Statistical mechanics of price stabilization and destabilization in financial markets

Jack Sarkissian

Managing Director, Algostox Trading

Abstract

We build theory of market equilibrium by minimizing market's free energy as a function of imbalance. A dark-soliton type of solution is found that describes a gradual transition from positive to negative imbalance as price increases, thus allowing for price uncertainty. Matching the empirical results, sudden security repricing by multiples of bid-ask spread results in a price crash with second-order type phase transition, while small repricing just leads to random walk. Lastly, we show how to use the developed theory to estimate price impact of a submitted order.

1. Order imbalance, execution imbalance and fair value

Finance theory teaches that price forms as a result of equilibrium between supply and demand. In classical theory this equilibrium takes form of the intersection between the supply and demand curves. If one of them moves, the intersection point moves accordingly, resulting in an immediate price shift. Since curves intersect in a single point, price in this model is represented by a single number. That number may fluctuate as a result of fluctuations in supply and demand, but it always remains a single number.

On microstructural level price is not a single number. Before a trade occurs there have to be an offer and a bid, which are generally positioned on different price levels. Only when the two equalize there will be a trade with a definite price, but until then all that can be said about price is that it's localized between bid and ask levels. This price uncertainty is an intrinsic feature of trading process arising due to quantum-chaotic interaction between trading orders [1-5]. It can be diminished by adding liquidity to the market [6], but it cannot be fully eliminated. Price therefore cannot be represented by a simple intersection of two lines. On microstructural level the classical notion of supply-demand equilibrium needs revision.

The main driver in establishing market equilibrium is the order imbalance:

$$J = \frac{N_b - N_a}{N_b + N_a}$$

where N_a is the selling volume and N_b is the buying volume. Order imbalance equals $+1$ in a market without sellers ($N_a = 0$), it equals -1 in a market without buyers ($N_b = 0$), and equals 0 in a balanced market where sellers and buyers are trading equal quantities ($N_a = N_b$).

Order imbalance isn't the only factor responsible for reaching equilibrium. For example, numerous corporate bonds have imbalanced order books for days and months without a single trade, which is why the other important factor is trade urgency. The two factors are combined in the quantity called execution imbalance and introduced in [7] as the difference between probabilities of execution p_a and p_b at ask and bid prices respectively:

$$I = p_a - p_b$$

Obviously, execution imbalance and order imbalance are related: $I = I(J)$. If the entire order imbalance is set to execute in a timeframe τ, then $I = \frac{J}{\tau}$. However, the real connection is more complicated due to multiple order levels, order cancellation, hidden liquidity, and even collective phenomena. Clearly the Taylor series expansion of $I(J)$ must begin with a linear term since $I(J = 0) = 0$ and $I(-J) = -I(J)$. We will use the linear approximation in this paper.

Let us now go further and extend the notion of execution imbalance to allow it to vary with price: $I = I(s)$. This quantity shows what imbalance would be observed if the security were traded at price s. Such extension is important since it now allows to establish the fair value: supply and demand must balance at fair value s_0, so that $I(s_0) = 0$. In this sense phrase "fair value" reflects not only the common sense, but also the definitions of IFRS[1] (International Financial Reporting Standards) [8]. It will be used in this sense throughout the article.

If price were a definite, then $I(s)$ would be a step-function as shown in Fig. 1:

[1] According to IFRS the fair value of an asset may be different from its market value. It takes into account not only market conditions, but also other information, particularly risk.

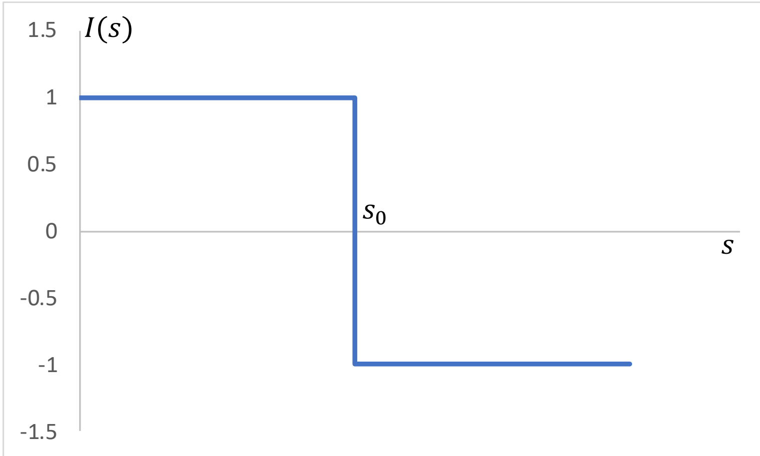

Fig. 1. Execution imbalance when security price is definite.

