Journal of Mathematical Economics 1 (1974) 39–49. © North-Holland Publishing Company

ON THE UNSTABLE BEHAVIOUR OF STOCK EXCHANGES

E.C. ZEEMAN*

University of Warwick, Coventry, Warwickshire, England

Received 29 August 1973

Some of the unstable behaviour of stock exchanges can be explained by a model based on catastrophe theory [Thom (1972), Zeeman (1971)]. A similar model can be applied to currencies, property markets, or any market that admits speculators.

The index. The simplest way of measuring the state of the market in a stock exchange is to choose some index, I, such as the Dow–Jones index. Let $J = \dot{I} = \mathrm{d}I/\mathrm{d}t$ denote the rate of change of our chosen index. Then $J = 0$ represents a static market, while $J > 0$ represents a bull market and $J < 0$ a bear market. In the case of a currency, let I denote exchange rate of that currency and again let $J = \dot{I}$. Then $J = 0$ represents a stable currency, while $J > 0$ represents a floating revaluation and $J < 0$ a floating devaluation.

The variable J can be regarded as a dependent variable, depending upon the rate of buying and selling of investors. At the same time there is a feedback because the knowledge of J in turn influences the investors, and this is what makes the dynamic interesting. To express this mathematically we need to introduce variables describing the activity of investors.

Main hypothesis. Broadly speaking there are two types of investors, fundamentalists and chartists. Fundamentalists are investors who act on the basis of estimates of large economic factors such as supply and demand, money supply, etc. Before a fundamentalist invests in a firm, he instructs his research team to assess its viability, its growth potential and market potential, etc. Chartists, on the other hand, are investors who base their investment policy upon the behaviour of the market itself, using the charts of recent behaviour to predict future behaviour. Speculators tend to be chartists.

We take as our main hypothesis that we can divide the investors into these two groups, and represent their activities by two variables C, F as follows. Let C be

*I am indebted to Alan Kirman for advice about some of the economic hypotheses, to Yung-Chow Wong for discussions about the Hong Kong stock exchange, and to Sharon Hintze for the terms fundamentalist and chartist.

the *proportion of the market held by chartists*, or in other words the proportion of speculative money in the market. Let F be the *excess demand for stock by fundamentalists*. In order to keep the model as simple as possible we do not introduce a separate variable for the excess demand by chartists for two reasons. Firstly the total excess demand can probably be represented by some function $f(J)$, and so the excess demand by chartists is merely the difference $f(J)-F$. Secondly the excess demand by chartists is, by definition of chartist, more like part of the internal mechanism of the market, compared with F which is more like an external driving force.

Dynamic flow. We ask the question: how are the variables C, F, J related? Of course the question is too naïve, because they will undoubtedly be affected by many other external factors. However, as a first approximation let us suppose that there is a dynamic relation between them, represented by an ordinary differential equation. If we wanted to add in the effect of external factors, we could, later, superimpose on this differential equation a stochastic noise term.

Take C, F, J are coordinates in 3-dimensional Euclidean space R^3. For intuitive convenience call C and F the horizontal coordinates, and call J the vertical coordinate. The differential equation will be given by a vector field X on R^3. The resulting flow, given by the solution curves of X, will represent the dynamic behaviour of the stock exchange or the currency. It is this flow that we seek to understand qualitatively. Once the flow is understood qualitatively, then it may be possible for economists to use data and build a quantitative model to describe a particular stock exchange or a particular portfolio, or a particular currency, during a particular period. The ultimate objective is prediction, and the design of more effective controls to reduce instability and avoid crashes.

The model. To build up the qualitative picture of the flow, we shall take as hypotheses a number of observed qualitative features of stock exchanges and currencies, and translate each feature into mathematics. The advantage of the translation process is as follows. Whereas it is easy to understand each feature by itself from a local and relatively static point of view, it is more difficult to simultaneously comprehend them all together, and grasp how they interact from a global dynamic point of view. By contrast the mathematical synthesis is relatively easy, because we can use deep theorems. It transpires that all mathematical features can be synthesised into a single mathematical concept, namely *the cusp catastrophe with a slow feedback flow*. This single mathematical entity can be visualised geometrically, and provides an overall grasp of the problem. Not only can the individual features be seen at a glance, but the more complicated patterns of change emerge clearly. Summarising: we insert seven disconnected elementary local hypotheses into the mathematics, and the mathematics then synthesises them for us and hands us back a global dynamic understanding.

