

The Research of Credit Risk of Corporate Bonds Based on Composite Quantile Regression

Changqing Liu

Department of Mathematics, Computer & Information Engineering, Baise University, Baise
Email: zpx2@163.com

Received: Jan. 10th, 2014; revised: Feb. 12th, 2014; accepted: Feb. 25th, 2014

Copyright © 2014 by author and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Abstract

In this paper, we establish the model, which changed the variance equation of the GARCH model with logarithmic transformation, to estimate the volatility of credit spread of corporate Bonds. It can ensure the independent identical distribution of model error. Also, the volatility is not negative after the logarithmic inverse transform. Owing to the asymmetry of model error, we use composite quantile regression, which is a more robust method, to estimate the modified GARCH model. The empirical analysis shows that the estimation of volatility is more effective with the modified model. Composite quantile regression is a very robust estimation method, and it is more effective to overcome the non-normality of the model error.

Keywords

Credit Risk of Corporate Bonds; Composite Quantile Regression; GARCH Model

基于复合分位数回归的股市风险研究

柳长青

百色学院, 数学与计算机信息工程系, 百色
Email: zpx2@163.com

收稿日期: 2014年1月10日; 修回日期: 2014年2月12日; 录用日期: 2014年2月25日

摘要

本文对GARCH模型的方差方程进行对数变换,得到的模型用于估计股市波动率。这样不仅能保证模型误差的独立同分布性,而且在做对数逆变换后也能保证波动率的非负性。针对模型误差非对称性,利用更为稳健的复合分位数回归方法估计变换后的GARCH模型。实证分析表明:变换后的模型对波动率的估计更为有效,并且复合分位数回归可以更有效的克服模型误差非正态性影响,是一种非常稳健的估计方法。

关键词

股市风险; 复合分位数回归; GARCH模型

1. 引言

股票市场投资是高风险投资,股票的价格和波动率是股票市场中的主要统计指标,对其进行研究可以分析股票市场的风险度量,有助于投资者认识股市风险,进而有助于投资者做出正确的投资决策。股票的价格是随着风险的变化而不断变化的。仅依靠传统直接比率分析的方法在分析风险时就缺乏一定的科学性[1]。波动率是解释股票风险一个重要因素,为了规避风险,了解风险特征,使投资者获取更大收益,对股票波动率准确估计与预测对风险控制具有极其深远的意义。但由于股票价格等金融时间序列数据往往是不平稳的,样本分布尖峰厚尾,异常值过多等特点,不满足 Gauss-Markov 条件,因此无论是利用极大似然估计或最小二乘估计处理模型稳健性都会很差,结果误差也很大。

分位数回归本身稳健性的特点决定了在处理样本尖峰厚尾,异常值过多的情形能发挥更大的优势。目前已有大量文献利用分位数回归方法研究经济时间序列波动率的问题[2] [3]。而复合分位数回归[4]则是分位数回归方法的一种更一般的推广形式,是一种更一般化的估计方法,它依据因变量的多个条件分位数对自变量进行回归,能够精确描述自变量对于因变量的条件变化以及条件分布形状的影响,而且不需要对误差项的分布做出任何具体假定。最近关于复合分位数回归的研究已有大量的文献。比如文献[5]对非参数回归模型提出了一个局部复合分位数回归统计推断方法。文献[6]对非线性回归模型则提出了一个加权的复合分位数回归统计推断方法。但是利用复合分位数回归方法研究经济时间序列波动率的问题目前还没有相关文献进行讨论。

基于此,本文引进复合分位数回归技术估计股票波动率。此外在处理过程中,本文对 GARCH 模型的方差方程进行对数变换,这样不仅保证了模型误差的独立同分布性,而且取对数逆变换后也能保证波动率的非负性。实证分析表明,利用复合分位数回归的方法估计变换后的模型,异常值的敏感程度更低,估计效果更好。