The meaning of this plot is that (a) if price were definite and (b) its value was known in the market, then any quantity of security offered at price lower than s_0 would be immediately bought, and any quantity bid at price higher than s_0 would be immediately satisfied. Our purpose is to establish equations that allow to determine the functional form of $I(s)$ given the realities of the markets and see if they allow solutions that reflect price uncertainty. To proceed, let us recap facts that we know about execution imbalance.

It was shown in [7] that execution imbalance is related to probability transfer between bid and ask levels. Coupling strength between the two price levels in coupled-wave model is proportional to execution imbalance and vanishes when imbalance is zero. Without coupling execution probabilities remain constant at 50% and 50%.

It was pointed in [7] that less liquid securities allow larger fluctuations of imbalance, while more liquid securities are less tolerant to fluctuations. Large imbalance signals a greater desire for trade, and a liquid market reacts faster. Frequent trading works to reduce imbalance by adjusting the price towards its fair value.

Further, execution imbalance is directly responsible for market crashes, price corrections, rebounds and rallies. We will collectively refer to such events as *critical events*. Critical events were studied experimentally in [9], where it was shown that as they unfold the market undergoes a self-controlled phase transition of the second order type, characterized by sharp buildup of ensemble correlations ρ_e (order parameter) and a spike of ensemble volatility (fluctuations). As example of this can be seen in Fig. 2.

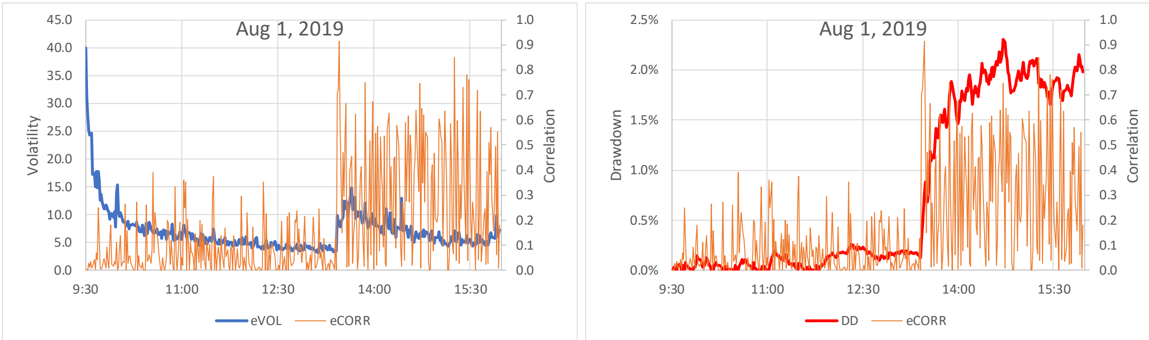

Fig. 2. Impact of drawdown (market drop) on 1-minute ensemble volatility (*e-VOL* index) and correlations (*e-CORR* index) for S&P 500 index

The main cause of market crashes is sudden market repricing, which is represented by the market drawdown D (drop from peak value). Phase transitions are observed only for large price shift ($D > D_c$) and do not occur for a small shift ($D <= D_c$) as demonstrated in [9] and shown schematically in Fig. 3. Repricing by a small amount only results in gradual (adiabatic) adjustment through random walk without rising correlations and volatility.

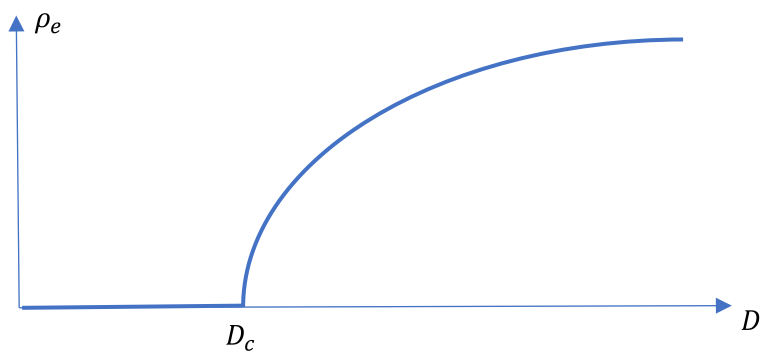

Fig. 3. Order parameter buildup for large drawdown $D > D_c$

To understand what is really happening on diagram Fig. 3, let us calculate the ensemble correlation [10] from the return operator [1-3]:

$$\hat{R} = \frac{1}{2}\begin{pmatrix} \sigma + \xi & \kappa \\ \kappa^* & \sigma - \xi \end{pmatrix}$$

