Comput Econ (2017) 50:111–140
DOI 10.1007/s10614-016-9584-1

Wavelets Analysis on Structural Model for Default Prediction

Lu Han[1] · **Ruihuan Ge**[1]

Accepted: 17 May 2016 / Published online: 3 June 2016

Abstract In recent years, to improve predictive ability of corporate defaults has become an important problem. In this paper, regarding on characteristics of listed companies, we sampled 100 companies according to industry types, constructed wavelet structural model, experimented with wavelet decomposition proceeds to get low frequency and high frequency sequence, built the prediction model for both sequences, and then using the prediction of future returns to reconstruct predictive returns, thus avoiding accumulated prediction process with earnings volatility of time series model, therefore enhanced the precision of default prediction. Finally we compared wavelet structural model with time series structural model based on the predictive default distance of China's listed companies.

Keywords Wavelet structural model · Time series analysis · Default prediction · Credit risk management

1 Introduction

Credit risk based on the characteristics of the debtor is often divided into sovereign, corporate, retail, etc. Retail debt is centered on customer credit, which includes short-term and intermediate-term credit to finance the purchase of commodities and services for consumption or to refinance debt incurred for such purposes. Corporate credit evaluation methods now are almost based on quantitative analysis. A good credit risk evaluation tool can help to grant credit to more creditworthy applicants and thus

✉ Lu Han
hanluivy@126.com

1 School of Management Science and Engineering, Central University of Finance & Economics, Beijing 100081, China

increases profit. Moreover, it can deny credit for the noncredit worthy applicants and thus decreases losses.

Today academic researches of forcasting enterprise default are based on two major milestone results, one is structure model, on behalf of Merton model and KMV model; the other is credit scoring. The structure model has entered the BaselIII and aslo become a market standard of monitoring the default risk. But in practice, the structure model needs enterprise's equity value and the fluctuation of equity value which limited scope of this model, and the prediction effect depends directly on estimation results of these two parameters.

This paper deeply discussed the methods for estimating parameters of structure equation model, put forward the wavelet structure model. Through wavelet transform decomposed the revenue sequence into the low frequency and high frequency, one can modeled the two sequences part respectively. As it is known that the low frequency part contains more information and the high frequency part has more noise, so modeling them with different equations can minimize cumulative calculation error of equity volatility, and can get the core information for prediction sequences reconstruction, so all of these gave a lot of improvement in the default prediction which can be verified with the Chinese listed company.

The structure of the rest in this paper is as follows: the next section puts forward the prior research of structural model and wavelet analysis. Section 3 describes our methods for default prediction in detail. Section 4 is about experiment studies; in this part we will present several robust check result of structural models in China's actual practice. Section 5 is the robust test. The Final section discusses the interesting results and gives some remarks.

2 Prior Research

The first researcher who gave a deep research on default prediction for large listed companies can be traced back to Merton (1974). He constructed the structure equations to price debt, through appropriately simplifying the corporate capital structure and dynamic change of corporate value of the company; he matched the price of corporate debt and corporate equity with options. And then KMV (Crosbie and Bohn 2002) company developed an empirical estimators Probability of Default on the basis of Metron model, which is known as the estimation of expected default frequency (EDF), and it can be seen as a consistent estimator for probability of default (PD), rather than using the cumulative normal distribution in the Merton model to calculate PD.

At present, the research of structural models for default prediction mainly focuses on the actual assessment and improvement in estimators. Ye et al. (2005) using a small sample with 22 enterprises to adjust the parameters of the KMV model, which makes the model better adapt to the situation in China. Lee (2011) used genetic algorithm to improve the best default point of KMV model. Camara and Popova (2011) applying several structure models to evaluate the default risk in the financial enterprises after the subprime crisis, and they found that KMV model had better accuracy in default prediction. Chen et al. (2010) experimented with large sample of 80 enterprises between 2004 and 2006 to build the KMV model, and in his research it can be found that structural models cannot give early warning of default risk to small and medium-sized enterprise in China. Though structural models are be seen as the most effective methods in the default prediction of large companies, it cannot avoid calculation of yield volatility which is also the key of these models.

Now, the majority studies focus on the calculation of yield volatility is time series modeling, which relies on linearity and symmetry assumptions. However, several authors have discussed in detail the inadequacy of linear models in capturing asymmetries. Importantly, Hamilton (2003) has settled that non-linear specifications should be seen as better candidate models than traditional linear approaches in capturing significantly much more stronger effects of oil shocks. Chiou and Lee (2009) argued that most of the time series models experience structural changes that when applied to real data, determine the break locations. Whilst there is general agreement in the literature that any inferences without consideration of regime switching phenomenon may well lead to unreliable results for many financial time series. Regimes switching models occurred as an alternative to standard GARCH models in allowing dynamic variables' behavior to depend on the state that takes place at any given point in time.

Undoubtedly, GARCH models worked well to capture leptokurtosis and volatility clustering generally observed in financial time series but they demonstrate some inaccuracies in terms of changes of time scales (Yalamova 2006). One major advantage afforded by wavelets analysis is the ability to perform local analysis e that is, to analyze a localized sub image area of a larger image (or signal). Therefore, wavelet analysis is capable of revealing aspects of data that other signal analysis techniques (like GARCH models) usually miss; aspects like trends, sharp spikes, discontinuities in higher derivatives, self-similarity.etc. Likewise, because it affords a different view of data than those presented by traditional techniques, wavelet analysis can often compress or de-noise a signal without appreciable degradation (Nguyen and Nabney 2010). In their brief history within the signal processing field, wavelets have already proven themselves to be a very useful tool for data de-noising (Chen et al. 1986) and deconvolution (separation between two convolved signals namely smooth and detail).Wavelet analysis provides better resolutions in the time domain since wavelet basis functions are time-localized, which is useful for capturing the changing volatility by Jagric Yogo (2008). Jagric (2002, 2003, 2004), Raihan et al. (2005), and Fernandez and Kutan (2005) use wavelet analysis to find out cyclical components. Crowley et al. (2005) use wavelet transforms (MODWT, and CWT) to analyze productivity cycles in Euro area, US and UK. Crivellini et al. (2003) apply wavelet analysis to analyze industrial output fluctuations in developing countries and concludes that time scale decomposition through wavelet analysis may reveal very different aspects in the characteristics and correlations of business cycle fluctuations. Yamada and Honda (2005) use wavelet analysis to predict business turning points of Nikkei 225 index and find that wavelet analysis can capture business peaks and troughs (minimum points) as an alternative structural break analysis. Yogo (2008) uses wavelet filter for business cycle component of US real GDP and concludes that wavelet filter is better than band-pass filter of Baxter and King (1999). Bowden and Zhu (2008) combines wavelet analysis with structural breaks and apply this combined model to agribusiness cycle.

In conclusion, it can be seen that structural model has good performance in the default prediction, but it has a number of parameters to estimate which directly influence the accuracy of the models. Wavelet analysis technology has been widely applied in the engineering field. Through joining time-frequency analysis, wavelet can make any signal decomposed into high frequency part and low frequency part which are independent, and according to the different sampling density, one can adjust the time window, so it can be used to check the characteristics of any signal under different magnification to effectively filter the noise. And it has be tested that wavelet has a good performance in the stock market. This article is based on these reasons, combines the structural model with wavelet analysis to better estimate the parameters, and then uses the actual data of listed companies to discuss the accuracy of structural model in China.

3 Preliminaries

Structure model requires the mark-to-market value for listed companies' credit assessment, which described the process of default as the explicit result of the deterioration with the companies' value. Then it can be simplified that the company owners' equity can be seen as call option, and the liability as put option. Once one made sure of the company valuation model and the company's capital structure, one can use the option pricing formula to price the equity and debt in order to predict default.

3.1 Assumption

Structural equation model usually needs to meet the following assumptions:

1. There are only two ways-equity (with the value S) and debt (with the principal D and maturity T) for the company (with the value V) to finance.
2. At any $t \ll T$, the value of a company is equal to the sum of debt value and equity value, which can be described as $V_t = S_t + D_t$.
3. The value of the company follows geometric Brownian motion, that is $dV = uVdt + \sigma_u VdZ$.
4. Before maturity, bond holders cannot force companies to bankruptcy. But at the maturity T, if the value of the company can cover the debt principal, it means that company has the payment ability, otherwise the value of the company is not enough to pay back the principal, which is $V < D$ and it occurs default.
5. When default occurs, bond holders have more priority than shareholders. So they can get the full value V of the company, otherwise, bond holders get their principals D.

According to these assumptions above, the share holders' profit and loss can be thought as a call option of the company's value which the strike price is D, and the bond holders' profit and losses can be thought as a put option that is the risk-free bond D, minus the company's value. Then based on this, one can predict default through pricing the value of equity and debt.

3.2 Models

According to the Black and Scholes option price, we can get the following relationship:

$$S = VN(d_1) - De^{-rT}N(d_2) \tag{1}$$

$d_1 = \frac{\ln(V/D)+(r+\sigma_V^2/2)\times T}{\sigma_V\sqrt{T}}$, $d_2 = d_1 - \sigma_V\sqrt{T}$, $N^{(*)}$ is the cumulative probability distribution function of standard normal distribution.