Hypothesis 1. J responds to C and F much faster than C and F respond to J. The main purpose of a stock exchange or money market is to act as a nerve centre so that prices (represented in our elementary model by the rate of change J of the index) can respond as swiftly and as sensitively as possible to supply and demand. Changes in C and F can cause changes in J within minutes, whereas changes in J have a much slower feedback on C and F. The response time for C can be a matter of hours, but is more likely to be days or weeks, while the response time for F can be weeks or months, due to the research involved.

What does this mean mathematically? The speed of response at a point p is represented by the length of the vector $X(p)$, and Hypothesis 1 implies that for most points p the vertical component X_J of X is much larger than the two horizontal components X_C and X_F. Consequently the flow lines are nearly vertical almost everywhere. Another way of saying this is that if we fix C and F, then the forces of supply and demand will cause J to rapidly seek a stable equilibrium position $J = J(C, F)$, where $X_J = 0$. Therefore we obtain a surface S of stable equilibria, given by J as a function of C and F (see fig. 3 below). We call S an *attractor* surface because, if the system starts at a point not on S, then the dynamic will carry the point towards S by a fast flow line that is almost vertical. On S itself the vertical component vanishes; therefore as the point approaches near S, the horizontal components, which previously had been relatively unimportant, now become dominant. Therefore on S, or more precisely near S, there will be a slow flow representing the feedback effect of J on C and F.

Mathematical digression. Some mathematically inclined readers may protest at the apparent imprecision of the words fast and slow, large and small, near and far, and so for their benefit we digress in order to make some precise mathematical statements. But first observe that the difference between large and small is a common phenomenon in applied mathematics, for instance a liquid is treated as a continuum in the large and as particles in the small. In our case the difference between weeks and minutes is so large a quantitative difference that it gives rise to a qualitative difference, which we wish to capture in the mathematics. However, it would be wrong to push this difference to a limit as in classical analysis, because the limiting statement would be both irrelevant and probably false. For instance when we say the flow lines are 'nearly vertical almost everywhere' we do not mean that they become *arbitrarily* close to the vertical *sufficiently* far from S. To state what we do mean precisely, it is necessary to use the differential-topological language of the qualitative theory of differential equations rather than the analytical language of the quantitative theory, as follows.

Let $\bar{S}$ be the subset of R^3 given by $X_J = 0$. Then we have a theorem: given any neighbourhood N of $\bar{S}$, there is a diffeomorphism of R^3 into itself, mapping each horizontal plane onto itself, and mapping all the flow lines outside N onto vertical lines. Therefore from the qualitative point of view we may intuitively think of the flow lines outside N as being vertical. Generically $\bar{S}$ will be a smooth surface without boundary, containing S as a subsurface, and meeting each vertical line in at least one and at most a finite number of points (see fig. 3).[1] Moreover $\bar{S}$ will separate R^3 into two regions, one region 'above' $\bar{S}$ given by $X_J < 0$, where the flow outside N is vertically downwards, and the other region 'below' $\bar{S}$ given by $X_J > 0$, where the flow outside N is vertically upwards.

Flows in 3-dimensions featuring this fast and slow response, and consequently possessing an attractor surface containing a slow feedback flow, have been studied in biological contexts [Zeeman (1971)]. Explicit algebraic examples are given there, which are related to the Van der Pol and Lienard equations, and the corresponding flow lines are illustrated in Zeeman (1971, figs. 7 and 9). This ends the digression.

We return to the surface S given by the stock exchange. The question that interests us here is to study the qualitative shape of S and the qualitative properties of the slow flow on S. In particular we shall investigate whether S has any folds or singularities, corresponding to multivaluedness of the function $J = J(F, C)$. If singularities do occur, then we can call upon the powerful classification theorem of Thom [see Thom (1972) and Zeeman (1971)], stating the only way in which they can occur. Summarising: Hypothesis 1 implies the existence of S, and to study the folds of S and the slow flow on S we need further hypotheses.

Hypothesis 2. If C is small then J is a continuous monotonic increasing function of F passing through the origin. In other words if the speculators are in a minority, and the market is dominated by well-informed investors, then an equal demand for buying and selling by the latter will cause the index to be static, $J = 0$; an excess demand will cause the index to rise, $J > 0$, and an excess supply will cause the index to fall, $J < 0$. Therefore the plane C = constant (small) will intersect S in a graph as in fig. 1a. In particular, since the graph goes through the origin, if we fix $F = 0$ then $J = 0$ is a stable equilibrium for the fast response of J.

