Modelling the spreading process of extreme risks via a simple agent-based model: Evidence from the China stock market

Jingru Ji‡, Donghua Wang ‡ and Dinghai Xu§

‡ Department of Finance, East China University of Science and Technology, Shanghai, 200237, China
§ Department of Economics, University of Waterloo, Ontario, N2L 3G1, Canada

Abstract

This paper focuses on investigating financial asset returns’ extreme risks, which are defined as the negative log-returns over a certain threshold. A simple agent-based model is constructed to explain the behavior of the market traders when extreme risks occur. We consider both the volatility clustering and the heavy tail characteristics when constructing the model. Empirical study uses the China securities index 300 daily level data and applies the method of simulated moments to estimate the model parameters. The stationarity and ergodicity tests provide evidence that the proposed model is good for estimation and prediction. The goodness-of-fit measures show that our proposed model fits the empirical data well. Our estimated model performs well in out-of-sample Value-at-Risk prediction, which contributes to the risk management.

Keywords: Agent-Based Model; Method of Simulated Moments; Extreme Risk; Value-at-Risk.

JEL Codes: C15, C52, G15
1 Introduction

The Agent-Based Model (ABM) has gained its popularity in recent years. One of its merits is that the ABM could provide supportive evidence from the aspect of behavioral finance. This makes the ABM different from some conventional econometric models (Farmer and Foley, 2009; LeBaron, 2000). Chen et al. (2012) classified the ABM into two types, namely the N-type design and the autonomous-agent design. For the autonomous-agent design, agents’ behavior is customized by artificial intelligence algorithms, which may lead the ABM to a black box. Thus, we cannot explain the causalities of the dynamic feedbacks (Westerhoff and Franke, 2012). For more details, see the work of LeBaron et al. (1999), Zhang et al. (2014), among others. Moreover, in general, the autonomous-agent design can not be estimated by traditional methods due to its model complexity (Recchioni et al., 2015). In this paper, we consider the N-type design as our fundamental structure. Basically, there are three N-type design ABMs: the adaptive belief system (Brock and Hommes, 1998), the interactive agent hypothesis (Kirman, 1993) and the ant type of system (Lux, 1995). Most of the new N-type design ABMs are based on these three fundamental mechanisms. For example, the FW model (Franke and Westerhoff, 2012) considers all the three mechanisms. In particular, we focus on a simple two-type design and consider these three fundamental mechanisms as a starting point. With few parameters and simple structures, it is possible to estimate the model and study the feedbacks of the agents.

For some ABMs, it is possible to obtain the closed-form moment conditions. For example, Ghonghadze and Lux (2016) derived the moment conditions of a simple asymmetric herding model, which was proposed by Alfarano et al. (2005). The generalized method of moments can be applied for the model estimation. For some ABMs, it is possible to simplify the model structure first before the estimation. For example, Li et al. (2016) transform a cross-market ABM into a threshold vector autoregression model, based on which one can estimate the model parameters using standard econometric methods. Recchioni et al. (2015) reconstructed the adaptive belief system model, which was proposed by Boswijk et al. (2007), and estimated it via minimizing the cumulated squared errors of the prices. However, not all ABMs can be estimated in such straightfor-
ward ways due to the complexity in the model structures. People consider alternatives in terms of estimating algorithms. In particular, when the analytical forms of moment conditions are not available, a popular alternative for estimating ABMs is the method of simulated moments (MSM), see for example, Grazzini and Richardi (2015), Winker et al. (2007), Franke (2009) and the references therein. In this paper, we adopt the MSM procedure for our model parameters’ estimation.

This paper focuses on investigating extreme risks on financial market. Therefore, two stylized facts, namely the volatility clustering effect and the heavy tail, should be taken into account. First, the volatility clustering effect is also referred to the autocorrelation effect, which indicates that periods of quiescence and turbulence tend to cluster together, (Lux 2000). Therefore, a homogeneous Poisson process is not suitable for modeling the occurrences of extreme risks. Instead, a self-exciting process proposed by Hawkes (1971) is more reasonable to depict the clustering property of extreme risks occurrences. From the aspect of behavioral finance, Lux (2000) explained the volatility clusters as the consequence of the market being subject to occasional temporary instability. In this paper, we introduce the concept of the market panics. This concept measures the risk degree of the market, based on which we can provide a mechanism to explain the occasional temporary instability. Therefore, our proposed framework is able to explain the clustering property of extreme risks occurrences properly.