Here σ, ξ and κ are normally distributed random numbers with different standard deviations but zero mean each. Then ensemble correlation equals

$$\rho_e = \frac{\langle r_i r_j \rangle}{\langle r_i^2 \rangle} =\approx \frac{\epsilon^2}{\sigma^2 + \epsilon^2} I^2 \tag{1}$$

where r_i is the return of i-th security, $I^2 \approx \langle I_i I_j \rangle$ is the average execution imbalance and $\epsilon = \sqrt{\xi^2 + \kappa^2}$ is the bid-ask spread. With the use of Eq. (1) correlation in Fig. 3 translates into imbalance $I \sim \pm\sqrt{\rho}$ with two branches for $D > D_c$, showing that for large drawdown (price shift) imbalance experiences bifurcation [11] as demonstrated in Fig. 4:

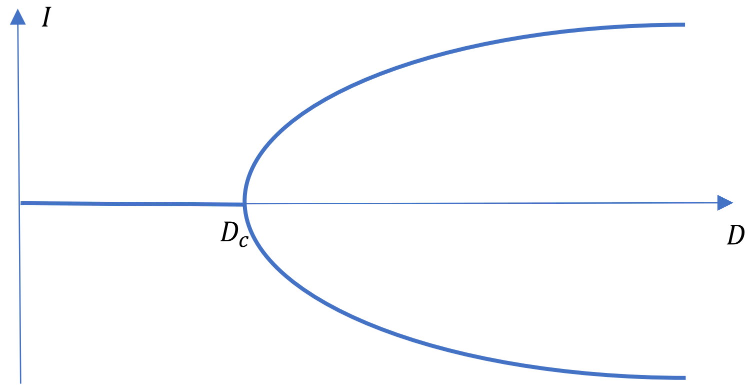

Fig. 4. Bifurcation in imbalance I for large drawdown $D > D_c$

Fig. 4 shows that under normal market conditions the equilibrium value of imbalance is $I = 0$. During the critical events the market has two equilibrium values and takes one of them depending on the direction. In market crashes the zero-imbalance equilibrium is replaced by $I \approx -1$. It is then replaced with $I \approx +1$ during the recovery stage, which however, has no effect on correlation due to the quadratic dependence. The bottom line is that during critical events the markets are in a highly imbalanced state and can flip between positive and negative values without having to go through zero value as can be seen in Fig. 5.

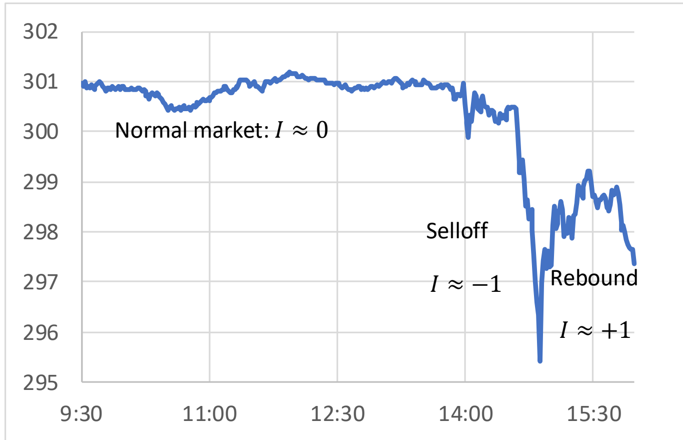

Fig. 5. S&P 500 on July 31, 2019. State with $I \approx -1$ changes to $I \approx +1$ without going through the balanced phase.

Such behavior can be schematically explained with diagram Fig. 1. If security's fair value suddenly shifts, the security will come under pressure with the corresponding imbalance sign. Price will respond to that imbalance by moving towards the new fair value. If price shift is large, pressure will continue for a substantial time. Searching for the new fair value securities can pass it, at which point the sign flipping occurs from $I \approx -1$ and $I \approx +1$. This continues until the market establishes the new fair value. The only question remains, whether it is possible to find the form of $I(s)$ with features similar to the one shown in Fig. 1 but also allowing for price uncertainty. For that, let's compose the free energy of the market as a statistical system.

2. Free energy of the market and equilibrium equation

Considerations of the first section show that imbalance plays primary role in price movements. In fact, it is the most basic factor underlying all other financial processes that influence price movements. In quantum theory of price formation, which relates to fundamental nature of trading, imbalance completely defines the quantum state of the security, considering the fact that phase fluctuates randomly [1-5]. We can therefore say that the free energy of the market[2] is also completely defined by imbalance.

[2] Word market can refer to market in a specific security or the market as a whole.

