NHH

INSTITUTT FOR FORETAKSØKONOMI

DEPARTMENT OF BUSINESS AND MANAGEMENT SCIENCE

FOR 12 2014

ISSN: 1500-4066

March 2014

Discussion paper

A Likelihood Ratio and Markov Chain Based Method to Evaluate Density Forecasting

BY

Yushu Li AND **Jonas Andersson**

Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS

A Likelihood Ratio and Markov Chain Based Method to Evaluate Density Forecasting

Yushu Li and Jonas Andersson

Department of Business and Management Science, Norwegian School of Economics, Norway

Abstract: In this paper, we propose a likelihood ratio and Markov chain based method to evaluate density forecasting. This method can jointly evaluate the unconditional forecasted distribution and dependence of the outcomes. This method is an extension of the widely applied evaluation method for interval forecasting proposed by Christoffersen (1998). It is also a more refined approach than the pure contingency table based density forecasting method in Wallis (2003). We show that our method has very high power against incorrect forecasting distributions and dependence. Moreover, the straightforwardness and ease of application of this joint test provide a high potentiality for further applications in both financial and economical areas.

JEL classification: C14, C53, C61

Key Words: Likelihood ratio test, Markov Chain, Density forecasting

I. Introduction

One of the most distinguished applications of the econometric model is forecasting, and the evaluating procedures to assess different forecasting methods occupy the same importance. There exists three categories in constructing the *ex ante* forecasting and evaluating the forecasting based on *ex post* realisation: point forecasting, interval forecasting and density forecasting. The point forecasting provides a single possible outcome such as the expected mean or median. The evaluation includes the Euclidean distance metric such as the mean squared error (*MSE*), mean average error (*MAE*) and in the decision-theoretical framework. There are several examples in the literature of the evaluation of a point estimate (Wallis, 1995; Diebold and Lopez, 1996; Gneiting, 2011). However, the point forecast only provides a possible outcome and ignores most of the uncertainty. A more advanced forecasting method is interval forecasting which can evaluate the probability that an outcome will fall within a stated interval (Granger *et al.*, 1989; Chatfield, 1993). The related research into the evaluation of interval forecasting exploded after Christoffersen (1998) proposed a complete theory to evaluate the interval forecast. The evaluation procedure in Christoffersen (1998) is based on the likelihood ratio test (*LRT*) and due to the additivity of the likelihood ratio test, the method can jointly test the unconditional coverage and independence by testing the correct conditional coverage. This test and its extensions (Clements and Taylor, 2003; Engle and Manganelli, 2004; Berkowitz *et al.*, 2011; Dumitrescu *et al.*, 2011) are most widely used to evaluate an interval forecast, especially in the value-at-risk (*VaR*) analysis which can be viewed as a one-sided interval forecast.

Although the interval forecasting identifies a certain degree of uncertainty of the outcomes, it can basically be viewed as a "two points" point forecast with each point being the interval endpoint. In contrast, density forecasting, which explicitly states the possible density distribution, makes future statistical inference of the *ex post* outcomes possible. Tay and Wallis (2000) carried out an explicit survey of the density forecasting and pointed out the necessity of a complete and accurate forecast probability statement in macroeconomic forecasting such as inflation and output growth, financial forecasting such as portfolio returns and risk management and volatility The literature on evaluating the uncertainty of the density forecast is limited and mainly based on the idea of the probability integral transform (*PIT*) or its extension (Diebold *et al.*, 1998; 1999; Tay and Wallis, 2000; Berkowitz, 2001). This test is actually a new type of goodness of fit test and it lacks an analytical illustration to evaluate the internal dependence of the data. Wallis (2003) subsequently proposed Pearson chi-squared based statistics which can evaluate the goodness of fit and independence at same time. However, Wallis (2003) mainly recast the likelihood ratio based interval forecast by Christoffersen (1998) into the framework of a Pearson chi-squared test and the detailed theory of the density forecast is not developed in this work. Moreover, the evaluation method of Wallis (2003) is a chi-squared test rather than a *LRT*, it lacks the property of additivity and makes the joint test for distribution and independence imprecise. This paper will extend the likelihood ratio based method of Christoffersen (1998) analytically to evaluate density forecasting. Due to the additivity of the *LRT*, our method can jointly test the unconditional distribution and independence. We show that this test has very high power against distributional bias and dependence and we provide the R program for ease of application.