According to the sensitivity analysis $\sigma_s = \frac{V}{S}(\frac{\partial V}{\partial S})\sigma_V$, there is $\frac{\partial V}{\partial S} = N(d_1)$,then we can get that:

$$\sigma_s = \frac{V}{S}N(d_1)\sigma_V \tag{2}$$

Because the value V of the company and its volatility cannot be evaluated easily, but we can calculate the value and the volatility σ_s of equity from capital market, then we can get V and σ_v through simultaneous Eqs. (1) and (2) the two equations. Therefore, it can be seen that S and σ_s are the key indicator for the model's accuracy. Furthermore, with the effect financial market, S can be obtained directly, so the calculation of σ_s becomes the core for structure model.

3.3 Kernel Parameters

In recent research, there always uses Garch(1,1) to estimate σ_s, the details can be found in the references (Yogo 2008). We gave the model below:

$$\begin{aligned} r_t &= \sqrt{h_t e_t} \\ h_t &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1} \\ e_t &\sim iidN(0,1) \end{aligned} \tag{3}$$

From the model, we can see that it needs repeat the prediction of future volatility, so it tends to expand the prediction calculation errors. Meanwhile, using this model with the passage of time, the volatility tends to be weaken and finally approaches zero, which can be called convergence.

Wavelet analysis is similar to Fourier analysis, the basic principle is to use a cluster of basis functions to represent or approach any signal, this cluster of basis function is called wavelet function system, which is the group of base stretching or shifting through wavelet function and its transform coefficients can be described as the characteristics of original signal.

Definition 1 Let $\psi \in L^2(R) \cap L^1(R)$ and $\hat{\psi}(0) = 0$, then the function $\{\psi_{a,b}\}$ which can be got through the way below is called Wavelet function.

$$\psi_{a,b}(x) = |a|^{-\frac{1}{2}}\,\psi\left(\frac{x-b}{a}\right) \quad b \in R, \quad a \in R^+ \tag{4}$$

And $\psi(x)$ is called basis wavelet or mother wavelet,a is called stretch factor,b is called shift factor. $\hat{\psi}(w)$ is the fourier transform[1] of $\psi(x)$.

Definition 2 Let $\psi(x)$ be the basis wavelet, $\{\psi_{a,b}\}$ is continuous wavelet got by (4), so for any function $f \in L^2(R)$, continuous wavelet transform $W_f(a,b)$ is defined as below:

$$W_f(a,b) = \langle f, \psi_{a,b} \rangle = |a|^{-\frac{1}{2}} \int_R f(x) \overline{\psi\left(\frac{x-b}{a}\right)} dx \tag{5}$$

Definition 3 Let $\psi \in L^2 \cap L^1$ and satisfying $C_\psi = \int \frac{\left|\hat{\psi}(w)\right|^2}{|w|} dw < +\infty$, $\hat{\psi}(w)$ is the fourier transform of $\psi(x)$, then $\psi(x)$ is called allowed wavelet. Allowed wavelet functions means that the wavelet has sufficient rate to decay, and with the mean of zero.

Theorem 1 *Let $\psi(x)$ be the allowed wavelet, and for any $f, h \in L^2(R, dx)$, there is $\iint_{R^2} W_f(a,b)\overline{W_h(a,b)}\frac{da}{a^2}db = C_\psi \langle f, h \rangle$. For any $f \in L^2(R)$, then $f(x)$ can be reconstructed by the following way: $f(x) = \frac{1}{C_\psi}\int_{-\infty}^{+\infty}\int_0^\infty W_f(a,b)\psi_{(a,b)}\frac{da}{a^2}db$.*

From these definitions and theorem above, we can find that any $f(x)$ can be reconstructed by its wavelet transform $W_f(a,b)$, wavelet transform can be seen as the process of $f(x)$ decomposition on its basis of wavelet, so there are multiple transforms for a. In the practice, we need only orthogonal wavelet basis to keep the bases off correlation which can be conducted by sampling discretely.

Definition 4 If stretch factor a and shift factor b are conducted by these rules: $a = a_0^{-m} \quad a_0 > 1, b = nb_0 a_0^{-m} \quad b_0 \in R \ m, n \in Z$ then the basis wavelet can be described as: $\psi_{m,n}(x) = a_0^{m/2}\psi(a_0^m x - nb_0)$, so discrete wavelet transform can be written as $DW_f = \int_{-\infty}^{+\infty} f(x)\overline{\psi_{m,n}(x)}dx = \langle f(x), \psi_{m,n}(x) \rangle$.In the process of discretization, if the stretch factor a is conducted by binary arithmetic operation $a_j = 2^j$, then the wavelet is called binary wavelet.

Theorem 2 *Let $\varphi(x)$ is the scale function of multiresolution analyses$\{V_j\}_{j\in Z}$ in $L^2(R)$, and satisfies: 1)$\{\varphi(x-n)\}_{n\in Z}$ is orthonormal basis of V_0; 2)$\varphi(x) = \sqrt{2}\sum_k p_k\varphi(2x-k)$, $\{p_k\}_{k\in Z} \in l^2$, let $\psi(x) = \sqrt{2}\sum_k p_{1-k}\varphi(2x-k) = \sqrt{2}\sum_k q_k\varphi(2x-k)$, then 1) $L^2(R) = \overline{\oplus_{j\in Z} W_j}$; 2)$W_j \perp W_{j'}$, $j \neq j'$; 3)$\{\psi_{j,k}(x)\}_{k\in Z}$ is the orthonormal basis of W_j, and also the orthonormal basis of $L^2(R)$.*

Therefore, any $f(x)$ has two way of wavelet transform: $f(x) = \sum_{m,n}\langle f, \psi_{m,n}\rangle \psi_{m,n}(x)$ and $f(x) = \sum_n \langle f, \varphi_{m_0,n}\rangle\varphi_{m_0,n}(x) + \sum_{m>m_0,n}\langle f, \psi_{m,n}\rangle\psi_{m,n}(x)$.On the one hand, let $\psi(x)$ be wavelet function and $\varphi(x)$ be scale function, then for any $N \in Z$, we can use f_N to approach V_N, while for all $j \in Z$, there is $V_j = V_{j-1} \oplus W_{j-1}$.

[1] Fourier transform: Let $f \in L(R)$,then fourier transform can be defined as $\hat{f}(w) = \left\langle f(x), e^{iwx}\right\rangle = \int_R f(x)\overline{e^{iwx}}dx = \int_R f(x)e^{-iwx}dx$.

So $f_N = f_{N-1} + g_{N-1}$, $f_{N-1} \in V_{N-1}$, $g_{N-1} \in W_{N-1}$, repeat this process, then $f_{N-1} = g_{N-1} + g_{N-2} + \ldots + g_{N-M} + f_{N-M}$. On the other hand, both $\psi(x) \in W_0$ and $\varphi(x) \in V_0$ belong to V_1 which is constructed by $\varphi_{1,k}(x) = 2^{1/2}\varphi(2x-k),\ k \in Z$, so there exists $\{p_k\}$ and $\{q_k\}$, which can make two-scale relations sense. Then we know that both $\varphi(2x)$ and $\varphi(2x-1)$ belong to V_1, $V_1 = V_0 \oplus W_0$, so it can get that $\varphi(2x-l) = \sum_k [a_{l-2k}\varphi(x-k) + b_{l-2k}\psi(x-k)]$, which is called the decomposition of $\varphi(x)$ and $\psi(x)$, in which $\{p_k\}$ and $\{q_k\}$ are called reconstruction sequence, $\{a_k\}$ and $\{b_k\}$ are called decomposition sequence. Finally $f_j \in V_j$ and $g_j \in W_j$ can be written by the only sequence below:

$$\begin{cases} f_j(x) = \sum_k c_{j,k}\varphi\left(2^j x - k\right) \\ c^j = \left\{c_{j,k}\right\} \in l^2 \end{cases} \tag{6}$$

$$\begin{cases} g_j(x) = \sum_k d_{j,k}\psi(2^j x - k) \\ d^j = \left\{d_{j,k}\right\} \in l^2 \end{cases}$$

The process of decomposition is

$$\begin{cases} c_{j-1,k} = \sum_l a_{l-2k}c_{j,l} \\ d_{j-1,k} = \sum_l b_{l-2k}c_{j,l} \end{cases}$$

The process of reconstruction is

$$c_{j,k} = \sum_l \left[p_{k-2l}c_{j-1,l} + q_{k-2l}d_{j-1,l}\right]$$

This is Mallat Algorithm. Both c^{j-1} and d^{j-1} are got by moving average with c^j using the weight of $\{a_k\}$ and $\{b_k\}$, while this process only takes the samples on even integer points, so it is also called down sample; meanwhile the process of reconstruction takes up sample which does convolution with c^{j-1} and d^{j-1} only on even integer points to get $\{p_k\}$ and $\{q_k\}$.

Based on Mallat Algorithm, we can do wavelet decomposition with $r(x)$, and then construct appropriate predictive models for $\{a_k\}$ and $\{b_k\}$ separately, finally using the predictive models to reconstruct $\{p_k\}$ and $\{q_k\}$ in order to get $r(x)'$, which can ensure that when $l \to \infty$, $r(x)' \to r(x)$.