Considering that $|I| \leq 1$, let us therefore write the expansion of free energy in imbalance and its derivative [12-14]:

$$F = F_I + F_S$$

where

$$F_I = a_1 I + \frac{a_2}{2} I^2 + \frac{a_3}{3} I^3 + \frac{a_4}{4} I^4$$

is the imbalance part and

$$F_S = b_1 \frac{dI}{ds} + \frac{b_2}{2} \left(\frac{dI}{ds}\right)^2$$

is the spread part (we will see later why we chose such name for it). While asymmetry between buying and selling generally exists, it is unessential for the effects studied here. We will therefore drop the odd terms from F_I:

$$a_1 \approx 0, a_3 \approx 0$$

$$F_I \approx \frac{a_2}{2} I^2 + \frac{a_4}{4} I^4$$

Therefore, the total free energy is:

$$F = \frac{a_2}{2} I^2 + \frac{a_4}{4} I^4 + b_1 \frac{dI}{ds} + \frac{b_2}{2} \left(\frac{dI}{ds}\right)^2 \qquad (2)$$

Interplay between F_I and F_S defines energy minimum and the stable distribution of $I(s)$. To ensure stability in I we must have

$$a_4 > 0$$

For $a_2 > 0$ energy F_I reaches minimum at $I = 0$. For $a_2 < 0$ the $I = 0$ becomes unstable as two other minima with lower free energy appear, Fig. 6:

$$I = \pm \sqrt{\frac{-a_2}{a_4}}$$

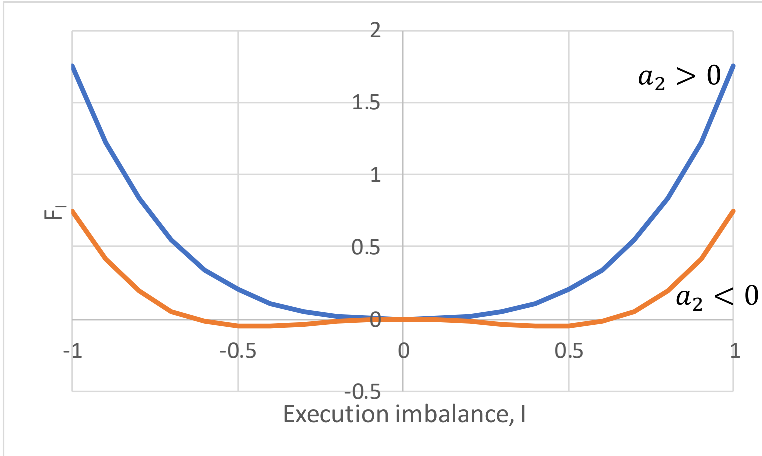

Fig 6. F_I as a function of I for $a_2 > 0$ and $a_2 < 0$

Since phase transitions are observed only for large shifts in fair price, this suggests that coefficient $a_2(s)$ remains positive in an interval around the fair price and changes to negative outside of that interval. We can therefore use the following approximation with positive coefficients:

$$a_2 \approx \alpha_0 - \alpha_2(s - s_0)^2$$

Naturally, since $|I| \leq 1$ in real systems coefficient a_2 is not going to drop below $-a_4$. The stability region is the range where $a_2(s) > 0$:

$$s_0 - \sqrt{\frac{\alpha_0}{\alpha_2}} < s < s_0 + \sqrt{\frac{\alpha_0}{\alpha_2}}$$

Within this region imbalance will fluctuate around zero and price will move in a random walk, Fig. 7. Outside of this region price will come under a strong pressure from imbalance and move directionally towards the region with $a_2 > 0$.

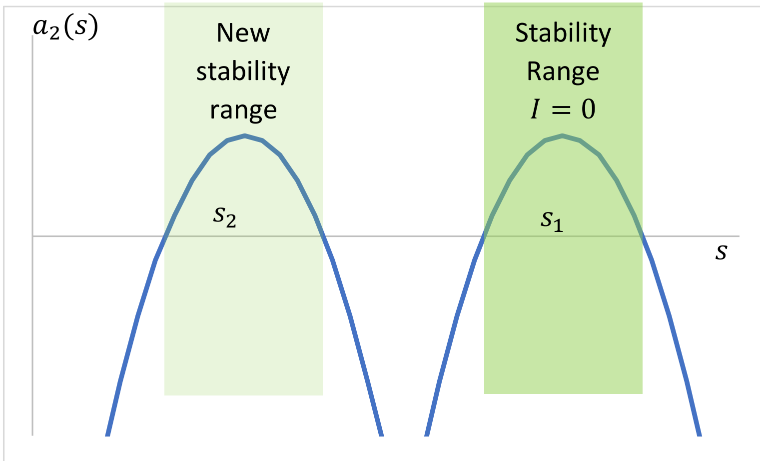

Fig. 7. Shift in coefficient a_2 before and after repricing

Since in finance the only physically valid situation is when $\frac{\partial I}{\partial s} \leq 0$ coefficient b_1 should be positive: $b_1 > 0$. To ensure stability in $\frac{\partial I}{\partial s}$ coefficient b_2 must be positive $b_2 > 0$.

