期权定价的波动率模型比较与探索*

王胜安，周　清**

(北京邮电大学理学院，北京　100876)

摘要:选取上证50ETF期权数据,利用信息含量的比较方法,对广义自回归条件异方差(GARCH)模型、Black-Scholes(B-S)隐含波动率模型、无模型、异质自回归波动率(HAR-RV)模型和"短期依赖"结合模型的波动率预测能力进行比较.实证结果表明,各模型的信息含量排序为ARMA-GARCH、B-S、HAR-RV、无模型和结合模型.对于单个模型而言,GARCH模型的预测精度相对较低,无模型隐含波动率表现最优,在综合考虑操作复杂程度和预测准确性的条件下HAR-RV模型则是最佳选择,结合模型在提升波动率预测稳定性方面表现较好.

关键词:波动率预测;信息含量;模型结合;模型比较

中图分类号:O211.63　　**DOI**:10.19789/j.1004-9398.2021.02.001

0　引　言

波动率是衡量期权风险大小的一种标准,通过建立模型来预测期权的波动率,可以达到提前调控,减少损失甚至是增加收益的目的.Engle[1]指出条件方差与滞后的残差平方有关,并建立了自回归条件异方差(ARCH)模型;Bollerslev[2]在ARCH模型的基础上,将模型拓展为广义的自回归条件异方差(GARCH)模型,解决了ARCH模型在拟合时滞后阶数过大的问题;Andersen等[3]利用高频记录的期权价格定义了新的实现波动率(realized volatility,RV),并指出该波动率不存在模型误差;Corsi[4]认为不同间隔的数据会造成不同的波动率,并建立已实现波动率异质自回归模型(HAR-RV);Black和Scholes[5]建立的B-S定价公式基础上反推出波动率,可达到对波动率预测的效果;Britten-Jones和Neuberger[6]为了排除模型对波动率的影响,提出了无模型的波动率预测方法;Fair和Shiller[7]提出了包含回归法,则适用于比较各个模型信息含量的多少.近年来也有很多国内学者开始研究波动率.袁臻[8]和刘倡[9]都比较了各个不同的波动率预测模型的优劣,后者在前者的基础上将近年才发展的无模型波动率加入比较;郑振龙和黄薏舟[10]指出在长期预测时无模型隐含波动率的表现要优于GARCH模型.

受现有研究启发,本文选取了国内期权市场的数据,利用来自不同类别的模型进行波动率预测,提出一种新的简单的模型结合方法,并比较各模型在预测波动率上表现的优劣.

1　模型介绍

1.1　RV模型

Anderson等[3]在研究中利用每5 min记录1次的高频数据,定义第t天的实现波动率为

$$\mathrm{RV}_t=\sqrt{\sum_{i=1}^{n}r_{t,i}^{\ 2}},\tag{1}$$

式中$r_{t,i}$表示第t天i时刻的对数收益率,RV_t为第t天的实现波动率.

1.2　GARCH模型

由Bollerslev[2]建立GARCH(p,q),具体的数学表达式为

$$r_t=\mu_t+\epsilon_t,\tag{2}$$

式中r_t为对数收益率,μ_t为漂移参数,ϵ_t是扰动项.σ_t^2为其方差,ϵ_t为波动率,令$r_t-\mu_t=\epsilon_t$,则

$$\epsilon_t=\sigma_t h_t,\tag{3}$$

$$\sigma_t^2=\alpha_0+\alpha_1\epsilon^2_{\ t-1}+\alpha_2\epsilon^2_{\ t-2}+\cdots+\alpha_p,\tag{4}$$

收稿日期:2020-03-23

* 国家自然科学基金(11871010,11971040)

** 通信作者:zqleii@bupt.edu.cn

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i \epsilon^2_{t-i} + \sum_{j=1}^{q}\beta_j \sigma^2_{t-j}. \quad (5)$$

式中h_t为均值为0、方差为1的独立同分布随机变量序列.$\alpha_0 > 0$且对所有的$i > 0, j > 0$有$\alpha_i \geqslant 0$;$\beta_j \geqslant 0$, $\sum_{i=1,j=1}^{\max(p,q)}(\alpha_i + \beta_j) < 1$.