4 Empirical Analysise

4.1 Data and Parameters

In this paper we use the CSMAR database (http://www.gtarsc.com/) as data source, and select listed companies in China which appeared on the market before 2009 and have not been off in 2009. There are 1697 listed companies (including 184 ST companies), and they belong to 13 industry groups.

This article mainly aimed at the default prediction of listed companies, Because financial listed companies have more liquidity with the assets and liabilities than the ordinary corporate entities, whose default risk may cause shocks in the financial system, there are many regulatory factors to restrict, so its default cannot be predicted only through the market information. Base on these, we don't take these financial listed companies in our experiment, then we sampled the rest of listed companies randomly according to the industry category, finally we selected 100 sample enterprises (including 50 ST enterprises, 50 enterprises not ST), enterprise code and name can be seen in Table 1.

Using the following way to estimate several key parameters in the model:

First, period. Structure model considers only default prediction problem which will be due in the next year. This section uses the listed companies' market data in 2009 to predict the default probability in 2010, and in Sect. 4 we use the market data in 2010 to validate the model's results.

Second, risk-free rate. In this section we use the 1-year deposit rates 2.25 % in 2009 as the risk-free rate.

Then, equity value. There are two kinds of shocks of a listed company in China for a long time-tradable shares and non-tradable shares. There mainly are two ways to calculate the equity value: (1) equity value = tradable shares × market price + non-tradable shares × conversion ratio; (2) equity value = tradable shares × market price + non-tradable shares × net assets per share. This paper adopts the second calculation method. The results are shown in Table 1.

Finally, default point. This paper follows the KMV model to determine the default point. Long Term debt is the debt with maturity more than 1 year, and written as LT for short; Short Term is the debt with maturity within a year, and written as ST for short. Then the default point can be determined in accordance with the following standards. The results are shown in Table 1.

$$\begin{aligned} DP &= ST + 0.5 \times LT & LT/ST < 1.5 \\ DP &= ST + (0.7 - 0.3 \times ST/LT) \times LT & LT/ST \geq 1.5 \end{aligned}$$

4.2 GARCH Structural Model

According to the closed price of a stock, we can get the yield sequences by $r_t = \log(\frac{p_t}{p_{t-1}})$ in 2009, with p_t on behalf of the day's closing price and p_{t-1} on behalf of the precious's closing price. Here take the stock with code 000713 as an example; its daily closing yield sequence is shown in Fig. 1 below:

Take autocorrelation test with r_t and r_t^2;the results are shown in Tables 2 and 3. From Table 2 we can see that there is no autocorrelation of r_t, and from Table 3 we can see there exists auto-correlation of r_t^2.

Give the LM test with r_t to test whether there is ARCH effect. The result is shown in the following Table 4.

From Table 4, we can find that there exists high-order ARCH effect, so it can be fit with GARCH (1,1) model. The result of GARCH (1,1) is showing in the following Table 5.

Table 1 Table of results

Code	Type	ST or not	Equity value (S)	Short liability (ST)	Long liability (LT)	Default point (DP)
000010	H	1	941187905	141,377,962	6,697,347	144,726,635
000017	C	1	1875924229	1,851,622,497	179,088,443	1,941,166,719
000022	F	0	10765687288	1,618,772,245	488,509,794	1,863,027,142
000036	C	1	8,114,644,398	875,650,683	499,512,129	1,125,406,747
000038	G	1	635,535,911	203,633,348	118,370,842	262,818,769
000050	C	0	6,771,382,001	1,673,296,529	2,704,090,935	3,064,171,224
000156	C	1	571,808,248	785,884,820	49,860,305	810,814,973
000158	C	0	8,582,021,366	1,799,837,841	145,783,158	1,872,729,420
000403	C	1	1,415,518,635	909,862,612	204,292,597	1,012,008,910
000408	C	1	2,558,022,303	379,895,953	0	379,895,953
000409	C	1	1,440,223,501	10,713,012	90,000	10,758,012
000430	K	1	959,586,065	328,428,099	16,623,893	336,740,045
000506	J	1	4,250,128,637	2,527,895,596	285,019,534	2,670,405,363
000517	J	1	5,683,319,392	1,821,762,969	1,534,000,000	2,588,762,969
000518	C	1	3,846,839,542	40,317,436	0	40,317,436
000539	D	0	35,096,320,198	8,748,800,713	9,359,931,772	13,428,766,599
000576	C	1	2,133,421,186	218,889,926	155,183,205	296,481,528
000585	C	1	2,275,547,278	297,675,111	217,887,964	406,619,093
000587	C	1	1,376,951,898	569,745,598	8,434,296	573,962,746
000591	C	0	2,879,928,467	1,641,019,411	37,898,927	1,659,968,875
000607	C	0	4,664,216,543	1,207,903,258	121,961,711	1,268,884,114
000616	J	0	9,824,423,175	3,142,236,925	1,471,385,742	3,877,929,796
000692	D	1	3,495,286,054	1,404,805,958	326,017,518	1,567,814,717
000713	A	0	4,056,577,617	847,358,914	13,791,296	854,254,562
000722	M	1	2,717,211,696	492,189,651	576,196,563	780,287,932
000805	G	1	21,526,888	153,216,399	10,000,000	158,216,399
000818	C	1	3,079,186,188	2,317,673,425	163,158,194	2,399,252,522
000851	C	0	3,585,806,480	755,209,013	41,138,523	775,778,275
000861	C	0	3,059,825,652	428,608,656	317,588,934	587,403,123
000892	G	1	1,243,983,694	11,737,693	0	11,737,693
000925	C	0	3,597,472,301	1,476,693,432	186,700,000	1,570,043,432
000935	C	1	2,171,901,646	361,235,040	124,717,303	423,593,691
000971	C	1	1,282,423,329	523,245,560	41,894,020	544,192,570
000998	A	0	5,626,133,737	1,000,857,427	43,468,657	1,022,591,756
002002	C	1	1,572,738,799	200,959,115	40,000	200,979,115
002013	C	0	1,630,707,627	240,089,505	30,000,000	255,089,505
002026	C	0	1,456,552,376	152,788,754	0	152,788,754
002123	C	0	6,020,492,468	539,407,595	177,648,983	628,232,087
002124	C	0	1,450,798,766	241,049,175	88,910,965	285,504,658

Table 1 continued

Code	Type	ST or not	Equity value (S)	Short liability (ST)	Long liability (LT)	Default point (DP)
002132	C	0	2,694,944,215	765,331,715	139,909,317	835,286,373
002162	C	0	2,658,764,683	712,260,117	95,991,240	760,255,737
002181	L	0	2,792,442,434	184,704,122	76,118	184,742,181
002198	C	0	640,452,514	11,905,495	0	11,905,495
002207	B	0	1,166,227,785	326,068,303	0	326,068,303
002236	C	0	2,525,574,782	331,880,848	0	331,880,848
002305	J	0	2,694,439,507	737,002,488	199,236,000	836,620,488
300008	K	0	910,565,488	43,116,514	873,165	43,553,096
300022	H	0	1,870,635,403	414,157,643	0	414,157,643
600019	C	0	267,142,248,023	70,721,946,401	29,201,535,339	85,322,714,071
600020	F	0	23,519,155,807	4,630,859,459	11,501,551,070	11,292,687,370
600057	C	1	1,041,051,267	969,741,513	12,709,230	976,096,128
600069	C	0	6,512,061,247	1,795,673,649	586,376,911	2,088,862,104
600070	C	0	1,811,658,995	633,386,753	34,314,124	650,543,815
600079	M	0	6,227,297,309	1,185,307,900	41,911,061	1,206,263,431
600084	H	1	5,972,993,002	1,697,793,844	570,250,467	1,982,919,078
600102	C	0	17,323,679,445	6,271,704,958	3,969,587,968	8,256,498,942
600115	F	1	69,579,245,230	35,663,041,000	32,742,512,000	52,034,297,000
600143	C	0	12,974,420,577	2,856,774,018	1,805,321,202	3,759,434,619
600178	C	0	9,353,218,558	1,536,738,139	287,447,947	1,680,462,113
600187	C	1	1,013,652,481	146,654,771	88,190,350	190,749,946
600234	H	1	1,149,715,681	417,187,172	89,189,754	461,782,049
600242	A	1	753,047,428	114,573,471	32,000,000	130,573,471
600252	M	0	6,664,955,165	531,652,596	395,976,979	729,641,086
600253	C	1	5,349,717,881	2,165,150,052	150,368,938	2,240,334,521
600259	C	1	2,129,694,958	793,926,538	38,567,565	813,210,321
600271	G	0	19,809,985,011	1,410,378,100	49,110,486	1,434,933,343
600285	C	0	2,721,797,210	347,373,705	31,142,281	362,944,845
600292	D	0	7,378,347,208	2,423,063,107	1,755,206,239	3,300,666,227
600297	C	0	3,769,792,269	804,166,352	353,550,662	980,941,683
600313	A	1	1,566,266,844	132,910,746	21,000,000	143,410,746
600338	C	1	2,929,591,138	497,309,134	93,323,092	543,970,680
600353	C	0	1,633,304,435	208,222,141	4,858,510	210,651,396
600355	C	0	1,696,178,324	142,382,619	1,101,682	142,933,460
600516	C	0	9,990,602,711	2,475,521,896	130,948,899	2,540,996,345
600556	C	1	1,003,637,388	135,551,747	650,000	135,876,747
600580	C	0	6,356,757,114	1,065,346,817	170,000,000	1,150,346,817
600596	C	0	12,541,221,631	876,980,762	655,882,641	1,204,922,082
600608	G	1	2,210,291,331	623,671,685	500,000	623,921,685

