Detecting Density Forecast Breakdowns

Yiru Wang*

This version: February 14, 2019

Abstract

In this paper, I propose a new method for evaluating density forecasting models and develop a formal test to capture the density forecast breakdown (DFB), a situation where the out-of-sample density forecast performance is significantly worse than its anticipated performance. This test provides separate statistics for parametric and non-parametric density forecasting models. For parametric models, it allows for dynamic misspecification and corrects parameter estimation uncertainty. For non-parametric models, it provides conditions under which the estimation uncertainty is asymptotically irrelevant. Furthermore, I provide a robust version of this test to analyze the evolution of the consistency of the out-of-sample density forecasting performance and its anticipated performance. The Monte Carlo results indicate that the test has good size property in moderately large samples and has power against changes in mean and variance, as well as changes in distribution type. Empirically, I find that (i) DFBs occurred during 2008Q2 to 2010Q2 in the one-year-ahead real GDP growth modeled with current and lagged financial and economic conditions; (ii) no DFB occurred in the one-day-ahead stock price return using GARCH(1,1) model, but DFBs occurred during 07/19/2008 to 11/09/2011 using the kernel estimation and forecasting procedure.

*Department of Econometrics and Business, University Pompeu Fabra, Spain. Contact address: Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain. Email: yiru.wang@upf.edu.
Acknowledgements: I am greatly indebted to my advisor Barbara Rossi for the invaluable guidance and support. I would also like to thank Atsushi Inoue, Christian T. Brownlees, Donghai Zhang, Enrique Sentana, Geert Mesters, Majid M. Al-Sadoon, Silvia Miranda-Agrippino, Simon van Norden for their helpful comments and/or sharing codes and data with me. The paper was presented in the econometrics seminars at Universitat Pompeu Fabra. I would also like to thank the participants in the seminars. All remaining errors are mine.

1 Introduction

Recently, density forecasting has gained more and more interest, as it provides a complete characterization of the uncertainty associated with a prediction; in contrast, its alternative, point forecasting, offers no characterization of the associated uncertainty. Density forecasting plays an important role in both macroeconomics and finance, see Tay and Wallis (2000) for a survey. Central banks have put great effort into generating density forecasts for key macroeconomic variables. For example, the Survey of Professional Forecasters, published by the Federal Reserve Bank of Philadelphia, provides density forecasts for inflation and output growth; the Bank of England also publishes its quarterly Inflation Report, which includes the average survey responses related to density forecasts of inflation, as well as a density forecast of inflation represented analytically by a specific non-normal probability distribution. In addition to significant applications in macroeconomics, density forecasting is intimately related to applications in quantitative finance, such as risk management. JP Morgan, Reuters, and Bloomberg regularly issue density forecasts of the change in the value of customized portfolios over a particular holding period, generating the measure commonly known as value at risk (VaR) as the nth percentile of the distribution. Many empirical studies have also been conducted to evaluate the distributions of stock returns, interest rates, and other financial data series.

While researchers are developing procedures to construct density forecasts, as well as new approaches to improve density forecast performance, evaluation tools related are greatly needed. It is important for a density forecaster to know how to judge whether a density forecasting model is good and how robust density forecast results are. However, evaluating density forecasts is a difficult task. The challenges that central banks and researchers face in evaluating point forecasts also exist in evaluating density forecasts. It is widely known that good in-sample performance does not imply good out-of-sample performance and that out-of-sample performance may not be stable over different time periods. Furthermore, density forecasts are estimates of the probability distribution of the possible future values of the targeted variable; thus, both the density forecasts themselves and the related evaluation functions exhibit different properties from those in point forecasts. Considering the increasing importance of density forecasting, this paper proposes a new method for evaluating density forecasting models and develops a formal test to answer the following questions: Does a density estimated in one period give good density forecasts across time periods? And how well does a density forecasting model adapt to changes in the economy? Note that the goal of this paper differs from either an absolute predictive density evaluation or an evaluation of density estimations. Rather, this paper focuses on the consistency of the future density forecasting performance and the anticipated performance based on past information.

Clements and Hendry (1998) described a forecast failure as a deterioration in forecast performance relative to the anticipated performance. Giacomini and Rossi (2009) generalized the definition to point forecasts using different schemes and developed a formal test for point forecast breakdowns. In this paper, I generalize this definition further, applying it to density forecasts, and develop a formal test to capture density forecast breakdowns. I formalize this idea by defining a density forecast breakdown as a situation in which the out-of-sample density forecast performance, judged by some evaluation function, is significantly worse than its anticipated performance; I further define the scoring surprise as the difference between out-of-sample density forecast performance and average in-sample performance. The evaluation function for density forecast performance is a logarithmic scoring rule that rewards a density forecast that assigns a high probability to the event that actually occurred; this is because evaluating density forecast performance amounts to understanding whether the realized values are indeed low-probability events. The test is built on the idea that, in the absence of a density forecast breakdown, the scoring surprise (i.e., the difference between out-of-sample density forecasting performance and average in-sample performance) should be close to zero.

The density forecast breakdown test statistic is built based on measures of global average scoring surprise, together with an appropriate asymptotic variance estimator required in the test statistic. On this basis, a robust version of the density forecast breakdown test statistic is provided based on measures of the entire time path of the local averaged scoring surprise, considering that useful information may be lost when averaging the global scoring surprise. The (robust) density forecast breakdown test can be applied in a wide range of procedures and is valid under general assumptions. It permits parametric estimation procedures including ordinary least squares (OLS), generalized method of moments (GMM), and maximum likelihood (ML), as well as non-parametric estimation procedures such as the kernel density estimation (KDE). In parametric procedures, the statistics correct parameter estimation uncertainty and allow for dynamic misspecification under both hypotheses. In non-parametric procedures, the estimation uncertainty is asymptotically irrelevant under certain assumptions on relative in-sample and out-of-sample sizes.

The test proposed in this paper is inspired by Giacomini and Rossi (2009). However, this paper is very different in the following three ways: (i) This paper focuses on density forecasting, which constructs estimates of the probability distributions. Taking the simple linear model with normal error terms as an example, constructing density forecasts requires both coefficient estimators and variance estimators, while constructing point forecasts does not require variance estimators. The inclusion of higher moments estimators makes it more complicated to study the asymptotic justification of the test statistic. (ii) In addition, evaluating model-based density forecasts requires more complicated estimation uncertainty correction. In the case of evaluating point forecasts, Giacomini and Rossi (2009) focused on the case in which parameter estimation uncertainty is asymptotically irrelevant, which occurs, for example, in the common situation where the loss function used for estimation is the same as that used for evaluation. However, in the case of evaluating density forecasts, the evaluation function and the loss function used for estimation can never be the same, which leaves the estimation uncertainty correction non-negligible. (iii) The test proposed in this paper can be applied not only to parametric density forecasting models but also to non-parametric models. Unlike the case of parametric procedures, estimation uncertainty cannot be corrected in the case of non-parametric models. Thus, this paper provides conditions under which the estimation uncertainty is asymptotically irrelevant.

Within the literature related to density forecast evaluation, one important strand is specification testing, see Diebold, Gunther and Tay (1998), Bai (2003), Hong and Li (2004), Corradi and Swanson (2006) for testing the correct specification of a parametric density in-sample, see Rossi and Sekhposyan (2017) for testing the correct specification of the out-of-sample predictive density, and see Knppel (2015), Berkowitz (2001), Hong, Li and Zhao (2007), Corradi and Swanson (2006) for additional approaches for assessing the correct calibration of predictive densities. This paper differs from specification testing in that, rather than correct specifications, this paper focuses on the consistency of the future density forecasting performance and the anticipated performance based on the past performances, which presents a representation that is closer to reality. In addition, the evaluation function adopted in this paper is the logarithmic scoring rule, which rewards a density forecast that assigns a high probability to the event that actually occurred, and the test proposed in this paper can be applied to both parametric and non-parametric density forecasting models. Another strand of related literature is that related to structural break testing, see Andrews (1993), Bai and Perron (1998), Hansen (2000), Elliott and Mller (2006), etc. Rather than solving the structural break testing problem, this paper has a different objective, focusing on the stability of the performance of density forecasts instead of the stability of parameters. The test proposed in this paper captures all kinds of changes that affect density forecasting performance, including both parameter changes and other types of changes. Furthermore, the test proposed in this paper, when applied to parametric density forecasting models, allows for dynamic misspecification, and the test can also be applied to non-parametric density forecasting procedures, which does not require the assumption of parametric models.

Monte Carlo simulation examples are provided to analyze the size and power properties of the density forecast breakdown test in finite samples with rolling and recursive schemes, considering parametric density forecasting models and non-parametric density forecasting procedures respectively. The Monte Carlo results indicate that the density forecast breakdown test has good size property in moderately large samples and that the density forecast breakdown test has power against changes in mean and variance, as well as changes in distribution type.

To illustrate the empirical usefulness of the test proposed in this paper, the author investigates whether there are density forecast breakdowns in the density forecasting of real gross domestic product (GDP) growth and of daily SP500 return. The first application is an assessment of the density forecasts of real DGP growth. Future real GDP growth is modeled as a linear function of current and lagged financial and economic conditions, because there has been evidence that the conditional distribution of real GDP growth is closely related to financial conditions, see Adrian, Boyarchenko and Giannone (2017). A simple OLS estimation procedure with a rolling-window scheme, together with a normal density forecasting model, are applied on quarterly real GDP growth data from 1973Q12018Q1. The empirical result based on measures of global scoring surpriseimplies that, for the one-year-ahead density forecasts of real GDP growth, the null hypothesis that there is no density forecast breakdown is rejected at a 5% significance level. Additionally, the empirical resultbased on measures of the entire time path of the local scoring surpriseimplies that the one-year-ahead density forecasts of real GDP growth experienced density forecast breakdowns during 2008Q22010Q2.

The second application is an assessment of the density forecasts of daily SP500 return. Daily SP500 returns, ranging from 01/03/2981 to 12/31/2018, are fitted with a generalized autoregressive conditional heteroskedasticity (GARCH) model with a rolling-window scheme, and the one-day-ahead density forecasts are constructed accordingly, with a normal error-term assumption. In addition to a GARCH model, a non-parametric kernel density estimation and forecasting procedure is also applied. The empirical resultbased on measures of global scoring surpriseimplies that, for the one-day-ahead density forecasts of SP500 return, the null hypothesis of no density forecast breakdown cannot be rejected at a 5% significance level using a GARCH(1,1) model, while the null hypothesis is rejected at a 5% significance level using kernel density estimation and forecasting procedure. This result corresponds to the existence of conditional time-varying volatility, which is not correctly captured by the simple kernel density estimation and forecasting procedure. Additionally, the robust density forecast breakdown test result implies that the breakdown occurs during 07/19/2008 to 11/09/2011.

The remainder of this paper is organized as follows. Section 2 discusses the details of the density forecast breakdown test for both parametric and non-parametric estimation and forecasting procedures, as well as the robust version of the density forecast breakdown test. Section 3 gives the results of the Monte Carlo simulation study. Section 4 applies the (robust) density forecast breakdown test to provide an assessment of the density forecasts of real GDP growth and the SP500 return. Section 5 concludes.

2 Detecting Density Forecast Breakdowns

2.1 Environment

Consider a stochastic process $\{Z_t : \Omega \to \mathbb{R}^{k+1}\}_{t=1}^{T}$ defined on a complete probability space $(\Omega; \mathbb{F}; P)$. The observed vector Z_t is partitioned as $Z_t = (y_t; X_t')'$, where $y_t : \Omega \to \mathbb{R}$ is the variable of interest and $X_t : \Omega \to \mathbb{R}^k$ is a vector of predictors. Let $1 \leq h < \infty$. Suppose we are interested in the true, but unknown, h step-ahead conditional predictive density for the scalar variable y_{t+h} based on $\mathbb{F}_t = \sigma(Z_1'; \ldots; Z_t')'$, which is the true information set available at time t.

Suppose that the researcher has divided the available sample of size T into an in-sample portion of size R and an out-of-sample portion of size P and obtained a sequence of h-step-ahead out-of-sample density forecasts of the variable of interest y_t using the information set $\mathcal{F}_t$, such that $R + P - 1 + h = T$ and $\mathcal{F}_t \subseteq F_t$. Note that this implies that the researcher observes a subset of the true information set, which allows for model misspecification. Which data constitute the in-sample window depends on the forecasting scheme. Three common forecasting schemes are allowed: (1) a fixed forecasting scheme, where the in-sample window includes observations indexed $1, \ldots, R$; (2) a rolling forecasting scheme, where the in-sample window at time t contains observations indexed $t - R + 1, \ldots, t$; and (3) a recursive forecasting scheme observations where the in-sample window includes observations indexed $1, \ldots, t$.