2. 研究方法 with 模型建立

2.1. 复合分位数回归基本原理

复合分位数回归最初是由 Zou 和 Yuan[4]提出的一种估计方法,其可以利用因变量的多个条件分位数对自变量做回归,得到其最小估计值。相比较极大似然或最小二乘估计,复合分位数回归不需要对模型误差的分布做出假定,应用条件更为宽松,挖掘的统计信息也更为丰富。

复合分位数回归可如下描述:考虑一组随机样本为 (y_i, x_i) , $i = 1, \dots, n$, 其中 y_i 为响应变量, x_i 为对应应用解释变量, 并且假定满足如下线性关系:

$$y_i = x_i \beta + \varepsilon_i, i = 1, \dots, n$$

其中 ε_i 为模型误差，并且有 $P(\varepsilon_i \leq b_k | x_i) = \tau_k, k = 1, \dots, K$ 。对任意的 $0 < \tau_k < 1$ ，定义检验函数为 $\rho_{\tau_k}(u) = u(\tau_k - I_{(u < 0)})$ ， $I_{(u < 0)}$ 为示性函数。则参数 β 的复合分位数回归估计为

$$(\hat{b}_1, \dots, \hat{b}_K, \hat{\beta}) = \arg \min \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k}(y_i - b_k - x_i \beta) \quad (1)$$

对模型(1)可以采用单纯形法求解，但对于大规模问题其收敛速度还是相对较慢。文献[7]提出了一种 Majorize-Minimize(MM)算法适应大规模问题求解。因此，本文基于 MM 算法对(1)式求最优解。

2.2. 模型建立

设 P_t 表示某种股票第 t 个投资期的收盘价，相应的对数收益为 $r_t = \log(P_t) - \log(P_{t-1})$ 。

考虑一般化模型

$$r_t = \mu_t + a_t \quad (2)$$

其中 $\mu_t = E(r_t | \Omega_{t-1})$ 为形如 $g(t)$ 等一般的未知非参数函数， Ω_{t-1} 表示 $t-1$ 之前的已有信息， a_t 表示随机扰动，并且有

$$a_t = \sigma_t \varepsilon_t \quad (3)$$

其中 $\varepsilon_t \sim N(0,1)$ 为高斯白噪声过程，则波动率为 $\sigma_t^2 = \text{Var}(r_t | \Omega_{t-1})$ 。取 $a_t^2 = \sigma_t^2 \varepsilon_t^2$ ，如果 σ_t^2 存在 ARCH 效应，可用多元 GARCH(p, q)模型 $\sigma_t^2 = f(a_{t-1}^2, \dots, a_{t-p}^2, \sigma_{t-1}^2, \dots, \sigma_{t-q}^2)$ 表示。为简化此问题，考虑如下可加参数模型：

$$\sigma_t^2 = c + \sum_{i=1}^p \phi_i a_{t-i}^2 + \sum_{j=1}^q \theta_j \sigma_{t-j}^2,$$

记 $v_t = (c + \sum_{i=1}^p \phi_i a_{t-i}^2 + \sum_{j=1}^q \theta_j \sigma_{t-j}^2)(\varepsilon_t^2 - 1)$ ，那么则有

$$a_t^2 = (c + \sum_{i=1}^p \phi_i a_{t-i}^2 + \sum_{j=1}^q \theta_j \sigma_{t-j}^2) + v_t \quad (4)$$

注意到 v_t 分布是非对称的且存在异方差。为了保证误差项满足独立同分布，对(3)式两边平方取对数可得：

$$\log(a_t^2) = \log(\sigma_t^2) + \mu_t + [\log(\varepsilon_t^2) - \mu_t] = \log(\sigma_t^2) + \mu_t + e_t \quad (5)$$