Hypothesis 3. If C is large this introduces an instability into the market. What does 'instability' mean mathematically? Suppose for the moment that $F = 0$, representing equal demands for buying and selling by the fundamentalists. Then $J = 0$ will again be an equilibrium position for J, but due to the presence of the large proportion of speculative money it will now become an *unstable* equilibrium. In other words, since J is the rate of change of the index, we are postulating in Hypothesis 3 that it is dynamically unstable for the index to remain constant. Any slight perturbation of the index up or down (by external forces) will at once be amplified by the chartists. If the index begins to rise then it will

[1] See footnote on p. 45.

quickly settle into a steadily rising state as a bull market, in other words J will quickly settle into a stable equilibrium position, $J > 0$. Conversely if the index begins to fall then J will quickly settle into a different stable equilibrium position as a bear market, $J < 0$. We emphasise that the word 'stable' here refers only to the fast response of J assuming C and F are kept fixed; if C and F are allowed to vary then the slow feedback flow on S may slowly return J to zero, or may slowly lead to a sudden change of sign of J, because in the long run a bull or bear market may not be able to sustain itself.

The critical consequence of Hypothesis 3 is that for large C and small F *the function* $J(C, F)$ *is* 2-*valued*, and so *the attractor surface* S *is* 2-*sheeted*.

Therefore the plane C = constant (large) will intersect S in the graph shown

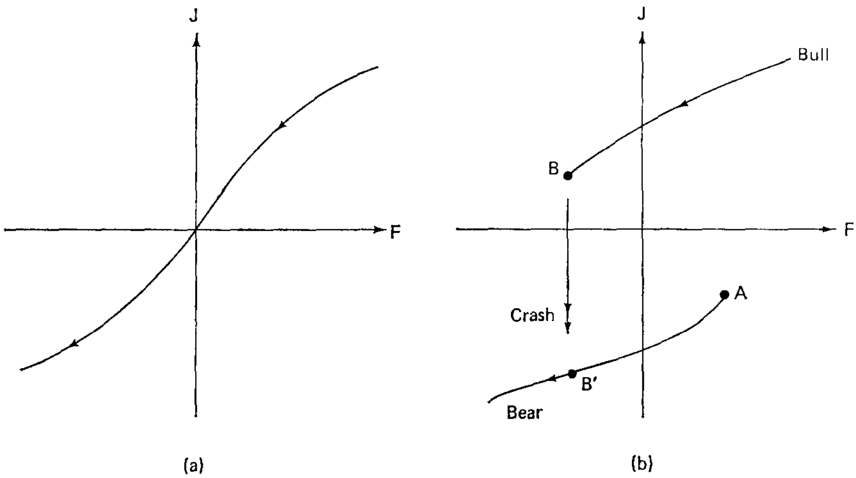

Fig. 1. (a) Small C, (b) large C.

in fig. 1b. Near the origin the graph will consist of the 2 monotonic curves marked bull and bear. If F is large positive, then the demand for buying by fundamentalists will override any external bearish pressures on the market, and guarantee a bull market. Therefore only the bull-graph extends along the positive end of the F-axis, and so the bear-graph must end somehow. Now the bear-graph cannot bend round and go back because this would make S 3-sheeted instead of 2-sheeted, nor can it join onto the bull-graph, because T-shaped junctions cannot occur in a family of sections of a smooth surface. Therefore the bear-graph must have an end point A. Similarly only the bear-graph extends along the negative end of the F-axis, and so the bull graph must have an end point B. Therefore the graph is disconnected.[2]

[2]Alternatively one can use Thom's theorem [Thom (1972), Zeeman (1971)] to deduce the conclusions of the last three sentences from the fact that S is 1-sheeted over some points and 2-sheeted over others.

We can deduce interesting market behaviour as a result of this disconnectedness. Suppose the fundamentalists decide to pull out of a bull market. The resulting slow flow is indicated by arrows on the graphs in fig. 1. If the market contains only a small proportion of speculative money (fig. 1a) then J will slowly fall. Since $\dot{I} = J$, this means that the rising index will gradually flatten as $J \to 0$, and gradually begin to drop as J becomes negative, giving a smooth maximum as in fig. 2a.