Secondly, it is well-known that financial asset returns exhibit significant excess kurtosis (De Grauwe and Grimaldi 2006). Therefore, the conventional normality assumption is not appropriate especially when modelling the tail part of the distribution. In this paper, we propose to use the generalized Pareto distribution (GPD) (Embrechts et al. 1997, Chavez-Demoulin et al. 2006, Grothe et al. 2014). One of the advantages of the GPD is that it assumes a flexible structure (via changing the shape parameter) to accommodate various tail behaviors. In the general framework of the Extreme Value Theory (EVT), we assume that the extreme risks over a certain threshold are independently and identically distributed from a GPD. This is also referred to the Peaks over Threshold (POT) approach. Furthermore, the EVT can even go beyond constant tails. Quintos et al. (2001) adopted a break test and found that the tail risk in financial series
exists time-variation. Empirically, Wagner (2003, 2005) proposed models to measure the time-varying tail shapes based on the EVT. In this paper, we are interested in investigating the left tail of the log-return distribution which corresponds to the events causing extreme losses.

Moreover, it is important for our proposed model to have a good performance in predicting the out-of-sample risk. For example, Chavez-Demoulin and McGill (2012) use the Value-at-Risk (VaR) to measure the risk degree of the market and back test the performance of the out-of-sample prediction via three likelihood ratio tests (Christoffersen, 1998). The empirical result indicates that the VaR estimates computed via the proposed model can accurately measure the out-of-sample risk, which contributes to the risk management.

In summary, the general contribution of the paper is twofold. Theoretically, we propose an original ABM that can capture many important market characteristics, such as the price limits, thick-tailness, volatility clustering effect. Empirically, we apply our model to the China stock market and explain well the traders’ behavior across different types (noise and risky traders). In general, the results show that our proposed model perform well in both in-sample fitting and out-of-sample prediction.

The rest of the paper is organized as follows. Section 2 introduces the general theoretical structure of the proposed ABM. Section 3 presents the MSM algorithm and its estimation result. In section 4, we discuss the stationarity and ergodicity tests, the goodness-of-fit, the model mechanism and forecasting performance. We conclude with section 5.

2 Model

In the paper, the extreme risks are defined as the left tail of the return distribution. In other words, we focus on the negative log-returns over a certain positive threshold $u$. Therefore, the series of exceedances can be denoted as

$$W_t = \max \{-R_t - u, 0\}$$  \hfill (1)
where \( W_t \) stands for the extreme risk level at time \( t \), and \( R_t \) corresponds to the continuous compounded return at time \( t \). For the observed dataset, if we denote \( V_t \) as the observed stock price at time \( t \), we can get \( R_t = \ln (V_t) - \ln (V_{t-1}) \).

We propose to accommodate the volatility clustering effect and heavy tail by introducing the heterogeneous beliefs and the herding mechanism of the market participants (Cont et al. 2011). In our model, the heterogeneity is caused by the investors’ risk preference (Wen et al. 2014). For simplicity, we assume two types of agents, namely risky traders and noise traders, whose demands at time \( t \) are defined as follows

\[
D^r_t = \text{Sign}_t \epsilon^r_t \tag{2}
\]

and

\[
D^n_t = \epsilon^n_t \tag{3}
\]

where the superscripts \( r \) and \( n \) stand for the risky traders and noise traders. \( \text{Sign}_t \) is defined as

\[
\text{Sign}_t = \begin{cases} 
1 & \text{prob } = 0.5 \\
-1 & \text{prob } = 0.5 
\end{cases} \tag{4}
\]

Considering the price limits, we assume that \( \epsilon^r_t \) follows a truncated GPD, whose cumulative distribution function (CDF) can be expressed as

\[
F_{(\xi, \delta_u, \delta_l, \sigma_r)}(x) = \begin{cases} 
\frac{1 - (1 + \xi \frac{x}{\sigma_r})^{-1/\xi}}{1 - (1 + \xi \frac{\delta_u}{\sigma_r})^{-1/\xi}} & \text{Sign}_t = 1 \\
\frac{1 - (1 + \xi \frac{x}{\sigma_r})^{-1/\xi}}{1 - (1 + \xi \frac{\delta_l}{\sigma_r})^{-1/\xi}} & \text{Sign}_t = -1 
\end{cases} \tag{5}
\]

where \( \delta_u \) represents the upper bound of the log-returns, \( \delta_l \) stands for the lower bound of the log-returns, and \( \xi \neq 0 \). Similarly, we assume that \( \epsilon^n_t \) follows a truncated normal
distribution, whose CDF can be expressed as

\[ F_{(0, \sigma^2_n)}(x) = \frac{\Phi\left(\frac{x}{\sigma_n}\right) - \Phi\left(\frac{\delta_u}{\sigma_n}\right)}{\Phi\left(\frac{\delta_u}{\sigma_n}\right) - \Phi\left(\frac{\delta_l}{\sigma_n}\right)} \]  

(6)

where \( \Phi(\cdot) \) is the CDF of \( N(0, 1) \).

As mentioned earlier, the GPD can capture various tail behaviors through changing its shape parameter \( \xi \). For instance, if \( \xi \) is greater than 0, \( \varepsilon^r_t \) is heavy tailed. In general, we expect that \( \varepsilon^r_t \) has a heavier tail distribution than that of \( \varepsilon^n_t \). In that case, risky traders are more likely to make aggressive decisions than noise traders. Therefore, if the market is dominated by risky traders, the occurrences of extreme risks increase.

