高校应用数学学报
2024, 39(1): 1-12

基于经验似然方法对分位数相关系数的区间估计

唐松乔, 李康强, 李 翔, 张立新

(浙江大学 数学科学学院, 浙江杭州 310058)

摘 要: 分位数相关系数是一种度量两个随机变量之间线性相关关系的非对称相关系数, 在统计, 金融, 化学等领域的特征选择问题中都扮演着重要的作用. 同时, 经验似然作为一种非参数方法, 被广泛应用于各类模型的统计推断问题. 对于任意两个随机变量, 从分位数相关系数的定义出发, 建立估计方程, 引入代入经验似然方法(PEL)和其修正版本(APEL), 分别得到渐近规则化的卡方分布和标准卡方分布, 从而得到分位数相关系数的区间估计. 数值模拟部分从覆盖概率, 置信区间长度和基于区间得分的平均损失三方面比较了两种基于经验似然的方法同其他已有方法的效果. 实证分析部分将提出的方法应用于一项来自福布斯排行榜的数据集.

关键词: 分位数相关系数; 经验似然; 区间估计

中图分类号: O212

文献标识码: A **文章编号**: 1000-4424(2024)01-0001-12

§1 引 言

分位数相关系数最早由文[1]提出, 它是一种可以在任意给定分位数水平$\tau \in (0,1)$处度量任何两个随机变量之间线性相关关系的非对称相关系数. 分位数相关系数被广泛应用于各种特征筛选和变量选择问题中. 例如, 文[2]开发了一种新的基于分位数相关系数的确定独立筛选程序(SIS), 能够解决超高维度下的稳健筛选和无模型筛选问题. 文[3]将这项工作进一步扩展, 解决了超高维回归下每个预测因子的相对重要性排序问题. 在生存分析中, 文[4]使用基于分位数相关系数的筛选指标来处理右删失数据. 文[5]提出了分位数相关系数的一种变体, 用于研究存在删失数据的分位数回归问题. 此外, 分位数相关系数在金融和化学等领域也有广泛的应用. 例如, 文[6]利用分位数相关系数研究了政治风险和墨西哥金融市场之间的非参数关系. 文[7]基于分位数相关系数采用稳健的, 无模型的特征筛选方法来选择一个挥发性化合物子集. 然而, 很少有研究关注分位数相关系数的统计推断问题, 而这一问题恰恰对于研究变量选择尤为重要.

收稿日期: 2022-12-01 修回日期: 2023-06-20
基金项目: 国家自然科学基金(U23A2064; 12031005)

经验似然是一种非参数统计推断方法, 最早由文[8-9]引入, 用来建立总体均值的置信域. 许多学者将这一思想推广到其他各种领域. 例如, 文[10]将其用于线性回归, 文[11]将其用于分位数, 文[12-13]将其用于M函数和分位数回归. 经验似然方法有许多吸引人的优点, 如域保留性, 变换不变性, 巴特利特校正, 无需估计协方差矩阵就可以轻松获得置信区间, 等等. 更多关于经验似然方法的文献可以参考文[14-15]. 近年来, 许多基于经验似然的方法被应用于各种相关系数. 例如, 文[16]将其应用于Pearson相关系数, 文[17-18]则将其应用于基尼相关系数. 然而, 当随机变量服从重尾分布或其他非正态分布时, 这些系数, 如Pearson相关系数, 并不能提供关于原始数据足够的信息. 而分位数相关系数很好地弥补了这一缺陷. 据了解, 目前为止尚没有基于经验似然的方法对分位数相关系数建立置信区间.

为了填补这一空白, 本文引入了代入经验似然(PEL)方法和它的修正版本(APEL)来进行区间估计. 在PEL方法中, 通过构建一个基于分位数相关系数定义的估计方程, 得到了一个规则化的卡方分布. 然而, 这个估计方程并不是最有效的, 所以推导出的渐近分布不是标准的卡方. 因此进一步引入APEL方法修正了这个估计方程, 得到了标准的卡方分布. 应用这两种方法非常方便, 且均不需要正态分布和对称分布的假设.

本文的组织结构如下. §2介绍了PEL方法和APEL方法, 并构建了相应的渐近理论结果. §3进行了数值模拟研究, 以评估所提出方法的效果. §4给出了一个实证分析. §5进行了总结和展望. 所有的理论证明都被放到了附录中.

§2 方法及理论结果

分位数相关系数的定义为

$$\mathrm{qcor}_\tau\{Y,X\}=\frac{\mathrm{qcov}_\tau\{Y,X\}}{\sqrt{(\tau-\tau^2)\sigma_X^2}},\tag{1}$$

其中$\mathrm{qcov}_\tau\{Y,X\}=\mathrm{E}\{\psi_\tau(Y-Q_{\tau,Y})(X-\mu_X)\}$是分位数协方差, $\psi_\tau(\omega)=\tau-I(\omega<0)$, $Q_{\tau,Y}$是Y的第τ无条件分位数. 由定义显然有$-1\le\mathrm{qcor}_\tau\{Y,X\}\le 1$. 令$Z_i=(X_i,Y_i), i=1,\cdots,n$, 是来自总体$(X,Y)$的独立同分布的样本. 从(1)中的定义可以得到估计

$$\mathrm{E}\left(\frac{\psi_\tau(Y-Q_{\tau,Y})(X-\mu_X)}{\sqrt{(\tau-\tau^2)\sigma_X^2}}-\mathrm{qcor}_\tau\{Y,X\}\right)=0.$$

可以对$\mathrm{qcor}_\tau\{Y,X\}$进行常规的经验似然构造

$$L(\mathrm{qcor}_\tau\{Y,X\})=\sup\left\{\prod_{i=1}^n p_i:p_i\ge 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i(V_i-\mathrm{qcor}_\tau\{Y,X\})=0\right\},\tag{2}$$