Table 1 continued

Code	Type	ST or not	Equity value (S)	Short liability (ST)	Long liability (LT)	Default point (DP)
600617	C	1	712,750,857	109,909,734	5,194,298	112,506,883
600658	G	0	3,954,300,262	1,216,224,196	118,296,171	1,275,372,281
600680	G	0	2,661,796,126	718,277,601	4,294,219	720,424,711
600681	L	1	418,958,041	522,593,393	34,538,306	539,862,546
600691	C	1	899,867,753	184,414,345	33,703,470	201,266,081
600705	M	1	857,118,931	11,039,338	0	11,039,338
600716	C	1	4,675,339,349	1,818,424,753	0	1,818,424,753
600722	C	1	3,774,857,766	1,579,024,778	1,517,505,959	2,337,777,757
600727	C	1	2,524,027,574	477,907,907	0	477,907,907
600728	G	1	1,826,249,475	432,779,329	59,544,851	462,551,754
600751	F	1	2,017,771,922	969,151,804	0	969,151,804
600800	C	1	2,534,486,975	746,018,201	269,992,515	881,014,458
600804	C	0	7,671,095,512	635,237,177	59,625,776	665,050,065
600859	H	0	17,984,360,612	3,631,019,495	838,012,044	4,050,025,517
600868	M	1	9,679,245,089	1,029,196,115	1,785,201,230	1,970,078,141
600887	C	1	27,367,009,684	9,037,426,350	404,778,435	9,239,815,568
600890	C	1	2,835,939,255	147,692,163	32,983,956	164,184,140
600970	E	0	27,058,309,181	14,340,354,262	374,232,615	14,527,470,569
600984	C	1	1,667,866,286	489,822,145	0	489,822,145
600988	C	1	434,551,405	131,028,807	0	131,028,807
600992	C	0	2,206,424,735	475,730,252	0	475,730,252
601666	B	0	51,147,288,004	6,417,054,063	1,413,269,833	7,123,688,980

Code	Type	ST or not	Time series model			
			Volatility of equity	Assets	Volatility of assets	Default distance
000010	H	1	0.13	9,552,600,000	0.47	2.10
000017	C	1	0.22	1,985,900,000	0.48	0.05
000022	F	0	0.09	13,587,000,000	0.32	2.70
000036	C	1	0.15	9,215,000,000	0.44	2.00
000038	G	1	0.29	637,540,000	0.53	1.11
000050	C	0	0.14	10,967,000,000	0.34	2.12
000156	C	1	0.30	810,820,000	0.54	0.00
000158	C	0	0.15	18,413,000,000	0.37	2.43
000403	C	1	0.27	1,426,300,000	0.45	0.65
000408	C	1	0.11	3,878,800,000	0.52	1.73
000409	C	1	0.15	7,631,500,000	0.42	2.38
000430	K	1	0.15	9,052,600,000	0.52	1.85
000506	J	1	0.15	6,861,100,000	0.32	1.91
000517	J	1	0.14	8,214,500,000	0.33	2.08

Table 1 continued

Code	Type	ST or not	Time series model			
			Volatility of equity	Assets	Volatility of assets	Default distance
000518	C	1	0.09	3,886,300,000	0.43	2.30
000539	D	0	0.11	53,596,000,000	0.36	2.08
000576	C	1	0.14	3,978,800,000	0.55	1.68
000585	C	1	0.09	4,878,800,000	0.42	2.18
000587	C	1	0.15	7,631,500,000	0.48	1.93
000591	C	0	0.17	5,463,600,000	0.32	2.18
000607	C	0	0.15	6,724,900,000	0.32	2.54
000616	J	0	0.18	24,616,000,000	0.33	2.55
000692	D	1	0.13	5,028,200,000	0.41	1.68
000713	A	0	0.12	5,726,200,000	0.37	2.30
000722	M	1	0.21	3,480,100,000	0.56	1.39
000805	G	1	0.28	158,220,000	0.52	0.00
000818	C	1	0.14	5,425,100,000	0.52	1.07
000851	C	0	0.13	5,924,300,000	0.38	2.29
000861	C	0	0.12	4,854,200,000	0.41	2.14
000892	G	1	0.09	9,621,000,000	0.52	1.92
000925	C	0	0.08	9,823,600,000	0.35	2.40
000935	C	1	0.14	4,134,900,000	0.46	1.95
000971	C	1	0.16	7,631,500,000	0.48	1.93
000998	A	0	0.14	9,852,000,000	0.38	2.36
002002	C	1	0.19	1,769,200,000	0.49	1.81
002013	C	0	0.12	2,653,100,000	0.36	2.51
002026	C	0	0.14	9,631,500,000	0.39	2.52
002123	C	0	0.04	7,934,700,000	0.41	2.25
002124	C	0	0.15	9,631,500,000	0.39	2.49
002132	C	0	0.14	5,634,600,000	0.38	2.24
002162	C	0	0.14	6,453,100,000	0.35	2.52
002181	L	0	0.10	5,324,100,000	0.33	2.93
002198	C	0	0.14	11,635,000,000	0.32	3.12
002207	B	0	0.14	9,214,500,000	0.34	2.84
002236	C	0	0.09	9,456,800,000	0.37	2.61
002305	J	0	0.16	8,934,400,000	0.36	2.52
300008	K	0	0.22	12,371,000,000	0.29	3.44
300022	H	0	0.25	8,649,800,000	0.37	2.57
600019	C	0	0.14	775,140,000,000	0.43	2.07
600020	F	0	0.11	83,569,000,000	0.27	3.20
600057	C	1	0.23	3,675,800,000	0.52	1.41
600069	C	0	0.14	11,294,000,000	0.34	2.40

Table 1 continued

Code	Type	ST or not	Time series model			
			Volatility of equity	Assets	Volatility of assets	Default distance
600070	C	0	0.14	9,351,800,000	0.27	3.45
600079	M	0	0.17	11,247,000,000	0.29	3.08
600084	H	1	0.15	7,911,800,000	0.54	1.38
600102	C	0	0.17	25,434,000,000	0.30	2.25
600115	F	1	0.11	100,570,000,000	0.34	1.42
600143	C	0	0.11	62,430,000,000	0.33	2.85
600178	C	0	0.16	47,295,000,000	0.46	2.10
600187	C	1	0.10	5,367,800,000	0.41	2.35
600234	H	1	0.18	7,631,500,000	0.56	1.68
600242	A	1	0.21	9,052,600,000	0.66	1.49
600252	M	0	0.17	9,978,400,000	0.37	2.51
600253	C	1	0.13	7,540,200,000	0.46	1.53
600259	C	1	0.16	5,026,300,000	0.36	2.33
600271	G	0	0.14	67,653,000,000	0.44	2.22
600285	C	0	0.14	9,264,700,000	0.42	2.29
600292	D	0	0.17	43,426,000,000	0.28	3.30
600297	C	0	0.09	8,787,900,000	0.36	2.47
600313	A	1	0.14	7,631,500,000	0.63	1.56
600338	C	1	0.17	3,461,500,000	0.64	1.32
600353	C	0	0.16	3,596,300,000	0.36	2.62
600355	C	0	0.13	2,046,800,000	0.45	2.07
600516	C	0	0.08	35,625,000,000	0.44	2.11
600556	C	1	0.25	3,378,800,000	0.67	1.43
600580	C	0	0.12	9,892,500,000	0.41	2.16
600596	C	0	0.11	23,212,000,000	0.29	3.27
600608	G	1	0.31	2,820,300,000	0.72	1.08
600617	C	1	0.28	9,052,600,000	0.68	1.45
600658	G	0	0.17	5,201,300,000	0.24	3.14
600680	G	0	0.12	3,366,200,000	0.27	2.91
600681	L	1	0.30	539,860,000	0.64	0.00
600691	C	1	0.20	9,052,600,000	0.62	1.58
600705	M	1	0.28	859,710,000	0.76	1.30
600716	C	1	0.17	6,453,300,000	0.45	1.60
600722	C	1	0.16	6,060,600,000	0.67	0.92
600727	C	1	0.09	4,138,800,000	0.72	1.23
600728	G	1	0.28	3,978,800,000	0.72	1.23
600751	F	1	0.18	3,645,800,000	0.73	1.01
600800	C	1	0.13	4,118,800,000	0.69	1.14