Next, we focus on modeling the dynamics of market fractions. This mechanism explains how agents switch their trading strategies at different time spots. Both transition probability approach (TPA) and discrete choice approach (DCA) are widely adopted in simple agent-based models [Franke and Westerhoff 2012]. In this paper, we adopt DCA to describe the switching mechanism, since it has been proved to be the most flexible mechanism in empirical study [Franke and Westerhoff 2012]. In detail, a DCA is based on the “Gibbs” probability

\[ n^i_t = \frac{\exp(\varphi U^i_t)}{\exp(\varphi U^r_t) + \exp(\varphi U^n_t)} \]  

(7)

where \( i = r, n \) and \( U^i_t \) stands for the utility of group \( i \) at time \( t \), and \( \varphi \) measures the intensity of choice. This setting can be equivalently expressed as

\[ n^r_t = \frac{1}{1 + P_t}, \quad n^n_t = 1 - n^r_t \]  

(8)

where the market panics \( P_t = \exp[-\varphi (U^r_t - U^n_t)] \). Because this expression cannot be directly computed, we measure \( P_t \) as

\[ P_t = \sum_{s=t-\tau}^{t-1} W_s \]  

(9)

where \( \tau \) is a positive integer, and \( W(\cdot) \) is defined in (1). (9) measures the aggregation of
the extreme risks during the former \( \tau \) periods. Substituting (9) into (8), we can find that the proportion of risky traders decreases as the market panics increases. This makes sense as the agents prefer more conservative strategies when the market risk arises.

Finally, the clearing mechanism of the stock should be specified. Empirical findings support that the excess demand (supply) linearly affects the price up (down) (Farmer and Joshi, 2002). Thus, considering the price limit of the market, the continuously compounded return at time \( t \) is of the form

\[
R_t = n_r D_r^t + n_n D_n^t
\]  

Note that \( R_t \) has the range \([\delta_l, \delta_u]\).

Based on the above data generating process (DGP), we can simulate the extreme risks with certain set of parameters. In section 3, we discuss the popularly used estimation method in the ABMs based on simulated moments.

## 3 The method of simulated moments

In this paper, the MSM is introduced to estimate the parameters given that the analytical forms of the moments in our model are not straightforward to achieve. The proposed model is applied to the China stock market data for empirical demonstration. Hence, we first preset the parameters \( \delta_u = \log (1.1) \) and \( \delta_l = \log (0.9) \), which are consistent with the price limit policy (±10%) in the China stock market. We set the threshold parameter, \( u \), to 0.01, which corresponds to the extreme risk level defined in this paper.\(^2\) We set \( \tau \) to be equal to 100, which is large enough to compute the aggregation of the market panics.\(^3\) Therefore, there are three parameters, denoted as a vector \( \theta = (\xi, \sigma_r, \sigma_n)' \), to be estimated.

The basic idea of the MSM is to minimize the discrepancies of the moment conditions between simulated and empirical data. In essence, we follow two key steps in this procedure: first, we simulate the data using our model as the DGP; then, we minimize

\(^2\)We have also experimented with other reasonable threshold levels. Our main conclusions of the paper are robust to such changes.

\(^3\)We have also checked other window lengths. The model is not sensitive to the selection of \( \tau \).
a loss function, which measures the distance between the artificial and empirical data.

Selecting a proper set of moments is one of the main challenges in all moment-based estimation procedures. In general, we would like to include the relevant moments that can capture the main properties of the extreme risks since this is our main target.

It is worth noting that $W_t$ is a marked point process. The most crucial properties of the series $W_t$ ($t = 1, 2, \cdots, T$) are the average magnitude of extreme risks and the average interval between two adjacent extreme risks. Suppose there are $N$ extreme risks till time $T$. The occurrence times of these extreme risks are denoted as $T_1, T_2, \cdots, T_N$ and the corresponding magnitudes are denoted as $X_{T_1}, X_{T_2}, \cdots, X_{T_N}$. We choose the mean of exceedances as the first moment condition

$$m_1 = \frac{1}{N} \sum_{i=1}^{N} X_{T_i} \quad (11)$$

$m_1$ captures the scale of the exceedances. The second moment condition that we use for estimation is the mean interval between two adjacent extreme risks, which can be written as

$$m_2 = \frac{1}{N-1} \sum_{i=1}^{N-1} (T_{i+1} - T_i) \quad (12)$$

$m_2$ measures the frequency of extreme risks.

As mentioned earlier, the heavy tail and the volatility clustering effects are also important features (McNeil, 1998; McNeil and Frey, 2000; McNeil et al., 2005) and should be taken into account in the estimated set of moments. Following Kon (2012), we capture the thick tails by choosing the kurtosis as the third moment condition, and we capture the clustering degree of exceedances by choosing the standard deviation of the interval between two adjacent extreme risks as the fourth moment condition. In summary, $\mathbf{m} = (m_1, m_2, m_3, m_4)'$ is our selected moments’ set for estimation.

