应用简报

非对称损失下的预测区间

文　平

(新疆财经大学应用数学学院, 乌鲁木齐, 830012)

摘　要

本文在回顾经典的点预测理论和区间预测理论的基础上给出了求最优预测区间的准则, 该准则不仅考虑了预测区间的可信程度, 而且将损失函数纳入进了求预测区间的框架. 在该框架下, 研究了损失函数满足某些特定条件时的预测区间, 结果发现与经典预测理论中的等尾取法和对称取法得到的预测区间是一致的.

关键词: 预测, 区间预测, 非对称损失函数.

学科分类号: O212.1.

§1. 引　　言

损失函数在预测理论与实践中扮演着十分重要的作用, 根据预测的有关理论, 最优点预测就是使期望损失达到最小的那个值. 显然, 损失函数不同, 最优点预测也会相应发生变化, 由此可见最优点预测与所取的损失函数密切相关. Granger (1969)研究了正态分布条件下损失函数为非对称的线性函数的最优预测问题, 从此开始, 在非对称损失函数条件下对最优点预测问题的研究不断出现并深入. Zeller (1986)在线性–指数损失函数(Linex loss Function)下讨论了贝叶斯估计与预测问题, 并且得到了非常好的具有分析性的结果, Christoffersen与Diebold (1996, 1997)讨论了非对称损失函数下非高斯过程的最优预测问题. Patton与Timmetmann (2007)则在非对称且非线性损失函数条件下研究了最优预测的特点与性质. 上述学者所做的工作有一个共同点, 即他们主要研究非对称损失函数下的最优点预测问题.

然而, 除了点预测外还有另外一类预测问题, 即区间预测. 到目前为止, 相对于点预测的研究, 区间预测还未得到应有的关注, 也未进行较深入的研究, 相关的文献也相对较少, Olive (2007)讨论了回归模型的预测区间; Demetrescu (2007)研究了非对称损失下的最优预测区间, 虽然他将损失函数纳入到了区间预测问题中, 但是他的讨论框架存在缺陷, 主要在于只考虑了预测区间内的期望损失而未考虑预测区间外的期望损失. 现在预测理论通常的做法是, 当预测分布得到后等尾的取预测区间或者关于最优点预测对称的取预测区间. 无

本文2010年11月22日收到, 2012年3月12日收到修改稿.

论是等尾取法还是对称取法都没有直接将损失函数纳入求预测区间的过程之中, 从而必然带有一定的随意性和主观性, 这与最优点预测的求法有着本质的区别和不同. 怎样将损失函数的纳入预测区间的求解过程中, 从而建立起构筑预测区间的框架就成为了一个重要的理论和实践问题. 本文就试图在这方面进行一些有意义的探讨, 在此基础上当损失函数取一些比较特殊的损失函数时求得预测区间并与传统求法所得的预测区间进行比较, 研究发现传统的预测区间只是该框架下所得预测区间的一种特殊情形.

§2. 损失函数及点预测

设损失函数$L(x)$仅依赖预测误差x, Granger (1999)指出损失函数应具有以下三个特点:

(1) $L(0) = 0$;

(2) $\min L(x) = 0$, 因而$L(x) \geq 0$;

(3) 当$x \geq 0$时, $L(x)$为单调非减函数, 当$x < 0$时, $L(x)$为单调非增函数.

损失函数的三个特点中第一条意谓着准确预测无损失, 第二条意谓着有预测误差就有损失, 第三条意谓着较大的预测误差将带来较大的损失. 这里, 为讨论方便起见, 假设损失函数是连续的. 常见的损失函数有以下几种.

1. 二次函数$L(x) = x^2$

该函数是最常用的一种损失函数, 它关于原点对称, 是一种对称的损失函数.

2. 线性 – 线性函数$L(x) = \begin{cases} ax, & x \geq 0 \\ bx, & x < 0 \end{cases}$

其中$a, b \geq 0$, 该损失函数最先被Granger (1969)使用. 特别当$a = b$时, 线性 – 线性损失函数变为$L(x) = a|x|$, 此时损失函数为对称的损失函数, 否则为非对称损失函数.

3. 线性 – 指数损失函数$L(x) = \exp(ax) - ax - 1$

该函数最早由Varian于1974年引入, Zellner (1986)在讨论贝叶斯估计与预测时使用了该函数, 参数a控制着非对称性的程度. 若$a > 0$, 较大的损失产生于正的预测误差, 若$a < 0$, 较大的损失将源于负的预测误差.

