Calibrating the L-A Model to

Chinese Stocks

Fei Lin
Sharalyn Chen
Chun Chen
Hechen Yu

Spring 2010
MS&E 444 Invest Practice
Management Science and Engineering, Stanford University
Advisor: Professor Kay Giesecke
1.0 INTRODUCTION

This paper studies the stocks that listed on Hong Kong and Shanghai stock exchanges and researches the short-sale restriction's effect on stock prices, based on the model for hard-to-borrow (HTB) stocks presented by Avellaneda and Lipkin's 2009 paper, “A Dynamic Model for Hard-to-Borrow Stocks”¹ (L-A model).

Hong Kong and Shanghai’s stock markets serve as a good pair of comparison for this study. A large number of corporations have their stocks traded in both of these two markets. However, due to regulatory difference, only those in Hong Kong can be shorted while those in Shanghai cannot. By calibrating the L-A model to the stock prices and comparing the differences in calibrated parameters, we hope to determine how the short-sale restrictions impact Shanghai stocks’ prices.

To understand the L-A model, it is essential to know the following definitions:

**Short selling**

Short selling means selling the stock that is not in inventory. Specifically, the seller sells a stock he does not own by borrowing from a buyer. After that, the clearing firm representing the seller must deliver the stock within a stipulated amount of time. To make delivery, the seller must buy the stock in the market or borrow it from a stock-loan desk.

**Hard-to-borrow stocks (HTB):**

Short positions in HTBs may be forcibly repurchased (bought in) by the clearing firms. In general, this buy-in will be made in order to cover shortfalls in delivery of stock following the Securities and Exchange Commission’s Regulation SHO.

**Buy–ins:**

Traders enter a short position without contracts are faced with forcibly repurchase to cover the short positions.

Following is a roadmap of the study.

1) Introduce the L-A model and related definitions.
2) Outline a systematic method for model calibration and discuss various difficulties encountered.
3) Examine the sensitivities of model parameters, i.e. their impact on the corresponding distribution of stock returns.
4) Present the calibration results and interpret how they describe the differences between the two markets.
5) Extend the L-A model by incorporating short interest data. Calibrate the extended model to US stock prices.
6) Propose future work based on the discussion above.
2.0 THE MODEL

Avellaneda and Lipkin’s paper proposes that hard-to-borrowness of a stock can lead to a type of bubble dynamics. For example, traders holding a short position of HTB stocks are exposed to a higher risk of forced buy-ins, which usually leads to a significantly excess demand of the stocks unmatched by supply at the current prices. This fact pushes up the prices of HTB stocks, which are however subject to crash, as the excess demand disappears after the forced buy-ins are completed.

As standard stock models are not able to describe this effect of hard-to-borrowness, the authors present a new model with a joint set of stochastic differential equations that correlates the stock price and the buy-in rate. More specifically, this model incorporates both buy-in’s temporary positive effect on stock price, and the soon-following impact of the downward jumps. The following is the L-A model. It has six parameters.

\[
\frac{dS_t}{S_t} = \sigma dW_t + \gamma \lambda_t dt - \gamma dN_{\lambda t}(t)
\]

\[
dX_t = \kappa dZ_t + \alpha (\bar{X} - X_t) dt + \beta \frac{dS_t}{S_t}, \quad X_t = \ln \left( \frac{\lambda_t}{\lambda_0} \right)
\]

where:
- \(S_t\): stock price at time \(t\)
- \(\sigma\) and \(\kappa\): respective volatilities
- \(dW_t\) and \(dZ_t\): standard Brownian motions
- \(\lambda_t\): buy-in rate
- \(dN_{\lambda t}\): Poisson process with intensity \(\lambda\) over \((t, t+dt)\)
- \(\gamma\): price elasticity of demand due to buy-ins
- \(\alpha\): speed of mean reversion
- \(\bar{X}\): long term equilibrium of \(X_t\)
- \(\beta\): impact of stock price change on buy-in intensity

3.0 CALIBRATION ALGORITHM

In this study, the L-A model is calibrated to stock price returns. The Calibration process is essentially a six-dimensional optimization problem. This section discusses three aspects of the calibration, the objective functions, the NaN problem, and the grid search.

