SOFT COMPUTING IN DECISION MAKING AND IN MODELING IN ECONOMICS



Optimal futures hedging strategies based on an improved kernel density estimation method

Xing Yu¹ · Xinxin Wang¹ · Weiguo Zhang² · Zijin Li³

Accepted: 23 August 2021 / Published online: 1 September 2021 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

In this paper, we study the hedging effectiveness of crude oil futures on the basis of the lower partial moments (LPMs). An improved kernel density estimation method is proposed to estimate the optimal hedge ratio. We investigate crude oil price hedging by contributing to the literature in the following twofold: First, unlike the existing studies which focus on univariate kernel density method, we use bivariate kernel density to calculate the estimated LPMs, wherein the two bandwidths of the bivariate kernel density are not limited to the same, which is our main innovation point. According to the criterion of minimizing the mean integrated square error, we derive the conditions that the optimal bandwidths satisfy. In the process of derivation, we make a distribution assumption locally in order to simplify calculation, but this type of local distribution assumption is far better than global distribution assumption used in parameter method theoretically and empirically. Second, in order to meet the requirement of bivariate kernel density for independent random variables, we adopt ARCH models to obtain the independent noises with related to the returns of crude oil spot and futures. Genetic algorithm is used to tune the parameters that maximize quasi-likelihood. Empirical results reveal that, at first, the hedging strategy based on the improved kernel density estimation method is of highly efficiency, and then it achieves better performance than the hedging strategy based on the traditional parametric method. We also compare the risk control effectiveness of static hedge ratio vs. time-varying hedge ratio and find that static hedging has a better performance than time-varying hedging.

Keywords Futures hedging \cdot Improved kernel density estimation \cdot ARCH model \cdot Lower partial moment \cdot Genetic algorithm \cdot Crude oil price

1 Introduction

Along with the expanding economic and business ties between countries and increasingly tense international situation, there are huge fluctuations in the prices of some important energy and a lot of uncertainty in the future, especially for the case with crude oil. During the past few days, international oil price fell sharply as a result of shock.

Xing Yu yuxing@mail.ccnu.edu.cn

On the one hand, OPEC, led by Saudi Arabia, and Russia have failed to reach an agreement on cutting output, and then Saudi Arabia launched a price war; on the other hand, the global spread of coronavirus pandemic creates a panic in the market. Take the crude oil price in March 9, 2020, as an example, the crude oil price went down 24%, which was the biggest one-day drop since the 1991 gulf war. In fact, as a kind of global commodity, crude oil can affect economic activities and financial markets, for example gold, oil and equities (Maghyereh et al. 2017), WTI crude oil futures returns and hedge funds (Zhang and Wu 2019), global crude oil market and China's commodity sectors (Meng et al. 2020), and so on. Therefore, under the background of highly volatile crude oil price, considering its complex risk transmission mechanism, people who need to hedge oil price risk are not limited to oil producers and refiners only, but also financial market participants and policy makers.

¹ School of Economics and Business Administration, Central China Normal University, Wuhan 430079, China

² School of Business Administration, South China University of Technology, Guangzhou 510640, China

³ Business School, University of Jinan, Jinan 250022, China

Hedging is one of the most important functions of futures markets. When hedging the risk of crude oil price, we have to establish a hedged portfolio. Computational problems arise when we embed spot and futures in a portfolio. The traditional parametric and semi-parametric methods usually assume that the joint distribution is known, which is likely to cause misspecification if we have no economic reason to prefer one functional form over another (Backus et al. 1998). For example, Feng et al. (2012) argue that the assumption of a certain type of distribution can cause biased results when studying carbon returns. By contrast, nonparametric kernel density does not require any prior information related to distributions and estimators are driven by real data (Li and Racine 2007), so the misspecification problem can be relieved to a large extent. For this purpose, kernel density estimation is adopted in this paper to fit the joint distribution of the hedged portfolio. There are a number of researches about financial problems by means of kernel density estimation. Bouezmarni and Rombouts (2010) adopted the gamma kernel density under the background of positive time series data for the sake of boundary problems and demonstrated the superiority of it. Harvey and Oryshchenko (2012) utilized kernel density estimations to describe probability density functions of stock market indexes. Shi et al. (2017) combined the Bayes discriminant approach based on the multivariate kernel density with the extension discriminant approach to advance the concreteness of discrimination. Yan and Han (2019) compared the performance of some normal mixture models and kernel density estimations in fitting the behavior of different stock returns. Since that the hedging research is related to the spot and futures returns, so we adopt bivariate kernel density estimation, at the same time, different from the existing literature which sets up a same bandwidth for different variables (Hazelton and Marshall 2009; Gramacki and Gramacki 2017), we assume two different bandwidth for spot and futures and find the optimal solutions by minimizing the mean integrated square error. In this process, normal distribution is assumed for simplifying calculation, but this assumption is solely used for acquiring optimal bandwidths and is local to some extent, which is different from the global distribution assumption in the traditional parameter method and has better performance empirically.

There is a condition for using the kernel density estimation that variables must be independent of each other, which is opposite to the fact that spot returns and futures returns are highly related. So we adopt autoregressive conditional heteroskedasticity (ARCH) model to separate two independent series from spot and futures returns, named noise terms in the model, and the density function of independent noised is estimated through kernel density. The ARCH model was introduced by Engle(1982), aiming to investigate the time-varying volatility of economic data and being used widely in financial market, especially in pricing financial derivatives and measuring investment risk. Giot and Laurent (2004) compared the performance of a model on the basis of the daily realized volatility and a daily ARCH type model, aiming to study the volatility of stocks and exchange rate returns. Catani and Ahlgren (2017) proposed a bootstrap combined equation-by-equation Lagrange multiplier test for ARCH errors in VAR models in order to overcome the difficulty of high dimensionality facing multivariate tests. Further, ARCH model also plays an important role in crude oil market volatility analysis. Cheong (2009) used ARCH model, which considers lots of crucial volatility facts just like clustering volatility, to discuss the time-varying volatility within some important crude oil markets. Nademi and Nademi (2018) conducted a price forecast of crude oil including OPEC, WTI and Brent by means of a semiparametric Markov switching AR-ARCH model. There is also one point we'd like to stress, although ARCH model is adopted, we do not want to research volatility, and the only purpose is to obtain two independent series.

For the risk management, a appropriate risk measure is consequential, and the adopted one in this paper is lower partial moment (LPM). The characteristics of LPM when measuring the risk include: (1) measurement of one-side risk, and the focus is the negative deviation from the target rate of return, that is, downside risk; in addition, by measuring the return characteristics of loss (Brogan and Stidham 2008), the lower partial moment can reflect the difference of investors' attitude towards profit and loss. (2) By setting different target rates of return and risk parameters, the LPM can contain the heterogeneity of investors. (3) LPM satisfies the subadditivity, monotonicity and transformation invariance as a coherent measure of risk. (4) Decision criterion based on the LPM conforms to the expected utility maximization criterion and the random dominant criterion, and it is not necessary to make special assumptions about the utility function. Due to the outstanding features of it, LPM has been the center of a large amount of studies. Demirer and Lien (2003) calculated the optimal hedging ratios and corresponding hedging performance as well as compared the results between short and long hedgers. Baghdadabad (2014) extended the n-degree A-DRM risk measures within the framework of n-degree LPM and then put up with a new MV model to evaluate the US investors' indications in respect to portfolio performance. Dai et al. (2017) calculated the optimal hedging ratios by means of minimizing the LPM. Jasemi et al. (2019) put up with an practical methodology to approximate the LPM of the first order to dealing with computational difficulties. In this paper, we deduce the hedging

strategy of crude oil futures based upon the lower partial moments (LPMs).

The rest of paper is structured as follows:Section 2 introduces kernel density estimation, deriving the functions which optimal bandwidths cater to. And then Section 3 introduces the ARCH model and solves the parameter estimation by genetic algorithm. We incorporate the kernel density into the LPMs and calculate the optimal hedging position in Section 4. Further, empirical analysis including the comparison between kernel density estimation and parametric method as well as static hedging and dynamic hedging is conducted in Section 5. Based on the research results, conclusions and suggestions for investors are provided in Section 6.