设Y_t为所要预测的随机过程, 其预测分布用该随机变量的累积分布函数$F_t(x)$表示, 设其概率密度函数为$f_t(x)$, 根据预测测理论, 最优点预测y_t^*就是要最小化期望损失, 即

$$y_t^* = \arg\min \mathsf{E}(L(Y_t - y_t^*)). \tag{2.1}$$

假如损失函数为二次函数, 很容易求出预测分布的均值为唯一的最优点预测. 假如损失函数为线性 – 线性函数, Granger (1969)证明了预测分布的$a/(a+b)$分位数的为最优点预测(注: 不一定唯一), 特别当线性 – 线性损失函数对称时, 预测分布的中位数就是最优点预测. 假如损失函数为线性 – 指数函数时, Zellner (1986)经过推导得到$y_t^* = (-1/a)$

$\cdot \log \mathsf{E}(e^{-aY_t})$为最优点预测. 特别当$Y_t$服从均值为$\mu$, 方差为$\sigma^2$的正态分布时, 最优点预测为$y_t^* = \mu - (a/2)\cdot\sigma^2$.

§3. 最优预测区间

如上所述, 点预测除与预测分布有关外还与所取的损失函数密切相关, 对于同一所要预测的随机过程, 不同的损失函数下的点预测的结果也将不同. 然而与点预测不同的是在传统预测理轮中, 预测区间的取法却带有很大的随意性, 一般和损失函数的取法无直接关系.

设所要预测的随机变量Y_t的累积分布函数为$F_t(x)$, 其概率密度函数为$f_t(x)$, 又设在损失函数$L(x)$下的最优点预测为y_t^*, 预测区间(y_t^l, y_t^u)的取法通常有以下两种方法.

1. 对于给定的正数为$\alpha\,(0<\alpha<1)$, 让Y_t落在预测区间(y_t^l, y_t^u)的概率为$1-\alpha$, 即$\mathsf{P}(y_t^l \le Y_t \le y_t^u) = 1-\alpha$, 则$(y_t^l, y_t^u)$就是可信程度为$1-\alpha$的预测区间.

显然这样的预测区间有很多个. 不妨设$\alpha_1+\alpha_2=\alpha$, 则y_t^l为Y_t的α_1分布位点, y_t^u为Y_t的$1-\alpha_2$分位点. 可见, 这种预测区间不唯一, 当预测分布已知时, 满足可信程度为$1-\alpha$的预测区间有无穷多个. 特别取$\alpha_1=\alpha_2=\alpha/2$, 则$y_t^l$为$Y_t$的$\alpha/2$分位点, y_t^u为Y_t的$1-\alpha/2$分位点, 这就是所谓的等尾取法. 等尾取法所得的预测区间在预测理论和实践中常被采用.

2. 对于给定的正数为$\alpha\,(0<\alpha<1)$, 让Y_t落在以最优点预测y_t^*为心, 以r为半径时的区间的概率为$1-\alpha$, 即$\mathsf{P}(y_t^*-r \le Y_t \le y_t^*+r) = 1-\alpha$, 从而$y_t^l = y_t^*-r$, $y_t^u = y_t^*+r$这样得到的预测区间(y_t^l, y_t^u), 这就是关于最优点预测的对称取法. 对称取法是另外一种求预测区间的方法.

预测区间的第一种取法与损失函数毫无关系, 且带有很大的随意性, 正因为如此, 预测区间不唯一. 预测区间的第二种取法显然与损失函数也无直接的关联. 这两种方法都只要求变量Y_t落在预测区间的概率为$1-\alpha$, 即只要求可信程度达到某一给定的数值. 那么如何将损失函数纳入预测区间的求法呢? 正如最优点预测最小优点预测的期望损失(预测风险), 最优预测区间应该最小化预测区间的期望损失(区间预测风险). 预测区间(y_t^l, y_t^u)的期望损失可以表示为

$$\mathsf{E}L(y_t^l, y_t^u) = \int_\Omega L(y_t^l, y_t^u, x)\mathrm{d}F_t(x). \tag{3.1}$$

这里L代表损失函数, Ω为Y_t的取值空间. 在求预测区间时, 除需要考虑期望损失外, 还要考虑预测区间的可信程度. 所以, 对于给定的正数$\alpha\,(0<\alpha<1)$和损失函数$L(x)$, 预测区间(y_t^l, y_t^u)的下界y_t^l和上界y_t^u应满足下列规划问题

$$\min \mathsf{E}L(y_t^l, y_t^u) = \int_\Omega L(y_t^l, y_t^u, x)\mathrm{d}F_t(x) \qquad \text{s.t.} \qquad \mathsf{P}(y_t^l \le Y_t \le y_t^u) = 1-\alpha. \tag{3.2}$$

当预测分布和损失函数已知时, 解上面的规划问题, 所得到的解就是所求的预测区间. 但是, 在一般的损失函数和预测分布下, 上述规划问题的解不容易求得. 下面考虑如下形式