3.1 Objective Functions

Two different objective functions are used in multiple steps to optimize the calibration of the model parameters:

i. Minimize \(\text{Max} \left[ \text{pdf}_{\text{data}}(r) - \text{pdf}_{\text{fitted}}(r) \right]\) where \(r\) denotes returns

ii. Minimize \((\text{Mean square error of (mean + variance + skewness + kurtosis)}_{\text{data}}\) and \((\text{mean + variance + skewness + kurtosis)}_{\text{fitted}}\))

The first step is to find the 20 best parameters groups by minimizing the objective function i). Then
objective function ii), as locally accurate, is used to determine the best group of parameters from the preliminary selection of 20 groups.

The mechanisms of the two objective functions are discussed below.

1)  \[ \text{Min} \ (\text{Max} \ |pdf_{data}(r) - pdf_{fitted}(r)|) \]

This objective function minimizes the maximum distance between the probability density functions (pdf) of the actual and fitted stock returns. More specifically, the Kernel smoothing density function is utilized to estimate the pdf of each set of stock returns’ distribution from their histograms.

The benefit of minimizing this objective function is that it matches the stock returns’ distribution between the fitted and actual date. However, with the smoothing mechanism of Kernel smoothing density function, the resulting pdfs are just estimations instead of the real underlying pdfs. Therefore, minimizing the objective function 1) only minimize the estimations of the stock pdfs and might not be accurate enough sometimes to calibrate the parameters. Furthermore, from our observation, minimizing the largest distance does not necessarily guarantee an overall best fit, which is another drawback of this function.

In addition, it is found out that there are quite a few parameter sets that generate similar values to the minimal value from the objective function. That’s why the second objective function is supplemented for the purpose of further screening.

2)  \[ \text{Min} \ (\text{Mean square error of} \ (\text{mean} + \text{variance} + \text{skewness} + \text{kurtosis})_{data} \text{ and} \ (\text{mean} + \text{variance} + \text{skewness} + \text{kurtosis})_{fitted}) \]

This objective function is selected as an aid to objective function 1) to select the best parameter set from the 20 groups yielded in the first step. It further matches the specific features of the actual and the fitted returns. Nevertheless, it is not recommended to use it directly to calibrate the parameters because moments greater than three are too volatile to fit in general.

Research has been done on selecting the elements in this objective function. Combinations of the first through seventh moments as well as skewness and kurtosis alone are tested and compared. It is found that the combination of mean + variance + skewness + kurtosis makes the best fit.

Moreover, from our testing results, we find that sometimes one moment becomes so large that it dominates the objective function. For most of time, the kurtosis is in order of magnitude larger than other factors and making the fitting worse for other factors. Therefore, we also use the objective function in its ratio form to offset this effect. For example, the “difference of mean” term in the objective function will then become “difference of mean” divided by mean of the actual stock returns.

3.2 NaN Problems

One of the most troublesome issues encountered when calibrating the model is deriving results of ‘Not a Number’ (NaN) from Matlab. After prudently debugging, it is discovered that such a problem results from blow-up of the value of \( \lambda \). In this section, the characteristic embedded in the model causing this phenomenon will be analyzed and ways to prevent this problem from happening is then presented.
To reveal the cause of the problem, the dynamic of the stock returns and intensity of $\lambda$ should be analyzed. In equation (2), $\alpha$ governs the mean-reverting process while $\beta$ relates $X_t$ to the dynamic of stock returns. When $\beta$ is relative large, $X_t$ will be very sensitive to movements of stock returns. Since sudden drops caused by increase in short interest does not happen frequently, $X_t$ tends to grow gradually with the increases in stock returns. In case when $\alpha$ is small, the mean-reverting process might not be strong enough to pull $X_t$ immediately back to its mean should the sudden drops take place. As $X_t$ deviates from its mean, it will enter a vicious spiral of continuing increasing. The growth of $X_t$ is further aggravated by the exponential conversion between $X_t$ and $\lambda$. With exponential conversion, a little increase in $X_t$ will give rise to a significant increase in value of $\lambda$. A greater $\lambda$ will in turn increase the stock return, causing $X_t$ to grow even higher through $\beta$. The cycle effects continues when $\alpha$ is small and $\beta$ is large until the program blows up. By our empirical study, the program blows up when $X_t$ reaches 13.

