



No.C2018003

2018-03-06

测量误差修正的已实现波动率模型在中国股票指数波动率预测中的应用

王天一 王雅婧 黄卓

摘要:

高频数据计算的已实现波动率本身存在着可观测量误差,传统高频数据计算的已实现波动率本身存在着可观测量误差,传统 HAR 模型的估计参数存在偏误。针对中国市场的高波动率特点,本文考察了 Bollerslev 等(2016)提出的 HARQ 模型和 Huang 等(2017)改进的 LogHARQ 模型对上证 50ETF 和沪深 300 股指波动率的建模效果。基于 2007-2016 年的实证结果,LogHARQ 模型的预测精度明显优于 HAR、HARQ 等线性模型和以及 LogHAR 等不考虑测量误差修正的对数线性模型。新模型在波动率的高波动时期表现更为突出。

关键词 波动率预测, 已实现波动率, 测量误差

测量误差修正的已实现波动率模型在中国股票指数波动率 预测中的应用

王天一¹ 王雅婧¹ 黄卓²

(¹对外经济贸易大学金融学院)

(²北京大学国家发展研究院)

【摘要】高频数据计算的已实现波动率本身存在着可观的测量误差，传统 HAR 模型的估计参数存在偏误。针对中国市场的高波动率特点，本文考察了 Bollerslev 等(2016)提出的 HARQ 模型和 Huang 等(2017)改进的 LogHARQ 模型对上证 50ETF 和沪深 300 股指波动率的建模效果。基于 2007-2016 年的实证结果，LogHARQ 模型的预测精度明显优于 HAR、HARQ 等线性模型和以及 LogHAR 等不考虑测量误差修正的对数线性模型。新模型在波动率的高波动时期表现更为突出。

关键词 波动率预测，已实现波动率，测量误差

中图分类号 F830 文献标识码 A

The volatility measurement error correction based LogHARQ model and its application in Chinese stock index volatility modeling

Abstract: There is considerable measurement error in the calculation of high frequency data, and the estimation parameters of the traditional HAR model are biased. With the characteristics of high volatility in Chinese market, this paper researches the modeling effects of the HARQ model by Bollerslev et al. (2016) and the LogHARQ model by Huang et al. (2017) based on the volatility of Shanghai 50ETF and CSI 300 index. Empirical results based on the high frequency data from 2007-2016 show that Log HARQ model is better than of the linear models like HAR, HARQ and logarithm linear models like LogHAR regardless of the measurement error correction in prediction accuracy. The new model performs better in volatility forecast, especially during periods of high volatility.

Key words: volatility forecast, Realized Volatility, measurement error

一、引言

波动率及其预测在资产定价和风险管理中起着至关重要的作用。自 Engle 和 Bollerslev (1986) 以来,传统的波动率建模一直依赖于 GARCH 类模型。随着技术的发展,高频数据的获取越来越容易,Andersen 和 Bollerslev(1998)提出了一种基于高频数据的非参数的已实现波动率(Realized Volatility)对波动率进行估计并由此开创了高频数据波动率建模的研究方向。近二十余年的研究表明,其相比使用日收益率的估计,已实现波动率的测度更为精确,并在诸如风险度量,投资组合构建,衍生品定价等问题的研究中显示出已实现测度有着可观的经济价值。

伴随着已实现测度计算方法的发展,除传统的 GARCH 模型之外,使用已实现测度进行波动率建模的方法逐渐发展起来。早期的文献如 Andersen 等(2003)提出使用 ARFIMA 模型对已实现波动率进行建模。王天一等(2012)对基于高频数据的波动率建模等进行了详细的文献评述。此后文献逐渐分为两类,一类文献借助传统的 GARCH 模型框架进行拓展,如 Engle 和 Gallo(2006)提出的 MEM 模型,Shephard 和 Sheppard(2010)提出的 HEAVY 模型以及 Hansen 等(2012)提出的通过已实现波动率和收益率的联合建模建立 Realized GARCH 模型。另一类是脱离 GARCH 框架的模型,代表性的研究如 Ghysels 等(2006)使用的基于混频数据抽样的 MIDAS 模型,Corsi(2009)提出的基于波动率瀑布效应的异质性自回归的 HAR 模型等。相比 MIDAS 模型,HAR 模型以其相对简单的结构受到研究者的高度关注。模型扩展性也比较强,例如 Andersen 等人(2007)年提出了 HAR-RV-J 和 HAR-RV-CJ 模型将波动率分为连续和跳跃成分分别给予权重,实证效果较传统的 HAR 模型有显著改进。王天一等(2012)在 Realized GARCH 模型引入厚尾分布改进模型对已实现测度的预测,预测精度较之前更高。Huang 等(2016)将 HAR 结构嫁接 Realized GARCH 模型,显著提高了后者对长期波动率预测的能力。