这里$V_i=\frac{\psi_\tau(Y_i-Q_{\tau,Y})(X_i-\mu_X)}{\sqrt{(\tau-\tau^2)\sigma_X^2}}$. 尽管总体均值$\mu_X$, 总体标准差$\sigma_X$和总体第$\tau$无条件分位数$Q_{\tau,Y}$是未知的, 但可以用$\bar{X},S_X,\widehat{Q}_{\tau,Y}$对它们分别进行估计; 这里

$$\bar{X}=n^{-1}\textstyle\sum_{i=1}^n X_i, S_X=\sqrt{n^{-1}\sum_{i=1}^n(X_i-\bar{X})^2}, \widehat{Q}_{\tau,Y}=\inf\{y:F_n(y)\ge\tau\}$$

代表$Y_1,\cdots,Y_n$的样本第τ分位数, $F_n(y)=n^{-1}\sum_{i=1}^n I(Y_i\le y)$表示经验分布函数. 将(2)中的总体版本用这些估计代替, 可以得到$\mathrm{qcor}_\tau\{Y,X\}$的代入经验似然构造(PEL)

$$\widehat{L}(\mathrm{qcor}_\tau\{Y,X\})=\sup\left\{\prod_{i=1}^n p_i:p_i\ge 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})=0\right\},$$

这里$\widehat{V}_i=\frac{\psi_\tau(Y_i-\widehat{Q}_{\tau,Y})(X_i-\bar{X})}{\sqrt{\tau-\tau^2}S_X}, i=1,\cdots,n$. 相应地, 在$\mathrm{qcor}_\tau(Y,X)$处的代入经验似然比为

$$R(\mathrm{qcor}_\tau\{Y,X\})=\sup\left\{\prod_{i=1}^n np_i:p_i\ge 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})=0\right\}.$$

利用Lagrange乘子法, 可以得到

$$p_i = \frac{1}{n}\frac{1}{1+\lambda(\widehat{V}_i - \mathrm{qcor}_\tau\{Y,X\})},\ i=1,\cdots,n,$$

这里λ满足

$$f(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_i - \mathrm{qcor}_\tau\{Y,X\}}{1+\lambda(\widehat{V}_i - \mathrm{qcor}_\tau\{Y,X\})} = 0. \tag{3}$$

紧接着便可以得到代入对数经验似然比

$$l(\mathrm{qcor}_\tau\{Y,X\}) = -2\log R(\mathrm{qcor}_\tau\{Y,X\}) = 2\sum_{i=1}^{n}\log\{1+\lambda(\widehat{V}_i - \mathrm{qcor}_\tau\{Y,X\})\}.$$

下面定理说明了该代入对数经验似然比的渐近性质.

定理2.1 假设$\mathrm{E}(X^4)<\infty$, 存在$\pi>0$使得条件密度$f_{Y|X}(\cdot)$在$[Q_{\tau,Y}-\pi, Q_{\tau,Y}+\pi]$上一致可积, 且密度函数$f_Y(\cdot)$是连续且大于0的. 如果$\mathrm{qcor}_\tau\{Y,X\}$是真值, 则有当$n\to\infty$时,

$$A\cdot l(\mathrm{qcor}_\tau\{Y,X\}) \xrightarrow{d} \chi_1^2,$$

这里χ_1^2表示自由度为1的标准卡方分布, 规则化常数$A=\sigma_0^2/\Omega$, 其中

$$\Omega = \frac{1}{\tau-\tau^2}\Big[\frac{\Sigma_{11}(\mathrm{qcov}_\tau\{Y,X\})^2}{4\sigma_X^6} - \frac{\Sigma_{13}\cdot\mathrm{qcov}_\tau\{Y,X\}}{\sigma_X^4} + \frac{\Sigma_{12}}{\sigma_X^2}\Big],$$

$$\sigma_0^2 = \mathrm{var}\Big[\frac{\psi_\tau(Y-Q_{\tau,Y})(X-\mu_X)}{\sqrt{\tau-\tau^2}\sigma_X}\Big],\ \mu_{X|Y} = \mathrm{E}[f_{Y|X}(Q_{\tau,Y})X]/f_Y(Q_{\tau,Y}),$$

$$\Sigma_{11} = \mathrm{E}(X-\mu_X)^4 - \sigma_X^4, \Sigma_{12} = \mathrm{E}[\psi_\tau(Y-Q_{\tau,Y})(X-\mu_{X|Y})]^2 - [\mathrm{qcov}_\tau\{Y,X\}]^2,$$

$$\Sigma_{13} = \mathrm{E}[\psi_\tau(Y-Q_{\tau,Y})(X-\mu_{X|Y})(X-\mu_X)^2] - \sigma_X^2\mathrm{qcov}_\tau\{Y,X\}.$$

利用定理2.1可以对$\mathrm{qcor}_\tau\{Y,X\}$进行统计推断, 但这之前需要先对规则化常数$A$进行估计. 本文利用Nadaraya-Watson回归去估计$m(y)=\mathrm{E}(X|Y=y)$, 估计量记作$\widehat{m}(y)$. 窗宽h的选择使用留一交叉验证. 进一步若随机向量(X,Y)有联合密度, 可以证明$\mu_{X|Y}=\mathrm{E}(X|Y=Q_{\tau,Y})$. 因此可以用$\widehat{\mu}_{X|Y}=\widehat{m}(\widehat{Q}_{\tau,Y})$估计$\mu_{X|Y}$, 这里$\widehat{Q}_{\tau,Y}$是$\{Y_1,\cdots,Y_n\}$的样本$\tau$分位数. 对于$\Omega$中包含的其他量, 包括$\mu_X$, σ_X^2, $\mathrm{qcov}_\tau\{Y,X\}$, Σ_{11}, Σ_{12}和Σ_{13}, 可以同文[1]所做的一样去估计. 最终可以得到Ω的一个估计, 记作$\widehat{\Omega}$. 此外令

$$\widehat{\sigma}_0^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^n\big(\frac{\psi_\tau(Y_i-\widehat{Q}_{\tau,Y})(X_i-\bar{X})}{\sqrt{\tau-\tau^2}S_X} - \frac{1}{n}\sum_{i=1}^n\frac{\psi_\tau(Y_i-\widehat{Q}_{\tau,Y})(X_i-\bar{X})}{\sqrt{\tau-\tau^2}S_X}\big)^2.$$

