International Journal of Forecasting 33 (2017) 416–432

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Density forecast evaluation in unstable environments

Gloria González-Rivera [a,*], Yingying Sun [b]

[a] Department of Economics, University of California, Riverside, CA, USA
[b] Bank of the West, San Francisco, CA, USA

ARTICLE INFO

Keywords:
Generalized autocontour-based testing
Structural breaks
Phillips curve

ABSTRACT

We propose a density forecast evaluation method in the presence of instabilities, which are defined as breaks in any conditional moment of interest and/or in the functional form of the conditional density of the process. Within the framework of the autocontour-based tests proposed by González-Rivera et al. (2011) and González-Rivera and Sun (2015), we construct Sup- and Ave-type tests, calculated over a collection of subsamples in the evaluation period. These tests have asymptotic distributions that are nuisance-parameter free and they are correctly sized and very powerful for detecting breaks in the parameters of the conditional mean and conditional variance. A power comparison with the tests of Rossi and Sekhposyan (2013) shows that our tests are more powerful across the models considered in their work. We analyze the stability of a dynamic Phillips curve and find that the best one-step-ahead density forecast of changes in inflation is generated by a Markov switching model that allows state shifts in the mean and variance of inflation changes as well as in the coefficient that links inflation and unemployment.

1. Introduction

In general, instability refers to changes in the parameters of a proposed forecasting model over the forecasting horizon. For clarification purposes, consider a simple model $y_{t+1} = \beta' x_t + \sigma\varepsilon_{t+1}$, with $\varepsilon_t \sim i.i.d.\, N(0, 1)$. The model is unstable over time if the slope coefficients β can change over the forecasting sample, either smoothly or abruptly, to contain one or more breaks. We may also entertain a time-varying variance such that σ is also subject to breaks, and we may have different conditional probability density functions, e.g., more or less thick tails, in different periods. This definition is general enough to account for most of the types of instability that are discussed in the current applied econometric literature. To date, the most comprehensive survey of the subject is that provided by Rossi (2014) in the *Handbook of Economic Forecasting*, which reports extensive empirical evidence of instabilities in macroeconomic and financial data. Some examples follow.

The instability of predictive regressions, in which the significance of predictive regressors varies over different subsamples, has been documented in studies of the predictability of stock returns (see Goyal & Welch, 2003; Paye & Timmermann, 2006; Rapach & Zhou, 2014), in exchange rate predictions (see Rogoff & Stavrakeva, 2008; Rossi, 2006) and in output growth and inflation forecasts (see Rossi & Sekhposyan, 2010; Stock & Watson, 2003). Naturally, linked to this evidence is the econometric issue of testing for parameter stability and structural breaks in the data, which has an illustrious history. From the Chow (1960) test to more recent works such as those of Andrews (1993), Andrews and Ploberger (1994) and Pesaran and Timmermann (2002), among others, testing for breaks has focused mainly on the behavior of the conditional mean. This paper aims to extend the testing for instabilities to

* Corresponding author.
E-mail address: gloria.gonzalez@ucr.edu (G. González-Rivera).

http://dx.doi.org/10.1016/j.ijforecast.2016.10.003

the *full* conditional density forecast that includes not only any conditional measure of interest (e.g., mean, variance, duration, etc.), but also the functional form of the assumed conditional density function. To the best of our knowledge, the literature on this question is very thin. We know only of the work by Rossi and Sekhposyan (R&S) (2013), who also tested for the instability of the density model, but using a different methodology. A comparison of the two approaches will be provided later.