Based on the selected moments, the standard loss function can be constructed as the following quadratic form

$$J = J (\mathbf{m}^S, \mathbf{m}^O; \Omega) = (\mathbf{m}^S - \mathbf{m}^O)' \Omega (\mathbf{m}^S - \mathbf{m}^O) \quad (13)$$

where $\mathbf{m}^S$ and $\mathbf{m}^O$ represent the moment vectors computed from the simulated and
observed data respectively. \( \Omega \) is a \( 4 \times 4 \) weighting matrix, which controls the penalty coefficients of the distances between \( m^S \) and \( m^O \). The weighting matrix plays an important role in the estimation procedure. First, the variance of \( m_i \) \( (i = 1, 2, \cdots, 4) \) caused by different samples from a certain population should be reciprocal to the penalty degree. Second, the dependence among \( m_i \) \( (i = 1, 2, \cdots, 4) \) should be considered in the estimation process. Following the standard way, we use the inverse of the variance-covariance matrix \( \Sigma \) as the weighting matrix. The \( \Sigma \) can be obtained through a bootstrap procedure. In particular, we first resample the data from the given population with a given random seed at the same length of the observed dataset, denoted as \( L \). Then, we compute the moment vector from the new sample. We repeat the above two steps \( N \) times (for example \( N = 5000 \)) and denote \( m^B(L,b) \) as the \( b^{th} \) moment vector computed with the random seed \( b \) at length \( L \). The \( \Sigma \) can be computed as

\[
\Sigma = \frac{1}{N} \sum_{b=1}^{N} \left[ m^B(L,b) - \frac{1}{N} \sum_{i=1}^{N} m^B(L,i) \right] \left[ m^B(L,b) - \frac{1}{N} \sum_{i=1}^{N} m^B(L,i) \right]'
\] (14)

For a given \( \theta \), a time series \( W_t \) \( (t = 1, 2, \cdots, 10000) \) can be simulated through the DGP introduced in section 2. We throw away the first 500 simulated observations to filter out the transient effect. Invoking the Nelder Mead algorithm (Nelder and Mead, 1965), we search for the optimal \( \theta \) to minimize the loss function. To reduce the variance from different simulated samples, we adopt random seeds ranging from 1 to 1,000 to simulate the samples and obtain the corresponding moment vectors. Formally, the optimization problem can be written as

\[
\hat{\theta}(a) = \arg \min_{\theta} J \left[ m^S(\theta; 10000, a), m^O(L) \right], a = 1, 2, \cdots, 1000
\] (15)

where \( a \) stands for the random seed. Based on (15), we can get the sample distribution of \( \hat{J}(a) \). Suppose there exists an \( \tilde{a} \), such that \( \hat{J}(\tilde{a}) \) is the median of \( \hat{J}(a) \). To filter out the run-specific effect, we set \( \tilde{\theta} := \hat{\theta}(\tilde{a}) \).

Our empirical study considers the applications to China market, which has the price limits of \( \pm 10\% \) as mentioned earlier. More importantly, our model is suitable for any market with price restrictions. CSI 300 that we use in this paper is for our empirical illustration for our theoretical model. Given the availability of the

---

\( ^4 \)More importantly, our model is suitable for any market with price restrictions. CSI 300 that we use in this paper is for our empirical illustration for our theoretical model. Given the availability of the
300 (CSI 300) daily level data from Jan. 4th, 2010 to Jan. 15th, 2016 as our in-sample data. The out-of-sample data starts from Jan. 18th, 2016 to Apr. 1st, 2016. As is shown in the upper panel of Figure 1, the black solid line represents the in-sample dataset, and the red dashed line represents the out-of-sample dataset. The middle panel of Figure 1 shows the log-returns of the CSI 300 with the same time interval presented in the upper panel. The Augmented Dickey-Fuller test shows that the log-return series rejects the unit root null hypothesis even at 1% significance level. Hence, the log-return series is stationary. The lower panel presents the 310 negative log-returns over the threshold \( u = 0.01 \) for the whole period. The blue vertical line separates the in-sample dataset from the out-of-sample dataset.

---

**Figure 1**: In the upper panel, the black solid line represents the in-sample dataset, and the red dashed line represents the out-of-sample dataset. The middle panel shows the log-returns of the CSI 300 with the same time interval presented in the upper panel. The lower panel shows the 310 negative log-returns over the threshold \( u = 0.01 \) for the whole period. The blue vertical line separates the in-sample dataset from the out-of-sample dataset.

