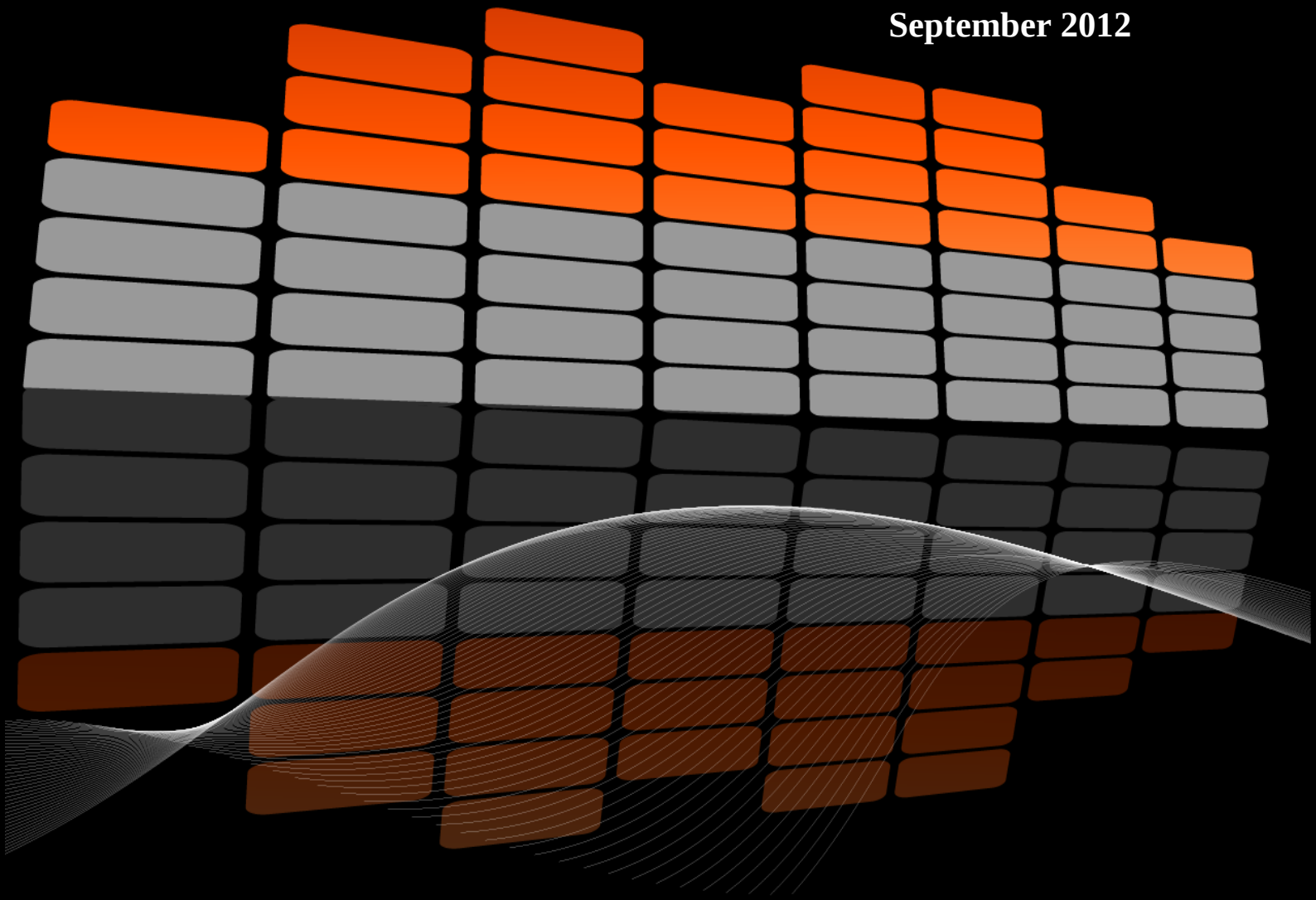
International Journal of Latest Trends in Finance & Economic Sciences

E-ISSN: 2047-0916

Volume 2 No. 3

September 2012



Web: excelingtech.co.uk

ojs.excelingtech.co.uk/index.php/IJLTFES

Editorial Board

Editorial Board

Editor in Chief

Managing Editor

Editorial Board Members

Abstracting / Indexing

International Journal of Latest Trends in Finance and Economic Sciences (IJLTFES) is abstracted / indexed in the following international bodies and databases:

- EBSCO
- Google Scholar
- DOAJ
- Scirus
- Getcited
- Scribd
- NewJour
- citeseerx

Table of Contents

Investor Behavior in Extreme Situations of Speculation and Crash: An Approach based on Iterated Prisoner's Dilemma

Rui Miguel Silva[*1], José António Filipe[#2], Ana Costa[#3]

[*1] *Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal*
E-mail: rui_silva03@hotmail.com

[#2] *Instituto Universitário de Lisboa (ISCTE-IUL), BRU/UNIDE, Lisboa, Portugal*
E-mail: jose.filipe@iscte.pt

[#3] *Instituto Universitário de Lisboa (ISCTE-IUL), DINÂMIA'CET, Lisboa, Portugal*
E-mail: acfc@iscte.pt

***Abstract* – This paper analyzes the investor behavior in situations of speculation and crash on stock markets. The investors' main behavioral features are addressed, notably those related to cognitive and decision-making matters, in order to obtain an individual and an aggregated behavioral profile of the investor in situations of extreme events.**

Keywords: Financial Crisis, Speculative Bubbles, Investor Behavior, Rationality, Iterated Prisoner's Dilemma

1. Introduction

The complexity and dynamics of the financial markets make this matter an extremely interesting subject for analysis. More specifically, the increased instability of the markets, characterized by periods of strong speculation and by crashes, affects macroeconomic and monetary stability. This reality has led in part to a rise in related studies and therefore allowed many new hypotheses. Different proposals have been made to shed light on the dynamics of the market and its recent shape.

This is the context in which we debate investors' behavior in extreme situations in the financial market. It is demonstrated that investors can develop cooperative attitudes in a speculative period (known as bubble) so as to maintain a favorable position that brings above average benefits; in addition, their actions are flawed in more unstable situations i.e. prior to and even at the exact moment of a crash.

Firstly, a methodological approach to the problem based on considerations about investors' behavior, notably cognitive and with intellectual limitations, is presented. Then some characteristics implicit to choice in uncertainty are defined.

This involves making a brief review of the intrinsic characteristics of extreme, but real, events based specially on the market crashes of 1929 and 2000.

Finally, a game theory model is presented for the problem under analysis, departing from the assumption that investors do not make use of any type of arbitrage or hedging strategies, and their investment decisions are geared to stocks (and in bonds only to minimize the risk); the investment in other instruments, like derivatives, is only considered in the case of a strong link to stocks. It is important to note that the choice of this type of approach (game theory) is essentially due to its focus on the behavior and decisions of the players individually and in group. The impact of the dynamics and actions of players on the market where the game is played is also analyzed.

2. Some features of investor behavior

Human beings have always been the direct or indirect focus of theories in the financial and social sciences generally. Human behavior and its respective features have been a key element for the success of several models, either from a micro or macro point of view.

This has played a major role in the research into financial markets. As can be seen in the first major studies on portfolio theory, made initially by Markowitz (1952) and later by Sharpe (1962) and Ross (1976), the definition and concretization of the investor's behavior was key to the efficiency and explanatory capacity of these models.

However, the lack of accuracy demonstrated by these models in several market situations opened the way to new approaches which essentially focus on the characteristic features of the investor and his respective decisions. This has provided new possibilities to understand some events.

2.1. The Rationality postulate and the deficiencies in information processing

The rationality and consequently the processing and use of information in the decision-making process are topics that have always intrigued economic and financial researchers. The association of the proposition of rationality to the economic man, made by several theorists, has been used, across the years, as a powerful assumption in the construction of several models.

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.001.12/ijltfes.2047.0916

Following this idea, it is assumed that the agent has the knowledge to make the best possible decisions in the existing environment and with the intrinsic limitations, supported by a well-organized and stable system of preferences, and in a context of perfect information, leading him to the best possible action (Simon, 1955).

It is therefore assumed that in a context of perfect information, the agent can process this information correctly and thus make the right decisions; an imperfect decision can be the result of an asymmetric information context.

However, despite the acclamation of the rationality postulate to be a vital assumption to several models, the Keynesian theory, for example, showed that good predictive models can be constructed in a way that is not based on this postulate (Blaug, 1992). As Arrow (1987: 70) refers: "*I don't know any serious derivation about the currency demand based on rational optimization*".

Therefore, an in-depth analysis of this issue is made, starting with an overview of the orthodox financial theory, followed by the ideas of the behaviorists, evolutionists and neuroeconomic theorists. The main objective is to provide an alternative theoretical background to support the non-appliance of the utilitarian agent in the subsequent model derived later on.

2.1.1. The Financial vision of Rationality

Markowitz (1952) was the first to bring a well designed approach for the selection of assets and construction of an asset portfolio to the financial theory. In his attempt to explain the allocation and selection of securities in a portfolio, he made a set of assumptions, notably the rationality principle. More specifically, it is assumed that within a context of perfect information the investor maximizes (or should maximize) the discounted expected returns, and diversifies (or should diversify) his funds among all the available securities, leading to a situation of maximum expected return and a mean-variance portfolio (Markowitz, 1952).

Some years later, the Capital Asset Pricing Model, known as CAPM, was developed in articles by William Sharpe (1964), John Lintner (1965) and Jan Mossin (1966). This model focuses on the relationship between the level of risk and the expected return of an asset and on the following and subsequent *equilibria*. The set of assumptions used is quite similar to what was used by Markowitz. All investors are rational mean-variance optimizers; hence, if all investors are rational, they will all analyze securities in the same way and share the same beliefs, which leads to homogeneous expectations (Bodie *et al*, 2009).

Another important asset pricing model is known as the Arbitrage Pricing Theory (APT) by Stephen Ross (1976). Though similar to CAPM, it is more general in the sense that the security returns are described through a factor or a set of factors related with the macroeconomic, financial or business sector environment. The main assumption is that a well-functioning security market does not allow the persistence of arbitrage opportunities because securities are not mispriced over a long period of time (Bodie *et al*, 2009; Ross, 1976).

Notwithstanding some other important models, it is turn lastly to the Efficient Market Hypothesis (EMH). Like the abovementioned assumptions, EMH assumes the market to be efficient and that individuals are rational. Basically, a market is efficient if the traded assets reflect all the available information in a given time, and if the price of the asset adjusts as quickly as possible to the new information; this leads to a random walk as the prices change unpredictably (Bodie *et al*, 2009).

2.1.2. The behaviorists critique and alternative

Despite the huge advances brought by the abovementioned theories to the evolution of financial and economic theory, they tend to fail in several situations because they are usually based on a normative analysis, which is concerned with the rational solution for the decision-making problem. This solution results from the definition of the ideal decisions to approach, rather than a descriptive analysis of the way in which real people actually make decisions (Kahneman and Riepe, 1998).

One of the critiques made by behavioral finance is that almost all investors suffer from biases of judgment and decision-making, sometimes called cognitive illusions. For this reason, the investor does not always process information correctly and tends to assume risks that do not acknowledge; this leads to incorrect probability distributions and inconsistent and systematically suboptimal decisions (Bodie *et al*, 2009; Kahneman and Riepe, 1998).

Overconfidence is one such bias. When the investor makes his own prediction, he often sets a very narrow confidence interval, thinking of specific quantities and anchoring too much in his own prediction. Unfortunately, few people are able to calibrate their predictions well and judgment errors are common. Moreover, this phenomenon is expected in dynamic environments where the agent systematically faces different problems and cannot learn with past examples as quickly as other agents in more stable environments (De Bondt, 1998; Kahneman and Riepe, 1998). If the investor is rational, the environment will be indifferent to his decision, making it well calibrated and leading to the same or similar behavior to that described in financial models.

Optimism is another important bias that supports the critique made by behavioral finance to the rationality postulate. The agent tends to rely too much on his own beliefs and talents so that he exaggerates the future outcome. Mixing optimism with overconfidence will generate an overestimation of the knowledge acquired and an underestimation of the risks, leading to an illusion of control in most events (Kahneman and Riepe, 1998; Shiller, 2000).

While these two biases are a great constraint to the investor's decision-making process, the hindsight bias can also play an important role because it encourages the agent to see the future as more predictable than it really is, and this will heighten overconfidence. If the event had been predicted, many of the bad situations would have been avoided because almost everyone would have modified their actions (Shiller, 2000).

Over-reacting to change events is another bias that is closely linked to the overconfidence phenomena. The investor believes that random moves are more likely to occur than systematic ones, impelling him to perceive patterns that do not exist; this indicates overconfidence in judgments about uncertain events (Kahneman and Riepe, 1998).

All four judgment biases are generated and amplified by certain types of anchor. In general, people tend to anchor too much because, when making ambiguous and complex decisions, they are influenced by the readily available information. The overconfidence and optimism biases may appear in situations where the investor uses quantitative anchors, e.g. the most recently remembered price or the nearest milestone to a major index. These anchors can lead to several judgment errors, creating an illusion prediction. On the other hand, moral anchors can be responsible for the hindsight bias because when the market is not working well, people tend to hold on to stories and intuitive reasons to embrace their investments and to see a more predictable world than actually exists. The fragility of these anchors lies in the agent's difficulty in using them to think ahead to contingent future decisions (De Bondt, 1998; Shiller, 2000).

Another limitation of the rational decision process is due to the heuristics used. In the original Greek definition, adopted by Duncker (1945), heuristic "serves to find out or discover" and is used to describe strategies such as "looking around" and "inspecting the problem". A few years later, Simon (1955) defined heuristics as strategies that facilitate decisions. More recently, the term has evolved, especially in the decision-making segment to denote strategies that help to find and to discover correct answers to problems in the probabilistic area of decision (Goldstein and Gigerenzer, 2002).

However, when dealing with optimizer behavior, the use of heuristics to solve problems sometimes leads to judgment errors and inefficient final outcomes. The representativeness heuristic is an example of this. In uncertain situations, a judgment is made by looking at familiar patterns and making an assumption that the future will resemble past patterns. In these cases, even without a sufficient consideration about these patterns, probabilities can be forgotten which results in overconfidence. Individuals dealing with uncertain environments such as financial markets may use this short-cut and make decision mistakes (Shiller, 2000; Tversky and Kahneman, 1974). In addition, recognition heuristic reflects a lack of information processing by the agent. It is used when the agent faces a choice between two or more objects. In these situations, the known object has a higher value in the individual's decision criteria. This heuristic relies on low cognitive ability and is often systematic. The problem focuses on the fact that the individual chooses the recognized object because he has more information about it and acting against the recognized object requires more cognitive effort (Pachur and Hertwig, 2006; Volz *et al*, 2006). Another heuristic that influences the decision-making process is known as the adjustment and anchoring heuristic. The anchoring process was examined above, but in this particularly case it is associated to the mental short cut of adjustment. In some uncertain situations, the agent estimates the final outcome, starting from a given initial value that is adjusted over time to yield the final result. Different initial or starting points obviously yield different estimations that are biased toward the initial values, in a phenomena caused by the anchoring. This problem is catalyzed essentially by insufficient adjustment and the existence of biases in the evaluation of events that are known as conjunctive (events that must occur in conjugation with others, like a multiple step plan); and disjunctive (events that are successful if at least one event is favorable) (Tversky and Kahneman, 1974).

2.1.3. Evolutionism approach

Evolutionism is another approach that represents a different way of analyzing the rationality postulate. The application of Darwin's theory of evolution to economic and social sciences has been controversial in recent years, principally because some authors consider it too mechanic and biological to be applied to the dynamics of sciences that deal with social and economic problems (Aldrich *et al*, 2007).

Despite these critiques, nowadays evolutionism is an important theory that can give a valid alternative to the rationality postulate.

The critique of the rationality postulate implied in the orthodox financial and economic theory is sustained, in the most general and simplified way, by the theory of Mayr (1988), known as paradigm of program-based behavior.

Mayr's theory essentially relies on the fact that an agent's behavior can be seen and guided by programs encoded to face different situations. These programs allow the agent to foresee and face the consequences of his potential choices in uncertain environments. These programs are constructed and mutated by a process of learning and evolution, through which they become more adapted to the relevant characteristics of given problems and environments. This process tends to eliminate and replace inadequate programs with new programs with different characteristics and knowledge in order to make decision-making more accurate. Thus, programs tend to be more adapted to the different problems and are a product of the agents' evolution and learning (Mayr, 1988; Vanberg, 2004).

The implication of this theory to this discussion relies on the possibility for specific actions to be not rational (from an optimizer way of thinking), even if

programs are well adapted to the particular problem and environment. It is allowed trough this theory the possibility of the existence of a systematic account for observed behaviors that can be considered as irrational and that are classified as anomalies (Vanberg, 2004).