The paper can be divided into the following sections: section II gives a brief introduction to the likelihood interval forecast, section III describes our evaluation method for density forecasting and section IV details the Monte Carlo experiment. The conclusion can be found in the last section.

II. Likelihood ratio and Markov chain based interval forecast

For *ex post* realisation $Y = (y_1, y_2 ... y_T)$, the *ex anti* interval forecast made at time $t-1$ is $C_{t|t-1}(p) = [L_{t|t-1}(p), U_{t|t-1}(p)]$ where p is the probability of coverage. Define the indicator variable $\{I_t\}_{t=1}^T$ as $I_t = \begin{cases} 1, & y_t \in C_t(p) \\ 0, & y_t \notin C_t(p) \end{cases}$, where $I_t = 1$ when the *ex post* realization lies inside $C_{t|t-1}(p)$ and $I_t = 0$ otherwise. Christoffersen (1998) constructed a test framework to evaluate whether $C_{t|t-1}(p) = [L_{t|t-1}(p), U_{t|t-1}(p)]$ is a "efficient" interval forecast with respect to the past information $\Psi_{t-1} = \{I_t, I_{t-1}, ...\}$ by testing whether $E(I_t | \Psi_{t-1}) = p$. The evaluation framework includes three tests:

1. **The unconditional coverage test statistic** LR_{uc} : to test whether the first order moment expected value of indicator sequence $\{I_t\}_{t=1}^T$ is equal to the coverage rate. The test ignores the dependence of I_t and the null hypothesis is $E(I_t) = p$, while the alternative hypothesis is $E(I_t) = \pi \neq p$. Define n_0 as the number of I_t which lie in the coverage rate and n_0 as the number of I_t which lie outside the coverage rate as $n_0 = Nr(I_t = 0)$, $n_1 = Nr(I_t = 1)$. The likelihoods under the null and alternative hypotheses are $L_p = (1-p)^{n_0} p^{n_1}$ and $L_{\hat{\pi}} = (1-\hat{\pi})^{n_0} \hat{\pi}^{n1}$ respectively, where the relative hit frequency $\hat{\pi} = \frac{n_1}{n_0 + n_1}$ is the maximum likelihood estimate (*MLE*) of π. Then the likelihood ratio based test statistic $LR_{uc} = -2\log\frac{L_p}{L_{\hat{\pi}}} \sim \chi^2(1)$ under null hypothesis. Christoffersen (1998) mentioned that the pure unconditional coverage test will have very low power and is inefficient when $\{I_t\}_{t=1}^T$ is clustered in a time dependent fashion. Thus this author further introduced an independence test and also a joint test for independence and unconditional coverage.
2. **The independence test statistic** LR_{ind}: to test whether I_t is distributed independently over the whole period. This test is important as it takes into account the higher-order moment dynamics of $\{y_t\}_{t=1}^T$, such as the autocorrelations. In practical applications

such as volatility forecasting, independence means that there is no cluster of violation in the volatile intervals and there is no tenuity in the tranquil intervals. The likelihood ratio based test statistic LR_{ind} is constructed using a two state first order Markov Chain. We will provide a detailed illustration of LR_{ind} later when we construct our density forecasting evaluation method which is based on a k state Markov chain.

3. **Conditional coverage test statistic** LR_{cc}: to test whether the forecasting interval has correct conditional coverage in the form of $E(I_t | \Psi_{t-1}) = p$. As the test of unconditional coverage and independence will not affect each other, this conditional coverage test is the combination of unconditional coverage test and independence test. Due to the additivity of the *LRT* statistics (Bera and McKenzie, 1985), we have $LR_{cc} = LR_{uc} + LR_{ind}$ which can jointly test the randomness and correct coverage, while the test of individual subcomponents can still be retained.