If on the other hand the market contains a large proportion of speculative money (fig. 1b), then there will be a delay while the market remains artificially bull until the boundary point B is reached. At this point the vertical fast flow will take over, indicated by the double arrow, causing a sudden drop in J from B to B′ on the bear part of the attractor surface. This will cause a sharp maximum

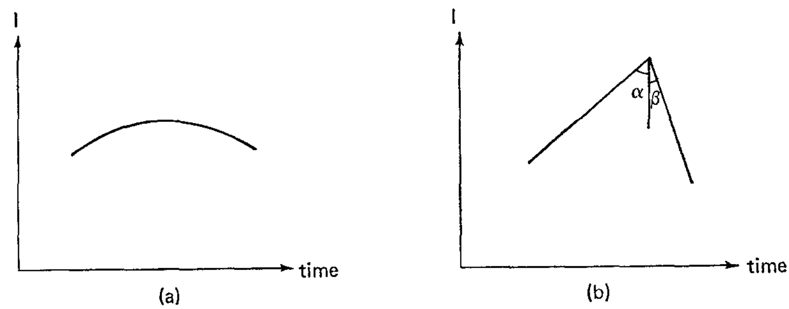

Fig. 2. (a) Small C, (b) large C.

in I as in fig. 2b, with angle $\alpha+\beta$, where $\cot\alpha+\cot\beta = BB'$. Therefore the larger the drop BB′, the smaller the angle β and the steeper the descent of the index. By and large nature encourages us to expect smooth graphs, and so we tend to regard fig. 2a as normal and call fig. 2b a *recession*. If BB′ is very large, causing a very steep descent, then we are liable to call the recession a *crash*.

Conversely fundamentalists often make large profits by quietly investing in a recovering market during the delay period while the market is remaining artificially bear, before it has reached the boundary point A. On the other hand, if the delay is exaggerated by a very high proportion of speculative money, then the fundamentalists' support may be insufficient to reach the boundary point A. This phenomenon occurs for example in a run against a reserve currency, when in spite of positive support by central banks, there is too much speculative money for the support to be effective, and so the currency may have to be devalued, even though in terms of its purchasing power it may then seem to be undervalued.

Synthesis. We now use the deep classification theorem of Thom [see Thom (1972) and Zeeman (1971)] to synthesise the information acquired so far into a

3-dimensional picture of the surface S. The theorem implies amongst other things that:

(i) S is a subsurface of a generic[3] surface $\bar{S}$;

(ii) a generic surface is smooth without boundary;

(iii) when a generic surface is projected orthogonally onto the (C, F)-plane, the only singularities that can occur are fold curves and cusp points;

(iv) the boundary of S equals the fold curves of $\bar{S}$.

Since S has sections as in fig. 1, we can deduce from the theorem that $\bar{S}$ must have a cusp point and therefore must be equivalent[4] to the cusp catastrophe surface, shown in fig. 3.

The cusp catastrophe is described in Thom (1972) and Zeeman (1971) and we now review its properties. The surface $\bar{S}$ shown in fig. 3 is given by the equation

$$J^3-(C-C_0)J-F = 0.$$

Here C_0 is the value of C at which the discontinuity shown in fig. 1b first begins. The surface S is the subsurface of $\bar{S}$ given by the inequality

$$3J^2+C_0 \geqq C.$$

The boundary ∂S of S is the fold curve of $\bar{S}$, given by $3J^2+C_0 = C$. The projection of ∂S onto the (C, F)-plane is the cusp shown in fig. 3, and has the equation

$$4(C-C_0)^3 = 27F^2.$$

Over the outside of the cusp S is single-sheeted and is the same as $\bar{S}$. Over the inside of the cusp S is 2-sheeted and $\bar{S}$ is 3-sheeted, the extra middle sheet being the complement $\bar{S}-S$. The complement $\bar{S}-S$ is given by $3J^2+C_0 < C$, and represents points of unstable equilibrium; in other words it is a *repellor* surface, the opposite of an attractor surface, and it is shown shaded in fig. 3. It follows from the theorem that the attractor surface S and repellor surface $\bar{S}-S$ together form the smooth surface $\bar{S}$, and are separated by the boundary ∂S. Although $\bar{S}$ is mathematically interesting it is irrelevant from the point of view of the application under consideration, because the system stays only on the attractor subsurface S.