近年来,国内学者对于高频数据波动率建模的研究逐年增加。其中基于 HAR 模型的研究多从波动率分解、加入外生变量以及采用滤波方法在模型基础上改进。在加入外生变量方面:张小斐等(2011)通过引入异质市场驱动因素建立 HAR-L-M 模型。马锋等(2016)以 HAR-RV, HAR-RV-J, HAR-RV-CJ, HAR-RV-TCJ 为基础,结合马尔可夫状态转换机制提出了马尔可夫状态高频波动率模型。刘晓倩等(2017)将代表隐含波动率的市场波动率指数(CVX)作为影响因子提出 HAR-CVX 模型。通过波动率变化和分解:文风华等(2012)基于市场微观结构的理论,同时考虑市场波动的杠杆效应和量价关系,构造了已实现波动率及交易量之长记忆异质自回归模型(LHAR-RV-V 模型)。孙洁(2014),翟慧等(2015)分别建立 HAR-CJN 模型和对数形式的 HAR-RV-CJ 模型。此外,龚旭等(2017)则运用 EMD 等方法将模型中的已实现波动率分解,并加入跳跃波动率成分,构建 HAR-RV-EMD-J 模型。

尽管国内外的研究均指出, HAR 模型在实证中的优越表现,但大都忽略了已实现测度

作为波动率的一个估计量的测量误差对估计系数的影响。基于这个观察, Bollerslev 等(2016)在 HAR 模型的基础上提出了 HARQ 模型, 对模型的系数进行调整。HARQ 模型明确考虑了测量误差的时变性质, 预测效果优于传统 HAR 模型。由于已实现测度数据中极端值比较多见, 波动率水平两极分化严重, 已有文献指出基于对数已实现测度的模型表现比基于已实现测度本身的模型更好。由于 HARQ 这类模型的误差修正方式依赖于模型结构, 对数线性设定下的 HARQ 模型并不能直接套用线性版本的结果。Huang 等(2017)基于跳跃检验的相关结论推导了对数线性版本下的 HARQ 模型结构, 即 LogHARQ 模型, 并考察了其在能源期货波动率建模时的表现。

由于我们国家股票市场波动率和波动率的变化都比较大, 该模型相比传统的线性 HAR 模型更能适应我国市场数据。为了验证模型的表现, 本文以沪深 300 股指和上证 50ETF 的高频数据为例对模型的预测能力进行了分析。结果表明: LogHARQ 模型相比 LogHAR 模型在预测能力上有明显的改善。同时 LogHARQ 模型的预测能力相比线性 HAR 模型和 HARQ 模型也有明显的改善。通过将已实现测度和度量已实现四次变差 RQ (Realized Quadqratic) 做分组统计表明, LogHARQ 模型在高波动率及波动率的高波动率下改进效果更好。

本文的组织结构如下: 第二部分简介 HARQ 和 LogHARQ 的模型设定。第三部分以沪深 300 股指和上证 50EFT 的高频数据为例, 给出模型的实证结果, 重点考察模型对于长记忆性的捕捉和预测能力, 并对结果进行讨论。第四部分对研究结果进行稳健性分析。第五部分总结全文。

二、模型简介

1、HARQ 模型

文献中已实现方差 Realized Variance(RV)的定义为日内收益率平方求和的结果, 即:

$$RV_t = \sum_{i=1}^M r_{t,i}^2$$

其中 M 为每个交易日的采样数。以我国市场为例, 计算 5 分钟已实现方差每日的采样数为 48 个。 $r_{t,i}$ 为第 t 个交易日的第 i 个分钟(对数)收益率。Bollerslev 等(2016)指出对于由于测量误差的存在, 以 RV 估计的波动率 AR(1)模型系数会发生向零的偏误。使用传统计量经济学的偏误公式, 他们提出了如下 ARQ 模型和 HARQ 模型:

$$ARQ: RV_t = \beta_0 + (\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}) RV_{t-1} + u_t$$

$$HARQ: RV_t = \beta_0 + (\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}) RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t$$