To fix the problem, the key point is to alter the values of $\alpha$ and $\beta$ so that the mean-reverting process is strong enough to offset the continuous effect of stock’s upward movement on $X_t$. It is also found out that such a combination of $\alpha$ and $\beta$ is related to the value of $\gamma$ since $\gamma$ governs the growth of stock return. Put another way, for a specific $\gamma$, there exists a few combinations of $\alpha$ and $\beta$ that will not result in the blow-up. By numerous trial and error, a variety of sets of feasible combinations of $\alpha$ and $\beta$ that will not give rise to the blow-up are found for each fixed $\gamma$. Most of the combinations have large $\alpha$ value and lower $\beta$ value to ensure a powerful mean-reverting process when $\gamma$ is relatively large.

3.3 Grid Search

After determining the objective functions to be used, the next step is to find out the sets of parameters to be fitted. The methodology of grid search is used in which for each parameter, a set of up and lower bounds is given and parameters are discretized within that range. Since there are six parameters to be fitted, a large number of parameter sets are needed by grid search. For example, if ten different values are tested for each parameter, then there will be totally $10^6$ combinations of parameters. Running such a huge set of parameter could be very time consuming.

After conducting sensitivity analysis discussed in the next section, it is discovered that the objective function is very sensitive to $\sigma$ and $\lambda$, which also determine the basic shape of the simulated return distributions, while the other four parameters have relatively less impact. In addition, from the analysis of the NaN problem, we know for a specific $\lambda$, there is a threshold ratio for $\alpha$ and $\beta$. Using these different features of the parameters and their relationships allows us to reduce the total number of parameter sets significantly. In particular, we focus on obtaining the precise values of $\sigma$ and $\lambda$ while testing only a limited number of combinations of the other four parameters. In the end, we reduce the grid size to approximately 9000, which accelerates the calibration process substantially.

4.0 FITTING

4.1 Sensitivity of Parameters

Since there are six parameters to be fitted, it is important to realize how each of them affects the
simulation results. A sensitivity analysis is undertaken for this purpose. One of the stocks is picked and the fitted parameters are altered to see how the outcome changes correspondingly. The outcome tested is the mean square error between Kernel Smooth density of actual stock returns and simulated sock returns. The results are depicted in diagrams below supplemented with detailed interpretations.

4.1.1 σ

The figure below shows a +/- 25% sensitivity test on σ causes the simulation results to increase by 300% and 250%. Obviously, the simulation result is very sensitive to σ. The following two figures show how σ affects the distribution of stock returns.

![Tested parameter: σ, range: +/-25%](image)

Figure 1: Sensitivity analysis of σ

![Kernel smoothing distribution](image)

Figure 2: The impact of σ on fitted stock returns

The above figure shows the kernel smoothed distribution of the simulated returns when decreasing σ by 10% and keeping other parameters unchanged. With a smaller σ, the peak of the distribution is higher, which agrees with intuition as a smaller σ will cause the stock returns to be less volatile, making the distribution thinner and taller.

4.1.2 γ

Figure 3 shows that the simulation result is relatively insensitive to γ in absolute value terms. The shape of the return distribution, however, is sensitive to γ. As seen in Figure 4, as we increase γ by 40%, the simulated distribution deviates from the true mean and leans right. One explanation for this phenomenon is that γ governs magnitude of gradual increase and sudden drop of stock returns. Thus, a larger γ will push the stock returns higher, which is reflected by the distribution leaning to the right. As the number of drops is quite rare, it is not easy to capture that by looking at the distribution.
To sum up, $\gamma$ governs the skewness of the distribution.

4.1.3 $\text{Xbar, } \alpha, \beta \text{ and } \kappa$

From the four figures below, it can be seen that the simulation results are insensitive to $\text{Xbar, } \alpha, \beta \text{ and } \kappa$. By scrutinizing and comparing the kernel smooth density, it is found that the four parameters impose no systematic impact on the shape of the distribution.
4.2 Imperfect Fit Due to Fat Tails

After comparing the KS density of actual and fitted stock returns, it is discovered that the L-A model is not able to capture the fat tails of the stock returns. There are two potential explanations. First, the L-A model is based on Gaussian assumptions as it consists two Brownian motion terms. It is discussed in many literature that Gaussian family is generally unfit to model fat tail returns. We tried to replace the Brownian motion terms with t-distributions and observed that the fat-tail problem was largely improved. Secondly, as seen from the evolving path of the stock price of Volkswagen, the stock has several sudden spikes and sharp drops due to short squeeze. As a result, very large positive and negative returns are generated, giving rise to the fat tails in the distribution. Looking at the first equation of the L-A model, one can see that the first and second terms govern the gradual growth of stock and the third term governs the sudden drop. There is no element associated with sudden increase in stock prices. This nature of the model could introduce return bias when calibrating stocks with sudden jumps.

