International Journal of Forecasting 31 (2015) 799–814

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Generalized autocontours: Evaluation of multivariate density models

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

[a] Department of Economics, University of California, Riverside, United States

[b] Department of Economics, Huazhong University of Science and Technology, China

ARTICLE INFO

Keywords:
Copula
Specification testing
Multivariate dynamic models
Density forecast

ABSTRACT

We propose a new tool, the Generalized Autocontour (G-ACR), as the basis for a battery of dynamic specification tests that are applicable (in-sample or out-of-sample) to univariate or multivariate random processes. We apply this methodology to the modeling of a multivariate system by specifying the dynamics of the marginal distributions of each process in the system, together with a copula that ties up the marginals to produce their multivariate distribution. We work with the probability integral transforms (PIT) of the system that, under a correct specification of the conditional model, should be i.i.d. U[0,1]. The dimensionality of the system is not a constraint, because the information contained in the vector of PITs is condensed into an indicator, which is the basis of the proposed tests. We construct hyper-cubes of different sizes within the maximum hyper-cube formed by a multidimensional uniform density $[0, 1]^n$, and assess the locations of the empirical PITs (duplex, triplex, n-plex of observations) within the corresponding population hyper-cubes. If the conditional model is correct, the volumes of the population hyper-cubes must be the same as those in their empirical counterparts. This approach allows the researcher to focus on different areas of the conditional density model, so as to assess the regions of interest. We estimate a trivariate model for a very large number of trades on the stocks of three large U.S. banks and find that the contemporaneous dependence among institutions is asymmetric, which implies that when liquidity drains (due to a lack of trading) in one institution, we should expect a concurrent effect among similar institutions. On the other hand, when liquidity is plentiful (due to dense trading), the trades on the stocks of the institutions are not correlated. We assess the models' performances by evaluating their one-step-ahead density forecasts of trades.

1. Introduction

The Generalized Autocontour (G-ACR) is a generalized version of the *autocontour* methodology proposed by González-Rivera, Senyuz, and Yoldas (2011, GR2011) for detecting misspecification in the dynamics of a time series model and departures from the assumed conditional density model. The G-ACR overcomes some of the limitations of the original methodology of GR2011. First, when the conditional density of interest departs from standard densities in financial econometrics, e.g. Normal, Student-t, Exponential, etc., the analytical expressions of the autocontours may be mathematically cumbersome to obtain, meaning that we need to resort to numerical methods in order to compute their density mass. The difficulty is compounded when the system is multivariate (González-Rivera & Yoldas, 2012). In contrast, the G-ACR is very easy to obtain for any density because it is based on

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

http://dx.doi.org/10.1016/j.ijforecast.2014.03.019

probability integral transforms (PIT) instead of standardized innovations, which are the basis of the original ACR. Second, GR2011 only consider univariate stochastic processes, with the dynamics restricted to the conditional mean and conditional variance, and a time-invariant functional form of the density of the standardized innovations of the model. The advantage of G-ACR is that it can be applied to either univariate or multivariate random processes. In a multivariate framework, the dimensionality of the system is not a constraint, because the information contained in the vector of PITs is condensed into an indicator, which constitutes the basis of the proposed tests. Furthermore, the components of the multivariate system may have different marginal densities, which could be tested individually; more importantly, though, the multivariate density, obtained as a copula function linking the marginals, can also be tested jointly. As a result, our statistics based on G-ACR are also useful diagnostics for correct copula specification. G-ACR does not restrict the dynamics of the model to any particular moment(s), and it is also applicable to cases where the predictive density does not have a closed form solution, e.g., multistep predictive densities in nonlinear models, and we have to resort to simulations or nonparametric methods, but where we can still obtain the PIT process from the simulated density. Third, the tests proposed by GR2011 have asymptotic variance–covariance matrices that do not all have closed-form solutions, with some combining parametric and nonparametric expressions. In contrast, and because of the simplicity of G-ACR, the asymptotic variances of the tests all have closed formulations that depend on only one parameter, the *a priori* specified probability level associated with the G-ACR.

