Computational Statistics and Data Analysis 124 (2018) 1–14

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Multivariate specification tests based on a dynamic Rosenblatt transform

Igor L. Kheifets

ITAM, Av. Camino a Santa Teresa 930, Col. Heroes de Padierna, Magdalena Contreras, Ciudad de Mexico, 10700, Mexico

A R T I C L E   I N F O

Article history:
Received 23 May 2017
Received in revised form 26 January 2018
Accepted 30 January 2018
Available online 13 February 2018

Keywords:
Diagnostic test
Joint distribution
Multivariate modeling
Rosenblatt transform
LSTAR model

A B S T R A C T

This paper considers parametric model adequacy tests for nonlinear multivariate dynamic models. It is shown that commonly used Kolmogorov-type tests do not take into account cross-sectional nor time-dependence structure, and a test, based on multi-parameter empirical processes, is proposed that overcomes these problems. The tests are applied to a nonlinear LSTAR-type model of joint movements of UK output growth and interest rate spreads. A simulation experiment illustrates the properties of the tests in finite samples. Asymptotic properties of the test statistics under the null of correct specification and under the local alternative, and justification of a parametric bootstrap to obtain critical values, are provided.

1. Introduction

Robust nonparametric methods are hard to implement in a multidimensional case, and parametric modeling is often called for. For example, linear and nonlinear VAR models with Gaussian innovations are often used in macroeconometrics, while multivariate volatilities, which can be described by different types of multivariate GARCH (MGARCH) or copula-based models, are popular in financial econometrics. The use of a misspecified parametric model may result in misleading conclusions, in particular, biased estimates of monetary policy effects and underestimation of the risk in financial models. Thus it is crucial to develop specification testing procedures for these models. In a multidimensional context, it is important to know not only the time structure of the random vectors but also the dependence between contemporaneous variables, and this dependence should be used, for example, for portfolio diversification. Hence, we should be able to test for the correct specification of the joint multivariate distribution conditional on past information.

There is a huge literature on testing multivariate normality, see Mecklin and Mundfrom (2004). For testing a general type of distribution in a dynamic setup, a dynamic version of the Rosenblatt Transform (cf. Rosenblatt, 1952), which is also a type of a Probability Integral Transform, PIT, allows us to approach all kinds of distributions in a unified manner. The idea is that, given the true conditional distribution, one can transform the data to independent and identically distributed (i.i.d.) uniform random variables and possibly further to normal i.i.d. Then, instead of testing the shape of the initial distribution, the uniformness and independence of the transformed data can be evaluated with histograms and correlograms, as suggested by Diebold et al. (1999) and Clements and Smith (2002) for multivariate density forecast evaluation. In practice, however, the distribution is known only up to parameters; therefore the method of Diebold et al. (1999) cannot be applied. The reason is that, when estimates are plugged in, the (dynamic) Rosenblatt Transform delivers only approximately i.i.d. uniforms; moreover, the asymptotic distribution may change, and even become model and case dependent, see Durbin (1973). Ignoring parameter uncertainty introduces severe size distortions in such tests, as documented in simulations by Kalliovirta and Saikkonen (2010). The problem usually is solved by either transforming the statistics of interest to make them convergent

E-mail address: igor.kheifets@itam.mx.

https://doi.org/10.1016/j.csda.2018.01.022
0167-9473/

to a known distribution (Khmaladze, 1981) or by approximating critical values by the parametric bootstrap (Andrews, 1997). For nonparametric testing of multivariate GARCH models, Bai and Chen (2008) developed a Kolmogorov-type testing procedure based on the dynamic Rosenblatt Transform. They applied a Khmaladze (1981) martingale transformation (K-transformation) to the empirical process to obtain a limiting distribution of statistics. However, in a univariate setup, Corradi and Swanson (2006) noted that Kolmogorov-type tests, based on a one-parameter empirical process, may not distinguish some important alternatives to the conditional distribution. In an i.i.d. setup with conditioning on covariates, Delgado and Stute (2008) proposed a consistent test using a two-parameter empirical process coupled with a Khmaladze martingale transformation. In a time series setup, where a Kolmogorov-type test does not capture misspecification in the dynamics, Kheifets (2015) proposed a test based on a multi-parameter empirical process and used a bootstrap to obtain critical values.