2.1.4. The role of emotions and the neuroeconomic analysis

The role of emotions in the decision-making process and questions related to the analysis of utilitarian rationality has received growing support in recent years. Neuroeconomics is one field that has devoted considerable effort to this area. One of the main points of research in neuroeconomics is the relation between brain activity and the choice and decision-making process under uncertain conditions. The neural reactions to some situations of choice can lead to a better understanding of how some decisions and actions are taken.

Damásio (1994) gives two examples that illustrate this problem. The first is that of Phineas Gage who lived in the mid 19th century in New England. He was a foreman working on the construction of a railroad. On a given day, when he was trying to detonate a pile of rocks, an iron bar was projected into his face, entering in the left side of his face and getting out by the top of his head. Phineas did not die and was fully conscious when he went to the hospital. Doctors today would know that this was a lesion in the Ventromedial Prefrontal Cortex and that the other important brain lobes were fully intact. Although he resumed normal life two months later, but never more was the same. The balance between the intellectual and instinctive sides had been destroyed and he became unpredictable and indecisive, displayed few emotions, made countless plans for the future which were easily abandoned. He was no longer able to work as a foreman, but the same problems arose when he did other jobs. He was unable to make decisions that were coherent with his knowledge. He died years later from a pathology known as *status epilecpticus.*

Damásio's second example is that of Elliot who had a brain tumor known as Meningioma; this was surgically treated by removing frontal lobe tissue but a lesion in the cortical region had damaged the Ventromedial Prefrontal Cortex (the temporal, occipital and pariental regions were intact, as were the basal ganglions and thalamus). Nevertheless, he made a good recovery but, like Phineas, was never the same again. He rarely got angry, and rare were the situations when he expressed emotions. This was caused by poor access to the social knowledge which is essential to more advanced reasoning. Some of the tests conducted revealed he was unable to make an efficient decision and sometimes no decision at all (Procrastination).

In these cases and others of lesions in the Ventromedial Prefrontal Cortex, patients show diminished emotional responsiveness and limited social emotions, closely associated to moral values. They also sometimes exhibit above average tolerance to anger and frustration which generally lead to bad or inefficient decisions. Notwithstanding, their general intelligence, logic reasoning and knowledge is unaffected (Koenigs *et al*, 2007).

This profile of a VMPC (Ventromedial Prefrontal Cortex) patient can be explained by Damásio's Somatic Marker (1994). In cases of decision-making which require the evaluation of future consequences, the somatic marker classifies the future action as good or bad. The somatic state makes the decisions quicker and more effective. Lesions to the Ventromedial Prefrontal Cortex cause the somatic signals guiding the action to fails. As a result, patients show indifference to possible future consequences of their actions, and are unable to see beyond the present. (Butman and Allegri, 2001; Damásio, 1994)

According to some empirical studies by Bechara *et al* (1994, 1996 and 1997), and Koenigs *et al* (2007), the main conclusion, is that VMPC patients have more utilitarian judgments and act more according to the economic and financial doctrine of rationality although may not be the best strategy (because this behavior does not take into account the importance of emotions in the decision-making process). In studies using card games, VMPC patients prefer to take risks and dangerous bets without considering the future outcomes of their actions. In the study by Koenigs *et al* (2007), VMPC patients have no difficulty in taking decisions in more emotional and stressful situations, which leads to more inefficient decisions/outcomes in a utilitarian way.

It can be concluded that the emotional side plays an important role in the decision-making process, leading to more efficient choices. It is known that uncontrolled emotions can lead to irrational behavior. But the reduction of emotions can lead to equal irrationally behavior (Damásio, 1994).

If emotions are responsible for irrational and rational decisions, the individual is not fully rational. But without emotions and with an increase in the utilitarian judgment, his decisions can be equally irrational and so the rationality postulate implied in most of the models cannot be correct.

2.2. The dynamics of the investor behavior

Arriving at this point, it is appropriate to analyze the particularities of the investor in a dynamic environment, highlighting and detailing the factors that determine his behavior in the market.

In investment dynamics, many decisions are made and this process is extremely intensive and demanding. It is important to define the features that determine the process to provide a profile of the investor. Thus, a brief review over the preferences of the investor's preferences, the way he makes a choice and the determinants affecting the decision-making process will be made.

2.2.1. The vision of orthodox financial theory

In financial theory, the investor's features are part of aggregated models that try to make a macro explanation of market behavior; this implies defining a more general, and less precise, set of assumptions for the investor.

In this context, the first assumption refers to the investor's preference to smooth his consumption, because of: (1) Time Consumption and the (2) Risk Dimension. The first is based on the fact that consumption is higher than the income in the early years of active life, because of situations like purchasing a house or a car. However, during those times, savings are constituted, that will be spent after retirement when the income is zero and consumption is positive. The risk dimension factor is based on the fact that the future is uncertain and that many states of nature can occur, which in turn makes it necessary to level off consumption so that it is not excessively concentrated in potentially unfavorable periods (Danthine and Donaldson, 2005).

Based on this, the process of decision-making can be divided into situations of certainty and uncertainty. While the assumption of rationality can be accepted in a situation of certainty (with the appropriate reservations because the choice depends on the framing of the problem) with every investor having a complete preference relation and the property of transitivity in a continuous relation, this is more difficult to occur in uncertain situations. In these situations (like a lottery), it is assumed that the preference relation is complete, transitive and continuous, with an independence of irrelevant alternatives. This last assumption is not common ground because it depends, for example, on the way in which the problem is placed (framing); this will be analyzed in more detail below (Danthine and Donaldson, 2005; Huang and Litzenberger, 1988).

Another assumption is that the investor is risk-averse, because he usually wants to avoid a fair gamble (when in an uncertain environment); his utility function is concave because, as the wealth increases, the utility from the additional consumption decreases (also known as decreasing marginal utility). However, despite not being assumed directly in the portfolio theory, this degree of risk-aversion can be measured in two ways: a) in terms of absolute risk aversion (ARA), that is, sensitivity to the amount and; b) relative risk-aversion (RRA), i.e. sensitivity to the proportion of wealth at stake. Thus, it is assumed that an investor will only play a fair game if there is a certainty equivalent, i.e. if there is an amount of money that is a certain equivalent to the investment that he could make (Holt and Laury, 2002).

Assuming the above, the problem for the investor is to maximize the expected utility of his wealth allocated in the possible investments. To do so, he integrates mean-variance preferences so that when there are investments with the same mean, he chooses the one with the smaller variance, and in the case of investments with the same variance, he chooses the one with the larger mean (Markowitz, 1952).

Every investor will generally possess the market portfolio and will invest in a risk-free asset (in order to respect the two fund separation), so the wealth will be allocated between the rf (risk free asset) and the tangency portfolio. But because the investor is risk-averse, he will only invest in the risky asset if his expected return is higher than the rf (McDonald and Siegel, 1986). Then it will be respect mean-variance dominance:

$$\text{Asset a dominates asset b if } \begin{cases} \mu_a \geq \mu_b \text{ and } \vartheta_a < \vartheta_b \\ \text{or} \\ \mu_a > \mu_b \text{ and } \vartheta_a \leq \vartheta_b \end{cases} \quad (1)$$

Moreover, he will look for changes in the composition of the portfolio in terms of the correlation; this implies that the construction of the portfolio will mainly take securities that have a correlation of between]-1;1[into consideration (Bodie *et al*, 2009; Markowitz, 1952).

It is assumed that a more risk-averse investor will allocate less wealth on the stock market; however, this sometimes depends on the intrinsic utility function of the investor.

Also, according to the CAPM, all investors possess the market portfolio and will therefore be pleased when the market goes up and sorrowful when it goes down; because they respect the law of decreasing marginal utility. This implies that what really matters to the investor is getting additional good payoffs in bad circumstances (of low market returns), which in turn makes the investor less enthusiastic about additional payoffs in good times. It can therefore be concluded that the investors like assets with low covariance with the market (Bodie *et al*, 2009).

2.2.2. An alternative based on Behavioral Economics and Finance

(a) Hyperbolic Discounting

Most decisions made by an investor involve a trade-off between outcomes/choices that will have effects on different periods; which in the real markets imply that the investor has to decide between investment options that may be more valuable in the future than in the present. This relation is captured in a conventional analysis by a discount function. With the help of this instrument, it is possible to measure the utility obtained from a series of future consumption situations, occurring at regular intervals, leading to the calculation of a Discounted Utility Function.

$$U' = \sum_{d=0}^{n} F(d)\, u\big(c(t+d)\big) \quad (2)$$

Where *F(d)* is the discount function, *t* the time of evaluation and *c(t+d)* the resources consumed at time *t+d*.

Thus, the discount function is a declining function of delay and often given by a discount rate *r*, which is the proportional change in the value of *F(d)* over a standard time period. It is also important to note that the decision maker is impatient and the rate of change of

F(d) is the pure rate of time preference. In addition, for rational decision makers, the rate in which money should be discounted must equal their marginal rate of substitution between the present and the future to the market interest rate.

Taking the following example: if we actually prefer 5€ in 3 months to 4€ in 2 months, then in 2 months time we will prefer the 5€ in 1 month to 4€ immediately unless there is a sudden need for cash. However, this may not occur with certainty and can imply an inconsistent time preference. Taking the examples given by Ainslie (1975), Ainslie (1991) and Read (2003) where we have a choice between two alternatives: a smaller-sooner (X) and a larger-later (Y); while the larger-later alternative is preferred when both are substantially delayed, when smaller-sooner alternative becomes imminent it undergoes a rapid increase in value and is briefly preferred. For example, the smaller-sooner reward can be the pleasure from a cigarette and the larger-later reward might be good health. The prospect of good health is preferred when looking one week ahead, but the desire for the cigarette grows faster than the desire for good health as time passes, until, for what may be a very brief period, the cigarette is preferred.

This kind of situation makes it difficult for the agent to plan the future and stick to it, which degenerates into procrastination. This example also shows that the discount rate does not always change proportionally to the value of *F(d)* over a standard time period like the one referred. Due to these time inconsistencies, a hyperbolic discount function can be the best way to illustrate this type of behaviour, instead of exponential discount functions as assumed when the decision maker is a rational agent, because it considers there may be a brief and temporary reversal in preferences (Read, 2003).It can therefore be said that individuals do not always smooth their consumption because, at one point in time, the agent may reverse his preferences (Steel and König, 2006).

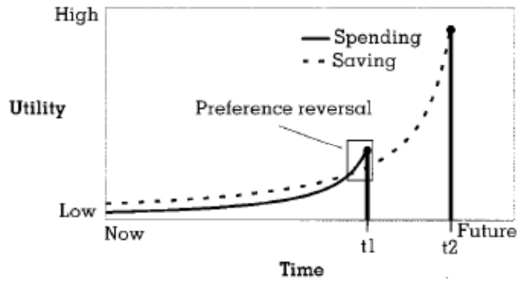


Figure 1: Possibility of reversal of preferences (Steel and König, 2006)

Another point that is not consensual is the consideration that money should be discounted at the prevailing market rate (Thaler, 1981). In fact, people do not apply the same rate to all decisions, being instead this rate highly domain dependent and even in the domain context dependent from the choice context (Chapman and Elstein, 1995). In addition to time inconsistency, several anomalies, linked to the constant discount theory can be summarized:

- **Delay effect**: if we elicit the present-value of a delayed outcome or the future value of an immediate outcome, then the longer the delay, the larger the obtained value of the discounting factor (Read, 2003);
- **Interval effect**: The difference between the delays of two outcomes is the interval between them. So discounting depends heavily on the length of this interval, in that longer intervals lead to smaller discount rates or larger discount functions (Read, 2001);
- **Magnitude effect**: This means that the discount rate is higher for smaller amounts (Green *et al*, 1997; Read, 2003; Shelley, 1993);
- **Direction effect**: the discount rate obtained by increasing the delay of an outcome is greater than that of reducing that same delay (Loewenstein, 1988; Read 2003);
- **Sign effect**: The discount rate is lower for losses than for gains (Antonides and Wunderink, 2001; Thaler, 1981);
- **Sequence effects**: A sequence is a set of dated outcomes all of which are expected to occur, such as one's salary or mortgage payments. People usually prefer constant or increasing sequences to decreasing ones, even when the total amount in the sequence is held constant (Chapman, 1996);

(b) Prospect Theory

For orthodox financial theory, the evaluation of outcomes and the decision-making process can be analyzed by taking the expected utility theory into consideration. In this theory, it is assumed that investors attempt to maximize the expected utility of their choices between risky options, giving weight to each outcome according to their probability and choosing the one with the highest weighted sum (Luce and Raiffa, 1952). It is also assumed that the psychological value of money or goods follows the rule of diminishing marginal utility, which is represented by a concave utility function, implying the presence of risk aversion (Levy, 1992).

Prospect Theory however posits a different way of analyzing this problem. It is assumed that the agents evaluate outcomes based on the deviations from a given reference point, instead of the level of net assets or value. The real deal, however, is the identification of this reference point. At the moment zero, it is usually assumed to be the *status quo*, but in some other cases it may be the aspiration level or another point. Allied to this, the agent is not always risk-averse and this varies depending on whether we are dealing with gains or losses (Kahneman and Tversky, 1979).

For example, an experiment by Kahneman and Tversky (1979) gave the choice of a certain outcome of $ 3000 vs. 80% chance of winning $ 4000 and 20%

chance of winning nothing. 80% of the respondents chose the certain outcome. However, when dealing with the same problem but in a negative frame, 92% chose to gamble when there was 80% chance of losing $ 4000 and 20% of losing nothing to a certain loss of $ 3000. In both cases the option with the lower expected value was chosen, which is incoherent with the expected utility theory and highlights the risk profiles. It suggests that individual utility functions are concave for the domain of gains and convex for the domain of losses; this is a pattern known as the reflection effect to the reference point. This implies that the sensitivity to changes in assets decreases as one moves further in either direction from the reference point (Kahneman and Tversky, 1979; Laury and Holt, 2000).

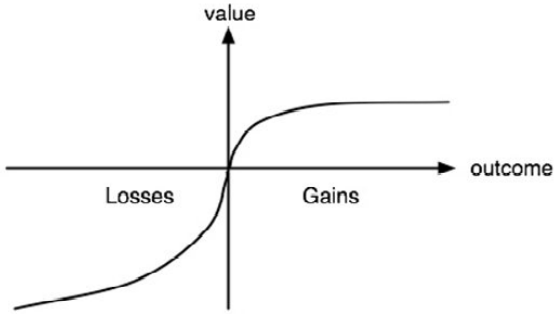


Figure 2: Prospect Theory utility function (Kahneman and Tversky, 1979)

However, the previous example shows that the propensity for risk depends on the way in which the problem is placed, i.e. the way it is framed. For example, in Kahneman (2002), to subjects were given the hypothetical choice between programs to outbreak a disease which was expected to kill 600 people. On the first attempt: program A corresponded to 200 people saved while in program B there was a 1/3 chance that 600 people would be saved (and no one die) and 2/3 that no people would be saved. On a second attempt: program A implied the death of 400 people and program B corresponded to a chance of 1/3 of people not dying and a 2/3 probability of people dying. The results showed that on the first attempt the majority of respondents chose program A, which indicates risk aversion. However, on the second attempt, program B was mainly chosen, which indicates risk-seeking behavior. It can be concluded that on the first attempt the possibility of certainly saving people was more attractive than a probability; on the other hand, respondents were averse to accepting the certain death of people and thus sought more risky alternatives (Kahneman, 2002).

Allied to this context, there are two types of effect that influence the decision-making process. First, the certainty effect, which impels the individuals to overweight outcomes which are certain relatively to outcomes that are merely probable. Also, they overweight low probabilities and underweight moderate or high probabilities. The latter effect is more pronounced. Therefore, extremely likely but uncertain outcomes are often treated as if they were certain; this is known as the pseudocertainty effect (Levy, 1992). Also changes on probabilities near to 0 or 1 have a greater impact on preferences than comparable changes in the middle probability range, leading to behaviors of subproportionality (Tversky and Kahneman, 1986).

(c) Mental Accounting

The mental accounting theory has proven to be a partially effective and efficient approach, along with the prospect theory, to understand the behaviour of agents and particularly of investors. For Kahneman and Tversky (1984), mental accounting is an outcome frame which specifies a set of elementary outcomes that are evaluated jointly and the manner in which they are combined; it is a reference outcome that is considered neutral or normal. It supports three important features: the prospect theory value function over gains and losses is used in relation to a reference point; both gain and loss functions display diminishing sensitivity; at the initial reference point (*status quo*) the agent is risk averse (Thaler, 1999).

One of the main proposals of this theory is that people behave according to the hedonic framing proposition, i.e. they segregate gains and integrate losses (because the respective functions are concave and convex) and more specifically, integrate smaller losses with larger gains and segregate small gains from larger losses (Thaler, 1985; Thaler, 1999). However, this proposition can sometimes fail, principally for the integration of losses, as Thaler and Johnson (1990) showed in their research. People sometimes think it is good to integrate losses, which intuitively implies that it should diminish the marginal impact and suggests that a prior loss makes them more sensitive to subsequent losses (Thaler, 1999).