2 An improved kernel density estimation method

There are parametric, semi-parametric and nonparametric methods to determine the probability density function of the sample data, and the common nonparametric methods include histogram and kernel density estimation. The concept of histogram estimation is simple, but the result is discontinuous, that is, the density value will suddenly drop to zero at the regional boundary, while the kernel density has the advantage of continuous estimation, and it is an efficient nonparametric density estimation method. The expression of kernel density is as follows:

$$\hat{f}(x_1, x_2) = \frac{1}{nh_1h_2} \sum_{i=1}^n K\left(\frac{X_{1i} - x_1}{h_1}, \frac{X_{2i} - x_2}{h_2}\right) \tag{1}$$

where *n* is the number of sample, h_1 and h_2 represent the bandwidths or smooth parameters. In the existing research, h_1 and h_2 are generally considered to be the same, i.e., $h_1 = h_2 = h$, while, in this paper, we are not assuming they're the same. X_{1i} and X_{2i} are the two given sample series, $K(\cdot, \cdot)$ is kernel function. Many studies have pointed out that different kernel functions have little effect on the accuracy of kernel density estimation, and there is asymptotic normality for kernel estimation in most samples, so Gaussian kernel is selected in this paper.

Kernel density fuses the form with observation point as the center, and the performance depends on the bandwidth selection. If the bandwidth is too small, the whole estimation, especially the tail, will appear interference and have a tendency to increase variance; if the bandwidth is too large, the distribution characteristics will be masked, and overaveraging will make the estimator have a large deviation. When considering estimation at a single point, a natural measure is the mean square error(MSE), defined as

$$MSE(\hat{f}(x_1, x_2)) = E(\hat{f}(x_1, x_2) - f(x_1, x_2))^2$$
(2)

By standard elementary properties of mean and variance,

$$MSE(\hat{f}(x_1, x_2)) = (E\hat{f}(x_1, x_2) - f(x_1, x_2))^2 + var\hat{f}(x_1, x_2)$$
(3)

The first and most widely used way of placing a measure on the global accuracy of \hat{f} is the mean integrated square error (MISE) (Silverman 1986), defined as

$$MISE(\hat{f}(x_1, x_2)) = \iint MSE(\hat{f}(x_1, x_2)) dx_1 dx_2$$

= $E \iint (\hat{f}(x_1, x_2) - f(x_1, x_2))^2 dx_1 dx_2$
= $\iint (E\hat{f}(x_1, x_2) - f(x_1, x_2))^2 dx_1 dx_2$
+ $\iint var\hat{f}(x_1, x_2) dx_1 dx_2$
(4)

which gives the MISE as the sum of the integrated square bias and the integrated variance.

Let $y_1 = X_{1i}, y_2 = X_{2i}, t_1 = \frac{y_1 - x_1}{h_1}, t_2 = \frac{y_2 - x_2}{h_2}$, and the kernel function $K(\cdot, \cdot)$ is a symmetric function satisfying:

$$\iint K(t_{1}, t_{2}) dt_{1} dt_{2} = 1,$$

$$\iint t_{1}K(t_{1}, t_{2}) dt_{1} dt_{2} = 0,$$

$$\iint t_{2}K(t_{1}, t_{2}) dt_{1} dt_{2} = 0$$
(5)

As was pointed out earlier, the calculation of bias is not determined by the size of sample (n) but rather the bandwidth (h_1, h_2) , of course, if the calculation of bandwidth depends on the *n*, then the bias will depend on *n* through its dependence on *h*. The approximation expression of bias is obtained as follows:

 $bias(x_1, x_2)$

$$= E\hat{f}(x_{1}, x_{2}) - f(x_{1}, x_{2})$$

$$= \iint \frac{1}{h_{1}h_{2}} K\left(\frac{y_{1} - x_{1}}{h_{1}}, \frac{y_{2} - x_{2}}{h_{2}}\right) f(y_{1}, y_{2}) dy_{1} dy_{2}$$

$$-f(x_{1}, x_{2})$$

$$= \iint K(t_{1}, t_{2})[f(x_{1} + h_{1}t_{1}, x_{2} + h_{2}t_{2}) - f(x_{1}, x_{2})] dt_{1} dt_{2}$$

$$+ \frac{1}{2}h_{2}^{2}t_{2}^{2}\frac{\partial^{2}f}{\partial x_{2}^{2}}] dt_{1} dt_{2}$$

$$= \frac{1}{2}h_{1}^{2}\frac{\partial^{2}f}{\partial x_{1}^{2}} \iint t_{1}^{2}K(t_{1}, t_{2}) dt_{1} dt_{2}$$

$$+ h_{1}h_{2}\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}} \iint t_{1}t_{2}K(t_{1}t_{2}) dt_{1} dt_{2}$$

$$+ \frac{1}{2}h_{2}^{2}\frac{\partial^{2}f}{\partial x_{2}^{2}} \iint t_{2}^{2}K(t_{1}, t_{2}) dt_{1} dt_{2}$$

$$= \frac{1}{2}h_{1}h_{2}\left(\frac{h_{1}}{h_{2}}\frac{\partial^{2}f}{\partial x_{1}^{2}}k_{1} + 2\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}}k_{2} + \frac{h_{2}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}}k_{3}\right)$$
(6)

where,

$$k_{1} = \iint t_{1}^{2} K(t_{1}, t_{2}) dt_{1} dt_{2},$$

$$k_{2} = \iint t_{1} t_{2} K(t_{1} t_{2}) dt_{1} dt_{2},$$

$$k_{3} = \iint t_{2}^{2} K(t_{1}, t_{2}) dt_{1} dt_{2}$$
(7)

By integrating the result above, we can get the following one:

$$\iint (E\hat{f}(x_{1}, x_{2}) - f(x_{1}, x_{2}))^{2} dx_{1} dx_{2}$$

$$\approx \frac{1}{4} h_{1}^{2} h_{2}^{2} \iint \left[\frac{h_{1}}{h_{2}} \frac{\partial^{2} f}{\partial x_{1}^{2}} k_{1} + 2 \frac{\partial^{2} f}{\partial x_{1} \partial x_{2}} k_{2} + \frac{h_{2}}{h_{1}} \frac{\partial^{2} f}{\partial x_{2}^{2}} k_{3} \right]^{2} dx_{1} dx_{2}$$
(8)

We now turn to the variance,

$$\begin{aligned} \operatorname{var}\widehat{f}(x_{1}, x_{2}) &= E(\widehat{f}(x_{1}, x_{2})^{2}) - (E\widehat{f}(x_{1}, x_{2}))^{2} \\ &= \frac{1}{nh_{1}h_{2}} \iint K^{2}(t_{1}, t_{2})f(x_{1} + h_{1}t_{1}, x_{2} + h_{2}t_{2}) dt_{1} dt_{2} \\ &\quad -\frac{1}{n} \left[\iint K(t_{1}, t_{2})f(x_{1} + h_{1}t_{1}, x_{2} + h_{2}t_{2}) dt_{1} dt_{2} \\ &\quad -f(x_{1}, x_{2}) + f(x_{1}, x_{2}) \right]^{2} \\ &= \frac{1}{nh_{1}h_{2}} \iint K^{2}(t_{1}, t_{2}) \left[f(x_{1}, x_{2}) + h_{1}t_{1} \frac{\partial f}{\partial x_{1}} \\ &\quad + h_{2}t_{2} \frac{\partial f}{\partial x_{2}} + \frac{1}{2}h_{1}^{2}t_{1}^{2} \frac{\partial^{2}f}{\partial x_{1}^{2}} + h_{1}h_{2}t_{1}t_{2} \frac{\partial^{2}f}{\partial x_{1}\partial x_{2}} \\ &\quad + \frac{1}{2}h_{2}^{2}t_{2}^{2} \frac{\partial^{2}f}{\partial x_{2}^{2}} \right] dt_{1} dt_{2} - \frac{1}{n} [f(x_{1}, x_{2}) + o(h_{1}h_{2})]^{2} \\ &= \frac{1}{nh_{1}h_{2}} f(x_{1}, x_{2}) \iint K^{2}(t_{1}, t_{2}) dt_{1} dt_{2} + o\left(\frac{1}{n}\right) \\ &\approx \frac{1}{nh_{1}h_{2}} f(x_{1}, x_{2}) \iint K^{2}(t_{1}, t_{2}) dt_{1} dt_{2} \end{aligned}$$

The result is obtained by using the approximation for the bias and assuming that h_1, h_2 is small and *n* is large. Further, we have

$$\iint var\hat{h}(x_1, x_2) \, dx_1 \, dx_2 = \frac{1}{nh_1h_2} \iint K^2(t_1, t_2) \, dt_1 \, dt_2$$
(10)

