questions

• How can we design order invariant tests of whether a multivariate predictive density coincides with the true (conditional) density function that are order invariant, i.e., how can we design test which do not depend on the ordering of variables in the forecast model?

• Which tests for proper calibration of density forecasts perform best in large dimensional settings?

Main Contributions

• We generalize existing tests for proper calibration of multivariate density forecasts to arbitrary dimensional problems.
• We derive new tests which are order invariant in general.
• We present a formal accounting of conditions under which different tests are order invariant.
• We develop adjusted versions of our tests that account for estimation uncertainty.
• We analyze size and power (against various deviations from the null hypothesis) of different tests in a Monte Carlo study.
• We present two applications (forecasting financial returns/macroeconomic variables) that demonstrates the usefulness of our new tests.

Theory

Background

Basic question: Does estimated/forecast distribution proper calibration

Theory in the multivariate case: order invariance

Problem in the multivariate case: distribution of UT under H0 is unknown.

Solution is based on the Rosenblatt transformation:

UT = F−1 d,V1,...,Vn,...,Vn (Y1,...,Yn) , \[ \mathcal{U}_{d}^{(k)} = F_{\mathcal{U},|d,...,1-k|}^{-1} \]

All terms are i.i.d. and independent of each other.

Existing Tests

General idea. Transform multivariate problem into a univariate one, i.e., aggregate the d components into a single one with known distribution.

• Diebold et al. (1999), stack all PITs \( S_{t} = \left( U_{t}^{1}, \ldots, U_{t}^{d} \right) \)
• Clements and Smith (2000, 2002), multiply all PITs \( U_{t}^{d} = \mathcal{F}_{d}(Y_{t}) \) \( \prod_{i=1}^{d} \mathcal{F}_{i}(Y_{t}) \)
• Ko and Park (2013), multiply location adjusted PITs \( U_{t}^{d} = \mathcal{F}_{d} - \mathcal{F}_{d}

Order Invariance

There are \( d! \) permutations possible (denoted by \( \pi_{k} \) for \( k = 1, \ldots, d! \)).

Definition 1. Let \( T(\pi_{k}) \) be a test statistic based on \( Y_{t}^{\pi_{k}} \) under permutation \( \pi_{k} \). We call a test statistic \( T(\pi_{k}) \) order invariant if \( T(\pi_{k}) = T(\pi_{k}) \) for \( \forall k \neq \).

New Tests

Alternative transformation I: \( Z_{d}^{1} = \sum_{i=1}^{d} \left( U_{t}^{i} - \frac{1}{n} \right) \left( Z_{i}^{(d)} - \frac{1}{n} \right) \) 
\( H_{0} \) implies that \( \sum_{i=1}^{d} \chi_{i}^{2} = 0 \) in case of \( \sum_{i=1}^{d} \chi_{i}^{2} \).

Alternative transformation II: \( Z_{d}^{2} = \sum_{i=1}^{d} \left( U_{t}^{i} - \frac{1}{n} \right) \left( Z_{i}^{(d)} - \frac{1}{n} \right) \) 
This is the sum of squares of all distinct “inverse PITs” for all possible permutations. In general, terms are not independent of each other according to \( \chi_{i}^{2} \) distribution under \( H_{0} \). Instead, distribution follows a mixture of \( \chi_{i}^{2} \) distributions.

Alternative transformation III: \( Z_{d}^{3} = \sum_{i=1}^{d} \left( U_{t}^{i} - \frac{1}{n} \right) \) 
Similar to \( Z_{d}^{2} \) but considers only the terms which are conditional on all but one variable, i.e., \( \sum_{i=1}^{d} \left( U_{t}^{i} - \frac{1}{n} \right) \).

Results

Monte Carlo Simulations

• Order-dependence can cause huge size-distortions if researcher wants to “cheat”.

• New tests perform equally well or better than existing tests against various alternatives and for all dimensions.

Macroeconomic BVAR

• FV-BVAR by Primiceri (2005) for unemployment rate, inflation, and short-term interest rate.
• Non-parametric methods for comparing PITs/alternative: approximation by Normal distribution.

1-step ahead forecasts

<table>
<thead>
<tr>
<th>( S )</th>
<th>( p )</th>
<th>( p^* )</th>
<th>( g^2 )</th>
<th>( g^{2*} )</th>
<th>( g^{21} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>0.374</td>
<td>0.667</td>
<td>0.002</td>
<td>0.200</td>
<td>0.276</td>
</tr>
<tr>
<td>1-3</td>
<td>0.522</td>
<td>0.216</td>
<td>0.769</td>
<td>0.158</td>
<td></td>
</tr>
<tr>
<td>2-1</td>
<td>0.402</td>
<td>0.644</td>
<td>0.005</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>0.385</td>
<td>0.184</td>
<td>0.055</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>3-2</td>
<td>0.405</td>
<td>0.314</td>
<td>0.366</td>
<td>0.112</td>
<td></td>
</tr>
<tr>
<td>2-3-1</td>
<td>0.484</td>
<td>0.558</td>
<td>0.271</td>
<td>0.164</td>
<td></td>
</tr>
</tbody>
</table>

Normal approximation

<table>
<thead>
<tr>
<th>( S )</th>
<th>( p )</th>
<th>( p^* )</th>
<th>( g^2 )</th>
<th>( g^{2*} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>0.032</td>
<td>0.110</td>
<td>0.058</td>
<td>0.001</td>
</tr>
<tr>
<td>1-3</td>
<td>0.027</td>
<td>0.116</td>
<td>0.154</td>
<td></td>
</tr>
<tr>
<td>2-1</td>
<td>0.032</td>
<td>0.125</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>0.017</td>
<td>0.150</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>3-2</td>
<td>0.035</td>
<td>0.164</td>
<td>0.056</td>
<td></td>
</tr>
<tr>
<td>2-3-1</td>
<td>0.089</td>
<td>0.149</td>
<td>0.008</td>
<td></td>
</tr>
</tbody>
</table>

Extensions

• Autocorrelation: it is straightforward to implement autocorrelation-robust tests (e.g., the case that multi-step ahead forecasts are analyzed). Estimated Parameters: adding “randomness” to the transformation (based on an idea by Durbin (1961)) can be used to allow for estimated parameters (relevant for in-sample evaluations).

Conclusions

• New tests are order invariant, applicable to high-dimensional problems, and have better power than existing tests.
• Issue of “cheating” can be very relevant in practice.
• In both applications, existing test results not unambiguous (across permutations).
• Many future applications: DSGE forecasts/electricity demand on connected markets/etc.

Main References

• Ishida, I., 2005. Scanning multivariate conditional densities with probability integral transforms. The University of Tokyo.

Acknowledgements

We would like to thank Malte Knüppel, Fabian Krüger, Mark Wolters, and the participants of seminars at RWTH Aachen University, Heidelberg University, University of Cologne, and Maastricht University for their helpful comments on an early draft of the paper. All errors are our own responsibility.