Nucleation of Market Bubbles

Arun Suresh

Abstract

This project explores the extent to which homogenious nucleation equations can be used to model financial bubbles. In doing so, it translates various aspects of classical and modern nucleation theory of bubbles into the language of financial markets - in order to better study the indicators of a financial collapse, and thereby better predict potential market collapses that may occur in the future. The project builds a model using training data from the housing bubble of 2007. Upon sufficient training/extrapolation, the model was applied to current market trends to observe that we are currently in a bubble economy. This verifies with a lot of recent speculations from various economists. We conclude with reasonable evidence that there is indeed a very natural connection between the nucleation and growth of physical and economic bubbles. We also acknowledge the need for more research in this budding field.

1 Introduction

The field of econo-physics is one that is rapidly growing and is the center for a lot of active research. This project in particular explores the applications of nucleation theory to financial market bubbles. In order to keep the project self contained, we introduce a few simple definitions.

Definition 1. ***Financial Bubble*** *A period of time in which market speculation causes the price of an asset to inflate far past its intrinsic value.*

Definition 2. ***Market crash*** *The financial market is said to crash when the market value of an asset sharply returns to the price before the bubble.*

Before understanding what nucleations of bubbles are in the physical sense, we introduce a priliminary definition that will aid our understanding. Consider a physical system, typically just matter in some state

Definition 3. ***First order phase transitions*** *A transition of phase during which the system either absorbs or releases a fixed amount of energy per unit volume. The temperature of the system stays constant as heat is added; the system is in a "mixed" phase regime in which some parts of the system have completed the transition and others have not.*

First order Phase transitions play an important role in science, and are useful in many technical applications. Simple examples are condensations, evaporation, boiling, etc.

Definition 4. ***Nucleation*** *The first step of the phase transition is essentially overcoming a free-energy barrier, which is the work done towards the formation of a small "embryo" and evolution of the embryo to become the nucleus of the new phase. This nucleus can only emerge from random thermal fluctuations within the old metastable phase. This initiating process of a first order phase transition is called nucleation.*

The following infographic provides a lot of insight into the nucleation process and the growth of bubbles.

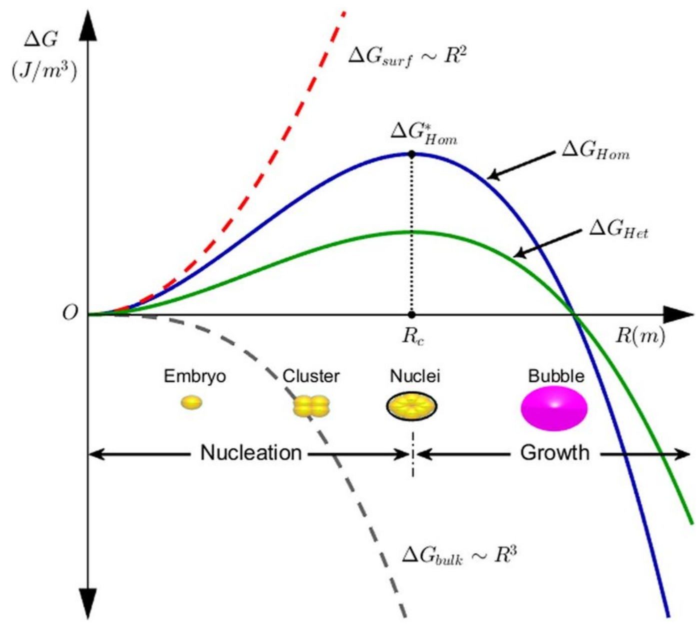

Figure 1: Physical formation of bubbles

A nucleation is said to be *Homogeneous* if the new phase is solely formed from the fluctuations within the old phase.

2 Building the model

2.1 Physical set up

In our project, we wish to use homogeneous/heterogeneous nucleation models, namely the ones pertaining to the change in Gibbs' free energy in a material going through a phase change, to gain a better understanding of how bubbles form in financial markets over time.
We will consider using the JMAK equation to model the phenomenon more effectively with respect to time. Presented below is the JMAK equation in its generality.

$$\frac{V_\beta}{V} = Y = 1 - e^{-K(t)^4} \tag{1}$$

and

$$K = \frac{\pi}{3}\dot{N}\dot{G}^3 \tag{2}$$

where,
V_β is the proportion of the volume that has already undergone nucleation,
V is the total volume of the substance
$\dot{N}$ is the rate of nucleation
$\dot{G}$ is the rate of growth of radius.

The equations for $\dot{N}$ and $\dot{G}$ stem from equations of their own. But before listing them, we introduce a few more preliminary equations.

$$\Delta G_V = G_V^{P_1} - G_V^{P_2} \tag{3}$$

is the change in Gibbs free energy per unit volume as the substance transitions from one phase P_1 to another P_2.