The expressions of MISE and AMISE can be obtained according to the analysis mentioned above:

$$\begin{split} MISE\hat{f}(x_{1}, x_{2}) \\ &= \frac{1}{4}h_{1}^{2}h_{2}^{2} \iint \left[\frac{h_{1}}{h_{2}}\frac{\partial^{2}f}{\partial x_{1}^{2}}k_{1} + 2\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}}k_{2} \right. \\ &\quad \left. + \frac{h_{2}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}}k_{3}\right]^{2}dx_{1}dx_{2} + \frac{1}{nh_{1}h_{2}}\iint K^{2}(t_{1}, t_{2})dt_{1}dt_{2} \\ &\quad \left. + o\left(h_{1}^{2}h_{2}^{2} + \frac{1}{nh_{1}h_{2}}\right)\right) \\ AMISE\hat{f}(x_{1}, x_{2}) \\ &= \frac{1}{4}h_{1}^{2}h_{2}^{2} \iint \left[\frac{h_{1}}{h_{2}}\frac{\partial^{2}f}{\partial x_{1}^{2}}k_{1} + 2\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}}k_{2} \right. \\ &\quad \left. + \frac{h_{2}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}}k_{3}\right]^{2}dx_{1}dx_{2} + \frac{1}{nh_{1}h_{2}}\iint K^{2}(t_{1}, t_{2})dt_{1}dt_{2} \end{split}$$

Then we can get the optimal window width h_1^* and h_2^* by calculating the follow equations:

(11)

$$\begin{cases} \frac{\partial AMISE\hat{f}(x_1, x_2)}{\partial h_1} = 0\\ \frac{\partial AMISE\hat{f}(x_1, x_2)}{\partial h_2} = 0 \end{cases}$$
(12)

That is, the optimal window widths satisfy:

$$\frac{1}{2}h_{1}h_{2}^{2} \iint \left[\frac{k_{1}h_{1}}{h_{2}}\frac{\partial^{2}f}{\partial x_{1}^{2}} + 2k_{2}\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}} + \frac{k_{3}h_{2}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}}\right]^{2} dx_{1} dx_{2} + \frac{1}{2}h_{1}^{2}h_{2}^{2} \iint \left[\frac{k_{1}^{2}h_{1}}{h_{2}^{2}}\left(\frac{\partial^{2}f}{\partial x_{1}^{2}}\right)^{2} \right] + \frac{2k_{1}k_{2}}{h_{2}}\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}}\frac{\partial^{2}f}{\partial x_{1}^{2}} - \frac{2k_{2}k_{3}h_{2}}{h_{1}^{2}}\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}}\frac{\partial^{2}f}{\partial x_{2}^{2}} \\- \frac{k_{3}^{2}h_{2}^{2}}{h_{1}^{3}}\left(\frac{\partial^{2}f}{\partial x_{2}^{2}}\right)^{2}\right] dx_{1} dx_{2} - \frac{1}{nh_{1}^{2}h_{2}}\iint \left[\frac{k_{1}h_{1}}{h_{2}}\frac{\partial^{2}f}{\partial x_{1}^{2}} + 2k_{2}\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}} + \frac{k_{3}h_{2}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}}\right]^{2} dx_{1} dx_{2} + \frac{1}{2}h_{1}^{2}h_{2}^{2}\iint \left[\frac{2k_{2}k_{3}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}} + 2k_{2}\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}} + \frac{k_{3}h_{2}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}}\right]^{2} dx_{1} dx_{2} + \frac{1}{2}h_{1}^{2}h_{2}^{2}\iint \left[\frac{2k_{2}k_{3}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}} + \frac{k_{3}h_{2}}{h_{1}}\frac{\partial^{2}f}{\partial x_{2}^{2}}\right]^{2} dx_{1} dx_{2} + \frac{k_{3}^{2}h_{2}}{h_{1}^{2}}\left(\frac{\partial^{2}f}{\partial x_{1}^{2}}\right)^{2} - \frac{2k_{1}k_{2}h_{1}}{h_{2}^{2}}\frac{\partial^{2}f}{\partial x_{1}\partial x_{2}}\frac{\partial^{2}f}{\partial x_{1}^{2}} \\+ \frac{k_{3}^{2}h_{2}}{h_{1}^{2}}\left(\frac{\partial^{2}f}{\partial x_{1}^{2}}\right)^{2} dx_{1} dx_{2} - \frac{1}{nh_{1}h_{2}^{2}}\iint K^{2}(t_{1}, t_{2}) dt_{1} dt_{2} = 0$$
(13)

The solutions of Eqs. (13) depend on the real density function. Assume that $\eta_1 \sim N(0, \sigma_1^2), \eta_2 \sim N(0, \sigma_2^2)$, and they are independent of each other. It should be emphasized that the normal assumption here is only a local assumption made in the derivation of the optimal window width, which is substantially different from the global assumption made in the parametric method. The joint density of η_1 and η_2 is

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\left(\frac{x_1^2}{2\sigma_1^2} + \frac{x_2^2}{2\sigma_2^2}\right)\right\}$$
(14)

We think this as the real density of population, and the derivative part contained in the above two equations can be expressed as follows:

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{x_2 - \sigma_1^2}{2\pi\sigma_1^5\sigma_2} \exp\left\{-\left(\frac{x_1^2}{2\sigma_1^2} + \frac{x_2^2}{2\sigma_2^2}\right)\right\}$$
$$\frac{\partial^2 f}{\partial x_1^2} = \frac{x_2^2 - \sigma_2^2}{2\pi\sigma_1\sigma_2^5} \exp\left\{-\left(\frac{x_1^2}{2\sigma_1^2} + \frac{x_2^2}{2\sigma_2^2}\right)\right\}$$
$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{x_1 x_2}{2\pi\sigma_1^3\sigma_2^3} \exp\left\{-\left(\frac{x_1^2}{2\sigma_1^2} + \frac{x_2^2}{2\sigma_2^2}\right)\right\}$$
(15)

At the same time, for the $\hat{f}(x_1, x_2)$ in Eq. (1), we adopt Gaussian kernel, and k_1, k_2 and k_3 are calculated as follows:

$$k_1 = 1, k_2 = 0, k_3 = 1 \tag{16}$$

Then, Eqs. (13) can be simplified as follows:

$$\begin{cases} nh_1^2h_2(3h_1^3\sigma_2^2 + h_1h_2^2\sigma_1^2) - 4\sigma_1^5\sigma_2^3 = 0\\ nh_1h_2^2(3h_2^3\sigma_1^2 + h_1^2h_2\sigma_2^2) - 4\sigma_1^3\sigma_2^5 = 0 \end{cases}$$
(17)

By solving the equations, we can obtain the new optimal window widths (h_1^*, h_2^*) , for which we can estimate the kernel density $\hat{f}(x_1, x_2)$:

$$\hat{f}(x_{1}, x_{2}) = \frac{1}{nh_{1}^{*}h_{2}^{*}} \sum_{i=1}^{n} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(\left(\frac{X_{1i} - x_{1}}{h_{1}^{*}}\right)^{2} + \left(\frac{X_{2i} - x_{2}}{h_{2}^{*}}\right)^{2}\right)\right\}$$
(18)

3 Independent sequences from ARCH Model

Since the sample data are not independent of each other in finance, insurance and other aspects, it would be a mistake to estimate the kernel density directly using the relevant data. Therefore, we use the ARCH model to fit the returns of spot and futures prices, and further to obtain the independent errors. Based on the independent errors, we estimate the optimal bandwidth for binary kernel density.

ARCH model is able to describe the time-varying volatility of economic data, and the generalized ARCH model can further depict the clustering of volatility, that is, volatility will change as time goes by as well as present an relatively high or low situation at some time. Of course, ARCH model is used here just for separating independent series and has nothing to do with volatility. The fundamental content of ARCH model is shown as follows:

$$\begin{cases} \mathbf{X}_{t} = \boldsymbol{\varphi} \mathbf{X}_{t-1} + \varepsilon_{t} \\ \varepsilon_{t} = \mathbf{D}_{t} \eta_{t} \\ \mathbf{D}_{t} = diag(\sqrt{h_{1t}}, \sqrt{h_{2t}}) \\ h_{1t} = w_{1} + A_{11}\varepsilon_{1,t-1}^{2} + A_{12}\varepsilon_{2,t-1}^{2} \\ h_{2t} = w_{2} + A_{21}\varepsilon_{1,t-1}^{2} + A_{22}\varepsilon_{2,t-1}^{2} \end{cases}$$
where
$$\mathbf{X}_{t} = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}, \varepsilon_{t} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \eta_{t} = \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \varphi = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

 $\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}$, and $w_1, w_2, A_{11}, A_{12}, A_{21}, A_{22}$ are constant param-

eters that should be estimated.