其中 RQ 是已实现四次变差, 其计算方式为:

$$RQ_t = \frac{M}{3} \sum_{i=1}^M r_{t,i}^4$$

HARQ 模型仅对日波动率系数进行了调整, Bollerslev 等(2016)给出的解释是: 1)在 HAR

模型中，主要的系数权重在日波动率上，周和月平均波动率的系数较小，2)周平均和月平均波动率本身是平均值的概念，其测量误差远小于日波动率。

2、LogHARQ 模型

与 Bollerslev 等(2016)类似，Huang 等(2017)使用 Bandorff-Nielsen 和 Shephard(2006)为跳跃波动率检验设计的渐进理论：

$$\ln RV_t = \ln IV_t + \eta_t \quad \eta_t : MN(0, c IQ_t IV_t^{-2})$$

修订了 ARQ 和 HARQ 模型，提出了在对数线性设定下使用的 LogARQ 模型：

$$\text{LogARQ} : \ln RV_t = \beta_0 + \left(\beta_1 + \beta_{1Q} \frac{RQ_{t-1}^{1/2}}{RV_{t-1}} \right) \ln RV_{t-1} + e_t$$

与 ARQ 模型不同的是，由于渐进方差的变化，我们除了需要计算 RQ 以外，还需要使用已实现波动率 RV 来近似积分波动率 IV。类似的，LogHARQ 模型的结构如下：

$$\text{LogHARQ} : \ln RV_t = \beta_0 + \left(\beta_1 + \beta_{1Q} \frac{RQ_{t-1}^{1/2}}{RV_{t-1}} \right) \ln RV_{t-1} + \beta_2 \ln RV_{t-1|t-5} + \beta_3 \ln RV_{t-1|t-22} + e_t$$

不做全部系数调整的理由和 Bollerslev 等(2016)一致。

三、实证检验

1、数据选取

本文采取沪深 300 股指和上证 50ETF 两个研究样本，时间选用沪深 300 股指自 2007 年 1 月 4 日发布以来至 2016 年 12 月 30 日。选取这两个数据的原因是以这两个数据对应了我国市场上两种重要的衍生工具，股指期货和指数期权。由于 2016 年 1 月 4 日和 7 日因熔断机制发生作用导致数据缺失，我们剔除了这两个交易日的数据。最后的数据集共包含 2403 个交易日的五分钟高频数据，数据采集自锐思高频数据库。其基本统计量见表 1：

表 1：描述性分析

	上证 50ETF				沪深 300 股指			
	2007-2011		2012-2016		2007-2011		2012-2016	
	RV	lnRV	RV	lnRV	RV	lnRV	RV	lnRV
均值	2.76E-04	-8.555	2.09E-04	-9.141	3.34E-04	-8.408	1.96E-04	-9.162
中位数	1.84E-04	-8.600	9.46E-05	-9.266	2.11E-04	-8.465	9.31E-05	-9.282
最大值	4.09E-03	-5.498	4.87E-03	-5.324	4.76E-03	-5.348	4.77E-03	-5.345
最小值	1.91E-05	-10.864	6.89E-06	-11.885	2.32E-05	-10.673	9.55E-06	-11.559
标准差	3.11E-04	0.817	4.29E-04	1.024	3.73E-04	0.876	3.80E-04	0.989
偏度	4.918	0.284	6.502	0.653	3.956	0.254	6.552	0.680
峰度	43.177	3.007	54.849	3.770	30.172	2.656	58.578	3.834
ADF	-8.765***	-5.491***	-7.540***	-4.347***	-8.472***	-5.598***	-7.540***	-4.347***