The *LRT* by Christoffersen (1998) has been followed by several developments in the literature (Clements and Taylor, 2003; Engle and Manganelli, 2004; Berkowitz *et al.*, 2011; Dumitrescu *et al.*, 2011) in terms of both theoretical extensions and applications. One reason for the popularity of this type of test is that it can discover both the first order moment (bias) and the second order moment (correlation dynamics) of the forecast. This idea can be extended to evaluate the density forecast and this paper will construct the likelihood and Markov chain based tests for evaluating the density forecast.

III. Likelihood ratio and Markov chain based density forecast evaluations

The main trend in evaluating the density forecast (Diebold *et al.*, 1998; 1999; Tay and Wallis, 2000; Berkowitz, 2001) is built on a seminar paper of Diebold *et al.* (1998) and is based on the *PIT*. The main idea is that when the *ex anti* forecasted distribution $\{s_t(y_t)\}_{t=1}^{T}$ is the correct forecasting, then for the *ex post* realisation $Y = (y_1, y_2 ... y_T)$, we have $z_t = \int_{-\infty}^{y_t} s_t(u)du \sim i.i.d.\ U(0,1)$. Deviation from *i.i.d.* means that the *ex anti* forecast under the assumption of independence fails to capture the underlying correlated dynamics for the model, and the deviation of $U(0,1)$ gives a wrong *ex anti* forecast distribution. However, their test for independence is built on the visual graph of correlograms, by checking if there exists serial correlation in z_t , and it lacks theoretical analysis. Thus, their test of whether z_t is uniformly distributed is actually a goodness of fit test. Just as the independence of I_t is an important aspect to evaluate whether the interval forecast is efficient, we need a more sophisticated method that just correlograms to additionally check the independence of the distribution. This idea of combining the goodness of fit and independence tests is later presented in Wallis (2003a), where they recast the interval evaluation method of Christoffersen (1998) into a framework of contingency table based Pearson chi-squared test. However, their test still concentrates on interval forecasting evaluation based on contingency tables for small samples and no theoretical derivation of the likelihood ratio based density forecasting evaluation is proposed. Our proposed method will fill this gap and extend the likelihood ratio evaluation method for interval forecast to density forecast. The method is still constructed using three aspects: a test for goodness of fit, a test for independence and a joint test for goodness of fit and independence.

1. **Unconditional density test statistic** LR_{ud} **:** consider the *ex post* outcome $Y = (y_1, y_2 ... y_T)$ which is generated by the distribution $f(y_t)$ and the *ex anti* forecasted density $s(y_t)$. The range of y_t is $[I_0, I_n]$ with $I_0 < y_t < I_n$. We divide $[I_0, I_n]$ into k mutually exclusive states as $[\underbrace{I_0, I_1}_{1}, ..., \underbrace{I_{k-1}, I_n}_{k}]$ and let the number of y_t lies in state i be n_i. Note that the interval forecasting is a special case where $k = 2$ and the test statistic LR_{uc} is actually based on the likelihood from a binomial distribution. To evaluate whether $s(y_t)$ is the unbiased density distribution is equivalent to testing $f(y_t) = s(y_t)$. Under the null hypothesis $f(y_t) = s(y_t)$, $N = (n_1, n_2, ..., n_k)$ follows a multinomial distribution $multinom(T, p_1 ... p_k)$ with event probability $p_i = \int_{I_{i-1}}^{I_i} s_y(u)du$. Thus, the likelihood function under the null hypothesis is $L(p) = \dfrac{T!}{n_1! ... n_k!} p_1^{n_1} ... p_k^{n_k}$ where $p_i = \int_{I_{i-1}}^{I_i} s_y(u)du$. The likelihood function under the alternative hypothesis is $L(\hat{p}) = \dfrac{T!}{n_1! ... n_k!} \hat{p}_1^{n_1} ... \hat{p}_k^{n_k}$, where $\hat{p}_i = n_i / T$ is the *MLE* of the event probability over the whole parameter space. The *LRT* statistic is $LR_{ud} = -2\log(L(p)/L(\hat{p}))$ and $LR_{ud} \sim \chi^2(k-1)$ under the null hypothesis. Just as the unconditional coverage test statistic LR_{uc} in interval forecast, LR_{ud} can only discover the biasedness of the forecasted distribution with the null hypothesis being

$s(y_t) = f(y_t)$, and it can be viewed as a pure goodness of fit test. When taking the past information $\Omega_{t-1} = \{y_t, y_{t-1}, ...\}$ into account and setting the null hypothesis as $s(y_t)|\Omega_{t-1} = f(y_t)$, LR_{ud} will have no power if there is dependence of the higher order moments in $\{y_t\}_{t=1}^{T}$.

