Sparse Nonparametric Dynamic Graphical Models

Un modello Grafico non parametrico, dinamico e sparso

Fabrizio Poggioni, Mauro Bernardi, Lea Petrella

Abstract We propose a Sparse Nonparametric Dynamic Graphical Model for financial application. We base our model on multiple CAViaR quantile regression models, and we address the issue of the quantile crossing for this type of semi-parametric models. We show how to jointly estimate the multiple quantile levels by exploiting the conditions on the parameters and setting the estimation as a linear constrained optimization problem. We employ the defined non-crossing Multiple CAViaR model as non-parametric estimation of the marginal distributions to get a sparse dynamic graphical model .

Abstract *Proponiamo un modello grafico non parametrico e dinamico per applicazioni finanziarie. Stimiamo modelli di regressione quantilica multipla di tipo CAViaR, ed affrontiamo il tema del non crossing pe modelli semi parametrici. Mostriamo come stimare congiuntamente i diversi livelli garantendo la propriet di non crossing e come usare questi come stime di distribuzioni marginali per ottenere un modello grafico che risulti sparso e dinamico.*

Key words: Multiple Quantile, Non-Crossing, Dynamic Graphical Model,. . .

1 Introduction

In recent years the theme of graphic models has developed in literature. A *Graphical Model* exploits the graph theory as well as the statical theory to describe the

Fabrizio Poggioni
MEMOTEF, Sapienza University of Rome, Italy e-mail: fabrizio.poggioni@uniroma1.it

Mauro Bernardi
Department of Statistical Sciences, University of Padua, Padua, Italy e-mail: mauro.bernardi@unipd.it

Lea Petrella
MEMOTEF, Sapienza University of Rome, Italy e-mail: lea.petrella@uniroma1.it

dependency and conditional dependence ratios of a set of random variables. In financial applications the Graphical model tool allows the advantage of being able to describe complex structures in a simple way. Following the crises of the last decade, it has been very successful as a tool to describe systemic risk as systems of connections in financial markets. Quantile Regression is another very widespread tool in literature used to analyze dependency in financial markets, see for example [1]. We propose in this work an application of graphical models in which we model the time conditional CDF with dynamic quantile regression. Estimating multiple quantile regression involves numerical problems, such as the so called non-crossing problem of the estimated quantiles and the consequent violation of one of the basic principles of the inverse distributions functions: the monotone property. Even if quantile crossing problem is a finite sample problem and should be negligible when the sample size is sufficiently large and the model is correctly specified, for a large number of estimated quantities and perhaps with non-linear specifications of the model quantile crossing may remain a relevant issue. A vast literature exists about the quantile crossing problem: [7] address the quantile-crossing problem using a support vector regression approach for nonparametric models; [4] propose a method for non parametric models in which they use an initial estimate of the conditional distribution function in a first step and solve the problem of inversion and monotonization simultaneously; [9] propose a stepwise method for estimating multiple quantile regression functions without crossing for linear and kernel quantile regression models; [3] propose a method to address the non-crossing problem by rearranging the original estimated non-monotone curve into a monotone rearranged curve; [2] propose a constrained version of quantile regression to avoid the crossing problem for both linear and nonparametric quantile curves. We focus our analysis on the so called semi-parametric Quantile Regression models, specifically on the Conditional Autoregressive Value at. Risk (CAViaR) Quantile Regression model specification introduced by [5], and its multi quantile extension. The existing methods for the non crossing that are applicable to semi-parametric regression models often force the estimate procedure to be step by step, thus less efficient than a joint estimation, and, mostly, these methods do not guarantee that the estimated parameters belong to a ”non crossing” parametric space. We believe that in the case of a regression model for which we have parametrical assumption, the quantile crossing issue can be dealt with more efficiently, moreover we find that ignoring the parametric conditions related to the non-crossing property can lead to serious estimation errors.