The testing methodology that we propose is based on the AutoContouR (ACR) device introduced by González-Rivera, Senyuz, and Yoldas (2011) and González-Rivera and Yoldas (2012), and later generalized by González-Rivera and Sun (2015). A brief summary of this latter work follows. The null hypothesis is a correctly specified density forecast (with the joint hypothesis of correct dynamics in the moments of interest and a correct functional form of the density). We calculate the Rosenblatt (1952) probability integral transforms (PIT) that are associated with the point forecasts. Under the null, the PITs must be i.i.d uniformly distributed U[0,1]. The generalized autocontour (G-ACR) is a device (set of points) that is very sensitive to departures from the null in either direction; consequently, it provides the basis for very powerful tests. More specifically, for a time series of PITs, we construct the G-ACRs as squares (in the univariate case) of different probability areas within the maximum square (area of 1), or as hyper-cubes (in the multivariate case) of different probability volumes within the maximum hyper-cube formed by a multidimensional uniform density $[0, 1]^n$ for $n \geq 2$. By statistical comparisons of the location of the empirical PITs and the volume of the empirical G-ACRs with the location and volume of the population G-ACRs, we are able to construct a variety of tests for correct density forecasts. Since the shapes of the G-ACRs can be visualized, we can extract information about where the rejection of the null hypothesis comes from, and how. This testing framework is the foundation of the new stability tests, the Sup- and Ave-type statistics, proposed in this paper. In a potentially unstable data environment, we form rolling subsamples within the forecasting sample. For every subsample, we apply a battery of G-ACR tests and construct Sup- and Ave-type statistics for detecting instabilities. Although the limiting distribution of these tests is a function of Brownian motions, the tests are nuisance-parameter free and their distributions can be tabulated.[1]

The R&S (2013) tests and our tests have similar null hypotheses, i.e., the constancy of the density model over the prediction sample, although their statistics allow for dynamic misspecification. The R&S tests follow the setup of Corradi and Swanson (2006), so that their tests, which are also based on the PITs $\{u_t\}$ of the proposed model, are a function of the distance between the empirical cumulative distribution function and that of the uniform distribution, which is a 45 degree line. Under dynamic misspecification, the PITs are still distributed uniformly in [0,1], but are no longer independent. Thus, the R&S tests borrow from Corradi and Swanson, in that the limiting variance of the statistics has to take into account the potential lack of independence. Their κ_P test is a Kolmogorov–Smirnov-type statistic, and their C_P is a Cramer-von Mises-type statistic, which mainly exploits the "identical distribution" property of the PITs under the null. Our G-ACR tests are based on the object "autocontour", and measure the independence, denseness and uniform distribution of the PITs over a collection of squares in a two-dimensional space (u_{t-k}, u_t). By construction, our tests exploit the independence properties of the PITs directly. The asymptotic distribution of the R&S tests is based on the statistical properties of an empirical process. Our tests are simpler, in that they rely on the properties of a binary indicator with well-defined moments. When the parameter uncertainty is non-negligible, the critical values of the R&S tests and the G-ACR tests are obtained by simulation.

The paper is organized as follows. Section 2 reviews the G-ACR approach in order to make the exposition self-contained, and introduces the new statistics with their asymptotic distributions. Section 3 assesses the finite sample properties (size and power) of the tests. We offer an extensive assessment by considering (i) fixed, rolling, and recursive estimation schemes; (ii) different ratios of prediction to estimation sample sizes; and (iii) break points that occur in different periods of the prediction sample. We also present a comparison of the power of our tests with those of Rossi and Sekhposyan (2013). Section 4 uses the tests to assess the stability of the Phillips curve from 1958 onwards by evaluating the models proposed by Amisano and Giacomini (2007). Section 5 concludes. Appendix A contains mathematical proofs and Appendix B describes the parametric bootstrap for correcting the size of the tests. We also provide a supplementary file with additional simulation materials (see Appendix C).

2. Statistics and asymptotic distributions

2.1. Construction of the statistics

The test statistics are based on the autocontour (ACR) and generalized autocontour (G-ACR) methodologies proposed by González-Rivera et al. (2011), González-Rivera and Yoldas (2012), and González-Rivera and Sun (2015), which provide powerful tests for the dynamic specification of the conditional density model in either in-sample or out-of-sample environments. In the present context, we adapt these tests to instances where instabilities may be present in the data, so that we will also be able to detect unstable periods beyond the evaluation of the density model.

Let Y_t denote the random process of interest with a conditional density function $f(y_t|\Omega_{t-1})$, where Ω_{t-1} is the information set available up to time $t - 1$. Observe that the random process Y_t could enjoy very general statistical properties, e.g., heterogeneity, dependence, etc. The researcher will construct the conditional model by specifying a conditional mean, conditional variance or other conditional moments of interest, and making distributional assumptions as to the functional form of $f(\cdot)$. Based on the conditional model, she will then proceed

[1] Although we focus on out-of-sample density forecasts, the methodology proposed in this paper can also be applied to in-sample specification testing.