2. **Independence test statistic** LR_{Indd} **:** Wallis (2003a) mentioned that the test for independence in the interval forecast can be extended to the density forecast without further analytical derivations. The following will provide a detailed illustration. The independence is tested against a k state first order Markov chain. Let $\pi_{ij} = \Pr(y_t \in \text{ state } j | y_{t-1} \in \text{ state } i)$. Then the Markov chain is specified with the transition probability matrix $\Pi = \begin{bmatrix} \pi_{11} \ ... \ \pi_{1k} \\ \pi_{i,j} \\ \pi_{k1} \ ... \ \pi_{kk} \end{bmatrix}$. Let n_{ij} denote that the number of events that a state i is followed by a state j as $n_{ij} = Nr(y_t; y_t \in j \,\&\, y_{t-1} \in i)$. Then the likelihood function under the alternative hypothesis for the whole process is $L(\Pi) = (\pi_{11}^{n_{11}} ... \pi_{1k}^{n_{1k}})...(\pi_{i1}^{n_{i1}} ... \pi_{ik}^{n_{ik}})...(\pi_{k1}^{n_{k1}} ... \pi_{kk}^{n_{kk}}) = \prod_{i=1}^{k}\prod_{j=1}^{k} \pi_{ij}^{n_{ij}}$ with $\hat{\pi}_{ij} = \dfrac{n_{ij}}{\sum_{j=1}^{k} n_{ij}}$ being the *MLE* of π_{ij}. Under the null hypothesis of independence, the outcome in the present time will not be influenced by the past information. Thus, when the outcome y_t lies in state j, the nearest outcome y_{t-1} has the same probability of lying in any state and this can be denoted as $\pi_{1j} = \pi_{2j} ... = \pi_{kj} = \pi_{\cdot j}$. Thus, $(\pi_{11}^{n_{11}} ... \pi_{1k}^{n_{1k}})...(\pi_{i1}^{n_{i1}} ... \pi_{ik}^{n_{ik}})...(\pi_{k1}^{n_{k1}} ... \pi_{kk}^{n_{kk}}) = \prod_{j=1}^{k} \pi_{\cdot j}^{n_{\cdot j}}$ where $n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$. As $\pi_{\cdot j}$ is actually the probability that an outcome lies in state j and $n_{\cdot j}$ is the number of outcomes that lies in state j, the MLE of $\pi_{\cdot j}$ is $\hat{\pi}_{\cdot j} = \dfrac{n_j}{T}$ with $n_j = n_{\cdot j}$. Therefore, the approximate likelihood function under the null hypothesis is $L(\hat{\Pi}_0) = \prod_{j=1}^{k} (\dfrac{n_j}{T})^{n_j}$ and the unrestricted approximate likelihood function is $L(\hat{\Pi}_1) = \prod_{i=1}^{k}\prod_{j=1}^{k} (\dfrac{n_{ij}}{\sum_{j=1}^{k} n_{ij}})^{n_{ij}}$. The *LRT* for independence

is then $LR_{Ind} = -2\log\frac{L(\hat{\Pi}_0)}{L(\hat{\Pi}_1)} \sim \chi^2((k-1)^2)$. We notice that $L(\hat{\Pi}_0) \propto L(\hat{p})$ and this relationship will simplify the joint test statistics in the following paragraph.