Mental accounting therefore predicts that, if for example, we buy, *s* stocks at *p* price, the investment will initially be worth *[s * p]* and will fluctuate in accordance with the evolution of stocks on the market. In fact, even with changes over time, which implies theoretical gains or losses, it only becomes a realized gain or loss when this position it is sold. An account will be opened with *[s * p]* and will be closed with the realized result, which can compensate or not for the initial investment. But because closing an account at a loss is painful to the investor, the prediction of mental accounting is that investors will be reluctant to sell securities that have declined in value. If at a given moment the investor has a need for cash, he will look at his asset portfolio (which for example contains *n* securities) and will sell those with a higher value than at the time of purchase. However, this hypothesis contradicts a rational analysis that postulates that the investor should sell the securities that were lower than their initial value. However, the assumption made by mental accounting theory can be supported on the example of Odean (1998) that, using data from transactions made by a big brokerage firm, had show that investors were more willing to sell one of their stocks that had increased in value than one that had decreased.

Another particularity analyzed by mental accounting is that sometimes the investor suffers from behaviour termed myopic loss aversion. This behaviour is analyzed in detail in the Equity Premium Puzzle of Benartzi and Thaler (1995); it focuses on the difference in the rate of return of equities and a safe investment like treasury bills, which historically has been very large (6% in the USA in the past 70 years) and which resulted in the appreciation of 1 dollar invested in equities almost 120 times the return from the dollar invested in treasury bills. However, T-Bills was the primary destination for investment in these years. The explanation for this puzzle was that the investor's loss aversion is strictly dependent on the frequency with which he reset his reference point (or how often he counts his money). The result was that people are indifferent as to investing in stocks or T-Bills if they only evaluate changes in their portfolio every 13 months. So the investor can suffer from myopic loss aversion because this myopic behaviour prevents him from using the best long term strategy and makes him think primarily about the present; which makes him evaluate the composition of the portfolio with great frequency. When the evaluation period is larger, the attractiveness of stocks increases (Thaler *et al*, 1997; Thaler, 1999).

3. The anatomy and history of Bubbles and Crashes

Bubbles and Crashes are unique situations which have been studied across the years. Economic and financial theorists' interest in this theme may be due to the fact that almost all propositions about investor rationality and then market rationality can be violated. Thus, in relation to the present article, a more realistic and improved investor profile can be constructed using this kind of situation because the efficient market hypothesis loses its descriptive validity.

According to the efficient market hypothesis, a crash occurs when a dramatic piece of information is revealed. However, this approach can be considered reductive as the piece of information that triggered the problem may be unknown; even when known, a period must have preceded the crash that created the necessary conditions.

In contrast with the market efficient hypothesis, it can be said that in these situations the market has entered an unstable phase so a small endogenous disturbance is sometimes enough, to trigger a shock (Sornette, 2003). The 1929 and 2000 crashes can be used to describe such unstable phase; in these years, the preview upward trend in stock prices never more was seen, being replaced by an unstable and undetermined fluctuation, with special emphasis on losses.

In fact, this situation is preceded by a rapid rise in market prices, known as a bubble, created by growing interaction and cooperation between investors that can last for months and even years. Investors are unaware of the cooperation relations that result from a general belief in a new state of affairs, triggered primarily by the growth of a given sector or industry. The expectations and beliefs generated tend to be accepted by the group of investors, which helps prices rise quickly and vertically in some days (Galbraith, 1954; Kindleberger *et al*, 2005; Sornette, 2003).

It can be concluded that the market's unstable position will lead to a collapse and the piece of information that trigged the reaction can be considered secondary (Sornette, 2003).

It will be now present a brief review of some historical stock market crashes to introduce the problem analyzed with the Iterated Prisoner's Dilemma.

However, the following criteria were used to select these historical events.

- First, these events took place in the United States of America. Despite the globalization on the financial markets, there are cultural, social and other differences between investors in different countries and these can lead to singular behaviors and practices that could skew the analysis.
- Second, these extreme events were primarily in the stock market. Events that originated in other security markets were excluded.
- Third, these events were preceded by long periods of speculation and formation of a bubble, followed by very destructive crash. This explains why Black Monday of 1987 is not in the list.

Two stock market crashes were chosen: The Great Crash of 1929 and the Dot-Com Crash of 2000. Both represent an optimal context of Bubble and Crash, meeting the above criteria. Both have long periods of speculation and strong crashes. The expectations and behavior of investors were quite similar, despite the industry or sector that leaded these beliefs.

The financial crisis of 2008 was not chosen, because it can be assumed that Financial Institutions, e.g. commercial and investment banks and other financial companies, created the speculative moments and not investors. Moreover, this event originated in the real estate market and not the stock market. Finally, the effects of the crisis are still ongoing and it is difficult to dissociate the subsequent effects from the recessive macroeconomic landscape.

3.1. The Great Crash of 1929

The 1920s could be called both a golden and a dark age. With World War I at an end, everyone was convinced that this would be a prosperous decade. Indicators for economic growth and development were improving; consumption was growing at a fast pace, and the level of prices was stable.

It was the time of the American dream which foresaw a better, richer and fuller for one and all. It was a vision of a social order in which everyone could reach their maximum potential and break the barriers of the old social hierarchy (Adams, 1931).

However, in the late 1910s, the rich were becoming richer more rapidly than the poor were becoming less poor. Because of this context, the financial and real estate markets provided the opportunity to get rich with the minimum effort, thus fulfilling the American dream and inverting the trend of the previous years (Galbraith, 1954). This led to the great demand in the real estate market across the USA, particularly in Florida and should have been a forewarning of what would happen in 1929.

The problem to accomplish this desire was that individuals simply believed that they were meant to be richer, regardless of their intellectuality restrictions, that is, cognitive limitations based on limited rationality and in the use of heuristics that could bias decision. Risk-taking and irrational decisions were the order of the day. An entire industry was born to provide services to investors on the stock market such as brokerage firms, investment banks and investment trusts. The creation of a bubble was inevitable, the same for the following crash (White, 1990).

3.1.1. The premonition

In the early 1920s, the first signs of speculative behavior and irrational illusion came from the real estate market. The main boom was in residential housing (White, 1990). Between 1921 and 1925 construction grew at a rapid pace with housing prices and building costs following the same trend. The boom was fueled by good macroeconomic conditions as well as the desire to become a home owner so as fulfill the American dream. This environment led to a change in the profile of investors with preference being given to short term profit as opposed to the constitution of savings.

Florida is the most outstanding example of the boom in the real estate market. The standard of living and its transport system developed quickly and the climate made it the perfect location for a speculative wave in real estate. Investors were easily influenced at the time and simply wanted an excuse to believe in something. And that excuse and belief came from the expectation that Florida would become a dream place, full of opportunities and rich people enjoying the local conditions. In addition to the formation of positive expectations, the real estate market began to increase, making real the expectations of the investor. With time, the price of land rose and land owners were soon making big profits. After some time, the reasons for the investment on these lands started to disappear, exceeded by the possibility of easy profits. The problem arose on the beginning of 1926, when the number of new investors and houses began to fall. The subsequent decline in house prices was the beginning of a slow crash; the Great Crash of 1929 that followed dashed all hopes of recovery (Galbraith, 1953; White, 2009).

The example of Florida reflects what happened across the USA and demonstrates that Americans were driven by the desire for get-rich-quick investment opportunities in the early 1920s.

3.1.2. From the prior years to the Great Crash – Euphoria and Mania

At the beginning of the 1920s, conditions were ripe for the expansion of the stock market. Despite the good macroeconomic environment, stock prices were low and dividends reasonable. Most companies were making high profits and this seemed the prevailing trend. Some of these were new large-scale commercial and industrial enterprises that took advantage of innovative processes and technologies. They were able to capture economies of scale and scope which made their production processes very efficient. Moreover, there was a great transformation in the utility sector in large part because of the rapid growth in the modern industrial enterprises. In hindsight, high returns could not be sustained because the markets were poorly developed and companies held unbalance structures (Chandler, 1977; White, 1990).

The greatest increase in the volume and prices of the stock market, particularly in the Dow Jones Industrial Average, began in 1927. In previous years, the stock market had flourished thanks to the growing interest of investors but prices were more volatile and the volume was relatively small. The slight growth until 1925 was followed by a very volatile period in 1926 due to expectations of an unstable macroeconomic scenario, before returning to growth in 1927.

This recovery was a product of a turnaround in the macroeconomic expectations, fueled several reasons, in which stood out the decrease in half percent made by the Federal Reserve in the discount rate, which increased the demand for Government Bonds. Commercial banks and some investors who held those bonds saw this as a good opportunity to sell and transfer their funds to the stock market (Galbraith, 1954).

At the start of 1928, stock prices started rising more quickly. Just as with the real estate boom in the early 1920s, investors only needed an excuse to believe in something and at this time they were convinced the stock market would bring them great wealth with little effort. With this it started a new “gold rush”, with stocks going up 10 or 20 basis points a day, with the utility and new technologies leading the gains (Galbraith, 1954).

There was a frenzied increase in the volume of trading. In June 1928, the volume surpassed the utopian mark of 5.000.000 stocks, rising further to over 6.000.000 in November (Galbraith, 1954). Also, according to Galbraith (1954) and Allen (1931), that year sealed the beginning of the speculative bubble, more specifically in March.

Given the context of the 1920s and especially after 1927, it became evident that those investors would need support and this provided a new market to explore. Regulations of commercial banks from the 19th century limited the provision of long-term loans; however, this was overcome by setting up wholly-owned securities affiliates that were allowed to enter in all aspects of the investment banking and brokerage business. On the

other hand, investors without enough capital to purchase a diversified portfolio of stocks saw a new industry of services develop, namely investment trusts. New ways of investing in the stock market also emerged e.g. margin deposits and negotiation i.e. the buyer, with margin, contracted a loan to buy a given number of stocks which remained in the possession of the broker as a guaranty of the loan. The buyer benefited from any value increase and with the same and fixed loan value. Whereas the investor was anxious to invest and benefit from the constant increases in the market, this kind of service sent supplementary funds to the market (Galbraith, 1954; Sornette, 2003; White, 1990).

3.1.3. The year of 1929 and the Great Crash

The year of 1929 began with a lull on the market. Despite the fact that, according with Galbraith (1954), the volume of stocks traded in January exceeded 5.000.000 in five days, in February, the decrease in the UK reference rate slowed the rhythm of trading. The first glimpse of what was to happen in October came in March that year. On 25th March stock prices fell, and with the rate of broker loans increased to 14%; the following day a wave of fear resulted in a volume of 8.000.000 stocks traded. Prices plummeted and both investors and brokers were panicking. The interest rate on brokers' loans reached 20% and telegrams began arriving requesting the delivery of the guaranty deposits. It was only Charles Mitchell's announcement that the Federal Reserve was obliged to stop a possible crisis that brought this panic under control (Galbraith, 1954). The power of information was doing its work. Also, 1929 would be characterized by the extreme flow of information from the most diverse sources, trying to bring calm and confidence to the markets.

Brokers' loans indicate the degree of speculation on the market, and they were reaching high levels by this time. However, the interest rate indexed to these loans was more volatile than at other times. This dichotomy demonstrates the conflict of expectations. On one hand, investors believed that the market would continue to rise. On the other, brokers were more uncertain (White, 1990).

Market behavior was "normal" until August with days of trading seen as the last of the great 1920s. But despite this behavior, some macroeconomic indicators were telling a slightly different story. In July, the industrial production index reached a peak and went into decline the following months. The problem was based on the fact that the stock market only feels the effects of this context with some delay, and only when the investors and all the market becomes aware of the macroeconomic situation. But at this time, investors were still confident (Galbraith, 1954).

In September and October the market started to slow and Galbraith (1954) and Allen (1931) stated that September represented the end of the golden days. Nevertheless, investors' expectations about the future were still quite optimistic in early October.

Though high on October 15, the situation started to change on October 19. The news was that stock prices were falling and the guarantee margins were rising, which meant that prices were getting so low that they no longer represented the guarantee on the loans. On October 21 the market was unstable but losses were covered at the end of the day. Then a normal idea started to circulate: sell stocks and buy gold. Despite the announcement by bankers that the market was fine, on October 23 there were continued losses and this led to the pre-crash on October, 24 (Galbraith, 1954; Sornette, 2003).

On that day, the volume of stocks traded reached 12.900.000 and panic started to set in. Prices started to fall and most transactions were to sell stocks. The uncertainty fueling the panic was only controlled at midday when a group of bankers met to discuss what to do and how to save the stock market. They decided to gather resources. But it was what the bankers told to investors and not the resources that stabilized the market. The relieved investors started trading again so as to be part of the new wave of prices increases and by the end of the day the majority of the losses were compensated. The wave of confidence in bankers restored calm and everything was thought to be back to normal. It was now important not to miss the opportunity to buy stocks that were cheaper than ever (Galbraith, 1954; Sornette, 2003).

Despite the restored levels of confidence, October 28 started with losses and was a very difficult day on the market. The volume was high and most stocks were falling. The Dow Jones Industrial Average dropped almost 40 points and the volume was very high. The bankers met again but unlike the previous meeting, now the concern was how not to help the market without increasing the wave of panic. But it was clear the next day that this idea had not worked. The final loss that day, known as Black Tuesday, was a little lower than the previous day but combined all the bad characteristics of the previous days . The volume of trading hit a historic maximum of 16.410.000 stocks and the Dow Jones Industrial Average dropped almost 30 points. The main company stocks continued to fall and stocks of trust funds were going to zero value as the volume of brokers' loans was decreasing. The bankers were held responsible for this situation and the help they had promised before did not come that day. Panic and fear set in (Galbraith, 1954; White, 1990 and 2004). Over the next days, there was some recovery and the real goal was to restore confidence among the leading actors on the market. Despite a brief recovery in December, this did not happen. The margin calls decreased by 25% and the volume of brokers' loans also decreased. Some companies went bankrupt and trusts funds were seen as a negative factor to the recovery because their stocks were in steady decline and became unsellable by November (Allen, 1931; Galbraith, 1954; White, 1990).

3.2. Dot-com bubble of 2000

The 1990s was one of the most prosperous times in the USA economic and financial history. The good macroeconomic indicators, the bullish market, the launch of Internet and the advances in the technological and biotechnological sectors made Americans believe the future would be prosperous.

Like the 1920s, these general conditions - economy, market and the emergence and development of a new sector – generated expectations and beliefs surrounding a New Economy. However, unlike the 1920s when most investors were just discovering the possibilities of the financial markets, especially the stock market, in the 1990s much of the population was familiar with the market and saw it as an inherent to the normal functioning of the economy. More specifically, investors were more knowledgeable and were not restricted to a particular class. It was just as normal to have an asset portfolio or invest in the stock market as to buy the groceries or pay the bills. It is therefore no surprise that the majority of the investors were excited about new, potentially lucrative investments.

The investors' appetite was precisely satisfied with the emergence of two new sectors, namely internet and technological industries. They brought new scope to the market and, more importantly, new stocks. Soon, the hope for a New Economy was built around these companies and, like the utility sector in the 1920s, they fuelled the main channels of investment. Not surprisingly, this period of enthusiasm was followed by a speculative bubble.

3.2.1. The rise of web companies and the investor profile

In the early 1990s, the US macroeconomic environment was unstable. According to the FED (Federal Reserve), the US economy was in recession and inflation and unemployment rates were rising. As a result, the real *per capita* consumption in 1991 was the lowest that decade.

The recovery began in 1992, coinciding with the IPO (Initial Public Offer) of American Online, the first big internet company. This act (IPO) became commonplace in the following years, taking place in well-known companies like Yahoo, Amazon or E-Bay (Liu and Song, 2001).

However, following the American Online IPO, the internet only appeared in the news again in November 1993. But at that time, very few people were aware of this new industry and even fewer had access to it. However, the computer and the possibility of accessing internet had such a powerful effect on people's lives that it gradually acquired as much importance as television. The Worldwide web was even more attractive because the creation of each new application or site gave a sense of contributing to the country's economic growth (Shiller, 2000).

Investors' interest in the potential and opportunities of the web triggered an exponential growth in IPOs for web companies, and with time this resulted this same interest resulted in enormous P/E (price over earnings) values and stock returns. The subsequent bubble generated by the expansion of dot-com companies and later tech and bio-tech companies was essentially a consequence of investors' and the population's new and different mindset in relation to previous decades; like in the 1920s, they saw an opportunity to get rich with the minimum effort (Shiller, 2000).

Also, this desire for investment and wealth was reflected in cultural values. A successful business person became much more revered than a brilliant scientist or artist. The success cases in the financial markets allied to the bullish trend increasingly gave the impression that investing in stocks was a quick and easy way to get rich with little effort. But the market was not only driven by individual investors; the growth of pension plans and mutual funds were raising demand for stocks, particularly in tech and dot-com stocks which were growing at a frantic pace (Shiller, 2000).

Just as in the 1920s, the stock market seemed to offer general investor a world of opportunities and their excessive optimism gradually made him neglect the risks and believe the market was more predictable than it really was (Liu and Song, 2001; Shiller, 2000; White, 2004).