这里 $e_t = \log(\varepsilon_t^2) - \mu_t$ ， $\mu_t = E(\log \varepsilon_t^2)$ 。由于 e_t 仍然是非对称的，利用极大似然估计估计结果不能保证其有效性。利用复合分位数回归的估计方法不需要对误差分布做假定，对异常值敏感度更低，结果更为稳健，因此接下来考虑利用复合分位数回归方法求解(5)式。类似的，对 $\log(\sigma_t^2)$ 定义

$$\log(\sigma_t^2) = c + \sum_{i=1}^p \phi_i a_{t-i}^2 + \sum_{j=1}^q \theta_j \sigma_{t-j}^2,$$

进行对数逆变换得到

$$\sigma_t^2 = \exp\left\{c + \sum_{i=1}^p \phi_i a_{t-i}^2 + \sum_{j=1}^q \theta_j \sigma_{t-j}^2\right\} \quad (6)$$

由(6)式可以看出对方差方程做对数变换还可保证波动率的非负性。将(6)带入(5)式即得到如下对数形式的 GARCH 模型：

$$\log(a_t^2) = c + \sum_{i=1}^p \phi_i a_{t-i}^2 + \sum_{j=1}^q \theta_j \sigma_{t-j}^2 + \mu_t + e_t \quad (7)$$

2.3. 模型估计算法

对波动率的求解，我们采用复合分位数回归的方法，对(7)式进行复合分位数回归，并且采用迭代向后拟合算法[8]处理模型中的可加部分。具体方法如下：

第一步：用极大似然方法对波动率做参数 GARCH 估计，进而得到波动率 σ_t^2 的初估计，记为 $\{\hat{\sigma}_{t,0}^2 : 1 \leq t \leq n\}$ 。

第二步：取迭代次数 $m=1$ ，做 a_t^2 关于 $a_{t-1}^2, \dots, a_{t-p}^2$ 和 $\hat{\sigma}_{t-1,m-1}^2, \dots, \hat{\sigma}_{t-a,m-1}^2$ 的向后拟合回归，这里取 c, φ_i, θ_j 的初估计 $\hat{c}_0, \hat{\varphi}_{i,0}, \hat{\theta}_{j,0}$ 为 0, $i=1, \dots, p, j=1, \dots, q$ 。运用复合分位数回归方法计算 $c_m, \varphi_{i,m}, \theta_{j,m}$ 的估计值。

$$\hat{c}_m^{(s)} = \arg \min \sum_{k=1}^K \sum_{t=1}^n \rho_{\tau_k} \left(\log(a_t^2) - \sum_{i=1}^p \hat{\varphi}_{im}^{(s-1)} a_{t-i}^2 - \sum_{j=1}^q \hat{\theta}_{jm}^{(s-1)} \hat{\sigma}_{t-j}^2 - \mu - c_m \right) \quad (8)$$

$$\hat{\varphi}_{im}^{(s)} = \arg \min \sum_{k=1}^K \sum_{t=1}^n \rho_{\tau_k} \left(\log(a_t^2) - \sum_{i \neq i} \hat{\varphi}_{im}^{(s-1)} a_{t-i}^2 - \sum_{j=1}^q \hat{\theta}_{jm}^{(s-1)} \hat{\sigma}_{t-j}^2 - \mu - \varphi_{im} a_{t-i}^2 \right) \quad (9)$$

$$\hat{\theta}_{jm}^{(s)} = \arg \min \sum_{k=1}^K \sum_{t=1}^n \rho_{\tau_k} \left(\log(a_t^2) - \sum_{i=1}^p \hat{\varphi}_{im}^{(s-1)} a_{t-i}^2 - \sum_{l \neq j} \hat{\theta}_{lm}^{(s-1)} \hat{\sigma}_{t-l}^2 - \mu - \theta_{jm} \hat{\sigma}_{t-l}^2 \right) \quad (10)$$

反复对(8)~(10)式迭代计算直到足够的精度，即得到 $c_m, \varphi_{i,m}, \theta_{j,m}$ 的向后拟合估计： $\hat{c}_m^*, \hat{\varphi}_{i,m}^*, \hat{\theta}_{j,m}^*$ 。