Table 1 continued

Code	Type	ST or not	Time series model			
			Volatility of equity	Assets	Volatility of assets	Default distance
600804	C	0	0.04	8,321,300,000	0.26	3.54
600859	H	0	0.13	21,944,000,000	0.33	2.47
600868	M	1	0.12	11,605,000,000	0.65	1.28
600887	C	1	0.14	27,453,000,000	0.39	1.70
600890	C	1	0.18	2,996,500,000	0.58	1.63
600970	E	0	0.04	49,762,000,000	0.30	2.36
600984	C	1	0.16	2,146,800,000	0.59	1.31
600988	C	1	0.19	9,939,400,000	0.63	1.57
600992	C	0	0.12	9,609,800,000	0.46	2.07
601666	B	0	0.13	52,465,000,000	0.37	2.34
Code	Type	ST or not	Wavelet structural model			
			Volatility of equity	Assets	Volatility of assets	Default distance
000010	H	1	0.13	9,052,600,000	0.83	1.19
000017	C	1	0.22	1,985,900,000	0.77	0.03
000022	F	0	0.12	13,597,000,000	0.31	2.78
000036	C	1	0.15	9,215,000,000	0.31	2.83
000038	G	1	0.29	637,540,000	0.78	0.75
000050	C	0	0.15	13,967,000,000	0.33	2.37
000156	C	1	0.30	810,820,000	0.82	0.00
000158	C	0	0.14	18,573,000,000	0.34	2.64
000403	C	1	0.28	1,426,300,000	0.73	0.40
000408	C	1	0.12	2,929,500,000	0.71	1.23
000409	C	1	0.16	7,631,500,000	0.77	1.30
000430	K	1	0.17	9,052,600,000	0.82	1.17
000506	J	1	0.11	7,861,100,000	0.34	1.94
000517	J	1	0.24	8,214,500,000	0.32	2.14
000518	C	1	0.10	3,886,300,000	0.34	2.91
000539	D	0	0.12	53,624,000,000	0.31	2.42
000576	C	1	0.14	3,878,800,000	0.82	1.13
000585	C	1	0.11	4,973,800,000	0.39	2.35
000587	C	1	0.15	7,631,500,000	0.84	1.10
000591	C	0	0.17	5,464,500,000	0.31	2.25
000607	C	0	0.17	6,726,300,000	0.30	2.70
000616	J	0	0.14	24,745,000,000	0.29	2.91
000692	D	1	0.13	5,028,200,000	0.35	1.97
000713	A	0	0.13	5,731,900,000	0.31	2.75

Table 1 continued

Code	Type	ST or not	Wavelet structural model			
			Volatility of equity	Assets	Volatility of assets	Default distance
000722	M	1	0.20	3,480,100,000	0.83	0.93
000805	G	1	0.29	158,220,000	0.81	0.00
000818	C	1	0.15	5,425,100,000	0.79	0.71
000851	C	0	0.14	5,941,700,000	0.33	2.63
000861	C	0	0.13	4,873,200,000	0.39	2.26
000892	G	1	0.10	7,631,500,000	0.81	1.23
000925	C	0	0.36	9,873,900,000	0.38	2.21
000935	C	1	0.14	5,523,100,000	0.42	2.20
000971	C	1	0.18	7,631,500,000	0.82	1.13
000998	A	0	0.13	9,858,400,000	0.37	2.42
002002	C	1	0.21	1,769,200,000	0.83	1.07
002013	C	0	0.16	2,694,200,000	0.38	2.38
002026	C	0	0.13	96,448,000,000	0.37	2.70
002123	C	0	0.13	7,962,700,000	0.38	2.42
002124	C	0	0.16	9,658,600,000	0.36	2.70
002132	C	0	0.14	5,687,300,000	0.35	2.44
002162	C	0	0.15	6,499,900,000	0.33	2.68
002181	L	0	0.13	5,355,300,000	0.31	3.11
002198	C	0	0.17	11,639,000,000	0.32	3.12
002207	B	0	0.13	9,222,500,000	0.30	3.22
002236	C	0	0.09	9,461,800,000	0.36	2.68
002305	J	0	0.34	8,962,100,000	0.29	3.13
300008	K	0	0.20	13,015,000,000	0.26	3.83
300022	H	0	0.32	8,734,200,000	0.32	2.98
600019	C	0	0.14	793,060,000,000	0.33	2.70
600020	F	0	0.13	88,241,000,000	0.27	3.23
600057	C	1	0.25	3,876,900,000	0.82	0.91
600069	C	0	0.15	14,208,000,000	0.32	2.67
600070	C	0	0.16	9,734,900,000	0.26	3.59
600079	M	0	0.18	13,646,000,000	0.27	3.38
600084	H	1	0.16	7,911,800,000	0.84	0.89
600102	C	0	0.18	30,527,000,000	0.28	2.61
600115	F	1	0.13	104,230,000,000	0.34	1.47
600143	C	0	0.12	77,241,000,000	0.32	2.97
600178	C	0	0.17	67,348,000,000	0.42	2.32
600187	C	1	0.13	4,465,300,000	0.39	2.45
600234	H	1	0.28	7,631,500,000	0.86	1.09
600242	A	1	0.20	9,052,600,000	0.72	1.37

Table 1 continued

Code	Type	ST or not	Wavelet structural model			
			Volatility of equity	Assets	Volatility of assets	Default distance
600252	M	0	0.17	10,245,000,000	0.36	2.58
600253	C	1	0.13	9,540,200,000	0.44	1.74
600259	C	1	0.19	4,988,200,000	0.37	2.26
600271	G	0	0.12	68,320,000,000	0.39	2.51
600285	C	0	0.16	9,652,600,000	0.40	2.41
600292	D	0	0.18	47,337,000,000	0.28	3.32
600297	C	0	0.05	8,588,900,000	0.31	2.86
600313	A	1	0.15	1,706,500,000	0.77	1.19
600338	C	1	0.33	3,461,500,000	0.82	1.03
600353	C	0	0.16	3,423,700,000	0.35	2.68
600355	C	0	0.19	2,356,800,000	0.48	1.96
600516	C	0	0.21	37,452,000,000	0.43	2.17
600556	C	1	0.27	35,428,800,000	0.86	1.16
600580	C	0	0.10	9,899,900,000	0.39	2.27
600596	C	0	0.10	20,912,000,000	0.28	3.37
600608	G	1	0.27	4,078,800,000	0.88	0.96
600617	C	1	0.28	9,052,600,000	0.82	1.20
600658	G	0	0.17	5,181,300,000	0.22	3.43
600680	G	0	0.11	3,836,200,000	0.25	3.25
600681	L	1	0.30	539,860,000	0.67	0.00
600691	C	1	0.20	9,052,600,000	0.78	1.25
600705	M	1	0.28	859,710,000	0.82	1.20
600716	C	1	0.16	6,453,300,000	0.44	1.63
600722	C	1	0.17	6,060,600,000	0.78	0.79
600727	C	1	0.10	4,178,800,000	0.82	1.08
600728	G	1	0.27	3,878,800,000	0.78	1.13
600751	F	1	0.18	3,976,200,000	0.81	0.93
600800	C	1	0.13	4,078,800,000	0.76	1.03
600804	C	0	0.02	9,651,300,000	0.26	3.58
600859	H	0	0.15	21,854,000,000	0.31	2.63
600868	M	1	0.11	11,605,000,000	0.75	1.11
600887	C	1	0.15	28,677,000,000	0.38	1.78
600890	C	1	0.19	2,996,500,000	0.73	1.29
600970	E	0	0.20	52,062,000,000	0.29	2.49
600984	C	1	0.17	2,446,800,000	0.76	1.05
600988	C	1	0.20	9,939,400,000	0.77	1.28
600992	C	0	0.11	10,549,000,000	0.41	2.33
601666	B	0	0.11	61,335,000,000	0.36	2.46

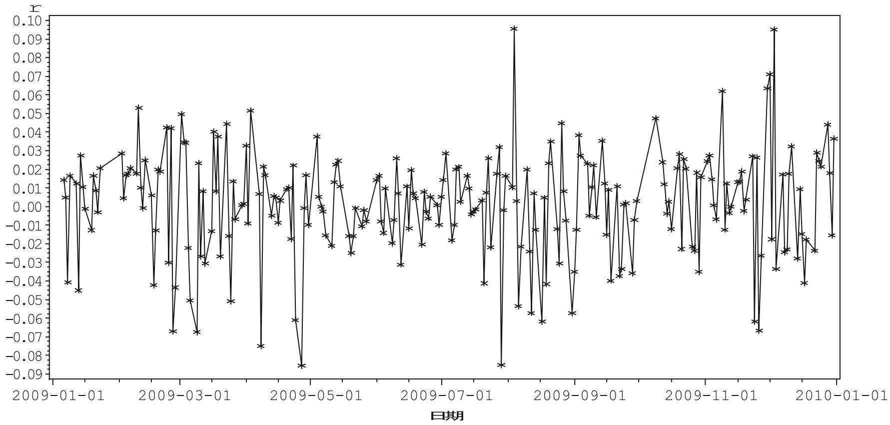

Fig. 1 Return series figure

Table 2 Autocorrelation of r_t

Name of Variable=r

Mean of Working Series 0.001461
Standard Deviation 0.02847
Number of Observations 241

Autocorrelations

Lag	Covariacnce	Correlation	−1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	Std Error
0	0.0008106	1.00000												*	*	*	*	*	*	*	*	*		0.00000
1	(0.0000051)	−0.00634																						0.06441
2	(0.0000480)	−0.05918										*												0.06441
3	0.0000240	0.02960												*										0.06464
4	(0.0001320)	−0.16281								*	*	*												0.06470
5	(0.0000823)	−0.10155									*	*												0.06637
6	0.0000981	0.12107												*	*									0.06719
7	(0.0000463)	−0.05711										*												0.06792
8	0.0000080	0.00992																						0.06812
9	0.0000247	0.03047												*										0.06813
10	(0.0000286)	−0.03533										*												0.06818