Figure 6: Volkswagen’s stock price over time

5.0 CALIBRATION RESULTS

5.1 Data

There are 61 stocks listed on both Hong Kong and Shanghai stock exchanges. The names of these stocks and their corresponding tickers can be found on a number of Chinese finance websites. The shares traded in Hong Kong are called “H shares” and the correspondents traded in Shanghai are called “A shares”. 51 out of the 61 H shares can be shorted in Hong Kong but none of the A shares can be shorted in Shanghai. The designated securities eligible for short selling is frequently updated by the Hong Kong stock exchange and can be found on its website. The data used by the first part of our analysis consists of stock prices only which we retrieved from Yahoo Finance. The short interest data used by the second part of our analysis which we incorporate into the L-A model is supplied by EVA.

5.2 Calibration Results – VOW

In this section, we first calibrate the model to Volkswagen, a typical hard-to-borrow stock in 2008, to see whether the fitted parameters have special characteristics. The sample period is from Jan 1 to Dec 31, 2008 during which the Volkswagen stock price rose from 150 Euro to almost 1000 Euro due to short squeeze.
The best fitted parameters are in the table below. This parameter is chosen among a grid of 13550 potential candidates. The first four moments are reasonably close to each other except for the skewness. As mentioned before, the Volkswagen stock price experienced drastic sudden increase which cannot be captured by the model. The fitted returns of the L-A model have negative skewness in general due to the sudden drops. Unless sudden climb is also incorporated into the model, fitted returns are not likely to be positively skewed. The parameter we should pay attention to is $\gamma$ as it is the only parameter in the model that can be associated with extreme price movements. Fitted $\gamma$ for Volkswagen has a value of one. This is a significant value and it reflects that the stock price has a “bubble”.

<table>
<thead>
<tr>
<th>Actual Returns</th>
<th>Fitted Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0065</td>
</tr>
<tr>
<td>STD</td>
<td>0.1171</td>
</tr>
<tr>
<td>Skewness</td>
<td>8.6589</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>105.7043</td>
</tr>
</tbody>
</table>

Table 1 Volkswagen calibration results

5.3 Calibrating Chinese Stocks

After examining the price movements of all the Chinese stocks listed in Shanghai and Hong Kong, we discover that Chinese stocks exhibit bubble effects in general, with A share prices exceeding H share prices. Since this bubble effect impacts all stocks, it is most likely due to systematic factors, rather than the degree of Hard-to-borrowness. Our calibration is hence divided into two categories. In one group we calibrate with data including bubble periods and in the other we do the opposite.

5.3.1 Calibration including bubble periods – Air China

Air China is a representative stock in this category because it has the typical bubble pattern that is common among many Chinese stocks between 2006 and 2010 as shown in the plot below.

Figure 7: Air China stock prices
Using a sample period from Aug 18, 2006 to Apr 19, 2010 and a grid size of 8800, the best fitted parameters are summarized in the table below. The differences in calibrated $\sigma$ and $\gamma$ can be explained by certain features of the stock prices. For example, the volatility of the realized returns is 4% for A share compared to 3.6% for H share. As a result, it is not surprising that the fitter $\sigma$ is also higher for A share. Secondly, the price ranges of the stock movements shed some light on the parameter $\gamma$. During the four-year period, A share fluctuates between 2.66 and 31.41 while H share fluctuates between 1.58 and 11.86. The fluctuation of the former is almost three times the size as that of the latter. One would reasonably expect A share to have a significantly higher $\gamma$ to account for such differences. The first four moments of the actual and fitted returns are approximately in-line except for A share kurtosis.

<table>
<thead>
<tr>
<th></th>
<th>$\sigma$</th>
<th>$\gamma$</th>
<th>$\bar{X}$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\kappa$</th>
<th>Grid Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Share</td>
<td>0.54</td>
<td>0.7</td>
<td>0</td>
<td>5.6</td>
<td>0.8</td>
<td>0.5</td>
<td>8800</td>
</tr>
<tr>
<td>H Share</td>
<td>0.4</td>
<td>-0.005</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.5</td>
<td>8800</td>
</tr>
</tbody>
</table>

Table 2 Air China calibration results

5.3.2 Calibration excluding bubble periods – China Railway Construction

China Railway Construction is a representative stock in this category. A shares and H shares are listed in Feb 2008, after the bubble period. From the following plot, one can see that their prices tend to track each other.