In this context, I am concerned with direct multi-step forecasting, where the predictors are lagged h periods (that is, the model specifies the relationship between y_t and X_{t-h}), and I define a parametric model, say $\phi^y_{t+h}(\cdot|\mathcal{F}_t, \Theta)$ to characterize the h step-ahead conditional predictive density for the scalar variable y_{t+h}. Needless to say, the model would be misspecified, unless $\mathcal{F}_t = F_t$. Considering that the researcher may observe a subset of the true information set ($\mathcal{F}_t \subseteq F_t$), I denote the pseudo-true conditional density of the scalar variable y_{t+h} by $\phi^{y,*}_{t+h}(.) = \phi(.|\mathcal{F}_t, \Theta^*)$, where the parameters in these densities correspond to the probability limits of the estimated parameters. As for the notations for density forecast, let's denote the time-t in-sample density estimate by $\hat{\phi}^{y,e}_t(.) = \phi^e(.|\mathcal{F}_t, \hat{\Theta}_t)$ and denote the time-t density forecast for the scalar variable y_{t+h} by $\hat{\phi}^{y,f}_{t+h}(.) \equiv \phi^f(.|\mathcal{F}_t, \hat{\Theta}_t)$. Then the sequence of P out-of-sample density forecast evaluated at the ex-post realizations be denoted as $\hat{\phi}^{y,f}_{t+h}(y_{t+h}) = \phi^f(y_{t+h}|\mathcal{F}_t, \hat{\Theta}_R), t = R, \ldots, T-h$. The parameter estimator $\hat{\Theta}_t$ depends on the forecasting scheme.[1] In Section 2.5, I extend the test to the case of non-parametric density estimation and forecasting procedures.

As for the evaluation of the density forecast, I adopt the so-called scoring rules. The scoring rules, i.e. loss functions whose arguments are the density forecast and the actual outcome of the variable (see, e.g., Winkler (1967), Diebold and Lopez (1996), Lopez (2001), Amisano and Giacomini (2007)), provide summary measures for the evaluation of probabilistic forecasts, by assigning a numerical score based on the predictive distribution and on the event or value that materializes, see Gneiting and Raftery (2007) for a review and theory of proper scoring rules.

In this context, I adopt the logarithmic scoring rule. The logarithmic scoring rule, proposed by Good (1952), is defined as $S(f, Y) = \log f(Y)$, where Y is the observed value of the variable, and $f(.)$ the density forecast. Intuitively, the logarithmic score rewards a density forecast that assigns a high probability to the event that actually occurred. The logarithmic score is also mathematically convenient, being the only scoring rule that is solely a function of the value of the density at the realization of the variable.

There are several reasons that I adopt the logarithmic scoring rule. (i) The logarithmic score itself is strictly proper, see Gneiting and Raftery (2007) for details.[2] (ii) In this context, the logarithmic scoring rule is preferable to probability integral transform (PIT) in the following perspectives. (a) This paper focuses

[1] For example, $\hat{\Theta}_t = \hat{\Theta}_R$ for $t = R, \ldots, T-h$ for fixed scheme.

[2] Gneiting and Raftery (2007) stated that "A scoring rule is proper if the forecaster maximizes the expected score for an observation drawn from the distribution F if he or she issues the probabilistic forecast F, rather than $G \neq F$. It is strictly proper if the maximum is unique."

on the consistency of the future density forecasting performance and the anticipated performance, rather than the correct specification. Thus the property that the PIT is uniform, independent and identically distributed if the density forecast is correctly specified cannot be made use of in this context. (b) Furthermore, this paper evaluates the out-of-sample performance relative to the in-sample performance, which is similar to Amisano and Giacomini (2007) who adopted the logarithmic scoring rule to compare the relative performance of two forecasting models. (iii) In addition, the logarithmic score is also mathematically convenient.

Thus I denote the scoring for the time-t density forecast for the scalar variable y_{t+h} by

$$S_{t+h}(\hat{\Theta}_t) \equiv S(\hat{\phi}^{y,f}_{t+h}(.), y_{t+h}) = \log \phi^f(y_{t+h}|\mathcal{F}_t, \hat{\Theta}_t)$$

and denote the in-sample scoring for period j, j varying over the in-sample window, by

$$S_j(\hat{\Theta}_t) \equiv S(\hat{\phi}^{y,e}_{j}(.), y_j) = \log \phi^e(y_j|\mathcal{F}_t, \hat{\Theta}_t)$$

For simplicity, I denote them as S_{t+h} and S_j from now on, and use $\cdot^*$ to refer to the counterparts evaluated with the pseudo-true parameter Θ^*.

2.2 Test

I define a ”Density Forecast Breakdown” as a situation where out-of-sample density forecast performance is significantly worsen than the in-sample average perfromance. I formalize this idea by defining a ”scoring surprise” at time $t+h$ as the difference between the out-of-sample scoring at time $t+h$ and the average in-sample scoring:

$$SS_{t+h} = S_{t+h} - \bar{S}_t \tag{1}$$

where $\bar{S}_t$ is the average in-sample scorings computed over the in-sample window implied the forecasting scheme: $\bar{S}_t = \frac{1}{R}\sum_{j=1}^{R} S_j$ for fixed scheme; $\bar{S}_t = \frac{1}{R}\sum_{j=t-R+1}^{t} S_j$ for rolling scheme; $\bar{S}_t = \frac{1}{t}\sum_{j=1}^{t} S_j$ for recursice scheme. Therefore, positive scoring surprise means better performance, while negative scoring surprise means worse performance.

Consider the out-of-sample mean of the scoring surprise:

$$\bar{SS}_{R,P} \equiv \frac{1}{P}\sum_{t=R}^{T-h} SS_{t+h} \tag{2}$$

I propose a test based on the idea that, if the density forecast is reliable, this mean should be close to zero. I construct the following null hypothesis of no density forecast breakdown [3]

$$H_0 : E\left[\frac{1}{P}\sum_{t=R}^{T} SS^*_{t+h}\right] = 0 \tag{3}$$

against the alternative that

$$H_a : E\left[\frac{1}{P}\sum_{t=R}^{T} SS^*_{t+h}\right] < 0 \tag{4}$$

where SS^*_{t+h} is the counterpart of SS_{t+h} evaluated with the pseudo-true parameter Θ^*.

The density forecast breakdown test statistics is

$$t_{R,P,h} = \frac{P^{1/2}\bar{SS}_{R,P}}{\hat{\sigma}_{R,P}} \tag{5}$$

where $\hat{\sigma}^2_{R,P}$ is the asymptotic variance estimator of $P^{1/2}\bar{SS}_{R,P}$.

A level α test rejects the null hypothesis whenever $t_{R,P,h} > z_\alpha$, where z_α is the $1-\alpha$-th quantile of a standard normal distribution. The asymptotic justification for the density forecast breakdown test is provided by Theorem 1.

2.3 Assumptions

I make the following assumptions for asymptotic justification for the density forecast breakdown test:

Assumption P 1. *$\{Z_t\}$ is a mixing sequence with α of size $-r/(r-2)$, $r > 2$.*

Assumption P 2. *(a) S_t is measurable and twice continuously differentiable with respect to Θ; (b) Under H_0, in the neighbourhood $\mathcal{N}$ of Θ^*, there exists a constant $D < \infty$ such that for all t, $sup_{\Theta\in\mathcal{N}}|\frac{\partial^2 S_t}{\partial\Theta\partial\Theta'}| < m_t$, for a measurable m_t such that $E[m_t] < D$.*

[3] The density forecast breakdown test is a one-sided test as a breakdown refers to worse performance. Similarly, if we are testing whether there is a density forecast improvement, then the alternative should be $H_a : E\left[\frac{1}{P}\sum_{t=R}^{T} SS^*_{t+h}\right] > 0$. If one is interested in testing whether there is a density forecast break, either a breakdown or an improvement, then a two-sided test can be constructed with the alternative $H_a : E\left[\frac{1}{P}\sum_{t=R}^{T} SS^*_{t+h}\right] \neq 0$.

Assumption P 3. *Under* H_0, $sup_{t\geq R}\left\|\hat{\Theta}_t-\Theta^*-B_t^*H_t^*\right\|\to^{a.s}0$, *where* $\hat{\Theta}$ *is* $k\times 1$. B_t^* *is a* $k\times q$ *matrix of column* k *such that* $sup_{t\geq 1}|B_t^*|<\infty$. $H_t^*=1/R\sum_{s=1}^{R}h_s^*$*(fixed scheme);* $H_t^*=1/R\sum_{s=t-R+1}^{t}h_s^*$*(rolling scheme);* $H_t^*=1/t\sum_{s=1}^{t}h_s^*$*(recursive scheme); for a* $q\times 1$ *orthogonality condition* h_s^* *such that* $\mathrm{E}[h_s^*]=0$.

Assumption P 4. $sup_{t\geq 1}\|[S_t^*,\partial S_t^*/\partial\Theta,h_t^{*\prime}]'\|^{2r}<\infty$, *where* $\partial S_t^*/\partial\Theta$ *is* $1\times k$.

Assumption P 5. $\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[\frac{\partial S_t^*}{\partial\Theta}\right]<\infty$ *for all* T.

Assumption P 6. $\mathrm{Var}\left(T^{-1/2}\sum_{t=1}^{T}S_t^*\right)>0$ *for all* T *sufficiently large.*

Assumption P 7. $R,P\to\infty$, $\frac{P}{R}\to\pi$, $0\leq\pi<\infty$.

Assumption P 1 restricts the dependence and allows heterogeneity in the data. This assumption follows Assumption A1 in Giacomini and Rossi (2009).

Assumption P 2 is adapted from Assumption A1 in West (1996) and Assumption A2 in Giacomini and Rossi (2009), implying that: (1). $\phi^{e/f}(y|\mathcal{F}_t,\Theta)\neq 0$ for all y; (2). $\phi^{e/f}(y|\mathcal{F}_t,\Theta)$ is twice continuous differentiable w.r.t Θ.[4]

Assumption P 3 follows Assumption A2 West (1996) and Assumption A3 in Giacomini and Rossi (2009), permitting a number of estimation procedures for the model's parameters, including OLS, (quasi-) maximum likelihood and GMM. But the difference lies in that: in point forecast, only coefficient parameters are included (see West (1996), Giacomini and Rossi (2009)); while in density forecast, variance parameter is also included.

For example, for OLS estimation of the parameters in the linear model $Y_s=X_s'\beta^*+\epsilon_s, s=1,\ldots,t$, the estimation uncertainty adopted in point forecast is $\left\|\hat{\beta}_t-\beta^*-B_t^*H_t^*\right\|\to^{a.s}0$ with $B_t^*=(\mathrm{E}(\frac{1}{t}\sum_{s=1}^{t}X_sX_s'))^{-1}$ and $h_s^*=X_s\epsilon_s$; while the density forecast requires the estimation uncertainty of the parameter vector $\Theta=[\beta'\sigma^2]'$. In density forecast,

$$\hat{\Theta}_t-\Theta^*=\begin{bmatrix}\hat{\beta}_t-\beta^*\\ \hat{\sigma}_t^2-\sigma^{*2}\end{bmatrix}=\begin{bmatrix}(\frac{1}{t}\sum_{s=1}^{t}X_sX_s')^{-1} & 0\\ -\frac{1}{t}\sum_{s=1}^{t}\epsilon_sX_s'(\frac{1}{t}\sum_{s=1}^{t}X_sX_s')^{-1} & 1\end{bmatrix}\begin{bmatrix}\frac{1}{t}\sum_{s=1}^{t}X_s\epsilon_s'\\ \frac{1}{t}\sum_{s=1}^{t}\epsilon_s\epsilon_s'-\sigma^{*2}\end{bmatrix}=\bar{B}_tH_t^*$$

[4]The reason is that, for a parametric model, the second derivative of $S_t=\log\phi(y_t|\mathcal{F}_t,\Theta)$ with respect to Θ is $\frac{\partial^2 S_t}{\partial\Theta\partial\Theta'}=\frac{\frac{\partial^2\phi(y_t|\mathcal{F}_t,\Theta)}{\partial\Theta\partial\Theta'}\cdot\phi(y_t|\mathcal{F}_t,\Theta)-(\partial\phi(y_t|\mathcal{F}_t,\Theta)/\partial\Theta)^2}{\phi(y_t|\mathcal{F}_t,\Theta)^2}$.

and it requires that $B_t^* = \begin{bmatrix} (E[\frac{1}{t}\sum_{s=1}^{t} X_s X_s'])^{-1} & 0 \\ 0 & 1 \end{bmatrix}$ and $h_s^* = \begin{bmatrix} X_s \epsilon_s' \\ \epsilon_s \epsilon_s' - \sigma^{*2} \end{bmatrix}$. [5]

Assumption P 5 is relaxed from Assumption A5 in Giacomini and Rossi (2009). In point forecast case, estimation uncertainty can be irrelevant if loss function is the same for estimation and evaluation.[6] But in density forecast case, $\frac{\partial S_t^*}{\partial \Theta} = \frac{\partial \phi(y_t|\Theta^*)/\partial \Theta}{\phi(y_t|\Theta^*)} \neq 0$, so estimation uncertainty cannot be omitted. Assumption P 5 ensures that the estimation uncertainty component converges.

Assumption P 7 follows Assumption A7 in Giacomini and Rossi (2009). It ensures that the test statistic has an asymptotically normal distribution when the in-sample and out-of-sample sizes go to infinity at the same rate, or that the in-sample size grows faster than the out-of-sample size.

2.4 Asymptotic variance estimator and main results

This section shows how to construct a valid asymptotic variance estimator for the density forecast breakdown test statistics.

