Positive feedback from the market: network externalities

Jan A Audestad

1. Market model

A model is not more than a model. A market model describes an idealisation and simplification of the real market – the model does not describe any concrete situation or activity within the market accurately. This is, of course, evident but nevertheless traditional economic theories and models are more often than not employed in order to state absolute facts about the market.

The theories of free markets, the effects of free competition, the theory of monopolies, the law of diminishing return, and so on are just models that may describe the general behaviour of particular types of market subject to some idealised assumptions. These theories are often concerned with determining stationary and stable outcomes of competition. Most of the traditional models do not take into account nonlinear and irreversible behaviours of the market.

This paper describes models (which again is not more than a model that may indicate the general market behaviour in certain cases) based on positive feedback or increasing return. These models were developed during the late 1980s but it took a long time before the models were actually accepted in the broad economic community.

See the paper by Brian Arthur published in Scientific American in 1990 for a simple discussion of positive feedback in economy.

A simple mathematical model for the *dynamics* of markets with positive feedback is derived below. In dynamic models we are concerned with how the system evolves as a function of time, for example, how the system evolves toward equilibrium, and not with equilibrium conditions themselves as determined by demand curves and utility functions.

It is hard to describe dynamic models without presenting them in mathematical form. The behaviour of the market is nonlinear and cannot be properly described using ordinary linear economic theory and models based on simple demand curves. The behaviour is not intuitive and simple reasoning may easily lead to wrong conclusions.

Feedback in a system (economic, technical, social, biologic etc.) means that the output signal from the system (in economic systems, for example, price or market share) is fed back to the input of the system. There are two types of feedback: *negative* and *positive*. By *negative feedback* we mean that the feedback signal counteracts any deviation of the output of the system from its equilibrium value: if the output signal increases (e.g., the price goes up), the negative feedback forces the output signal back (i.e., reduces the price) to its equilibrium value; if the output signal decreases, the feedback again forces the signal back to equilibrium. Systems with negative feedback are always stationary; that is, after an initial transient period, the output signal is a constant function of time. Negative feedback exists in all natural and technical control systems (cybernetic systems).

Positive feedback means that any change in the output signal is amplified. If the output signal increases, the feedback forces it to increase further. Similarly, if the output signal decreases, the feedback forces it to decrease further. The destiny of systems with positive feedback may thus be that the output signal approaches a fixed limit (or saturates). In some systems, the positive feedback will cause the output signal to start oscillating (periodically or chaotically) between given values. Which behaviour the system will follow (saturation or oscillation), depends on the nonlinearities and the delay in the feedback system.

We shall only be concerned with systems that saturates; that is, reaches an upper limit.

The models below are derived for products of which the buyer only buys one copy during the whole lifetime of the product. Do such products exist and are they important?

The answer is yes. There are many such products. Telecommunications subscription is one of them. We only subscribe to a single copy of a newspaper; we receive electricity from a single producer; we usually only buy a single copy of a book, film or music record; and we have usually one or just a few insurance policies.

Sometimes we buy more than one copy of the product (for example, telecommunications to the home and to the cottage). However, again we may argue that there is an upper bound on the number of copies being sold in a population of a certain size so that the simple model still describes the market with sufficient accuracy. At least, the model enables us to understand the basic behaviour of such markets.

2. Mathematical model of a market without feedback

Let us start with a mathematical description of the evolution of a market without feedback. This makes it easier to understand than the more complex models where positive feedback is introduced.

Let there be N customers likely to buy one copy each of the product where N is independent of time; that is, the size of the market is independent of time. The number of products sold at time t is $S(t)$ where $S(t)$ is then a number between 0 and N. Since N is a constant we can get rid of it in the calculation that follows by *normalising* the market by introducing the new *normalised* variable

$$s(t) = S(t)/N$$

Then $s(t)$ is a number between 0 an 1 where the upper limit corresponds to the case where everyone have purchased one copy of the product. The relative number of persons having purchased the good is then $s(t)$ and the relative number of persons who are still likely to buy the product is $1 - s(t)$. The result for a population of N at time t is then found by simply multiplying $s(t)$ and $1 - s(t)$ by N.

In an ordinary market, the number of copies sold per unit time is proportional to the number of people not having purchased a copy already:

$$ds(t)/dt = a(1 - s(t)) \qquad (1)$$

where a is a constant of proportionality. The value of a depends on price, utility and other economic factors. The derivative $ds(t)/dt$ is the relative number of people purchasing a copy during the infinitesimal time interval dt. Since we are concerned with the temporal behaviour (or dynamics) of the market, we may also view $ds(t)/dt$ as a velocity, namely the speed by which the market evolves as a function of time.