We denote by γ the interfacial free energy. With the help of this and ΔG_V, we can now identify a critical radius r^* that would differentiate an embryo from a nucleus. Thus, this is also the radius at which "bubbles form"

$$r^* = \frac{2\gamma}{\Delta G_V} \tag{4}$$

and the associated critical change in gibbs energy (barrier for bubble formation) is given by

$$\Delta G^* = \frac{16\pi\gamma^3}{3(\Delta G_V)^2} \tag{5}$$

Finally, assuming that we have a spherical bubble of volume V, we have

$$r = \sqrt[3]{\frac{3V}{4\pi}} \tag{6}$$

as the radius of the bubble.

With all of these equations, we are finally ready to define $\dot{N}$ and $\dot{G}$. They are defined as follows.

$$\dot{N} = ae^{\frac{\Delta G^*}{f_0 T}} \tag{7}$$

$$\dot{G} = \frac{dr}{dt} \tag{8}$$

where, a, f_0 are some positive scaling factors involving fractional multiples of π, which upon inclusion would only further complicate the equation.

Presented below is a plot of the Avrami equation.

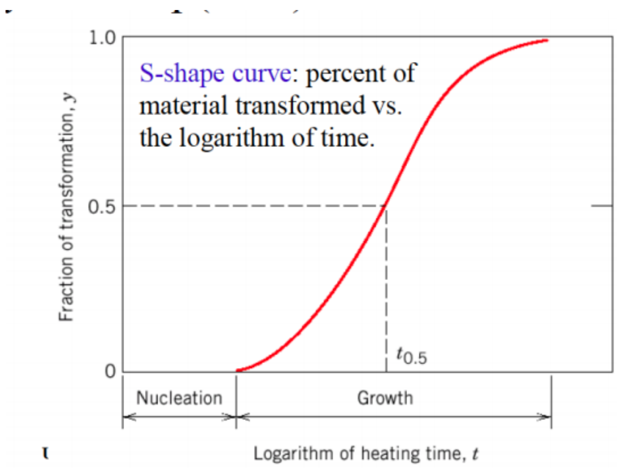

Figure 2: Plot of Y vs $\log(t)$

2.2 Translating to economics terms

We devote this section to translate the equations seen in the above subsection, into economic terms - by identifying economic variables and their correlation to physical variables.

Notation 1. *Denote by* V_j $j \in \overline{1,n}$ *the market cap of a company in an industry, and denote by* V_i $i \in \overline{1,m}$ *the market cap of an industry.*

A simple observation is that

$$\sum_{j=1}^{n} V_j = V_i \tag{9}$$

In this sense, we can redefine our physical variables as follows:
V_β: The summed market cap of companies (in an industry) in the bubble phase
V: The summed market cap of the industry.

It is clear that $0 \leq \frac{V_\beta}{V} \leq 1$.

This covers the left hand side of the JMAK equation. For the right hand side, we study k. We can then redefine $\dot{N}$ and $\dot{G}$ as follows

$\dot{N}$: Rate at which stocks transition into the bubble phase.
$\dot{G}$ Rate at which the market cap of a company grows.

This allows us to think of bubble nucleation as companies entering the bubble phase. We can also define a quantity called "prepice" that is analogous to the radius of the bubble in the physical sense

$$r = \sqrt[3]{\frac{3V_j}{4\pi}} \tag{10}$$

that is proportional to the cube-root of the market cap.
We can also redefine γ the surface interface energy as market forces that would prevent bubbles from forming, and in the same spirit, redefine ΔG_V, the change in gibbs free energy per unit volume, as factors that assist in the growth of a bubble.
In a more formal economic sense, we can define

$$\gamma_j = \frac{B_j}{V_j}\sqrt[3]{C_i} \tag{11}$$

as the book value to market-cap ratio times the cube root of the average Cost (barriers to market entry) of the industry to which it belongs. The bigger the value of γ (or, in other words, the more stable a company/industry is) the harder it should be for the company to become a bubble. ΔG_V can also be modified as

$$\Delta G_V = |\nu_{actual} + \nu_{implied}|\left|\frac{F_j}{V_j}\right| \tag{12}$$

where ν_{actual} and $\nu_{implied}$ are the actual and implied volatility of the company; and F_j is the free cash flow for the company. The higher the value of ΔG_V (that is to say, the more the volatility or Cash flow of a company), the easier it should be for the company to become a bubble.

With these definitions in hand, we can now define the "critical prepice" as the price value (or the critical radius in the physical sense) at which a company enters a bubble.

$$r^* = \frac{\zeta\gamma_j}{\Delta G_{Vj}} \tag{13}$$

Along with this definition, we also make an implicit assumption that a company can either be inside a bubble ($r > r^*$) or outside a bubble ($r < r^*$). We can also associate a ΔG^* to this set up, which is more or less defined through r^* (or can be thought of as the critical combination of F and ν at which the company enters the bubble phase).

$$\Delta G^* = b\frac{\gamma^3}{\Delta G_V^2} \tag{14}$$

where b is a positive scaling factor.

