Order Invariant Evaluation of Multivariate Density Forecasts

Jonas Dovern & Hans Manner

Heidelberg University & University of Graz

jonas.dovern@awi.uni-heidelberg.de — +49 (6221) 54 2958

Abstract

Existing tests for proper calibration of multivariate density forecasts based on Rosenblatt probability integral transforms can be manipulated by by a change in the ordering of variables in the forecasting model. We derive tests that do not depend on the ordering of variables. 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 have superior power relative to existing approaches. We use the tests to evaluate forecasts from multivariate GARCH models for stock market returns and from a macroeconomic Bayesian VAR model.

Motivation

- More and more application use multivariate models to form predictive densities.
- Often, the joint predictive density is of primary interest (e.g., when multiple input variables enter a decision problem)
- Existing tests for proper forecast calibration in multivariate setups have serious limitations:
  - Sensitive to the ordering of variables ⇒ "Prone to cheating".
  - Focus only on bivariate case.
- Issue of dependence of test statistic on ordering of variables not yet 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? In other words: How can we design tests 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?
- How can we take dynamic misspecification and estimation uncertainty into account?

Main Contributions

- We generalize existing tests for proper calibration of multivariate density forecasts to settings of arbitrary dimension.
- We derive new tests which are order invariant in general.
- We develop versions of our tests that account for estimation uncertainty and dynamic misspecification.
- We analyze size and power (against various deviations from H_0) of different tests in MC studies.
- We present two applications (forecasting financial returns/macroeconomic variables) that demonstrates the usefulness of our new tests.

Theory

Background

Basic question: Does (predictive) distribution $F_t(Y_{t+h}|\Omega_{t-1}, \theta_0)$ coincide with the true (conditional) distribution $G_t(Y_{t+h}|\Im_{t-1})$?

- We allow for dynamic misspecification ($\Omega_{t-1} \subset \Im_{t-1}$)
- Potentially, θ_0 can be replaced by an estimate $\hat{\theta}$.
- Sample $\{Y_t, \Omega_{t-1}\}_{t=1}^n$, of which the first R observations can be used to estimate θ_0 and remaining P observations are for evaluation.

Testable implication is that of **proper calibration**: Statistical consistency between F_t and realized observations.

In the univariate case, if $F_t = G_t$, then so-called probability integral transforms (PITs) are uniformly distributed:

$$U_t = \int_{-\infty}^{Y_t} f_t(Y)dY = F_t(Y_t) \sim \mathcal{U}(0,1)$$

Any appropriate test (e.g., Neyman smooth test, NST) can be used to test uniformity.

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

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

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

Under H_0, all terms are $\mathcal{U}(0,1)$ and independent of each other.

Test of H_0 is possible by **transforming multivariate problem into a univariate one**, i.e., aggregating the d components into a single one with known distribution and applying any goodness-of-fit test (we use NST).

Existing Tests

- Diebold et al. (1999), stack all PITs: $S_t = [U_t^{d|d-1,\ldots,1}, \ldots, U_t^1]'$
- Clements and Smith (2000, 2002), multiply all PITs: $CS_t = g(Y_t) = \prod_{i=1}^d U_t^{i|1:i-1}$
- Ko and Park (2013), multiply location adjusted PITs: $KP_t = g(Y_t) = \prod_{i=1}^d (U_t^{i|1:i-1} - 0.5)$

Order Invariance

Definition 1. *Denote the $d!$ possible permutations of the variables by π_k for $k = 1, \ldots, d!$. Let $T(\pi_k)$ be a test statistic based on $\{Y_t\}_{t=R+1}^n$ 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 $Z_{t,d}^2 \sim \chi_d^2 \Rightarrow$ Test uniformity of $U_t^{Z^2} = F_{\chi_d^2}(Z_t^2)$. In Gaussian settings this is equal to the 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. Distribution follows directly from distribution of Z_t^{2*}.