---

*other market data (with price limits), we believe that our model can be easily applied to such data as well.*

*5 Our model can not predict the out-of-sample dataset well if the in-sample dataset is a non-ergodic time series. For example, if the in-sample dataset is from 2010 to 2014, the model may fail to predict the big fluctuations in 2015.*
Table 1 presents the summary statistics for the in-sample negative log-returns (the whole dataset) and exceedances (with the threshold $u = 0.01$). Note that the left tail of the daily log-returns is mapped into the positive domain for convenience. It is worth mentioning that the tail of the in-sample exceedances show a heavier tail than that of the log-returns. Table 2 reports the estimation result and standard error for each parameter. Moreover, last column, $J$, provides the information on the loss function. The graphical representations for the objective function are provided in Figure 2, and the graphical representations for Table 2 are provided in Figure 3.

**Table 1: In-sample summary statistics**

<table>
<thead>
<tr>
<th>Total negative log-returns</th>
<th>Excess negative log-returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0086%</td>
</tr>
<tr>
<td>St. deviation</td>
<td>0.0164</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.6824</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.0848</td>
</tr>
</tbody>
</table>

**Table 2: Model parameter estimation results**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\xi$</th>
<th>$\sigma_r$</th>
<th>$\sigma_n$</th>
<th>$J$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est.</td>
<td>2.0523</td>
<td>0.0014</td>
<td>0.0556</td>
<td>1.2361</td>
</tr>
<tr>
<td>St. deviation</td>
<td>0.1010</td>
<td>0.0002</td>
<td>0.0033</td>
<td>1.3884</td>
</tr>
</tbody>
</table>

![Figure 2: The objective function over the domain of each parameter.](image)
Figure 3: The distribution of each estimated parameter and $J$.

Figure 2 plots the objective function over the domain of each parameter of the model. The true parameters’ values are taken from Table 2. The objective function is evaluated with the simulated data under a fixed random seed. All parameters converge nicely. The red line in Figure 3 (a) represents the $\hat{\xi}(\tilde{a})$, and Figure 3 (b), (c) and (d) are the plots for the $\sigma_r$, $\sigma_n$ and $J$ respectively. $\xi$ measures the thick tailness of the risky group. As is expected, $\hat{\xi}$ is greater than 0 and is significantly different from 0, which indicates that the demand of the risky group has a heavy tail distribution. We conclude that the demand of the risky group is more volatile than that of the noise group, which coincides with our model setting described in section 2. $\sigma_r$ and $\sigma_n$ measure the uncertainty of each strategy, and both of the two parameters are significantly different from 0. We infer that the heterogeneity of the noise group is larger, since $\hat{\sigma}_n$ is greater than $\hat{\sigma}_r$. From the distribution of the parameters, we can conclude that each parameter has a good convergence. Figure 3 (d) plots the distribution of $\hat{J}(a)$. As is reported in Table 2, the median loss 1.2361, and the standard deviation, 1.3884, is relatively large. The reason causing this is due to some outliers of the tail of distribution $\hat{J}(a)$. We discuss this issue in the next section via a formal $J$ test based on a bootstrap procedure. We would introduce a more formal way to check our estimation in the next section. Moreover, Grazzini and Richardi (2015) mentioned that testing the stationarity and ergodicity for
an ABM is essential, hence, we also discuss the stationarity and ergodicity in the next section.

4 Model Diagnostics and Extreme Risk Analysis

4.1 Stationarity and Ergodicity Tests

Both stationarity and ergodicity are fundamental properties of a DGP (Grazzini, 2012). Stationarity indicates that the ABM converges to a statistical equilibrium state, in other words, the moments of the ABM are stationary. Ergodicity means that this statistical equilibrium is unique, which guarantees that the moment properties of the proposed DGP can be analyzed by using one long time series.

Both stationarity and ergodicity can be checked via the Runs Test (Wald and Wolfowitz, 1940). To test the stationarity of a particular moment, we first use a fixed random seed and the parameters reported in Table 2 to generate a long time series and compute the overall moment $m_i^A$ which is defined as the $i^{th}$ moment over the whole time series. Then, we split the time series into equal-size windows and compute the $i^{th}$ moment of $j^{th}$ window $m_{i,j}^W$ which is referred to the window moment. The $j^{th}$ symbol for the stationarity test can be expressed as

$$Sym_{i,j}^S = \begin{cases} 1 & m_{i,j}^W > m_i^A \\ 0 & \text{otherwise} \end{cases}$$

(16)

Next, we define a “run” as a consecutive of 1s or 0s whose preceding and following symbols are different (or even no symbol at all). We use the notation $U$ to represent the number of runs. Following Wald and Wolfowitz (1940), $U$ asymptotically converges to a normal distribution, whose mean and variance are

$$E (U) = \frac{2ab}{a+b} + 1$$

(17)

$$Var (U) = \frac{2ab(2ab - a - b)}{(a+b)^2 (a+b - 1)}$$

(18)

*Thanks to one anonymous referee to point out the important tests.
where $a$ stands for the number of symbol 1s and $b$ stands for the number of symbol 0s. The null hypothesis of stationarity cannot be rejected if the p-value of the two-tail test is greater than the significance level.