With this we are finally ready to define $\dot{N}$ and $\dot{G}$ in a way that is analogous to the physical set-up. Their equations are as follows.

$$\dot{N} = f c_0 e^{\frac{\Delta G^*}{T}} \tag{15}$$

$$\dot{G} = \frac{dr}{dt} \tag{16}$$

where

$$c_0 = n\frac{r_i^{*3}}{V_i^2} \tag{17}$$

with n being the number of companies in the bubble phase; and

$$T = \frac{V_i}{1+\psi} \tag{18}$$

where ψ is the federal interest rate[1].
Putting it all together, we have our JMAK equation for financial market bubble nucleation as

$$\frac{V_\beta}{V} = Y = 1 - e^{-k^2(t)^4} \tag{19}$$

where

$$k = a\dot{N}\dot{G}^3 \tag{20}$$

The reader can find more information about the variables and their qualitative justification in the Appendix.

[1] We add 1 to the interest rate in order to avoid divide by zero errors. The minimum observed interest rate was -0.80

3 Data

Data collection for our model was generally split into two phases. In order to apply our model to historical financial bubbles, we obtained data from Fred Economic data, a resource powered by the federal reserve bank of St. Louis. An example plot of the change in the volatility index as a time series data.

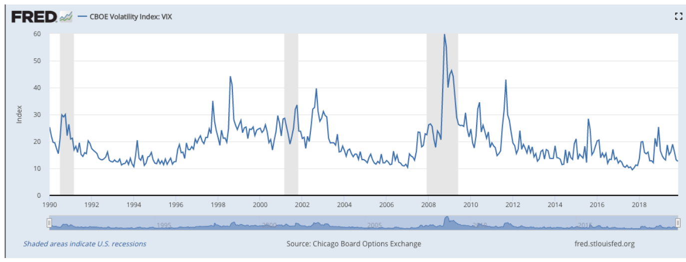

Figure 3: Time series data of the volatility index

The grey bars observed pertain to recessions that the US economy went through, and is typically correlated with sharp spikes in the volatility index. This confirms our intuition that a large volatility index might lead to the easier formation of a financial bubble. Similar data for other variables were procured. This constituted the first phase of data collection, which provided us with intuitive motivation and trends of various variables we considered in our model.

The second phase of data collection was done using the Bloomberg terminal API. Raw data was collected for the variables from 12/31/1989 to 11/21/2019 for 2500 stocks from the New York stock exchange. The data was collected for various industries and was categorized by industry name. Each Industry file contained 9 sheets for the various variables we considered with each column representing a stock and every row representing a day.

Definition 5. ***Time slice*** *We refer to a single row of the data-set as a time slice. It represents a day in real life terms.*

Data cleaning was then performed - however due to complications such as survivorship bias (data was procured for companies that only survived until 2019) and missing data, a lot of the rows were deleted. It was important to delete the entire time slice because removing just the one stock with NA would mean that we skew our model in favor of stocks without NAs in them by giving them more weight (this would necessarily create significant errors when we average quantities over the entire market). This process carried intrinsically with it, another major issue - we were forced to consider a subset of our data-set with as little NAs as possible (to be still able to produce meaningful interpretations). Since missing values were less common in recent times - we sampled and worked with the housing crisis of 2007 to test our model.

In addition to the Bloomberg terminal API, we also used data from the federal reserve to obtain the federal funds rate that we used as the interest rate in our model.

3.1 Implementation

We implemented the model using R. R was able to efficiently handle the data and was very robust and fast compared to python. After reading the data using R, we subset the data so that we were looking at values only in a specified time range. Initially to study the model during times of instability in the market, this interval was set to July 2007 to July 2009. This time frame encompassed a significant part of the housing crisis. Historically, it is known that the bubble ended up crashing in Dec 2007. So, starting at July 2007 gave us ample time to study the behavior of various variables before and after crash.

Since the central aspect of our model is our so called "critical prepice" r^* which is analogous to the critical radius of a physical bubble - we were very interested in observing the trend of r^*. We picked a random stock from our list of stocks to study. Some lists did not show interesting behavior because of a couple of reasons.

- The stocks was too big or too small
- The stock had a lot of NAs or repeated values
- The stock crashed too quickly

This narrowed our search down to a couple interesting stocks. It is worth mentioning that we have not tested out all of the stocks in our list. This subset of "interesting stocks" stems only from a subset of the total stocks. Using data from the subset, we computed r^* for that stock at every time slice. We plotted this alongside r which prepice of the company at every time slice. Discussion of these plots and other results are done in the next section.