那么$\widehat{A}=\widehat{\sigma}_0^2/\widehat{\Omega}$就是规则化常数$A$的一个估计. 基于上述讨论, $\mathrm{qcor}_\tau\{Y,X\}$在水平$100(1-\alpha)\%$处的渐近置信区间可构造为

$$\big\{\mathrm{qcor}_\tau\{Y,X\} : \widehat{A}\cdot l(\mathrm{qcor}_\tau\{Y,X\}) \le \chi_1^2(\alpha)\big\},$$

这里$\chi_1^2(\alpha)$表示χ_1^2的上α分位数.

同时可以考虑假设检验问题

$$H_0 : \mathrm{qcor}_\tau\{Y,X\} = x_0 \quad \text{versus} \quad H_a : \mathrm{qcor}_\tau\{Y,X\} \ne x_0, \tag{4}$$

x_0是$[-1,1]$间的一个具体的常数. 如果$\widehat{A}\cdot l(x_0) > \chi_1^2(\alpha)$, 则拒绝原假设.

尽管上述PEL方法易于操作, 但需要先对A进行估计. 为了避免估计这个常数, 接下来将对估计方程做一些修正. 由文[1]的证明, 可以得到

$$\begin{aligned}&\sqrt{n}\Big[\frac{1}{n}\sum_{i=1}^{n}\psi_\tau(Y_i-\widehat{Q}_{\tau,Y})(X_i-\bar{X}) - \mathrm{qcov}_\tau\{Y,X\}\Big]\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi_\tau(Y_i-Q_{\tau,Y})(X_i-\mu_{X|Y}) - \mathrm{qcov}_\tau\{Y,X\}] + o_p(1).\end{aligned} \tag{5}$$

显然有

$$\sqrt{n}(S_X^2-\sigma_X^2)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[(X_i-\mu_X)^2-\sigma_X^2]+o_p(1).$$

使用Delta方法和一些代数运算, 可以得到

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\psi_\tau(Y_i-Q_{\tau,Y})(X_i-\mu_{X|Y})}{\sqrt{\tau-\tau^2}\sigma_X}-\frac{1}{2}\mathrm{qcor}_\tau\{Y,X\}\big[1+\frac{(X_i-\mu_X)^2}{\sigma_X^2}\big]\right\}+o_p(1)\\&\xrightarrow{d}N(0,\Omega).\end{aligned}$$

记$V_i^c(\mathrm{qcor}_\tau\{Y,X\})=\frac{\psi_\tau(Y_i-Q_{\tau,Y})(X_i-\mu_{X|Y})}{\sqrt{\tau-\tau^2}\sigma_X}-\frac{1}{2}\mathrm{qcor}_\tau\{Y,X\}\big[1+\frac{(X_i-\mu_X)^2}{\sigma_X^2}\big]$, 对$\mathrm{qcor}_\tau\{Y,X\}$定义修正代入经验似然(APEL)

$$\widehat{L}_c(\mathrm{qcor}_\tau\{Y,X\})=\sup\left\{\prod_{i=1}^{n}p_i:p_i\geq 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\widehat{V}_i^c(\mathrm{qcor}_\tau\{Y,X\})=0\right\},$$

这里

$$\widehat{V}_i^c(\mathrm{qcor}_\tau\{Y,X\})=\frac{\psi_\tau(Y_i-\widehat{Q}_{\tau,Y})(X_i-\widehat{\mu}_{X|Y})}{\sqrt{\tau-\tau^2}S_X}-\frac{1}{2}\mathrm{qcor}_\tau\{Y,X\}\big[1+\frac{(X_i-\bar{X})^2}{S_X^2}\big]$$

是V_i^c的估计值. 类似地, 利用Lagrange乘子法, 可以得到

$$p_i=\frac{1}{n}\frac{1}{1+\lambda_c\widehat{V}_i^c(\mathrm{qcor}_\tau\{Y,X\})},\ i=1,\cdots,n,$$

这里λ_c满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_i^c(\mathrm{qcor}_\tau\{Y,X\})}{1+\lambda_c\widehat{V}_i^c(\mathrm{qcor}_\tau\{Y,X\})}=0.$$

相应地, 关于$\mathrm{qcor}_\tau\{Y,X\}$的对数经验似然比为

$$l_c(\mathrm{qcor}_\tau\{Y,X\})=2\sum_{i=1}^{n}\log\{1+\lambda_c\widehat{V}_i^c(\mathrm{qcor}_\tau\{Y,X\})\}.$$

此时可以建立该修正的代入对数经验似然比的理论性质.

定理2.2 假设定理2.1中的条件成立. 如果$\mathrm{qcor}_\tau(Y,X)$是真值, 那么当$n\to\infty$时,

$$l_c(\mathrm{qcor}_\tau\{Y,X\})\xrightarrow{d}\chi_1^2.$$

可以看到定理2.1中的常数A在定理2.2中消失了. 这主要是因为从$\mathrm{qcov}_\tau\{Y,X\}$中提取出的式(5)是无偏的, 因此避免了产生修正项, 使得结果更加精确.