In the ergodicity test, we use the Runs Test to check whether two samples come from the same population. The first sample element denoted as $f_{i,j}$ is formed by the $i^{th}$ moment for the $j^{th}$ window from a long time series which is generated via a fixed random seed and the parameters reported in Table 2. The second sample element denoted as $g_{i,j}$ is formed by the $i^{th}$ moment from different time series generated by different random seeds $j$ and the same parameters reported in Table 2. Then, we use the set $H$ to record the union of $\{f_{i,j}\}$ and $\{g_{i,j}\}$, and each element of the set $H$ is arranged in an ascending order. Therefore, the $k^{th}$ symbol for the ergodicity test of the $i^{th}$ moment is

$$ Sym^E_{i,k} = \begin{cases} 
1 & h_{i,k} \in \{f_{i,j}\} \\
0 & h_{i,k} \in \{g_{i,j}\} 
\end{cases} \quad (19) $$

Then, we can follow the remaining steps of the Runs test to check the ergodicity for a specific moment. Following Grazzini (2012), we set the length of the time series as 1 million, and the length of the windows as 1000. Moreover, in the ergodicity tests, we use 1000 different random seeds to generate different time series, each of which is of length 1000.

Table 3: Stationarity and ergodicity tests for each moment

<table>
<thead>
<tr>
<th>Moment condition</th>
<th>Stationarity test</th>
<th>Ergodicity test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1$</td>
<td>0.1632***</td>
<td>0.6069***</td>
</tr>
<tr>
<td>$m_2$</td>
<td>0.1530***</td>
<td>0.8405***</td>
</tr>
<tr>
<td>$m_3$</td>
<td>0.5494***</td>
<td>0.0772**</td>
</tr>
<tr>
<td>$m_4$</td>
<td>0.6616***</td>
<td>0.4605***</td>
</tr>
</tbody>
</table>

\(^7\)Thanks to one anonymous referee to point out that the choice of the window length could potentially affect the ergodicity test results. We have experimented other window sizes including 500, 800, 2000, 3000, 4000 and 5000. We do not observe significance changes on our ergodicity test results. We have to admit that when the window size is too small, the ergodicity test may fail. However, in this paper, we do not recommend to use small window sizes, since the testing statistics may not be reliable based on small samples in our case. It would be interesting to investigate the choice of optimal window length. This is beyond the scope of this paper, and we will leave this for future research.
Table 3 reports the result of the stationarity and ergodicity tests for each moment. 

"**" indicates the significance at 1% level, "***" at 5% level and "****" at 10% level. As is shown in the second column, all the four selected moments are stationary at 10% significance level. The third column shows that the third moment condition is ergodic at 5% significance level, while other moment conditions are ergodic at 10% significance level.

4.2 The Goodness-of-Fit

In this subsection, we provide a simulation method to test the goodness-of-fit of our estimated model. In fact, Franke and Westerhoff (2012, 2016) provide two different approaches to test the goodness-of-fit. Although the approach proposed by Franke and Westerhoff (2016) is more straightforward and easy to understand, it requires a huge computational scale. Therefore, this paper utilizes the more convenient approach introduced by Franke and Westerhoff (2012) to test the goodness-of-fit.

Figure 4: The blue dashed line and the black solid line in Figure 4 represent the distributions of $J_B$ and $J_S$ respectively. Considering extreme cases with a significance level of 5%, we use the red dashed line to represent the 95% quantile of $J_B$, which is 7.0585.
The essence of this test is to compare the distributions of $J$ from the artificial and from the observed data: if the artificial $J$ is within the range of $J$ from the real-world series, we cannot reject the null hypothesis that the two distributions are the same. In particular, the $J$ values from the observed data can be simulated via a bootstrap method as follows

$$J^B := J \left[ m^B (L, b) \right] \quad (20)$$

The $J$ values from the simulated data can be simulated via our proposed DGP as follows

$$J^S := J \left[ m^S \left( \hat{\theta}, L, a \right) \right] \quad (21)$$

The blue dashed line and the black solid line in Figure 4 represent the distributions of $J^B$ and $J^S$ respectively. Considering extreme cases with a significance level of 5%, we use the red dashed line to represent the 95% quantile of $J^B$, which is 7.0585. If the distribution of $J^S$ falls substantially beyond the corresponding value of the red dashed line, the null hypothesis is rejected.

The probability of $J^S$ less than the 95% quantile of $J^B$ is 0.9054, which means that our model has a p-value of 0.9054. Thus, we can not reject the null hypothesis even at 10% significance level. In fact, this result indicates that the two $J$ distributions are not so different from each other, and the estimated DGP can explain the observed dataset well.