注：单位根检验采取滞后 5 阶，表中***表示在 1%水平上显著。峰度为原始峰度。

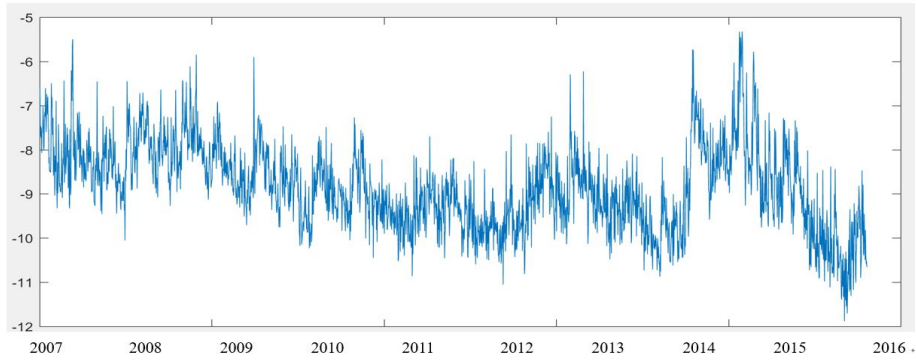


图 1: 上证 50ETF 的 lnRV 序列

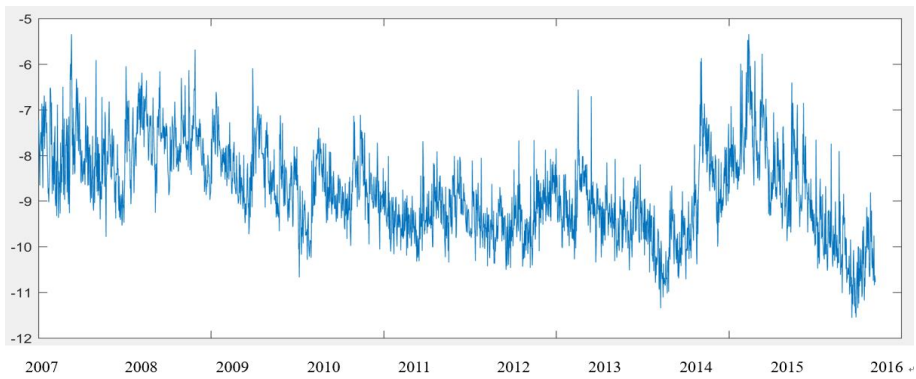


图 2: 沪深 300 指数的 lnRV 序列

日度 lnRV 的描述性统计如表 1 所示, ADF 单位根检验在 1%显著性水平下拒绝存在单位根的原假设, 显示所研究的波动率序列为平稳时间序列, 因此可以进一步分析和建模。以 2011 年末为分界点, 从数据分段上看: 第二段数据的方差波动性比第一个阶段更高一些, 可能是 2015 年股灾导致的波动率异常波动造成。从图 1 和 2 中我们可以看到, 取对数后的日度 RV 相对更为平滑, 但仍可以看出 2015 年股灾期间, 无论是上证 50ETF 还是沪深 300 股指, lnRV 对应的 RV 较大。因此对 2007-2016 年分段分别对前后五年数据进行分析对比。

2、模型估计

如上文所述, LogHARQ 模型的估计仍然可以使用传统的最小二乘估计方法来实现。表 2 给出了两个指数的全样本估计参数, 分别考虑了四种不同的模型设定: 不含测量误差调整的 LogAR 和 LogHAR, 含误差调整的 LogARQ 和 LogHARQ。

表 2: 全样本的估计结果

	上证 50ETF				沪深 300 股指			
	LogAR	LogHAR	logARQ	LogHARQ	LogAR	LogHAR	LogARQ	LogHARQ
β_0	-1.966**	-0.502**	-1.624**	-0.070	-1.703**	-0.425**	-1.577**	-0.135
s.e.	(0.123)	(0.134)	(0.180)	(0.184)	(0.115)	(0.116)	(0.171)	(0.166)
β_1	0.778**	0.318**	0.820**	0.370**	0.807**	0.353**	0.822**	0.386**
s.e.	(0.014)	(0.027)	(0.021)	(0.032)	(0.013)	(0.027)	(0.020)	(0.031)
β_2		0.402**		0.390**		0.381**		0.374**
s.e.		(0.044)		(0.044)		(0.042)		(0.043)

β_3	0.223**		0.237		0.219**		0.228
s.e.	(0.037)		(0.037)		(0.035)		(0.035)
β_{1Q}		1.27E-03**	1.54E-03**		4.99E-04**		1.08E-03**
s.e.		(4.49E-04)	(4.27E-)		(4.36E-04)		(4.08E-
Adjusted R2	0.471	0.479	0.482	0.496	0.456	0.473	0.456