2 Non-Crossing MQ-CAViaR as a constrained problem

We show in this chapter how to include the non crossing conditions in the estimation problem as linear constraints to the regression parameters. For the assumptions necessary for the correct specification and estimation of the models we refer to [8]. We consider here a *Symmetric Caviar* specification for two different quantile levels:

$$q_{t_i,\tau_A} = w_A + \phi_A q_{t_{i-1},\tau_A} + \gamma_A |x_0|, \tag{1}$$

$$q_{t_i,\tau_B} = w_B + \phi_B q_{t_{i-1},\tau_B} + \gamma_B |x_0|. \tag{2}$$

Let $\tau_A > \tau_B$, the non crossing condition would be, trivially, $q_{t_i,\tau_A} > q_{t_i,\tau_B}$, $i = 1,2,...,T$. Putting directly this condition as a constraint would not be efficient, contrariwise we consider the assumption made on the stochastic process X_t in order to find necessary conditions on the parameters that satisfy the non crossing issue. Let consider the time series y_t with $t = 1,\dots,T$ and a grid of p different quantile levels, the estimation problem is:

$$\begin{aligned}
&\min_{\omega,\gamma,\phi} \ \textstyle\sum_{j=1}^{p} \sum_{t=1}^{T} w_{t,\tau_j}(y_t - q_{t,\tau_j}) \\
&\quad \text{s.t.} \\
&\qquad \gamma_{\tau_j} \leq \gamma_{\tau_{j+1}}, \quad \forall j \\
&\qquad \phi_{\tau_j} \geq \phi_{\tau_{j+1}}, \quad j < j^* \\
&\qquad \phi_{\tau_j} \geq 0, \quad \forall j
\end{aligned} \tag{3}$$

where w_{t,τ_j} is the check function for the quantile level τ_j at time t, $\tau \in [0,1]$ levels are sorted so that $\tau_j < \tau_{j+1}$, and j^* corresponds to τ^*, the quantile level where the autoregressive terms reach their possible minimum ϕ^*. It is necessary to specify that not all the conditions are included in the constraints of the estimation problem, it remains a necessary condition: $w_A(\sum_{1=0}^{n-1}\phi_A^i) - w_B(\sum_{1=0}^{n-1}\phi_B^i) + (\phi_A^n A_0 - \phi_B^n B_0) > 0$, where n is the number of the observations. This last condition can be written more simply, however ensuring this relationship for each integer n means adding a non linear constrain to the estimation problem. Following the tests carried out we choose to proceed this way: we first solve the previous linear constrained problem, then we check if the estimated parameters also meet the non included conditions. If they don't, we estimate the fully constrained problem. Most of the time we did not need to repeat the estimation, thus we believe this way can be more efficient instead of solving the fully constrained problem. It is worth noting that in the "not fully constrained" problem we don't check if our estimated quantiles exhibit crossing to verify the requirement of monotonicity for CDF, whose absence would be just a necessary condition for the monotonicity requirement, instead we check the estimated parameters whose conditions are necessary and sufficient for the non-crossing (monotonicity) requirements. Indeed, forcing the non crossing condition with the constraints $q_{t_i,\tau_A} > q_{t_i,\tau_B}$, or adjusting the estimates that exhibits crossing in this sense, does not guarantee that the non-crossing conditions are satisfied. In the case in which a model with a complex parametric structure is used, for which the non crossing conditions are not exactly known we recommend using simulations of the estimated model to ensure that no crossing occurs instead of just checking that the estimates does not exhibits crossings.