Following this micro analysis, we wanted to understand industrial behavior during this time, and understand how the industrial averages behave with respect to our model. Treading into macro side of things, the process involved reading through every company variable at a given time slice, and computing r^* and other necessary variables for each stock and storing it in an array. After doing this - the numbers were averaged to produce the industrial average variables for that time slice. This process was then repeated for every time slice and values at every time slice was stored in respective industry arrays for each variable. This lets us think of the entire industry as a single stock, and thus we could perform the same type of analysis as before. However, in this case, we also compute other things such as n that denotes the number of stocks in the bubble phase at a given time slice, c_0 that gives us the number of companies per dollar are getting into the bubble phase at a given time (In the physical case, it represented the number of bubbles per unit volume) and also V_β the sum of market cap of all stocks in the bubble phase.

After performing the same routine and computing all the additional variables, it was time to fit the industry values to our model to see how well it performed. This presented a lot of new problems. For example, in the physical model $\dot{G}$ was considered to be a constant, but by the financial definition, it is not a constant because the prepice fluctuates very rapidly and any smooth approximation of it will certainly have a non-zero second derivative - especially when the entire market is unstable and various stocks are changing phase. However what we did know is that the prepice values, that scales as $\sqrt[3]{V_j} > 1$, are very small compared to some of the other variables that they are multiplied by; especially really large values like V_j and ΔG^*. So fluctuations in $\frac{dr}{dt}$ do not necessarily affect the model to a huge extent. This gave us confidence to consider $\dot{G} = \overline{\frac{dr}{dt}}$ where $\bar{\cdot}$ denotes the average over time. Another issue that we faced is that since many companies enter and exit the bubble phase over time, and since many small companies never entered the bubble phase - it is expected that the length of the c_0 and n arrays is much lower compared to the length of other arrays such as the market cap which had non-NA value for every point of time. This means, we would have to artificially introduce values into these smaller arrays to be able to bring them to the same dimension as other variables so that we would be able to plot them. We did not want to do this, and thus we averaged out the c_0 and n values and treated them as constants in our model. We are still looking for a more efficient way to work around this issue.

Once we had the model ready, we had to fit our data to some physical model. We found a few research papers that gave us the figures for $\dot{N}$ and $\dot{G}$ for bubbles in liquid sodium. However, we refrained from using these values because fitting our model to this particular theoretical model would mean that we are intrinsically assuming that the abstract surface in which companies bubble has the same properties as that of liquid sodium; and that would be an unjustified assumption. So, we had to find a way to reasonably extract the $\dot{N}$ and $\dot{G}$ values for this hypothetical abstract surface. We did this in the following fashion. Given our financial equation

$$\frac{V_\beta}{V_i} = 1 - e^{-Kt^4} \tag{21}$$

with $K = \frac{\pi}{3}\dot{N}\dot{G}^3$
But from our computations of variables with the financial model, we know exactly what all variables except K was. This allowed us to solve for K. But before solving for K, we divided $\frac{V_\beta}{V_i}$ by the maximum value in that array so that we have a maximum value of 1 and a minimum value of 0 just like the theoretical model. To this modified $\frac{V_\beta}{V_i}$ we solved for K as

$$K = -\frac{\ln(1 - \frac{V_\beta}{V_i})}{t^4} \tag{22}$$

Once again, since the length of $\frac{V_\beta}{V_i}$ as an array was less than the length of t, we only did this only for the t values for which we had a non-NA value for $\frac{V_\beta}{V_i}$. But once again since this didn't encompass the entire time frame and thus had a shorter length than the rest of the arrays, we were faced with the same problem as we did with n and c_0. Faced with the same choices, we decided to average over all the values in K and treat it as a constant. The value of K was reported to be 5.23×10^{-9}

Using this constant we were able to construct our "physical model" for bubbles in this hypothetical surface.

4 Results and Interpretations

Our initial assumption was that whenever $r^* > r$ the stock was not in a bubble and when $r > r^*$, the stock was in a bubble. This is not what we observed. The reason for this is because - r^* value is very sensitive to changes in free cash flow in the market. However, the model does not account for negative free cash flow, and thus there were instances where we got a negative r^* value, which clearly does not make sense. Therefore, we ended up considering the absolute value of free cash flow instead of cash flow itself. This created a discrepancy in our analysis of r vs. r^*. However, this does not mean that the model is not working. We were able to observe many other important results that still provide some meaningful insight into formation and nucleation of market bubbles. What we mean by this is - instead of relative comparison between r and r^*, we can simply look at abrupt changes in r^*. This would still make meaningful sense because - a sudden increase in r^* means that a company just went from a state where it could easily form a bubble to a state where it could not easily form a bubble - most probably indicating a crash. One can observe such a correlation in the data plots. Again, in most cases one observes that a sudden decrease in r^* value correlates well with the company stabilizing itself. This indicates to us that abrupt changes in r^* is still a decent indicator of whether a company/stock is in a bubble or not.

Presented below are the results that we observed upon plotting the "prepice" and the "critical prepice" against time of two different stocks. In addition to this we also overlayed with the stock's trend during the same time range. The latter plot was obtained from Yahoo finance.