3.2.2. The speculative wave: Evolution of the Nasdaq, web and tech companies and investor behavior

The motives that triggered this situation must be analyzed before the Nasdaq speculative bubble can be understood. Clearly this speculative wave cannot be explained simply by the behavior of web companies. Like on the 1920s, an analysis of the companies' and investors' behaviors proves to be the most efficient approach.

The speculative wave that was seen on the Nasdaq Composite Index in the late 1990s is mostly associated with the huge surge in IPO's, the dramatic rise in web companies' stock prices as well as the interest and expectations of the investors in this sector. Let's look at an example. The Nasdaq Composite Index rose from 755 points at the beginning of 1995 to 5.000 points in March 2000, i.e. a valorization of 522%. The speculative bubble can also be isolated and seen at the end of 1998 and beginning of 1999 when the return rates of the Nasdaq frequently reached values of over 10% (Liu and Song, 2001; Sornette, 2003).

This evolution can be analyzed in two phases. Firstly until 1997, it was almost entirely explained by the rise of the sector and the expectations and beliefs generated among investors that this was the sector of the future. These expectations changed the natural course of the market, triggering an abnormal demand for web stocks and thus a dramatic rise in prices. Secondly, after 1998 in particular, the market and companies reacted and responded to this situation (Shiller, 2000).

Companies that were already part of the index at this time like Yahoo and e-Bay, were successful and

improving their results, prices and market share and were therefore giving investors the right and expected signs. This was the ideal environment for more companies to join the market, though in many cases it proved too soon.

This rush to the market by web and tech companies, many of which had been operating for only a few years or months, boosted offer and gave investors the impression that the market was developing fast, driving them to buy more and more stocks, often speculatively, bringing a wave of money into the market. However, even though many of these new companies in the IPO process were not as strong as the stock price reflected, their price was rising every day. In conclusion, the illusion that was created of the sector and the market was unfounded. So why did these companies start the IPO process and why did they enter the market so soon?

The answer to both these questions lies with the investor. Firstly, internet and tech stocks were irrationally overpriced. The recent performance of these companies on the market and the future growth prospects combined with investors' beliefs made some young companies precipitate entry into the market to take advantage of these high prices. However, their stock prices reflected investors' beliefs and expectations for the sector rather than their actual performance because these companies were too young and their financial structure was unbalanced. In time the stocks of the entire sector and index became overvalued (Liu and Song, 2001; Schultz and Zaman, 2000).

A second explanation was the rush to grab market share. In an industry with enormous potential, an IPO provides capital for a company to invest in marketing and R&D even after losing money for several quarters as well as the possibility to acquire other companies and improve market share. As a result, the increase in the market share brings economies of scale, implying lower costs and more efficient development; in the mid-term, this improves the results of the company and its stock price (Liu and Song, 2001; Schultz and Zaman, 2000).

It can be concluded that the market changed mostly because of investors' positive expectations for internet and tech companies, which tried to gain from this by going to the market thus giving investors the impression of an expanding sector. It did not take long before stock prices ceased to be based on fundamentals, but on the beliefs of companies and investors; jointly, they caused stock prices to rise and this brought benefits to both parties.

However, the volume of short selling clearly demonstrates that as the Nasdaq improved, overpricing and speculation increased. For example, an average web firm in 2000 had almost 6 times as much of its public float shorted (Hand, 2000).

3.2.3. The year of 2000 and the Crash[1]

In 2000, tech and dot-com stocks were still increasing despite forecasts of a rise in the interest rate. When all other indices were decreasing, on January 4 the Nasdaq rose to a record of more than 4.000 points.

But Nasdaq's volatility and fragility of started to be seen on January 7 when Lucent Technologies, a maker of telephone equipment, warned about lower than expected profits and sales. After this announcement, investors started the typical strategy of rotating new and old economy stocks in their portfolio. This gathered pace and by April became more frenzied and was closely linked with rising levels of myopic risk aversion. However, the channeling of almost all available money (such as dividend and tax gains) to dot-com and tech stocks continued as the falls in the market were seen as normal corrections.

But this time, some analysts were underestimating investors' strength and power, and continued to believe that fundamentals were strongest than psychological moods. Nevertheless, most financial analysts were avoiding dot-com stocks.

Moreover, the volume of short selling remained high, with an average of 2.4 billion shares shorted; this indicated a strong bearish mood among the aggressive group of investors even though Nasdaq stocks rose 2 or even 3 digits.

Another curious circumstance was the increase in the return rates of Nasdaq stocks despite rises in treasury bonds yields (which went from 4,8% in 1998 to 6,3% in 2000), indicating the possible speculative effect, once the investors' expectations, based on the good past performance of stocks and in the expected high consumption on the sectors to which they give support, were skewing their predictions.

The investment fever continued in February with investors holding record credit in margin debt trade. It is interesting to note that the last time there had been such a high volume of credit in the hands of investors was precisely in September 1987 - the month before Black Monday, 1987.

In March, just before the fall, Nasdaq rose to a record of 5.000 points, up from 3.000 points just four months earlier.. However, greater returns bring a serious increase in volatility, which ultimately increases the risk and consequently the costs of margin debt.

In March 10, Nasdaq reached the 5.000 point mark for the last time. In the 3 following days, it registered point drops, setting the index at 4.500 points on the so-called "correction days".

Allied to these situations, the FED began to express concern about the over-speculation on the market, indicating that these new economy companies were too

[1] This point 2.3 was made essentially using news from the economy and markets section of the New York Times and New York Daily News, from January to April

dependent on the old economy and therefore a risk to the economy.

The anxiety and uncertainty started to spread significantly when on March, 20 the Nasdaq recorded its biggest historical percentage loss, though was later exceeded negatively on March 30. Even the most optimistic investor began to question whether such a large number of corrections in such short period of time was normal?

The events of April confirmed that this was not a period of correction or adjustment but the burst of a bubble. After all, the traditional laws of economy applied to the Nasdaq. Whereas on March 10 the Nasdaq was 24% higher than January, in April the gain was only 12%. Almost immediately the margin debt rates started to increase even more than in March and the credit lenders became more suspicious about Nasdaq's behavior and future. Also, the mutation and roll-over in the composition of individual portfolios became more common, particularly among naive investors who bought tech and dot-com stocks just because they were rising.

Even for the more skeptical investor, the worst was confirmed on April 4. The market opened with countless sell orders which rapidly led to an almost 14% fall in the Nasdaq; and the volume for historical records, all of this on a day without any significant bad news. The market only recovered when the rumor that hedge funds were buying stocks and bringing liquidity to the market began to circulate, but the panic was already installed. Fear was abated when the day closed only 2% down.

The market volatility was beginning to hit most investors who were losing capital and running out of cash to cover their losses. On the other hand, some aggressive investors saw this as a unique opportunity to buy stocks and to gain with the possible recuperation. However, even the biggest tech and dot-com companies were announcing losses and the commercial banks began to refuse money to invest in dot-com stocks, starting a run on convertible bonds.

But the decline continued and by April 12 the Nasdaq had already lost more than 25% since its peak, closing the day more than 7% down at 3.769 points - the lowest at close of day since January. Already fully aware of what was happening, on April 14 Nasdaq recorded its biggest 1 day loss ever, down more than 10% to 3.321 points. The week closed with a 7 day fall of 25%, the worst week in its history.

The Nasdaq never again reached the levels witnessed in this period and it continued to fall for a few more months. In addition, the USA subsequently went into macroeconomic recession and innumerous tech and dot-com companies went bankrupt.

4. An Iterated Prisoners Dilemma approach

This research addresses investor behavior in the stock market in extreme situations of speculation and crash. In this context, the game chosen to obtain a significant explanatory efficiency must contain more than two individuals and, in this case, a finite but indeterminate number of investors. Despite the possibility of constructing a model with players acting either individually or as different groups, it was chosen the first alternative (N players acting individually) because joining individuals in groups can be complex given the need to access an enormous amount of information in order to form groups with a higher percentage of similarities.

Moreover, a one period static model was discarded due to its lack of efficiency in situations in which behaviors and actions tend to evolve over time and in response to the actions of adversaries. Therefore, a game was created with T infinite periods of time. However, the temporal horizon of the present game will be comprehended between $[0+d;T-d]$, $d \neq 0$, implying the analysis of a sub-game. This procedure was selected because the objective is to focus on the speculation and crash periods, which are only a fraction of the time T. Hence, there are d periods of time before the speculation period and d periods after the crash, which implies that the game has a partial but not a final result, because the game itself will continuously evolve to other states after the end of the sub-game analyzed.

The aim is also to analyze the appearance of both cooperative and non cooperative behaviors as the game matures, what excludes games that do not consider the possibly of an evolution in aggregate behaviors and subsequent equilibriums. The context of information is asymmetric and imperfect; it is perceived and used gradually by the players, which does not imply *à priori* that they have advantage over the others. Thus, allied to this, the game is sequential because the investors do not act at the exactly same period of time, opening possibilities for the application of strategies that mutate in response to other players' actions.

Finally, it is assumed that the investor is not fully rational; this implies that despite the prior objective of optimizing his results, his actions can lead him to inefficient outcomes. Accordingly, the following game will not be based on a payoff function that expresses the result of the game for the player, but on a function that will explain the incentive to cooperate and defect.

Moreover, as investor preferences are not stable and rigid, actions can vary considerably in different time periods, making preferences closer to a hyperbolic function which considers the possibility of preference reversal.

Given the abovementioned assumptions, an Iterated Prisoner's Dilemma game (IPD) was selected and applied to N players and for the given temporal horizon mentioned, with non zero sum result, which indicates that the benefits and incentives to cooperate are not necessarily the same for defecting.

In the basic form (for 2 players), the IPD assumes that each player can choose to cooperate or defect; the game can be repeated or iterated as many times as needed in a sequential fashion, implying that the

strategies used can mutate according with each player's previous action. It is important to note that the players do not know the length of the game, thus invalidating an end behavior effect which may arise in super-games with finite time periods (Selten and Stoecker, 1986).

Thus, the game can be presented in the following matrix form.

Table 1: Standard Payoff Matrix (canonical form) of the IPD for 2 players

	Cooperate	**Defect**
Cooperate	R R	T S
Defect	S T	P P

In addition, the game will only be an IPD if the following assumptions are respected:

a) $T > R > P > S$;

b) $R > \frac{S+T}{2}$

It is assumed that circumstantial cooperative equilibriums may occur, but that these will not be dominant and stable (Aumann, 1959). As the number of iterations increase, a Nash Equilibrium can be reached but only if the players have monotonic preferences, which is easier to achieve with 2 players.

However, the 2 player form of the game is considered to be reductive because, when dealing with real-life situations, more realistic results can be obtained with an N player game (Davis *et al*, 1976).

It has been therefore selected the more usual approach, namely an N player IPD, which implies the following:

- Each player faces two choices: cooperation or defection;
- The defection (D) is a dominant pure strategy for each player and it will be better if he always choose that option;
- The equilibriums achieved are not stable in some cases, principally in cases of cooperation (C).

Thus, the game can be presented as follows.

Table 2: Matrix presentation of the IPD for N players

Nº of Cooperators	0	1	…	X	…	N-1
Cooperate	C_0	C_1	…	C_x	…	C_{N-1}
Defect	D_0	D_1	…	D_x	…	D_{N-1}

As in the 2 player form, with N players the game will only be an IPD if the following conditions are achieved:

a) $D_x > C_x$ for $0 \leq x \leq N-1$

b) $D_{x+1} > D_x$ and $C_{x+1} > C_x$ for $0 \leq x < N-1$

c) $C_x > (D_x+C_{x-1})/2$ for $0 < x < N-1$

In a concrete model, the C_{N-1} and D_{N-1}will be payoff functions that translate the incentive to actions of cooperation and/or defection.

One of the important features of this model is the possibility of mutation in the behavior and actions made across the game (dealing only with pure alternatives) and this can be expressed with resource to the strategies used.

Nevertheless, players in the real world do not know the actions taken by others in real time, and there is a delay that can be caused by innumerous factors. Hence, the investor only knows the adversary moves with a ⌊ period delay, improving his knowledge of the game with time (memory), i.e. he will learn as the game evolves. This learning ability is a very important factor to avoid the possibility of superrational players. Moreover, even with the premise of learning and delay, some mistakes can be made by the player because he is unable to process all the available information and therefore selects information using anchors and heuristics; this can lead to judgment errors.

4.1. The application of the ITD to investor behavior in extreme financial events

After contextualizing the model to be applied (IPD), on the following pages it is defined the problem and the parameters of the model before analyzing the results obtained.

4.1.1. Problem definition

The above mentioned examples of financial crashes of 1929 and 2000 demonstrate that the investor faces two distinct situations. First, the context of a speculative bubble whereupon the investor increased or maintained his positions in overvalued stocks, especially in companies belonging to the new sector of the time (utility in the 1920s and dot-com and new technologies in the late 1990s). At that time he was being driven both by the desire to maximize profits and emotional considerations like euphoria and mania. Also, despite the short duration of the speculative bubble generated, the returns and the volatility implied in the stocks is linked more to increasing demand by investors than to other factors, thus suggesting a more deterministic trend in these periods than others in which the random walk prevails.

Secondly, there are crash situations and these have a different profile. Unlike the bubble context, the investor tries to avoid losses at all cost. However, this feature does not appear unexpectedly so the transition made between speculation and crash is not sudden. In the months before the crash in 1929 and 2000, a market scenario compounded with more volatility and rising but more unstable trend in prices was observed. It can be also assumed that some investors were starting to

have doubts about the real value of the stocks in their portfolios by this time. However, initially the defection from these positions was made by a minority; most investors started to sell their positions at the moment of the crash, thus decreasing the liquidity and increasing the volatility on the market.

Given these two contexts, the problem to be applied to the IPD can be presented as follows: the investor has two choices, both on speculation and crash. He can cooperate with the rest of the investors to maintain the speculative bubble and the rising trend, or he can defect and invest in other kind of assets, which means that he is not interested in maintaining the situation of speculation. Thus, the players in the game can take the following actions:

- **Cooperation (C)**: can be seen as a collaboration between investors to maintain (even if unconsciously) the speculative bubble by investing or sustaining positions on stocks that are overvalued;
- **Defection (D)**: logically this is the opposite situation, i.e., the investor is not interested in maintaining the bubble and thus takes two possible actions: the investor does not want to invest in this kind of stock, or he has these stocks but does not want to maintain his position and therefore sells them and does not support the trend.

Defection can also result in the possibility of leaving the game or remaining but with positions in different assets.

4.1.2. The formalization of the game

(a) Players

The number of players in the game is indeterminate but is a finite set of dimension N. Also, for each player $i \in N$ is a nonempty set A_i of actions available that are pure: cooperation or defection. The players have the following characteristics:

- They are not fully rational, acting more in line with the Simon's (1955) postulate of limited rationality which implies that at some time in the game they may not optimize their actions; this may lead to the maintenance of long periods of cooperation;
- They do not have monotonic preferences or a stable set of preferences, which means that there is not a relation $\succ_i$ on A. Instead they may have hyperbolic preferences which allows for mutation in the preference set and reversal of a preference A over preference B in a given time period T_i;
- The risk profile of the players/investors respects the Prospect Theory utility function, which means that they are not always risk averse. The degree will depend on and vary according to whether they are dealing with gains or losses.
- Although a player can exit the game when he defects, however that is not a dominant attitude across all group members, i.e., when they defect, they can still be part of the game but with investments in other assets, or they can even observe and then enter again at a later time.

(b) Time

This game considers an infinite time period T. However, the game begins a few periods before the beginning of the bubble and continues when this analysis ends, evolving in more *d* time periods. This also implies that the model occurs in a sub-game; this does not represent a problem because, as argued by Friedman (1991), a game that begins in a given time period that does not coincide with the time period T_0 may have all the same characteristics as a game and realize the same *equilibria*.

(c) Payoff or incentive function

The payoff function in the IPD works as a mathematical translation of the incentive to cooperate. Thus, there are two payoff functions: one for cooperation and another for defection. However, neither of them is static and stable, varying with the number of cooperators, which also varies according to the strategies used by each player.

The payoff functions are denominated C_x and D_x and belong to a space set N of the number of cooperators between $\{0,1,2,\ldots,N-1\}$

(d) Information and strategies

The game is played in an imperfect information context in which it is assumed that players have to make decisions at several moments without knowing all the game history or the adversaries' choices (Fiani, 2004). However, as the investor is not fully rational, it is implied that even decisions made in a perfect information context would not be supported by all known information because of their cognitive limitations.

In addition, players' actions are supported by the use of Tit-for-Tat strategy. However, because the game begins at $0+d$ time periods, it is impossible to know when the first move of defection really happened. Nevertheless, it is important to say that players take the choices made by adversaries into consideration, but with a $\lfloor$ lag period which is not standard for all players. This must be assumed because, without lagging, the game was on a short period stabilized on a Nash Equilibrium of defection (and therefore the preferences monotonic). Also, they will not remember all the previous moves from the lag period because of the amount of information, and it is therefore assumed that only a few moves prior to the lag period will be remembered.