### 4.3 The mechanism of the model

In this subsection, we study the dynamic mechanisms of the model and explain why extreme risks spread. Overall, two important mechanisms contribute to the stylized facts, namely the aggregation of panics and the motion of the market fractions.

Since our ABM passes the stationarity and ergodicity tests, it is reasonable to analyze the model property using one long time series. Based on the estimated parameters, we reproduce the series $W_t$ at length 10000 and plot the evolutionary process of panics and market fractions respectively.
Figure 5: Figure 5 (a) and (b) plot the evolutionary process of market panics and its probability density function (PDF) at different time spots respectively. (c) and (d) plot the evolutionary process of risky traders’ proportion and its PDF respectively. (e) shows the details of our proposed DGP within a certain time period.

Figure 5 (a) and (b) plot the evolutionary process of market panics and its probability density function (PDF) at different time spots respectively. (c) and (d) plot the evolutionary process of risky traders’ proportion and its PDF at different time spots respectively. (e) shows the details of our proposed DGP within a certain time period. Figure 5 (b) shows that the panics are centered at approximately 0.2 with a long heavy tail on the right. Based on our model setting, when market panics increase, the risky traders’ proportion $n_r^t$ decreases. In other words, a risky strategy is preferred only if the market seems at a low risk degree, which is measured by the past performance of extreme risks. Figure 5 (d) shows that the risky traders’ proportion $n_r^t$ fluctuates in a relatively wide range $[0.5, 1]$. This indicates that the proposed switching mechanism is vital in explaining the microstructure of the market.

Finally, we would like to explain the formation of the two important stylized facts respectively, i.e. the heavy tail and the volatility clustering effect. As is mentioned earlier, the distribution of $D_t^n$ is normal, while the distribution of $D_t^t$ is heavy tailed. Hence, it is not hard to see that a combination of $D_t^n$ and $D_t^t$ follows a heavy tailed
distribution at each time spot. Then, we would explain the volatility clustering effect via a simple example. As is shown in figure 5 (e), the blue solid line represents the market panics and the vertical blue bars stand for the extreme risks within the period 6400-6600. The two red dashed lines split the period 6400-6600 into three parts. In the range 6400-6440, the market panics is relatively small. According to (8), risky traders may increase in the later period and thus lead to a heavier tail for the total demand. In other words, more extreme risks should occur and the market panics should increase in the later period. As is expected, we see a cluster of extreme risks within the range 6440-6540. Since the market panics arises in this range, the market fraction of risky traders should decrease in the later period and thus trigger a lighter tail for the total demand. That is to say, few extreme risks should occur and the market panics should decrease in the later period. The range 6540-6600 verifies our deduction. In short, the volatility clustering effect is influenced by the market panics.

4.4 Out-of-sample forecasting

Out-of-sample performance is also an important measure for model evaluation. There are two problems to be specified before predicting the out-of-sample risk. First, a proper risk measurement should be chosen as our prediction tool. Hong et al. (2009) propose that VaR has become a standard synthetic measure to evaluate extreme risks. Following Hong et al. (2009), we use the VaR as our risk measure in this paper. Second, we should select the length of the out-of-sample dataset. As mentioned earlier, our out-of-sample dataset starts from Jan. 18th, 2016 to Apr. 1st, 2016. This is because the length of the out-of-sample dataset should be large enough to test the VaR. It should possess the similar properties to the in-sample dataset as well. In other words, the interval of the out-of-sample dataset can not be too long, since the property of the daily data may change as time increases.

We compute the one step ahead VaR at level $p$ based on our estimated model. At time $t$, we can obtain the historical information till time $t$, thus, it is possible to simulate a large amount of scenarios of $W_{t+1}$ based on our estimated model. Then, we sort all

---

8In this paper, for illustration purpose, we just use the static window for forecasting. The rolling-window forecasting procedure can be done based on iterating/updating our static window prediction algorithm.
the simulated scenarios and find out the \( p \) quantile as our VaR estimates at level \( p \) on time \( t + 1 \). Empirically, we set \( p=90\% \) and 95\% to predict the risk.

Figure 6: The blue dotted line in Figure 6 represents the 90\% VaR estimates and the red dashed line represents the 95\% VaR estimates.

The blue dotted line in Figure 6 represents the 90\% VaR estimates and the red dashed line represents the 95\% VaR estimates. To back test the prediction result, we introduce the violation indicator variable \( I_{t+1}^p \), which is defined as

\[
I_{t+1}^p = \begin{cases} 
1 & \text{if } \text{VaR}_{t+1}^p < \ln(V_t) - \ln(V_{t+1}) - u \\
0 & \text{otherwise}
\end{cases}
\]  