1.3 HAR-RV模型

Corsi[4]所建立的模型中,实际波动率的预测表达式为

$$\mathrm{RV}^{(\mathrm{d})}_{t+1\mathrm{d}} = c + \beta^{(\mathrm{d})}\mathrm{RV}^{(\mathrm{d})}_t + \beta^{(\mathrm{w})}\mathrm{RV}^{(\mathrm{w})}_t + \beta^{(\mathrm{m})}\mathrm{RV}^{(\mathrm{m})}_t + \omega_{t+1\mathrm{d}}, \quad (6)$$

式中$\mathrm{RV}^{(\mathrm{d})}_t$、$\mathrm{RV}^{(\mathrm{w})}_t$、$\mathrm{RV}^{(\mathrm{m})}_t$分别表示第$t$天的日、周、月观测的实际波动率,周实际波动率为过去5 d的日波动率的平均值,月波动率为过去22 d日波动率的均值.$\beta^{(\mathrm{d})}$、$\beta^{(\mathrm{w})}$、$\beta^{(\mathrm{m})}$代表参数项、ω代表误差项、c为常数项.

1.4 B-S隐含波动率模型

由B-S定价公式可知,公式中共存在6个参数,而在实际的市场中除了波动率这一项参数之外,其余5个均可以被直接记录,因此,根据B-S定价公式就可以反推出该期权波动率.一个期权就能算出一个隐含波动率,而实际市场中存在着很多不同行权价但到期日相同的期权,因此,一般选择多个期权波动率的平均值来作为同一到期日不同行权价的期权的隐含波动率.

1.5 无模型

Britten-Jones和Neuberger[6]指出无模型波动率预测仅仅依赖于当前的期权价格,并不依赖于任何模型.因此,理论上避免了模型设定所带来的误差.

设定:T为有限的时间集,h为时间间隔,S_0为初始价格,股票价格从有限集里取值.行权价格$\boldsymbol{K} = \{K: K = S_0 u^i, i = 0, \pm 1, \pm 2, \cdots, \pm M\}$,其中$u > 1$.期权的价格为$C(t, K)$,$t$时刻的股票价格为$S_t$,交割日期为$T$,交割价格为$K$.

$$E\left[\max\left(S_t - K, 0\right)\right] = C(t, K), t \in T, K \in \boldsymbol{K}, \quad (7)$$

$$\mathrm{E}_0\left[\int_0^t\left(\frac{\mathrm{d}S_t}{S_t}\right)^2\right] = 2\int_0^{\infty}\frac{C(t,K) - \max(S_0 - K, 0)}{K^2}\mathrm{d}K, \quad (8)$$

式中$\max(S_0 - K, 0)$是当前到期的期权价格$C(0, K)$,公式(8)的左端是波动率的平方,只要将各个参数值带入等式,即可算出期权的隐含波动率.

1.6 波动率模型

Fair和Shiller[7]提出包含回归法可用于比较不同模型信息含量,从而达到比较不同模型的目的.定义${}_{t-s}\hat{Y}_{1t}$为模型1在t-s时刻对Y_t的预测,s为预测前的长度,建立回归方程

$$Y_t - Y_{t-s} = \alpha + \beta({}_{t-s}\hat{Y}_{1t} - Y_{t-s}) + \gamma({}_{t-s}\hat{Y}_{2t} - Y_{t-s}) + u_t. \quad (9)$$

如果模型1和模型2中都不包含Y_t前s步的预测信息的话,则γ、β都应该为0,则α为s步预测值的平均值.如果2个模型都含有s步的预测信息,则γ、β应该都不为0.根据上面的信息,验证具体模型是否包含实际波动率的信息,在恰当的置信水平下检验各模型系数的显著性.

2 实证研究

2.1 数据选取

本文选取2019年6月到期的50ETF数据,选取的时间节点为2018年10月25日—2019年4月16日.数据参数含:标的资产每5 min的开盘价、收盘价、最高价和最低价,相同到期日,不同行权价的期权每5 min的价格数据.对于只需要日度数据信息的模型,本文采用每个交易日的开盘价、收盘价、最高价和最低价的均值作为每个交易日的数据.同时选择一年期的国债利率作为无风险利率(r),文中所有数据均来自wind数据库.