利用定理2.2, 可以建立对于$\mathrm{qcor}_\tau\{Y,X\}$在水平$100(1-\alpha)\%$处的渐近置信区间

$$\{\mathrm{qcor}_\tau\{Y,X\}:l_c(\mathrm{qcor}_\tau\{Y,X\})\leq\chi_1^2(\alpha)\}.$$

类似地可以考虑假设检验问题(4), 若$l_c(x_0)>\chi_1^2(\alpha)$, 则拒绝原假设.

注2.1 定理2.1和定理2.2都可应用于一元分位数回归模型$Y=a_0+b_0X+\epsilon$, 这里$a_0=Q_{\tau,Y-b_0X}$. 具体来说, 可以检验是否协变量X与响应变量Y的τ分位数相关, 即

$$H_0:b_0=0\quad\text{versus}\quad H_a:b_0\neq 0.$$

根据文[1, 引理1], 分位数相关系数和分位数回归的斜率参数间有密切的联系. 具体来说, 他们证明了$\mathrm{qcov}_\tau\{Y,X\}=\varrho(b_0)$, 这里$\varrho(b)=\mathrm{E}[\psi_\tau(\epsilon-Q_{\tau,\epsilon+bX}+bX)X]$是一个连续单调递增函数. 特别地, $\varrho(b)=0$当且仅当$b=0$. 这说明检验是否$b_0=0$可以转化为检验是否$\mathrm{qcov}_\tau\{Y,X\}=0$, 而后者相当于在式(4)中令$x_0=0$.

§3 数值模拟

本节将对本文提出方法(PEL和APEL)的有限样本表现进行比较. 此外, 本节也加入了与利用分位数相关系数估计的渐近分布构造的置信区间的比较(NA). 为了体现Pearson相关系数和分位数相关系数的区别, 本节也将本文方法同文[16]提出的IFEL方法进行比较, 他们的工作是基于经验似然方法对Pearson相关系数(corr)做区间估计. 本节对分位数相关系数的真值等于或不等于零的情况均进行了讨论. 具体来说, 本节使用Monte Carlo模拟来计算覆盖概率和95%置信度下的置信区间的平均长度. 此外, 本节采用了文[19]中讨论的损失函数的概念, 通过共同考虑覆盖概率和区间宽度来评估不同的区间. 样本量取$n = 30, 50, 100$. 接下来将在以下例子中分别考察上述方法在三个不同的分位数水平$\tau = 0.2, 0.5, 0.9$处的表现.

- 例1: $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$.
- 例2: $X \sim t_5$, $Y \sim t_5$, X, Y独立.
- 例3: $\begin{pmatrix} X \\ Y \end{pmatrix} \sim 0.8 \times N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) + 0.2 \times N\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$, 这是一个混合正态分布, 其中80%的观测来自于第一个正态分布, 20%的观测来自第二个正态分布.
- 例4: $X \sim N(0, 1)$, $Y \sim t_2$, X, Y独立.
- 例5: $X \sim N(0, 1)$, $Y \sim \chi^2_1$, X, Y独立.
- 例6: $X \sim N(0, 1), Y = X + X_1$, 其中$X_1 \sim N(0, 1)$且$X, X_1$独立.

注3.1 当基础分布为正态分布时, 可以很容易地生成具有特定非零分位数相关系数的数据. 例如对于例6, 有

$$\mathrm{qcor}_\tau\{Y, X\} = \frac{\frac{1}{2\sqrt{\pi}} \mathrm{e}^{-\frac{Q_\tau^2}{4}}}{\sqrt{\tau - \tau^2}},$$

这里Q_τ是N$(0, 2)$的第τ分位数.

对于例1-例5, 分位数相关系数真值为零. 例1考虑标准正态分布, 例3是为了评估基础分布错定时区间估计的表现. 例2和例4考察了基础分布是重尾的情况. 因为分位数方法的一个优点是当基础分布是非对称分布时仍可以很好地工作, 因此例5考虑了卡方分布. 例6则分析了分位数相关系数不为零时的情形.

所有的模拟结果都是基于每个例子的1000次重复, 结果见表1-6. 可以发现, 随着样本量n的变大, 平均长度变得更加精确, 而且大多数情况下的覆盖概率都在0.95左右. 当真实分位数相关系数为零时, 在重尾和不对称分布的情况下, IFEL方法的覆盖概率低于基于分位数相关系数的方法, 见表2, 表4和表5. 对于NA方法, 从平均损失的角度来看, 当取中位点时, 其表现总是差于APEL. 虽然与NA相比, APEL在极端分位点(如0.9)的覆盖概率略小而平均损失略大, 这主要是由于APEL方法的准确度很大程度上依赖于$\mu_{X|Y}$估计的准确度, 而由于N-W估计方法的特性, 在极端分位点处可用的样本量会相对中位点处较少. 从模拟结果中也可以看到, 随着样本量的

增加, 在极端分位点处两者的差距会快速缩小. 而对于分位数相关系数非零时, 从表6可知, 本文提出的两种方法均比IFEL方法表现出明显的优势. 随着样本量增长到100时, 从平均损失的角度来看, APEL方法总是优于NA方法. PEL方法和APEL方法相比, APEL在大多数情况下的平均损失都低于PEL, 见表1, 表3和表6. 当真实分位数相关系数为零且分位数水平较极端时, PEL方法的覆盖概率倾向于低于置信度0.95, 特别是在小样本的情况下. 然而, APEL方法则表现得更加稳健. 总而言之, 本文提出的两种方法均表现出明显的优势, 而APEL方法则比PEL方法更加稳健.