[3]*Generic* is a technical mathematical term: generic surfaces arise from generic systems. Here we define X to be generic if $\int_0^r X_J \, dJ$, $r \in R$, regarded as a map from the (C, F)-plane into $C^\infty(R)$, is transversal to the natural stratification of $C^\infty(R)$. Transversal maps are open dense in the space of all C^∞-maps, endowed with the Whitney C^∞-topology. Therefore any system can be approximated arbitrarily closely by a generic system, and so we are justified in taking as our model a generic surface.

[4]*Equivalent* means the following: two surfaces S_1, S_2 are equivalent if there is a diffeomorphism of R^3 onto itself throwing S_1 onto S_2, and throwing vertical lines to vertical lines. In particular we can incorporate any required scaling of C, F, J into the diffeomorphism.

D

We now investigate the slow flow on S, and for this we need four more hypotheses describing the feedback effect of J on C and F. The dotted flow lines shown in fig. 3 will be the consequence of these hypotheses.

Recall that $J = \dot{I}$, where $\dot{I}$ denotes the rate of change of the index I.

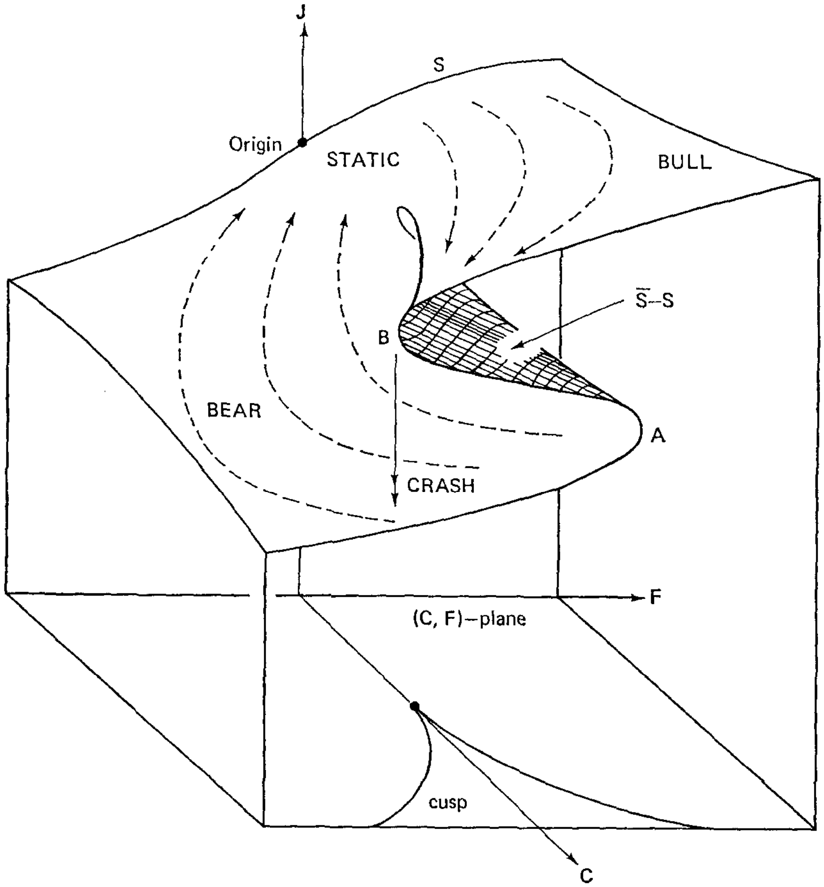

Fig. 3. Note: For convenience we have drawn the (C, F)-plane not through the origin $J = 0$, but below the origin as $J = k$, where k is a negative constant.

Hypothesis 4. *$\dot{C}$ has the same sign as J.* In other words chartists follow the trend. A bull market attracts the chartists, and so the proportion of speculative money in the market goes up, while a bear market repels chartists, and so the proportion of speculative money goes down.

Hypothesis 5. $\dot{F} < 0$ *after a large rise in I, even though I may still be rising.* In other words fundamentalists, with the experience that a bull market does not last indefinitely, and the knowledge of the industrial capacity behind the market, tend to judge when the market has become overvalued, and tend to cash in while the index is still rising before it has reached its maximum. Therefore if the system has been following a slow curve during which J has been positive and increasing, then $\dot{F}$ will become negative.

Hypothesis 6. $\dot{F} < 0$ *after a short fall in I.* Fundamentalists often pursue a policy of choosing a margin for each stock, and then selling if there is a fall in price of magnitude greater than that margin. In other words the selling price is the maximum price to date minus the margin. This policy of cutting losses would give rise to Hypothesis 6. In particular if J is large negative then $\dot{F} < 0$.