* Is the items for delay of data, eg. the right of lag 0 has ten delays which will affect the lag 1
| is on behalf of the threshold of delay, from which we can select the variables out off

Table 3 Autocorrelation of r_t^2

Name of Variable=r2

Mean of Working Series 0.000813
Standard Deviation 0.01394
Number of Observations 241

Autocorrelations

Lag	Covariacnce	Correlation	−1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	Std Error
0	0.0000019	1.00000												*	*	*	*	*	*	*	*	*		0.00000
1	0.0000001	0.04228												*										0.06442
2	0.0000003	0.17692												*	*	*	*							0.06453
3	0.0000002	0.09358												*	*									0.06651
4	0.0000003	0.16606												*	*	*								0.06706
5	0.0000008	0.42210												*										0.06874
6	0.0000002	0.08719												*	*									0.06885
7	0.0000001	0.05674												*										0.06931
8	(0.0000002)	−0.09734									*	*												0.06950
9	0.0000001	0.02649												*										0.07006
10	0.0000000	0.02469																						0.07010

* Is the items for delay of data, eg. the right of lag 0 has ten delays which will affect the lag 1
| is on behalf of the threshold of delay, from which we can select the variables out off

Table 4 LM Tests of r_t

Q and LM tests for ARCH disturbances				
Order	Q	P > Q	LM	P > LM
1	0.5454	0.4602	0.5634	0.4529
2	7.4276	0.0244	7.3402	0.0255
3	9.5662	0.0226	8.8753	0.0310
4	16.0044	0.0030	12.9619	0.0115
5	16.3866	0.0058	12.9691	0.0237
6	18.5835	0.0049	13.3627	0.0376
7	19.5461	0.0066	13.5569	0.0596
8	22.0221	0.0049	18.8789	0.0155
9	22.2924	0.0080	18.9054	0.0260
10	22.5574	0.0125	19.4069	0.0354
11	22.6075	0.0201	19.4680	0.0532
12	22.6440	0.0309	19.4951	0.0773

From the table above, the GARCH (1, 1) model gets through the test and has small AIC. So it can be said that GARCH (1, 1) has a good fitting effect of the volatility of sequence r_t.

Continue to take LM test with the sequence residual error, it is can be found that the residual error has no ARCH effect. Thereby we can get the yield sequence volatility model as follows: $\sigma_t^2 = 0.0000894 + 0.1670\varepsilon_{t-1}^2 + 0.7313\sigma_{t-1}^2$.

Based on aggregation formula of GARCH model, it can be obtained the predicted earnings volatility model as follows: $\sigma_{t+h,t}^2 = 0.0000894 \times \frac{1-(0.167+0.7313)^h}{1-(0.167+0.7313)} + (0.167 + 0.7313)^h \sigma_t^2$. So the total volatility is equal to $\sigma^2 = \sum_{h=1}^{252} \sigma_{t+h,t}^2$ in future year, which is obtained by cumulating the every day's volatility.

From this process, we can see that the volatility will have huge error with the time pass, so this model can only fit short time prediction. Just as this process, other stocks' volatility can be predicted just as this, which is shown in Table 1.

4.3 Wavelet Structural Model

Wavelet structural model takes the different way to predict the yield sequence, which can avoid accumulation process. Here still give the stock (code 000713) as an example. For r_t, according to the Formula (5), we adopt Mallat algorithm using db3 wavelet to decompose the sequence, which is asymmetric wavelet, and it is suitable for processing large fluctuations of the sequence, then it can get the corresponding low frequency sequence and high frequency sequence. The results are shown in Fig. 2, in which ca3 is on behalf of the three layer decomposition's low frequency coefficient sequences and cd1 to cd3 are on behalf of the corresponding decompositions' high frequency coefficient sequences. With the wavelet decomposition, low frequency coefficient sequence tends to contain the trend of the model, and so it always has useful information, just

Table 5 Statistic of GARCH(1,1)

GRACH estimates					
SSE	0.19538	Obs	241.00000		
MSE	0.00081	Uncond var	0.00088		
Log likelihood	523.90461	Total R-square	–		
SBC	(1025.87000)	AIC	−1039.80920		
Normality test	8.89240	P > Chisquare	0.01170		
Variable	DF	Estimate	SE	T value	$P > \|t\|$
Intercept	1	0.00109	0.00171	0.63000	0.52630
ARCH0	1	0.00009	0.00005	1.64000	0.10110
ARCH1	1	0.16700	0.06400	2.61000	0.00910
GARCH1	1	0.73130	0.10730	6.82000	<0.0001

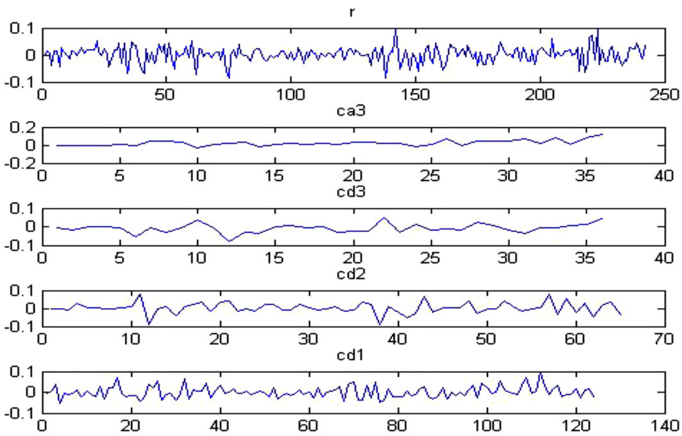

Fig. 2 Wavelet decomposition of r_t

as which can be seen from the Fig. 2 there is a growth trend of ca3; high frequency coefficient sequences are usually noisy disturbance information.

Take stationary test and pure randomness test with ca3, cd3, cd2 and cd1 respectively. Upon examination, in addition to the ca3, the other sequences are all white noise sequences. The results of stationarity and pure randomness test with ca3 can be found in Tables 6, 7 and 8.

From these tables above, it can be found that ca3 autocorrelation coefficient is tailing, and partial autocorrelation coefficient is not tailing which is suitable for ARIMA model. The modeling experiments results are shown in Table 9. According to the smallest Akaike information criterion, we choose to build AR (2) model as a low-frequency coefficients ca3's prediction model, the result is $ca3_t = 1 - 0.27373 \times ca3_{t-2}$.

According to the ca3's prediction model AR(2), we can use the coefficient sequences $\{p_k\}$ and $\{q_k\}$ which is shown as cd3, cd2 and cd1 to reconstruct the yield sequence$r_t$using the Eq. (5). The layers of the reconstructed results are shown in Fig. 3.

Table 6 Autocorrelations of ca3

Autocorrelations

Lag	Covariacnce	Correlation	−1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	Std Error
0	0. 0043140	1. 00000											|	*	*	*	*	*	*	*	*	*	*	0. 00000
1	0. 0014537	0. 33697											|	*	*	*	*							0. 16667
2	0. 0009584	0. 22216											|	*	*	*								0. 18462
3	0. 0015677	0. 36340											|	*	*	*	*							0. 19191
4	0. 0013874	0. 32160											|	*	*	*								0. 21015
5	0. 0011446	0. 26532											|	*	*	*								0. 22341
6	0. 0012717	0. 29479											|	*	*	*	*	*						0. 23200
7	0. 0003921	0. 09089											|	*	*									0. 24218
8	0. 0003704	0. 08585											|	*	*									0. 24312
9	0. 0015119	0. 35047											|	*	*	*	*	*						0. 24396
10	0. 0002883	0. 06683											|	*										0. 25757
11	(0. 0001330)	−0. 03083										*	|											0. 25805
12	0. 0000874	0. 02025											|											0. 25815

* Is the items for delay of data, eg. the right of lag 0 has ten delays which will affect the lag 1
| is on behalf of the threshold of delay, from which we can select the variables out off

Table 7 Partial autocorrelations of ca3

Autocorrelations

Lag	Covariacnce	Correlation	−1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	Std Error
0	0. 0043140	1. 00000											|	*	*	*	*	*	*	*	*	*	*	0. 00000
1	0. 0014537	0. 33697											|	*	*	*	*							0. 16667
2	0. 0009584	0. 22216											|	*	*	*								0. 18462
3	0. 0015677	0. 36340											|	*	*	*	*							0. 19191
4	0. 0013874	0. 32160											|	*	*	*								0. 21015
5	0. 0011446	0. 26532											|	*	*	*								0. 22341
6	0. 0012717	0. 29479											|	*	*	*	*	*						0. 23200
7	0. 0003921	0. 09089											|	*	*									0. 24218
8	0. 0003704	0. 08585											|	*	*									0. 24312
9	0. 0015119	0. 35047											|	*	*	*	*	*						0. 24396
10	0. 0002883	0. 06683											|	*										0. 25757
11	(0. 0001330)	−0. 03083										*	|											0. 25805
12	0. 0000874	0. 02025											|											0. 25815

* Is the items for delay of data, eg. the right of lag 0 has ten delays which will affect the lag 1
| is on behalf of the threshold of delay, from which we can select the variables out off

Reconstructed yield sequence comparing with the original yield sequence is shown in Fig. 4, the reconstructed sequence is marked with circles and the original sequence is marked with asterisk. Through calculating the daily return value error between the two is only 0.026164.

