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

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

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 i) 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 develop adjusted 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 Simulation as well as two applications based on multivariate GARCH-based models for stock market returns and on a macroeconomic Bayesian vectorautoregressive 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 $d \leq 2$ or $d \leq 3$.
  - Sensitive to the ordering of variables ⇒ "Prone 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 $\hat{F}_t$ coincide with the true distribution F_t?

One important condition is **proper calibration**:

- Statistical consistency between $\hat{F}_t$ and the realized observations y_t for $t = 1, \ldots, n$

In the univariate case, if $\hat{F}_t = F_t$, then so-called probability integral transforms (PITs) are uniformly distributed:

$$U_t = \int_{-\infty}^{Y_t} \hat{f}_t(y)dy = \hat{F}_t(Y_t) \sim \mathcal{U}(0,1)$$

Test uniformity of $\{U_t\}_{t=1}^n$ with Kolmogorov-Smirnov (KS) or Neyman smooth test (NST).

Problem in the multivariate case: distribution of U_t under H_0 is unknown.

Solution is based on the **Rosenblatt transformation**:

$$U_t^1 = \hat{F}_{Y_1}(Y_{1,t}),\ U_t^{2|1} = \hat{F}_{Y_2|Y_1}(Y_{2,t}),\quad \ldots \quad,\ U_t^{d|d-1,\ldots,1} = \hat{F}_{Y_d|Y_{d-1},\ldots,Y_1}(Y_{d,t})$$

All terms are $\mathcal{U}(0,1)$ 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): $S_t = [U_t^{d|d-1,\ldots,1}, \ldots, U_t^1]'$
- Clements and Smith (2000, 2002), multiply all PITs (P): $P_{t,d} = g(Y_t) = \prod_{i=1}^d U_t^{i|1:i-1}$
- Ko and Park (2013), multiply location adjusted PITs (P^*): $P^*_{t,d} = g(Y_t) = \prod_{i=1}^d (U_t^{i|1:i-1} - 0.5)$

Order Invariance

There are $d!$ permutations possible (denoted by π_k for $k = 1, \ldots, d!$).

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 $T(\pi_k) = T(\pi_j)$, $\forall\, k \neq j$.*

New Tests

Alternative transformation I: $Z_t^2 = \sum_{i=1}^d \left(\Phi^{-1}\left(U_t^{i|1:i-1}\right)\right)^2$

H_0 implies that $Z^2_{t,d} \sim \chi^2_d \Rightarrow$ Test uniformity of $U_t^{Z^2} = F_{\chi^2_d}(Z_t^2)$. (In Gaussian setting equal to transformation proposed by Ishida (2005).)

Alternative transformation II: $Z_t^{2*} = \sum_{i=1}^d \sum_{k=1}^{2^{d-1}} \left(\Phi^{-1}\left(U_t^{i|\gamma_i^k}\right)\right)^2$

This is the sum of squares of all distinct "inverse PITs" for **all** possible permutations. In general, terms are not independent of each other → no χ^2 distribution under H_0. Instead, distribution follows a mixture of χ^2 distributions.

Alternative transformation III: $Z_t^{2\dagger} = \sum_{i=1}^d \left(\Phi^{-1}\left(U_t^{i|-i}\right)\right)^2$