本文均采用样本内预测方法,会预测出第22个样本后的93个数据,但只选取前30个作为1个月度的数据来进行比较.

2.2 GARCH模型

建立ARMA-GARCH模型,并带入期权数据可得预测结果见图1.可知,整体上预测序列整体趋势与实际一致,但数值要大于实际值,产生误差的原因是:预测出的数值不断地被代入模型拟合过程中,会把预测误差也代入进去,导致误差逐渐变大.

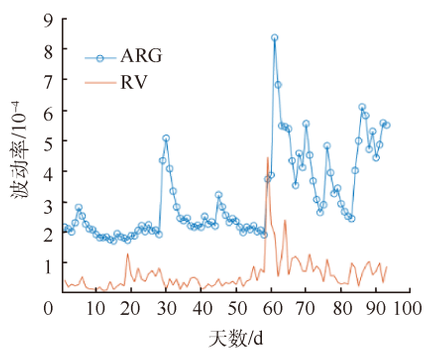


图1 ARMA-GARCH模型预测结果

注:ARG为ARMA模型下广义自回归条件异方差;RV为实现波动率.

2.3　HAR-RV模型

根据已有研究所建模型和本文所选数据,利用matlab进行实践,所得结果见图2.该模型预测结果在走势上与实际值基本相同,只是在数值大小上存在一定的差异.

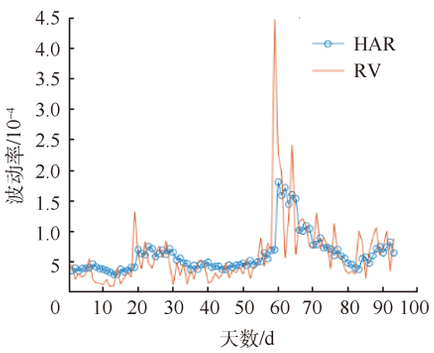


图2　HAR-RV模型预测结果

注:HAR为异质自回归.

2.4　B-S隐含波动率模型

根据B-S公式反推出隐含波动率,本文在设定各个行权价期权的预测波动率的平均值作为同一标的资产,同一到期日期权的波动率预测数据,其预测结果见图3.该模型的预测结果与ARMA-GARCH模型相似,整体趋势上与实际值接近,但是数值上整体大于实际值.

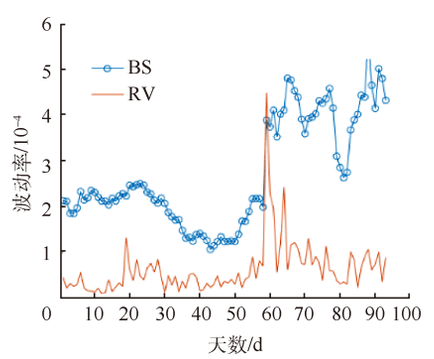


图3　B-S模型下隐含波动率预测结果

注:BS为B-S模型下隐含波动率.

2.5　无模型隐含波动率

在该模型的计算过程中,由于所收集的数据有限,无法直接根据恐惧指数的编制方法来计算隐含波动率.因此,本文采取郑振龙和黄薏舟[14]提出的编制方法.故要先对公式(8)离散化处理,得到

$$E_0^F\left[\int_0^T\left(\frac{\mathrm{d}F_t}{F_t}\right)^2\right]\cong 2\sum_{i=1}^{m}\frac{C^F\left(T,K_i\right)-\max\left(0,F_0-K_i\right)}{K_i}\Delta K. \tag{10}$$

令$B(t,T)$表示在T时刻支付1元的国债在t时刻的价格;E_0^F表示以$B(t,T)$为单位的风险中性世界的期望,等式左端仍为波动率的平方;$C^F\left(T,K_i\right)$为交割日期为T,交割价格为K_i的期权在0时刻的远期价格;F_t表示标的资产在t时刻的远期价格,$\max(0,F_0-K_i)$相当于是$C^0\left(T,K_i\right)$,$\Delta K=(K_{\max}-K_{\min})/m$,$K_i=K_{\min}+i\Delta K$,$i=0,1,2,\cdots,m$.将公式(10)转化为现货价格表达式之后,即可算出无模型隐含波动率,公式为

$$E_0^F\left[\int_0^T\left(\frac{\mathrm{d}S_t}{S_t}\right)^2\right]\cong 2\sum_{i=1}^{m}\frac{C\left(T,K_i\right)\mathrm{e}^{rT}-\max\left(0,S_t\mathrm{e}^{rT}-K_i\right)}{K_iK_i}\Delta K_i, \tag{11}$$