表 1 关于$\mathrm{qcor}_\tau\{Y,X\}$和$\mathrm{corr}\{Y,X\}$区间估计的覆盖概率, 平均长度和平均损失, 例1

		覆盖概率			平均长度			平均损失		
n	方法	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
30	NA	0.921	0.927	0.906	0.752	0.748	0.755	1.040	1.028	1.061
	PEL	0.904	0.923	0.803	0.846	0.769	0.882	1.152	1.027	1.532
	APEL	0.932	0.957	0.866	0.783	0.728	0.815	1.026	0.893	1.215
	IFEL		0.922			0.633			0.911	
50	NA	0.931	0.944	0.929	0.577	0.576	0.576	0.711	0.725	0.734
	PEL	0.917	0.945	0.865	0.632	0.584	0.666	0.849	0.741	1.067
	APEL	0.935	0.953	0.903	0.592	0.561	0.631	0.734	0.637	0.906
	IFEL		0.907			0.517			0.756	
100	NA	0.934	0.947	0.949	0.400	0.399	0.404	0.496	0.483	0.487
	PEL	0.953	0.952	0.908	0.424	0.403	0.448	0.549	0.518	0.616
	APEL	0.953	0.959	0.931	0.410	0.393	0.429	0.506	0.448	0.544
	IFEL		0.954			0.379			0.452	

表 2 关于$\mathrm{qcor}_\tau\{Y,X\}$和$\mathrm{corr}\{Y,X\}$区间估计的覆盖概率, 平均长度和平均损失, 例2

		覆盖概率			平均长度			平均损失		
n	方法	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
30	NA	0.921	0.907	0.922	0.762	0.740	0.767	1.042	1.116	0.981
	PEL	0.893	0.924	0.837	0.843	0.775	0.926	1.279	1.051	1.539
	APEL	0.925	0.948	0.871	0.782	0.714	0.804	1.030	0.895	1.142
	IFEL		0.905			0.605			0.945	
50	NA	0.938	0.926	0.934	0.576	0.567	0.573	0.739	0.774	0.770
	PEL	0.909	0.914	0.868	0.629	0.591	0.671	0.869	0.843	1.088
	APEL	0.947	0.948	0.914	0.592	0.551	0.625	0.712	0.688	0.796
	IFEL		0.902			0.497			0.743	
100	NA	0.943	0.940	0.941	0.398	0.398	0.401	0.483	0.505	0.502
	PEL	0.919	0.950	0.911	0.427	0.408	0.459	0.544	0.501	0.665
	APEL	0.951	0.957	0.940	0.410	0.387	0.431	0.512	0.440	0.523
	IFEL		0.925			0.369			0.488	

表 3 关于$\mathrm{qcor}_\tau\{Y,X\}$和$\mathrm{corr}\{Y,X\}$区间估计的覆盖概率, 平均长度和平均损失, 例3

		覆盖概率			平均长度			平均损失		
n	方法	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
30	NA	0.919	0.922	0.918	0.764	0.754	0.772	0.944	1.062	1.072
	PEL	0.891	0.935	0.832	0.853	0.777	0.911	1.238	1.058	1.455
	APEL	0.926	0.961	0.886	0.782	0.728	0.831	0.995	0.868	1.281
	IFEL		0.922			0.639			0.920	
50	NA	0.931	0.945	0.908	0.580	0.580	0.585	0.743	0.758	0.791
	PEL	0.907	0.941	0.880	0.633	0.589	0.679	0.808	0.726	1.030
	APEL	0.941	0.959	0.903	0.597	0.563	0.625	0.714	0.643	0.867
	IFEL		0.940			0.515			0.690	
100	NA	0.932	0.949	0.944	0.401	0.402	0.405	0.505	0.499	0.498
	PEL	0.934	0.948	0.908	0.424	0.404	0.456	0.536	0.484	0.626
	APEL	0.956	0.955	0.932	0.412	0.394	0.432	0.507	0.458	0.587
	IFEL		0.931			0.381			0.497	

表 4 关于$\mathrm{qcor}_\tau\{Y,X\}$和$\mathrm{corr}\{Y,X\}$区间估计的覆盖概率, 平均长度和平均损失, 例4

		覆盖概率			平均长度			平均损失		
n	方法	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
30	NA	0.923	0.939	0.915	0.799	0.757	0.821	1.075	0.966	1.118
	PEL	0.919	0.930	0.847	0.885	0.778	0.979	1.214	1.035	1.458
	APEL	0.940	0.954	0.904	0.823	0.736	0.898	1.018	0.949	1.230
	IFEL		0.865			0.570			0.989	
50	NA	0.930	0.940	0.933	0.606	0.582	0.636	0.765	0.763	0.831
	PEL	0.936	0.941	0.899	0.663	0.588	0.782	0.876	0.739	1.088
	APEL	0.956	0.951	0.938	0.619	0.564	0.694	0.747	0.637	0.860
	IFEL		0.890			0.461			0.719	
100	NA	0.941	0.942	0.951	0.411	0.401	0.435	0.490	0.496	0.534
	PEL	0.940	0.947	0.936	0.439	0.405	0.510	0.529	0.482	0.668
	APEL	0.947	0.947	0.945	0.418	0.395	0.459	0.492	0.462	0.536
	IFEL		0.927			0.347			0.471	

表 5 关于$\mathrm{qcor}_\tau\{Y,X\}$和$\mathrm{corr}\{Y,X\}$区间估计的覆盖概率, 平均长度和平均损失, 例5

		覆盖概率			平均长度			平均损失		
n	方法	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
30	NA	0.893	0.926	0.925	0.671	0.742	0.848	1.101	0.966	1.035
	PEL	0.869	0.923	0.845	0.714	0.755	0.999	1.387	1.020	1.529
	APEL	0.918	0.961	0.904	0.699	0.715	0.912	1.079	0.901	1.216
	IFEL		0.892			0.584			0.916	
50	NA	0.915	0.941	0.939	0.533	0.568	0.646	0.763	0.722	0.798
	PEL	0.883	0.948	0.885	0.567	0.581	0.782	0.859	0.713	1.070
	APEL	0.927	0.942	0.923	0.551	0.552	0.689	0.730	0.641	0.833
	IFEL		0.896			0.479			0.739	
100	NA	0.931	0.934	0.961	0.381	0.398	0.441	0.503	0.520	0.535
	PEL	0.933	0.929	0.920	0.402	0.399	0.512	0.506	0.509	0.630
	APEL	0.942	0.943	0.936	0.392	0.391	0.465	0.496	0.455	0.557
	IFEL		0.920			0.362			0.497	