3. Fair value and spread

The general equilibrium equation is reached when [12, 13]:

$$\frac{d}{ds}\frac{\delta F}{\left(\delta \frac{dI}{ds}\right)} - \frac{\delta F}{\delta I} = 0$$

For the free energy described by Eq. (2), we must have

$$a_2 I + a_4 I^3 - b_2 \frac{d^2 I}{ds^2} = 0$$

In this equation a_2 is a function of s. To begin, we will consider it constant and generalize later.

Apart from the trivial solution $I(s) = 0$, this equation has a few other solutions. If $a_2 < 0$ then

$$I(s) = -I_0 \tanh\left(\frac{s - s_0}{\Delta s}\right) \tag{3}$$

where $I_0 = \sqrt{\frac{|a_2|}{a_4}}$ and $\Delta s = \sqrt{\frac{2b_2}{|a_2|}}$. This dark-soliton type of solution is shown in Fig. 8. For $\Delta s \to 0$ it transforms into the step-function shown in Fig. 1.

For $a_2 \geq 0$

$$I(s) = -I_0 \tan\left(\frac{s - s_0}{\Delta s}\right)$$

where $I_0 = \sqrt{\frac{a_2}{a_4}}$ and $\Delta s = \sqrt{\frac{2b_2}{a_2}}$. This solution is periodic and does not satisfy the condition $\frac{dI}{ds} < 0$. It will not realize in real markets. Since the $\tan(x)$ function diverges for arguments close to $\pm\pi/2$ this would lead to formation of numerous stability regions across the price with a discontinuity between them. As a result, the system will choose the trivial solution $I(s) = 0$.

Another, bright-soliton type of solution of Eq. (2) is

$$I(s) = \frac{I_0}{\cosh\left(\frac{s - s_0}{\Delta s}\right)}$$

with $I_0 = \pm\sqrt{\frac{-a_2}{a_4}}$ and $\Delta s = \sqrt{\frac{b_2}{a_2}}$. This solution also violates the condition $\frac{dI}{ds} < 0$.

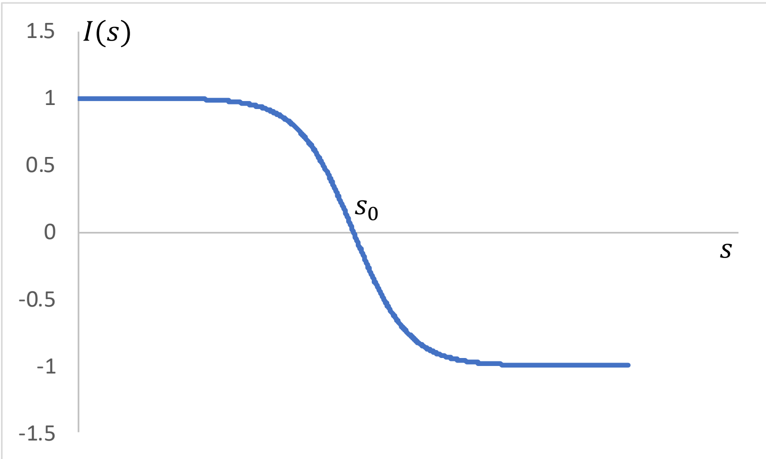

Fig. 8. Graphic representation of solution Eq. (3)

We must conclude that the only physically valid solution is the one provided by Eq. (3). It is similar to the one presented in Fig. 1, but allows for price uncertainty. Apart from the discrete $I = \pm 1$ format, intermediate values of imbalance are allowed, with $I(s \ll s_0) = -1$ and $I(s \gg s_0) \gg 1$, and $I(s_0) = 0$. This is the type of solution that we were looking for.

Note, that solution Eq. (3) is valid for any s_0. Remembering that coefficient a_2 is in fact a function of s, the form of $a_2(s)$ specifies the position of s_0.

Let us consider a model coefficient α_2 with a rectangular profile as shown in Fig. 9:

$$\begin{cases} a_2 > 0, & s_b < s < s_a \\ a_2 = -a_4, & s \le s_b \; and \; s \ge s_a \end{cases}$$

Here s_b represents the bid price and s_a represents the ask price.