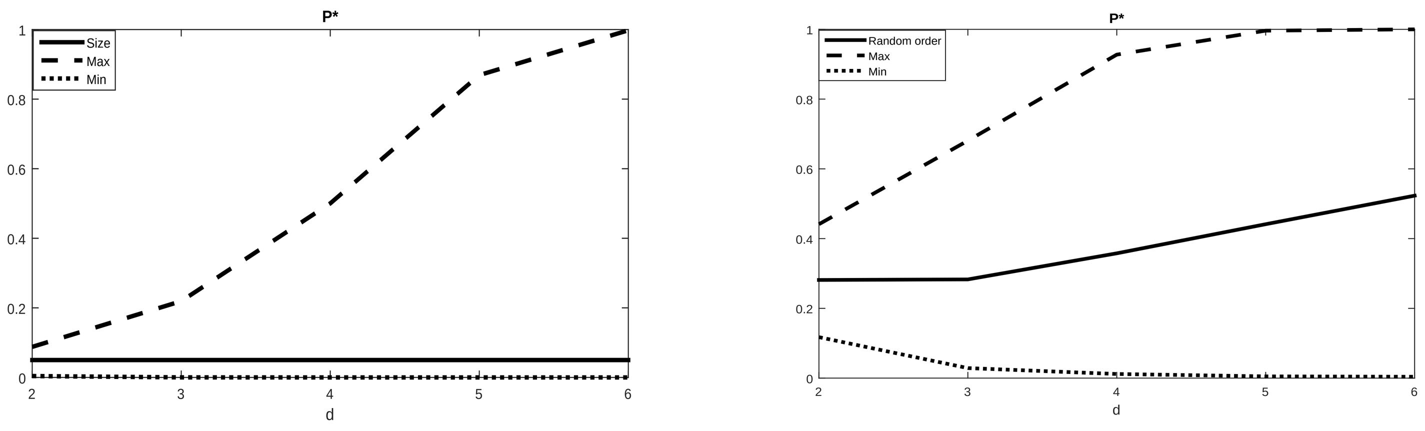
Similar to Z_t^{2*} but considers only the terms which are conditional on all but one variable, i.e., $U^t_{i|\{1,\ldots,d\}\setminus i}$. Distribution follows directly from distribution of Z_t^{2*}.

	S	P	P^*	Z^2	Z^{2*}	$Z^{2\dagger}$
Reference	DHT (1999)	CS (2000)	KP (2013)			
Order invariant?						
Independence	✓	✓	✓	✓	✓	✓
Gaussianity				✓	✓	✓
In general					✓	✓
Feasible for large d?	✓	✓	✓	✓		✓

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

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

1-step-ahead

non-parametric densities	S	P	P^*	Z^2	Z^{2*}	$Z^{2\dagger}$
1-2-3	0.374	0.667	0.022	0.006	0.230	0.276
1-3-2	0.552	0.216	0.769	0.158		
2-1-3	0.402	0.644	0.005	0.004		
2-3-1	0.385	0.184	0.055	0.083		
3-1-2	0.366	0.314	0.366	0.112		
3-2-1	0.484	0.556	0.271	0.164		

Normal approximation	S	P	P^*	Z^2	Z^{2*}	$Z^{2\dagger}$
1-2-3	0.032	0.110	0.058	0.001	0.001	0.000
1-3-2	0.027	0.116	0.154			
2-1-3	0.032	0.125	0.021			
2-3-1	0.007	0.150	0.005			
3-1-2	0.005	0.166	0.009			
3-2-1	0.009	0.149	0.008			

4-step forecasts

non-parametric densities	S	P	P^*	Z^2	Z^{2*}	$Z^{2\dagger}$
1-2-3	0.052	0.042	0.208	0.034	0.063	0.223
1-3-2	0.051	0.038	0.194	0.088		
2-1-3	0.010	0.067	0.024	0.022		
2-3-1	0.020	0.007	0.000	0.007		
3-1-2	0.122	0.004	0.755	0.039		
3-2-1	0.032	0.008	0.240	0.129		

Normal approximation	S	P	P^*	Z^2	Z^{2*}	$Z^{2\dagger}$
1-2-3	0.020	0.004	0.595	0.000	0.000	0.002
1-3-2	0.057	0.142	0.267			
2-1-3	0.028	0.008	0.605			
2-3-1	0.086	0.144	0.238			
3-1-2	0.122	0.007	0.097			
3-2-1	0.099	0.004	0.305			

Extensions

- Autocorrelation: it is straightforward to implement autocorrelation-robust tests (e.g., in 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

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

Diebold, F. X., Hahn, J., 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.

Durbin, J., 1961. Some methods of constructing exact tests. Biometrika 48, 41–55.

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.

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

Primiceri, G. E., 2005. Time varying structural vector autoregressions and monetary policy. Review of Economic Studies 72 (3), 821–852.

Acknowledgements

We would like to thank Malte Knüppel, Fabian Krüger, Maik 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.