In this paper, we consider nonparametric testing of a multivariate distribution specification. We study the consequences of using a Kolmogorov-type test in this setup. We find that Kolmogorov-type tests result in missing both dynamic and cross-section dependence. To overcome this problem, we consider tests based on a multivariate empirical process, adapting the weak convergence results of Kheifets (2015) to a multivariate case. As well as a Kolmogorov test, our PIT-based procedure can test nonlinear models and capture deviations in marginal distribution. Besides that, our test includes two ingredients: a dynamic check (similar to Kheifets, 2015) and a cross-section check. Thus our technique may be used not only for testing but also for investigating sources of misspecification. Our test complements the parametric tests of Kalliovirta and Saikkonen (2010) and Gonzalez-Rivera and Yoldas (2012). We avoid bandwidth selection, and our test statistics have a parametric rate of convergence, unlike the smoothing techniques of Hong and Li (2005), Li and Tkacz (2011), and Chen and Hong (2014), who use kernels to estimate the conditional distribution function and spectrum.

The contribution of the paper is the following: We develop the test and apply it to a model of UK growth and interest spreads and perform a set of Monte Carlo experiments to study the performance of the test in finite samples, where we observe results similar to Kolmogorov tests' in some cases and improvement in others. We derive asymptotic properties of the test under the null and the local alternative, taking into account the parameter estimation effect, and justify the use of bootstrapped critical values.

The remainder of the paper is organized as follows. Section 2 introduces specification test statistics, based on the dynamic Rosenblatt Transform. The empirical application is in Section 3. Monte Carlo experiments are shown in Section 4. Section 5 concludes. Asymptotic properties of the test are listed in the Appendix.

2. The test statistics

2.1. Our proposal

We now explain our methodology in detail. Suppose that a sequence of $d \times 1$ vectors $Y_1, Y_2, \ldots, Y_T$, where $Y_t = (Y_{t1}, Y_{t2}, \ldots, Y_{td})'$, $t = 1, \ldots, T$, is given. Let Ω_t be the information set at time t (not including Y_t), i.e., the σ-field of $\{Y_{t-1}, Y_{t-2}, \ldots\}$.

We consider a family of joint distributions $F_t(y|\Omega_t, \theta)$, conditional on the past information, parameterized by $\theta \in \Theta$, where $\Theta \subseteq R^L$ is a finite dimensional parameter space. Apart from allowing the conditional (information) set Ω_t to change with time, we permit change in the functional form of the distribution using subscript t in F_t. Our null hypothesis of correct specification is as follows:

H_0: The multivariate distribution of Y_t conditional on Ω_t is in the parametric family $F_t(y|\Omega_t, \theta)$ for some $\theta_0 \in \Theta$.

Note that by specifying the multivariate conditional distribution, we specify many properties of the data simultaneously, such as multivariate and univariate marginal distributions, univariate conditional distributions, time and cross-section dependence, symmetry, and moments. Many dynamic models can be written in the form of a conditional distribution. Examples include (non)linear vector autoregressive (VAR) and MGARCH models with i.i.d. parametric innovations, copula-based models with parametric marginals and possibly time-varying copula functions, and discretely sampled continuous-time models represented by a stochastic differential equation.

We now describe how to use PIT. In a simple univariate unconditional testing, we have that if $Y \sim F(\cdot)$, then $U = F(Y)$ is uniform, which is the base of the Kolmogorov test. More precisely, the null hypothesis of the Kolmogorov test is that $F(Y)$ is uniform. If we are interested in conditional distribution testing, which is the case when we have covariates or dynamics, we use the fact that if $Y_t|\Omega_t \sim F_t(\cdot|\Omega_t)$, then $U_t = F(Y_t|\Omega_t)$ is uniform and i.i.d. The distribution needs to be absolutely continuous. For a discrete distribution, one may use a different transform, see Kheifets and Velasco (2013, 2017).

In multivariate setup, $F(Y)$, which seems to be an obvious generalization of the PIT, is not generally uniform for a vector $Y \sim F(\cdot)$. $F(Y)$ is related to copula functions, which describe multivariate dependence without specifying marginals. The properties of $F(Y)$ were studied in Genest and Rivest (2001). In this paper we want to check the specification of the multivariate distribution; therefore, testing the specification of the copula function is not sufficient. For a multivariate PIT, we need to define the conditioning sets properly. Following Rosenblatt (1952) and Diebold et al. (1999), construct a long univariate series by stacking sequentially Y_t to get a $n = Td$ long univariate series, which we denote

$$\{Z_k\}_{k=1}^n := \{\ldots, Y_{t1}, Y_{t2}, \ldots, Y_{td}, \ldots\}_{t=1}^T. \tag{1}$$

In other words, $Z_k = Y_{t\ell}$ for $t = \lceil k/d \rceil$ and $\ell = k - td$, where $k = 1, \ldots, n$, $t = 1, \ldots, T$, $\ell = 1, \ldots, d$, and $\lceil x \rceil$ denotes the smallest integer not less than x. There are many ways to order and stack $Y_{t\ell}$; for example, $\{\ldots, Y_{t2}, Y_{t1}, \ldots, Y_{td}, \ldots\}$