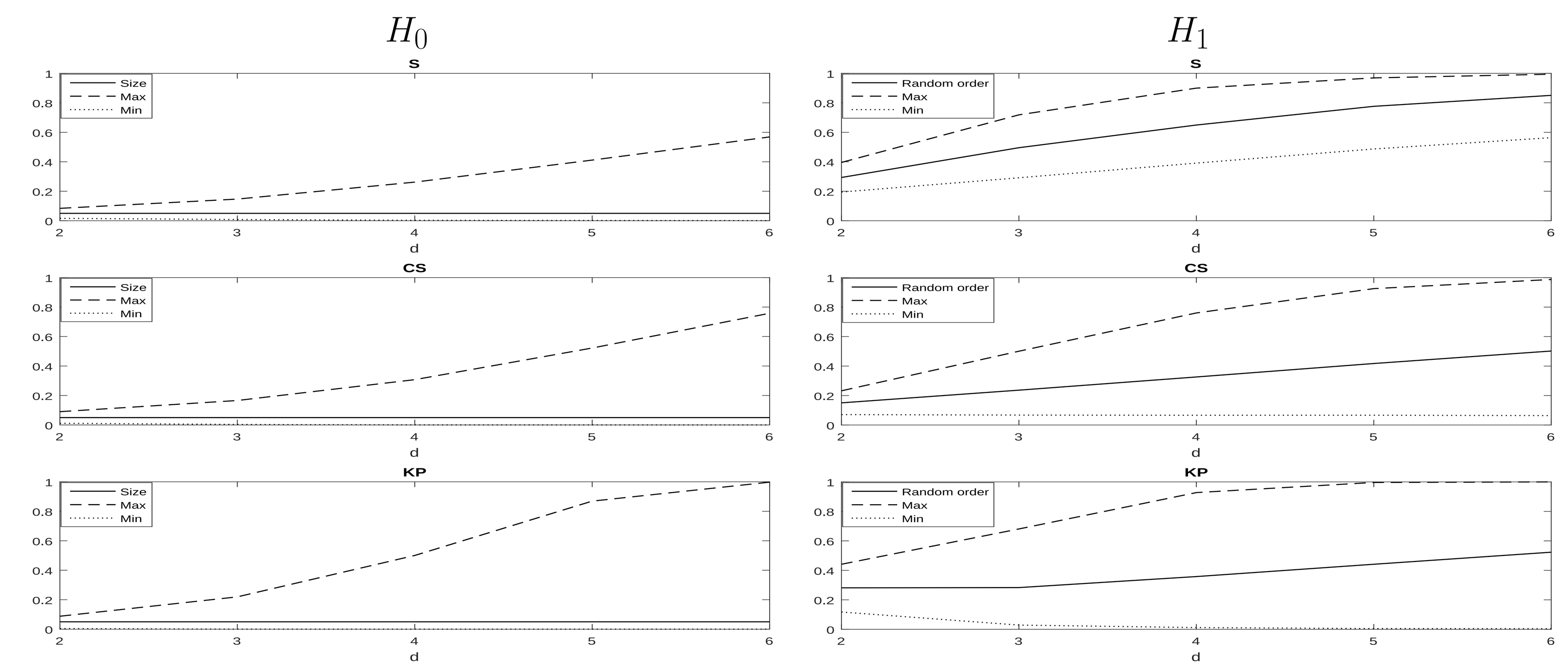
	S	P	P^*	Z^2	Z^{2*}	$Z^{2\dagger}$
Reference	DHT (1999)	CS (2000)	KP (2013)			
Order invariant?						
Independence	✓	✓	✓	✓	✓	✓
Gaussianity				✓	✓	✓
In general					✓	✓
Feasible for large d?	✓	✓	✓	✓		✓

Dynamic misspecification and estimated parameters: NST can be adjusted to account for both features relying on results in West (1996) and West and McCracken (1998).

Results

Monte Carlo Simulations

- Order-dependence offers much room for distortion of rejection frequencies if researcher wants to "cheat".



- New tests perform equally well or better in terms of power than existing tests against various alternatives and for all dimensions.
- Dynamic misspecification and parameter uncertainty lead to severe size distortions if not accounted for - modified tests are well sized.

Macroeconomic BVAR

- TVP-BVAR by Primiceri (2005) for unemployment rate, inflation, and short-term interest rate.
- PITs are based on-parametric methods/approximated based on Normal distribution and we take potential dynamic misspecification and estimation uncertainty into account.

	Normal approximations						Non-parametric densities					
$h=1$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$	S	CS	KP	Z^2	Z^{2*}	$Z^{2\dagger}$
$u_t - \Delta p_t - i_t$	0.032	0.110	0.058	0.001	0.001	0.000	0.374	0.667	0.022	0.006	0.230	0.107
$u_t - i_t - \Delta p_t$	0.027	0.116	0.154				0.552	0.216	0.769	0.158		
$\Delta p_t - u_t - i_t$	0.032	0.125	0.021				0.402	0.644	0.005	0.004		
$\Delta p_t - i_t - u_t$	0.007	0.150	0.005				0.385	0.184	0.055	0.083		
$i_t - u_t - \Delta p_t$	0.005	0.166	0.009				0.366	0.314	0.366	0.112		
$i_t - \Delta p_t - u_t$	0.009	0.149	0.008				0.484	0.556	0.271	0.164		
$h=4$	DHT	CS	KP	Z_t^2	Z_t^{2*}	$Z_t^{2\dagger}$	DHT	CS	KP	Z_t^2	Z_t^{2*}	$Z_t^{2\dagger}$
$u_t - \Delta p_t - i_t$	0.164	0.058	0.692	0.060	0.058	0.140	0.493	0.006	0.730	0.593	0.442	0.531
$u_t - i_t - \Delta p_t$	0.310	0.449	0.396				0.413	0.302	0.063	0.516		
$\Delta p_t - u_t - i_t$	0.167	0.080	0.685				0.346	0.015	0.319	0.562		
$\Delta p_t - i_t - u_t$	0.598	0.625	0.731				0.262	0.000	0.346	0.551		
$i_t - u_t - \Delta p_t$	0.680	0.020	0.508				0.595	0.001	0.850	0.553		
$i_t - \Delta p_t - u_t$	0.662	0.016	0.667				0.501	0.081	0.565	0.440		

Notes: The table shows the p-values for NST for different transformations and all possible permutations of the data. For those transformations that yield order-invariant test statistics, we only report one p-value.

Conclusions

- New tests are order invariant, applicable to high-dimensional problems, they can be adjusted to account for dynamic misspecification and parameter uncertainty, 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 potential applications: DSGE forecasts, electricity demand on connected markets, . . .

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.

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.

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

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

West, K., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1087.

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

Acknowledgements

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.