的损失函数

$$L(y_t^l, y_t^u, x) = L_1(y_t^l - x) + L_2(x - y_t^u) + k(y_t^u - y_t^l). \tag{3.3}$$

这里k为一个非负常数, $L_1(x)$, $L_2(x)$为单调非减函数且当$x \le 0$时$L_1(x) = L_2(x) = 0$. $L_1(x)$可以看成低估所产生的损失, $L_2(x)$可以看成高估所产生的损失. 一般来讲, 预测区间越长, 精度也就越差, 精度差也会造成损失, 所以$k(y_t^u - y_t^l)$反映精度所造成的损失. 在损失函数(3.3)下, 期望损失可以用数学期望加以表示

$$\mathsf{E}L(y_t^l, y_t^u) = \mathsf{E}_{-\infty}^{y_t^l}[L_1(y_t^l - x)] + \mathsf{E}_{y_t^u}^{+\infty}[L_2(x - y_t^u)] + k(y_t^u - y_t^l). \tag{3.4}$$

我们考虑一种最简单之情形, 即$L_1(x)$, $L_2(x)$均为线性函数的情形, 这时,

$$L_1(y_t^l - x) = \begin{cases} a(y_t^l - x), & x < y_t^l; \\ 0, & x \ge y_t^l, \end{cases}, \qquad L_2(x - y_t^u) = \begin{cases} b(x - y_t^u), & x \ge y_t^u; \\ 0, & x < y_t^u, \end{cases} \tag{3.5}$$

则期望损失为

$$\mathsf{E}L(y_t^l, y_t^u) = \int_{-\infty}^{y_t^l} a(y_t^l - x)\mathrm{d}F_t(x) + \int_{y_t^u}^{+\infty} b(x - y_t^u)\mathrm{d}F_t(x) + k(y_t^u - y_t^l),$$

从而

$$\frac{\partial \mathsf{E}L(y_t^l, y_t^u)}{\partial y_t^l} = aF(y_t^l) - k, \qquad \frac{\partial \mathsf{E}L(y_t^l, y_t^u)}{\partial y_t^u} = -b(1 - F(y_t^u)) + k.$$

令

$$\frac{\partial \mathsf{E}L(y_t^l, y_t^u)}{\partial y_t^l} = \frac{\partial \mathsf{E}L(y_t^l, y_t^u)}{\partial y_t^u} = 0,$$

则有$\alpha F(y_t^l) - b(1 - F(y_t^u)) = 0$, 考虑到$\mathsf{P}(y_t^l \le Y_t \le y_t^u) = 1 - \alpha$, 即$F(y_t^u) - F(y_t^l) = 0$, 从而有

$$F(y_t^l) = \frac{b}{a+b}\alpha, \qquad 1 - F(y_t^u) = \frac{a}{a+b}\alpha,$$

即最优预测区间的下界为预测分布的下侧$[b/(a+b)]\alpha$个分位点, 最优预测区间时上界为预测分布的上侧$[a/(a+b)]\alpha$分位点.

特别地, 若$a = b$, 则有

$$F(y_t^l) = \frac{1}{2}\alpha, \qquad 1 - F(y_t^u) = \frac{1}{2}\alpha,$$

此时, 最优预测区间与等尾取法所得的预测区间相同, 可见等尾取法是形为(3.5)的损失函数当$a = b$时的特殊情形.

我们考虑一种更为一般的情形, 假设形如(3.3)式的损失函数$L_1(x)$, $L_2(x)$关于$x = 0$对称且可导, Y_t的概率密度函数$f_t(x)$关于其均值μ对称, 则最优预测区间为$(\mu - r, \mu + r)$, 其中r满足

$$\int_{\mu - r}^{\mu + r} f_t(x)\mathrm{d}x = 1 - \alpha.$$

根据Granger (1969)的结论, 当损失函数关于$x=0$对称且Y_t的概率密度函数关于其均值μ对称时, 最优点预测为Y_t的均值μ, 即$y_t^*=\mu$. 下面证明若$\mathrm{P}(\mu-r\le Y_t\le\mu+r)=1-\alpha$, 则$(\mu-r,\mu+r)$为满足(3.2)的最优预测区间.