The intuition is that the average scoring surprise is a weighted average of in-sample and out-of-sample scoring, with weights depending on R, P and on forecasting schemes, and that the average estimation uncertainty component is a weighted average of in-sample and out-of-sample $h_t(\Theta^*)$ as defined in Assumption P 3, with weights depending on R, P and on forecasting scheme and computed using B_t^* as defined in Assumption P 3. Then the asymptotic variance estimator is simply a (rescaled) HAC estimator of the variance and covariance of these two weighted averages.

For simplification, I give the algorithms for the fixed scheme in detail. Algorithms for rolling and recursive schemes are provided in detail in the Appendix A.2.

Algorithm 1 *Construct the following: (1) a $1 \times T$ vector S^* of in-sample and out-of-sample scorings, with element $S_t^*, t = 1, \ldots, T$, and its counterpart S with element $S_t, t = 1, \ldots, T$:*

$$S^* \equiv \Big[\underbrace{S_1^*(\Theta^*), \ldots, S_R^*(\Theta^*)}_{R}, \underbrace{S_{R+1}(\Theta^*), \ldots, S_{R+h-1}(\Theta^*)}_{h-1}, \underbrace{S_{R+h}^*(\Theta^*), \ldots, S_T^*(\Theta^*)}_{P} \Big]$$

[5] $\bar{B}_t = \begin{bmatrix} (\frac{1}{t}\sum_{s=1}^{t} X_s X_s')^{-1} & 0 \\ -\frac{1}{t}\sum_{s=1}^{t} \epsilon_s X_s' (\frac{1}{t}\sum_{s=1}^{t} X_s X_s')^{-1} & 1 \end{bmatrix} \to \begin{bmatrix} (E[\frac{1}{t}\sum_{s=1}^{t} X_s X_s'])^{-1} & 0 \\ 0 & 1 \end{bmatrix} = B_t^*$

[6] If loss function is the same for estimation and evaluation, then $\mathrm{E}[\frac{\partial L_t(\beta^*)}{\partial \beta}] = 0$. It is shown in Giacomini and Rossi (2009) that estimation uncertainty is irrelevant if this assumption holds.

$$S \equiv \left[\underbrace{S_1(\hat{\Theta}_R), \ldots, S_R(\hat{\Theta}_R)}_{R}, \underbrace{S_{R+1}(\hat{\Theta}_R), \ldots, S_{R+h-1}(\hat{\Theta}_R)}_{h-1}, \underbrace{S_{R+h}(\hat{\Theta}_R), \ldots, S_T(\hat{\Theta}_R)}_{P} \right]$$

and the corresponding vector $\tilde{S}$ of demeaned scorings, where $\tilde{S}_t \equiv S_t - T^{-1} \sum_{j=1}^{T} S_j$; (2) a $1 \times T$ vector of weights, depending on forecasting scheme, with elements $\omega_t^S, t = 1, \ldots, T$:

$$\omega^S \equiv \left[\underbrace{-\frac{P}{R}, \ldots, -\frac{P}{R}}_{R}, \underbrace{0, \ldots, 0}_{h-1}, \underbrace{1, \ldots, 1}_{P} \right]$$

so that

$$\sum_{t=1}^{T} \omega_t^S S_t^* = \sum_{t=R}^{T} SS_{t+h}^*$$

Algorithm 2 *Construct the following: (1) a $(k+1) \times T$ vector of in-sample and out-of-sample h^*, as defined in Assumption P 3, with element $h_t^*, t = 1, \ldots, T$, and its counterpart h with element $h_t, t = 1, \ldots, T$:*

$$h^* \equiv \left[\underbrace{h_1^*(\Theta^*), \ldots, h_R^*(\Theta^*)}_{R}, \underbrace{h_{R+1}^*(\Theta^*), \ldots, h_{T-h}^*(\Theta^*)}_{P-1}, \underbrace{O, \ldots, O}_{h} \right]$$

$$h \equiv \left[\underbrace{h_1(\hat{\Theta}_R), \ldots, h_R(\hat{\Theta}_R)}_{R}, \underbrace{h_{R+1}(\hat{\Theta}_R), \ldots, h_{T-h}(\hat{\Theta}_R)}_{P-1}, \underbrace{O, \ldots, O}_{h} \right]$$

(2) a $1 \times T$ vector of weights ω^{h*}, depending on forecasting scheme, with elements $\omega_t^{h*}, t = 1, \ldots, T$ and its counterpart ω^h with elements $\omega_t^h, t = 1, \ldots, T$:[7]

$$\omega^{h*} \equiv \left[\underbrace{-\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_R^*}{R}, \ldots, -\frac{\sum_{t=R}^{T-h} D_{t+h}^* B_R^*}{R}}_{R}, \underbrace{O, \ldots, O}_{T-R} \right] \qquad D_{t+h}^* = \frac{\partial SS_{t+h}^*(\Theta^*)}{\partial \Theta}$$

$$\omega^{h} \equiv \left[\underbrace{-\frac{\sum_{t=R}^{T-h} D_{t+h} B_R}{R}, \ldots, -\frac{\sum_{t=R}^{T-h} D_{t+h} B_R}{R}}_{R}, \underbrace{O, \ldots, O}_{T-R} \right] \qquad D_{t+h} = \frac{\partial SS_{t+h}(\hat{\Theta}_R)}{\partial \Theta}$$

[7] B_t is the consistent estimate of B_t^*.

so that

$$\sum_{t=1}^{T}\omega_t^{h*}h_t^* = \sum_{t=R}^{T}D_{t+h}^*B_t^*H_t^*$$

Then the asymptotic variance estimator is simply a (rescaled) HAC estimator of the variance and covariance of these two weighted averages:

$$\hat{\sigma}^2_{R,P} = \frac{T}{P}(V_T^{SS} + V_T^{hh} + 2V_T^{Sh}) \tag{6}$$

with V_T^{SS},V_T^{hh},V_T^{Sh} defined as follows:

$$\begin{aligned}
V_T &= \begin{bmatrix} V_T^{SS} & V_T^{Sh} \\ V_T^{Sh} & V_T^{hh} \end{bmatrix} \\
V_T^{SS} &= \frac{1}{T}\sum_{t=1}^{T}(\omega_t^S\tilde{S}_t)^2 + \frac{2}{T}\sum_{j=1}^{p_T}v_{T,j}\sum_{t=j}^{T}\omega_t^S\tilde{S}_t\omega_{t-j}^S\tilde{S}_{t-j} \\
V_T^{hh} &= \frac{1}{T}\sum_{t=1}^{T}\omega_t^h h_t h_t'\omega_t^{h\prime} + \frac{2}{T}\sum_{j=1}^{p_T}v_{T,j}\sum_{t=j}^{T}(\omega_t^h h_t h_{t-j}'\omega_{t-j}^{h\prime} + \omega_{t-j}^h h_{t-j}h_t'\omega_t^{h\prime}) \\
V_T^{Sh} &= \frac{1}{T}\sum_{t=1}^{T}\omega_t^S\tilde{S}_t h_t'\omega_t^{h\prime} + \frac{2}{T}\sum_{j=1}^{p_T}v_{T,j}\sum_{t=j}^{T}(\omega_t^S\tilde{S}_t h_{t-j}'\omega_{t-j}^{h\prime} + \omega_{t-j}^S\tilde{S}_{t-j}h_t'\omega_t^{h\prime})
\end{aligned} \tag{7}$$

where bandwidth p_T and weights $v_{T,j}$ are appropriately chosen (as Newey and West (1987)). The use of a HAC estimator for the asymptotic variance is motivated by the possible presence of serial correlation in the scorings and the estimation uncertainty.

Theorem 1. ***(Generalization of Asymptotic Justification of the Density Forecast Breakdown Test)*** *Given Assumption P 1-7, if V_T defined in eq (7) is p.d., then under H_0 in (3), $t_{R,P,h} \xrightarrow{d} \mathcal{N}(0,1)$, where $t_{R,P,h}$ is defined in* (5) *and $\hat{\sigma}^2_{R,P}$ defined in* (6)*.*

2.5 An extension to non-parametric models

In this section, I extend the test to the case of non-parametric method. I use $\phi_{t+h}^y(\cdot|\mathcal{F}_t)$ to characterize the h step-ahead conditional predictive density for the scalar variable y_{t+h}, not imposing any parametric form. As for the notations for density forecast, let's denote the time-t in-sample density estimate by $\hat{\phi}_t^{y,e}(.) = \phi^e(.|\mathcal{F}_t)$ and denote the time-t density forecast for the scalar variable y_{t+h} by $\hat{\phi}_{t+h}^{y,f}(.) \equiv \phi^f(.|\mathcal{F}_t)$.

Then the sequence of P out-of-sample density forecast evaluated at the ex-post realizations be denoted as $\hat{\phi}^{y,f}_{t+h}(y_{t+h}) = \phi^f(y_{t+h}|\mathcal{F}_t), t = R, \ldots, T-h$. Considering that the researcher may observe a subset of the true information set ($\mathcal{F}_t \subseteq F_t$), I denote the pseudo-true conditional density of the scalar variable y_{t+h} by $\phi^{y,*}_{t+h}(.) = \phi(.|\mathcal{F}_t)$.

Unlike the parametric case where the scoring is a function of the parameter vector Θ, with non-parametric density estimation and forecasting, the scoring for the time-t density forecast for the scalar variable y_{t+h} is a function of the estimated density function which follows

$$S_{t+h}(\hat{\phi}_t(.)) \equiv S(\hat{\phi}^{y,f}_{t+h}(.), y_{t+h}) = \log \phi^f(y_{t+h}|\mathcal{F}_t)$$

and denote the in-sample scoring for period j, j varying over the in-sample window, by

$$S_j(\hat{\phi}_t(.)) \equiv S(\hat{\phi}^{y,e}_j(.), y_j) = \log \phi^e(y_j|\mathcal{F}_t)$$

For simplicity, I denote them as S_{t+h} and S_j from now on, and use $\cdot^*$ to refer to the counterparts evaluated with the pseudo-true density $\phi^{y,*}(.)$.

A fundamental difference between the parametric case and the non-parametric case is that one cannot correct for estimation uncertainty in the non-parametric case. Here, I propose the assumptions for the non-parametric estimation and forecasting procedures to ensure that the estimation uncertainty component is asymptotically irrelevant.

Assumption NP 1. *(a) The Hardamard Derivative of the scoring $S_t(\phi^*)$ exist, denoted as D^*_ϕ, and $sup_{t\geq 1} \left\| D^*_\phi \right\| < \infty$. (b) $sup_{t\geq 1} \|S_t(\phi^*)\|^{2r} < \infty$. (c) $\text{Var}\left(T^{-1/2}\sum_{t=1}^{T} S_t(\phi^*)\right) > 0$ for all T sufficiently large.*

Assumption NP 2. $\sup_y |\hat{\phi}_t(y) - \phi^*(y)| = Op(R^{\alpha_\phi}), \forall t = R, \ldots, T$. *$R, P \to \infty$, $\sqrt{P}R^{\alpha_\phi} \to 0$.*

Assumption NP 1 is an assumption similar to Assumptions P 2, P 4, P 5 and P 6 in the parametric case. Apart from the regular conditions, it ensures that the functional derivative of the scoring exists and is bounded.

Assumption NP 2 can be replaced with more detailed assumptions when adopting different non-parametric estimation methods. Below is a list of detailed assumptions for the KDE case, which is a very commonly used non-parametric estimation method.

Example (kernel density estimation and forecasting): Suppose we are interested in the true, but unknown, h step-ahead unconditional predictive density for the scalar variable y_{t+h} based on $\mathbb{F}_t = \sigma(y_1'; \ldots; y_t')'$, which is the true information set available at time t. At each period t, I adopt kernel density estimation with kernel function $K(.)$ to get the density estimator $\hat{\phi}_t^y(.)$[8] and also use it as the density forecast for y_{t+h}. I make the following assumptions:

Assumption NP-kernel 1. *$\{y_t\}$ is i.i.d, and have a three-time differentiable pdf $\phi(y)$ with $\inf_y \phi(y) \geq \delta > 0$.*

Assumption NP-kernel 2. *The kernel function $K(.)$ is bounded and satisfies: (a) $\int K(u)du = 1$. (b) $K(u) = K(-u)$. (c) $\int u^2 K(u)du = \kappa_2 > 0$.*

Assumption NP-kernel 3. *The estimation window size R and the corresponding bandwidth parameter h_R satisfy that, as $R \to \infty$, $h_R \to 0$ and $Rh_R \to \infty$.*

Assumption NP-kernel 4. *$R, P \to \infty$, $\frac{P \ln R}{Rh_R} \to 0$ and $Ph_R^4 \to 0$.*

Assumption NP-kernel 5. *(a) The Hardamard Derivative of the scoring $S_t(\phi^*)$ exist, denoted as D_ϕ^*, and $sup_{t\geq 1} \left\| D_\phi^* \right\| < \infty$. (b) $sup_{t\geq 1} \left\| S_t(\phi^*) \right\|^{2r} < \infty$. (c) $\mathrm{Var}\left(T^{-1/2} \sum_{t=1}^{T} S_t(\phi^*) \right) > 0$ for all T sufficiently large.*

Assumption NP-kernel 1 can be relaxed to allow the data to be mixing, because time series dependence has no effect on the asymptotic bias and variance of the kernel estimator as KDE is averaging the data locally in the y-dimension, where there is no time-series dependence.