表 6 关于$\mathrm{qcor}_\tau\{Y,X\}$和$\mathrm{corr}\{Y,X\}$区间估计的覆盖概率, 平均长度和平均损失, 例6

		覆盖概率			平均长度			平均损失		
n	方法	0.2	0.5	0.9	0.2	0.5	0.9	0.2	0.5	0.9
30	NA	0.942	0.937	0.937	0.541	0.492	0.563	1.539	1.926	1.164
	PEL	0.948	0.936	0.948	0.546	0.501	0.606	1.638	2.040	1.254
	APEL	0.953	0.960	0.909	0.555	0.500	0.612	1.198	1.193	1.563
	IFEL		0.907			0.348			2.295	
50	NA	0.954	0.950	0.940	0.403	0.369	0.426	1.058	1.475	1.006
	PEL	0.957	0.952	0.947	0.409	0.375	0.436	1.178	1.346	1.268
	APEL	0.959	0.951	0.942	0.416	0.370	0.453	0.941	1.157	1.009
	IFEL		0.933			0.272			1.718	
100	NA	0.949	0.952	0.947	0.270	0.251	0.292	1.073	1.203	0.858
	PEL	0.953	0.957	0.950	0.273	0.249	0.292	1.178	1.202	1.017
	APEL	0.951	0.961	0.943	0.274	0.253	0.303	0.949	0.992	0.843
	IFEL		0.936			0.195			1.710	

§4 实证分析

本节将具体应用提出的PEL和APEL方法, 研究一个公司的某项指标与其他指标的相关性. 由于一些指标, 如销售值在不同的公司之间可能会有非常大的差异, 故使用分位数回归模型更为合理. 而分位数相关系数作为一个有用的工具, 可以在参数估计或进行其他统计推断之前进行变量选择.

这个数据集包括1986年福布斯500强名单中的79家公司, 包含七个变量, 分别是资产额, 销售额, 市场价值, 利润, 现金流, 雇员人数和公司所处的市场类型. 前六个是数字变量, 最后一个是属性变量. 为了讨论的方便, 这里把销售量的变量记为Y, 利润记为X_1, 雇员人数记为X_2. 该数据集可以在`https://dasl.datadescription.com/datafile/companies`找到.

具体来说, 考虑在两个分位数水平$\tau_1 = 0.5$和$\tau_2 = 0.75$下销售量和利润之间的分位数相关系数. 计算可得在τ_1和τ_2的样本分位数相关系数分别为0.170和0.284, 而样本Pearson相关系数为0.814. 首先基于数值模拟部分提到的NA方法, PEL方法, APEL方法和IFEL方法做假设检验, 结果都拒绝了零假设, 这意味着两变量之间存在一定的相关性. 接下来使用这四种方法计算95%置信区间, 结果见表7. 可以看到, 在两个分位数水平下, NA方法, PEL方法和APEL方法下的估计得的置信区间长度都比较相似. 这表明销售额和利润之间的分位数相关系数是适中的, 它们之间并没有太强烈的关联. 而IFEL方法则给出了一个相对更宽的置信区间, 相对而言损失了较多的信息.

表 7 销售额和利润之间各相关系数的95%置信区间

	置信区间		区间长度	
方法	0.5	0.75	0.5	0.75
NA	(0.063,0.277)	(0.104,0.464)	0.214	0.360
PEL	(0.085,0.304)	(0.142,0.511)	0.219	0.369
APEL	(0.039,0.266)	(0.053,0.438)	0.227	0.385
IFEL	(0.342,0.940)		0.598	

直观上来说, 销售量可能与雇员人数有很高的相关性. 接下来在与之前相同的分位数水平下考虑这两个变量. 在τ_1和τ_2的样本分位数相关系数分别为0.476和0.663, 而样本Pearson相关系

数为0.924. 相应的假设检验的结果也是拒绝零假设. 用上述四种方法计算95%置信区间, 结果见表8. 结果表明可以有把握地认为这两个变量之间存在一定的相关性. 此外, 三种基于分位数的方法表现较为接近, 均体现出该相关性在较高的分位数水平上更强, 这意味着对于销售额最高的公司来说, 它们与雇员人数的相关性更高. 这些信息是分位数方法所独有的, 而其它方法, 如IFEL方法, 则无法体现出来.

表 8 销售额和雇员人数之间各相关系数的95%置信区间

	置信区间		区间长度	
方法	0.5	0.75	0.5	0.75
NA	(0.360,0.592)	(0.567,0.759)	0.232	0.191
PEL	(0.385,0.594)	(0.572,0.773)	0.210	0.201
APEL	(0.391,0.596)	(0.568,0.774)	0.205	0.206
IFEL	(0.789,0.970)		0.181	

§5 结论和展望

分位数相关系数是在很多回归模型中实现特征选择的一个重要工具, 在统计, 金融和化学等很多学科中都有广泛应用. 然而, 像经验似然这样的非参数方法用于分位数相关系数的统计推断还没有得到详细的研究. 本文提出了两种基于经验似然的方法对分位数相关系数进行区间估计, 分别推导出渐近的规则化卡方分布和渐近的标准卡方分布. 数值研究表明, 与其他已有的方法相比, 本文提出的方法不仅更加稳健, 而且在重尾和不对称分布的情况下可以提供更多的信息. 对于两种基于经验似然的方法的比较, APEL方法则相对更加稳健.