第三步：计算

$$\hat{\sigma}_{t,m}^2 = \exp \left\{ \hat{c}_m + \sum_{i=1}^p \hat{\varphi}_{i,m} a_{t-i}^2 + \sum_{j=1}^q \hat{\theta}_{j,m} \hat{\sigma}_{t-j,m-1}^2 \right\}$$

并且取 $m=m+1$ ，将估计值 $\hat{\sigma}_{t,m}^2$ 返回到第二步，循环 N 次，当 $|\hat{\sigma}_{t,N}^2 - \hat{\sigma}_{t,N-1}^2| < \eta$ 时停止循环，其中 η 为充分小的正数。此时相应的复合分位数回归的最终估计记为 $\hat{c}_N, \hat{\varphi}_{i,N}, \hat{\theta}_{j,N}$ 。则最终得到股票波动率的估计为

$$\hat{\sigma}_t^2 = \exp \left\{ \hat{c}_N + \sum_{i=1}^p \hat{\varphi}_{i,N} a_{t-i}^2 + \sum_{j=1}^q \hat{\theta}_{j,N} \hat{\sigma}_{t-j,N-1}^2 \right\}$$

3. 实证分析

3.1. 样本选取和基本统计信息

本文研究对象为上海证券综合指数和深圳成分指数的收益序列。选取 2000 年 1 月 28 日到 2011 年 12 月 30 日的 144 个交易日月末收盘数据。数据来源 Sohu 网页数据库(<http://business.sohu.com/>)。首先对两序列进行正态性检验。如下表 1 列出了两个收益序列的偏度、峰度以及 Bera-Jarque 检验统计量的结果。

由表 1 可以看出，两个序列均为正偏，这表明有较多的低于自身均值的收益率。另外其峰度值均大于 3，表现出明显的尖峰厚尾性。统计量的检验结果在 1% 的显著性水平上拒绝了收益率序列的正态性假定。

Table 1. The basic information of returns of securities

表 1. 证券收益序列的基本统计信息

证券简称	标准差	峰度	偏度	统计量 T
上证指数	0.03064	5.36353	1.33482	17.70
深成指数	0.03661	5.59504	1.69232	18.79

3.2. 收益时序模型平稳性和条件异方差性检验

对上述上海综合指数和深圳成分指数的收益序列利用带有截距项不带趋势项显著性水平 1%(T 统计量临界值约为-3.4574)的 ADF 检验验证其平稳性。如表 2 所示两收益序数据均为平稳的。利用极大似然估计方法对上述两序列做非参数回归,通过反复对比观察,选取的相应非参数自回归模型拟合结果以及模型检验结果见表 2。结果显示所有的残差序列平稳性检验显著,自相关和偏相关系数分析都没有截尾和拖尾存在。这表明利用非参数 1 阶自回归模型拟合该两收益序列是合适的。

另外,表 2 还给出了对两收益序列的残差序列的异方差性做 LM-ARCH 检验的结果。结果表明这两个收益序列均具有显著的条件异方差特性。对于有明显 ARCH 效应的扰动项我们利用 2.3 节提出的方法拟合波动率

3.3. 波动率的估计与评价

接下来对两个收益序列的残差序列分别用模型(4)和模型(7)建立 GARCH 模型,并且利用 2.3 节给出的复合分位数回归方法分别做估计。由于无法观测到真实的波动率,因此无法用估计的波动率与真实波动率进行比较。对此我们采用 Peter 和 Alexander[8]提出的量化指标来对比其优化程度:

$$L_2 = \frac{1}{n} \sum_{i=1}^n (\hat{\sigma}_i^2 - a_i^2)$$

经过反复计算以及比较各类 GARCH 模型的 AIC 值,最终得出的结果是上海综合指数收益序列的差残差序列选取的模型滞后阶数为(1,1)时最优,深圳成分指数收益序列的差残差序列选取的模型滞后阶数为(2,1)时最优。如下表 3 给出了分别基于模型(4)与模型(7)的 L_2 值,在复合分位数回归方法中,取分位点分别为 $\tau_1 = 0.25$, $\tau_2 = 0.5$ 和 $\tau_3 = 0.75$ 。由表 3 可以看出,相对于模型(4),基于模型(7)拟合波动率的逼近程度更高一些。另外还可以看出基于模型(4)与模型(7)利用复合分位数回归得到 L_2 值都比较小,这也表明复合分位数回归可以有效的克服模型误差非正态影响,对异常值的敏感程度更低,是一种非常稳健的估计方法。

4. 结论

复合分位数回归方法是经典分位数回归方法的一个有意义的推广形式,该方法为计算时序波动率提

Table 2. The estimation of model parameters and the test of residual sequences

表 2. 模型参数估计与残差序列检验

证券简称	参数估计			残差估计检验				
	ADF 检验	最优模型拟合	残差 ADF 检验	LM ARCH 检验	一阶自相关	二阶自相关	一阶偏自相关	二阶偏自相关
上证指数	-3.4787	AR(1)	-173353**	0.0000**	-0.053	0.042	0.067	-0.027
深成指数	-2.9288	AR(1)	-15.3962**	0.0000**	0.052	0.038	-0.056	-0.023

注: **表示在 1%显著性水平下显著。

Table 3. The comparison of the two estimation methods based model (4) and model (7)

表 3. 基于模型(4)与模型(7)的两种估计方法比较

证券简称	基于模型(4)的 L_2 值	基于模型(7)的 L_2 值	最优拟合方法(下降比)
上证指数	0.0052676	0.0047258	模型(7) (10.28%)
深成指数	0.0072486	0.0066955	模型(7) (7.63%)

提供了一个新框架。基于复合分位数回归方法，可以直接对任意水平的条件分位点进行建模，而不依赖于特定的分布形式，更适用于厚尾分布数据的应用。本文选取 GARCH 模型的方差方程做对数变换，得到的模型用于估计股票收益波动率。这样不仅能保证模型误差的独立同分布性，而且在做对数逆变换后也能保证波动率的非负性。

另外，针对模型误差非对称性，本文利用更为稳健的复合分位数回归方法估计变换后的 GARCH 模型。实证分析表明，变换后的模型对波动率的估计更为有效。

基金项目

广西高校科研立项项目(2013LX147)。

参考文献 (References)

- [1] 戴国强, 孙新宝 (2011) 我国企业债券信用利差宏观决定因素研究. *财政研究*, **12**, 63-71.
- [2] 王新宇, 赵绍娟 (2008) 基于分位数回模型的沪深股市风险测量研究. *中国矿业大学学报*, **3**, 416-421.
- [3] 张钰 (2008) 基于分位数回模型的证券市场风险研究. *统计与决策*, **3**, 416-421.
- [4] Zou, H. and Yuan, M. (2008) Composite quantile regression and the oracle model selection Theory. *The Annals of Statistics*, **36**, 1108-1126.
- [5] Kai, B., Li, R. and Zou, H. (2010) Local composite quantile regression smoothing: An efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society Series B*, **72**, 49-69.
- [6] Jiang, X.J., Jiang, J.C. and Song, X.Y. (2012) Oracle model selection for nonlinear models based on weighted composite regression. *Statistica Sinica*, **22**, 1479-1506.
- [7] Hunter, D.R. and Lange, K. (2000) Quantile Regression via an MM Algorithm. *Journal of Computational and Graphical Statistics*, **9**, 60-77.
- [8] Peter, B. and Alexander, J.M. (2002) An algorithm for nonparametric GARCH modeling. *Computational Statistics & Data Analysis*, **40**, 665-683.