式中r为无风险利率,$\Delta K_i=K_{i+1}-K_i$,其余参数含义与公式(8)及(10)一致,具体的预测结果见图4.该模型在近一个月左右的预测线更接近于实际曲线,而后半段曲线则与实际值相差较大,这一点与郑振龙和黄薏舟[10]的研究结果相近.

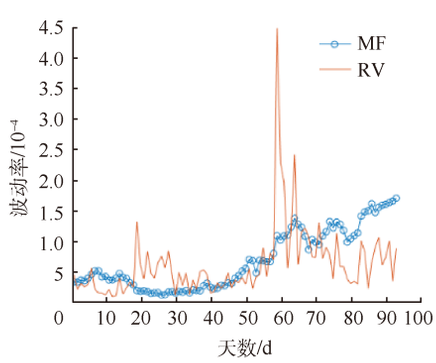


图4　无模型波动率预测结果

注:MF为无模型波动率.

2.6　结合模型

单个模型在预测时,会存在模型不稳定的情况,每个模型不稳定的触发条件不相同.使用合理的方法,结合各个不同模型的预测结果,在解决单个模型不稳定问题的同时,不影响模型预测效果,甚至可以提高结合模型对样本的解释能力.

选择绝对误差作为某一时刻的各单个模型预测精度的评价函数,绝对误差大的预测精度差,反之亦然.基于此,提出一种简单的模型结合方法.

该方法的核心思想是:在某一次预测中一个模型对于当期波动率的预测精度最高,则在预测下一期的波动率时仍会选择该模型.并且每次进行预测

时,都要基于当期的信息,找到预测精度最高的模型,从而选出下一个预测模型.这样一个每期都进行判断和选择的过程,即为本文提出的"短期依赖"动态模型结合法,具体的结合步骤如下:

(1)计算实现波动率来作为基准,使用上文提到的各个预测模型,进行样本内预测并保存结果.

(2)与上文的预测事件起点一致,计算当前时刻各个模型预测与实际结果的绝对误差.

(3)选择当前期绝对误差最小的模型,以该模型的下一期预测值作为结合模型,预测实际波动率.

(4)预测下下一期的波动率时,重新计算绝对误差,再次选出下下期的预测模型,重复上述操作,即可获得结合模型对未来一个月波动率的预测结果.

利用数学软件实现上述操作,即可获得图5结果.与各单个预测模型的结果相比,该结合方法的预测效果明显优于各单个预测模型.

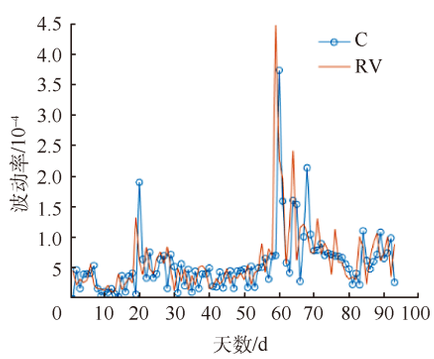


图5　结合模型的预测结果

注:C为结合模型.

2.7　回归结果分析

根据Corsi[4]研究结果,本文同样使用包含回归法进行模型信息含量的分析,公式为

$$\sigma_t^{RV}=\alpha+\beta^{ARG}\sigma_t^{ARG}+\beta^{HAR}\sigma_t^{HAR}+\beta^{BS}\sigma_t^{BS}+\beta^{MF}\sigma_t^{MF}+\beta^{C}\sigma_t^{C}+\varepsilon_t,\quad(12)$$

式中σ_t^{RV}为t时刻的实际波动率,σ_t^{ARG}为ARMA-GARCH模型预测的t时刻波动率,σ_t^{HAR}为HAR-RV模型预测的t时刻波动率,σ_t^{BS}为B-S模型预测的t时刻波动率,σ_t^{MF}为无模型波动率预测的t时刻波动率,σ_t^{C}为结合模型预测的t时刻波动率.