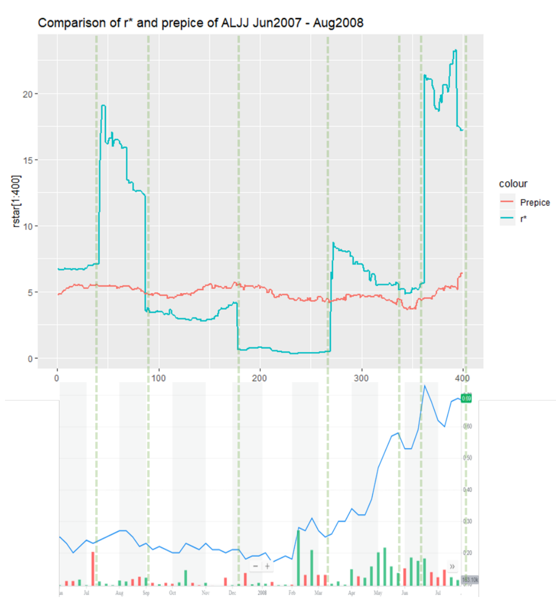

Figure 4: ALJJ from Jun2007 - Aug 2008

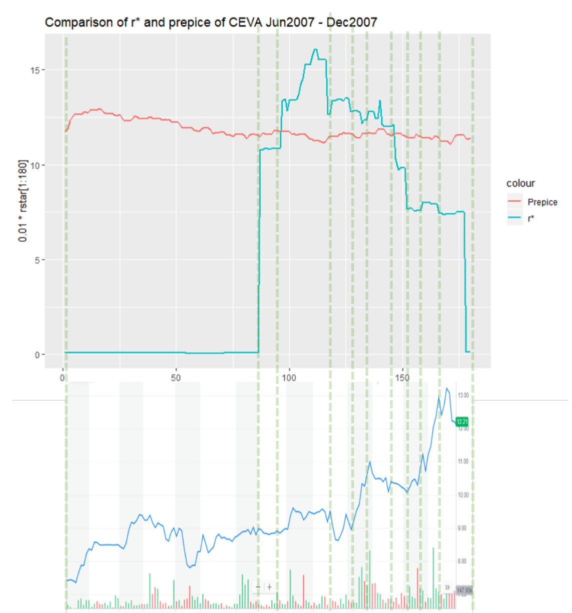

Figure 5: CEVA from Jun 2007 - Dec 2007

As we can see at most of the abrupt change in the r^* value, the stock is either at the top of a mountain and proceeds to go down for some time, or it is in the bottom of a mountain and proceeds to climb up for some time. This directly relates to the change in phase of the stock - where it gets into a bubble or gets out of a bubble. This is by no means a rigorous analysis of the accuracy of the model, but it does provide some significant insights about changes in phase of a certain stock - especially informing that the variables that make up r^* do indeed play a significant role in determining the local minimum and maximum of prices. This can certainly be improved upon with more time and computational resources.

Presented below are more plots of r^* and r for a comparatively bigger company Activision over two stretches of time. We once again observe similar trend in as the other companies. Various spikes in r^* value indicate the relative growth/fall of the company. It is important to know that the last huge spike in the first plot corresponds to the release of the game Call of Duty and thus the company grew by great margins towards the end of that time range. The last spike in the second plot corresponds to the time when Blizzard incorporated (a bigger company) inherited Activision, thus creating a huge bump in Activision's value. This tells us that our analysis is scale invariant