Since that the distribution of η_t is unknown, so here the quasi-likelihood estimation method is adopted. That is, we maximize the following criterion function to obtain the quasi-likelihood estimation of parameters.

$$L = \frac{1}{n} \sum_{t=1}^{n} \left(-\frac{1}{2} \ln(\det(\mathbf{D}_{t} \Gamma \mathbf{D}_{t})) - \frac{1}{2} \varepsilon_{t}^{T} (\mathbf{D}_{t} \Gamma \mathbf{D}_{t})^{-1} \varepsilon_{t} \right)$$
(20)

Then we deduce the concrete form of criterion function, as we all know,

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} \eta_{1t}\sqrt{h_{1t}} \\ \eta_{2t}\sqrt{h_{2t}} \end{pmatrix}, \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} X_{1t} - \varphi_1 X_{1t-1} \\ X_{2t} - \varphi_2 X_{2t-1} \end{pmatrix}$$
(21)

Let
$$\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$
. We have
 $\varepsilon_t^T (\mathbf{D}_t \Gamma \mathbf{D}_t)^{-1} \varepsilon_t$

$$= \left(\eta_{1t} \sqrt{h_{1t}} \quad \eta_{2t} \sqrt{h_{2t}} \right) \begin{pmatrix} \frac{1}{h_{1t}} & 0 \\ 0 & \frac{1}{h_{2t}} \end{pmatrix} \begin{pmatrix} \eta_{1t} \sqrt{h_{1t}} \\ \eta_{2t} \sqrt{h_{2t}} \end{pmatrix}$$

$$= \eta_{1t}^2 + \eta_{2t}^2$$
(22)

In this way, the likelihood function can be expressed as:

$$L = -\frac{1}{2n} \sum_{t=1}^{n} (\ln(h_{1t}h_{2t}) + \eta_{1t}^2 + \eta_{2t}^2)$$
(23)

and

$$\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = D_t^{-1} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} \frac{X_{1t} - \varphi_1 X_{1t-1}}{\sqrt{h_{1t}}} \\ \frac{X_{2t} - \varphi_2 X_{2t-1}}{\sqrt{h_{2t}}} \end{pmatrix}$$
(24)

So, it yields

$$\eta_{1t}^2 = \frac{(X_{1t} - \varphi_1 X_{1t-1})^2}{h_{1t}}, \eta_{2t}^2 = \frac{(X_{2t} - \varphi_2 X_{2t-1})^2}{h_{2t}}$$
(25)

Then, the likelihood function is shown as follows:

$$L = -\frac{1}{2n} \sum_{t=1}^{n} \left(\ln(h_{1t}h_{2t}) + \frac{(X_{1t} - \varphi_1 X_{1t-1})^2}{h_{1t}} + \frac{(X_{2t} - \varphi_2 X_{2t-1})^2}{h_{2t}} \right)$$
(26)

In parallel, we know that,

$$h_{1t} = w_1 + A_{11}(X_{1t-1} - \varphi_1 X_{1t-2})^2 + A_{12}(X_{2t-1} - \varphi_2 X_{2t-2})^2 h_{2t} = w_2 + A_{21}(X_{1t-1} - \varphi_1 X_{1t-2})^2 + A_{22}(X_{2t-1} - \varphi_2 X_{2t-2})^2$$
(27)

Finally, based on the given data, we can rewrite the likelihood function as follows

$$L = -\frac{1}{2n} \sum_{t=1}^{n} (Y_{1t} + Y_{2t} + Y_{3t} + Y_{4t})$$
(28)

where,

$$Y_{1t} = \ln(w_1 + A_{11}(X_{1t-1} - \varphi_1 X_{1t-2})^2 + A_{12}(X_{2t-1} - \varphi_2 X_{2t-2})^2)$$

$$Y_{2t} = \ln(w_2 + A_{21}(X_{1t-1} - \varphi_1 X_{1t-2})^2 + A_{22}(X_{2t-1} - \varphi_2 X_{2t-2})^2)$$

$$Y_{3t} = \frac{(X_{1t} - \varphi_1 X_{1t-1})^2}{w_1 + A_{11}(X_{1t-1} - \varphi_1 X_{1t-2})^2 + A_{12}(X_{2t-1} - \varphi_2 X_{2t-2})^2}$$

$$Y_{4t} = \frac{(X_{2t} - \varphi_2 X_{2t-1})^2}{w_2 + A_{21}(X_{1t-1} - \varphi_1 X_{1t-2})^2 + A_{22}(X_{2t-1} - \varphi_2 X_{2t-2})^2}$$

To estimate the parameters in the ARCH model, Alzghool and Al-Zubi (2018) adopted semi-parametric methods including quasi-likelihood and asymptotic quasi-likelihood estimation. For the problem of numerical implementation of model structure choice, approach, which is based on genetic algorithm, is proposed. It is a heuristic search algorithm used for solving optimization and modeling tasks by random selection, combination and variation of the required parameters with the use of mechanisms that resemble biological evolution. A distinctive feature of genetic algorithm is an emphasis on the use of "crossing" operator, which makes an operation of recombination of solution candidates, whose role is similar to that of crossing in living nature. In this paper, GA is used to tune the parameters that maximize quasi-likelihood.

4 Lower Partial Moments

LPM is associated with downside risk, according to Bawa and Linderberg (1997) and Lien and Tse (2001); its expression is shown as follows:

$$L(c, m, r_p) = E[max(0, c - r_p)]^m$$
(29)

where *c* is the expected return and *n* is the power of the shortfall, the higher *c* is, the investors expect a higher return; *m* represents the risk aversion coefficient, if m < 1, the investors appetite for risk, and if m > 1, the investors are risk-aversion. In particular, let m = 0, the LPM is the equal of value-at-risk (VaR); when m = 1, the LPM is equivalent to conditional value at risk (CVaR); when c = 0 and m = 2, the LPM is similar to semi-variogram of Markowitz. In addition, r_p is the hedged portfolio return, and $r_p = r_s - Hr_f$, in which r_s is the spot return, r_f is the futures return and *H* is the hedged ratio.

Based on ARCH model, we can express r_s and r_f as follows:

 $r_s = r_1 + \sqrt{h_1}\eta_1$

and

 $r_f = r_2 + \sqrt{h_2}\eta_2$

Then we incorporate the noise into LPM:

$$L = E[max(0, c - r_1 - \sqrt{h_1}\eta_1 - H(r_2 + \sqrt{h_2}\eta_2))]^m$$

= $\iint_{D_1} [c - r_1 - \sqrt{h_1}x_1 - H(r_2 + \sqrt{h_2}x_2)]^m$
 $f(x_1, x_2) dx_1 dx_2$ (30)

Here, $D_1 = c - r_1 - \sqrt{h_1}x_1 - H(r_2 + \sqrt{h_2}x_2) \ge 0$. $f(x_1, x_2)$ are the joint density of η_1 and η_2 . Whenever the joint distribution of r_s and r_f is known, we can apply numerical methods to find the optimal hedge ratio. Due to the fact that the true distribution of rs and rf is unknown, so we adopt an indirect method to estimate the distribution of the hedged portfolio returns considering any given c. Specifically, for a given c, we construct the data series for η_1 and η_2 from the data of r_s and r_f , and then apply nonparametric methods to estimate the distribution of set as follows.

5 Minimum LPM Hedged Ratios

Further, we incorporate the calculated kernel density into the LPM. For the calculation of optimal hedging ratios, traditional approach called static hedging figures out a constant value by minimizing the risk measure, which originated from Johnson (1960) and Stein (1961), who select an optimal futures position to minimize the variance of the hedged portfolio. Then Ghosh (1993) adopted the error correction model to calculate the constant hedge ratio based on the cointegration theory. Although the static hedging strategy has been widely used in existing literature, it ignores the time-varying characteristic of the (co)-variance between the spot and futures returns. Qu et al. (2019) investigated the dynamic hedging performance of China's CSI 300 index futures, utilizing the high-frequency intraday information with RMVHR-based models. So we calculate the optimal hedging ratios of static and dynamic hedging, respectively.