Equation (1) is just the common linear model of a market.

We could, of course, used a discrete model since there will always be an integer number of customers buying the product during any time interval. For some people this is a more intuitive model but since we are considering large markets it is simpler to treat the problem as a continuous problem from a mathematical viewpoint. This produces no significant error in the result.

The solution of this differential equation is found simply by separation of the variables:

$$\frac{ds(t)}{1-s(t)} = a\,dt$$

Integrating both sides of the equation then gives:

$$-\ln(1 - s(t)) = at + C$$

where C is a constant of integration. Solving for $s(t)$ we get:

$$s(t) = 1 - ce^{-at}$$

where $c = e^{-C}$ is another representation of the constant of integration. We see that the size of the initial market is $s(0) = s_0 = 1 - c$ or $c = 1 - s_0$. Finally we then find:

$$s(t) = 1 - (1 - s_0)\,e^{-at} \tag{2}$$

The market behaviour is as shown in Figure 1.

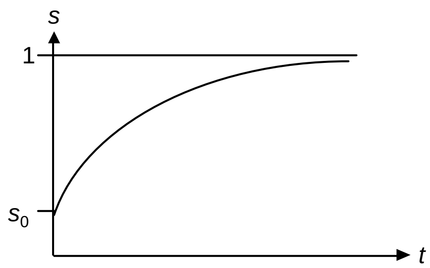

Figure 1 Market without feedback

Observe that this market starts growing even if there are no customers initially ($s_0 = 0$). The market then evolves as follows:

$$s(t) = 1 - e^{-at} \tag{3}$$

This is an important observation because it is a property that is not shared by all markets as we shall see in the next section.

3. Market with positive feedback

Positive feedback from the market means that the number of customers likely to purchase a product also depends on how many customers that have already purchased the product. There are many products like this, for example, the attractiveness of a restaurant for a hungry person passing by is likely to depend on the number of customers already dining in the restaurant; the initial growth of internet subscriptions was driven by positive feedback – the more people being connected the more attractive is the service; certain branches such are car retailers, antiquarian bookshops and antique dealers are likely to lump together in the same area. In Paris, most of the shoes are sold in one street, and almost all antique dealers are lumped together in the same building.

Sometimes this type of feedback is called *network externality* or *network effect* because the feedback is caused by a formal or informal network among the customers. Such networks may be based on acquaintance where the buyer may seek advice from other people before buying the product of a particular brand. Furthermore, the benefit the customer finds in owning the product increases with the number of people already owning it. Telecommunication products belong to this class of market.

First we shall look at a market where there is only one provider of the goods (a monopoly) in order to discover the basic behaviour caused by positive feedback. In the next section, the behaviour of a market with two providers of the same good is analysed. The analysis in this and the next section is simple and does not describe the complex performance of real markets. However, the analysis uncovers some properties that are common to all markets with positive feedback.

Let us then analyse a very simple model where the feedback from the market is given by the relative number of customers $s(t)$ having purchased the good at time t. In other words,

$$ds(t)/dt = b\ s(t)(1 - s(t)) \tag{4}$$

where b is a constant of proportionality (the market coupling). The factor $1 - s(t)$ on the right-hand side of the equation is just the relative number of people that have not bought the good at time t. The factor $s(t)$ is the relative number of people having bought the good and represents the feedback from the market. In other words, the number of new customers in time dt is proportional to both the number of people not having purchased the good (normal market behaviour) and the number of people having bought the good (feedback). Observe again that $ds(t)/dt$ is the velocity by which the market is evolving with time.

First observe that the behaviour of this type of market is nonlinear (in this case a square function in $s(t)$). All markets with positive feedback are nonlinear. This makes the analysis of this type of market mathematically challenging. The market described by the above equation is one of the few models that can be solved by simple integration of the differential equation. This is one reason why just this model was chosen for the mathematical analysis. Another reason is, of course, that the model seems to offer an intuitively plausible description of a market with positive feedback.