Using a sample period from Mar 13, 2008 to Apr 19, 2010 and a grid size of 8800, the best fitted parameters are summarized in the table below. The comparison of the first four moments and the kernel smoothed distributions of the actual and fitted illustrates that both parameter sets fit the data well. A share and H share have the same $\gamma$ values as their prices move closely with each other.
Different from previous scenarios, $\gamma$ is equal to 0.015 in both cases, significantly smaller than what one would expect for stocks with bubble patterns.

<table>
<thead>
<tr>
<th>China Railway Con: 3/13/2008 – 4/19/2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
</tr>
<tr>
<td>A Share</td>
</tr>
<tr>
<td>H Share</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A Share R</th>
<th>A Share Fit</th>
<th>H Share R</th>
<th>H Share Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.0003</td>
<td>0.0000</td>
<td>0.0001</td>
</tr>
<tr>
<td>STD</td>
<td>0.0264</td>
<td>0.0181</td>
<td>0.0327</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.0162</td>
<td>-0.0351</td>
<td>0.0205</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>6.2201</td>
<td>3.0105</td>
<td>10.5326</td>
</tr>
</tbody>
</table>

Table 3 China Railway Construction calibration results

In the L-A model, $\sigma$ and $\gamma$ have direct impact on stock prices while $\alpha, \beta, \kappa,$ and $\bar{X}$ have indirect impact. It is hence difficult to interpret the meaning of $\alpha, \beta, \kappa,$ and $\bar{X}$ individually, but they can be interpreted together. Collectively, they impact the movement of the “buy-in” intensity $\lambda$. As shown in the left graphs below, even though the fitted $\alpha, \beta$ and $\kappa$ are largely different between A share and H share, the net effect of these parameters bounds the movement of $\lambda$ between 0 and 40 for both cases. In addition, when fitted $\gamma$ is small, the correspondent $\lambda$ would be big and vice versa. In this case, the Poisson jump term, as it is multiplied by $\gamma$, would have minimal impact and would no longer characterize “sudden drops”. As a result, instead of being the “buy-in” intensity, more appropriately, $\lambda$ should be referred to as the intensity of the stock’s movement.

5.4 Summary of Calibration Results

In addition to Air China and China Rail Construction, other stocks are calibrated for each category. The fitted parameters are summarized in the table below.
Kolmogorov-Smirnov test has been performed to compare the distributions of the actual and fitted returns. For each company, after the best fitted parameters are chosen, fitted returns are simulated and compared to the actual returns 2000 times. Within each comparison, K-S test gives an indication whether the two distributions are the same or different. The null hypothesis is that two distributions are the same. The ‘KS Reject Ratio’ is calculated as the number of rejections of the null hypothesis over 2000. Hence, the bigger is the ratio, the stronger is the indication that the actual and fitted returns follow different distributions.

As illustrated, majority of stocks in category 1 fails the KS test while majority of stocks in category 2 passes the KS test. There are two implications:

• It is possible that stocks with bubble patterns are harder to calibrate. The grid size currently used is not yet big enough to locate the optimal parameter sets.
• It is also possible that for certain stocks, the optimal parameters would never be found as they might not be hard-to-borrow and the L-A model would be unfit.

At the end of report, we include future work which can potentially reduce the KS rejection ratio.