鉴于变量选择问题的重要性, 在未来将该工作运用于一些高维分位数回归模型是一种可能, 例如文[20-21]. 借鉴文[22]的思路, 经验似然方法也可以被运用于随机删失情形, 因此, 也可以考虑将本文的方法应用于像文[23]和文[24]等提出的删失分位数回归模型.

附录

定理2.1的证明需要如下两个引理.

引理1(见[1, 定理1]) 在定理2.1的条件下, 有$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widehat{V}_i - \mathrm{qcor}_\tau\{Y,X\}) \xrightarrow{d} \mathrm{N}(0,\Omega)$.

引理2 在定理2.1的条件下, 有$\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i - \mathrm{qcor}_\tau\{Y,X\})^2 \xrightarrow{p} \sigma_0^2$.

证 注意到$-1 \le \mathrm{qcor}_\tau\{Y,X\} \le 1$. 因此

$$\left|\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i - \mathrm{qcor}_\tau\{Y,X\})^2 - \frac{1}{n}\sum_{i=1}^{n}(V_i - \mathrm{qcor}_\tau\{Y,X\})^2\right| \le O_p(1)\cdot\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{V}_i - V_i\right|$$

$$= O_p(1)\cdot\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\psi_\tau(Y_i - \widehat{Q}_{\tau,Y})(X_i - \bar{X})}{\sqrt{\tau-\tau^2}S_X} - \frac{\psi_\tau(Y_i - Q_{\tau,Y})(X_i - \mu_X)}{\sqrt{\tau-\tau^2}\sigma_X}\right|$$

$$\le O_p(1)\cdot\frac{1}{n}\left[\sum_{i=1}^{n}\left|\mu_X - \bar{X}\right|\right] = o_p(1).$$

由大数定律

$$\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})^2\xrightarrow{p}\sigma_0^2.$$

综上引理2得证.

定理2.1的证明 令$W_n=\max\limits_{1\le i\le n}\left|\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\}\right|$. 显然$W_n=O(1)$, a.s.

令$S_n=\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})^2$. 由(3)可得

$$\begin{aligned}0=|f(\lambda)|&=\frac{1}{n}\Big|\sum_{i=1}^{n}(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})-\lambda\sum_{i=1}^{n}\frac{(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})^2}{1+\lambda(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})}\Big|\\&\ge\frac{|\lambda|}{n}\sum_{i=1}^{n}\frac{(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})^2}{1+\lambda(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})}-\frac{1}{n}\Big|\sum_{i=1}^{n}(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\}\Big|\\&\ge\frac{|\lambda|S_n}{1+|\lambda|W_n}-\Big|\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i)-\mathrm{qcor}_\tau\{Y,X\}\Big|.\end{aligned}$$

由引理1可知$\frac{|\lambda|S_n}{1+|\lambda|W_n}=O_p(n^{-1/2})$, 进而得到$|\lambda|=O_p(n^{-1/2})$.

令$\gamma_i=\lambda(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})$, 有$\max\limits_{1\le i\le n}|\gamma_i|=O_p(n^{-1/2})O(1)=O_p(n^{-1/2})$. 则式(3)可改写为

$$\begin{aligned}0=f(\lambda)&=\frac{1}{n}\sum_{i=1}^{n}\left[(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})(1-\gamma_i+\frac{\gamma_i^2}{1+\gamma_i})\right]\\&=\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\}-S_n\lambda+\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})\frac{\gamma_i^2}{1+\gamma_i}.\end{aligned}$$

由以上讨论知, 最后一项$\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})\frac{\gamma_i^2}{1+\gamma_i}=O_p(n^{-1})$, 因此

$$\lambda=S_n^{-1}(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})+\beta,\ \beta=O_p(n^{-1}).$$

再根据Taylor展开

$$l(\mathrm{qcor}_\tau\{Y,X\})=2\sum_{i=1}^{n}\gamma_i-\sum_{i=1}^{n}\gamma_i^2+2\eta_n,$$

这里

$$|\eta_n|\le C\sum_{i=1}^{n}|\lambda(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})|^3\le C|\lambda|^3n=O_p(n^{-1/2}).$$

代入λ, 并结合引理2可得

$$\begin{aligned}l(\mathrm{qcor}_\tau\{Y,X\})=&2n\lambda(\frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\}))-nS_n\lambda^2+2\eta_n\\=&\frac{n(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})^2}{S_n}-nS_n\beta^2+2\eta_n\\&-2S_n^{-1}(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})\beta n\\=&\frac{n(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i-\mathrm{qcor}_\tau\{Y,X\})^2}{\Omega}\frac{\Omega}{S_n}+o_p(1)\xrightarrow{p}\frac{\Omega}{\sigma_0^2}\chi_1^2,\end{aligned}$$

这说明随着$n\to\infty$, $A\cdot l(\mathrm{qcor}_\tau\{Y,X\})\xrightarrow{d}\chi_1^2$.

定理2.2的证明 定理2.2的证明同定理2.1类似, 故在此略去.

注3.1的证明 因为X, X_1独立, 有$Y = X + X_1 \sim N(0, 2)$. 进而

$$\begin{aligned}
\mathrm{qcor}_\tau\{Y, X\} &= \frac{\mathrm{E}\{\psi_\tau(Y - Q_{\tau,Y})(X - \mathrm{E}X)\}}{\sqrt{\tau - \tau^2}} \\
&= \frac{\int_{-\infty}^{+\infty}[\int_{-\infty}^{Q_\tau - y} -x \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x]\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y}{\sqrt{\tau - \tau^2}} \\
&= \frac{\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{Q_\tau^2 - 2Q_\tau y + 2y^2}{2}}\mathrm{d}y}{\sqrt{\tau - \tau^2}} \\
&= \frac{\frac{1}{2\sqrt{\pi}}\mathrm{e}^{-\frac{Q_\tau^2}{4}}}{\sqrt{\tau - \tau^2}}.
\end{aligned}$$

最后一步由事实$\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2}\mathrm{d}x = \sqrt{\pi}$辅以一些代数运算可得.