As a brief introduction to G-ACR, explained in detail in the forthcoming sections, suffice it to say that the basis of our testing techniques is the construction of hyper-cubes of different sizes within the maximum hyper-cube formed by a multidimensional uniform density $[0, 1]^n$. We assess the locations of the empirical PITs (duplex, …, n-plex of observations) within the corresponding population hyper-cubes. If the multivariate model is correct, the volumes of the population hyper-cubes must be the same as those of their empirical counterparts. Our tests evaluate these differences statistically as either rejecting or failing to reject the proposed density model. This approach also permits us to focus on different areas of the conditional density in order to assess those regions of interest. There is also a graphical visualization aspect of our approach that is very helpful for guiding the modeling.

As an illustration of the proposed methodology, we will specify a multivariate model for the numbers of stock trades of three large U.S. banking institutions: Bank of America, JP Morgan Chase, and Wells Fargo. Though the number of trades is a discrete random variable, these three big banks show almost continuous trading, so that, for a given interval of time, the number of trades is large enough that the data can be considered to be almost continuous. For instance, at the 5 min frequency (from January 3 to June 30, 2011), the median number of trades is 1757 trades for Bank of America, 1300 for JP Morgan, and 1210 for Wells Fargo. Dynamic trading is important because it reflects the arrival of news, and is intimately related to issues of liquidity risk and market microstructure; see O'Hara (1995) and Madhavan (2000) among others. We proceed by specifying an autoregressive system for the number of trades of each bank. We will use different distributional assumptions for the marginal densities of each component of the system, but we are most interested in the modeling of contemporaneous correlations of the trades, as they may have implications for the risk that these large institutions pose to the banking system and beyond. We use a copula function to understand the contemporaneous correlation among the three banks. Heinen and Rengifo (2007) also implemented a copula approach, but restricted themselves to a normal copula where the dependence is contained in a correlation coefficient. As the recent crisis has shown, the correlation among institutions varies during episodes of low or high liquidity. We explore the possibility of an asymmetric contemporaneous correlation such that the correlation may be different when the number of trades is large (the market is very active) or small (the market is slow). We assess the model in an out-of-sample environment by evaluating the one-step-ahead density forecasts of the number of trades. The modeling and forecasting exercises will allow us to showcase the proposed testing methodology, and the visualization techniques in particular, used to drive the specification exercise.

The paper is organized as follows. In Section 2, we introduce G-ACR and present the testing methodology for univariate and multivariate models. In Section 3, we offer extensive Monte Carlo simulations to assess the size and power of the tests within the context of multivariate processes with and without contemporaneous correlation. In Section 4, we provide an empirical illustration dealing with a trivariate system for the number of trades for the three large banks mentioned above, and in Section 5 we conclude. The Appendix contains mathematical proofs.

2. Generalized autocontour and test statistics

We introduce a device — the generalized autocontour — as the basis for constructing statistical tests for the null hypothesis of a well-specified conditional density model, whether univariate or multivariate.

2.1. Generalized autocontour: G-ACR

Following Diebold, Gunther, and Tay (1998) among others, if the proposed predictive density model for Y_t, i.e. $\{f_t^*(y_t|\Omega_{t-1})\}_{t=1}^T$, coincides with the true conditional density $\{f_t(y_t|\Omega_{t-1})\}_{t=1}^T$, then the sequence of probability integral transforms (PIT) of $\{Y_t\}_{t=1}^T$ w.r.t $\{f_t^*(y_t|\Omega_{t-1})\}_{t=1}^T$, i.e. $\{u_t\}_{t=1}^T$, must be i.i.d. $U(0, 1)$, where $u_t = \int_{-\infty}^{y_t} f_t^*(v_t|\Omega_{t-1})dv_t$. Thus, the null hypothesis $H_0 : f_t^*(y_t|\Omega_{t-1}) = f_t(y_t|\Omega_{t-1})$ is equivalent to the null hypothesis $H_0' : \{u_t\}_{t=1}^T$ is i.i.d. $U(0, 1)$.

We construct the G-ACR under i.i.d. uniformity for both univariate and multivariate predictive densities. We start with the univariate case. Within the process $\{u_t\}_{t=1}^T$, we choose any vector $(u_t, u_{t-k}) \subset R^2$. Under $H_0' : \{u_t\}_{t=1}^T$ i.i.d. $U(0, 1)$, the G-ACR$_{\alpha_i,k}$ is defined as the set of points