As it can be seen, although in low frequency coefficient prediction is using predictive model, taking off the noisy information the low frequency part also becomes more

Table 8 Check for white noise of ca3

To lag	Chi-square	DF	P > Chi-square	Autocorrelations					
6	23.39	6	0.0007	0.337	0.222	0.363	0.322	0.265	0.295
12	30.67	12	0.0022	0.091	0.086	0.350	0.067	(0.031)	0.020

Table 9 Test ARIMA of ca3

Minimum information criterion

Lags	MA0	MA1	MA2	MA3	MA4	MA5	MA6
AR0	(5.2343)	(5.6505)	(5.5515)	(5.5074)	(5.4859)	(5.4272)	(5.5171)
AR1	(5.6765)	(5.6363)	(5.5654)	(5.4976)	(5.4224)	(5.3544)	(5.4227)
AR2	(5.7430)	(5.5433)	(5.4997)	(5.4150)	(5.3420)	(5.2600)	(5.3259)
AR3	(5.6290)	(5.5561)	(5.4616)	(5.4885)	(5.3941)	(5.3301)	(5.3123)
AR4	(5.5986)	(5.4998)	(5.4013)	(5.3922)	(5.2950)	(5.2336)	(5.2195)
AR5	(5.5863)	(5.4874)	(5.4080)	(5.3771)	(5.3073)	(5.3252)	(5.2881)
AR6	(5.6945)	(5.6469)	(5.6111)	(5.5323)	(5.4571)	(5.3757)	(5.3424)

Error series model: AR(0)

Minimum table value: AIC(2,0) = −5.74299

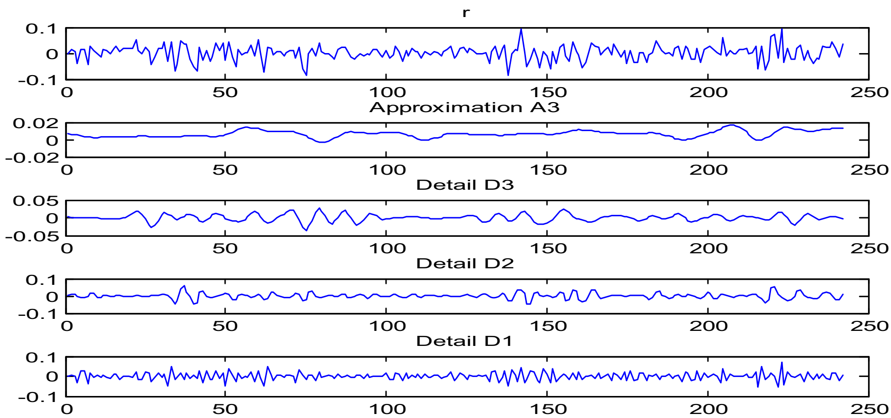

Fig. 3 Wavelet reconstruction of r_t

accurate and efficient, while its high frequency coefficient information is not reduced, so the reconstruction of the sequence can highly approach the original series.

Furthermore, for a stock, it can be considered that its volatility characteristics will not change in short term, that is, its high-frequency disturbance sequence will not be significantly different, so we can use low-frequency sequence predictive model to predict future earnings trends, and then use the data in the future to do successive convolution with high frequency coefficient sequence, to achieve the forecast earnings in the coming year. The prediction results are shown in Fig. 5.

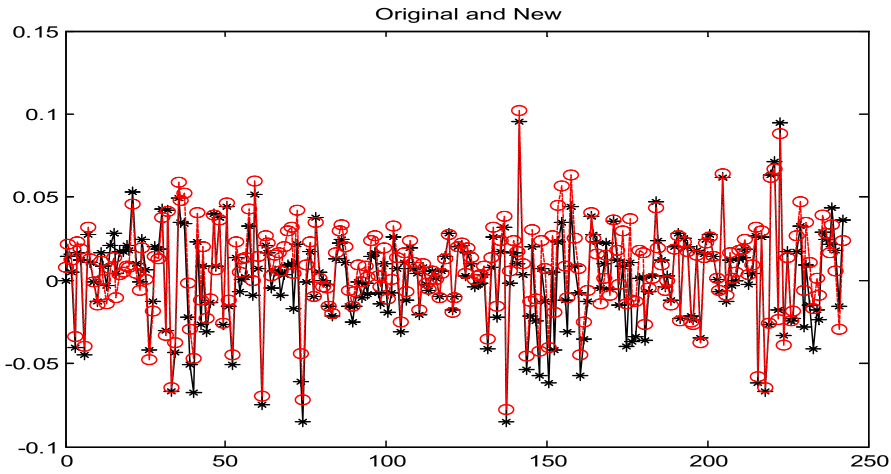

Fig. 4 Original r_t and new r_t

Finally, we can get the standard deviation $\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2}$ of yield forecasts for the coming year, which can be seen as a consistent estimate of volatility.

The volatility of other sample enterprises can also be calculated following similar processes, which are shown in Table 1.

5 Robust Tests

According to the modeling process in the third part, we determine the parameters using Matlab. Based on simultaneous Eqs. (1) and (2), we can solve assets V and its volatility σ_V, the results are also shown in Table 1.

Although calculations of default probability in structural models are very different, now researchers generally agree that the calculation method of KMV is more fit able. KMV model uses the default distance which can be seen in Eq. (6) to judge for the possibility of default, the default distances of selected samples are shown in Table 1.

$$DD = \frac{V - DP}{V \times \sigma_v} \quad (7)$$

On the one side, there are great difficulty with collecting the information of a company's actual default, on the other side, it is a regulation that if there are any financial struggle or abnormal situation with the listed company, which can lead investors difficult to judge its prospects and interests may be impaired, its stock must be special treatment with a "ST" mark in China. Therefore in these robust checks, we use "ST" representing breach of companies. In order to validate the two models, we do the paired T tests. The results are shown in Tables 10 and 11.

As can be seen from the tables above, at the 95 % confidence interval, the two models can identify ST enterprises from good enterprises, so it can be said that both

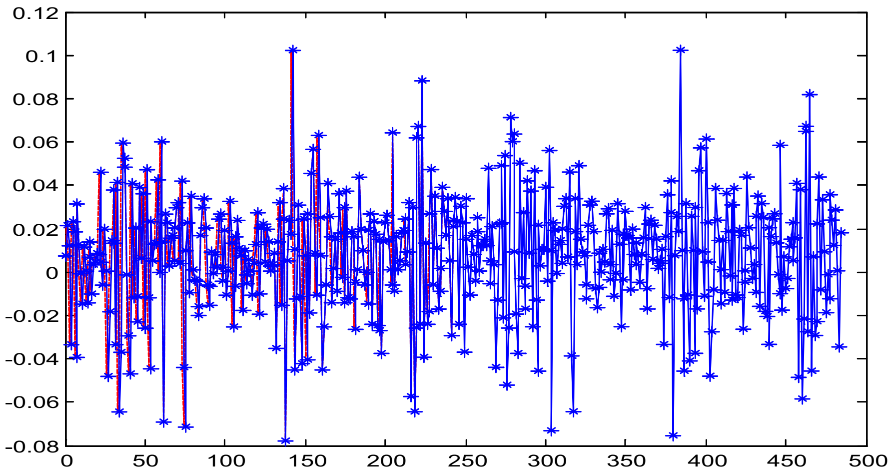

Fig. 5 Forecast of r_t

models have effective judgment in default prediction with listed companies in China, thus it can be inferred that structural model is effective in risk assessment with Chinese listed companies.

Furthermore, in order to compare the two structural models, we design two sets of experiments.

Firstly, test the distinguish ability. We use the data in 2009. According to Garch structural model and Wavelet structural model, we construct the paired group between non-ST and ST companies to do paired samples t-test. The results are shown in Tables 12 and 13.

The result can be seen from the Table 12 is that at 95 % confidence interval Wavelet structural model has smaller default distance for ST enterprises than time series model; meanwhile, the result can be seen from Table 13 is that at 95 % confidence interval wavelet structure model has greater default distance for non ST enterprises than time series model. Therefore, it is can be said that Wavelet structure model has better performance in default distinguish for both ST companies and non-ST companies.

Secondly, test prediction ability. We use the data in 2010 just like distinguish test, compare the two group predictive power based on the ST and non-ST companies. There is only one company which is non-ST in 2009 and is ST in 2010 with the stock code 600355. For time series model it gets the default distance is 2.067, and for wavelet structural model it has the default distance is 1.957. Because of the limited sample size, though it is not statistically inferred that wavelet structural model is superior to time series model, as far as in this case, wavelet structural model has better predictive power than time series model which has the default distance closer to 0. Also, we do paired t test to compare the two models' prediction ability with these companies which are ST in 2009 and non-ST in 2010, there are 13 companies, the results of paired t test are shown in the Table 14 below.