The differential equation can be solved by simple analytic methods. Again the variables can be separated as follows:

$$\frac{ds(t)}{s(t)(1-s(t))} = b dt$$

The left-hand side of the equation can be written as a sum of two terms:

$$\frac{ds(t)}{s(t)} + \frac{ds(t)}{1-s(t)} = b dt$$

The solution is found by integrating term by term

$$\ln s(t) - \ln (1 - s(t)) = bt + A \text{ or } \ln \frac{s(t)}{1-s(t)} = bt + A \text{ or } \frac{s(t)}{1-s(t)} = c_0 e^{bt}$$

where A is the constant of integration and $c_0 = e^A$. This gives

$$s(t) = \frac{c_0 e^{bt}}{1 + c_0 e^{bt}}$$

This can be written as

$$s(t) = \frac{s_0}{s_0 + (1 - s_0)e^{-bt}} \tag{5}$$

where $s_0 = s(0)$ is the initial market size and, as is easily seen, $c_0 = s_0 /(1 - s_0)$.

This function is shown in Figure 2. The curve represents the evolution with time of the modelled market with positive feedback. Note that the dynamics of the market is such that $s(t)$

? 1 as t ? 8; that is, the final state of the market is such that everyone owns a copy of the product. Note also that $s(t) = 1$ for all t is the solution of the differential equation with $s_0 = 1$.

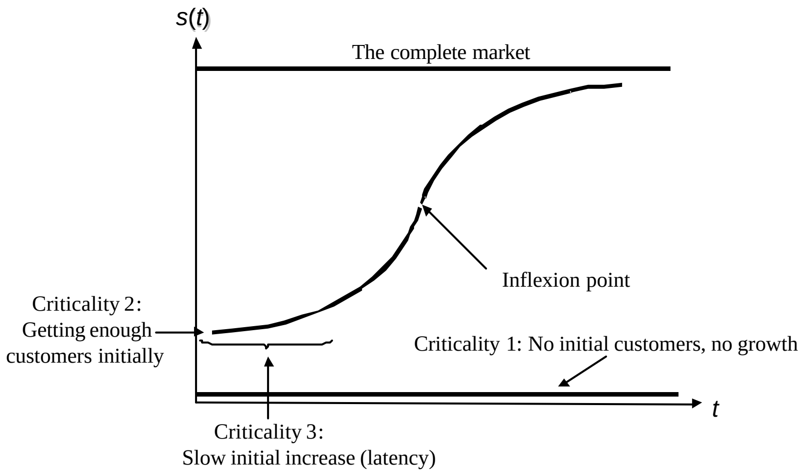

Figure 2 Market with positive feedback

Criticalities 1, 2 and 3 summarises the three most important observations we can make from this model. Each of them represents a strategic challenge. These observations are true for many other nonlinear and more complex market models.

Critically 1 states that *if there are no initial customers having purchased the good, there will be no customers in the future*. This is so because the solution of the equation is $s(t) = 0$ if $s_0 = 0$. We also see immediately that this equilibrium solution is unstable since any perturbation (that is, any arbitrarily small deviation of s_0 from 0) will cause $s(t)$ to increase since $ds/dt > 0$ for $t > 0$ if $s_0 > 0$.

This means that in a market with positive feedback we must first establish a customer base (s_0) before the product is launched. This is referred to as Criticality 2 in the figure. The strategic challenge is then to determine how this can be done and how big s_0 should be. One alternative is to give away for free the first few copies of the product, for example, launching a trial service in order to monitor the response from the market. This is often done with technically advanced products that are likely to become popular within small groups of techno-freaks. Another approach can be to bind the product to a complementary product that is not subject to network externalities.

Criticality 3 indicates that even with initial customers, the market first increases very slowly. This may be regarded as a latency period. The length of the latency period depends on the constant of proportionality (the market coupling constant b) and the number of initial customers s_0. The strategic challenge is then to be patient enough and not terminate the product too early. This is, of course, a real challenge, particularly if the strategy is such that products must show positive return after a short time. It is therefore necessary to determine whether a product will be subject to positive feedback from the market before it is launched in order to take the right decision.

The long latency period is believed to be common for a large number of markets with positive feedback (see C. Shapiro and H. R. Varian, *Information Rules: A Strategic Guide to the Network Economy*, Havard Business School Press, 1999).

The inflexion point may also be regarded as a criticality. At the inflexion point the gradient (ds/dt) in market increase takes its maximum value. Beyond the inflexion point the relative increase in the market size diminishes. By some decision makers this may be regarded as a strategic challenge since it signals that something negative starts happening with the product, namely that the popularity of the product becomes gradually smaller. The reason for the decline in demand is a natural consequence of the dynamics of the market and has nothing to with attractiveness (the size of b).

However, note that this is just a simple model. The reality is much more complex as will be discussed in Section 5 below.