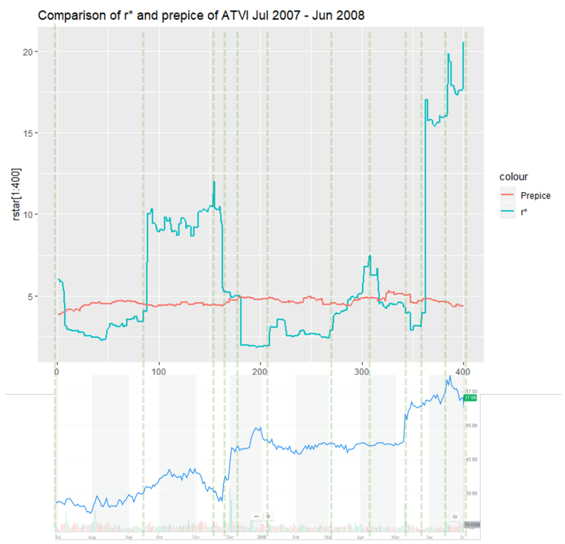

Figure 6: ATVI from Jun 2007 - Aug 2008

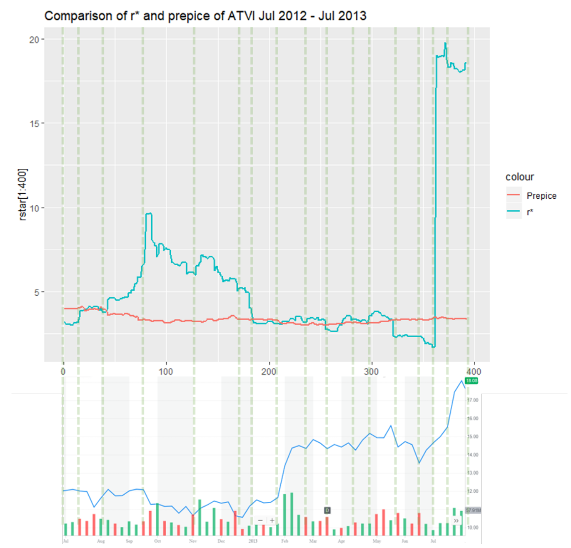

Figure 7: ATVI from Jun 2012 - Dec 2013

We can now plot the variables of the induvidual stocks to the theoretical model in order to see the extent of the fit. It is clear that during the bubble phase - even though there are strong fluctuations, the output of the model is still enveloped strictly by the physical model. This tells us that the trend followed by the model is definitely in line with the physical model. The fluctuations pertain to individual stocks in different phases of the market growth/collapse. During the non-bubble phase, we observe in case of ATVI from 2012 - 2013 (Figure 10) the trend is very oscillatory and does not follow the theoretical model at all. This yields more evidence towards the validity of the model.

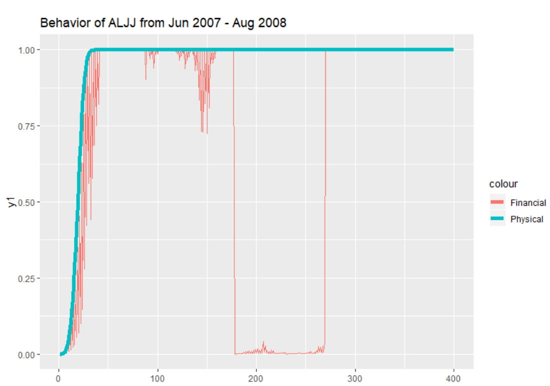

Figure 8: ALJJ from Jun 2007 - Aug 2008

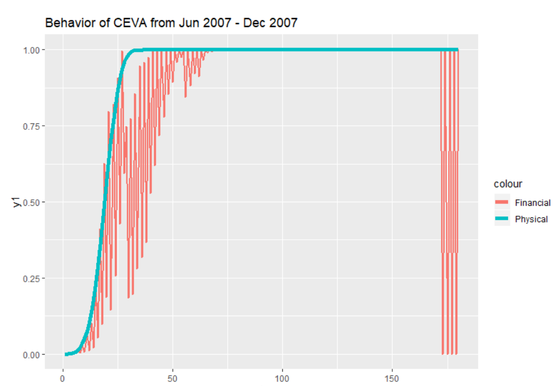

Figure 9: CEVA from Jun 2007 - Dec 2007

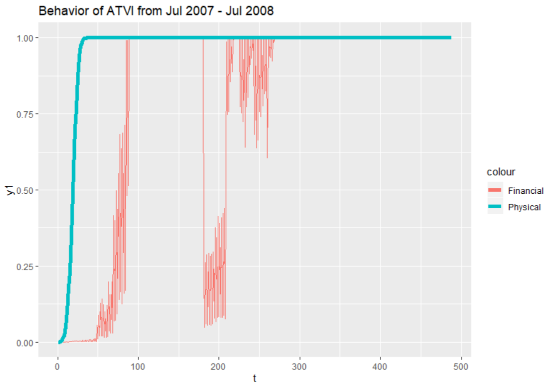

Figure 10: ATVI from Jun2007 - Jul 2008

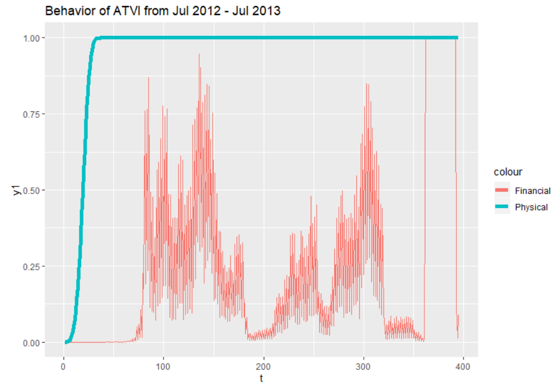

Figure 11: ATVI from Jul 2012 - Jul 2013

After observing the trends and fit in individual stocks, we tried to fit the industrial average to the theoretical model. In order to compare and contrast, we picked two industries - the technology industry and the Industrial entities industry (the choice was random). The industrial trends were computed for two different time frames. One during the actual housing bubble. The other during the last 10 months (From Jan 2019 - Oct 2019). The fits are presented below