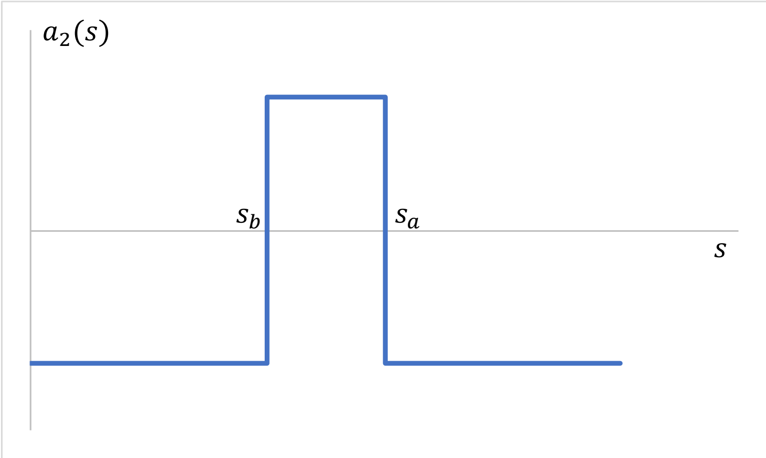

Fig. 9. Coefficient $a_2(s)$ with rectangular profile

For this case the imbalance is:

$$\begin{cases} I(s) = -\tanh\left(\frac{s - s_b}{\Delta s}\right), & s \le s_b \\ I(s) = -\tanh\left(\frac{s - s_a}{\Delta s}\right), & s \ge s_a \\ I(s) = 0, & s_b < s < s_a \end{cases} \quad (4)$$

This solution is shown in Fig. 10. In the limit of $s_b \to s_0$ and $s_a \to s_0$, Eq. (4) will converge to Eq. (3) positioned at the fair value s_0.

Practically, there is no requirement for interval $s_a - s_b$ to be small. It can be larger for illiquid securities, allowing for a larger window of price uncertainty. Let's remember that while spreads in equity markets are as thin as a few basis points ($1\ bp = \frac{1}{100}\ of\ 1\%$), they can reach 100's bp in much less liquid fixed income markets.

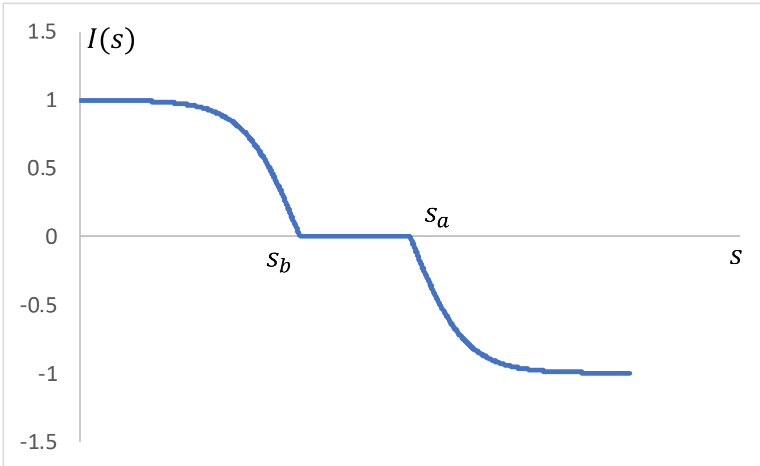

Fig. 10. Graphic representation of solution Eq. (4)

Finally, we can use the Eq. (3) to calculate the spread:

$$\epsilon = 2\,\Delta s = \sqrt{\frac{8b_2}{a_4}}$$

Presence of spread in Fig. 8 as opposed to Fig. 1 is warranted by the F_s part of the free energy. This is why we called it the spread part.

4. Price dynamics after a fair value shift

Now that we have the functional dependence of imbalance on price, we can study price dynamics under the influence of imbalance. Eq. (3) describes a steady state of the market. After a shift of fair value, the new equilibrium will be centered at the new fair value. Transition to the new equilibrium happens gradually and in a manner coordinated with price, so $I = I(s,t)$. Leaving the dynamics of $I(s,t)$ outside of scope of this paper, let's consider price dynamics in a given $I(s)$. Practically, this corresponds to a situation in which a broad announcement is made stating the new security price and leaving little space for speculation. Examples of such situations are cash acquisitions in equity markets and early redemptions in fixed income. These situations do not cause critical events but are nevertheless real.

Suppose that the fair value shifts from s_0' to s_0 by an amount $\delta s_0 = s_0' - s_0$. The dynamic equation for price under the influence of imbalance follows from the quantum coupled-wave theory [1-4] and is

$$\tau \frac{ds}{dt} = \frac{\epsilon}{2} I$$

Remembering that $\epsilon = 2\Delta s$, we get the following equation for the remaining deviation $\delta s = s - s_0$:

$$\sinh\left(\frac{\delta s(t)}{\Delta s}\right) = e^{-\frac{t}{\tau}} \sinh\left(\frac{\delta s_0}{\Delta s}\right)$$

If the price shift is large $|\delta s_0| \gg \Delta s$, price converges to the new fair value linearly:

$$\delta s(t) = \delta s_0 - \Delta s \frac{t}{\tau}$$

This equation means that with each trade price gets closer to the fair value by the amount of semi-spread[3].