Examples of telecommunications products with network externality are:

- The facsimile service that was launched in the early 1980s depends on how many facsimile machines exist and should therefore be subject to strong network effects (the initial market size disadvantage). On the contrary, this service developed rapidly apparently with virtually no latency period.

  The facsimile service had already existed for almost one hundred years but was not taken into use (except by newspapers (the telephoto service)) because there were only modest need for the service in industry and government. Furthermore, standards for a common information transfer protocol (except for telephoto) did not exist. What happened around 1980 was that a common international standard was developed by the International Telecommunications Union, the globalisation of the industry had just started, office automation had made enormous advances, and there was a growing need for electronic transfer of documents within the businesses. On the other hand, there was little need for transfer of documents between different companies. Therefore, the externality was not strong at all. Facsimile transmission took first place within companies so that an initial set of users had already been established when it became evident that facsimile transmission between different companies and enterprises also was an enormous advantage in the new business regimes that had just been created.
- Similarly, the SMS service had a latency period of more than five years. However, when someone found out that this was a simple means for sending messages, the set of initial users was in fact everyone owning a mobile phone. Therefore, the rise of the service was enormous.
- The introduction of IP version 6 has been delayed because of network effects. The advantage of IPv6 over IPv4 is not very big in the current internet because of large overcapacity. Furthermore, an IPv6 network owned by a single operator is almost without value. The problem is then how to get started; that is, the problem is related to criticalities 1 and 2. The tunnelling of IPv6 datagrams through IPv4 networks may offer one solution to this dilemma. This means that considerable advantage may exist even for a few IPv6 networks since they may be interconnected by tunnelling.
- The picturephone is an example of a product subject to network effects that never took off at a large scale. Apparently the coupling to the market (the advantage of the product as judged by the users) was too small. Few people and companies saw big advantages in a service where the advantage over telephony was just to show the picture of the communicating persons at a much higher price than for a telephone call. Web camera on PCs and multimedia in the IP network may reintroduce the picturephone; but then we can hardly justify to call the new service picturephone?

The value of a network depends on its size.

Sarnoff's law states the obvious fact that the value of a broadcast network is proportional to the number of viewers (law of linearity). Metcalf's law states that the value of a communications network is quadratic in the number of subscribers in the network (or more precisely, $N(N-1) = N^2$ if N is large) since every member will have the benefit of interconnecting with $N-1$ other members pairwisely. However, Reed's law is even bolder stating that the value of the network is 2^N (i.e., exponentially) since this is the number of groups that can be formed in the network. The capability of forming groups with different purposes and scopes may be the most important aspect of some networks. Formation of groups is, for example, one of the most important events taking place in he Internet(e.g., Facebook). Note that N for the Internet (more than one billion people) is so big that these laws represent no practical measure concerning the value of Internet but provides a clue to what size may mean to a network depending upon its mission (membership (linear), capability of forming pairs (quadratic), and capability of forming groups (exponential)).

4. Winner-takes-all markets

Let us now analyse a market with positive feedback where two providers **A** and **B** are offering the same (or substitutable) product but with the additional condition that customers may churn between the two providers. We also assume that both providers have identical coupling to the market in order to make the model simple.

The churning rate from **A** to **B** (i.e., the net number of customers that **A** looses to or captures from **B**) can be written as a nonlinear function $f(s_A(t),s_B(t))$ where $s_A(t)$ and $s_B(t)$ are the market shares of **A** and **B**, respectively, at time t. Note that that $0 \le s_A(t) + s_B(t) \le 1$ as well as $0 \le s_A(t) \le 1$ and $0 \le s_B(t) \le 1$. The churning rate from **B** to **A** is then $f(s_B(t),s_A(t)) = -f(s_A(t),s_B(t))$ because **B** captures the customers that **A** looses, and vice versa.

The churning rate function f used in this context has the following properties:

- the function is antisymmetric, that is, $f(x,y) = -f(y,x)$
- $-1 \le f(x,y) \le 1$
- $f(1,0) = f(0,1) = 0$, that is, there is now churning if one of the operators have captured all customers.
- $f(x,x) = 0$, that is, the churning rate is zero if both providers have equally many customers (this is a consequence of the antisymmetry).