为了检验各单个模型是否包含实际波动率的信息,对所得数据进行单变量回归,并进行假设检验,单变量回归结果如表1所示.在5%的置信水平下,序号1、3均不能拒绝系数为0的原假设,则说明ARMA-GARCH模型、B-S隐含波动率模型在预测未来一个月波动率上的表现较差,包含的实际波动率信息较少.而HAR-RV模型、无模型波动率和结合模型的预测结果是包含了实际波动率的信息的.

表1　单变量回归结果

序号	变量名	估计值	标准误	P	R
1	β^{ARG}	-4.90×10^{-2}	0.072	0.498	0.129
	α	5.19×10^{-5}	0	0.004	
2	β^{HAR}	7.66×10^{-1}	0.321	0.024	0.411
	α	3.13×10^{-6}	0	0.851	
3	β^{BS}	4.54×10^{-1}	0.297	0.137	0.278
	α	-5.90×10^{-7}	0	0.375	
4	β^{MF}	-1.28	0.335	0	0.585
	α	8.14×10^{-5}	0	0	
5	β^{C}	5.64×10^{-1}	0.151	0.001	0.576
	α	1.90×10^{-5}	0	0.014	

为了比较不同模型信息含量之间的差异,则要2个模型之间进行双变量回归,通过R值即可比较出不同模型信息含量的差异,结果表2所示.相比于用ARCH-GARCH模型预测结果进行单变量回归,加入其他任意一个模型进行双变量回归后,拟合的R值都有很明显的提升,这说明其余模型都包含了

表2　双变量回归结果

序号	变量名	估计值	标准误	P	R
1	β^{ARG}	-1.250×10^{-1}	0.068	0.077	0.511
	β^{HAR}	9.930×10^{-1}	0.332	0.006	
	α	1.980×10^{-5}	0	0.288	
2	β^{ARG}	-3.100×10^{-2}	0.072	0.665	0.289
	β^{BS}	4.300×10^{-1}	0.306	0.171	
	α	-4.700×10^{-6}	0	0.521	
3	β^{ARG}	-1.090×10^{-1}	0.058	0.071	0.647
	β^{MF}	-1.427	0.33	0	
	α	0	0	0	
4	β^{ARG}	8.500×10^{-2}	0.041	0.040	0.595
	β^{C}	5.100×10^{-1}	0.107	0	
	α	3.150×10^{-5}	0	0.044	
5	β^{HAR}	6.790×10^{-1}	0.379	0.084	0.419
	β^{BS}	1.500×10^{-1}	0.332	0.655	
	α	-2.650×10^{-5}	0	0.700	
6	β^{HAR}	-4.210×10^{-1}	0.508	0.415	0.559
	β^{MF}	-1.688	0.597	0.009	
	α	0	0	0.011	
7	β^{HAR}	9.800×10^{-2}	0.391	0.803	0.577
	β^{C}	5.300×10^{-1}	0.206	0.016	
	α	1.540×10^{-5}	0	0.338	
8	β^{BS}	9.900×10^{-2}	0.276	0.722	0.588
	β^{MF}	-1.229	0.369	0.003	
	α	5.790×10^{-5}	0	0.391	
9	β^{BS}	2.220×10^{-1}	0.263	0.406	0.591
	β^{C}	5.290×10^{-1}	0.158	0.002	
	α	-2.900×10^{-6}	0	0.618	
10	β^{MF}	-8.100×10^{-1}	0.412	0.060	0.644
	β^{C}	3.380×10^{-1}	0.185	0.079	
	α	5.340×10^{-5}	0	0.009	

ARMA-GARCH模型所不包含的信息,即这所有的模型中ARMA-GARCH模型包含信息量最少.

在HAR-RV模型的基础上,加入其他模型进行回归,回归结果表明,加入B-S隐含波动率模型后,拟合优度提升并不大,且从系数显著性检验的P值可知,只有HAR-RV模型的系数能在5%的显著水平下拒绝系数为0的原假设,这说明HAR-RV模型包含了B-S模型所不包含的信息.加入无模型波动率和结合模型的波动率后,R值有明显的提升,且此时HAR-RV模型的系数均不能通过非0检验,这说明无模型波动率和结合模型中均包含了HAR-RV模型所不包含的信息.