For a small price shift $|\delta s_0| \ll \Delta s$ price converges to the new fair value exponentially and within the characteristic trade time τ:

$$\delta s(t) = e^{-\frac{t}{\tau}}\, \delta s_0$$

Exponential convergence works much slower than linear because $|\delta s_0| \ll \Delta s$. The two regimes can be clearly seen in Fig. 11.

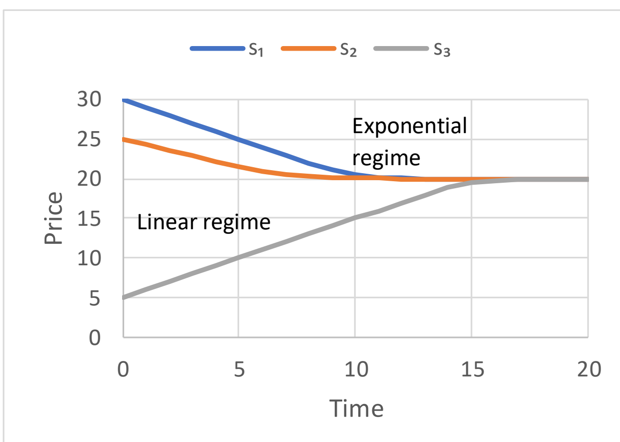

Fig 11. Price convergence to $s_0 = 20$ for $s(t)$ with different parameters δs_0, Δs and τ.

[3] From traders' perspective the entire spread is paid in each transaction. In reality only half is paid, and the perception of entire paid spread arises because of the illusion that the other half is also available.

5. Market impact

All variables in this model are measurable quantities, so one can find the parameters that calibrate the model to market data, then use this information in trade execution and decision making. For example, Eq. (3) can be reversed to express impact of a newly submitted order on price.

If a security is traded at its fair value and has a balanced order book, then submitting an order will create imbalance J. If the order is to be executed within time horizon T then the created execution imbalance $I = \frac{J}{T}$. This imbalance can be balanced by a hypothetical order from the opposite side: $J + J' = 0$. Since such order is missing, the only way for the market to come back to balance is to shift the price by amount:

$$\delta s = \Delta s \operatorname{atanh}(I)$$

Therefore the price impact of an order of size n added to top of the book of total size N will be

$$\delta s = \Delta s\, T \operatorname{atanh}\left(\frac{J}{T}\right) = \Delta s\, T \operatorname{atanh}\left(\frac{n}{NT}\right)$$

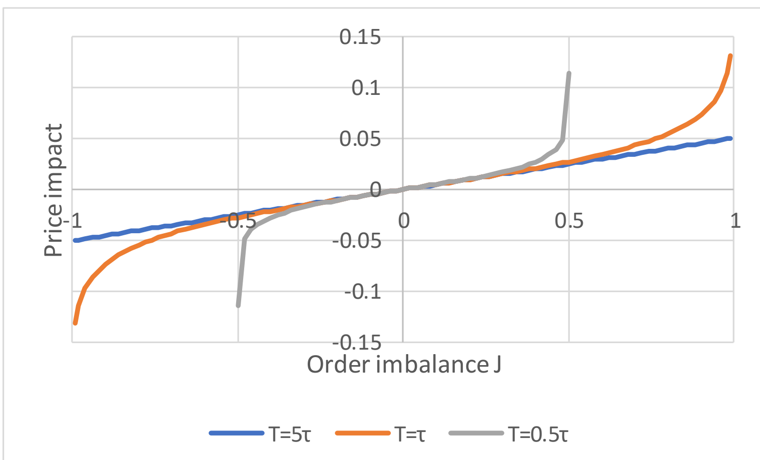

Fig 12. Price impact as a function of order imbalance induced by an order executing uniformly over time horizons $T = 5\tau$, $T = \tau$ and $T = 0.5\tau$.

Fig. 12. shows price impact as a function of order imbalance induced by an order executing uniformly over time horizons $T = 5\tau$, $T = \tau$ and $T = 0.5\tau$. Uniform execution means that the order is split into equally-sized child orders that are submitted in equal time intervals over time T. We see that no matter the time

horizon, small orders execute similarly[4]. The difference arises (a) either for large orders or (b) short time horizon as compared to average trade time.

Large orders will produce smaller impact if they are spread over longer time. Submitting orders so fast that they match all available liquidity ($T = \tau$) will cost approximately the effective spread per order. Submitting orders that not only match all available liquidity but drain it with a residual order leftover ($T < \tau$) will result in order accumulation with amplified negative impact on price. For better execution the residual has to be cleared or at least hidden before more orders can be submitted. Apart from mathematical rigor, these behaviors make plain common sense that also matches the trading realities.