Assumption NP-kernel 2 and NP-kernel 3 are standard assumptions for kernel function and bandwidth parameter respectively, which ensure the consistency of the estimated kernel density estimator.

Assumption NP-kernel 5 is required to ensure the functional analysis of the average density estimation uncertainty component, which converges to 0 asymptotically. It works similarly as Assumption P 4, P 5 and P 6 in the parametric case.

Theorem 2. *(**Generalization of Asymptotic Justification of the Density Forecast Breakdown Test in Non-parametric Settings**)*

Given Assumption NP 1 and NP 2, if V_T defined below is p.d., then under H_0 in (3), $t_{R,P,h} \xrightarrow{d} \mathcal{N}(0,1)$, where $t_{R,P,h}$ is defined in (5) *and $\hat{\sigma}_{R,P}^2$ is defined as*

[8] With kernel density estimation and forecasting procedure, $\hat{\phi}_t^{y,e}(.)$ and $\hat{\phi}_t^{y,f}(.)$ should take the same expression.

$$\hat{\sigma}^2_{R,P} = \frac{T}{P}\left\{\frac{1}{T}\sum_{t=1}^{T}(\omega_t^S\tilde{S}_t)^2 + \frac{2}{T}\sum_{j=1}^{p_T}\upsilon_{T,j}\sum_{t=j}^{T}\omega_t^S\tilde{S}_t\omega_{t-j}^S\tilde{S}_{t-j}\right\}$$

where bandwidth p_T and weights $\upsilon_{T,j}$ are appropriately chosen (as Newey and West (1987)).

2.6 A robust version of density forecast breakdown test

The density forecast breakdown test discussed above focuses on the global performance of the density forecast model. However, the relative out-of-sample performance, compared with the in-sample performance, may itself be time-varying. Thus, on the basis of the density forecast breakdown test, a robust version of the density forecast breakdown test statistic is provided based on measures of the entire time path of the local averaged scoring surprise, considering that useful information may be lost when averaging the global scoring surprise. Giacomini and Rossi (2010) constructed the Fluctuation test for forecast comparisons in unstable environment. The robust density forecast breakdown test is actually a Fluctuation test (Giacomini and Rossi (2010)) applied on the density forecast breakdown test, i.e. the density forecast breakdown test statistics computed over the rolling out-of-sample windows of size M (M is defined formally below).

The robust density forecast breakdown test statistic is built on measures of the local averaged scoring surprise, and tests whether it equals zero at each point in time. Against to the null hypothesis in (3) for the density forecast breakdown test, it focuses on the robust null hypothesis H_0^R:

$$H_0^R : \mathrm{E}\left[SS^*_{t+h}\right] = 0, \forall t = R, \ldots, T-h \tag{8}$$

Define the local out-of-sample mean of the scoring surprise:

$$\bar{S}S_{R,M,t} \equiv \frac{1}{M}\sum_{j=t-M/2}^{t+M/2-1} SS_j, \qquad t = R + M/2, \ldots, T - M/2 + 1 \tag{9}$$

And I make the following assumptions:

Assumption R 1. $\{P^{-1/2}\sum_{t=R}^{R+[\tau P]} SS^*_{t+h}\}$ *obeys a Functional Central Limited theorem.*

Assumption R 2. $M/P \rightarrow \mu \in (0, \infty)$ *as* $M \rightarrow \infty$*,* $P \rightarrow \infty$*.*

The assumptions differ from Giacomini and Rossi (2010) in the following ways: (i) The null hypothesis of the robust density forecast breakdown test focuses on the (pseudo) true parameter/density, with the estimation uncertainty corrected through the appropriate asymptotic variance estimator, while Giacomini and Rossi (2010) followed the framework of Giacomini and White (2006), reserving the estimates under both hypothesis (ii) The robust density forecast breakdown test requires that the in-sample size $R \to \infty$ as it corrects for estimation uncertainty, while Giacomini and Rossi (2010) assumes that $R < \infty$.

The following proposition describes the asymptotic justification of the robust density forecast breakdown test.

Proposition 1. *(The Robust Density Forecast Breakdown Test) Suppose the assumptions for the density forecast breakdown tests hold. And suppose Assumption R 1 and R 2 hold.*

Let

$$\mathcal{F}_{t,R,M} = \frac{M^{-1/2} \sum_{j=t-M/2}^{t+M/2-1} SS_j}{\hat{\sigma}_R}, \qquad t = R + M/2, \ldots, T - M/2 + 1 \tag{10}$$

where $\hat{\sigma}_R^2$ is the asymptotic variance estimator of $P^{1/2}\bar{SS}_R$.

Under the null hypothesis $H_0 : \mathrm{E}\left[SS_{t+h}^*\right] = 0$ for all $t = R, \ldots, T-h$:

$$\mathcal{F}_{t,R,M} \Rightarrow [\mathcal{B}(\tau + \mu/2) - \mathcal{B}(\tau - \mu/2)]/\sqrt{\mu} \tag{11}$$

where $t = [\tau P]$, $M = [\mu P]$ and $\mathcal{B}(\cdot)$ is a standard univariate Brownian motion. Then,

$$\sup_t \mathcal{F}_{t,R,M} \xrightarrow{d} \sup_\tau [\mathcal{B}(\tau + \mu/2) - \mathcal{B}(\tau - \mu/2)]/\sqrt{\mu} \tag{12}$$

The critical values for a significance level α are κ_α, where κ_α solves

$$Pr\left[\sup_\tau |[\mathcal{B}(\tau + \mu/2) - \mathcal{B}(\tau - \mu/2)]/\sqrt{\mu}| > \kappa_\alpha\right] = \alpha \tag{13}$$

Note that the robust density forecast breakdown test can still be applied for both parametric and non-parametric density estimation and forecasting procedures, with corresponding SS defined in section 2.2 and 2.5.

3 Monte Carlo Evidence

In this section, I analyze the size and power properties of the density forecast breakdown test in finite samples, considering both parametric and non-parametric estimation/forecasting methods.

3.1 Size properties

3.1.1 Simulation design

This section describes the design of the simulation study of the size properties of the density forecast breakdown test. To investigate the different forecasting procedures, I consider the following two scenarios with the data generating process (DGP) and different estimation and forecasting procedures described in detail below.

(i) **Scenario S1: DGP with parametric model:** I generate the data from the DGP described below, and adopt a parametric estimation and forecasting model.

(ii) **Scenario S2: DGP with non-parametric model:** I generate the data from the DGP described below, and adopt a non-parametric estimation and forecasting model.

The DGP considered in the size study is

$$y_t = \beta_0^* + x_t\beta_1^* + \epsilon_t, \qquad \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^{*2}) \tag{14}$$

for $t = 1, \ldots, T$, where y_t, x_t are scalars, and consider that $\beta_0^* = 0$, $\beta_1^* = 0$, $\sigma^{*2} = 1$, and i.i.d. regressors and errors: $x_t, \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, 1)$, independent of each other.

Suppose I am interested in the h-step ahead density forecast of y_{t+h}, and I consider the following parametric and non-parametric estimation and forecasting procedures.

For the parametric case, I apply the OLS estimation on the following regression at each period with the corresponding in-sample observations:

$$y_t = \beta_0^* + x_t\beta_1^* + \epsilon_t$$

As for forecasting, given the OLS estimator $\hat{\beta}_{0,t}$, $\hat{\beta}_{1,t}$ and $\hat{\sigma}_t$ at each period t, I assume normality of ϵ_t, and the h-step ahead density forecast of y_{t+h} is

$$\hat{\phi}^f_{t+h}(y) = \phi(y|\hat{\beta}_{0,t}, \hat{\beta}_{1,t}, \hat{\sigma}^2_t) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2_t}} e^{-\frac{(y-\hat{\beta}_{0,t}-x_t\hat{\beta}_{1,t})^2}{2\hat{\sigma}^2_t}}$$

For the non-parametric case, I apply the kernel density estimation (KDE) at each period with the corresponding in-sample observations:

$$\hat{\phi}^e_t(y) = \frac{1}{Rh_R}\sum_{t=1}^{R}\kappa(\frac{y_t - y}{R}), \qquad h_R = cR^{-1/5}$$

where h_R is the bandwidth parameter.

As for forecasting, the h-step density forecast is the same as the density estimator at period t:

$$\hat{\phi}^f_{t+h}(y) = \hat{\phi}^e_t(y)$$

For the in-sample and out-of-sample sizes (R, P), I consider different combinations given by $R, P \in \{50, 100, 200, 400\}$. For each pair (R, P), I conduct 5000 Monte Carlo replications, and generate $T = R + P$ data as in (14).

I consider the density forecast breakdown test for the rolling window scheme and recursive window scheme, using the general asymptotic variance estimator of Theorem 1 and 2, with truncation lags for the HAC estimator set to be 0.

3.1.2 Simulation result

Table 1 and 2 give the rejection frequencies of the density forecast breakdown test for various (R, P) pairs for **Scenario S1** and **Scenario S2** respectively. The results imply that the density forecast breakdown test has good size properties for large in-sample and out-of-sample sizes in moderate large samples for both parametric and non-parametric estimation and forecasting procedures.

Table 1: Size of density forecast breakdown test, nominal size 0.05, Scenario S1

R	P	Rolling	Recursive
50	50	0.0568	0.0894
50	100	0.0412	0.0774
50	200	0.0338	0.0658
50	400	0.0220	0.0610
100	50	0.0792	0.0768
100	100	0.0636	0.0698
100	200	0.0496	0.0548
100	400	0.0404	0.0548
200	50	0.0752	0.0752
200	100	0.0646	0.0686
200	200	0.0570	0.0590
200	400	0.0560	0.060
400	50	0.0792	0.0770
400	100	0.0634	0.0632
400	200	0.0576	0.0588
400	400	0.0538	0.0546

Note: This table reports rejection frequencies over 5000 Monte Carlo replications of the density forecast breakdown test, using the general asymptotic variance estimator of Theorem 1, implemented with rolling and recursive schemes. R and P denote the in-sample and out-of-sample sizes, respectively.

3.2 Power properties

3.2.1 Simulation design

This section describes the design of the simulation study of the power properties of the density forecast breakdown test. I consider the following scenarios with various sources of density forecast breakdowns (DGP-P1, GDP-P2, DGP-P3) and different estimation and forecasting procedures described in detail below.

(i) **Scenario P1: DGP-P1 with parametric model:** I assume there is a change in the mean and generate data from the DGP-P1 described below, and adopt a parametric estimation and forecasting model.

(ii) **Scenario P2: DGP-P2 with parametric model:** I assume there is a change in the variance and generate data from the DGP-P2 described below, and adopt a non-parametric estimation and forecasting model.

(iii) **Scenario P3: DGP-P3 with parametric model:** I assume there is a change in the distribution

Table 2: Size of density forecast breakdown test, nominal size 0.05, Scenario S2

R	P	Rolling	Recursive
50	50	0.0682	0.0738
50	100	0.0720	0.0644
50	200	0.0938	0.0706
50	400	0.1226	0.0736
100	50	0.0590	0.0624
100	100	0.0580	0.0542
100	200	0.0692	0.0586
100	400	0.0632	0.0614
200	50	0.0658	0.0692
200	100	0.0542	0.0570
200	200	0.0578	0.0534
200	400	0.0554	0.0512
400	50	0.0676	0.0678
400	100	0.0552	0.0572
400	200	0.0586	0.0568
400	400	0.0554	0.0534

Note: This table reports rejection frequencies over 5000 Monte Carlo replications of the density forecast breakdown test, using the general asymptotic variance estimator of Theorem 2, implemented with rolling and recursive schemes. R and P denote the in-sample and out-of-sample sizes, respectively.

type and generate data from the DGP-P3 described below, and adopt a parametric estimation and forecasting model.

(iv) **Scenario P4: DGP-P1 with non-parametric model:** I assume there is a change in the mean and generate data from the DGP-P1 described below, and adopt a non-parametric estimation and forecasting model.

(v) **Scenario P5: DGP-P2 with non-parametric model:** I assume there is a change in the variance and generate data from the DGP-p2 described below, and adopt a parametric estimation and forecasting model.

(vi) **Scenario P6: DGP-P3 with non-parametric model:** I assume there is a change in the distribution type and generate data from the DGP-P3 described below, and adopt a non-parametric estimation and forecasting model.

DGP-P1: Change in mean

Consider a one-time change in mean in DGP:

$$y_t = \alpha^{\beta} \cdot 1(t > T_b) + \epsilon_t, \qquad \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1) \tag{15}$$

Let $(R, P) = (200, 100)$, and let $T_b = R + 0.15P$. α^{β} controls the size of the change in mean. The power curve is obtained by letting α^{β} vary between 0 and 2 and considering 5000 Monte Carlo replications for both rolling and resursive shcemes.

DGP-P2: Change in variance

Consider a one-time change in variance in DGP:

$$\begin{aligned} y_t &= \epsilon_t, \qquad \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_t^{*2}) \\ \sigma_t^{*2} &= 1 + \alpha^{\sigma} \cdot 1(t > T_b) \end{aligned} \tag{16}$$

Let $(R, P) = (200, 100)$, and let $T_b = R + 0.15P$. α^{σ} controls the size of the change in variance. The power curve is obtained by letting α^{σ} vary between 0 and 2.5 and considering 5000 Monte Carlo replications for both rolling and resursive shcemes.