注：表中**表示在 5%水平上显著，标准误为 robust 标准误。

估计结果显示对于两个指数的四种模型设定，波动率 HAR 项 ($\beta_1, \beta_2, \beta_3$) 的系数均在 5%水平上显著，与传统 HAR 文献吻合。对于 LogARQ 和 LogHARQ 模型而言，测量误差调整项 β_{1Q} 项系数为正且在 5%水平上显著，说明调整项在统计上是有意义的。虽然调整模型的中一阶滞后项 $\ln RV_{t-1}$ 的系数是随 $RQ_{t-1}^{1/2}RV_{t-1}^{-1}$ 的变化而变化，但是考虑到 $RQ_{t-1}^{1/2}RV_{t-1}^{-1}$ 始终为正且模型的 β_1 估计值大于对应的无调整模型的 β_1 ，我们仍然可以做出考虑测量误差修正确实可以使得一阶滞后项的系数增强。

3、滚动窗口样本外预测

本文预测将同时采用滚动窗口 Rolling Window (RW)和延展窗口 Increasing Window (IW)两种方法，两者的区别在于：滚动窗口每次估计都是用固定长度的数据，每新加入一个最新的观测值就将估计窗口中最老的观测值去掉；扩展窗口每次估计的时候保持估计窗口的起点不变，估计窗口随着新数据的加入逐渐延展。在我们的实证中，滚动窗口的窗宽和延展窗口的初始窗宽均设为 1000，每天更新估计系数并基于其做一步预测。由于 $E(\exp(X))$ 和 $\exp(E(X))$ 并不相等，为了能和线性 HAR 模型作比较，我们假设：LogHAR 和 LogHARQ 模型的残差服从正态分布，从而我们可以使用对数正态分布的性质获得 F_t ：

$$F_t = \exp\left(\beta_0 + \left(\beta_1 + \beta_{1Q} \frac{RQ_{t-1}^{1/2}}{RV_{t-1}}\right) \ln RV_{t-1} + \beta_2 \ln RV_{t-1|t-5} + \beta_3 \ln RV_{t-1|t-22} + \frac{\sigma^2}{2}\right)$$

在预测评价方式上，我们选取文献中常用的均方误差损失函数(MSE)，拟似然函数型损失函数(QLIKE)，前者是对称型的似然函数，对波动率的高估和低估给予相同的惩罚，后者是非对称的损失函数，对波动率的低估予以更大的惩罚。：

$$MSE = \sum_{t=\tau}^T (RV_t - F_t)^2$$

$$QMLE = \sum_{t=\tau}^T \left[\frac{RV_t}{F_t} - \ln\left(\frac{RV_t}{F_t}\right) - 1 \right]$$

其中 F_t 是模型给出的预测值， RV_t 是真实波动率的度量。两种似然函数均为稳健损失函数，即使用波动率的无偏估计作真实波动率指标测试预测模型的表现可以得到一致的比较结果。

基于全样本的模型预测效果如表 3 所示，HARQ 在所有线性模型中表现较好，logHARQ

模型在所有以 Log 模型中表现最好，且预测效果整体上优于 HARQ 模型。

表 3 全样本的预测效果(相对损失函数)

		上证 50ETF				沪深 300 股指			
线性版本		AR	HAR	ARQ	HARQ	AR	HAR	ARQ	HARQ
MSE	RW	1.050	1.000	0.952	1.035	1.107	1.000	1.148	1.101
	IW	1.077	1.000	0.997	0.966	1.150	1.000	1.093	1.002
QLIKE	RW	1.349	1.000	1.160	1.006	1.439	1.000	1.321	1.035
	IW	1.406	1.000	1.175	0.963	1.551	1.000	1.150	0.939
log 版本		LogAR	LogHAR	LogARQ	LogHARQ	LogAR	LogHAR	LogARQ	LogHARQ
MSE	RW	1.132	1.000	1.072	0.986	1.186	1.000	1.128	0.989
	IW	1.029	1.000	0.998	0.986	1.074	1.000	1.046	0.995
QLIKE	RW	1.228	1.000	1.141	0.990	1.224	1.000	1.144	0.981
	IW	1.151	1.000	1.089	0.990	1.139	1.000	1.080	0.983

注：表中线性版本数据为各线性版本数据与线性 HAR 模型对应数据的比值，例如：HARQ 对应列为 HARQ 模型的相对损失函数对应值与 HAR 模型的相对损失函数对应值的比值。表中 log 版本数据为各 log 版本数据与 LogHAR 模型对应数据的比值，例如：LogHARQ 对应列为 LogHARQ 模型的相对损失函数对应值与 LogHAR 模型的相对损失函数对应值的比值。