3. **Conditional density test statistic** LR_{cd} **:** to test whether the conditional forecasted density distribution based on the past information $s(y_t)|\Omega_{t-1}$ provides efficient forecasting of the data distribution. As the conditional coverage test statistic LR_{cc} in the situation of interval forecasting, this test can be viewed as a combination of a goodness of fit test and a test for independence: we test whether $s(y_t) = f(y_t)$ and whether $\{y_t\}_{t=1}^{T}$ is independent. The test statistic can be constructed based on the additivity of the *LRT* (Bera and McKenzie, 1985): the test statistic to test a joint hypothesis is the sum of the test statistics which test the components of the null hypothesis separately. Then the test statistic LR_{cd} which can jointly test the independence and goodness of fit is $LR_{cd} = LR_{ud} + LR_{Ind}$. As :

$$LR_{ud} = -2\log\frac{L_p}{L_{\hat{\pi}}} = -2\log\left(\frac{\frac{T!}{n_1!\ldots n_k!}p_1^{n_1}\ldots p_k^{n_k}}{\frac{T!}{n_1!\ldots n_k!}\hat{p}_1^{n_1}\ldots \hat{p}_k^{n_k}}\right) = -2[\log(p_1^{n_1}\ldots p_k^{n_k}) - \log(\hat{p}_1^{n_1}\ldots \hat{p}_k^{n_k})]$$

$$LR_{Ind} = -2\log\frac{L(\hat{\Pi}_0)}{L(\hat{\Pi}_1)} = -2\log\left(\frac{\prod_{j=1}^{k}(\frac{n_j}{T})^{n_j}}{\prod_{i=1}^{k}\prod_{j=1}^{k}(\frac{n_{ij}}{\sum_{j=1}^{k}n_{ij}})^{n_{ij}}}\right) = -2[\log(\prod_{j=1}^{k}(\frac{n_j}{T})^{n_j}) - \log(\prod_{i=1}^{k}\prod_{j=1}^{k}(\frac{n_{ij}}{\sum_{j=1}^{k}n_{ij}})^{n_{ij}})]$$

and $\hat{p}_j = \frac{n_j}{T}$, then $LR_{cd} = LR_{ud} + LR_{Ind}$ can be simplified as

$$LR_{cd} = -2[\log(p_1^{n_1}\ldots p_k^{n_k}) - \log(\prod_{i=1}^{k}\prod_{j=1}^{k}(\frac{n_{ij}}{\sum_{j=1}^{k}n_{ij}})^{n_{ij}})] \sim \chi^2(k(k-1)) \text{ where } p_i = \int_{I_{i-1}}^{I_i} s_y(u)du.$$

Compared with LR_{ud}, which only has power against biased unconditional forecasted density and ignores the internal dependence of $\{y_t\}_{t=1}^{T}$, LR_{cd} has power against both misspecified density forecasting and internal correlation of the data series. Therefore, instead of only testing the first order moment unbiasedness of the forecasted distribution, LR_{cd} can discover whether there exists higher order moment dynamics such as conditional heteroscedasticity dependent upon the data generating process.

The LR_{cd} test can then be applied to evaluate the efficiency of the density forecasting. Under the null hypothesis $s(y_t)\left|\Omega_{t-1} = f(y_t)\right.$, or $s(y_t) = f(y_t)$ and $\{y_t\}_{t=1}^{T}$ is independent, $LR_{cd} = LR_{ud} + LR_{Ind} \sim \chi^2(k(k-1))$.

To investigate the efficiency of the test statistics LR_{ud}, LR_{Indd} and LR_{cd}, a Monte Carlo study is carried out in the next section. Due to that LR_{ud} is essentially the goodness of fit test, we will compare it with two other popular goodness of fit tests include the Kolmogorov-Smirnov (*KS*) test and the *PIT* test, as no comparison of those methods have previously been made.

IV. Monte Carlo study

The null hypothesis in the Monte Carlo study is that the forecasted density distributions $s(y_t)$ are separately a normal distribution, t distribution and truncated Cauchy distribution, and the observations $\{y_t\}_{t=1}^{T}$ are independent. These three distributions are chosen because based on the density function graph, they look very similar to each other and formal tests are needed. The data generation process (*DGP*) for $\{y_t\}_{t=1}^{T}$ will be designed to check how the tests will perform from both size and power perspectives and this process can be divided into two cases:

Case 1: $y_t \sim i.i.d.\ N(0,1.2),\ \ i.i.d.\ t(6),\ \ i.i.d.\ TCauchy(0,1,\text{-}4,\text{-}4)$