DGP-P3: Change in distribution type

Consider a one-time change in distribution type in DGP:

$$\begin{aligned} & y_t = \epsilon_t \\ & \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_t^{*2}), \qquad t = 1, \ldots, T_b \\ & \epsilon_t \overset{i.i.d}{\sim} Logistic(\mu^*, s^*), \quad t = T_b + 1, \ldots, T \end{aligned} \tag{17}$$

Let $(R, P) = (200, 100)$, and let $T_b = R + 0.15P$. The parameter μ^* controls the location of the density and the parameter s^* controls the scaling of tails. The location parameter μ^* is set to be 0, and the power curve is obtained by letting s^* vary between 0.5 to 1.5 and considering 5000 Monte Carlo replications for both rolling and resursive shcemes.

Suppose I am interested in the h-step ahead density forecast of y_{t+h}, and I consider the following parametric and non-parametric estimation and forecasting procedures.

For the parametric case, I apply the OLS estimation on the following regression at each period with the corresponding in-sample observations:

$$y_t = \beta_0^* + x_t \beta_1^* + \epsilon_t$$

As for forecasting, given the OLS estimator $\hat{\beta}_{0,t}$, $\hat{\beta}_{1,t}$ and $\hat{\sigma}_t$ at each period t, I assume normality of ϵ_t, and the h-step ahead density forecast of y_{t+h} is

$$\hat{\phi}^f_{t+h}(y) = \phi(y|\hat{\beta}_{0,t}, \hat{\beta}_{1,t}, \hat{\sigma}_t^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} e^{-\frac{(y-\hat{\beta}_{0,t}-x_t\hat{\beta}_{1,t})^2}{2\hat{\sigma}_t^2}}$$

For the non-parametric case, I apply the kernel density estimation (KDE) at each period with the corresponding in-sample observations:

$$\hat{\phi}^e_t(y) = \frac{1}{Rh_R}\sum_{t=1}^{R} \kappa(\frac{y_t - y}{R}), \qquad h_R = cR^{-1/5}$$

where h_R is the bandwidth parameter.

As for forecasting, the h-step density forecast is the same as the density estimator at period t:

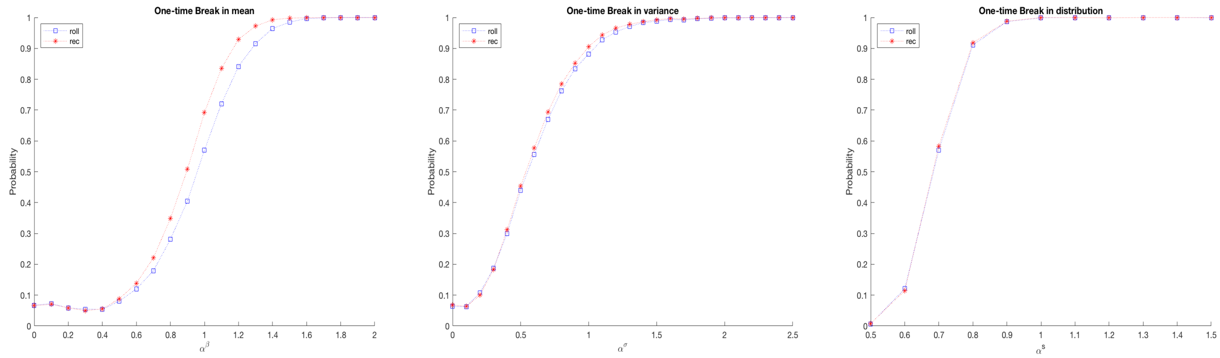
$$\hat{\phi}^f_{t+h}(y) = \hat{\phi}^e_t(y)$$

For the in-sample and out-of-sample sizes, I consider let $(R, P) = (200, 100)$. And I conduct 5000 Monte Carlo replications, and generate $T = R + P$ data in each GDP.

I consider the density forecast breakdown test for the rolling window scheme and recursive window scheme, using the general asymptotic variance estimator of Theorem 1 and 2, with truncation lags for the HAC estimator set to be 0.

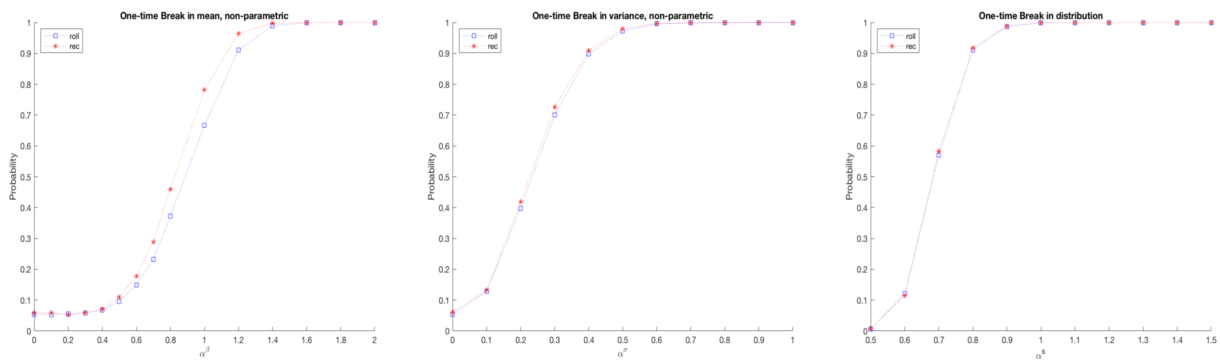
3.2.2 Simulation result

Figure 1(a), 1(b) and 1(c) show the power curves of the density forecast breakdown test with a parametric estimation and forecasting model described in **Scenario P1, Scenario P2 and Scenario P3**. Similarly, Figure 2(a), 2(b) and 2(c) show the power curves of the density forecast breakdown test with a non-parametric estimation and forecasting model described in **Scenario P4, Scenario P5 and Scenario P6**. These power curves imply that the density forecast breakdown test has power against changes in mean, variance and distribution type, with both parametric and non-parametric estimation and forecasting procedures.



(a) A one-time change in mean (b) A one-time change in variance (c) A one-time change in distribution type

Figure 1: Power function, Scenario 1,2,3



(a) A one-time change in mean (b) A one-time change in variance (c) A one-time change in distribution type

Figure 2: Power function, Scenario 4,5,6

4 Empirical Analysis

To illustrate the empirical usefulness of the test proposed in this paper, I investigate whether there are density forecast breakdowns in the density forecasting of real gross domestic product (GDP) growth and of daily stock price return.

4.1 Assessment of density forecasting procedure of real GDP growth

This section provides an empirical assessment of the density forecasting procedure of real GDP growth based on a parametric model. I model the future real GDP growth as a function of current financial and economic conditions, since there has been evidence that conditional distribution of GDP growth is closely related to financial conditions, see Adrian, Boyarchenko and Giannone (2016).

I use real GDP data from the Bureau of Economic Analysis (BEA) to compute real GDP growth.[9] The real GDP growth is calculated as the percentage change from the same quarter one year ago. And I also use it for economics condition measurement. To gauge the financial conditions, I use the National Financial Conditions Index (NFCI)[10], which is a weighted average of 105 measures of financial activity, each expressed relative to their sample averages and scaled by their sample standard deviations. The NFCI provides a weekly estimate of U.S. financial conditions in money markets, debt and equity markets, and the traditional and shadow banking systems. When the NFCI is positive, financial conditions are tighter than average. Brave and Butters (2012) show that the National Financial Conditions Index (NFCI) is a highly predictive and robust indicator of financial stress at leading horizons of up to one year. The NFCI is converted into the quarterly frequency by averaging the weekly observations within each quarter.[11] The data covers 181 periods, ranging from 1973Q1 to 2018Q1.

The model used for estimation and forecasting is as follows:

$$
\begin{aligned}
y_{t+h} &= c + \beta^y(L)y_t + \beta^f(L)f_t + \epsilon_{t+h}, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2) \\
\beta^y(L) &= \beta_0^y + \beta_1^y L + \cdots + \beta_p^y L^p \\
\beta^f(L) &= \beta_0^f + \beta_1^f L + \cdots + \beta_q^f L^q
\end{aligned} \tag{18}
$$

where y_t refer to the real dgp growth at period t and f_t refer to the financial condition at period t, and h refers to the density forecast horizons.

I cut the whole sample in half and set the in-sample window size to be $R = 90$, that is from 1973Q1 to 1995Q2. For the horizons, I consider a 1-quarter-ahead density forecast ($h = 1$) and a 1-year-ahead density forecast ($h = 4$). I apply the rolling window scheme and consider different combinations of lags

[9] Downloaded from FRED.

[10] The NFCI is computed by the Federal Reserve Bank of Chicago.

[11] For the attribution of weeks to overlapping quarters I follow the convention of Federal Reserve Economic Data (FRED) Economic Data, which is the source of our data. Weeks that start in one quarter and end in the next one are fully assigned to the latter quarter.

included in (18), as well as the numbers of lags chosen by BIC. And I set the truncation lags for HAC estimation to be $T^{1/3}$.

Table 3 displays the result of a density forecast breakdown (one-sided) test. I find that the null hypothesis that there is no density forecast breakdown cannot be rejected for the 1-quarter-ahead density forecast. But for the 1-year-ahead density forecast, we can reject, for most of the cases at a 5% significance level, the null hypothesis that there is a density forecast breakdown.

Table 3: Results of density forecast breakdown test

		$h = 1$: 1-quarter horison	$h = 4$: 1-year horison
p	q	p-value	p-value
0	0	0.2970	0.0505
0	3	0.2775	0.0401
3	0	0.4686	0.0328
3	3	0.4858	0.0456
BIC	BIC	0.4989	0.0324

Note: The table reports the p-values of the density forecast breakdown (one-sided) test. I use a rolling window scheme with $T = 181$ and $R = 90$ for real-time data. The density forecast horizons are $h = 1, 4$ quarters. p and q are the numbers of lags included for economic condition and financial consition respectively. For example, $p = 0$ and $q = 0$ refer to the case where I include only the current economic condition y_t and the current financial condition f_t. The row labeled "BIC" reports results for the case where the lag lengths are chosen by the BIC with a maximum of 7 lags.

I plot the scoring sequence in Figure 3. It is clear that there is a sharp jump around the 2009Q2, which gives a natural guess that the density forecast breakdown results from the 2008 financial crisis.

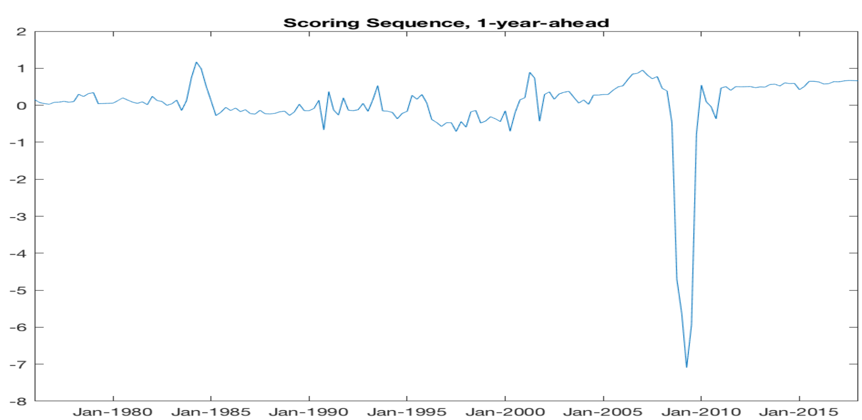


Figure 3: Standardized score sequence of 1-year-ahead model

Furthermore, I apply the robust version of the test on the 1-year-ahead density forecasting model of the real GDP growth. Figure 4 displays the robust density forecast breakdown test statistics across time. It implies that the density forecasting of real GDP growth experiences breakdowns during 2008Q2-2010Q2 (where $F_{R,M,t} < k_{0.05}$), which corresponds to the financial crisis.

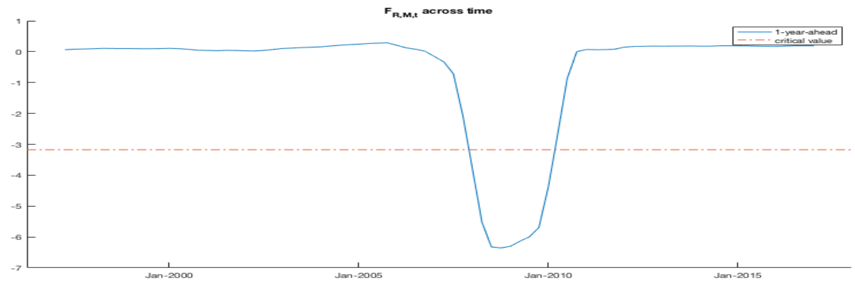


Figure 4: Robust density forecast breakdown test statistics $F_{R,M,t}$, real GDP growth

4.2 Assessment of density forecasting procedure of stock price return

This section provides an empirical assessment of the density forecasting procedures of stock price return.

The data used is the S&P 500 adjusted price, downloaded from Yahoo! Finance. The returns are calculated as the percentage change of daily S&P 500 adjusted prices. The data covers 9581 periods, ranging from 01/05/1981 through 12/31/2018. The data are plotted in Figure 5.