表 4 以 2011 年年末为分界点分段对上证 50ETF 和沪深 300 股指的模型预测效果进行分析。首先，对比每个研究样本分别对应的两个阶段，可以看出：每个研究样本对应的前三列 HARQ/HAR、LogHAR/HAR、LogHARQ/HAR 三者的数值自左至右逐渐缩小，由此可以得出 HARQ、LogHAR 和 LogHARQ 模型的预测精度较传统的 HAR 模型逐步增强，且 LogHARQ 模型表现最优，其相对 HAR 模型改进约在 15%-30%，相对 LogHAR 模型改进在 5%以内。其次，对比前后两个时间段的 LogHARQ/HAR，LogHARQ/LogHAR 的值，可以得出：LogHARQ 模型在两个时间段内预测精度都较高，模型在两个阶段均适用。

表 4：分段数据预测效果

		上证 50ETF				沪深 300 股指			
		HARQ	LogHAR	LogHARQ	LogHARQ	HARQ	LogHAR	LogHARQ	LogHARQ/
		/HAR	/HAR	/HAR	/LogHAR	/HAR	/HAR	/HAR	LogHAR
2007-2011									
MSE	RW	1.030	0.817	0.802	0.981	0.976	0.793	0.789	0.995
	IW	0.983	0.748	0.733	0.979	0.971	0.743	0.735	0.989
QLIKE	RW	0.984	0.899	0.870	0.968	0.921	0.853	0.861	1.009
	IW	0.960	0.855	0.827	0.967	0.902	0.809	0.807	0.998
2012-2016									
MSE	RW	1.106	0.789	0.774	0.981	1.046	0.882	0.854	0.967
	IW	1.146	0.818	0.803	0.982	1.053	0.892	0.862	0.967
QLIKE	RW	0.961	0.779	0.754	0.968	0.895	0.823	0.814	0.989
	IW	1.029	0.822	0.798	0.971	0.919	0.858	0.847	0.988

注：表中 HARQ/HAR、LogHAR/HAR、LogHARQ/HAR 数据分别对应为各 Log 版本数据与线性 HAR 模型对应数据的比值，例如：LogHARQ/HAR 列为 LogHARQ 模型的相对损失函数对应值与线性 HAR 模型的相对损失函数对应值的比值。LogHARQ/LogHAR 对应值为 LogHARQ 模型相对损失函数对应值与 LogHAR 模型的损失函数对应值的比值。

在此基础上，以高 5%的水平为分界线对 RQ 采用分层的方法对模型的预测效果进行进一步的分析，如表 5 所示。从两个表中可以看出，在各时间段的各种分层情况下 HARQ/HAR、LogHAR/HAR、LogHARQ/HAR 三者的数值依次减小，除了得出与表 4 相同结论外，模型在高 5%RQ 的情况下相对于低 95%RQ 表现更好，由此证实 LogHARQ 在波动率具有高波动时发挥更好的预测效果。总的来说，通过考虑测量误差的异方差性和对已实现波动率取对数处理，LogHARQ 模型明显优于其他模型，且对于两个阶段均适用，能够有效提高对我国股票市场波动率的预测精度。

表 5: RQ 分层预测效果

		上证 50ETF				沪深 300 股指			
		HARQ	LogHAR	LogHARQ	LogHARQ	HARQ	LogHAR	LogHARQ	LogHARQ
		/HAR	/HAR	/HAR	/LogHAR	/HAR	/HAR	/HAR	/LogHAR
低 95%RQ (2007-2011)									
MSE	RW	0.968	0.834	0.820	0.983	0.902	0.814	0.820	1.000
	IW	0.956	0.790	0.777	0.984	0.887	0.779	0.780	0.994
QLIKE	RW	0.969	0.907	0.888	0.979	0.896	0.863	0.884	1.015
	IW	0.953	0.871	0.853	0.979	0.875	0.822	0.831	1.003
高 5%RQ (2007-2011)									
MSE	RW	1.552	0.671	0.647	0.964	1.468	0.655	0.585	0.950
	IW	1.149	0.492	0.460	0.935	1.447	0.543	0.487	0.947
QLIKE	RW	1.214	0.767	0.593	0.773	1.208	0.739	0.596	0.907
	IW	1.058	0.653	0.490	0.751	1.176	0.668	0.555	0.909
低 95%RQ(2012-2016)									
MSE	RW	0.945	0.815	0.795	0.975	0.965	0.893	0.881	0.982
	IW	0.986	0.851	0.830	0.976	0.977	0.908	0.896	0.981
QLIKE	RW	0.945	0.787	0.763	0.969	0.880	0.823	0.821	0.996
	IW	1.015	0.833	0.810	0.972	0.905	0.860	0.857	0.995
高 5%RQ(2012-2016)									
MSE	RW	1.793	0.679	0.686	1.011	1.141	0.871	0.822	0.950
	IW	1.807	0.684	0.693	1.013	1.141	0.872	0.824	0.951
QLIKE	RW	1.311	0.606	0.569	0.939	1.120	0.815	0.702	0.883
	IW	1.323	0.606	0.574	0.947	1.121	0.819	0.707	0.885