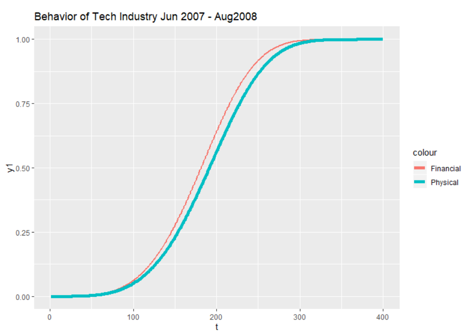

Figure 12: Tech Industry from Jun 2007 - Aug 2008

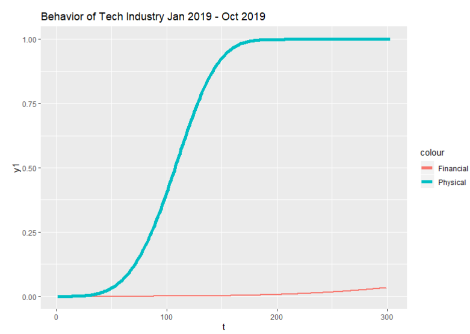

Figure 13: Tech Industry from Jan 2019 - Oct 2019

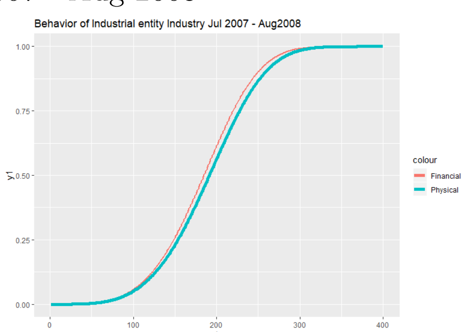

Figure 14: Industrial entities from Jun2007 - Aug 2008

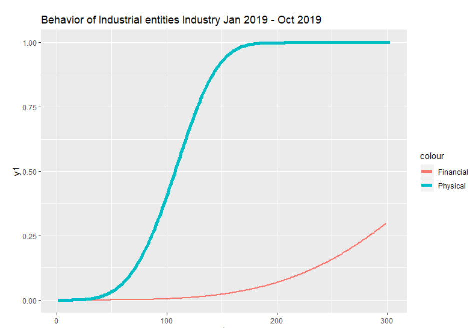

Figure 15: Industrial entities from Jan 2019 - Oct 2019

The total percentage error between the physical and financial model was 8.01% for the Tech industry and 11.2% for the Industrial entities industry. Addressing the speculation given by many economists that we are currently in a bubble, the model does predict that we are slowly entering a bubble. (more info in next section)

5 Discussion and Conclusions

In the course of this project, we were able to understand nucleation theory, and then translate it to economic terms and build an economic model of this physical phenomenon. Even though we did not obtain the results we were expecting to observe; due to computational, theoretical and data based complications; but we were still able to observe meaningful trends in our variables and have a systematic way of quantifying the change of phases of stocks. We also observe that on a macro-scale the model is working and is performing just as expected - this informs us that the errors that we did observe on the micro scale do not seriously hinder the effectiveness of the model. Since the industrial averages do not contain any of the noise that is present in individual stocks, this means that most of the noise that was present in the data was averaged out while considering the industry as a whole. What is more important is that the averaged out values follow the physical model closely. This is crucial for the model because it means that on a macro-scale the model is indeed working as intended.

Another important outcome of the model was that once we were able to fit the data to a certain time frame with a bubble, we had the power to predict if the industry or a certain stock entered or exited a bubble at any given time - we were able to successfully verify this for stocks and industries as well. Using our model, we were also able to predict if our current economy is in a bubble or not - and evidence from the model suggest that we are gradually entering the bubble confirming the speculations of various economists in the US. It is infact observed that the industrial entities industry is entering a bubble faster than the technology industry. This could be motivated by the political climate in the united states and recent changes in national funding for various government sectors (with an increase in funding for sectors such as coal, construction, army - and a cut in IT resources and other aspects of tech industry).

More analysis is definitely needed to understand and tweak the model in the micro-scale; but the results produced in the course of this project provides promise to this endeavour. We might just be scratching the tip of a very large un-turned iceberg; but it definitely signifies that there is meaningful progress to be made.

5.1 Further Directions

Model Side: As was noted before, there are "constants" like c_0, f_0 and n that change through time and this indicates that there exists exogenous variables that are not accounted for. It is definitely in the agenda of the project to redefine r^* to better reflect the micro scale behavior. The model could also be generalized to apply to other markets to act as study cases. We can also correct for survivor-ship bias by redefining some of our variables.
Implementation Side: In the implementation side of things, there is certainly lot more work to be made in the effective cleaning of data. There is a significant amount of work pertaining to tackling the problems of mismatching dimensions - thereby allowing is to avoid a lot of averaging out of significant variables. It is also very important to make this project scalable with data, computational power and time. The R code also needs severe optimization because every single run that was performed for this project took close to 35 minutes to compile fully.

There is certainly more scope for this project and we hope that in the foreseeable future with effort and time, we could further this research in the various areas pointed above.