参考文献:

[1] Li Guodong, Li Yang, Tsai Chih-Ling. Quantile correlations and quantile autoregressive modeling[J]. Journal of the American Statistical Association, 2015, 110(509): 246-261.

[2] Ma Xuejun, Zhang Jingxiao. Robust model-free feature screening via quantile correlation[J]. Journal of Multivariate Analysis, 2016, 143: 472-480.

[3] Xu Kai. Model-free feature screening via a modified composite quantile correlation[J]. Journal of Statistical Planning and Inference, 2017, 188: 22-35.

[4] Liu Yi, Chen Xiaolin. A new robust model-free feature screening method for ultra-high dimensional right censored data[J]. Communications in Statistics-Theory and Methods, 2022, 51(6): 1857-1875.

[5] Pan Jing, Zhang Shucong, Zhou Yong. Variable screening for ultrahigh dimensional censored quantile regression[J]. Journal of Statistical Computation and Simulation, 2019, 89(3): 395-413.

[6] Gkillas Konstantinos, Gideon Boako, Dimitrios Vortelinos, et al. Non-parametric quantile dependencies between volatility discontinuities and political risk[J]. Finance Research Letters, 2020, 32: 101074.

[7] Zakari Ibrahim Sidi, Assi N' Guessan, Alexandre Dehaut, et al. Volatile compounds selection via quantile correlation and composite quantile correlation: a whiting case study[J]. Open Journal of Statistics, 2016, 6(6): 995.

[8] Owen Art. Empirical likelihood ratio confidence intervals for a single functional[J]. Biometrika, 1988, 75(2): 237-249.

[9] Owen Art. Empirical likelihood ratio confidence regions[J]. The annals of statistics, 1990, 18(1): 90-120.

[10] Owen Art. Empirical likelihood for linear models[J]. The Annals of Statistics, 1991, 19(4): 1725-1747.

[11] Adimari Gianfranco. An empirical likelihood statistic for quantiles[J]. Journal of Statistical Computation and Simulation, 1998, 60(1): 85-95.

[12] Zhang Biao. Empirical likelihood confidence intervals for M-functionals in the presence of auxiliary information[J]. Statistics & probability letters, 1997, 32(1): 87-97.

[13] Zhang Biao. Quantile processes in the presence of auxiliary information[J]. Annals of the institute of Statistical Mathematics, 1997, 49(1): 35-55.

[14] Chen Songxi. On the accuracy of empirical likelihood confidence regions for linear regression model[J]. Annals of the Institute of Statistical Mathematics, 1993, 45(4): 621-637.
[15] Owen Art. Empirical likelihood[M]. London: Chapman and Hall/CRC, 2001.
[16] Hu Xinjie, Aekyung Jung, Qin Gengsheng. Interval estimation for the correlation coefficient[J]. The American Statistician, 2020, 74(1): 29-36.
[17] Wang Dongliang, Zhao Yichuan. Jackknife empirical likelihood for comparing two Gini indices[J]. Canadian Journal of Statistics, 2016, 44(1): 102-119.
[18] Sang Yongli, Dang Xin, Zhao Yichuan. Jackknife empirical likelihood methods for Gini correlations and their equality testing[J]. Journal of Statistical Planning and Inference, 2019, 199: 45-59.
[19] Gneiting Tilmann, Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation[J]. Journal of the American statistical Association, 2007, 102(477): 359-378.
[20] Wang Huixia Judy, Ian W McKeague, Qian Min. Testing for marginal linear effects in quantile regression[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2018, 80(2): 433-452.
[21] Tang Songqiao, Wang Huiyu, Yan Guanao, et al. Empirical likelihood based tests for detecting the presence of significant predictors in marginal quantile regression.[J] Metrika, 2022: 1-31.
[22] Wang Qihua, Jing Bingyi. Empirical likelihood for a class of functionals of survival distribution with censored data[J]. Annals of the Institute of Statistical Mathematics, 2001, 53(3): 517-527.
[23] Leng Chenlei, Tong Xingwei. Censored quantile regression via Box-Cox transformation under conditional independence[J]. Statistica Sinica, 2014, 24(1): 221-249.
[24] Wang Huixia Judy, Wang Lan. Locally weighted censored quantile regression[J]. Journal of the American Statistical Association, 2009, 104(487): 1117-1128.

Empirical likelihood based interval estimation of quantile correlation

TANG Song-qiao, LI Kang-qiang, LI Xiang, ZHANG Li-xin
(School of Math. Sci., Zhejiang Univ., Hangzhou 310058, China)

Abstract: Quantile correlation is an asymmetric correlation coefficient describing the linear relationships between two random variables. It plays a vital role in feature screening problems in many subjects such as Statistics, Finance, and Chemistry. Besides, empirical likelihood is a famous non-parametric method that is widely used in the statistical inference of several models. In this paper, the estimation equation is firstly established from the definition of the quantile correlation of any two random variables, then two proposed methods, plug-in empirical likelihood(PEL) and its adjusted version(APEL), are introduced to make interval estimation with establishing asymptotic scaled chi-squared distribution and standard chi-squared distribution, respectively. Simulation studies compare these two empirical likelihood-based methods with other methods in terms of coverage probability, interval length, and the average loss based on interval score. A real data set from Forbes magazine is adopted in the application to illustrate the presented methods.

Keywords: quantile correlation; empirical likelihood; interval estimation
MR Subject Classification: 62G05; 62G20