<table>
<thead>
<tr>
<th>Company</th>
<th>Sample Period</th>
<th>σ</th>
<th>γ</th>
<th>X̄</th>
<th>α</th>
<th>β</th>
<th>κ</th>
<th>Grid Size</th>
<th>KS Reject Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1: Calibration including bubble periods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volkswagen</td>
<td>1/1/08-12/31/08</td>
<td>0.2296</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.1365</td>
<td>0.0169</td>
<td>13550</td>
<td>97.2%</td>
</tr>
<tr>
<td>Air China A</td>
<td>8/18/06 - 4/19/10</td>
<td>0.54</td>
<td>0.7</td>
<td>0</td>
<td>5.6</td>
<td>0.8</td>
<td>0.5</td>
<td>8800</td>
<td>3.95%</td>
</tr>
<tr>
<td>Air China H</td>
<td>8/18/06 - 4/19/10</td>
<td>0.4</td>
<td>-0.005</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>0.5</td>
<td>8800</td>
<td>100%</td>
</tr>
<tr>
<td>AnG Steel A</td>
<td>1/1/03-4/19/10</td>
<td>0.32</td>
<td>0.9</td>
<td>3</td>
<td>0.9</td>
<td>0.1</td>
<td>0.3</td>
<td>8800</td>
<td>100%</td>
</tr>
<tr>
<td>AnG Steel H</td>
<td>1/1/03-4/19/10</td>
<td>0.46</td>
<td>0.3</td>
<td>-0.5</td>
<td>6.4</td>
<td>0.8</td>
<td>0.1</td>
<td>8800</td>
<td>100%</td>
</tr>
<tr>
<td>TsingTao A</td>
<td>1/1/03-4/19/10</td>
<td>0.23</td>
<td>0.015</td>
<td>3</td>
<td>100</td>
<td>10</td>
<td>0.5</td>
<td>8800</td>
<td>100%</td>
</tr>
<tr>
<td>TsingTao H</td>
<td>1/1/03-4/19/10</td>
<td>0.32</td>
<td>0.015</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0.8</td>
<td>8800</td>
<td>100%</td>
</tr>
<tr>
<td>Yanzhou A</td>
<td>1/1/03-4/19/10</td>
<td>0.32</td>
<td>0.03</td>
<td>2</td>
<td>2.5</td>
<td>5</td>
<td>0.1</td>
<td>8800</td>
<td>100%</td>
</tr>
<tr>
<td>Yanzhou H</td>
<td>1/1/03-4/19/10</td>
<td>0.36</td>
<td>0.045</td>
<td>3</td>
<td>50</td>
<td>5</td>
<td>0.8</td>
<td>8800</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Category 2: Calibration excluding bubble periods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHN Rail Co A</td>
<td>3/13/08-4/19/10</td>
<td>0.28</td>
<td>0.015</td>
<td>3</td>
<td>100</td>
<td>10</td>
<td>0.3</td>
<td>8800</td>
<td>1.4%</td>
</tr>
<tr>
<td>CHN Rail Co H</td>
<td>3/13/08-4/19/10</td>
<td>0.4</td>
<td>0.015</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0.8</td>
<td>8800</td>
<td>2.35%</td>
</tr>
<tr>
<td>Bank of Com A</td>
<td>5/15/07-4/19/10</td>
<td>0.37</td>
<td>0.035</td>
<td>2</td>
<td>0.5</td>
<td>1</td>
<td>0.8</td>
<td>8800</td>
<td>0.7%</td>
</tr>
<tr>
<td>Bank of Com H</td>
<td>5/15/07-4/19/10</td>
<td>0.28</td>
<td>0.025</td>
<td>3</td>
<td>100</td>
<td>10</td>
<td>0.3</td>
<td>8800</td>
<td>89.55%</td>
</tr>
<tr>
<td>CHN Rail Grp A</td>
<td>3/25/08-4/19/10</td>
<td>0.28</td>
<td>0.055</td>
<td>1</td>
<td>50</td>
<td>5</td>
<td>0.5</td>
<td>8800</td>
<td>3.6%</td>
</tr>
<tr>
<td>CHN Rail Grp H</td>
<td>3/25/08-4/19/10</td>
<td>0.39</td>
<td>0.045</td>
<td>-0.5</td>
<td>5</td>
<td>1</td>
<td>0.1</td>
<td>8800</td>
<td>2.85%</td>
</tr>
<tr>
<td>METALLURGICAL A</td>
<td>9/24/09-4/19/10</td>
<td>0.21</td>
<td>-0.025</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>0.1</td>
<td>8800</td>
<td>7.2%</td>
</tr>
<tr>
<td>METALLURGICAL H</td>
<td>9/24/09-4/19/10</td>
<td>0.26</td>
<td>0.005</td>
<td>3</td>
<td>100</td>
<td>10</td>
<td>0.5</td>
<td>8800</td>
<td>26.65%</td>
</tr>
</tbody>
</table>

Table 4 Summary of calibration results

6.0 EXTENSION TO THE L-A MODEL

Advised by EVA, in this section, we extend the current L-A model to include short interest data. We will calibrate this extended model to six stocks traded in the US market to see whether we can detect
any relationship between short interest data and stock prices.