Case 2: $$\begin{aligned} &y_t = n_t\sqrt{h_t};\ h_t = 0.4 + 0.15y_{t-1}^2 + 0.45h_{t-1} \\ &n_t \sim i.i.d.\ N(0,1.2),\ \ i.i.d.\ t(6),\ \ i.i.d.\ TCauchy(0,1,\text{-}4,\text{-}4) \end{aligned}$$

The *DGP* in Case 1 is used for the goodness of fit and independence test as there exists no higher order moment dependence in the data series, and we can investigate both size and power of all five test statistics. The *DGP* in Case 2 is used to investigate the power of the tests as there exists higher order moment dependence in the GARCH(1,1) process. The GARCH(1,1) model is also the typical model which has been applied in previous research to evaluate interval forecasting (Christoffersen, 1998; Clements and Taylor, 2003) and density forecasting (Diebold *et al.*, 1998; Bao *et al.*, 2007. To save space, we present only the results for a sample size of 1500; this sample size can already produce an unbiased size in most of the tests, instead of 4000 in Diebold *et al.* (1998). The number of states *k* is initially chosen as the integer value of $1 + \log_2(T)$ and the interval length for each state interval is chosen as identical. If there exist empty bins based on the initial division, we combine the nearby bins until each bin contains observations. The Monte Carlo replication time is 1000 and results based on the *DGP* from Case 1 are as follows:

Table 1: Size and power of the tests when *DGP* is independent

S	t			Norm			TCauchy		
DGP	t	Norm	TCauchy	t	Norm	TCauchy	t	Norm	TCauchy
KS	0.048	0.740	0.999	0.735	0.001	1.000	1.000	0.926	0.042
PIT	0.050	0.827	1.000	0.736	0.001	1.000	1.000	0.926	0.042
LR_{ud}	0.036	0.975	1.000	0.994	0.018	1.000	1.000	1.000	0.025
LR_{Indd}	0.032	0.084	0.092	0.026	0.082	0.099	0.026	0.097	0.111
LR_{cd}	0.030	0.742	1.000	0.874	0.055	1.000	1.000	1.000	0.107

In Table 1, the first line denotes the forecasted distributions $s(y_t)$ as the t distribution, normal distribution, and truncated Cauchy distribution. For each distribution, we evaluate the size and power for all five tests when the *DGP* (in the second line) are from Case 1. The underlined values correspond to the size of the tests. For a Monte Carlo simulation time of 1000, the approximate 95% confidence interval for the estimated size at 5% significant levels is $0.05 \pm 1.96 * \sqrt{\frac{0.05(1-0.05)}{1000}} = (0.0365,\ 0.0635)$. Table 1 shows that when the aim is to test the goodness of fit, the size is unbiased in almost all tests. The LR_{ud} test has the highest power in the goodness of fit test while LR_{cd} has the next highest power. Thus, although LR_{cd} is a joint test for biased distributions and independence, it still has high power when used as a pure goodness of fit test. The test of independence, LR_{Indd} , shows a slightly inflated level of bias but this stays at an acceptable level. We next investigate the power of all tests when the *DGP* is from Case 2; the results are shown in Table 2.

Table 2: Power of the tests when *DGP* is not independent

S	t			Norm			TCauchy		
DGP	t	Norm	TCauchy	t	Norm	TCauchy	t	Norm	TCauchy
KS	0.070	0.914	1.000	0.969	0.005	1.000	1.000	0.969	1.000
PIT	0.071	0.934	1.000	0.970	0.005	0.974	0.886	0.969	1.000

LR_{ud}	0.062	0.986	1.000	0.999	0.243	1.000	1.000	1.000	0.954
LR_{Indd}	0.875	0.946	0.978	0.886	0.944	1.000	1.000	0.943	0.980
LR_{cd}	0.850	0.998	1.000	0.999	0.936	1.000	1.000	1.000	0.991