6. Discussion

Question remains that how much impact does one have to produce to initiate a critical event? The answer is that the impact has to take price well outside its spread. How many multiples of spread that impact has to be depends on a security. For example, looking at SPDR S&P 500 ETF Trust (ticker SPY) tracking the S&P 500 index this number is about 25 bp. Given that the spread is 3-4 bp, the required price shift is over 6-8 times multiple. So, 6-8 radical transactions in a security will very likely produce a small crash in this security. Capital requirements are determined by existing liquidity, all of which has to be consumed in a joint effort with all other market participants.

Looking at the whole picture, we can say that we found a quantitative description of the market equilibrium with price uncertainty. These results come as a natural consequence of the laws that govern the markets as a statistical system. Throughout the article we built the theory step by step relying only on experimental/empirical facts and without bringing in individual observations/beliefs. We entirely avoided unquantifiable notions such as “trader mood”, “market sentiment”, “amount of information”, etc., that are so widely used in abstract models. All parameters used in this paper can be computed and calibrated to market data.

The presented theory seems to bypass the quantum description of the markets [1-5]. True, we did not directly derive it form the quantum theory of price formation. Numerous factors are believed to affect prices and imbalance could be just one of them. The fact that market’s free energy can be entirely defined by imbalance is not a trivial realization. It comes only after the fact that the amplitudes of wavefunction, corresponding to bid and ask prices are determined by imbalance while phase is random: $\psi_b = \sqrt{\frac{1-I}{2}}e^{i\phi}$

[4] We of course disregard the commissions/rebates, which is outside of our purpose here.

and $\psi_a = \sqrt{\frac{1+I}{2}} e^{i\phi}$. This is why the free energy can be taken as a function of I without any accompanying parameters.

Although the presented description is quantitative, it is phenomenological. It is unable to provide important characteristics, such as spread statistics, its connection with trading volume and price volatility. It is not therefore viable by itself and compliments the quantum theory by providing the functional form of $I(s)$.

Having provided the description of events that lead to critical events, we did not provide the description of critical events themselves. A significant simplification was that price was set to evolve in a fixed, albeit shifted, imbalance field. While this is a realistic situation, typical for M&As, early redemption of bonds, equity buyback and earnings announcements, these are not critical events. In critical events the field $I(s, t)$ evolves in time in a coordinated manner with $s(t)$. As a result of simplification, important attributes of critical events were missed, for example the oversell/overbuy caused by induced speculative order flow, the subsequent price oscillations, the mechanism of sign flipping in imbalance. All these features have to come up in a more complex kinetic model of critical market events.

6. References

[1] J. Sarkissian, "Quantum Coupled-Wave Theory of Price Formation in Financial Markets: Price Measurement, Dynamics and Ergodicity" (2019), https://ssrn.com/abstract=3459288

[2] J. Sarkissian, "Coupled mode theory of stock price formation" (2013), https://arxiv.org/abs/1312.4622

[3] J. Sarkissian, "Price Measurement in Financial Markets and Quantum Coupled-Wave Model of Price Dynamics", American Physical Society Annual Meeting, (2019)

[4] J. Sarkissian, "Quantum theory of securities price formation in financial markets" (2016), https://ssrn.com/abstract=2765298

[5] D. Orrell, "Quantum Economics: The New Science of Money", Icon Books (2018)

[6] J. Sarkissian, "Spread, volatility, and volume relationship in financial markets and market maker's profit optimization" (2016), http://ssrn.com/abstract=2799798

[7] J. Sarkissian, "Dynamic Coupling in Order Book: Quantum Framework" (2017), https://ssrn.com/abstract=3095073

[8] IFRS 13, "Fair Value Measurement", International Accounting Standards Board (2011)

[9] J. Sarkissian, "Self-controlled phase transitions during market crashes and price corrections" (2019), https://ssrn.com/abstract=3512362

[10] J. Sarkissian, J. Sebold, "Measurement of Current Market Correlations Based on Ensemble Statistics" (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3308669

[11] G.G. Malinetsky, A.B. Potapov, "Modern Problems of Nonlinear Dynamics" (Editorial URSS, Moscow, 2000)

[12] L.D. Landau, E.M. Lifshitz, "Statistical Physics". Vol. 5, 3rd ed., (Butterworth-Heinemann, 1980)

[13] B.Ya. Zel'dovich, N.V. Tabirian, "Orientational optical nonlinearity of liquid crystals", Sov. Phys. Usp. 28 (1985)

[14] V.A. Belyakov, V.E. Dmitrienko, "The blue phase of liquid crystals", Sov. Phys. Usp. 28 (1985)