5.1 Optimal hedged ratios based on the static Hedging

The optimal hedged ratios are calculated based on the whole sample data. Based on Eq. (30), the expression of LPMs is written as follows:

$$L = E[max(0, c - r_1 - \sqrt{h_1\eta_1 - H(r_2 + \sqrt{h_2\eta_2})})]^m$$

= $\sum_{i=1}^n \iint_{D_2} [c - r_{1i} - \sqrt{h_{1i}x_1 - H(r_{2i} + \sqrt{h_{2i}x_2})}]^m$
= $\frac{1}{nh_1^*h_2^* 2\pi} \exp\left\{-\frac{1}{2}\left[\left(\frac{X_{1i} - x_1}{h_1^*}\right)^2 + \left(\frac{X_{2i} - x_2}{h_2^*}\right)^2\right]\right\} dx_1 dx_2$

where $D_2: c - r_{1i} - \sqrt{h_{1i}}x_1 - H(r_{2i} + \sqrt{h_{2i}}x_2) \ge 0$. Let

$$I_{1} = \int_{-\infty}^{D_{3}} \frac{1}{2n\pi h_{1}^{*}h_{2}^{*}} [c - r_{1i} - \sqrt{h_{1i}}x_{1i} - H(r_{2i} + \sqrt{h_{2i}}x_{2i})]^{m} \exp\left\{-\frac{1}{2}\left(\frac{X_{1i} - x_{1}}{h_{1}^{*}}\right)^{2}\right\} dx_{1}$$

Here, $D_3: \frac{c - r_{1i} - H(r_{2i} + \sqrt{h_{2i}}x_2)}{\sqrt{h_{1i}}}$, then we have

$$I_{1} = \int_{0}^{+\infty} \frac{1}{\sqrt{h_{1i}} 2n\pi h_{1}^{*} h_{2}^{*}} u^{m} \exp\left\{-\frac{1}{2} \left(\frac{\sqrt{h_{1i}} X_{1i} - c + u + r_{1i} + H(r_{2i} + \sqrt{h_{2i}} x_{2})}{\sqrt{h_{1i}} h_{1}^{*}}\right)^{2}\right\} du$$
(31)

Therefore, the LPMs are expressed by

$$L = \sum_{i=1}^{n} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\frac{X_{2i} - x_2}{h_2^*}\right)^2\right\} I_1 \, dx_2 \tag{32}$$

We can obtain the optimal hedged ratio by calculating $\frac{\partial L}{\partial H} = 0$, that is, the optimal hedged ratio satisfies the following equation:

.

$$\sum_{i=1}^{n} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2} \left(\frac{X_{2i} - x_2}{h_2^*}\right)^2\right\} \frac{\partial I_1}{\partial H} \, dx_2 = 0 \tag{33}$$

According to Eq. (31), we have

$$\frac{\partial I_1}{\partial H} = \int_0^{+\infty} u^m \exp{-\frac{1}{2} \left(\frac{A}{\sqrt{h_{1i}}h_1^*}\right)^2 \frac{A(r_{2i} + \sqrt{h_{2i}}x_2)}{-h_1^*(h_{1i})^{\frac{3}{2}}}} \quad (34)$$

where $A = \sqrt{h_{1i}}X_{1i} + u - c + r_{1i} + H(r_{2i} + \sqrt{h_{2i}}x_2)$.

For the different values of m, we can deduce the condition that the optimal hedge ratio satisfies. The results are shown in the following proposition.

Proposition 1 Suppose a hedger want to hedge the downside risk measured by LPMs with a static hedging strategy. The optimal hedge ratio H^* , therefore, satisfies the following conditions:

• when m = 0, the optimal hedged ratio H^* is solved from the following equation

$$\sum_{i=1}^{n} \exp\left\{-\frac{1}{2} \frac{\left(aH^{*}+b\right)^{2}}{h_{1}^{*2}h_{1i}+h_{2}^{*2}H^{*2}h_{2i}}\right\}$$

$$\frac{ah_{1}^{*2}h_{1i}-bH^{*}h_{2}^{*2}h_{2i}}{\left(h_{1}^{*2}h_{1i}+H^{*2}h_{2}^{*2}h_{2i}\right)^{\frac{3}{2}}}=0$$
(35)

where $a = \sqrt{h_{2i}}X_{2i} + r_{2i}$, $b = \sqrt{h_{1i}}X_{1i} - c + r_{1i}$. X_{1i}, X_{2i} are the return series of spot and futures for the given data. h_1^*, h_2^* are the best bandwidths estimated based on Eqs. (17). And, h_{1i}, h_{2i} are obtained from Eq. (27).

• when m = 1, the optimal hedged ratio H^* is solved from the following equation

$$\sum_{i=1}^{n} \int_{-\infty}^{+\infty} \frac{v}{\sqrt{h_{2i}}} \exp\left\{-\frac{1}{2}\left(\frac{a-v}{\sqrt{h_{2i}}h_{2}^{*}}\right)^{2}\right\}$$

$$\Phi\left(\frac{-b-H^{*}v}{\sqrt{h_{1i}}h_{1}^{*}}\right) dv = 0$$
(36)

• when m = 2, the optimal hedged ratio H^* is solved from the following equation

$$\sum_{i=1}^{n} \int_{-\infty}^{+\infty} \sqrt{\frac{2\pi}{h_{2i}}} (bv + H^{*}v^{2})$$

$$\exp\left\{-\frac{1}{2} \left(\frac{a-v}{\sqrt{h_{2i}}h_{2}^{*}}\right)^{2}\right\} \Phi\left(\frac{-b-H^{*}v}{\sqrt{h_{1i}}h_{1}^{*}}\right) dv$$

$$+ \sum_{i=1}^{n} \frac{h_{1}^{*2}h_{2}^{*}h_{1i}\sqrt{2\pi h_{2i}} (ah_{1}^{*2}h_{1i} - bH^{*}h_{2}^{*2}h_{2i})}{(h_{1}^{*2}h_{1i} + H^{*2}h_{2}^{*2}h_{2i})^{\frac{3}{2}}}$$

$$\exp\left\{-\frac{1}{2} \frac{(aH^{*} + b)^{2}}{h_{1i}^{*2}h_{1i} + h_{2}^{*2}H^{*2}h_{2i}}\right\} = 0$$
(37)

5.2 Optimal hedged ratios based on the dynamic Hedging

Different from the static hedging, the optimal hedged ratio in every day changes according to the market states. The LPMs in day k (k = 1, 2, 3...n) is expressed as follows:

$$L_{k} = E[max(0, c - r_{1} - \sqrt{h_{1}}\eta_{1} - H(r_{2} + \sqrt{h_{2}}\eta_{2}))]^{m}$$

$$= \sum_{i=1}^{n} \iint_{D_{2}} [c - r_{1k} - \sqrt{h_{1k}}x_{1} - H(r_{2k} + \sqrt{h_{2k}}x_{2})]^{m}$$

$$= \frac{1}{nh_{1}^{*}h_{2}^{*}} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[\left(\frac{X_{1i} - x_{1}}{h_{1}^{*}}\right)^{2} + \left(\frac{X_{2i} - x_{2}}{h_{2}^{*}}\right)^{2}\right]\right\}$$

$$= \sum_{i=1}^{n} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\frac{X_{2i} - x_{2}}{h_{2}^{*}}\right)^{2}\right\} I_{2} dx_{2}$$
(38)

where

$$H_{2} = \int_{-\infty}^{D_{3}} \frac{1}{2n\pi h_{1}^{*}h_{2}^{*}} \left[c - r_{1i} - \sqrt{h_{1i}}x_{1} - H(r_{2i} + \sqrt{h_{2i}}x_{2})\right]^{m} \exp\left\{-\frac{1}{2}\left(\frac{X_{1i} - x_{1}}{h_{1}^{*}}\right)^{2}\right\} dx_{1}$$

and $D_3: \frac{c-r_{1k}-H(r_{2k}+\sqrt{h_{2k}x_2})}{\sqrt{h_{1i}}}$ Then we can get the optimal hedged ratio H_k by calculating the first-order condition of $\frac{\partial L_k}{\partial H_k} = 0$, that is, the optimal hedged ratio satisfy the following equation:

$$\sum_{i=1}^{n} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2} \left(\frac{X_{2i} - x_2}{h_2^*}\right)^2\right\} \frac{\partial I_2}{\partial H_k} \, dx_2 = 0 \tag{39}$$

Here,

$$\frac{\partial I_2}{\partial H_k} = \int_0^{+\infty} u^m \exp\left(-\frac{1}{2} \left(\frac{A}{\sqrt{h_{1k}}h_1^*}\right)^2 \frac{A(r_{2k} + \sqrt{h_{2k}}x_2)}{-h_1^*(h_{1k})^{\frac{3}{2}}}\right)$$

and

$$A = \sqrt{h_{1k}}X_{1i} + u - c + r_{1k} + H(r_{2k} + \sqrt{h_{2k}}x_2)$$

For the different values of m, we can deduce the condition that the optimal dynamic hedge ratio in day k satisfies. The results are shown in the following proposition.