4.1.3. The model and results

The initial problem investor's face is to preserve the speculative bubble, maintaining their positions or

investing in assets that are overvalued, which implies taking cooperative actions. However, as time evolves, more investors will share similar investment decisions and this implies a stronger incentive to cooperate than to defect. But some investors realize the situation can be unstable as the bubble matures, and see defection as a more appealing incentive.

Given this problem, the following payoff functions can be applied, proposed by Seo, Cho and Yao (2000).

$$C_x = \frac{1}{2}x^2 - k \; for \; 0 \leq x \leq N \quad (3)$$

$$D_x = \sqrt{2x} \; for \; 0 \leq x \leq N \quad (4)$$

Note that x expresses the number of cooperators in the game in each time period T_i and k is an unknown variable that indicates the exogenous incentive to defection, which can include any type of information (even a crucial piece that can trigger the crash), being seized only by some players. However, they will start to defect when this parameter k becomes too big, making the incentive to cooperate smaller than the respective one to defect.

Also, as referred, the players make use of a Tit-for-Tat strategy, which implies that the decisions made will take the actions taken by other players into consideration, but only a small number of moves and with temporal lag.

$$l = \; y, for \; all \; T_i \quad (5)$$

$$m = (\omega * T) - l, for \; all \; T_i \quad (6)$$

Therefore, based on the idea of Axelrod (1987), the equation (5) indicates the number of lag periods in all the time T_i period, and equation (6) provides a quantification of how many periods of information a player can remember prior to the lag, where ω (a constant) is the percentage of the fullness of periods prior to the lag .

However, the players will only remember the moves made by some adversaries, mainly because it is a context of incomplete information and the investor cannot assimilate an unreal amount of information.

Thus, despite the natural dominance of defection, as shown in Figure 3, with a small k, as the number of cooperators grows, the incentive to cooperate will increase more quickly than to defect, thus implying an intersection between both incentives at a given point in time in which cooperation thereafter becomes more appealing.

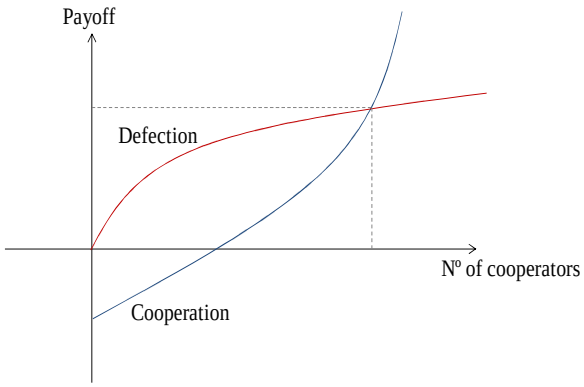


Figure 3: The evolution of the payoff result for Cooperation and Defection across a growing number of cooperators

Therefore, the game can be analyzed in two distinct parts.

The first begins on the initial intersection point described in Figure 4 as point (1). The k value at this time is small and investors are making use of a Tit-for-Tat strategy. As mentioned, the moment at which the game starts is not relevant to the present study. Thus, the period starting at point (1) is linked to sentiments of euphoria and mania, as seen in the 1929 and 2000 bubbles, when an increasing number of investors were investing in overvalued positions.

The number of cooperators starts rising rapidly and this may be due to the desire to make gains with stocks belonging mostly to sectors of the new economy, as seen with the utility sector in 1929 and dot-com and technologies in 2000. Thus, as the existing cooperators maintain their positions, partially due to the strategy being used, new players begin cooperating so as to enjoy the evident returns. The rising trend in cooperating players causes the speculative bubble to grow, as does the incentive that sustains it.

This scenario can be characterized as a minimal equilibrium of cooperation because the number of cooperators is bigger than that of defectors, and the growth of the incentive to cooperate is more accentuated than the one linked to defection. Hence, with the maintenance of a low k, the number of cooperators continues to rise to a point at which the equilibrium reaches its strongest position (point (2) on Figure 4). Thus, the peak of the speculative bubble (or minimal equilibrium of cooperation) in point (2) coincides with moments seen in the two events previously analyzed. In the Great Crash of 1929, it refers to late 1928 when the volume of stocks traded exceeded utopian marks for that time of 5 and 6 million stocks. In the crash of 2000, this is the moment when Nasdaq reached 5.000 points.

Thereafter, the equilibrium becomes more unstable. In the events described, the markets became more unstable and volatile after the peak, and investors, banks and states became more anxious and nervous. The irregularity of the market can be seen as a result of a rising number of investors starting to defect. The

explanation for this defection may reside in the value of parameter k. This parameter contains pieces of information that indicate that the bubble is not stable and that it is better to start leaving the positions held on overvalued assets before a stressful drop in the market. However, this information is only perceived by some investors. Thus, some cease to cooperate and the equilibrium becomes more volatile.

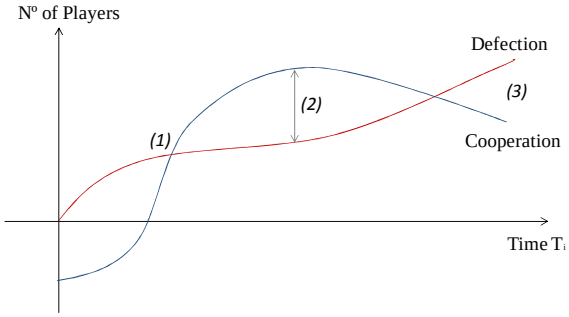


Figure 4: The evolution of the number of players cooperating or defecting across time: (1) indicates the beginning of the minimal equilibrium of cooperation; (2) Point in time at which the minimal equilibrium of cooperation is strongest; (3) Beginning of the minimal equilibrium of defection which will become stronger but not stable as a solution of the game

This moment marks the beginning of the end of that period. As more players become aware of the key pieces of information, they start to see the cooperative payoff diminishing at a faster pace than that of defection; this is explained by the value of k, which is reducing the incentive. There is a causality relation in which more k implies less cooperation and therefore a downward trend in the cooperative incentive.

As the players are using a Tit-For-Tat strategy, they start to realize some defective actions in some other players and therefore also start also to replicate their actions and choose defection; this implies a reversal of preferences that had been relatively stable for a long period of time (period coincident with speculation). This implies another intersection between the number of players cooperating and defecting. This intersection will leads to moment (3) and can be associated to the beginning of the crash.

As referred, the moment after the peak of the bubble is related to an environment of increasing volatility and anxiety among market agents. The more unstable variations of the market can be interpreted as more defective actions made by players. Hence, the game reaches a new equilibrium when more players are defecting than cooperating; this is a minimal defection equilibrium which coincides with the moments related with the crash on the market. As more players defect, the incentives decline but the incentive to cooperate becomes smaller more quickly than to defect. Thus, the gain of stability in the equilibrium coincides with the fall of the market, implying also domination over the previous one.

Nevertheless, this equilibrium will not be stable for a long period of time because investors do not have a rigid set of preferences, implying the inexistence of a Nash point of equilibrium. However, in future time periods the dominating equilibrium will evolve to other types of state.

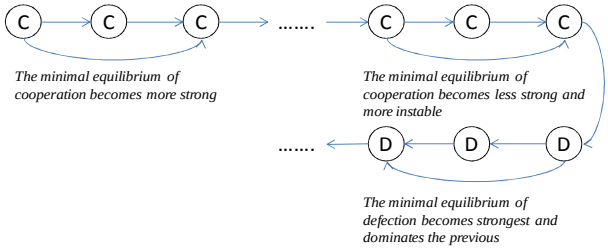


Figure 5: An illustrative diagram of transition between the equilibriums of the game

As established in Figure 5, the problem has three phases. The first involves the growth of the minimal equilibrium of cooperation, coinciding with the expansion of the speculation period. The second period is characterized by increasing instability on the markets, associated to more defective actions made by investors and making the equilibrium volatile. The final period marks the transition between equilibriums and coincides with the crash on the markets and implies the choice to defect dominates for most players.

5. Concluding notes

This research aimed to address the behavior of investors in extreme situations on the stock market. The main purpose was to shed light on some features of the investor profile in these situations so as to understand actions taken individually and as a group.

Thus, to obtain a realistic investor profile in those situations, the main features of the investor were approached in a number of ways. First, the postulate of rationality was analyzed and the existing literature on asset pricing and portfolio theory was set against the ideas of behaviorists and neuroscientists; they confirm first that the economic agent is not fully rational, being maybe closer to Simon's (1955) notion of limited rationality and second, that feasible models and theories can be constructed which do not take pure rationality as the central premise the. In addition to behavioral economics and finance, the evolutionism approach and neuroeconomics also seem to agree that the utilitarian agent is not the most efficient way to address some problems.

The decision-making process is another key feature, particularly with regard the investor's set of preferences. If the options offered in the real world are systematically changing and the actual individual has a dynamic and mutated set of preferences which may be biased by the influence of limited rationality and information processing, then the process of decision making has to be more complex than shown and derived by some theories. Considerations about the utility

function and the set of preferences, the degree of risk aversion or the process of accounting gains and losses by the economic agent become so complex that it is difficult to resume all in simple axioms. However, it was showed that the complexity of the agent's profile implies that new models, and the one analyzed in this study, have to take this kind of consideration into account.

Given the above, the model generated through the Iterated Prisoner's Dilemma gives rise to the hypothesis of the existence of minimal equilibriums of cooperation (in the speculative bubble event) and defection (at previous moments and during the crash period), which are not stable because of players' perceptions of key information and their impact on the set of preferences. However, it was found that there equilibriums play a successive role if the aim is to see the game as a whole and not as a sub-game. It should also be stressed that when individuals act as a group, they do not secure cooperation, and self-interest is a key factor in the decisions made. Nevertheless, investors' motivations in the speculative bubble event can degenerate on attitudes of cooperation, as commonly seen at times on commons tragedies in natural resources. In conclusion, the proposed hypothesis launched was corroborated by the model used. However, the results can vary in line with the type of event analyzed.

References

[1] Adams, J. T. (1931), *Epic of America*, Little, Brown, and Company;

[2] Ainslee, G. (1975), Specious Reward: A behavioral theory of impulsiveness and impulse control, *Psychological Bull Vol. 82, pp.* 463-496

[3] Ainslee, G. (1991), Derivation of "Rational" economic behavior from Hyperbolic Discount Curves, *American Economic Review Vol. 81, pp.* 334-340

[4] Aldrich, H.E., G.M. Hogdson, D.L. Hull, T. Knudsen, J. Mokyr and V.J. Vanberg (2008), In Defense of Generalized Darwinism, *Journal of Evolutionary Economics Vol. 18 No. 5, pp.* 577-596

[5] Allen, F. L. (1931), *Only yesterday, an Informal History of the Nineteen-Twenties*, Harper and Row, New York

[6] Antonides, G. and S. R. Wunderink (2001), Subjective time preference and willingness to pay for an energy-saving durable good, *Zeitschrift Fur Sozialpsychologie Vol. 32 No. 3, pp.* 133-141

[7] Arrow, K.J. (1987), Economic theory and the hypothesis of rationality, in Eatwell, J.M., M. Milgate and P. Newman (Eds.), *The New Palgrave: a dictionary of Economics*, London: Macmillian, pp. 69-74

[8] Aumann, R. J. (1959), Acceptable points in general cooperative n-person games, in R. D. Luce and A. W. Tucker, (Eds.), *Contributions to the Theory of Games IV - Annals of Mathematics Study*, 40, Princeton University Press, Princeton New Jersey, pp. 287-324

[9] Axelrod, R. (1987), The evolution of strategies in the Iterated Prisoner's Dilemma, in L. Davis (Ed.), *Genetic Algorithms and Simulated Annealing*, Morgan Kaufmann, San Mateo, pp. 32-41

[10] Bechara, A., A. Damásio, H. Damásio and S.W. Anderson (1994), Insensitivity to future consequences following damage to human prefrontal cortex, *Cognition Vol. 50, pp.* 7-12

[11] Bechara, A., H. Damásio, D. Tranel and A. Damásio, (1996), Failure to respond autonomically to anticipate the future outcomes following damage to prefrontal cortex, *Cognition Vol. 6, pp.* 215-225

[12] Bechara, A., H. Damásio, D. Tranel and A. Damásio, (1997), Deciding advantageously before knowing the advantageous strategy, *Science Vol. 275, pp.* 1293-1295

[13] Benartzi, S., and R. Thaler (1995), Myopic Loss aversion and the equity premium puzzle, *The Quarterly Journal of Economics Vol. 110 No. 1, pp.* 73-92

[14] Berenson, A. (2000), Investing – Technology defies another dictum, *New York Times*, February 6

[15] Berenson, A. (2000), A record volume as stocks surge on the big board, *New York Times*, March 17

[16] Berenson, A. (2000), As some look for more fireworks, the Dow takes a dip, *New York Times*, March 18

[17] Berenson, A. (2000), So much for the theory that old-economy stocks are finally regaining their place in the sun, *New York Times*, March 23

[18] Blaug, M. (1992), *The Methodology of Economics, or How economists explain*, Cambridge University Press

[19] Bodie, Z., A. Kane and A.J. Marcus (2009), *Investments*, McGraw – Hill (8th edition)

[20] Buttman, J. and R.F. Allegri (2001), A Cognição Social e o Córtex Cerebral, *Psicologia: Reflexão e Crítica Vol. 14 No. 2, pp.* 275-279

[21] Camerer, C. F, G. Loewenstein and M. Rabin (2004), *Advances in Behavioral Economics*, Princeton University Press, New Jersey

[22] Campbell, J. Y. and J. Ammer (1993), What moves the stock and bond markets? A variance decomposition for long-term asset returns. *Journal of Finance* Vol. 48, pp. 3-37

[23] Carlson, M. (2006), A Brief History of the 1987 Stock Market Crash with a Discussion of the Federal Reserve Response, *Finance and Economic Discussion Series, Division of Research & Statistics and Monetary Affairs Federal Reserve Board*, pp. 2-24

[24] Chandler, A. D. Jr. (1977), *The Visible Hand: The Managerial Resolution in American Business*, Harvard University Press

[25] Chapman, G. B., and S. A. Elstein (1995), Valuing the future – Temporal Discounting of Health and Money, *Medical Decision Making Vol. 15 No. 4, pp.* 373-386