3 The Graphical model: a Sparse Gaussian Copula VAR

A *Graphical Model* exploits the graph theory as well as the statical theory to describe the dependency and conditional dependence ratios of a set of random variables. Nowadays the literature about Graphical Models is huge ,a wide but not exhaustive overview on the subject is [6]. Beyond the obvious advantages of being able to graphically represent dependency structures, we must be able to say something about the dependency and the conditional dependence of random variables. Formally, if X; Y; Z are continuous random variables which admit a joint distribution, we say that X is conditionally independent of Y given Z , and write $X \perp Y|Z \iff f_{XY|Z}(x,y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z)$. A graph $G = (V,E)$ is a *conditional independence graph* if it respects this property. In order to exploit the property of conditioned independence we will use the Gaussian copula tool as a graphical model (Gaussian Graphical Model). The Copula approach allows to model multivariate associations and dependencies separately from the univariate marginal distributions of the observed variables. The theorem by [10] shows that every multivariate distribution can be represented in terms of a copula function, which *couples* the univariate marginal distributions, i.e. $F(Y_1,\ldots,Y_d) = C(F_1(Y_1),\ldots,F_d(Y_d))$. In order to exploit the property of conditioned independence we will use the Gaussian copula tool as a graphical model (Gaussian Graphical Model), the we assume the copula function to be: $C(Y_1,\ldots,Y_d) = \Phi_d(\Phi^{-1}(F_1(Y_1)),\ldots,\Phi^{-1}(F_d(Y_d)))$, where Φ^{-1} is the univariate standard Gaussian quantile function, and Φ_d is an n-variate Gaussian CDF with mean 0_p and covariance matrix P. Moreover , in the light of the complex relationships that link financial returns we want to describe the conditional dependence of the random variables *between* and *within* time. We choose a VAR specification for the multivariate Gaussian Copula-regression model, thus the structure of the model is of the type: $Y = XB + E$. Where Y denotes the $n \times q$ random response matrix X represents a $n \times k$ regression coefficient matrix containing lagged values of Y, B is the $k \times q$ regression matrix and E is the $n \times k$ error terms matrix. Under Gaussian assumption the multivariate series E is assumed to be distributed as a $N(0,\Sigma)$. As it is known, the estimation of multivariate regression matrices can lead to numerical problems especially for high number of marginals. We employ here the MRCE (multivariate regression with covariance estimation) methodology proposed by [11] which consists in a joint estimation of the parameters in B and in Σ, adding two penalties to the negative log-likelihood function g to obtain sparse estimates for both the matrices. The penalty is of the Lasso type, the penalized estimation problem for $(\hat{B},\hat{\Omega})$ is:

$$\min_{\Omega,\Pi}\{g(B,\Omega) + \lambda_1 \sum_{j' \neq j} |\omega_{j',j}| + \lambda_2 \sum_{j=1}^{p} \sum_{i=1}^{q} |\phi_{j,k}|\} \tag{4}$$

where $\lambda_1 \geq 0, \lambda_2 \geq 0$ are the tuning parameters, $\omega_{j',j}$ are entries of Ω^{-1} and $\phi_{j,k}$ are entries of B. The problem in 4 is not convex, but solving for either Ω or P with the other fixed is a convex problem. Then [11] proposed a cyclical-coordinate descent algorithm for an efficient computation, the authors also implemented an Rpackage for *MRCE* methodology called MRCE. The MRCE methodology allows us to get sparse estimates of the parameters of a Copula VAR model, this will improve the interpretability of the results and the stability of the predictions.

4 Empirical application

We collected 4000 daily observation of 24 USA banking institutions. We consider a grid of 103 quantile levels $\tau = 0.0001, 0.001, 0.01, 0.02, \ldots, 0.98, 0.99, 0.999, 0.9999$ and for these levels we estimate a symmetric Non Crossing-CAViaR model for each series by solving the constrained problem and checking the non crossing parametric condition as explained in section 2. Then we can use the estimated multiple conditional quantile to get the estimation $\hat{F}(y_t)$ for $t = 1, \ldots, T$. To get the estimate $\hat{F}(y_t)$ we can proceed directly using the 103 estimated quantile at time t to get the estimated probability $\hat{u}_t = \hat{F}(y_t)$ by linear interpolation. Alternatively we can try to get more precision by estimating $\hat{F}(y_t)$ with a smoothing spline.