References

[1] Porter, David A., Kenneth E. Easterling, and Mohamed Sherif. Phase Transformations in Metals and Alloys, (Revised Reprint). CRC press, 2009.

[2] Chicago Board Options Exchange, CBOE Volatility Index: VIX [VIXCLS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/VIXCLS, December 10, 2019.

[3] Board of Governors of the Federal Reserve System (US), Effective Federal Funds Rate [DFF], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DFF, December 10, 2019.

[4] M. Donne, M.P. Ferranti, The growth of vapor bubbles in superheated sodium, International Journal of heat and mass transfer, ISSN: 0017-9310, Vol: 18, Issue: 4, Page: 477-493

[5] Activision Blizzard, Inc. (ATVI). December 10, 2019. Yahoo! Finance. Retrieved from https://finance.yahoo.com/chart/ATVI?p=ATVI

[6] ALJ Regional Holdings, Inc. (ALJJ). December 10, 2019. Yahoo! Finance. Retrieved from https://finance.yahoo.com/chart/ALJJ?p=ALJJ

[7] CEVA, Inc. (CEVA). December 10, 2019. Yahoo! Finance. Retrieved from https://finance.yahoo.com/chart/CEVA?p=CEVA

Appendix

Table 1: Variables

Symbol	**Physics Meaning**	**Economics Parallel**
$\mathbf{Y} = \frac{V_\beta}{V}$	Ratio of the volume of bubbles in a system to the volume of the entire system.	Ratio of the market cap of bubbling assets in an industry to the market cap of the whole industry.
$\dot{N}$	Rate of nucleation (i.e. rate of new bubbles forming in the system).	Rate of companies changing from normal phase to bubble phase.
$\dot{G}$	Rate of radius growth of bubbles in the system.	Rate of "prepice" growth of stocks in the bubble phase.
$V = \frac{4\pi r^3}{3}$	Volume as a function of the radius.	Market cap as a function of the "prepice."
γ	Interfacial free energy; multiplied by area to give surface tension.	Market forces that prevent bubbles from forming.
ΔG_V	Free energy per unit volume	Market forces that cause bubbles to form.
r	Radius of bubble	The "prepice" is a variable proportional to the cubic root of market cap.
r^*	Critical radius; the radius at which a bubble is formed.	Critical prepice; the prepice at which an asset transitions from to the bubble phase.
ΔG^*	Critical free energy; the free energy at which a bubble is formed.	The ratio of balancing markets forces that indicates a transition to the bubble phase.
c_0	Number of atoms in a system per unit volume.	The ratio of companies entering bubble phase and the market cap of an industry.
n	Number of atoms in a system.	Number of companies in bubble phase.
T	Temperature	Variable that measures the macroeconomic environment with the ratio between the market cap of an industry and interest rate.
L_V	Latent heat of fusion per unit volume.	N/A
ΔT	Undercooling	N/A
T_m	Melting temperature.	N/A
V_B	N/A	Book Value
ν	N/A	Volatility (implied or actual)
F_j	N/A	Free Cash Flow
C_j	N/A	Cost of revenue

Table 2: Variables

Variable	Positive or Negative	Correlation Justification	Qualitative Justification
V_i	Positive	$V_i \uparrow \Rightarrow T \uparrow \Rightarrow \dot{N} \uparrow$	An increase in the market cap of an industry should increase the rate of new bubbles forming because there is more total capital to supply each bubble.
V_j	Positive	$V_j \uparrow \Rightarrow \gamma_j \downarrow \Rightarrow \Delta G^* \downarrow \Rightarrow \dot{N} \uparrow$	An increase in the market cap of a company should increase the rate of nucleation because a bubble forms when market cap reaches a critical point.
F_j	Positive	$F_j \uparrow \Rightarrow \Delta G_V \uparrow \Rightarrow \Delta G^* \downarrow \Rightarrow \dot{N} \uparrow$	An increase in free cash flow should increase the rate of nucleation because this is a commonly used metric that is known to inflate the value of a stock.
ν	Positive	$\nu \uparrow \Rightarrow \Delta G_V \uparrow \Rightarrow \Delta G^* \downarrow \Rightarrow \dot{N} \uparrow$	An increase in volatility (both implied and actual) should increase the rate of nucleation because historical data shows that volatility is increased during asset bubbles.
C_j	Negative	$C_j \uparrow \Rightarrow \gamma_j \uparrow \Rightarrow \Delta G^* \uparrow \Rightarrow \dot{N} \downarrow$	An increase in the cost of revenue decreases the rate of nucleation because this metric is hypothesized to deflate the value of a stock.
B_j	Negative	$B_j \uparrow \Rightarrow \gamma_j \uparrow \Rightarrow \Delta G^* \uparrow \Rightarrow \dot{N} \downarrow$	An increase in book value decreases the rate of nucleation because companies with a higher value of tangible assets are less likely to develop into bubbles.