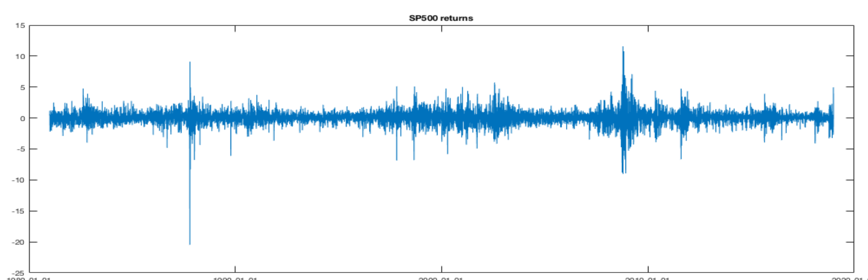


Figure 5: SP500 return

The sample is split in half, i.e. in-sample and out-of-sample periods for model estimation and density forecast evaluation. There are 4298 in-sample observations (01/05/1981 - 12/31/1997) and 5283 out-of-sample observations (01/01/1998 - 12/31/2018). Note that I apply rolling window scheme and recursive window scheme, which, compared with the fixed window scheme, are more widely used in practice and adapt better to the changes in the economy.

I fit the SP500 return with a GARCH model, and construct the one-day-ahead forecast according, as in the literature, GARCH models are widely used to model asset price volatility dynamics. The applications of (G)ARCH models on financial data dated back to Engle (1982), Bollerslev (1986), Taylor (1986). The (G)ARCH models, dealing with the assumption of non-stationarity and treating heteroscedasticity as a variance to be modelled, have become widespread tools for analysis of financial data. Diebold, Gunther and Tay (1997) studied the stock price return using a GARCH model. They selected an MA(1)-GARCH(1,1) model based on both the Akaike and Schwarz information criteria. Besides, there are many studies showing that GARCH type models are the best for forecasting stock market volatility.

I fit the SP500 return with a GARCH(1,1) model

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1)$$
$$\sigma_t^2 = w + a y_{t-1}^2 + b\sigma_{t-1}^2$$

At each period t, I apply the MLE on the corresponding in-sample observations. As for forecasting, given the MLE estimators $\hat{w}_t$, $\hat{a}_t$ and $\hat{b}_t$ at each period t, I assume normality of ϵ_t, and the one-day-ahead density forecast of y_{t+1} is

$$\hat{\phi}_{t+1}^f(y) = \phi(y|\hat{w}_t, \hat{a}_t, \hat{b}_t) = \frac{1}{\sqrt{2\pi\hat{\sigma}_{t+1}^{f2}}} e^{-\frac{y^2}{2\hat{\sigma}_{t+1}^{f2}}}$$

where $\hat{\sigma}_{t+1}^f$ is constructed using MLE estimators $\hat{w}_t$, $\hat{a}_t$ and $\hat{b}_t$.

In addition to a GARCH(1,1) model, a non-parametric kernel density estimation and forecasting procedure is also applied. I apply the KDE at each period t on the corresponding in-sample observations, and the one-day-ahead density forecast is the same as the density estimator at period t:

$$\hat{\phi}_{t+1}^f(y) = \hat{\phi}_t^e(y) = \frac{1}{Rh_R}\sum_{t=1}^{R} \kappa(\frac{y_t - y}{R}), \qquad h_R = cR^{-1/5}$$

Table 4 gives the results of a density forecast breakdown test, which is based on measures of global scoring surprise. The results imply that, for the one-day-ahead density forecasts of SP500 return, the null hypothesis of no density forecast breakdown cannot be rejected at a 5% significance level using a GARCH(1,1) model, while the null hypothesis is rejected at a 5% significance level using KDE and the forecasting procedure. This result corresponds to the existence of conditional time-varying volatility, which is not correctly captured by the simple KDE and forecasting procedure.

Table 4: Results of Density Forecast Breakdown Test, p-value

horizon	$h = 0$		$h = 1$	
scheme	roll	rec	roll	rec
GARCH(1,1)	0.1225	0.0334	0.1074	0.0495
kernel	0.0000	0.0104	0.0000	0.0073

Note: The table reports the p-values of the density forecast breakdown (one-sided) test. I use a rolling window scheme with $T = 181$ and $R = 90$ for real-time data. The density forecast horizons are $h = 0, 1$ quarters. In the KDE, kernel function is chosen to be normal.

Additionally, I apply the robust version of the test on kernel density estimation and forecasting model of the stock price return. Figure 6 displays the robust density forecast breakdown test statistics, which is based on measures of local scoring surprise. It implies that the density forecasting of SP500 return experiences breakdowns during 2008.07.19-2011.11.09 (where $F_{R,M,t} < k_{0.05}$), which corresponds to the financial crisis.

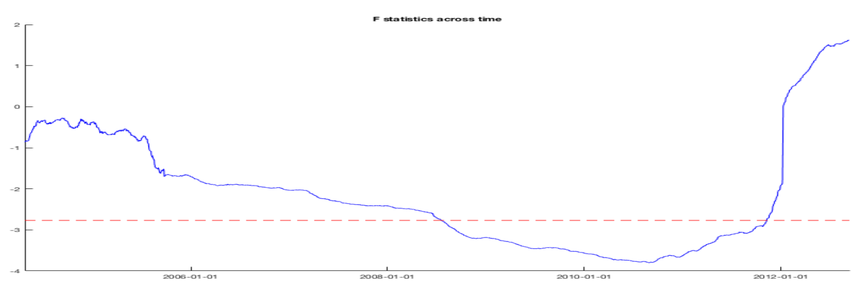


Figure 6: Robust density forecast breakdown test statistics $F_{R,M,t}$, SP500 return

5 Conclusion

This paper develops a formal test to capture density forecast breakdowns. The (robust) density forecast breakdown test statistic is built based on measures of (local) average scoring surprise, together with an appropriate asymptotic variance estimator required in the test statistic. The (robust) density forecast breakdown test can be applied in a wide range of procedures and is valid under general assumptions. It permits parametric estimation procedures including OLS, GMM and ML, as well as non-parametric estimation procedures such as KDE.

Compared to the literature related to density forecast evaluation, this paper shades light on the following perspectives: (i) Rather than specification testing, this paper focuses on the consistency of the future density forecasting performance and the anticipated performance based on past information. (ii) The test proposed in this paper captures all kinds of changes that affect density forecasting performance, including both parameter changes and other types of changes. (iii) The test proposed in this paper can be applied not only to parametric density forecasting procedures but also to non-parametric procedures. The test statistic corrects parameter estimation uncertainty and allow for dynamic misspecification under both hypotheses in parametric density forecasting. As for non-parametric procedures, this paper provides conditions under which the estimation uncertainty is asymptotically irrelevant.

The Monte Carlo results indicate that the density forecast breakdown test has good size property in moderately large samples and that the density forecast breakdown test has power against changes in mean and variance, as well as changes in distribution type.

To illustrate the empirical usefulness of the test proposed in this paper, the author investigates whether there are density forecast breakdowns in the density forecasting of real gross domestic product (GDP) growth and of daily SP500 return. There is evidence that the one-year-ahead density forecasts of real GDP growth, modeled as a linear function of current and lagged financial and economic conditions, experienced density forecast breakdowns during 2008Q2-2010Q2. In addition, the density forecasting of SP500 return experiences no density forecast breakdown using a GARCH(1,1) model, but experiences breakdowns using non-parametric KDE during 07/19/2008 to 11/09/2011.

References

Adrian, T., Boyarchenko, N. and Giannone, D.: 2016, Vulnerable Growth., *American Economic Review (forthcoming)* .

Amisano, G. and Giacomini, R.: 2007, Comparing density forecasts via weighted likelihood ratio tests., *Journal of Business & Economic Statistics* **25**.

Andrews, D.: 1993, Tests for Parameter Instability and Structural Change with Unknown Change Point., *Econometrica: Journal of the Econometric Society* .

Bai, J.: 2003, Testing Parametric Conditional Distributions of Dynamic Models., *Review of Economics and Statistics* **85**.

Bai, J. and Perron, P.: 1998, Estimating and Testing Linear Models with Multiple Structural Changes, *Econometrica* **66**, 47–78.

Berkowitz, J.: 2001, Testing density forecasts, with applications to risk management., *Journal of Business & Economic Statistics* **19**.

Bollerslev, T.: 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **31**.

Clements, M. and Hendry, D.: 1998, *Forecasting Economic Time Series.*, Cambridge University Press.

Corradi, V. and Swanson, N.: 2006, Predictive Density Evaluation., *Handbook of Economic Forecasting.*, Vol. 1, pp. 197–284.

Diebold, F., Gunther, T. and Tay, A.: 1997, Evaluating Density Forecasts.

Diebold, F., Gunther, T. and Tay, A.: 1998, Evaluating Density Forecasts with Applications to Finance and Management., *International Economic Review* **39**, 863–883.

Diebold, F. and Lopez, J.: 1996, Forecast Evaluation and Combination., *Handbook of Statistics*, Vol. 14, pp. 241–268.

Elliott, G. and Mller, U.: 2006, Efficient Tests for General Persistent Time Variation in Regression Coefficients., *The Review of Economic Studies* **73**.

Engle, R.: 1982, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation., *Econometrica: Journal of the Econometric Society* .

Giacomini, R. and Rossi, B.: 2009, Detecting and Predicting Forecast Breakdowns., *The Review of Economic Studies* **76**.

Giacomini, R. and Rossi, B.: 2010, Forecast Comparisons in Unstable Environments., *Journal of Applied Econometrics* **25**.

Giacomini, R. and White, H.: 2006, Tests of Conditional Predictive Ability., *Econometrica* **74**.

Gneiting, T. and Raftery, A.: 2007, Strictly Proper Scoring Rules, Prediction, and Estimation., *Journal of the American Statistical Association* **102**.

Good, I.: 1952, Rational Decisions., *Journal of the Royal Statistical Society* .

Hansen, B.: 2000, Testing for Structural Change in Conditional Models., *Journal of Econometrics* **97**.

Hong, Y. and Li, H.: 2004, Nonparametric Specification Testing for Continuous-time Models with Applications to Term Structure of Interest Rates., *The Review of Financial Studies* **18**.

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**.

Knppel, M.: 2015, Evaluating the Calibration of Multi-step-ahead Density Forecasts Using Raw Moments., *Journal of Business & Economic Statistics* **33**.

Li, Q. and Racine, J.: 2007, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Lopez, J.: 2001, Evaluating the Predictive Accuracy of Volatility Models., *Journal of Forecasting* **20**.

Newey, W. and West, K.: 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* **55**, 703–708.

Rossi, B. and Sekhposyan, T.: 2017, Alternative Tests for Correct Specification of Conditional Predictive Densities.

Tay, A. and Wallis, K.: 2000, Density Forecasting: a Survey., *Journal of forecasting* **19**.

Taylor, S.: 1986, Modelling Financial Time Series., *Wiley, Chichester* .

West, K.: 1996, Asymptotic Inference about Predictive Ability., *Econometrica: Journal of the Econometric Society* .

White, H.: 2001, *Asymptotic Theory for Econometricians*, Academic press.

White, H. and Domowitz, I.: 1984, Nonlinear Regression with Dependent Observations., *Econometrica: Journal of the Econometric Society* .

Winkler, R.: 1967, The Quantification of Judgment: Some Methodological Suggestions., *Journal of the American Statistical Association* **62**.

Wooldridge, J. and White, H.: 1988, Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes., *Econometric Theory* **4.**

Appendix

A Proof

A.1 Notation

Let $\cdot^*$ denote the counterparts evaluated at the pseudo parameter Θ^*; let $\tilde{\cdot}$ denote the demeaned counterparts. We omit the parameters of a function for simplicity. For example, $\bar{S}_t$, S_{t+h} and SS_{t+h} is used for $\bar{S}_t(\hat{\Theta}_t)$, $S_{t+h}(\hat{\Theta}_t)$ and $SS_{t+h}(\hat{\Theta}_t)$.

Define $D_{t+h} \equiv \frac{\partial S_{t+h}}{\partial \Theta} - \frac{\partial \bar{S}_t}{\partial \Theta} = \frac{\partial SS_{t+h}}{\partial \Theta}$, for $t = R, \ldots, T-h$, which is a $1 \times k$ vector; define $D^*_{t+h} \equiv \frac{\partial SS^*_{t+h}}{\partial \Theta}$; $\tilde{D}^*_{t+h} \equiv D^*_{t+h} - E[D^*_{t+h}]$.