The market model now consists of a set of two coupled first order differential equations.

$$ds_A(t)/dt = b\ s_A(t)(1 - s_A(t) - s_B(t)) + f(s_A(t),s_B(t)) \quad (6)$$

$$ds_B(t)/dt = b\ s_B(t)(1 - s_A(t) - s_B(t)) - f(s_A(t),s_B(t)) \quad (7)$$

In this set of equations, we have assumed that the coupling to the market (b) is the same for both providers in order to make the analysis of the equations somewhat simpler. The first term on the right-hand side of each equation is the same as in equation (1) in Section 3. However, now the number of customers not having purchased the good is $1 - s_A(t) - s_B(t)$. The feedback term is given by the total number of customers the provider has captured at time t.

We cannot solve these equations in a closed form. However, we shall use a powerful method from nonlinear dynamics called fixed point analysis.

Figure 3 shows the *phase plane portrait* of the two coupled differential equations (6) and (7). The dependent variables are used as abscissa ($s_A(t)$) and ordinate ($s_B(t)$) in the diagram. The time dependence is represented as a vector field of velocities. For a given initial condition, the solution will develop along a trajectory to which the velocities $(ds_A(t)/dt,ds_B(t)/dt)$ are tangents (as indicated by the arrows). At a given point $(s_A(t),s_B(t))$ in the phase plane the velocity $(ds_A(t)/dt,ds_B(t)/dt)$ is found directly from equations (6) and (7) without solving the differential equation. However, in order to determine functions $s_A(t)$ and $s_B(t)$ as explicit functions of time we have to solve the differential equations. Most likely this cannot be done by analytic methods. We have to use numerical integration of the equations providing us with a set of curves representing the market evolution for various initial conditions.

However, we may draw important conclusion about the performance of the system without solving the differential equations as explained below.

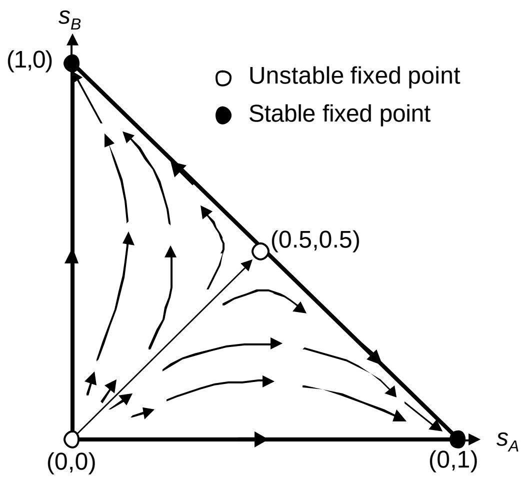

Figure 3 Phase plane portrait of equations (6) and (7)

The market shares s_A and s_B must satisfy the following three boundary conditions:

$$0 \le s_A(t) + s_B(t) \le 1$$

$$0 \le s_A(t) \le 1$$

$$0 \le s_B(t) \le 1$$

This gives rise to a region in the phase plane in which these conditions are fulfilled; that is, the region in which a solution to equations (6) and (7) satisfying the boundary conditions exists. This region is bounded by the following three straight lines:

$$s_A(t) + s_B(t) = 1$$

$$s_A(t) = 0$$

$$s_B(t) = 0$$

The region in which solutions consistent with the boundary conditions exist is enclosed by the bold lines in Figure 3. The figure also contains the four marked points (0,0), (0,1), (1,0) and (0.5,05) called fixed points. In a fixed point, the velocity (($s_A(t)/dt$,$s_B(t)/dt$) is zero; that is, $s_A(t)/dt = 0$ and $s_B(t)/dt = 0$ are satisfied simultaneously. Since the velocity is zero, the system cannot leave such a point if it has been trapped in it. There are three types of fixed points:

- If the system is in a *stable fixed point* and is subject to a small perturbation away from the fixed point, the system will move back again to the fixed point. The vectors in the velocity field close to the fixed point are then pointing toward the fixed point.
- If the system is in an *unstable fixed point* and is subject to a small perturbation away from the fixed point, the system will continue to move away from the fixed point. All the velocity vectors are pointing away from an unstable fixed point

- For some fixed points called *saddle points* some of the trajectories lead into the fixed point while most of them will move away from the fixed point. In the figure, (0.5,0.5) is a saddle point. A saddle point is an unstable fixed point since almost all perturbations will cause the system to move away from the point. In the figure, it is only a perturbation along the line $s_A(t) = s_B(t)$ that will bring the system back to the fixed point.