Proposition 2 Suppose a hedger want to hedge the downside risk measured by LPMs with a dynamic hedging strategy. The optimal hedge ratio H_k^* in day k, therefore, satisfies the following conditions:

• when m = 0, the optimal dynamic hedged ratio H_k^* satisfy the following equation

$$\sum_{i=1}^{n} \exp\left\{-\frac{1}{2} \frac{(aH_{k}^{*}+b)^{2}}{h_{1}^{*2}h_{1k}+h_{2}^{*2}H_{k}^{*2}h_{2k}}\right\}$$

$$\frac{ah_{1}^{*2}h_{1k}-bH_{k}^{*}h_{2}^{*2}h_{2k}}{(h_{1}^{*2}h_{1k}+H_{k}^{*2}h_{2}^{*2}h_{2k})^{\frac{3}{2}}}=0$$
(40)

where $a = \sqrt{h_{2k}}X_{2i} + r_{2k}, b = \sqrt{h_{1k}}X_{1i} - c + r_{1k}.$

• when *m* = 1, the optimal dynamic hedged ratios satisfy the following equation

$$\sum_{i=1}^{n} \int_{-\infty}^{+\infty} \frac{v}{\sqrt{h_{2k}}} \exp\left\{-\frac{1}{2}\left(\frac{a-v}{\sqrt{h_{2k}}h_2^*}\right)^2\right\}$$

$$\Phi\left(\frac{-b-H_k^*v}{\sqrt{h_{1k}}h_1^*}\right) dv = 0$$
(41)

• when m = 2, the optimal dynamic hedged ratios satisfy the following equation

$$\sum_{i=1}^{n} \int_{-\infty}^{+\infty} \sqrt{\frac{2\pi}{h_{2k}}} (bv + H_{k}^{*}v^{2}) \exp\left\{-\frac{1}{2}\left(\frac{a-v}{\sqrt{h_{2k}}h_{2}^{*}}\right)^{2}\right\}$$

$$\Phi\left(\frac{-b-H_{k}^{*}v}{\sqrt{h_{1k}}h_{1}^{*}}\right) dv$$

$$+ \sum_{i=1}^{n} \frac{h_{1}^{*2}h_{2}^{*}h_{1k}\sqrt{2\pi h_{2k}} (ah_{1}^{*2}h_{1k} - bH_{k}^{*}h_{2}^{*2}h_{2k})}{(h_{1}^{*2}h_{1k} + H_{k}^{*2}h_{2}^{*2}h_{2k})^{\frac{3}{2}}}$$

$$\exp\left\{-\frac{1}{2}\frac{(aH_{k}^{*}+b)^{2}}{h_{1}^{*2}h_{1k} + h_{2}^{*2}H_{k}^{*2}h_{2k}}\right\} = 0$$
(42)

6 Empirical Study

In this section, we achieve the following tasks. First, descriptive statistics for spot and futures returns. Second, estimation of relevant parameters in ARCH model through genetic algorithm. Third, optimal hedged ratios and corresponding effectiveness are calculated according to different objective return (c) and risk aversion coefficient (m) of LPMs, and comparisons, including kernel density versus parametric method under the framework of static hedging, static hedging versus dynamic hedging by kernel density, kernel density versus parametric method in dynamic hedging, are made. The conclusions are at the end.

6.1 Data

According to the ex ante versus ex post method (Alizadeh et al. 2015; Ghoddusi and Emamzadehfard 2017), we divide the history day data of WTI crude oil into two parts

for the sake of static hedging research. The former part for the in-sample analysis covers the period between January 2, 2015, and April 7, 2018, while the latter part for the outof-sample analysis covers from April 8, 2018 to October 11, 2019. For the dynamic hedging, in order to simplify the calculation, we select 100 samples included in the sample data mentioned above to accomplish the test. The in-sample analysis covers the period between January 2, 2015, and March 16, 2015, while the out-of-sample analysis covers from April 8, 2018, to June 4, 2018. The optimal bandwidths calculated of in-sample and out-of-sample are $h_1^* = 0.2405, h_2^* = 0.0881$ and $h_1^* = 0.1992, h_2^* = 0.0701$, respectively. Here is the descriptive statistic of the whole data in Figure 1:

From Fig. 1, we can clearly notice the volatility clustering among the estimators of noise. Further, we test the ARCH effects which are shown in Table 1.

For Table 1, the upper gives summary statistics on returns while the lower presents the results of ARCH effect test. It is obvious that there exists positive or negative skewness or kurtosis among the in-sample and out-ofsample data, especially for the case with futures returns in sample which have the largest skewness and kurtosis, that's to say, it is more appropriate to adopt kernel density to estimate the distribution of returns rather than normal assumption. In addition, the LM(K) statistic delineates the existence of ARCH effect for spot and futures returns, which identifies the rationality of our usage of ARCH model to fit the return data and obtain the independent noise series.

6.2 Parameter estimation of ARCH model

Genetic algorithm is adopted in this paper to solve the parameter estimation problem of ARCH model, which has been widely used as a high-efficiency optimization instrument. The GA was proposed first by Holland (1975), which operates directly on the structure object without the limitation of derivative and continuity of function. According to Abdullah et al. (2018), the GA can conduct a multidirectional search within crowds of candidate solutions, which allows the seeds of possible success to be spread uniformly over the whole solution space and make itself achieve success in the process of optimizing compared to single search point-based algorithms. Genetic algorithm is a kind of stochastic algorithm, developing randomly generated individuals for better solution by iterative process, and the definition of the survival of the fittest of this algorithm is a process to find the optimal offspring, and the ultimately generated individual is the optimal solution within the optimization process. Each individual represents a solution of the optimization problem, and the fitness is



Fig. 1 Estimators of noise from in-sample to out-of-sample

Table 1 Descriptive statistic of returns and Engle tests

		In-sample			Out-	of-sample		
		spot	fut	tures	spot		futures	
Mean -0.0008		_(0.0016	0.00	91	0.0041		
Median -0.0265		_(-0.0024		082	-0.008	88	
Standard deviation 0.7280		0.7280	0.2667		0.5549		0.1844	
Variance 0.5300		0.5300	0.0711		0.3080		0.0340	
Excess kurtosis		4.3167	13	.8038	7.8448		9.5273	
Skewness	Skewness –0		0.6	5398	-0.0619		-0.1502	
Range		5.7676	3.7390		5.4997		2.0740	
Minimum		-3.0707	-1.4327		-3.2471		-1.308	39
Maximum		2.6969	2.3	3063	2.25	26	0.7651	
Engle test	Q-statistic	P-value	Q-statistic	P-value	Q-statistic	P-value	Q-statistic	P-value
LM(4)	216.20	0.000	216.20	0.000	5.2838	0.000	6.1253	0.000
LM(6)	258.09	0.000	258.09	0.000	15.151	0.000	8.3496	0.000
LM(8)	M(8) 332.75 0.000		332.75	0.000	17.650	0.000	12.356	0.000
LM(10)	413.66	0.000	413.66	0.000	18.553	0.000	12.698	0.000

used as the evaluation index. Fitness represents the survival chance of the individuals. The higher the fitness is, the higher the probability of the individual entering the next iteration. In practical optimization problems, fitness is usually the value of objective function. During iteration, new individuals are generated by crossover operators and mutation operators, and two different generations are generated by random combination and exchange of elements in a pair of individuals by crossover operators, while the mutation operator adds some small random changes to the offspring. Genetic algorithm can set reinitialization after each convergence to ensure that the most suitable individuals are retained in the iteration process and new random individuals can be created at the same time, so as
 Table 2
 Parameter estimators of static hedging

Parameter	w_1	<i>w</i> ₂	A_{11}	A_{12}	A_{21}	A ₂₂	ϕ_1	ϕ_2
in-sample	0.0012	0.0133	0.0354	-0.0622	-1.1058	-0.8922	0.1944	-0.1108

to reduce the risk of premature convergence of the algorithm. The parameters of ARCH model for in-sample data are estimated and are presented in Table 2.