A.2 Algorithm

Algorithm 1 *Construct the following: (1) a $1 \times T$ vector S^* of in-sample and out-of-sample scorings, with element $S^*_t, t = 1, \ldots, T$, and its counterpart S with element $S_t, t = 1, \ldots, T$:*

$$S^* \equiv \left[\underbrace{S^*_1(\Theta^*), \ldots, S^*_R(\Theta^*)}_{R}, \underbrace{S_{R+1}(\Theta^*), \ldots, S_{R+h-1}(\Theta^*)}_{h-1}, \underbrace{S^*_{R+h}(\Theta^*), \ldots, S^*_T(\Theta^*)}_{P} \right]$$

$$\text{fixed} \quad S \equiv \left[\underbrace{S_1(\hat{\Theta}_R), \ldots, S_R(\hat{\Theta}_R)}_{R}, \underbrace{S_{R+1}(\hat{\Theta}_R), \ldots, S_{R+h-1}(\hat{\Theta}_R)}_{h-1}, \underbrace{S_{R+h}(\hat{\Theta}_R), \ldots, S_T(\hat{\Theta}_R)}_{P} \right]$$

$$\text{rolling} \quad S \equiv \left[\underbrace{S_1(\hat{\Theta}_R), \ldots, S_R(\hat{\Theta}_R)}_{R}, \underbrace{S_{R+1}(\hat{\Theta}_{R+1}), \ldots, S_{R+h-1}(\hat{\Theta}_{R+h-1})}_{h-1}, \underbrace{S_{R+h}(\hat{\Theta}_R), \ldots, S_T(\hat{\Theta}_{T-h})}_{P} \right]$$

$$\text{recursive} \quad S \equiv \left[\underbrace{S_1(\hat{\Theta}_R), \ldots, S_R(\hat{\Theta}_R)}_{R}, \underbrace{S_{R+1}(\hat{\Theta}_{R+1}), \ldots, S_{R+h-1}(\hat{\Theta}_{R+h-1})}_{h-1}, \underbrace{S_{R+h}(\hat{\Theta}_R), \ldots, S_T(\hat{\Theta}_{T-h})}_{P} \right]$$

and the corresponding vector $\tilde{S}$ of demeaned scorings, where $\tilde{S}_t \equiv S_t - T^{-1} \sum_{j=1}^{T} S_j$; (2) a $1 \times T$ vector of weights, depending on forecasting scheme, with elements $\omega^S_t, t = 1, \ldots, T$:

$$\text{fixed} \quad \omega^S \equiv \left[\underbrace{-\frac{P}{R},\ldots,-\frac{P}{R}}_{R},\underbrace{0,\ldots,0}_{h-1},\underbrace{1,\ldots,1}_{P}\right]$$

$$\text{rolling}(P<R) \quad \omega^S \equiv \left[\underbrace{-\frac{1}{R},\ldots,-\frac{P}{R}}_{P},\underbrace{-\frac{P}{R},\ldots,-\frac{P}{R}}_{R-P},\underbrace{-\frac{P-1}{R},\ldots,-\frac{P-h+1}{R}}_{h-1},\underbrace{1-\frac{P-h}{R},\ldots,1-\frac{1}{R}}_{P-h},\underbrace{1,\ldots,1}_{h}\right]$$

$$\text{rolling}(P\geq R) \quad \omega^S \equiv \left[\underbrace{-\frac{1}{R},\ldots,-\frac{R}{R}}_{R},\underbrace{-\frac{R}{R},\ldots,-\frac{R}{R}}_{h-1},\underbrace{0,\ldots,0}_{P-R-h-1},\underbrace{1-\frac{R-1}{R},\ldots,1-\frac{1}{R}}_{R-1},\underbrace{1,\ldots,1}_{h}\right]$$

$$\text{recursive} \quad \omega^S \equiv \left[\underbrace{-a_{R,0},\ldots,-a_{R,0}}_{R},\underbrace{a_{R,1},\ldots,a_{R,h-1}}_{h-1},\underbrace{1-a_{R,h},\ldots,1-a_{R,P-1}}_{P-h},\underbrace{1,\ldots,1}_{h}\right]$$

$$a_{R,j} = \frac{1}{R+j}+\frac{1}{R+j+1}+\cdots+\frac{1}{T-h}$$

so that

$$\sum_{t=1}^{T}\omega_t^S S_t^* = \sum_{t=R}^{T} SS_{t+h}^*$$

Algorithm 2 *Construct the following: (1) a $(k+1)\times T$ vector of in-sample and out-of-sample h^*), as defined in Assumption P 3, with element $h_t^*, t=1,\ldots,T$, and its counterpart h with element $h_t, t=1,\ldots,T$:*

$$h^* \equiv \left[\underbrace{h_1^*(\Theta^*),\ldots,h_R^*(\Theta^*)}_{R},\underbrace{h_{R+1}^*(\Theta^*),\ldots,h_{T-h}^*(\Theta^*)}_{P-1},\underbrace{O,\ldots,O}_{h}\right]$$

$$\text{fixed} \quad h \equiv \left[\underbrace{h_1(\hat\Theta_R),\ldots,h_R(\hat\Theta_R)}_{R},\underbrace{h_{R+1}(\hat\Theta_R),\ldots,h_{T-h}(\hat\Theta_R)}_{P-1},\underbrace{O,\ldots,O}_{h}\right]$$

$$\text{rolling} \quad h \equiv \left[\underbrace{h_1(\hat\Theta_R),\ldots,h_R(\hat\Theta_R)}_{R},\underbrace{h_{R+1}(\hat\Theta_{R+1}),\ldots,h_{T-h}(\hat\Theta_{T-h})}_{P-1},\underbrace{O,\ldots,O}_{h}\right]$$

$$\text{recursive} \quad h \equiv \left[\underbrace{h_1(\hat\Theta_R),\ldots,h_R(\hat\Theta_R)}_{R},\underbrace{h_{R+1}(\hat\Theta_{R+1}),\ldots,h_{T-h}(\hat\Theta_{T-h})}_{P-1},\underbrace{O,\ldots,O}_{h}\right]$$

(2) a $1\times T$ vector of weights ω^{h*}, depending on forecasting scheme, with elements $\omega_t^{h*}, t=1,\ldots,T$

$$
\begin{aligned}
\text{fixed} \quad \omega^{h*} &\equiv \Bigg[\underbrace{-\frac{\sum_{t=R}^{T-h} D^*_{t+h}B^*_R}{R},\ldots,-\frac{\sum_{t=R}^{T-h} D^*_{t+h}B^*_R}{R}}_{R},\underbrace{O,\ldots,O}_{T-R}\Bigg]\\
\text{rolling}(P<R) \quad \omega^{h*} &\equiv \Bigg[\underbrace{-\frac{D^*_{R+h}B^*_R}{R},\ldots,-\frac{\sum_{t=R}^{T-h} D^*_{t+h}B^*_t}{R}}_{P},\underbrace{-\frac{\sum_{t=R}^{T-h} D^*_{t+h}B^*_t}{R},\ldots,-\frac{\sum_{t=R}^{T-h} D^*_{t+h}B^*_t}{R}}_{R-P},\underbrace{-\frac{\sum_{t=R+1}^{T-h} D^*_{t+h}B^*_t}{R},\ldots,-\frac{D^*_T B^*_{T-h}}{R}}_{P-1},\underbrace{O,\ldots,O}_{h}\Bigg]\\
\text{rolling}(P\geq R) \quad \omega^{h*} &\equiv \Bigg[\underbrace{-\frac{D^*_{R+h}B^*_R}{R},\ldots,-\frac{\sum_{t=R}^{2R-1} D^*_{t+h}B^*_t}{R}}_{R},\underbrace{-\frac{\sum_{t=R}^{2R} D^*_{t+h}B^*_t}{R},\ldots,-\frac{\sum_{t=P}^{T-h} D^*_{t+h}B^*_t}{R}}_{P-R},\underbrace{-\frac{\sum_{t=P+1}^{T-h} D^*_{t+h}B^*_t}{R},\ldots,-\frac{D^*_T B^*_{T-h}}{R}}_{R-1},\underbrace{O,\ldots,O}_{h}\Bigg]\\
\text{recursive} \quad \omega^{h*} &\equiv \Bigg[\underbrace{b^*_{R,0},\ldots,b^*_{R,0}}_{R},\underbrace{b^*_{R,1},\ldots,b^*_{R,P-1}}_{P-1},\underbrace{0,\ldots,0}_{h}\Bigg]\\
b^*_{R,j} &= \frac{D^*_{R+h+j}B^*_{R+j}}{R+j}+\frac{D^*_{R+h+j+1}B^*_{R+j+1}}{R+j+1}+\cdots+\frac{D^*_T B^*_{T-h}}{T-h}
\end{aligned}
$$

so that

$$\sum_{t=1}^{T}\omega_t^{h*}h_t^* = \sum_{t=R}^{T} D^*_{t+h}B^*_t H^*_t$$

and the counterpart ω^h with elements $\omega_t^h, t = 1,\ldots,T$ by replacing $D^*_\cdot, B^*_\cdot$ with $D_\cdot, B_\cdot$ evaluated with estimated parameters.

A.3 Lemmas

Lemma 1. *(a) For $0 \leq a < 0.5$, $P^{-1/2}\sum_{t=R}^{T} t^{-1+a} \to 0$; (b) $P^{-1/2}\sum_{t=R}^{T} t^{-1/2} = O(1)$.*

Proof: See West (1996) Lemma 1.

Lemma 2. *For $0 \leq a < 0.5$, (a) $\sup_t |t^a H^*_t| \xrightarrow{p} 0$; (b) $\sup_t |t^a(\hat{\Theta}-\Theta^*)| \xrightarrow{p} 0$.*

Proof: (a) is shown in West (1996) Lemma 3(a). (b) By Assumption 3 with new definition of B_t and H^*_t, it still holds that $\sup_t |B_t - B^*_t| \xrightarrow{p} 0$.

$$
\begin{aligned}
\sup_t |t^a(\hat{\Theta}-\Theta^*)| &\equiv \sup_t |t^a B_t H^*_t|\\
&\leq \sup_t |t^a(B_t - B^*_t)H^*_t| + \sup_t |t^a B^*_t H^*_t|\\
&\leq \sup_t |B_t - B^*_t| \sup_t |t^a H^*_t| + \sup_t |B^*_t| \sup_t |t^a H^*_t|\\
&\xrightarrow{p} 0
\end{aligned}
$$

Lemma 3. *(a)* $R_1 \equiv P^{-1/2}\sum_{t=R}^{T}\tilde{D}^*_{t+h}B^*_t H^*_t = op(1)$*; (b)* $R_2 \equiv \frac{1}{2}P^{-1/2}\sum_{t=R}^{T}(\hat{\Theta}_t - \Theta^*)'\frac{\partial^2 SS_{t+h}(\bar{\Theta}^*)}{\partial\Theta\partial\Theta'}(\hat{\Theta}_t - \Theta^*) = op(1)$*, where* $\bar{\Theta}^*$ *is an intermediate point between* $\hat{\Theta}_t$ *and* Θ^**.*

Proof: (a) is analogous to proof of Lemma 1 (a) in Giacomini and Rossi (2009) by replacing L with S in this context and redefining B, D in this context.

(b) For some a, $0 < a < 0.5$, C a positive constant, m_t defined in Assumption A2(b) and denoting by $\bar{m}_t$ the mean of the m_t's over the relevant in-sample window at time t, we have

$$\begin{aligned}
R_2 &= \frac{1}{2}P^{-1/2}\sum_{t=R}^{T} t^{1-a}(\hat{\Theta}_t - \Theta^*)'\left(t^{a-1}\frac{\partial^2 SS_{t+h}(\bar{\Theta}^*)}{\partial\Theta\partial\Theta'}\right)(\hat{\Theta}_t - \Theta^*) \\
&\leq C \sup_{R\leq t\leq T}|t^{0.5-0.5a}(\hat{\Theta}_t - \Theta^*)|^2 P^{-1/2}\sum_{t=R}^{T} t^{a-1}|\frac{\partial^2 SS_{t+h}(\bar{\Theta}^*)}{\partial\Theta\partial\Theta'}| \\
&\leq C \sup_{R\leq t\leq T}|t^{0.5-0.5a}(\hat{\Theta}_t - \Theta^*)|^2 P^{-1/2}\sum_{t=R}^{T} t^{a-1}\left(|\frac{\partial^2 S_{t+h}(\bar{\Theta}^*)}{\partial\Theta\partial\Theta'}| + |\frac{\partial^2 \bar{S}_t(\bar{\Theta}^*)}{\partial\Theta\partial\Theta'}|\right) \\
&\leq C \sup_{R\leq t\leq T}|t^{0.5-0.5a}(\hat{\Theta}_t - \Theta^*)|^2 P^{-1/2}\sum_{t=R}^{T} t^{a-1}(m_{t+h} + \bar{m}_t) = op(1)
\end{aligned}$$

using Lemma 1(a) and Lemma 2(b), Assumption 2(b).

Lemma 4. $\frac{T}{P}V_T^{SS*} \equiv var\left(P^{-1/2}\sum_{t=1}^{T}\omega_t^S\tilde{S}_t^*\right) > 0$ *for all* T *sufficiently large.*

Proof: The proof is analogous to proof of Lemma 2 in Giacomini and Rossi (2009) by replacing L with S in this context.