From Table 14, it is can be seen that default distance of wavelet structural model has larger mean than one of time series model. Because of sample size limitation, the

Table 10 Pair T-test of time series model

Variable	St_type	N	Lower CL mean	Mean	Upper CL mean	Lower CL SD	SD	Upper CL SD	SE	Min	Max
dd_ga	0	50	2.4404	2.5571	2.6736	0.3428	0.4103	0.5113	0.0580	2.0663	3.5388
dd_ga	1	50	0.3057	0.4717	1.6378	0.4881	0.5843	0.7281	0.0826	0.0000	2.3776
dd_ga	diff (1–2)		2.1347	2.0863	1.2856	0.4430	0.5048	0.5869	0.1010		

Variable	Method	Variances	DF	t-Value	P> |t|
dd_ga	Pooled	Equal	98	10.75	<0.0001
dd_ga	Satterthwaite	Unequal	87.9	10.75	<0.0001
dd_ga	Coohran	Unequal	49	10.75	<0.0001

Table 11 Pair T-test of wavelet analysis model

Variable	St_type	N	Lower CL mean	Mean	Upper CL mean	Lower CL SD	SD	Upper CL SD	SE	Min	Max
dd_wa	0	50	2.6281	2.7491	2.8701	0.3556	0.4257	0.5305	0.0602	1.9570	3.8333
dd_wa	1	50	1.0746	1.2602	1.4458	0.5455	0.6530	0.8137	0.0923	(0.0000)	2.9107
dd_wa	diff (1–2)		1.2701	1.4889	1.7077	0.4837	0.5512	0.6408	0.1102		

<table>
<tr><th>Variable</th><th>Method</th><th>Variances</th><th>DF</th><th>t-Value</th><th>P> |t|</th></tr>
<tr><td>dd_wa</td><td>Pooled</td><td>Equal</td><td>98</td><td>13.51</td><td><0.0001</td></tr>
<tr><td>dd_wa</td><td>Satterthwaite</td><td>Unequal</td><td>84.3</td><td>13.51</td><td><0.0001</td></tr>
<tr><td>dd_wa</td><td>Coohran</td><td>Unequal</td><td>49</td><td>13.51</td><td><0.0001</td></tr>
</table>

Table 12 Identify ST companies of these two models

Variable	St_type	N	Lower CL mean	Mean	Upper CL mean	Lower CL SD	SD	Upper CL SD	SE	Min	Max
dd	ga	50	1.3057	1.4717	1.6378	0.4881	0.5843	0.7281	0.0826	(0.0000)	2.3776
dd	wa	50	1.0746	1.2602	1.4458	0.5455	0.6530	0.8137	0.0923	(0.0000)	2.9107
dd	diff (1–2)		(0.0340)	0.2115	0.4574	0.5437	0.6196	0.7203	0.1239		
Variable		Method		Variances		DF		t-Value		P> |t|	
dd		Pooled		Equal		98		1.71		0.0912	
dd		Satterthwaite		Unequal		96.8		1.71		0.0917	
dd		Coohran		Unequal		49		1.71		0.0942	

Table 13 Identify Not ST companies of these two models

Variable	St_type	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL SD	SD	Upper CL SD	SE	Min	Max
dd	ga	50	2.4404	2.5570	2.6736	0.3428	0.4103	0.5113	0.0580	2.0663	3.5388
dd	wa	50	2.6281	2.7491	2.8701	0.3556	0.4257	0.5305	0.0602	1.9570	3.8333
dd	diff		(0.3580)	(0.1920)	(0.0260)	0.3669	0.4181	0.4861	0.0836		

Variable	Method	Variances	DF	t-Value	P> |t|
dd	Pooled	Equal	98	−2.3	0.0237
dd	Satterthwaite	Unequal	97.9	−2.3	0.0237
dd	Coohran	Unequal	49	−2.3	0.0259

Table 14 Forecast check of these two models

Variable	St_type	N	Lower CL mean	Mean	Upper CL mean	Lower CL SD	SD	Upper CL SD	SE	Min	Max
dd	ga	50	1.7323	1.9245	2.1167	0.2281	0.3181	0.5251	0.0882	1.4194	2.3524
dd	wa	50	1.8656	2.1299	2.3941	0.3136	0.4373	0.7218	0.1213	1.4729	2.9107
dd	diff (1–2)		(0.5150)	(0.2050)	0.1041	0.2985	0.3823	0.5319	0.1500		

Variable	Method	Variances	DF	t Value	P> |t|
dd	Pooled	Equal	24	−1.37	0.1835
dd	Satterthwaite	Unequal	21.9	−1.37	0.1847
dd	Coohran	Unequal	12	−1.37	0.1959

results are not very notable, but it can also be said that at the 80 % confidence interval wavelet structural model has better predictive power than time series model in default prediction.

6 Conclusion

Credit risk management is one of the most important problems which commercial bank should face with. Normally, the optimal way of credit risk management is to forecast the default accurately before the loan.

In the recent study, it is always using the structure model to forecast the default of the listed companies, because this model can realize the company's market value by mark-to-market, thus it is can be inferred that this model can give the overall information for a company, so it is more accurate, and be widely used in practice. But with the application process of structural model, it needs to estimate the company's equity value. According to the recent research, it usually use the time series modeling the volatility for the prediction of equity value, but with this method it can't avoid iterative calculation, so with the accumulation of time interval, the prediction will have large deviation.

Based on these, this paper puts forward a new method which is named wavelet structural model for default prediction. Regarding on characteristics of listed companies, the author sampled 100 companies according to industry types to construct wavelet structural model. The process of wavelet structural model is that: firstly apply wavelet decomposition on the proceeds, and then built different models separately for low frequency part and high frequency part, finally reconstruct the predictive return. So through this process, wavelet structural model can avoid accumulated calculate process of the volatility in time series model.

Checking with the actual situation of Chinese listed companies, it can be found that the wavelet structural model is more sensibility and more precision than time series model. But just as other structure models, wavelet structural model still cannot be able to avoid the calculation of the value of equity, so it has strong dependence with market environment, which limits its applications with the small and medium-sized enterprises.

Acknowledgements The work was supported by the National Social Science Foundation of China under Grant No. 13CTJ004, the National Natural Science Foundation of China under Grant No. 71232003, and the fundamental research funds for the Central Universities.

References

Altman, E. I. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, *21*, 1721–1742.

Baxter, M., & King, R. G. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics*, *81*, 575–593.

Bowden, R., & Zhu, J. (2008). The agribusiness cycle and its wavelets. *Empirical Economics*, *34*(3), 603–622.

Camara, A., Popova, L., & Simkins, B. A. (2011). Comparative study of the probability of default for global financial firms. *Journal of Banking & Finance*, *2*, 131–147.

Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, *59*, 383–403.
Chen, X. H., Wang, X. D., & Wu, D. (2010). Credit risk measurement and early warning of SMEs: An empirical study of listed SMEs in China. *Decision Support Systems*, *49*, 301–310.
Chiou, J. S., & Lee, Y. H. (2009). Jump dynamics and volatility: Oil and the stock markets markets. *Energy*, *34*, 788–796.
Crivellini, M., Gallegati, M., Gallegati, M., & Palestrini, A. (2003). Industrial output fluctuations in developed countries: A time-scale decomposition analysis[C]. 4th Eurostat and DG ECFIN Collequim on Modern Tools for Business Cycle Analysis.
Crosbie, P. J., & Bohn, J. R. (2002). Modeling default risk [EB]. http://www.kmv.com.
Crowley, P. M., Maraun, D., & Mayes, D. (2005). *Analysing productivity cycles in the euro area, US and UK using wavelet analysis [M]*. Mimeo: Texas A and M University.
Fernandez, V., & Kutan, A. M. (2005). Do regional integration agreements increase business cycle convergence? Evidence from APEC and NAFTA[Z], William Davidson Institute, Working Paper, No. 765.
Hamilton, J. (2003). What is an oil shock? *Journal of Econometrics*, *113*, 363–398.
Jagric, T. (2002). Measuring business cycles. *Eastern European Economics*, *40*(1), 63–87.
Jagric, T. (2003). Business cycle in central and east European countries. *Eastern European Economics*, *41*(5), 6–23.
Jagric, T. (2004). Method of analyzing business cycles in a transition economy: The case of Slovenia. *The Developing Economies*, *XLII–1*, 42–62.
Lee, W. (2011). Redefinition of the KMV model's optimal default point based on genetic algorithms-evidence from Taiwan. *Expert Systems with Applications*, *38*, 10107–10113.
Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, *29*, 449–470.
Nguyen, H. T., & Nabney, I. T. (2010). Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. *Energy*, *35*, 3674–3685.
Raihan, S. M., Wen, Y., & Zeng, B. (2005). Wavelet: A new tool for business cycle analysis [Z]. FRBSL, Working Paper Series, 050A.
Yalamova, R. (2006). Wavelet test of multifractality of Asia-Pacific index price series? *Asian Academy of Management Journal of Accounting and Finance*, *1*, 63–83.
Yamada, H., & Honda, Y. (2005). Do stock prices contain predictive information on business turning points? A wavelet analysis. *Applied Financial Economics Letters*, *1*, 19–23.
Ye, Q., Jing, N., & Xu, L. (2005). Research on credit risk measurement based on the capital market theory. *Economist*, *2*, 112–117.
Yogo, M. (2008). Measuring business cycles: A wavelet analysis of economic time series. *Economics Letters*, *100*(2), 208–212.