在B-S隐含波动率模型的基础上,加入其他模型进行回归,在加入无模型和结合模型之后,拟合结果的R值有了很明显的提升,这说明无模型波动率和结合模型均包含了B-S模型中所不包含的信息.

比较上述的信息含量后,再比较无模型波动率和结合模型所包含的信息量,即可知所有模型的信息含量.多变量回归分析结果如表3.观察各单个模型系数的P值,在5%的置信水平下,只有结合模型的系数不显著为0,而无模型隐含模型的系数在10%的置信水平下不显著为0,其余模型均不能拒绝系数为0的原假设,这说明在更严格的条件下,结合模型的信息含量要多于无模型波动率.在一系列包含回归检验后,各个模型信息含量由低到高排序为ARMA-GARCH模型、B-S隐含波动率模型、HAR-RV模型、无模型波动率、结合模型.至此,本文在直观和数学角度上都证明了结合模型和无模型波动率所含信息量要优于其余模型.

表3 6个变量回归结果

变量名	估计值	标准误	P	R
β^{ARG}	-3.600×10^{-2}	0.073	0.631	0.737
β^{HAR}	-2.800×10^{-1}	0.73	0.704	—
β^{BS}	2.210×10^{-1}	0.321	7.498	—
β^{MF}	−1.154	0.59	0.063	—
β^{C}	5.770×10^{-1}	0.233	0.021	—
α	4.360×10^{-5}	0	0.575	—

注:—处无数据.

3 实证结论

本文结合研究所选数据,对5个模型进行实证,在这个基础上基于“短期依赖”提出了一种新的模型结合方法,该方法主要用于提升模型对下一期波动率的预测能力.运用包含回归法对这些模型信息含量进行了比较,在直观的比较图形和实证中都得出了各模型中信息含量最高的为ARMA-GARCH、最低的为结合模型.即基于“短期依赖”提出的模型结合方法,是所有模型中信息含量最高的,说明该模型结合方法有提出意义.

参 考 文 献

[1] ENGLE R F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation[J]. Econometrica,1982,50(4):987-1007.

[2] BOLLERSLEV T. Generalized autoregressive conditional heteroskedasticity[J]. Journal of Econometrics,1986,31(3):307-327.

[3] ANDERSEN T G,BOLLERSLEV T,DIEBOLD F X,et al. The distribution of realized stock return volatility[J]. Journal of Financial Economics,2001,61(1):43-76.

[4] CORSI F. A simple approximate long-memory model of realized volatility[J]. Journal of Financial Econometrics,2009,7(2):174-196.

[5] BLACK F,SCHOLES M. The pricing of options and corporate liabilities[J]. Journal of Political Economy,1973,81(3):637-654.

[6] BRITTEN-JONES M,NEUBERGER A. Option prices,implied price processes,and stochastic volatility[J]. The Journal of Finance,2000,55(2):839-866.

[7] FAIR R C,SHILLER R J. Comparing information in forecasts from econometric models[J]. American Economic Review,1990,80(3):375-389.

[8] 袁臻.波动率预测模型与比较[D].上海:上海交通大学,2013.

[9] 刘倡.波动率预测模型的比较研究[D].重庆:重庆大学,2016.

[10] 郑振龙,黄薏舟.波动率预测:GARCH模型与隐含波动率[J].数量经济技术经济研究,2010,27(1):140-150.

Comparison and exploration of different volatility models' forecasting performance

WANG Sheng'an, ZHOU Qing

(School of Science, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract: This paper compares the performance of different models in predicting volatility and establishes a new combined model for better and more stable results. We select many models including Generalized Autoregressive Heteroscedastic(GARCH) model, Black-Scholes(B-S) model, Heterogeneous Autoregressive model of Realized Volatility(HAR-RV) model, model-free volatility and combined model. To demonstrate the compare results and the advance of the combined model, the option data from 50 ETF index for empirical analysis was used. Intuitionistic and numerical results both show that combined model which proposed has the best accuracy and stability in predicting volatility. For a single model, the GARCH model has the worst prediction performance, and the model-free implied volatility performs best.

Keywords: volatility predicting; information content; combined model; model comparison

(责任编辑:马田田)