The fixed points are thus found by solving the two algebraic equations resulting from setting $ds_A(t)/dt = ds_B(t)/dt = 0$. This gives

$$b\ s_A(t)(1 - s_A(t) - s_B(t)) + f(s_A(t),s_B(t)) = 0 \tag{8}$$

$$b\ s_B(t)(1 - s_A(t) - s_B(t)) - f(s_A(t),s_B(t)) = 0 \tag{9}$$

It is evident that the points (0,0), (0,1) and (1,0) are fixed points where (0,0) is unstable and (0,1) and (1,0) are stable fixed points. Since $f(0.5,0.5) = 0$ and $0.5 + 0.5 = 1$, we see easily that (0.5,0.5) is also a fixed point. This fixed point is unstable (in fact, a saddle point). We have furthermore assumed that the churning function is such that it there are no other solutions to equations (8) and (9).

From Figure 3 we see that there are two stable fixed points and all markets satisfying equations (6) and (7) and the particular churning condition will in the long run end up in one of the stable states (0,1) or (1,0); that is, one of the providers will finally dominate the entire market. This situation is referred to as *winner-takes-all* markets. There are quite a number of such markets. Brian Arthur offers several examples of winner-takes-all markets. Examples are the competition between the video recording standards VHS and Beta which was won by VHS, and the competition between the data transmission standards X.25 and the Internet which was won by the Internet. In these markets there was no room for more than one standard.

Finally, observe that adding the differential equations (6) and (7) gives the following single equation:

$$d(s_A(t) + s_B(t))/dt = b\ (s_A(t) + s_B(t))(1 - s_A(t) - s_B(t)) \tag{10}$$

This is the same as equation (4). This implies that the total market $s_A(t) + s_B(t)$ behaves in the same way as shown in Figure 2. The development of the overall market is thus as if there is only one provider. This is only true since the coupling constant (b) to the market is the same for the two providers. However, even if this is not the case, the overall market will grow in a way similar to that shown in Figure 2. This implies (and is in fact true for a vide variety of models) that we must expect that there is a certain latency time before the market really starts growing caused by the positive feedback.

5. Markets without churning

We may model the market with feedback but without churning as shown by equation (11):

$$ds_i(t)/dt = b_i\ s_i(t)(1 - \sum s_i(t)) \tag{11}$$

The sum is taken over all K competitors in the market. We see that the fixed points of this equation are, in addition to the origin, all points in the $K - 1$ dimensional plane $1 - \sum s_i(t) = 0$. All points in this plane are stable equilibriums (or rather quasi-stable equilibriums since the velocity field is zero in the plane).

This example is obvious: the asymptotic state is given uniquely by the initial market share of each competitor. However, it leads us directly into the more complex theory of Brian Arthur where the existence of multiple equilibriums is not so obvious. We may say that equation (11) is a trivial example of this case.

6. Theory of Brian Arthur

Above we have studied the behaviour of markets which can be analysed by simple mathematical means. The purpose was to provide some insight into the effects positive feedback may have on the economy. The main conclusions are:

- The feedback causes the market to grow very slowly initially. This is so because the attractiveness of the product depends on the number of copies sold at any time as well as the number of customers likely to by it.
- Such markets are often of the type in which one of the providers has captured all customers (winner-takes-all).
- As shown by equation (11), there are also markets where the end result may be any of a number of stable equilibria.

The dynamic model can be extended to more complex cases where several providers compete in the same market. Brian Arthur (and others) has shown that, in this case, there may be several stable fixed points allowing a mixture of providers to exist simultaneously. While the linear market is characterised by a single stable state, the markets with positive feedback may end up in one of several end states.

The method used by Brian Arthur is based on what is called "Polya's nonlinear urn problem". The potential buyers can be modelled as an urn filled with balls of different colours where each colour represents one competitor. The balls are picked one by one at random and placed on a table creating a colour pattern showing who has bought the product of different competitors. This pattern is fed back to the urn such that the proportion of balls of a given colour in the urn is the same as that on the table. The probability of picking a ball with a given colour is thus proportional to the number of balls of that colour already picked.

Churning may easily be introduced by allowing the ball to change colour or keep the original colour on the way from the urn to the table in accordance with a churning probability. The churning probability is also a function of the colour pattern on the table. Furthermore, it is possible to introduce additional stochastics by also allowing random changes in the colour distribution of balls in the urn.

The asymptotic solution to this problem is that there ultimately may be several possible colour patterns on the table each representing a stable sharing of the market among the competitors. There is thus not just one equilibrium solution but several. This is what Brian Arthur and collaborators proved. The existence of the urn allows us to apply the theory to markets where the demand for the product is an arbitrary number of copies per customer.