[26] Chapman, G. B. (1996), Temporal discounting and utility for health and money, *Journal of Experimental Psychology Learning Memory and Cognition Vol. 22 No. 3, pp.* 771-791
[27] Damásio, A. (1994), *O Erro de Descartes – Emoção, Razão e Cérebro Humano,* Publicações Europa-América, Mem Martins
[28] Danthine, J. P. and J. B. Donaldson (2005), *Intermediate Financial Theory*. Elsevier Academic Press, (2nd Edition)
[29] Davis, J. H., P. R. Laughlin and S. S. Komorita (1976), The social psychology of small groups, *Annual Review of Psychology Vol. 27, pp.* 501-542
[30] De Bondt, W.F.M. (1998), A portrait of the individual investor, *European Economic Review Vol. 42, pp.* 831-844
[31] Donovan, A. (2000), The neediest cases: A wider reach for Wall-Street's Boom, *New York Times,* January 26
[32] Duncker, K. (1945), On problem solving, *Psychological Monographs Vol. 58 No. 5*
[33] El-Faizy, M. (2000), Shares to be doled out privately, but sale to public still year away, *New York Daily News,* January 3
[34] El-Faizy, M. (2000), Techs keep Nasdaq rolling and renewed rate fears ruin Dow's new year, *New York Daily News,* January 4
[35] El-Faizy, M. (2000), New dose of interest-rate fears fuels markets' flop, *New York Daily News,* January 25
[36] El-Faizy, M. (2000), Nasdaq down, Dow up with focus on the FED, *New York Daily News,* March 21
[37] El-Faizy, M. (2000), FED hike can't halt stocks, Dow, Nasdaq soar despite quarter-point rise in short-term rates, *New York Daily News*, March 22
[38] El-Faizy, M. (2000), Dow, Nasdaq hurdle downhill, suffer biggest 1-day point loss, *New York Daily News,* April 15
[39] Fiani, R. (2004), *Teoria dos Jogos para cursos de administração e economia*, Elsevier
[40] Fleming, J., C. Kirby and B. Ostdiek (1998) Information and volatility linkages in the stock, bond, and money markets. *Journal of Financial Economics* Vol. 49, pp. 111-137
[41] Friedman, J. W. (1991), *Game theory with applications to economics*, Oxford University Press, London (2nd Edition)
[42] Galbraith, J. K. (1954), *The Great Crash 1929,* Houghton, Mifflin Company
[43] Gilpin, K. N. (2000), Demand for Technology shares drives Nasdaq to record, *New York Times,* February 8
[44] Gilpin, K. N. (2000), Shares close lower on declines by Microsoft and Banks, *New York Times,* March 28
[45] Gilpin, K. N. (2000), Shares prices fall as analyst reallocates model portfolio, *New York Times,* March 29
[46] Gilpin, K. N. (2000), Nasdaq posts its first gain of the week, *New York Times,* April 1
[47] Gilpin, K. N. and M. Richtel (2000), Nasdaq goes into free fall before making a recovery, *New York Times,* April 5
[48] Gilpin, K. N. (2000), April 23-29: Ritalin for Nasdaq?, *New York Times,* April 30
[49] Goldstein, D.G. and G. Gigerenzer (2002), Models of Ecological Rationality: The Recognition Heuristic, *Psychological Review Vol. 109 No. 1, pp.* 75-90
[50] Green, L., J. Myerson and E. McFadden (1997), Rate of temporal discounting decreases with amount of reward, *Memory and Cognition Vol. 25 No.* 5, pp. 715-723
[51] Gross, D. (2000), Desperation of the Dot-coms, *New York Times,* April 7
[52] Hand, J. R. M. (2000), The role of book income, web traffic, and supply and demand in the pricing of U.S. Internet Stocks, *European Financial Review Vol. 5 No. 3, pp. 295-317*
[53] Hershey Jr., R. D. (2000), The Markets: Investors Flee Technology; Dow rises and Nasdaq falls, *New York Times,* January 7
[54] Hershey Jr., R. D. (2000), Nasdaq surges as Dow drops after Greenspan comments, *New York Times,* February 18
[55] Hershey Jr., R. D. (2000), A Nasdaq correction, now back to business, *New York Times,* March 19
[56] Hershey Jr., R. D. (2000), Flight from technology bolsters pillars of old economy, *New York Times,* March 30
[57] Holt, C. A., and S. K. Laury (2002), Risk Aversion and Incentive Effects, *The American Economic Review Vol. 92 No. 5, pp.* 1644-1654
[58] Huang, C., and R. H. Litzenberger (1988), *Foundations for Financial Economics,* Prentice – Hall
[59] Levy J. S. (1992), An Introduction to Prospect Theory, *International Society of Political Psychology Vol. 13 No. 2, pp.* 171-186
[60] Lintner, J. (1965), The valuation of Risk assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, *Review of Economics and Statistics Vol. 47 No. 1, pp.* 13-37
[61] Kahneman, D. and A. Tversky (1979), Prospect Theory: An analysis of decision under risk, *Econometrica Vol. 47, pp.* 263-291
[62] Kahneman, D. and A. Tversky (1984), Choices, Values and Frames, *American Psychologist Vol. 39 No. 4, pp.* 341-350
[63] Kahneman, D. and M. W. Riepe (1998), Aspects of Investor Psychology: beliefs, preferences, and biases investment advisors should know about, *Journal of Portfolio Management Vol. 24 No. 4, pp.* 52-65
[64] Kahneman, D. (2002), A perspective on Judgment and Choice, *American Psychologist Vol. 58 No. 9, pp.* 697-720
[65] Kindleberger, C. P. and R. Z. Aliber (2005), *Manias, Panics and Crashes: A History of Financial Crises,* Palgrave Macmillan
[66] Koenigs, M. L., R. Young, D. Adolphs, F. Tranel, M. Cushman, M. Hauser and A. Damásio (2007),

Damage to the prefrontal cortex increases utilitarian moral judgments, *Nature Vol. 446, pp.* 908-911

[67] Laury, S. K., and C. A. Holt (2000), Further reflections on Prospect Theory, *Discussion Paper,* Georgia University Press

[68] Liu, Q. and F. Song (2001), The Rise and Fall of Internet stocks: Should Financial Analysts be Blamed, *SSRN Working Paper Series*

[69] Loewenstein, G. F. (1988), Frames of Mind in Intertemporal Choice, *Management Science Vol. 34 No. 2, pp.* 200-214

[70] Luce, R. D. and H. Raiffa (1957), *Games and decisions,* Wiley, New York

[71] Madrick, J. (2000), When Greenspan speaks and the stock market doesn't listen, the loss of credibility is dangerous, *New York Times,* March 16

[72] Markowitz, H. (1952), Portfolio Selection, *The Journal of Finance Vol. 7 No. 1*, pp. 77-91

[73] Marino, V. (2000), A Technology tide in Europe lifted International Funds, *New York Times,* April 9

[74] Mayr, E. (1988), *Toward a New Philosophy of Biology – Observations of an Evolutionist*, Harvard University Press, London

[75] McDonald, R. and D. Siegel (1986), The Value of waiting to invest, *The Quarterly Journal of Economics, pp.* 707-726

[76] Michel-Kerjan, E., and P. Slovic (2010), *The Irrational Economist, Making decisions in a dangerous world,* Public Affairs, New York

[77] Morgenson, G. (2000), Investing longtime best bet is being trampled by the bulls, *New York Times,* January 15

[78] Morgenson, G. (2000), The FED has no surprises and the market has a reason to rally, *New York Times,* March 22

[79] Morgenson, G. (2000), Buying on margin becomes a habit, investors turn to credit in a bull market, *New York Times,* March 24

[80] Morgenson, G. (2000), At Waterloo, a Technology Stock's retreat, *New York Times,* April 16

[81] Morgenson, G. (2000), Forces of nature, immutable truths, *New York Times,* April 02

[82] Norris, F. (2000), Nasdaq market at 5000: Volatility as way of life, *New York Times,* March 10

[83] Norris, F. (2000), A hard fall as a highflier revises figures, *New York Times,* March 21

[84] Norris, F. (2000), Modigliani's message: it's a bubble, and bubbles will burst, *New York Times,* March 31

[85] Norris, F. (2000), Market Place: In only a few weeks Nasdaq falls 25,3% from its pinnacle, *New York Times,* April 13

[86] Mossin, J. (1966), Equilibrium in a Capital Asset Market, *Econometrica Vol. 34 No. 4, pp.* 768-783

[87] Odean, T. (1998), Are investors reluctant to realize their losses, *Journal of Finance Vol. 53 No. 5, pp.* 1775-1797

[88] Oppel Jr., R. A. (2000), Dell Computer profit that meets lower forecast, *New York Times,* February 11

[89] Pachur, T. and R. Hertwig (2006), On the Psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use, *Journal of Experimental Psychology: Learning, Memory and Cognition Vol. 32, pp.* 983-1002;

[90] Read, D. (2001), Is time-discounting hyperbolic or subadditive?, *Journal of Risk and Uncertainty Vol. 23 No.* 1, pp. 5-32.

[91] Read, D. (2003), Intertemporal Choice, *Working Paper LSEOR 03.58*

[92] Ross, S. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory Vol. 13, pp.* 341-360

[93] Schoolman, J. (2000), Dow makes a comeback, profit-takers give Nasdaq small loss, *New York Daily News,* February 29

[94] Schultz, P. and M. Zaman (2000), Do the individuals closest to internet firms believe they are overvalued?, *Journal of Financial Economics Vol. 59 No. 3, pp. 347-381*

[95] Seo, Y., S. Cho and X. Yao (2000), The impact of payoff function and local interaction on the N-Player Iterated Prisoner's Dilemma, *Knowledge and Information Systems Vol. 2, pp.* 461-478

[96] Selten, R. and R. Stoecker (1986), End behavior in sequences of finite Prisoner's Dilemma Supergames, a learning theory approach, *Journal of Economic Behavior and Organization Vol. 7, pp.* 47-70

[97] Sharpe, W. (1964), Capital Asset Prices: A theory of Market Equilibrium, *The Journal of Finance Vol. 19 pp. 3, pp.* 425-442

[98] Shefrin, H. and M. Statman (2000), Behavioral Portfolio Theory, *Journal of Financial and Quantitative Analysis Vol. 35 No. 2, pp.* 127-151

[99] Shelley, M. K. (1993), Outcome Signs, Question Frames and Discount Rates, *Management Science Vol. 39 No.* 7, pp. 806-815

[100] Shiller, R. (1982), Consumption, asset markets, and macroeconomic fluctuations, *Carnegie Rochester Conference Series on Public Policy Vol.* 17, pp. 203-238

[101] Shiller, R. J., and A. E. Beltratti (1992) Stock prices and bond yields: Can their comovements be explained in terms of present value models? *Journal of Monetary Economics Vol.* 30, pp. 3-51

[102] Shiller, R. (2000), *Irrational Exuberance,* Princeton University Press, Chichester, West Sussex

[103] Skinner, B. F. (1989), Review of Hull's Principles of Behavior, *Journal of the experimental analysis of behavior Vol. 51 No. 2, pp.* 287-290

[104] Simon. H.A. (1955), A behavioral model of rational choice, *Quarterly Journal of Economics Vol. 69, pp.* 99-118

[105] Sornette, D. (2003), *Why Stock Markets Crash: Critical Events in Complex Financial Systems*, Princeton University Press
[106] Steel, P. and C. J. König (2006), Integrating theories of motivation, *Academy of Management Review Vol. 31 No. 4*, pp. 889-913
[107] Thaler, R. (1981), Some empirical evidence of dynamic inconsistency. *Economics Letters*, *Vol. 8*, pp. 201-207
[108] Thaler, R. (1985), Mental Accounting and consumer choices, *Marketing Science Vol. 4*, pp. 199-214
[109] Thaler, R. and E. J. Johnson (1990), Gambling the house money and trying the break even: The effects of prior outcomes on risky choice, *Management Science Vol. 36 No. (6), pp.* 643-660
[110] Thaler, R., A. Tversky, D. Kahneman and A. Schwartz (1997), The effect of myopia and loss aversion on risk taking: an experimental test, *Quarterly Journal of Economics, pp.* 647-661
[111] Thaler, R. (1999), Mental Accounting Matters, *Journal of Behavioral Decision Making Vol. 12*, pp. 183-206
[112] Tversky, A. and D. Kahneman (1974), Judgment under uncertainty: Heuristics and Biases, *Science Vol. 185, pp.* 1124-1131
[113] Tversky, A. and D. Kahneman (1986), Rational choice and the framing of decisions, *Journal of Business Vol. 59 No. 4, pp.* 251-278
[114] Vanberg, V.J. (2004), The rationality postulate in economics: its ambiguity, its deficiency and its evolutionary alternative, *Journal of Economic Methodology Vol. 11 No. 1, pp.* 1-29
[115] Volz, K.G., L.J. Schooler, R. Schubotz, M. Raab, G. Gigerenzer and D.Y. Von Crammon (2006), Why you think Milan is larger than Modena: neural correlates of Recognition Heuristic, *Journal of Cognitive Neuroscience Vol. 18, pp.* 1924-1936
[116] White, E. N. (1990), The Stock Market Boom and Crash of 1929 Revisited, *The Journal of Economic Perspectives Vol. 4 No. 2, pp.* 67-83
[117] White, E. N. (2004), Bubbles and Bursts: The 1990s in the mirror of the 1920s, *NBER Working Paper No. 12138*
[118] White, E. N. (2009), Lessons from the Great American Real Estate Boom and Bust of the 1920s, *NBER Working Paper No. 15573*

Conceptualizing and Measuring the Economic issues in the Evaluation of Socio-ecological Resilience: A Commentary

Alberto Mancini[1], Luca Salvati[2], Adele Sateriano[3], Giuseppe Mancino[4], Agostino Ferrara[4]

[1] *Via Portuense 507, I-00147 Rome, Italy*

[2] *Italian Agricultural Research Council, Centre for the Study of Plant-Soil Interactions (CRA-RPS), Via della Navicella, 2-4, I-00184 Rome, Italy*

[3] *Via C. Facchinetti 85, I-00159 Rome, Italy*

[4] *Department of Crop, Forest and Environmental Sciences, University of Basilicata*

Viale dell'Ateneo Lucano 10, I-85100 Potenza, Italy

[1]alberto.mancini.it@gmail.com

[2]luca.salvati@entecra.it (corresponding author)

[3]adsateri@tin.it

[3]giuseppe.mancino@unibas.it

[3]agostino.ferrara@unibas.it

Abstract **- Resilience is the capacity of an ecosystem to tolerate disturbance without collapsing into a state controlled by a different set of environmental conditions and processes. Despite the growing importance of socioeconomic resilience, this concept has not been yet carefully defined or satisfactory measured within the more general issue of socio-ecological resilience of both natural (forest) and anthropogenic (agricultural) systems. Investigating socioeconomic resilience in a rapidly changing landscape is important for sustainable land management under intense and increasing human pressure, like observed in the Mediterranean region. This paper presents an overview of definitions and indicators of the socioeconomic resilience and comments on some possible measurements of the concept taken from the parallel ecological literature. The study is intended to contribute to this deserving issue in the light of the (increasingly complex) relationships between the environment, the economic systems, and the social sphere.**

Keywords - *Desertification risk, Resilience, Forest ecosystems, Human systems, land cover.*

Introduction

Despite the growing importance of socioeconomic resilience during the current period of 'economic' and 'social' global crisis, this concept has not been yet carefully defined or satisfactory measured within the more general issue of socio-ecological resilience. This concept has received increasing interest in Europe, particularly in the Mediterranean region, whose landscape resulted from the complex interaction between society, economy, and the ecosystem (Lepart and Debussche 1992; Lavorel 1999, Thompson 2005).

The 'resilience' term, as originally illustrated by Elton (1958), refers to the amplitude of changes brought about by disturbance and by dynamics of post-disturbance recovery. Holling (1973) popularized this term within the broader framework of 'ecosystem stability'. He provided a definition of 'resilience' as the amount of disturbance that an ecosystem could withstand without changing self-organized processes and structures, defined as alternative stable states. The 'resilience' concept does not necessarily imply a return to the pre-existing state, but could be referred to as the capacity to respond to opportunities which arise as a result of change (Holling 2001).

In this perspective, Folke (2006) described 'resilience' as the set of opportunities that disturbance opens up in terms of recombination of evolved structures and processes, renewal of the system and emergence of new paths (Brand and Jax 2007). 'Resilience' hence expresses the adaptive capacity that allows for continuous development like a dynamic interplay between sustaining and developing with change (Carpenter et al. 2001).

This paper presents an overview outlining some possible meanings and measures of socioeconomic resilience by debating the implicit definitions

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.002.12/ijltfes.2047.0916

proposed in the (rather) restricted literature that deals with socioeconomic resilience and commenting on some possible measurements of the concept taken from the parallel ecological literature.

In the first section we present a general definition of the concept from both the ecological and the socioeconomic side. A summary methodological introduction on how to measure socioeconomic resilience of human and natural systems was introduced afterwards. The relationship between ecosystem management and socio-economic resilience was further discussed in the light of complex relationships between the environment, the economic systems, and the social sphere. In the final section we outlined an economically-oriented vision of the 'resilience' concept as an original contribution to the complex research issue of sustainability.

1. Main definitions

In social science literature reviews on resilience, perhaps the most traditional meaning of socioeconomic resilience is the ability of a regional economy to maintain a pre-existing state (typically assumed to be an equilibrium state) in the presence of some kinds of exogenous shocks. Although only a few studies explicitly use the term "resilience" most of the economic literature that deals with the idea of resilience is concerned with the extent to which a regional or national economy that has experienced an external shock is able to return to its previous level and/or growth rate of output, employment, or population (Blanchard and Katz 1992, Rose and Liao 2005, Briguglio et al. 2006, Feyrer, Sacerdote, and Stern 2007).

Another interpretation could be the idea of path-dependence, or historical "lock-in" processes; this concept is based on the assumption that an economy has multiple equilibria, not all of which are efficient (in a static or dynamic sense). This suggests a concept of regional economic resilience in which resilience is the ability of an economy to avoid becoming locked into such a low-level equilibrium or, if in one, to transition quickly to a "better" equilibrium.

A long-term, holistic perspective, in contrast, would emphasize the structure of relationships among macroeconomic variables that persists over a long period of time and the economic, political, and social institutions that condition this structure (Reich 1997). As an example, a social structure is not static; although it persists for a long time, it evolves in ways that ultimately threaten firms' profitability and long-term macroeconomic growth.

The study of resilience would then be the study of the rise, stability, and eventual decay of the institutions that underlie long-term regional economic growth. An economy would be resilient to the extent that its social structure was stable or to the extent that it was able to make a rapid transition from one structure to another. So, in general, socioeconomic resilience is defined as the ability of a nation or a region to recover successfully from shocks to its economy that either throw it off its growth path or have the potential to throw it off its growth path but do not actually do so.

Economic systems that experience negative shocks may exhibit three different kinds of responses. Some of these may have returned to or exceeded their previous growth within a relatively short period of time (definition concerns dealt with below); these regions might be called economically resilient. Some may not have been thrown off their growth path at all; these regions might be called shock-resistant. Finally, some regions may have been unable to rebound and return to or exceed their previous path; these might be called non-resilient.

2. General methodologies

Economically resilient and non-resilient economies can be identified using data on aggregate economic performance, while shock-resistant countries can be identified using data on industry performance or other information on non-industry shocks. Human skill may also be regarded as an important factor here but is rather difficult to calculate in a way comparable with the other factors.