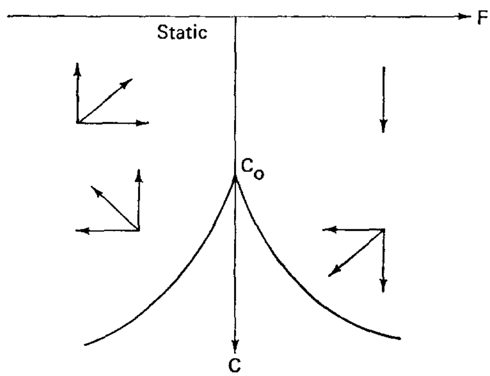

Fig. 4

Hypothesis 7. $\dot{F} > 0$ *if I has been falling for some time and is beginning to flatten out.* In other words a recovering bear market is an attractive investment for fundamentalists because prices are liable to be undervalued, and small losses in the short run are liable to be offset by much greater gains in the long run. In terms of J this occurs when $J < 0$ and $\dot{J} > 0$.

The slow flow on S. Fig. 4 shows the changes in C and F implied by Hypotheses 4–7 using the values of J given by S. The sums of components indicate the direction of the slow flow vectors. The resulting slow flow lines on S are shown dotted in fig. 3.

Global dynamics. We can now deduce a global dynamic behaviour of a stock exchange. If there are no external forces the market will settle into a stable and relatively static equilibrium. Suppose now there is a stochastic noise in the form of an influx of fundamentalist money. This frequently happens to a market which has been static for some time, and often is caused by fundamentalists moving their money away from another market which is experiencing a period of instability, into the stock exchange under consideration, because, due to the latter's recent spell of stability, the fundamentalists have had time to research into its potentialities. The system then becomes caught in the slow flow with the following sequence of consequences. First the rising index attracts the chartists, and the bull market accelerates. If the proportion of chartist money becomes too high then the fundamentalists may begin to cash in. If sufficiently many cash in, then after some delay a recession will occur, causing chartists to withdraw. Finally there is a slow recovery as the fundamentalists begin to reinvest.

The higher the proportion of speculative money before the recession, the longer the delay and the steeper the slope of the recession (the slope $=\cot\beta$ in fig. 2b). If C is very large the slope will be very steep, and the recession will be called a crash. Recall that $C = C_0$ gives the position of the cusp point. As a corollary of Thom's theorem, for $C \geqq C_0$ we have the formulae:

$$\text{delay} = K(C-C_0)^{\frac{3}{2}},$$
$$\text{slope of recession} = L(C-C_0)^{\frac{1}{2}},$$

where K, L are non-vanishing smooth functions of C. In the case of the cusp catastrophe the functions K, L are constant $K = 2/3\sqrt{3}$, $L = 1/\sqrt{3}$ but in general the surface S is only equivalent to, rather than the same as, that of the cusp catastrophe, and so K, L may not be constant. Nevertheless since K, L do not vanish at C_0, we can as a first approximation take them to be constant near C_0, giving surprisingly precise formulae that might be useful for testing quantitatively. For instance near the cusp $\cot\beta = 2\cot\alpha$ (see fig. 2).

Of course the above sequence of events could be slowed down or accelerated, or interrupted, by stochastic noise caused by external forces. If the noise is large then its unpredictability may obscure the underlying dynamic, giving the impression that a stock exchange is only stochastic.

If a government wanted to introduce controls to prevent a runaway bull market from crashing, it could impose a temporary heavy tax on the sale of stock proportional to length of time held. This would discourage selling in general, and in particular might persuade fundamentalists to hold on to their long-term investments during the tax period, while the proportion of speculative money slowly subsided. If the model was correct, this would steer the market back to normality by a continuous rather than a discontinuous path.

Note added in proof: A similar economic model, also based on the cusp catastrophe, has been developed by the anthropologist Michael Thompson, to explain the periodic cycles of credit confidence underlying the complex pig-giving ceremonies in the New Guinea Highlands.

References

Thom, R., 1972, Stabilité structurelle et morphogénèse (Benjamin, New York).
Zeeman, E.C., 1971, Differential equations for the heartbeat and nerve impulse, in: C.H. Waddington, Towards a theoretical biology, vol. 4 (E.U.P.) 8–67.