Since the short interest ratios are not available, utilization, which corresponds closest to short interest, is used instead. Utilization is defined as:

\[
\text{Utilization} = \frac{\text{stocks out on loan}}{\text{stocks available to lend}}
\]

6.1 Algorithm

We assume utilization is the true value for \( \{\lambda_t\}_{t \geq 0} \). This extra assumption allows us to use a more efficient algorithm to find the optimal parameter set which is explained in detail later. Maximum Likelihood Estimation is used to find the initial parameter set. This initial set is in “good neighborhood” hence grid search will no longer be required. The second step is to solve the minimization problem shown below using Matlab built-in optimization functions.

\[
\text{Min} \quad \text{(Max}[\text{cdf}_{\text{actual}}(r) - \text{cdf}_{\text{fitted}}(r)])
\]

subject to \( \sigma \geq 0, \gamma \geq 0, \kappa \geq 0 \)

where \( \text{cdf}_{\text{actual}}(r) \) is the cumulative distribution function of the actual returns and \( \text{cdf}_{\text{fitted}}(r) \) is the equivalent for the fitted returns.

Let the vector of parameters for this problem to be \( \theta = \{\sigma, \gamma, \bar{X}, \alpha, \beta, \kappa\} \). The two likelihoods \( L_1 \) and \( L_2 \), given the successive stock prices \( \{S_0, S_1, \ldots, S_T\} \) and buy-in rates \( \{\lambda_0, \lambda_1, \ldots, \lambda_T\} \), are given by:

\[
\log L_1(\theta) = -\frac{T}{2} \log(2\pi) - T \log(\sigma) + \log \left((1 - \lambda_t dt) \exp\left(\frac{(r_t - \gamma \lambda_t dt - \gamma)}{2\sigma^2}\right) + \lambda_t dt \exp\left(\frac{(r_t - \gamma \lambda_t dt - \gamma)}{2\sigma^2}\right)\right)
\]

\[
\log L_2(\theta) = -\frac{T}{2} \log(2\pi) - T \log(\kappa) - \sum_{t=1}^{T} \frac{(dX_t - \alpha(\bar{X} - X_t)dt - \beta r_t)^2}{2\kappa^2}
\]

where \( dt = \frac{1}{252}, \quad r_t = \frac{S_t - S_{t-1}}{S_{t-1}}, \quad X_t = \log\left(\frac{\lambda_t}{\lambda_0}\right), \) and \( dX_t = X_t - X_{t-1} \).

The estimators of \( \sigma, \gamma \) can be obtained by maximizing \( \log L_1(\theta) \) using the matlab function fmincon. The estimators of \( \alpha, \beta, \) and \( \kappa \) can be written down in terms of the given stock returns and buy-in rates:

\[
\hat{\alpha} = \frac{AG - EC}{BG - FC}, \quad \hat{\beta} = \frac{AF - BE}{CF - BG}, \quad \hat{\kappa} = \sqrt{\frac{\sum_{t=1}^{T} (dX_t - \hat{\alpha}(X_t - X_t)dt - \hat{\beta} r_t)^2}{T}}, \quad \bar{X} = \frac{\sum_{t=1}^{T} X_t}{T}
\]

where \( A = \sum_{t=1}^{T} dX_t (X_t - \bar{X}) dt, \quad B = \sum_{t=1}^{T} (X_t - \bar{X})^2 dt, \quad C = \sum_{t=1}^{T} r_t (X_t - \bar{X}) dt, \quad E = \sum_{t=1}^{T} dX_t r_t, \quad F = \sum_{t=1}^{T} (X_t - \bar{X}) dt \times r_t, \quad G = \sum_{t=1}^{T} (r_t)^2 \).
After obtaining the initial parameter set \( \hat{\theta} = \{ \hat{\sigma}, \hat{\gamma}, \hat{\bar{X}}, \hat{\alpha}, \hat{\beta}, \hat{\kappa} \} \), matlab function fmincon is used to find the optimal parameter set. In the end, the Kolmogorov-Smirnov test is conducted to check if the simulated stock returns and actual returns are from the same distribution. The result will be 1 if the test rejects the null hypothesis at the 5% significance level; otherwise, it will return 0.

6.2 Results
The following table shows the optimal parameter sets and the corresponding K-S test results. Three stocks with high utilizations and three stocks with low utilizations are chosen.