(22)

where \( V_t \) stands for the observed underlying price at time \( t \). For simplicity, we denote the length of the out of sample as \( L' \). Thus, we get \( \lambda_1^p = \sum_{i=1}^{L'} I_i^p \) and \( \lambda_0^p = L' - \lambda_1^p \). Our evaluations of the VaR estimates at level \( p \) are based on three aspects. First, we compare the ratio \( \lambda_1^p / L' \) with the value \( 1 - p \). Theoretically, we assume that the discrepancies between the two ratios shall not be too big. This idea is called the unconditional coverage test (Christoffersen, 1998) and can be checked via

\[
LR_{ac} = 2 \ln \left( \frac{(\lambda_1^p / L')^{\lambda_1^p} (\lambda_0^p / L')^{\lambda_0^p}}{(1 - p)^{\lambda_1^p} p^{\lambda_0^p}} \right) \sim \chi^2(1)
\]

(23)
Second, we need to check whether each violation is independent with the others. Denote \( \lambda_{ij}^p \) as the number of transitions from state \( i \) to state \( j \) based on the VaR estimates at level \( p \), where \( i, j = 0, 1 \). The independence among violations can be tested via the following likelihood ratio test

\[
LR_{\text{ind}} = \frac{\left( \frac{\lambda_{00}^p}{\lambda_{00}^p + \lambda_{01}^p} \right)^{\lambda_{00}^p} \left( \frac{\lambda_{01}^p}{\lambda_{00}^p + \lambda_{01}^p} \right)^{\lambda_{01}^p} \left( \frac{\lambda_{10}^p}{\lambda_{10}^p + \lambda_{11}^p} \right)^{\lambda_{10}^p} \left( \frac{\lambda_{11}^p}{\lambda_{10}^p + \lambda_{11}^p} \right)^{\lambda_{11}^p}}{\left( \frac{\lambda_{10}^p}{\lambda_{10}^p + \lambda_{01}^p + \lambda_{11}^p} \right)^{\lambda_{10}^p + \lambda_{11}^p} \left( \frac{\lambda_{00}^p}{\lambda_{00}^p + \lambda_{10}^p + \lambda_{01}^p + \lambda_{11}^p} \right)^{\lambda_{00}^p + \lambda_{10}^p}} \sim \chi^2 (1) \quad (24)
\]

Finally, a joint test, which is also often referred to as the so-called conditional coverage test, is constructed. This can be expressed as

\[
LR_{\text{cc}} = LR_{\text{uc}} + LR_{\text{ind}} \sim \chi^2 (2) \quad (25)
\]

Table 4: The out-of-sample back test

<table>
<thead>
<tr>
<th>Test</th>
<th>( p=90% )</th>
<th>( p=95% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of violations</td>
<td>5 (5)</td>
<td>2 (2.5)</td>
</tr>
<tr>
<td>Unconditional</td>
<td>0.999***</td>
<td>0.737***</td>
</tr>
<tr>
<td>Independent</td>
<td>0.999***</td>
<td>0.987***</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.999***</td>
<td>0.945***</td>
</tr>
</tbody>
</table>

Table 4 reports the result of the three likelihood ratio tests at both 90\% and 95\% level. "***" indicates the significance at 1\% level, "**" at 5\% level and "*" at 10\% level. Taking 90\% VaR back testing result as an example, the number of violations is 5, while the expected number of violations is 5, which is reported in the parenthesis. We can conclude that both the ratio of violations and the independence among different violations satisfy our assumptions. In other words, we can not reject the null hypotheses even at 10\% level on the first two tests. Furthermore, the joint test also shows a statistically significant result. Similar conclusions can be made on the 95\% VaR estimates.

In summary, our estimated ABM has a reasonably good performance in VaR estimates.
5 Conclusion

We propose a simple ABM to explain the behavior of the market traders when extreme risks occur under price limits. Based on the heterogenous assumption, traders are classified into noise group and risky group. At each time spot, the proportion of each group fluctuates based on market panics. We explain the heavy tail via the heterogenous behavior across groups. The volatility clustering effect can be well explained by the change of market panics.

We can apply the proposed ABM to the observed data. Our empirical study uses the CSI 300 daily observations, whose negative exceedances over the threshold 0.01 are very heavy tailed. Four important moment conditions that can capture the main characteristics of negative exceedances are selected as the criteria of the estimation. We adopt the MSM algorithm to estimate the model parameters, and test the stationarity and ergodicity of the selected moments. The goodness-of-fit measures show that our proposed model fits the empirical data well. Moreover, we study the dynamic mechanisms of the model, and find the causality of both heavy tail and volatility clustering of negative exceedances. Finally, based on the estimated model, a Monte Carlo method is performed to predict the out-of-sample VaR at both 90% and 95% level. Three likelihood ratio tests show that our model has a good performance in predicting the out-of-sample VaR.

In future research, we would focus on developing the model into a higher dimensional case. Therefore, it is possible to analyze the risk contagions and the spreading process of market panics amongst different markets. Furthermore, we can also construct an ABM that takes the trading volume into account.

References


Wald A., Wolfowitz J., 1940. On a test whether two samples are from the same population. The Annals of Mathematical Statistics 11(2), 147-162.