注：表中 HARQ/HAR、LogHAR/HAR、LogHARQ/HAR 数据分别对应为各 Log 版本数据与线性 HAR 模型对应数据的比值，例如：LogHARQ/HAR 列为 LogHARQ 模型的相对损失函数对应值与线性 HAR 模型的相对损失函数对应值的比值。LogHARQ/LogHAR 对应值为 LogHARQ 模型相对损失函数对应值与 LogHAR 模型的损失函数对应值的比值。

四、稳健型检验

为了检验 LogHARQ 模型的稳健型，下面将 1 分钟 RK 视为真实收益率，分别对不同频率的 RV 预测效果进行分析。表 6 利用 5 分钟、10 分钟以及 15 分钟频率下 RV 进行预测，表示了 LogHARQ 模型较线性 HARQ 模型预测精度的改进程度，可以看出：对于上证 50ETF 和沪深 300 股指而言，LogHARQ 模型基于不同频率的改进较为稳定，且远小于 1，因此

LogHARQ 模型可较好的适用于基于不同频率 RV 的波动率预测。

表 6: 不同频率 RV 的 Log 模型与线性模型对比

LogHARQ/HARQ		上证 50ETF			沪深 300 股指		
		RV5	RV10	RV15	RV5	RV10	RV15
MSE	RW	0.694	0.689	0.654	0.796	0.850	0.831
	IW	0.694	0.689	0.688	0.799	0.854	0.840
QLIKE	RW	0.837	0.816	0.674	0.975	0.932	0.860
	IW	0.831	0.808	0.711	0.981	0.951	0.893

注：表中数据为将 1 分钟 RK 视为真实收益率，采用 5 分钟、10 分钟以及 15 分钟频率下 RV 预测的情况下，LogHARQ 模型与线性 HARQ 模型损失函数的比值。

表 7 分析了在 RV 采用 5 分钟、10 分钟以及 15 分钟频率下，LogHARQ 模型与 5 分钟频率 RV 下 LogHARQ 模型的对比，对于上证 50ETF 和沪深 300 股指而言，相比 5 分钟、10 分钟以及 15 分钟 RV 的预测效果，LogHARQ 模型与 5 分钟频率下 LogHARQ 模型的比值依次减小，即 5 分钟 RV 的频率下，LogHARQ 模型的预测精度最优。

表 7: Log 模型下不同频率 RV 对比

LogHARQ/ LogHARQ-RV5		上证 50ETF			沪深 300 股指		
		RV5	RV10	RV15	RV5	RV10	RV15
MSE	RW	1.000	1.016	1.041	1.000	1.052	1.054
	IW	1.000	1.015	1.040	1.000	1.052	1.054
QLIKE	RW	1.000	1.034	1.028	1.000	0.996	0.957
	IW	1.000	1.035	1.030	1.000	0.995	0.958

注：表中数据为将 1 分钟 RK 视为真实收益率，采用 5 分钟、10 分钟以及 15 分钟频率下 RV 预测的情况下，LogHARQ 模型与基于 5 分钟 RV 的 LogHARQ 模型损失函数的比值。

五、结论

本文针对我国股市波动率高且波动率变化迅速的特点，使用 Huang 等(2017)提出的 LogHARQ 模型对沪深 300 和 50ETF 回报率序列进行建模和预测。实证结果表明：LogHARQ 模型无论是相比于原有的线性模型，还是其他的 Log 模型，预测效果都是最好的，且在波动率的高波动时期(2012-2016)更为适用；同时该模型在高 RQ 时表现更优。我国股市的高波动率对投资组合、风险管理和衍生品定价等方面具有重要影响，利用本文构建的 LogHARQ 模型除对于完善波动率的建模有重要的理论意义外，对投资者获得更为精准的波动率预测、合理优化投资组合配置与有效地规避风险具有重要的现实意义，通过更加准确地预测波动率帮助投资者度量市场风险并做出正确的投资决策。