<table>
<thead>
<tr>
<th>Stock</th>
<th>( \sigma )</th>
<th>( \gamma )</th>
<th>Xbar</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \kappa )</th>
<th>KS Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Utilization</td>
<td>MYL</td>
<td>0.30628</td>
<td>0.47833</td>
<td>0.59946</td>
<td>0.83026</td>
<td>-0.20937</td>
<td>0.062753</td>
</tr>
<tr>
<td></td>
<td>GRMN</td>
<td>0.53519</td>
<td>0.20898</td>
<td>-0.03102</td>
<td>7.4123</td>
<td>-0.40229</td>
<td>0.076586</td>
</tr>
<tr>
<td></td>
<td>MGM</td>
<td>0.57843</td>
<td>0.16064</td>
<td>-0.00404</td>
<td>1.6727</td>
<td>-0.13389</td>
<td>0.086211</td>
</tr>
<tr>
<td>Low Utilization</td>
<td>PEP</td>
<td>0.18374</td>
<td>4.44845</td>
<td>0.54541</td>
<td>10.2863</td>
<td>-0.52841</td>
<td>0.147464</td>
</tr>
<tr>
<td></td>
<td>RE</td>
<td>0.2667</td>
<td>0.69302</td>
<td>0.74866</td>
<td>7.6965</td>
<td>-0.81051</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>ABT</td>
<td>0.20721</td>
<td>5.87447</td>
<td>0.330319</td>
<td>11.1984</td>
<td>-0.33382</td>
<td>0.199646</td>
</tr>
</tbody>
</table>

Table 5 Summary of calibration results for the extended model

From the table we can see that the K-S test did not reject the null hypothesis, i.e., the assumption that the simulated stock returns based on the best fitted parameter sets and the actual returns are from the same distribution is acceptable. All \( \beta \)s are negative, indicating that the stock returns have negative effects on the utilization. The higher the stock returns are, the smaller the utilizations will be. The low-utilization class tend to have larger \( \gamma \), suggesting that if the buy-in happens, the low-utilization class will likely experience a bigger bubble than the high-utilization class. One of the fundamental assumptions of the L-A model is that \( \beta \) is greater than zero, indicating a positive correlation between buy-in intensity and prices changes. Our empirical result shows the contrary.

7.0 CONCLUSIONS
Out analysis demonstrates that Hard-to-borrow stocks tend to have a significant \( \gamma \) value after the L-A model calibration. From calibrating the Chinese stocks, we learn that calibration excluding bubble periods yields small \( \gamma \) values. This is the case for all eight stocks considered in the category. Calibration including bubble periods, however, gives mixed results. We observe that the existence of bubble alone does not guarantee a significant \( \gamma \) value. Most often, bubble must be also accompanied by extreme price movements to be considered as a potential Hard-to-borrow stock. The extension of the L-A model to incorporate short interest data yields meaningful results. We show that stock returns have negative effects on the utilization rate, which is the opposite of what is proposed in the L-A’s paper for a hard-to-borrow stock, implying that the six fitted US stocks are likely not hard-to-borrow.

8.0 FUTURE WORK
8.1 Improve grid search

Due to the dimension of the optimization, it is possible that the parameters from our calibration are not globally optimized. One future improvement is to implement finer grids, especially for the cases where the K-S test is not passed.

8.2 Improve the objective function

The best fitted parameters derived from the current objective functions do not guarantee that the simulated returns will pass the K-S test. One future improvement is to use the K-S test as the objective function, avoiding fitting probability density functions and higher moments which could be extremely volatile.

8.3 Obtain additional data

Avellaneda and Lipkin’s model option prices as if the stock pays a continuous dividend, reflecting a modified form of Put-Call Parity. A stock that does not pay such a dividend may have calls subject to early exercise. This is an effective way to evaluate if a stock is having heavy short interest rate or not, which is an important criteria for stock selection. Short interest rate data and option prices are not available for our study, making it very difficult to derive meaningful results to directly explain the relationship between short interest and stock prices. The study can definitely be improved with additional data.

8.4 Try different markets

Emerging market is less mature and standardized. For example, the Shanghai stock exchange is heavily regulated by Chinese government. Any policy change may result in price changes that are hard to predict or impossible to be explained by models. A good example is the years from 2007 to 2009. Since the L-A model is recently proposed, it might be more meaningful to calibrate it to stocks in mature market first. Once the study is complete for the developed markets, it can be conducted for emerging market as a comparison.

9.0 REFERENCES


10.0 APPENDIX

Plots for calibrating the extended model
Figure 11: MYL

Figure 12: GRMN
Figure 13: MGM

Figure 14: PEP
Figure 15: RE

Figure 16: ABT