In Table 2, the null hypothesis is unchanged but the *DGP* includes a second order moment dependence and therefore, all values in Table 2 correspond to the power perspective. The underlined values in Table 2 show that LR_{ud}, KS and PIT have almost no power against dependence while both LR_{Indd}, LR_{cd} have very high power in all cases. Thus, the LR_{ud}, KS and PIT tests can only discover whether the unconditional distributions are correct and cannot investigate the internal dependence of the data. LR_{Indd} can be used to explore whether the observations are independent and LR_{cd} can jointly test the independence and the biasedness of the distribution. Combined with Table 1, we can conclude that LR_{cd} is the most efficient test in three aspects: it is unbiased under the null hypothesis and it has power against both dependence and incorrect distributions. In applications, we can carry out the evaluations step by step. The first step is to apply LR_{cd} to jointly test the independence and goodness of fit. If the null hypothesis is not rejected, we can conclude that $s(y_t)$ is the proper distribution and the outcomes are independent. However, if we reject the null hypothesis, we can further investigate whether the rejection is due to the dependence or incorrect distribution by applying LR_{ud} and LR_{Ind} separately.

V. Conclusion

This paper has proposed a test framework for the evaluation of density forecasting. The evaluation is built on the *LRT* statistics and the Markov chain. It is an extension of the interval forecasting evaluations of Christoffersen (1998). We show that in order to evaluate the efficiency of the forecasted distributions, the joint test LR_{cd} can test both the goodness of fit and the dependence, while LR_{ud}, KS and PIT only have power against biased distributions. The test statistic LR_{cd} is constructed using the additivity of the likelihood tests. We also show that the components for LR_{cd}, which are LR_{ud} and LR_{Indd}, have very high power against goodness of fit and dependence separately. We propose a three step evaluation process which can explore the full character of the underlying data generating process.

References

Bao, Y., Lee, T.H. and Saltoglu, B. (2007). “Comparing density forecast models”, *Journal of Forecasting*, 26(3), 203-225.

Bera, A.K. and McKenzie, C.R. (1985). “Alternative forms and properties of the score test”, *Journal of Applied Statistics*, 13, 13-25.

Berkowitz, J. (2001). “Testing density forecasts with applications to risk management”, *Journal of Business and Economic Statistics*, 19, 465-474.

Berkowitz, J., Christoffersen P. and Pelletier, D. (2011). “Evaluating value-at-risk models with desk-level data,” *Management Science*, 57(12), 2213-2227.

Chatfield, C. (1993). “Calculating interval forecasts”, *Journal of Business and Economic Statistics*, 11, 121-135.

Christoffersen, P.F. (1998). “Evaluating interval forecasts”, *International Economic Review,* 39, 840-841.

Clements, M. and Taylor, N. (2003). “Evaluating interval forecasts of high frequency financial data”, *Journal of Applied Econometrics,* 18, 445-456.

Diebold, F.X. and Lopez, J.A. (1996). “Forecast evaluation and combination”, *Handbook of Statistics 14: Statistical Methods in Finance*, Maddala GS, Rao CR (eds). North-Holland: Amsterdam.

Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998). “Evaluating density forecasts with applications to financial risk management”, *International Economic Review*, 39, 863-883.

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, 661-673.

Dumitrescu, E.L., Hurlin, C. and Madkour, J. (2011). “Testing interval forecasts: A GMM-based approach”, *Journal of Forecasting*, Published online in Wiley Online Library. (wileyonlinelibrary.com) DOI: 10.1002/for.1260.

Engle, R.F. and Manganelli, S. (2004). “CAViaR: Conditional autoregressive value-at-risk by regression quantiles”, *Journal of Business and Economics Statistics*, 22, 367-381.

Gneiting, T. (2011). “Making and evaluating point forecasts”, *Journal of the American Statistical Association,* 106:494, 746-762.

Granger, C.W.J., White, H. and Kamstra, M. (1989). “Interval forecasting: an analysis based upon ARCH quantile estimators”, *Journal of Econometrics* , 40, 87-96.

Tay, A.S. and Wallis, K.F. (2000). “Density forecasting: a survey”, *Journal of Forecasting,* 19, 235-254.

Wallis, K.F. (1995). “Large-Scale Macroeconometric Modeling,” *Handbook of Applied Econometrics*, Pesaran, M.H. and Wickens, M.R.(eds). Oxford: Blackwell, 312-355.

Wallis, K.F. (2003). “Chi-square tests of interval and density forecasts, and the Bank of England’s Fan Charts”, *International Journal of Forecasting*, 19, 165-175.