参考文献

- [1] Engle R F, Bollerslev T. Modeling the Persistence of Conditional Variances[J]. Econometric Reviews, 1986, 5(1):1-50.
- [2] Andersen T G, Bollerslev T. Answering the Skeptics: Yes, Standard Volatility Models do

- Provide Accurate Forecasts[J]. *International Economic Review*, 1998, 39(4):885-905.
- [3] Andersen T, Bollerslev T, Diebold F, Labys P. Modeling and forecasting realized volatility[J]. *Econometrica*, 2003, 71(2): 579-625.
- [4] 王天一, 黄卓. 基于高频数据的波动率建模及应用研究评述[J]. *经济学动态*, 2012(03):141-146.
- [5] Engle R F, Gallo G M. A multiple indicators model for volatility using intra-daily data[J]. *Journal of Econometrics*, 2006, 131(1):3-27.
- [6] Shephard N, Sheppard K. Realising the future: forecasting with high-frequency-based volatility (HEAVY) models[J]. *Journal of Applied Econometrics*, 2010, 25(2):197-231.
- [7] Hansen P R, Huang Z, Shek H H. Realized GARCH: a joint model for returns and realized measures of volatility[J]. *Journal of Applied Econometrics*, 2012, 27(6):877-906.
- [8] Ghysels E, Santa-Clara P, Valkanov R. Predicting volatility: getting the most out of return data sampled at different frequencies[J]. *Journal of Econometrics*, 2006, 131(1-2):59-95.
- [9] Corsi F. A Simple Approximate long-memory model of realized volatility[J]. *Journal of Financial Econometrics*, 2009, 7(2): 174-196.
- [10] Andersen T G, Bollerslev T, Diebold F X. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility[J]. *Review of Economics & Statistics*, 2007, 89(4):701-720.
- [11] 王天一, 黄卓. 高频数据波动率建模——基于厚尾分布的 Realized GARCH 模型[J]. *数量经济技术经济研究*, 2012,29(05):149-161.
- [12] Hansen P R, Zhuo H. Exponential GARCH modeling with realized measures of volatility[J]. *Journal of Business & Economic Statistics*, 2016, 34(2): 269-287.
- [13] 张小斐, 田金方. 异质金融市场驱动的已实现波动率计量模型[J]. *数量经济技术经济研究*, 2011,28(09):140-153.
- [14] 马锋, 魏宇, 黄登仕, 夏泽安. 基于马尔科夫状态转换和跳跃的高频波动率模型预测[J]. *系统工程*, 2016,34(01):10-16.
- [15] 刘晓倩, 王健, 吴广. 基于高频数据 HAR-CVX 模型的沪深 300 指数的预测研究[J]. *中国管理科学*, 2017,25(06):1-10.
- [16] 文风华, 刘晓群, 唐海如, 杨晓光. 基于 LHAR-RV-V 模型的中国股市波动性研究[J]. *管理科学学报*, 2012,15(06):59-67.
- [17] 孙洁. 考虑跳跃和隔夜波动的中国股票市场波动率建模与预测 [J]. *中国管理科学*, 2014,22(06):114-124.
- [18] 瞿慧, 杨洋. 沪深 300 指数波动的多状态平滑转移异质自回归模型[J]. *统计与决策*, 2015 (09):34-37.
- [19] 龚旭, 文风华, 黄创霞, 杨晓光. 下行风险、符号跳跃风险与行业组合资产定价[J]. *中国管理科学*, 2017,25(10):1-10.

- [20] Bollerslev T, Patton A J, Quaedvlieg R. Exploiting the errors: A simple approach for improved volatility forecasting[J]. *Journal of Econometrics*, 2016, 192(1):1-18.
- [21] Zhuo H., Tianyi W., Yajing Wang. Forecasting the Volatility of Natural Gas Futures with Realized Variance: Does Measurement Error Matter? working paper. 2017
- [22] Barndorff-Nielsen O E, Shephard N. Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation[J]. *Journal of Financial Econometrics*, 2006, 4(1):1-30.