Economically resilient and non-resilient economies can in principle be identified by examining their economic performance over a period of time. Criteria for a negative economic shock can be defined and pre- and post-shock growth rates and levels of economic performance can be measured. A region which post-shock growth rate is at least as high as its pre-shock growth rate and that achieves its pre-shock level of economic performance within a specified time period can be considered resilient, while a region that experiences a negative shock and does not meet these criteria can be considered non-resilient.

A continuous variable can be developed to measure socioeconomic resilience, e.g. the number of years (or quarters if the data permit) it takes to return to the previous growth path, percentage of lost employment (or other relevant measure) replaced within a standard period of time, or some other measure that takes into account the relationship between post-shock performance and the size of the shock (Hill. et al. 2008). To implement a socioeconomic resilience measure is necessary to address a series of measurement issues, such as the following:

• What measure(s) of economic performance should be used, e.g. gross domestic product, employment, earnings, income (and for all of these, total or per worker or per capita)?

• Should the growth rate for a region be measured in absolute terms, relative to the national average, or relative to the average in the relevant economic zoning, census region, or administrative division (or all of these)?

• How far back in time should growth paths be traced?

• For how many years should growth paths and shock periods be measured? Should the same number of years be used to define pre-shock, shock, and post-shock periods, or should the lengths of these periods be allowed to differ?

• How should growth paths and shocks be measured (e.g., average annual growth or the slope of a regression line through all observations during a time period)?

• How large does a negative deviation have to be (relative to the region's previous performance and/or national average performance) to count as a negative shock?

• How should a region's pre-shock level of economic performance be defined (e.g., peak or average performance during the pre-shock period)?

3. Measuring socioeconomic resilience in natural and anthropogenic systems

The first challenge faced in measuring socioeconomic resilience is to define spatial areas that reflect patterns of human activity. Areas defined according to ecological criteria (such as hydrological basins or bio-geographical regions) do not effectively capture these patterns. Any spatial definition of socioeconomic systems is to some degree subjective; these are open systems in which people, money, goods, and services continually cross any boundary adopted. Further, if socioeconomic systems are defined in a spatial hierarchy (international, national, regional, and local), interactions occur among all levels. The theoretical basis for socioeconomic resiliency rests on the concept of social well-being, which is defined as a composition of three factors: economic resiliency, social and cultural diversity (e.g. population size, human skill mix) and civic infrastructure (McCool et al. 1997).

A provisional index of economic resiliency could be thus developed directly from measures of diversity in employment or income among economic sectors. Social and cultural diversity can be measured by using data on lifestyle diversity. Because there was no direct way to measure civic infrastructure, in this analysis the population density is used as a proxy, according to Barkley et al. (1996). The socioeconomic resilience index is developed mixing three factors: economic resiliency, population density, and lifestyle diversity. In this context the measures for both economic resiliency and lifestyle diversity are calculated using a diversity index (Shannon and Weaver 1949):

$$D = -\sum_{i=1}^{n} (E_i * \log E_i)$$

where, D is the diversity index of an area; i is the i-th industry; n is the number of industries; Ei is the proportion of total employment in the area located in the ith industry; logEi is the logarithm (base 10) of Ei . The indices are normalized so that all numbers ranged between 0 (no diversity) and 1 (perfect diversity). The composite rating of socioeconomic resiliency is determined by combining the results of economic resiliency, population density, and lifestyle diversity. the socioeconomic resiliency rating based is assigned on the sum of the ratings for the three factors; that is, the three factors are equally weighted (Horne and Haynes 1999).

3.1. Economic-oriented resilience

We said that socioeconomic resilience refers to the policy-induced ability of an economy to recover from or adjust to the negative impacts of adverse exogenous shocks and to benefit from positive shocks. In the study from Briguglio (2007), economic resilience is associated with actions undertaken by

policy-makers and private agents which enable a region to withstand or recover from the negative effects of shocks. Actions which enable a region to better benefit from positive shocks are also considered to lead to economic resilience. The term is used in two senses in this analysis, respectively related to the ability to recover quickly from a shock and to withstand the effect of a shock. The ability of a regional economic system to recover from the effects of adverse shocks is associated with the flexibility of an economy, enabling it to bounce back after being adversely affected by a shock. On the other hand, this ability will be enhanced when the economy possesses discretionary policy tools which it can utilize to counteract the effects of negative shocks.

The ability to withstand shocks relates to the capability to absorb shocks, so that the end effect of a shock is neutered or rendered negligible. This type of resilience occurs when the economy has in place mechanisms to reduce the effects of shocks. For example, the existence of a flexible, multi-skilled labour force could act as an instrument of shock absorption, as negative external demand shocks affecting a particular sector of economic activity can be relatively easily met by shifting resources to another sector enjoying stronger demand.

3.2. Constructing a measure of resilience

This section illustrates an attempt to construct a composite index of economic resilience. However care was taken to base the choice on a set of desirable criteria related to appropriate coverage, simplicity and ease of comprehension, affordability, suitability for international comparisons and transparency. The compilation of the index encountered a number of problems with regard to data collection, the most important of which were associated with shortage of data and non-homogenous definitions across countries. A resilience index should be aimed at measuring the effect of shock-absorption or shock counteraction policies across countries. It is hypothesised that the variables that capture these effects are the following:

- macroeconomic stability;
- microeconomic market efficiency;
- good governance;
- social development.

Macroeconomic stability relates to the interaction between an economy's aggregate demand and aggregate supply. If aggregate expenditure in an economy moves in equilibrium with aggregate supply, the economy would be characterised by internal balance, as manifested in a sustainable fiscal position, low price inflation and an unemployment rate close to the natural rate, as well as by external balance, as reflected in the international current account position or by the level of external debt. These can be all considered to be variables which are highly influenced by economic policy and which could act as good indicators of an economy's resilience in facing adverse shocks.

In this study the macroeconomic stability aspect of the resilience index proposed is constructed on the basis of three variables:

- the fiscal deficit to GDP ratio: the government budget position is suitable for inclusion in the resilience index because a healthy fiscal position would allow adjustments to taxation and expenditure policies in the face of adverse shocks.

- the sum of the unemployment and inflation rates: those indicators are also considered to be suitable indicators of resilience. This is because price inflation and unemployment are strongly influenced by other types of economic policy, including monetary and supply-side policies. They are associated with resilience because if an economy already has high levels of unemployment and inflation, it is likely that adverse shocks would impose significant costs on it.

- the external debt to GDP ratio: the adequacy of external policy may be gauged through the inclusion of this indicator. This is considered to be a good measure of resilience, because a country with a high level of external debt may find it more difficult to mobilize resources in order to offset the effects of external shocks.

Microeconomic market efficiency is constructed on the variables that composing the Economic Freedom of the World Index (Gwartney and Lawson 2005), entitled "regulation of credit, labour and business" which is aimed at measuring the extent to which markets operate freely, competitively and efficiently across countries. In the financial market this index assesses the extent to which the banking industry is dominated by private firms; foreign banks are permitted to compete in the market; credit is supplied to the private sector; and controls on interest

rates interfere with the market in credit. All these relate to the degree of interference by government in the financial market, which could preclude the economy from reacting flexibly to shocks.

Similar considerations apply in the case of the labour market. Here interference relates to unduly high unemployment benefits (which could undermine the incentive to accept employment), dismissal regulations, minimum wage impositions, centralised wage setting, extensions of union contracts to non-participating parties and conscription. All these are viewed as possibly precluding work effort, thereby limiting the ability of a country to recover from adverse shocks. Bureaucratic control of business activities are also thought to inhibit market efficiency.

Good governance is essential for an economic system to function properly and hence to be resilient. Governance relates to issues such as rule of law and property rights. This is considered to be useful in the context of the present exercise in deriving an index of good governance. This concept covers five components: judicial independence; impartiality of courts; the protection of intellectual property rights; military interference in the rule of law; and political system and the integrity of the legal system.

A composite index was computed by taking a simple average of the four components previously described: macroeconomic stability, microeconomic market efficiency, good governance and social development. All observations of the components of the index were standardised using a linear transformation. This transforms the values of observations in a particular variable array so that they take a range of values from 0 to 1 (Briguglio 2006).

4. Ecological resilience, ecosystem management and socioeconomic resilience

One of the objectives for the ecosystem management is to encourage socioeconomic resilience defined as the ability of human institutions to adapt to change (Haynes et al. 1996). These institutions include both communities and economies. A community is defined as a sense of place, organization, or structure (e.g. Galston and Baehler 1995). An economy is defined by transactions among people that allocate scarce resources among alternative uses, and may exhibit different spatial configurations than communities. With the concept of socioeconomic resilience, this vision recognize that change is inherent in human systems. Social and economic factors are continuously changing - population grows, people migrate, social values evolve, and new technologies and knowledge are created.

In the study from Horne and Haynes (1999) the challenge is how to develop a measure of socioeconomic resilience that is useful for understanding the extent to which changes in policies for land management may affect the human systems coincident with those lands (Quigley et al. 1996). The interest of this analysis stems from a long-held concern about the relation between ecosystem management practices and the economic well-being of nearby residents.

In this analysis it is assumed that the relation between diversity and resilience in social and economic systems is similar to that in the ecological literature (e.g. Moffat 1996); that is, a system with higher diversity is less affected by change than a system with lower diversity and the former therefore has higher resilience. Socioeconomic systems with high resiliency are defined as those that adapt quickly as indicated by rebounding measures of socioeconomic well-being. People living in areas characterized by high resilience have a wide range of skills and access to diverse employment opportunities.

Thus if specific firms or business sectors experience downturns, unemployment rates rise only briefly until displaced people find other employment. Systems with low resilience have more lingering negative impacts, such as unemployment or out-migration rates that remain high for several years. The terms “high” and “low” should not be thought of as “good” or “bad,” but simply as a reflection of the ability of a socioeconomic system to respond to changes in social or economic factors. Note that having greater diversity (and higher resilience) does not eliminate the possibility of wide fluctuations for single economic entities or sectors.

5. Concluding remarks

The approach previously described is potentially interesting but rather narrow in scope and very difficult to measure across countries. Other components of the ‘resilience’ dimension can be mentionned at this point. We made just two examples in the follows.

Social development is an essential component of economic resilience. Social development in a country can be measured in a number of ways. Variables relating to income such as its dispersion and the proportion of the population living in poverty, the long-term unemployment rate — indicating the proportion of the population with low skills and inadequate employment prospects — and the proportion of the population with low level of education could be useful, additional indicators to the illustrated evaluation system.

Economic resilience can also be viewed to be determined by a series of other (non-social, non-institutional) factors apart from those mentioned above. It may be argued, for example, that it could be useful to consider the effects of environmental management in this regard (see also paragr. 5). The environment can be an important source of vulnerability by giving rise to shocks of an adverse nature, principally by rapid events, such as earthquakes and floods. In turn, these would have important repercussions on the economy and society.

In conclusion, only a thorough analysis involving multiple research dimensions from environmental, economic, social, and political studies and also multiple assessment scales (from local to sovra-national) may assure a conceptual definition and a reliable operational description of the socioeconomic resilience.

Acknowledgments

This study was supported by the 7FP project LEDDRA – Land and Ecosystem Degradation and Desertification: Assessing the Fit of Responses – ENV.2009.2.1.3.2. (started on march 2010). Agostino Ferrara is responsible for WorkPackage 3 “Land and ecosystem degradation and desertification in forests/shrublands”.

References

[1] Adger, WN (2000). Social and ecological resilience: are they related? Progress in Human Geography 24: 347-364.

[2] Horne AL, Haynes RW (1999). Developing Measures of Socioeconomic Resiliency in the Interior Columbia Basin. General Technical Report PNW-GTR-453, USDA FS, Pacific Northwest Research Station, Portland, OR.

[3] Barkley, DL, Henry, MS, Bao, S (1996). Identifying “spread” versus “backwash” effects in regional economic areas: a density functions approach. Land Economics 72(3): 336-357.

[4] Berkes. F, Colding, J, Folke, C (2003). Navigating social–ecological systems: building resilience for complexity and change. Cambridge, UK. Cambridge University Press.

[5] Blanchard, O, and Lawrence FK (1992). Regional Evolutions, Brookings Papers on Economic Activity 1992, no. 1, pp. 1-75. Washington: Brookings Institution.

[6] Brand FX, Jax K (2007). Focusing the meaning(s) of resilience: Resilience as a descriptive concept and a boundary object. Ecology and Society 12(1): 23.

[7] Briguglio, L (2007). Economic Vulnerability and Resilience Concept and measurements. Working paper, Economics Department, University of Malta.

[8] Briguglio, L (2006). Conceptualizing and Measuring Economic Resilience. Working paper, Economics Department, University of Malta.

[9] Carpenter S, Walker B, Anderies J, Abel N (2001). From metaphor to measurement: resilience of what to what? Ecosystems 4: 765-781.

[10] Chapin, FSIII, Kofinas, GP and Folke C. (2009). Principles of Ecosystem Stewardship: Resilience-Based Natural Resource Management in a Changing World. New York, Springer.

[11] Dell, B, Hopkins, AJM, Lamont, BB (1986). Resilience in Mediterranean-type ecosystems. Dordrecht. Dr W Junk Publishers.

[12] Folke, C (2006). Resilience: The emergence of a perspective for social–ecological systems analyses. Global environmental Change 16: 253-267.

[13] Folke, C, Carpenter, S, Walker, B, Scheffer, M, Elmqvist, T, Gunderson, L, Holling, CS (2005). Regime shifts, resilience and biodiversity in ecosystem management. Annual Review of Ecology Evolution and Systematics 35: 557-581.

[14] Folke, C, Carpenter, SR, Walker, B, Scheffer, M, Chapin, T, Rockström, J (2010). Resilience thinking: integrating resilience, adaptability and transformability. Ecology and Society 15: 20.

[15] Gwartney, J, Lawson, R (2005). Economic Freedom of the World 2005 Annual Report. Vancouver: Fraser Institute.

[16] Holling, CS (1973). Resilience and stability of ecological systems. Annual Review of Ecology and Systematics 4: 1-23.

[17] Holling, CS (2001) Understanding the complexity of economic, ecological and social systems. Ecosystems 4: 390-405.

[18] Lavorel, S (1999). Ecological diversity and resilience of Mediterranean vegetation to disturbance. Diversity and distributions 5: 3–13.

[19] Lepart J, Debussche, M (1992) Human impact on landscape patterning: Mediterranean examples. In: Hansen, AJ, di Castri, F (Eds.), Landscape Boundaries for Biotic Diversity and Ecological Flows. Springer, New York, pp 76–106.

[20] Perrings CA (2006) Resilience and stability of ecological systems. Ann Rev Ecol System 4: 1-23.

[21] Reich, M (1997). Social Structure of Accumulation Theory: Retrospect and Prospect. Review of Radical Political Economics 29: 1-10.

[22] Rose, A, Shu-Yi L (2005). Modeling Regional Economic Resilience to Disasters: A Computable General Equilibrium Analysis of Water Service Disruptions. Journal of Regional Science 45: 75-112.

[23] Shannon, CE, Weaver, W (1949). The mathematical theory of communication. Urbana, IL: The University of Illinois Press. 125 p.

[24] Walker, B, Gunderson, L, Kinzig, A, Folke, C, Carpenter, S, Schult, L (2006). A Handful of Heuristics and Some Propositions for Understanding Resilience in Social-Ecological Systems. Ecology and Society 11: 13.

[25] Westley, F (2002). The devil is in the dynamics. In: Gunderson, L.H. and Holling, C.S. (Eds). Panarchy: understanding transformation in human and natural systems. Washington, D.C., USA. Island Press. 333-360.

Nanotechnology and Processes: Nano-Photovoltaic Panels as an Innovation in Energy Market

José Chavaglia[*], José António Filipe[1#], Manuel Alberto M. Ferreira[2#], Manuel Pacheco Coelho[+]

[*]Instituto Universitário de Lisboa (ISCTE-IUL)
Lisboa Portugal
jnchavaglia@gmail.com

[#]Department of Quantitative Methods
Instituto Universitário de Lisboa (ISCTE-IUL), BRU – IUL
Lisboa Portugal
[1]*jose.filipe@iscte.pt*
[2]*manuel.ferreira@iscte.pt*

[+]SOCIUS & ISEG/UTL – Portugal
coelho@iseg.utl.pt

***Abstract*— Nanotechnology can be a powerful weapon in creating competitive advantages in the energy market, through the use of the photovoltaic nano-panels, which may reduce production costs and simultaneously meet the socio-environmental requirements. It is a way to produce clean energy. Moreover, today the adoption of nanotechnology in energy production can make this kind of energy very interesting along the years, and fractal theory may help to show the effects. Nanotechnology may, in fact, be responsible for unimaginable gains, both economically and to preserve the planet.**

Keywords **- Nanotechnology, photovoltaic nano-panels, solar energy, process.**

1. Introduction

In the new era in which competition has become global, technological innovations are the way to companies overcoming the challenges in the context of competition, by applying innovations to their production processes.