6.3 Empirical results of static hedging

Static hedging means that the optimal hedged ratios and effectiveness are calculated according to all the sample data, with all the sample data as a whole. At the same time, the results based on the kernel density estimation are compared with the ones under parameter method which assumes a normal distribution through the standardization of sample according to the center limit theorem. Their results of in-sample test are shown in Tables 3 and 4, and Table 5 depicts the result of out-of-sample test.

From Tables 3 and 4, we can firstly confirm that all the hedging effectiveness is bigger than zero, so the model constructed by us to solve the hedging problem is effective. Then we can see that situations differ with the change of risk aversion coefficient. For the case m = 0, compared with kernel density estimation, the hedged ratios are relatively smaller while the effectiveness is higher for the most data of results of parametric. For example, when c = 0.01, the former position and effectiveness are 0.34 and 0.48 while the latter ones are 0.21 and 0.51. When m = 1, it is difficult to tell which one is better, because two results are similar. For the case c = -0.002, there are a relatively smaller position and a relatively higher effectiveness in parametric method, but the opposite is true for the case c = -0.005. Different from the previous results, when m = 2, the result of kernel density estimation achieves a

 Table 3 Optimal LPM hedge ratios based on the kernel density estimation

-	better	perfo	ormar	ice.	When	c =	-0.01,	the	position	is	larg	ger

while the effectiveness is lower in parametric method; in addition, for the same efficiency, the positions calculated by parametric method are generally larger. Next we turn to the out-of-sample results which are shown in Table 5.

For the results using kernel density estimation, whether m = 0, m = 1 or m = 2, the hedging effectiveness of outof-sample test is higher than the ones of in-sample test generally; on the contrary, effectiveness from out-of-sample test becomes smaller compared with the results of insample for parametric method. Finally, combination of the in-sample analysis is likely that kernel density estimation represents the real distribution characteristics of data in financial market better and achieves a better hedging performance.

6.4 The empirical results of dynamic hedging

Dynamic hedging means that the calculation of the optimal hedged ratios and effectiveness is based on the single daily data; for the all the observations, we can get *n* results. Here c = 0. At first, we compare static hedging and dynamic hedging under the framework of kernel density. The results of in-sample and out-of-sample test as well as the comparison with static hedging are shown in Figures 2 and 3, in which the value represented by the straight line is the result of static hedging with same target return(c = 0) and risk aversion coefficient(*m*).

From Figs. 2 and 3, we find that for the in-sample result, considering the optimal hedging ratios, there are half the points above and half the points below the straight line, so there is no particular benefit to using one approach over the

c	Hedging	ratios		Effectiveness				
	$\mathbf{m} = 0$	m = 1	m = 2	$\mathbf{m} = 0$	m = 1	m = 2		
0.01	0.34	0.32	0.17	0.48	0.85	0.79		
0.005	1.12	1.48	0.16	0.59	0.88	0.76		
0.002	0.96	0.86	0.08	0.52	0.81	0.75		
0	0.58	0.33	0.13	0.43	0.74	0.76		
-0.002	0.94	0.93	0.86	0.64	0.86	0.82		
-0.005	0.44	0.40	0.08	0.35	0.69	0.79		
-0.01	0.66	0.53	0.06	0.30	0.37	1.00		

Table 4 Optimal LPM hedge ratios based on the parametric method

c	Hedged ratio			Effectiveness			
	$\mathbf{m} = 0$	m = 1	m = 2	$\mathbf{m} = 0$	m = 1	m = 2	
0.01	0.21	0.33	0.25	0.51	0.86	0.86	
0.005	0.49	0.69	0.98	0.61	0.89	0.94	
0.002	0.50	0.25	0.01	0.59	0.79	0.82	
0	0.92	0.72	0.23	0.73	0.87	0.86	
-0.002	0.61	0.16	0.72	0.60	0.74	0.92	
-0.005	0.96	0.91	0.21	0.70	0.87	0.86	
-0.01	0.42	0.34	0.30	0.46	0.70	0.87	

Table 5 Comparison based on the out-of-sample test

c	Kernel d	lensity esti	mation	Parametric method			
	$\mathbf{m} = 0$	m = 1	m=2	$\mathbf{m} = 0$	m = 1	m = 2	
0.01	0.49	0.89	0.87	0.46	0.84	0.68	
0.005	0.65	0.92	0.95	0.61	0.87	0.57	
0.002	0.63	0.81	0.79	0.51	0.77	0.55	
0	0.75	0.90	0.85	0.31	0.62	0.57	
-0.002	0.64	0.72	0.93	0.63	0.82	0.72	
-0.005	0.72	0.89	0.85	0.16	0.48	0.68	
-0.01	0.42	0.74	0.88	0.33	0.58	0.99	

other. Then we turn to the effectiveness; it is obvious that the effectiveness obtained by static hedging is higher than most of results of dynamic hedging. The similar conclusion can be acquired from the out-of-sample test, that is, static hedging strategy achieves better performance. In addition, it is of crucial importance that, whether for optimal hedging ratios or for effectiveness, the results of dynamic hedging are discrete and unstable, what's more, there are many invalid points that the effectiveness is below the zero. Further, we incorporate the calculated optimal hedged ratios into the portfolios $r_p = r_s - Hr_f$, finding different wealth paths, and there is a descriptive statistic about returns shown in Table 6 and 7.

From Tables 6 and 7, we can see that all the mean and most of medians of static hedging strategy are bigger than those of dynamic hedging strategy, as for variance, although the values of static hedging are little bigger for the in-sample test, the opposite is the truth with out-ofsample test, that is, static hedging strategy achieves better performance. In a word, we think static hedging based on the whole sample is a more appropriate hedging strategy. The above content compares the performance of static and dynamic hedging under the framework of kernel density; more importantly, in order to prove the superiority of our improved kernel density, the comparison between kernel density and parametric method should be made under the framework of dynamic hedging. The optimal hedging ratios and efficiency based on the in-sample data as well as the results from out-of-sample data are shown in Table 8. Because there is a large amount of data in results, difficult to show in pictures, so we compare them in a statistical sense.

For the in-sample results, by comparing mean and median, it is easy to find that the optimal hedging ratios calculated by kernel density are smaller, which means lower cost, than those obtained by parametric method, while the efficiency calculated by kernel density is higher. For out-of-sample efficiency, we also find kernel performs better. So the conclusion can be drawn that the strategy based on the kernel density achieves better hedging performance with lower cost compared with parametric method, which proves the superiority of our improved kernel density again.



Fig. 2 In-sample hedging performances of static and dynamic hedging



Fig. 3 Out-of-sample hedging performances of static and dynamic hedging

	$\mathbf{m} = 0$		m = 1		m = 2		
	Static	Dynamic	Static	Dynamic	Static	Dynamic	
Mean	0.00002	-0.00062	-0.00005	-0.0077	-0.00012	-0.00140	
Variance	0.00078	0.00013	0.00065	0.00018	0.00060	0.00013	
Median	-0.00008	-0.00160	0.00007	-0.00110	-0.00018	-0.00180	
Minimum	-0.11910	-0.02570	-0.10910	-0.03770	-0.11350	-0.02570	
Maximum	0.09320	0.02930	0.08180	0.02970	0.08460	0.02970	

Table 7 Out-of-sample hedging
returns of static and dynamic
hedging

Table 6 In-sample hedgingreturns of static and dynamic

hedging

	m = 0		m = 1		m=2	m = 2		
	Static	Dynamic	Static	Dynamic	Static	Dynamic		
Mean	0.00154	0.00088	0.00020	-0.00022	0.00025	-0.00028		
Variance	0.00004	0.00084	0.00039	0.00063	0.00035	0.00057		
Median	0.00010	0.00038	-0.00028	-0.00008	-0.00009	-0.00300		
Minimum	-0.11650	-0.1170	-0.11460	-0.11290	-0.1130	-0.11780		
Maximum	0.07200	0.1048	0.0714	0.04630	0.07080	0.04880		