先证一种特殊情形. 设Y_t的概率密度函数$f_t(x)$关于$x=0$对称, 显然Y_t的均值$\mu=0$, 构造拉格朗函数

$$\begin{aligned}F(y_t^l,y_t^u,\lambda) &= \int_{-\infty}^{y_t^l} L_1(y_t^l-x)\mathrm{d}F_t(x)+\int_{y_t^u}^{+\infty} L_2(x-y_t^u)\mathrm{d}F_t(x)+k(y_t^u-y_t^l)\\ &\quad +\lambda\Big(\int_{y_t^l}^{y_t^u} f(x)\mathrm{d}x-1+\alpha\Big),\end{aligned}$$

则有

$$\frac{\partial F}{\partial y_t^l}=\int_{-\infty}^{y_t^l} L_1'(y_t^l-x)\mathrm{d}F_t(x)-ky_t^l-\lambda f(y_t^l), \tag{3.6}$$

$$\frac{\partial F}{\partial y_t^u}=-\int_{y_t^u}^{+\infty} L_2'(x-y_t^u)\mathrm{d}F_t(x)+ky_t^u+\lambda f(y_t^u), \tag{3.7}$$

$$\frac{\partial F}{\partial \lambda}=\int_{y_t^l}^{y_t^u} f(x)\mathrm{d}x-1+\alpha. \tag{3.8}$$

令

$$\frac{\partial F}{\partial y_t^l}=\frac{\partial F}{\partial y_t^u}=\frac{\partial F}{\partial \lambda}=0,$$

则若$-r$满足(3.6)式, 由于损失函数及概率密度函数所具有的对称性, r则满足(3.7)式, r还必须满足(3.8)式, 故有$\mathrm{P}(-r\le Y_t\le r)=1-\alpha$, 所以满足$\mathrm{P}(-r\le Y_t\le r)=1-\alpha$的区间$(-r,r)$为最优预测区间.

若Y_t的均值$\mu\ne0$, 设$Y_t'=Y_t-\mu$, 则根据上面推导Y_t'的最优预测区间为$(-r,r)$, 其中r满足$\mathrm{P}(-r\le Y_t'\le r)=1-\alpha$, 从而$Y_t$的最优预测区间为$(\mu-r,\mu+r)$, 其中$r$满足$\mathrm{P}(-r\le Y_t-\mu\le r)=1-\alpha$, 即$\mathrm{P}(\mu-r\le Y_t\le\mu+r)=1-\alpha$.

可见, 经典预测理论中的等尾取法和对称取法皆可纳入我们求最优预测区间的理论框架. 这不仅说明求最优预测区间理论框架的正确性, 而且说明在该框架下可以大大增加所求预测区间的现实符合性. 如前所述, 经典预测理论中, 预测区间的取法带有随意性, 其根本原因就是没有将损失函数纳入求最优预测区间的理论框架. 现在就可根据实际情况选定损失函数, 从而确立最优预测区间. 这样就就从根本上避免了随意性, 使预测区间由损失函数和预测分布唯一确定. 另外, 这样的求预测区间的方法就与求最优点预测的方法形成了一致性.

§4. 结　　论

本文在回顾经典的点预测理论和区间预测理论的基础上, 给出了求最优预测区间的准

则, 该准则不仅要考虑预测区间的可信程度, 更重要的是将损失函数纳入进了求预测区间的过程, 这样做不仅减少了和避免了预测区间取法的随意性而且增加了现实符合性. 在这种框架下, 最优预测区间从本质上看就是某个规划问题的解, 当损失函数满足某些条件时, 规划问题解与经典预测理论中的等尾取法和对称取法是一致的, 这充分说明我们了所建立起的理论框架的正确性.

参 考 文 献

[1] Christoffersen, P.F. and Diebold, F.X., Further results on forecasting and model selection under asymmetric loss, *Journal of Applied Econometrics*, **11**(1996), 561–572.

[2] Christoffersen, P.F. and Diebold, F.X., Optimal prediction under asymmetric loss, *Econometric Theory*, **13**(1997), 808–817.

[3] Elliott, G. and Timmermann, A., Optimal forecast combinations under general loss functions and forecast errordistributions, *Journal of Econometrics*, **122**(2004), 47–79.

[4] Granger, C.W.J., Prediction with a generalized cost of error function, *Operational Research Quarterly*, **20**(1969), 199–207.

[5] Granger, C.W.J., Outline of forecast theory using generalized cost functions, *Spanish Economic Review*, **1**(1999), 161–173.

[6] Zellner, A., Bayesian estimation and prediction using asymmetric loss functions, *Journal of the American Statistical Association*, **81**(1986), 446–451.

[7] Patton, A.J. and Timmermann, A., Properties of optimal forecasts under asymmetric lossand nonlinearity, *Journal of Econometrics*, **140**(2007), 884–918.

[8] Olive, D.J., Prediction intervals for regression models, *Computational Statistics & Data Analysis*, **51**(2007), 3115–3122.

[9] Demetrescu, M, Optimal forcast intervals under asymmetric loss, *Journal of Forcasting*, **26**(2007), 227–238.

[10] Winkler, R, A decision theoretic approach to interval estimation, *Journal of American Statistical Association*, **67**(1974), 187–191.