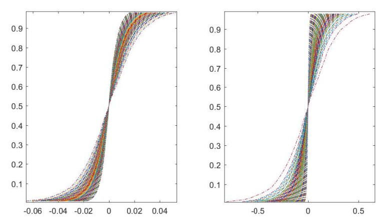

Fig. 1 Some of the time conditional density functions estimated on 103 quantile levels. A period of low volatility in the left graph,a period of high volatility in the right.

Once obtained the series $u_{i,t} = F_{i,t}^{-1}(Y_{i,t})$ with $t = 1, \ldots, 4000$ and $p = 1, \ldots, 24$, we can proceed with the estimation of the parameters of the Copula-VAR function employing the MRCE algorithm to get sparse estimates. We choose a Lag-2 VAR specification and we use the cross validation for the chose of the LASSO shrinking parameters λ_1 and λ_2. Fig. 1 shows some of the estimated cumulative marginal distribution, in Fig.2 the estimated Sparse Graphical model.

5 Conlusions and Discussion

We have identified the parametric space that ensures the non crossing condition for some of the models belonging to the CAViaR specification: this allow the possibility of jointly estimating a considerable number of semi parametric quantile regression models and use them to get non parametric estimation of the marginals for a Copula model. Despite the advantages of knowing the exact parametric space that ensure non-crossing, depending on the parametric assumption of the Multiple Quantile re-

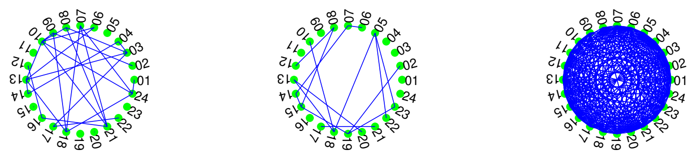

Fig. 2 A circular net representation of the estimated Copula-VAR parameters. From left to right: the regression parameters for the LAG 1, the regression parameters for the LAG 2, and the estimated correlation matrix. Note that the first two networks are intended to be directional.

gression model the constraints to the estimation problem can make the computational part very hard, it would be useful to find algorithms that can efficiently satisfy these conditions. Finally, it is possible to enrich the analysis by adding in the specifications of the models some additional exogenous variable, maintaining the property of non-crossing. The Multivariate part can also be extended by adding exogenous variables, in this case maintaining the same proposed estimation method.

References

1. Bernardi, Mauro and Gayraud, Ghislaine and Petrella, Lea and others:Bayesian tail risk interdependence using quantile regression. Bayesian Analysis & International Society for Bayesian Analysis (2015)
2. Bondell, Howard D and Reich, Brian J and Wang, Huixia: Non-crossing non-parametric estimates of quantile curves. Biometrika & JSTOR (2010)
3. Chernozhukov, Victor and Fernández-Val, Iván and Galichon, Alfred: Quantile and probability curves without crossing. Econometrica & Wiley Online Library (2010)
4. Dette, Holger and Volgushev, Stanislav: Non-crossing non-parametric estimates of quantile curves. Journal of the Royal Statistical Society & Wiley Online Library (2008)
5. Engle, Robert F and Manganelli, Simone: CAViaR: Conditional autoregressive value at risk by regression quantiles. Journal of Business & Economic Statistics & Taylor & Francis (2004)
6. Koller, Daphne and Friedman, Nir: Probabilistic graphical models: principles and techniques. MIT press (2009)
7. Takeuchi, Ichiro and Furuhashi, Takeshi: Non-crossing quantile regressions by SVM. MIT press (2009)
8. White Jr, Halbert L and Kim, Tae-Hwan and Manganelli, Simone: Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR. ECB Working Paper, (2008)
9. Wu, Yichao and Liu, Yufeng: Stepwise multiple quantile regression estimation using non-crossing constraints. Statistics and Its Interface (2009)
10. Sklar, M: Fonctions de répartition á n dimensions et leurs marges. Université Paris 8 (1959)
11. Rothman, Adam J and Levina, Elizaveta and Zhu, Ji : Journal of Computational and Graphical Statistics. & Taylor & Francis (2010)