7 Conclusion

The LPMs measures an individual hedger's downside risk, as opposed to the two-sided risk measure. This study proposed an improved kernel density estimation to estimate the optimal hedge ratio of crude oil futures hedging based on LPMs. Our goal in this paper is twofold: (a) Due to the correlation between spot and futures returns, we extend the kernel method to the bivariate case. Furthermore, different from the existing literature, for the spot and futures returns, we assume different optimal bandwidths which are calculated by minimizing the mean integrated square error. (b) In order to get independent time series, we adopt ARCH model which relevant parameters are estimated by means of genetic algorithm. The purpose of this treatment is to satisfy the independent sequence requirement of binary kernel density estimation. In the part of empirical analysis, comparisons, including kernel density versus parametric method under the framework of static hedging, static hedging versus dynamic hedging by kernel density, kernel density versus parametric method in dynamic hedging, are made.
 Table 8
 The comparison

 between kernel density and
 parametric method

Sample	m	Content	Method	Mean	Max	Min	Median
Sample	III	Content	Wiethou	Ivicali	IVIAN	WIIII	wiculan
In-sample	$\mathbf{m}=0$	Position	Kernel	0.45	1.12	0.23	0.66
			Parametric	0.56	1.34	0.30	0.71
Hedging		Efficiency	Kernel	0.34	0.55	0.09	0.37
			Parametric	0.22	0.47	0.05	0.29
Position	m = 1	Position	Kernel	0.57	1.34	0.32	0.62
			parametric	0.66	1.21	0.34	0.64
And		Efficiency	Kernel	0.31	0.55	0.12	0.37
			parametric	0.25	0.58	0.07	0.26
Efficiency	m = 2	Position	Kernel	0.67	0.97	0.35	0.56
			parametric	0.71	1.01	0.32	0.59
		Efficiency	Kernel	0.24	0.46	0.09	0.27
			parametric	0.12	0.31	0.04	0.14
	$\mathbf{m}=0$	Efficiency	Kernel	0.49	0.67	0.23	0.45
Out-of-sample hedging efficiency			Parametric	0.44	0.69	0.21	0.42
	m = 1	Efficiency	Kernel	0.56	0.81	0.35	0.54
			Parametric	0.32	0.66	0.12	0.35
	m = 2	Efficiency	Kernel	0.26	0.60	0.09	0.23
		-	Parameter	0.23	0.71	0.05	0.22

Empirical results reveal that, at first, the hedging strategy based on the kernel density estimation method is of highly efficiency, and then it achieves better performance than the hedging strategy based on the traditional parametric method (normal) under the framework of both static hedging and dynamic hedging, that is, smaller hedged ratios and higher effectiveness, which proves the superiority and robustness of our improved kernel density fully. What's more, in accordance with the comparison of optimal positions, effectiveness and returns, we come to the conclusion that the results of static hedging strategy are better and more stable due to the incorporation of more sample points while the results of dynamic hedging strategy are inefficient, discrete and unstable.

Last but not least, when calculating optimal bandwidths, normal distribution is assumed for simplifying calculation, which is local to some extent in kernel density and is different from the global distribution assumption in the traditional parameter method. So how to avoid dependence on distributions altogether and obtain the optimal bandwidths through simple calculation in the case of higher dimensions will be challenging and rewarding.

Acknowledgements This paper is supported by Funds for International Cooperation and Exchange of the National Natural Science Foundation of China (71720107002); National Natural Science Foundation of China (No.71501076, 71971086); Guangdong Basic and Applied Basic Research Foundation (No. 2019B151502037); Financial Service Innovation and Risk Management Research Base of Guangzhou; The Raising initial capital for High-level Talents of Central China Normal University (30101190001); Fundamental Research Funds for the Central Universities (CCNU19A06043, CCNU19TD006, CCNU 19TS062, No. 2019ZD13); Fundamental Research Funds for the Central Universities (Innovation Funding Projects)(2020CXZZ047); Humanities and Social Science Planning Fund from Ministry of Education (Grant No.21YJC790148).

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Abdullah Y, Birdal S, Ufuk Y (2018) Maximum likelihood estimation for the parameters of skew normal distribution using genetic algorithm. Swarm and Evolutionary Comput 38(2):127–138
- Alizadeh AH, Huang CY, Dellen SV (2015) A regime switching approach for hedging tanker shipping freight rates. Energy Econom 49(3):44–59
- Alzghool R, Al-Zubi LM (2018) Semi-parametric estimation for ARCH models. Alexandria Eng J 57(1):367–373
- Angle R (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50:987–1007
- Backus D, Foresi S, Zin S (1998) Arbitrage opportunities in arbitragefree models of bond pricing. J Business and Econom Statistics 16:13–24
- Baghdadabad MRT (2014) Average drawdown risk reduction and risk tolerances. Res Econom 68(3):264–276
- Bawa VS, Linderberg E (1997) Capital market equilibrium in a mean lower partial moment frame work. J Financial Econom 5(2):189–200
- Bouezmarni T, Rombouts JVK (2010) Nonparametric density estimation for positive time series. Comput Statistics Data Anal 54(2):245–261

- Brogan AJ, Stidham S (2008) Non-separation in the mean Clowerpartial-moment portfolio optimization problem. Eur J Operat Res 184:701–710
- Catani PS, Ahlgren NJC (2017) Combined Lagrange multiplier test for ARCH in vector autoregressive models. Econom Statistics 1:62–84
- Cheong CW (2009) Modeling and forecasting crude oil markets using ARCH-type models. Energy Policy 37(6):2346–2355
- Dai J, Zhou HG, Zhao SQ (2017) Determining the multi-scale hedge ratios of stock index futures using the lower partial moments method. Physica A: Statistical Mech Appl 466(15):502–510
- Demirer R, Lien D (2003) Downside risk for short and long hedgers. Int Rev Econom Finance 12:25–44
- Feng ZH, Wei YM, Wang K (2012) Estimating risk for the carbon market via extreme value theory: an empirical analysis of the EU ETS. Appl Energy 99:97–108
- Ghoddusi H, Emamzadehfard S (2017) Optimal hedging in the US natural gas market: the effect of maturity and cointegration. Energy Econom 63(3):92–105
- Ghosh A (1993) Hedging with stock index futures: estimation and forecasting with error correction model. J Futures Market 13(7):743–752
- Giot P, Laurent S (2004) Modelling daily Value-at-Risk using realized volatility and ARCH type models. J Empirical Finance 11(3):379–398
- Gramacki A, Gramacki J (2017) FFT-based fast bandwidth selector for multivariate kernel density estimation. Comput Statistics & Data Anal 106:27–45
- Harvey A, Oryshchenko V (2012) Kernel density estimation for time series data. Int J Forecast 28(1):3–14
- Hazelton ML, Marshall JC (2009) Linear boundary kernels for bivariate density estimation. Statistics & Probab Lett 79(8):999–1003
- Holland J (1975) Adaptation in natural and artificial system: an introduction with application to biology, control and artificial intelligence. University of Michigan Press, Ann Arbor
- Jasemi M, Monplaisir L, Jam PA (2019) Development of an efficient method to approximate the risk measure of lower partial moment of the first order. Comput Ind Eng 135:326–332

- Johnson LL (1960) The theory of hedging and speculation in commodity futures. Rev Econom Stud 27(3):139–151
- Li Q, Racine JS (2007) Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton, NJ
- Lien D, Tse YK (2001) Hedging downside risk: futures vs. options. Int Rev Econom Finance 10(2):159–169
- Maghyereh AI, Awartani B, Tziogkidis P (2017) Volatility spillovers and cross-hedging between gold, oil and equities: Evidence from the gulf cooperation council countries. Energy Econom 68(10):440–453
- Meng J, Nie H, Jiang YH (2020) Risk spillover effects from global crude oil market to China's commodity sectors. Energy 117208. https://doi.org/10.1016/j.energy.2020.117208
- Nademi A, Nademi Y (2018) Forecasting crude oil prices by a semiparametric Markov switching model: OPEC, WTI, and Brent cases. Energy Econom 74:757–766
- Qu H, Wang TY, Zhang Y, Sun PF (2019) Dynamic hedging using the realized minimum-variance hedge ratio approach - Examination of the CSI 300 index futures. Pacific-Basin Finance J 57:101048. https://doi.org/10.1016/j.pacfin.2018.08.002
- Shi JL, Zhu SH, Zhou YY, Li RH (2017) Bayes-extension discriminant method of two populations based on multivariate kernel density estimation. Procedia Comput Sci 122:780–787
- Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
- Stein JL (1961) The simultaneous determination of spot and futures prices. Am Econom Rev 51:1012–1025
- Yan HH, Han LY (2019) Empirical distributions of stock returns: mixed normal or kernel density? Physica A: Statistical Mech Appl 514:473–486
- Zhang YJ, Wu YB (2019) The time-varying spillover effect between WTI crude oil futures returns and hedge funds. Int Rev Econom Finance 16(5):156–169

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.