A.4 Proof of Theorem 1

Apply a second order mean value expansion on the numerator of $t_{R,P,h}$ in (5), we have

$$
\begin{aligned}
&\sqrt{P}\left[\frac{1}{P}\sum_{t=R}^{T} SS_{t+h}(\hat{\Theta}_t) - \mathrm{E}[\frac{1}{P}\sum_{t=R}^{T} SS_{t+h}(\Theta^*)]\right] \\
&= P^{-1/2}\sum_{t=R}^{T}\left(SS_{t+h}(\Theta^*) - \mathrm{E}[SS_{t+h}(\Theta^*)]\right) + P^{-1/2}\sum_{t=R}^{T}\frac{\partial SS_{t+h}(\Theta^*)}{\partial\Theta}(\hat{\Theta}_t - \Theta^*) \\
&\quad + \frac{1}{2}P^{-1/2}\sum_{t=R}^{T}(\hat{\Theta}_t - \Theta^*)'\frac{\partial^2 SS_{t+h}(\bar{\Theta}^*)}{\partial\Theta\partial\Theta'}(\hat{\Theta}_t - \Theta^*) \\
&= P^{-1/2}\sum_{t=R}^{T}\left(SS_{t+h}(\Theta^*) - \mathrm{E}[SS_{t+h}(\Theta^*)]\right) + P^{-1/2}\sum_{t=R}^{T}\mathrm{E}\left[\frac{\partial SS_{t+h}(\Theta^*)}{\partial\Theta}\right](\hat{\Theta}_t - \Theta^*) \\
&\quad + P^{-1/2}\sum_{t=R}^{T}\left(\frac{\partial SS_{t+h}(\Theta^*)}{\partial\Theta} - \mathrm{E}\left[\frac{\partial SS_{t+h}(\Theta^*)}{\partial\Theta}\right]\right)(\hat{\Theta}_t - \Theta^*) + \frac{1}{2}P^{-1/2}\sum_{t=R}^{T}(\hat{\Theta}_t - \Theta^*)'\frac{\partial^2 SS_{t+h}(\bar{\Theta}^*)}{\partial\Theta\partial\Theta'}(\hat{\Theta}_t - \Theta^*) \\
&= P^{-1/2}\sum_{t=R}^{T}\left(SS_{t+h}(\Theta^*) - \mathrm{E}[SS_{t+h}(\Theta^*)]\right) + P^{-1/2}\sum_{t=R}^{T}\mathrm{E}\left[D^*_{t+h}\right]B^*_t H^*_t + P^{-1/2}\sum_{t=R}^{T}\tilde{D}^*_{t+h}B^*_t H^*_t \\
&\quad + \frac{1}{2}P^{-1/2}\sum_{t=R}^{T}(\hat{\Theta}_t - \Theta^*)'\frac{\partial^2 SS_{t+h}(\bar{\Theta}^*)}{\partial\Theta\partial\Theta'}(\hat{\Theta}_t - \Theta^*)
\end{aligned} \tag{A.1}
$$

Apply Lemma 3, it holds that

$$
\begin{aligned}
&\sqrt{P}\left(\frac{1}{P}\sum_{t=R}^{T} SS_{t+h}(\hat{\Theta}_t) - \mathrm{E}[\frac{1}{P}\sum_{t=R}^{T} SS_{t+h}(\Theta^*)]\right) \\
&= P^{-1/2}\sum_{t=R}^{T}\left(SS_{t+h}(\Theta^*) - \mathrm{E}[SS_{t+h}(\Theta^*)]\right) + P^{-1/2}\sum_{t=R}^{T}\mathrm{E}\left[D^*_{t+h}\right]B^*_t H^*_t + op(1)
\end{aligned}
$$

Given Assumption P 5, the term $P^{-1/2}\sum_{t=R}^{T} E\left[D^*_{t+h}\right]B^*_t H^*_t$ retains. So it holds that

$$\sqrt{P}\left(\frac{1}{P}\sum_{t=R}^{T}SS_{t+h}(\hat{\Theta}_t)-\mathrm{E}[\frac{1}{P}\sum_{t=R}^{T}SS_{t+h}(\Theta^*)]\right)$$
$$=P^{-1/2}\sum_{t=R}^{T}\left(SS_{t+h}(\Theta^*)-\mathrm{E}[SS_{t+h}(\Theta^*)]\right)+P^{-1/2}\sum_{t=R}^{T}\mathrm{E}\left[D_{t+h}^*\right]B_t^*H_t^*+op(1)$$
$$\overset{under H_0}{=}P^{-1/2}\sum_{t=R}^{T}SS_{t+h}(\Theta^*)+P^{-1/2}\sum_{t=R}^{T}\mathrm{E}\left[D_{t+h}^*\right]B_t^*H_t^*+op(1)$$
$$\overset{under H_0}{=}\begin{bmatrix}1 & 1\end{bmatrix}P^{-1/2}\begin{bmatrix}\sum_{t=R}^{T}SS_{t+h}(\Theta^*)\\ \sum_{t=R}^{T}\mathrm{E}[D_{t+h}^*]B_t^*H_t^*\end{bmatrix}+op(1)$$

Now we show that, under H_0,

$$\left(\frac{T}{P}V_T\right)^{-1/2}P^{-1/2}\begin{bmatrix}\sum_{t=R}^{T}SS_{t+h}(\Theta^*)\\ \sum_{t=R}^{T}\mathrm{E}[D_{t+h}^*]B_t^*H_t^*\end{bmatrix}\xrightarrow{d}\mathcal{N}\left(O,I_2\right)$$

With notations defined above, we can write

$$\left(\frac{T}{P}V_T\right)^{-1/2}P^{-1/2}\begin{bmatrix}\sum_{t=R}^{T}SS_{t+h}(\Theta^*)\\ \sum_{t=R}^{T}E[D_{t+h}^*]B_t^*H_t^*\end{bmatrix}=V_T^{-1/2}T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega_t^S S_t^*\\ \sum_{t=1}^{T}\omega_t^{h*}h_t^*\end{bmatrix}=V_T^{-1/2}T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega_t^S \tilde{S}_t^*\\ \sum_{t=1}^{T}\omega_t^{h*}h_t^*\end{bmatrix}$$

where the last equality holds under H_0 as $T^{-1/2}\sum_{t=1}^{T}\omega_t^S\mathrm{E}[S_t^*]=T^{-1/2}\sum_{t=1}^{T}E[\frac{1}{P}\sum_{t=R}^{R}SS_{t+h}^*]=0$.

Now we show that,

$$V_T^{*-1/2}T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega_t^S \tilde{S}_t^*\\ \sum_{t=1}^{T}\omega_t^{h*}h_t^*\end{bmatrix}\xrightarrow{d}\mathcal{N}\left(O,I_2\right),\qquad V_T^*=var\left(T^{-1/2}\begin{bmatrix}\sum_{t=1}^{T}\omega_t^S \tilde{S}_t^*\\ \sum_{t=1}^{T}\omega_t^{h*}h_t^*\end{bmatrix}\right)$$

We verify that the zero-mean vector sequence $\left\{[V_T^{*-1/2}\omega_t^S\tilde{S}_t^*,V_T^{*-1/2}\omega_t^{h*}h_t^*]'\right\}_{t=1}^{T}$ satisfies the conditions of Wooldridge and White (1988) Central Limit Theorem for mixing processes (also see Theorem 5.20 in White (2001)).

- $g_t=[V_T^{*-1/2}\omega_t^S\tilde{S}_t^*,V_T^{*-1/2}\omega_t^{h*}h_t^*]'$ is a function of only a finite number of leads and lags of Z_t, it follows from Lemma 2.1 of White and Domowitz (1984) that it is mixing of the same size as Z_t, which is mixing with α of size $-r/(r-2), r>2$.
- By Assumption P 4, we have $|\tilde{S}_t^*|^{2r}<\infty$. And $|\omega_t^S|<\infty$ for all t.(It holds for the fixed and

rolling schemes by Assumption P 7; it holds for recursive scheme by the fact that $a_{m,j} < a_{m,0} \rightarrow ln(1+\pi) < \infty$.) Since V_T^* is p.d., it holds that $E|V_T^{*-1/2}\omega_t^S \tilde{S}_t^*|^{2r} < \infty$.

- By defining $\lambda_t = V_T^{*-1/2}\omega_t^{h*} = V_T^{*-1/2}\frac{1}{T}\sum_{j=1}^T \mathrm{E}[\partial S_j^*]P_{t,j}$, where $P_{t,j}$ is defined in the proof Lemma 1, and by the fact that $|\lambda_{t,i}| < \infty$ for all t, i (using Assumption P 5, $P_{t,j}$ having bounded component shown in the proof of Lemma 1, and that V_T^* is p.d.), we can write

$$E|V_T^{*-1/2}\omega_t^{h*}h_t^*|^{2r} = E|\lambda_t h_t^*|^{2r} = E|\sum_{i=1}^q \lambda_{t,i}h_{i,t}^*|^{2r} \leq \left[\sum_{i=1}^q |\lambda_{t,i}|(|E|h_{i,t}^*|^{2r})^{1/2r}\right]^{2r} < \infty$$

  where the first inequality uses Minkowski's inequality and the second inequality uses Assumption P 4.

This implies that we can apply Wooldridge and White (1988) Central Limit Theorem for mixing processes on the zero-mean vector sequence $\left\{[V_T^{*-1/2}\omega_t^S\tilde{S}_t^*, V_T^{*-1/2}\omega_t^{h*}h_t^*]'\right\}_{t=1}^T$, and get

$$V_T^{*-1/2}T^{-1/2}\begin{bmatrix}\sum_{t=1}^T \omega_t^S \tilde{S}_t^* \\ \sum_{t=1}^T \omega_t^{h*}h_t^*\end{bmatrix} \xrightarrow{d} \mathcal{N}(O, I_2), \qquad V_T^* = var\left(T^{-1/2}\begin{bmatrix}\sum_{t=1}^T \omega_t^S \tilde{S}_t^* \\ \sum_{t=1}^T \omega_t^{h*}h_t^*\end{bmatrix}\right)$$

The result then follows from the fact that $V_T - V_T^* \xrightarrow{p} 0$, owing to consistency of $\hat{\Theta}_t$ for Θ^* under H_0.

A.5 Proof of Theorem 2

$$\begin{aligned}
&\sqrt{P}\left\{\frac{1}{P}\sum_{t=R}^T SS_{t+h}(\hat{\phi}_t) - \mathrm{E}[\frac{1}{P}\sum_{t=R}^T SS_{t+h}(\phi^*)]\right\} \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^T (SS_{t+h}(\phi^*) - \mathrm{E}[SS_{t+h}(\phi^*)]) + \frac{1}{\sqrt{P}}\sum_{t=R}^T \left(SS_{t+h}(\hat{\phi}_t) - SS_{t+h}(\phi^*)\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^T (SS_{t+h}(\phi^*) - \mathrm{E}[SS_{t+h}(\phi^*)]) + \frac{1}{\sqrt{P}}\sum_{t=R}^T \left(S_{t+h}(\hat{\phi}_t) - S_{t+h}(\phi^*) + \frac{1}{R}\sum_{j=t-R+h+1}^t S_j(\hat{\phi}_t) - S_j(\phi^*)\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^T (SS_{t+h}(\phi^*) - \mathrm{E}[SS_{t+h}(\phi^*)]) \\
&\quad + \frac{1}{\sqrt{P}}\sum_{t=R}^T D_\phi^*(y_{t+h})\left(\hat{\phi}_t(y_t) - \phi^*(y_t)\right) + \frac{1}{R}\sum_{j=t-R+h+1}^t D_\phi^*(y_j)\left(\hat{\phi}_t(y_j) - \phi^*(y_j)\right)
\end{aligned} \tag{A.2}$$

The proof of the asymptotic justification of the first component follows from reasonings analogous to those in the proof of Theorem 1. As for the second component, here we show that the second component is asymptotically irrelevant in the case of kernel density estimation.

Proof of Example (kernel density estimation)

Given Assumption NP-kernel 1 to NP-kernel 3, according to Theorem 1.4 in Li and Racine (2007), it holds that

$$\sup_y |\hat{\phi}_t(y) - \phi(y)| = O(\sqrt{\frac{\ln R}{Rh_R}} + h_R^2) \qquad \forall t$$

See Li and Racine (2007) for detailed proof.

Given Assumption NP-kernel 5, $D_\phi^*(.)$ is bounded.

Thus, it then holds that

$$\begin{aligned}
&\frac{1}{\sqrt{P}} \sum_{t=R}^{T} D_\phi^*(y_{t+h}) \left(\hat{\phi}_t(y_t) - \phi^*(y_t)\right) + \frac{1}{R} \sum_{j=t-R+h+1}^{t} D_\phi^*(y_j) \left(\hat{\phi}_t(y_j) - \phi^*(y_j)\right) \\
&\leq \frac{1}{\sqrt{P}} \sum_{t=R}^{T} |D_\phi^*(y_{t+h})| | \left(\hat{\phi}_t(y_t) - \phi^*(y_t)\right) | + \frac{1}{R} \sum_{j=t-R+h+1}^{t} |D_\phi^*(y_j)| | \left(\hat{\phi}_t(y_j) - \phi^*(y_j)\right) | \\
&\leq \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \sup_y |D_\phi^*(y)| \sup_y |\hat{\phi}_t(y) - \phi(y)| + \frac{1}{R} \sum_{j=t-R+h+1}^{t} \sup_y |D_\phi^*(y)| \sup_y |\hat{\phi}_t(y) - \phi(y)| \\
&\leq \frac{1}{\sqrt{P}} P \sup_y |D_\phi^*(y)| \max_t \sup_y |\hat{\phi}_t(y) - \phi(y)| 2 \\
&= \sqrt{P} O(\sqrt{\frac{\ln R}{Rh_R}} + h_R^2)
\end{aligned}$$

Therefore, given Assumption NP-kernel 4, $\frac{P \ln R}{Rh_R} \to 0$ and $Ph_R^4 \to 0$ imply $\sqrt{P} O(\sqrt{\frac{\ln R}{Rh_R}} + h_R^2) = op(1)$.