A run to technological innovations becomes quite visible, especially in the last decade in which the life cycle of products has decreased considerably. Many developments in terms of innovations have been seen in communications, medicine, robotics, computing, energy or many other sectors of society.

It is very relevant to see the role that internet took over, the development of mobile devices, new medical imaging, the self-service centers, banks, etc. The truth is that people have become dependent on new technologies and companies became dependent on the process of constant search for innovation. But it is not possible to say in what extent this is vital. The fact is that just as some people do better professionally than others, some companies also do better in the competitive environment than others. Notably, in both cases, both individuals and companies that stand out have something in common: the ability to innovate.

This scenario is to highlight one of the great scientific revolutions in human history, nanotechnology, which opens up unimaginable possibilities in various fields of human reality and in various fields of science. Today the influence of nanotechnology can be seen in areas such as medicine (implantation of nano-robots, for example), agriculture (for pest control) or many others. The possibilities are endless on these fields.

The use of nanotechnology in the production processes of industry is evidenced in this article, in particular in the production of electric energy produced by photovoltaic panels. Emphasis is placed on the competitive advantage associated with the use of nanotechnology to solar energy production for companies in this market segment.

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.003.12/ijltfes.2047.0916

The American firm "Nanosolar", which studies are sponsored by major companies like Google or IBM, or resulting from the allocation of benefits offered by the Department of Energy, is leading the race for energy production derived from nanotechnology. It has named this technology as "nano-photovoltaic panels".

The use of this product has been showing a reduction in total costs on firms in relation to other types of solar energy, and meet government requirements and social use of energy from clean sources.

Despite the high feasibility for the economy and the environment, there are some considerations regarding the ethical and moral limits on nanotechnology that should be taken into account.

2. The Emergence of Nanotechnology

The first person to conceptualize Nanotechnology was Richard P. Feynman, although he has not used this term in his speech to the *American Physical Society* on December 29, 1959, where he made the first comments on the subject. The word "nanotechnology" was first used by Professor Norio Taniguchi (1974) to define the fabrication of a scale of 1 nm. Nanotechnology is the potential ability to create things from the smallest element, using the techniques and tools that are being developed today to place every atom and molecule in place. The use of nanometer implies the existence of a system of molecular engineering, which will likely generate the subversion of the factory manufacturing model as it is known.

However, nanotechnology will offer, in addition to products of higher quality at lower cost, a range of possibilities to produce new means of production and new types of resources and factors. This is a manufacturing system that could produce more manufacturing systems (plants that produce other plants) in a quick, cheap and clean way. The means of production may be reproduced exponentially. So in just a few weeks, power would pass from just some to several billion nanofactories. Thus, this represents a kind of revolutionary technology, manufacturing, powerful, but also with many potential risks, besides the benefits it has (see Euroresidentes, 2011, p. 01).

In Brazil, the budget of the Ministry of Science and Technology for the next four years is R680 million to this area. Overall it is estimated that only developed countries will allocate a sum of around USD 5.5 billion. An important example of successful application of nanotechnology is *Empresa Brasileira de Agropecuária* (Embrapa). It has been working with nanotechnology in various research centers and has already released some products. One of the most notable is perhaps the "electronic tongue", a device that combines chemical sensors with nanometer-thick with a computer program which detects flavors and aromas and serve to quality control and certification of wines, juices, coffees and other products (see DIEESE, 2008, p. 03).

As suggested earlier, the subversion of the industrial model was directly related to the use of nanotechnology in the various branches of economic activities. However this study gives an outline for the manufacture of electric power, which generically McKibben considered itself as a country's economy (see McKibben, 2009, p. 24).

Nanotechnology is therefore contributing to the transformation of traditional models, being the way goods and services are produced, or in the way the production is conducted and made.

3. Nano-photovoltaic Panels as an Innovation

Companies need to be dynamic in the development of innovations and thus creating competitive advantages through its production processes so they can create economic value, and consequently generate their viability based on the market in which they operate. For McDonough III (2009, p. 04) "in the current buoyant economy, organizations must continually reinvent what they are and what they do [...]". This means that they need to constantly maintain market differentiation, through deliberate strategies in order to obtain competitive advantages that provide monopoly profits, even if temporary, in this environment that requires from companies a high degree of competitiveness.

The competition is part of a dynamic and evolutionary operation of the capitalist economy. The evolution of this economy is seen as over time based on an uninterrupted process of introduction and diffusion of innovations in a broad sense, ie any changes in the economic space in which these companies operate, whether changes in products, processes, sources of raw materials, forms of productive organization, or in their own markets, including in terms of geography (see Schumpeter, Brazilian version, 1982, p. 65).

According to McAfee and Brynjolfsson (2008, p. 78) "the result is that an innovator with a better way of doing things can grow at unprecedented speeds and dominate the industry".

In contrast, at the same time, which seeks a high level of competitiveness, it is understood that there is a need for companies to retain the existing common resources for their optimization in the future because we can no longer count on such a supply of natural resources to meet the continuing huge demand, given the level of production that humanity has achieved over the last century.

To Nogami and Passos (1999, p. 03), from the harsh reality of scarcity arises the necessity of choice. Since it is not possible to produce everything that people want, mechanisms must be created to somehow show to the societies the path to decide what goods to produce and what needs are met.

The concepts that are consistent with the possibility of economic and technological efficiency allow us to make an analysis of the act of production through the combination of forces and inputs in the production unit. All the means or methods of production indicate some of these combinations. Production methods vary in how such combinations occur, or by objects or by the combined ratio of their quantities. Every concrete act of production incorporates some combination thereof. It could also be considered as a combination enterprise itself, and even the production conditions of the whole economic system (see Schumpeter, 1982, p. 16).

The possibility of economic and technological efficiency reflects a producing combination of forces and inputs to reach an interesting production level for the company. All the means or methods of production indicate different combinations. Production methods vary in how such combinations occur, or by objects or by the combined ratio of their quantities. Every concrete act of production incorporates a kind of combination. A company can itself be considered a combination by itself, and even the production conditions of the whole economic system (see Schumpeter, 1982, p. 16).

Thus, it can be said that any company that uses clean energy in its production of goods and services, in particular the use of photovoltaic panels generated by nanopanels, generates competitive advantage by breaking the closed circle of the economy by creating a new mechanism of generation of market value, since it is a new way to produce through a new combination of available resources.

In a direct way, besides the fact that the use of photovoltaics is already an innovation in itself, when it is combined with nanotechnology, its power of subversion, which can also be interpreted as its ability to break paradigms in the energy industry, takes the form of a powerful competitive weapon of production units.

Given this context, nanotechnology combined with solar energy production becomes interesting for the production units, whether public or private.

4. Deployment of a Photovoltaic system

4.1. Manufacture of a Nano-Solar Panel

In order to describe how a nanopanel is got, first semiconductor nanoparticles have to be produced (approximately 20mn of size, which is the equivalent of 200 atoms in diameter). Later, aluminum sheets are placed in presses, similar to those used in paper graphics. These aluminum sheets can be dynamic in the way they are used, due to their length and their width. This makes a product much more adaptable to the needed formats. After that, a thin layer of semiconducting ink is painted on aluminum substrate. Then another press deposits layers of cadmium sulfide and sulfur and zinc oxide (ZnO and CdS). Zinc oxide layer is non-reflective to ensure that sunlight is able to reach the semiconductor layer. Finally, the leaf is set in solar cells. Unlike other methods of manufacture of panels that require a special location for manufacturing, on this system nano-panels can be produced outdoors.

4.2. System Operation

To produce energy derived from solar radiation is necessary to understand the functioning of a photovoltaic cell. The photovoltaic solar energy is obtained through direct conversion of light into electricity, the so-called "photovoltaic effect" (see Becquerel, 1839, the discoverer of this effect).

The photovoltaic cell works when light reaches the photovoltaic panel and moves electrons which circulate freely from atom to atom, forming the electric current.

The photovoltaic cell is a practical application of photoelectric effect. When the light falls on certain

substances, takes off electrons that circulating freely from atom to atom, form a chain that can be stored. The photovoltaic cell that transforms the light into electricity will continue to generate power according to the level of radiation emitted, i.e. while the panel receive the light it will continue to generate electricity.

For Nascimento (2004, p. 02) "the photovoltaic cell does not store electrical energy, but maintains a flow of electrons established in an electrical circuit as long as there is light on it".

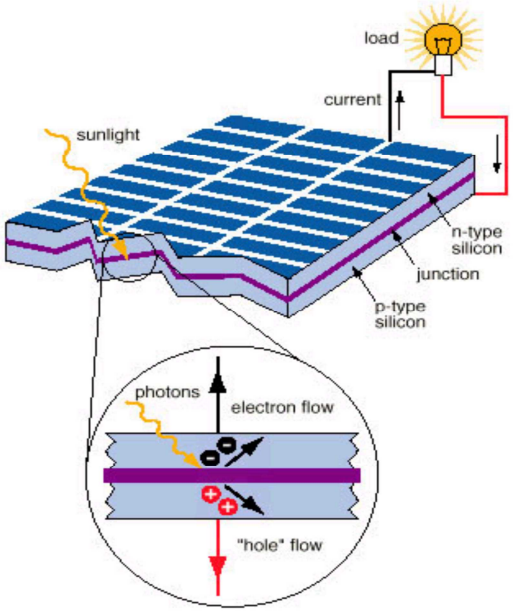


Figure 1- Energy Conversion
Direct conversion of solar radiation into electricity
(Source: Australian CRC for Renewable Energy Ltd)

4.3. The Installation of Photovoltaic System in a Company

A photovoltaic installation in an organization can be independent or have a connection to the electricity grid. For this connection to the network, it is necessary to establish a point of reception. The point is the point of power system network of the public service, which makes the connection of the installation. The sine wave inverter is used on systems connected to the power system network. Most homes use alternating current of 120 Volts 60 Hz. The sine inverter transforms the direct current photovoltaic system (ranging usually between 12Vcd-360Vcd) in 120 VAC, 60 Hz and synchronizes with the electrical network (see Figure 2).

The photovoltaic panels are easy to install, being a relatively simple process, regardless of the installation location, since there is a flexibility in choice of location, especially with the use of nano-panels. There are several possible locations for its installation, such as rooftops, or even on the ground or in the buildings' painting. On the roofs, the support structure will be secured to the roof structure, or in the ink of the tiles and structures. For the panels installation on the ground a suitable structure should be used for fixing them to the ground.

In photovoltaic installation for enterprise consumption, the connection is made to the electrical system of the building of the production unit. Considering the photovoltaic installation, aiming to sell the surplus energy to private or public network, it is essential to record it in the system to account the power sent to the power grid. This record quantifies the energy that is released and the energy that is billed to the electrical network. The registry will preferably be close to the main electricity recording in the company.

Consequently, the materials used in photovoltaic network deployment have the same characteristics and the same lifetime of equipment already used in other electrical installations. They differ only in the batteries of energy that have a lifespan of 5 to 10 years and act as accumulators for use in schedules without the presence of solar radiation.

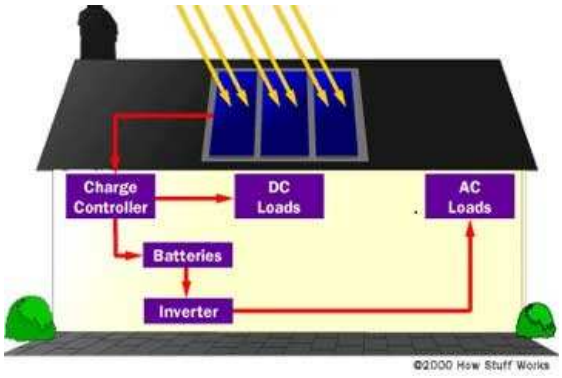


Figure 2 - Photovoltaic system installation.
Schema of a complete photovoltaic installation.
(Source: Toothman, J. and Aldous, S., 2000)

4.4. Costs of Solar Photovoltaic Energy

A company in a competitive market like the one of nowadays has necessarily to be very competitive. To be competitive, the company should prioritize the maximization of its profits, and the minimization of its cost of production.

In fact, a competitive advantage reflects a costs reduction in production. According to Varian (2006, p. 386) "the short-term cost function is defined as the minimal cost to achieve a given level of production, with only the adjustment of variable factors of production". Costs can be divided in two types, fixed costs and variable costs. Fixed costs are costs that remain constant, regardless of the degree of occupation of the productive capacity of the company. These costs derive from the very existence of the company. The variable costs are the costs arising from the degree of occupation of productive factors on the company. For Sandroni (2007, p. 219) they are "a part of the total cost that vary according to the degree of occupation of the productive capacity of the enterprise: for example, cost of raw materials, wages for production and others".

Considering this, the cost arising from the use of photovoltaics fits both fixed costs (resulting from the installation, on equipment required for the generation of energy) but also fits into variable costs (resulting from the level of use of generated energy, varying accordingly to the allocation of resources and the level of production).

Type of energy	Cost per kWh (R$)	Cost per kWh (€)
Electric power network	0,34	0,12
Photovoltaics (silicon Panel)	10,49	3,75
Photovoltaics (nano-Panel)	0,92	0,34

Figure 3 - Cost per energy type
(Source: Adapted from Scientific American 2008)

Considering figure 3, showing the costs relating to the use of some types of energy: the electricity from the public network, the solar photovoltaic energy through silicon panels and energy generated by photovoltaic nano-panels. The costs of solar power generated by nano-panels is between the other two ones. Therefore, cheaper energy means reduction of variable costs and this type may be a very competitive future type of energy.

Compared with the use of the energy supplied by the local public network, energy derived from solar panels still presents a higher cost of operation. However, the difference in costs between them has been dropping consistently and it happens even if compared to photovoltaic solar energy generated by silicon panels, which presents a much greater production cost than the energy generated by nano-photovoltaic panels. It can be also considered that there are extra costs generated by hydroelectric power that are subjective, because they affect the environment where the entrepreneur belongs. Allied to this, in Brazil a strong trend of increased costs with the production of hydroelectric power can be observed. In Brazil the total cost of energy production has been performing considerably above the country's inflation in the period.

There is an advantage of using photovoltaic energy through the use of nano-panels once there are reasons that qualify the use of photovoltaic energy as a generator of competitive advantages in the market. The company that owns this type of technology is likely to have a reduction of its variable costs, when compared with the solar energy generated by photovoltaic silicon solar panels. There is also a perceived competitive advantage considering the conservation of the environment when it is compared with the energy provided by the local public network.

4.5. A New Opportunity for Power Generation Based on Nanotechnology

In addition to the already existent nano-panels, another opportunity to revolutionize the market is already in sight. Some scientists through a research in nanotechnology have developed devices capable of recharging nanoscale devices without the need for bulky quantities of energy provision as the battery and cables connected directly to the electric network.

To convert the mechanical energy of body motion, muscle stretching or water into electricity, there are "nanogenerators" that may make possible a new class of implantable medical devices, sensors and portable auto-feeding electronic devices (see Wang, 2010, p. 01).

The system works as follows: the current-producing nanogenerators flexion and then release zinc oxide nanowires that are simultaneously piezoelectric and semiconductors. In summary, nanogenerators convert mechanical energy into electrical energy. This is undoubtedly a giant step in the search for viable solutions to the economy and society in general in several areas.

Until now many possibilities derived from nanotechnology have been presented to the production of goods and services in the economy. However a great barrier to the full use of these devices has been in fact the power source, either by its absence by itself, and by replacing the energy storage system with batteries, because the existent ones are too expensive and generate a high degree of pollution when they are discarded. The impact of this change may be very considerable, for example, in the production of goods and services in isolated locations as it is the case of the Amazon region.

Being interesting for being used inside the body because the zinc oxide is not toxic, the nanogenerators could also be used whenever the mechanical power-hydraulic movement of either the sea water, the wind or the movement of a foot inside a shoe, for example, is available. Nanowires may grow on crystal substrates but not only, they can grow also in any film based on polymers. One day the use of flexible polymer substrates may allow portable devices to be powered by the movement of their users (see Wang, 2010, p. 02).

Of course this is just speculation yet. However it is a possibility to be considered in the future and would represent a revolution in the way man produces power and in the way people may satisfy some needs and consumers satisfy some desires.

5. Marketing representation through nanotechnology innovation in the energy sector

The importance of nanotechnology associated with energy production in industrial production is already considerable and evident. However a question hangs in the air: is it possible to validate the benefits of the use of nanotechnology in the production of goods and services?

In fact, considering the nanotechnology used in energy production so far presented in market terms allows a new stage of development in the world economy. After all, goods just may be reproduced at big scales if they have marketing value.

"Theory of geometry of fractals" can be used to explain the possible performance of an innovation derived from the nanotechnology in the market.

Having into account the geometry of fractals, fractal is a word coined by Mandelbrot to describe the irregular geometry. It represents something fractured - from the Latin fractus, fractured source. Fractality is repeating geometric patterns in different scales, revealing increasingly smaller versions of themselves (Taleb, 2009, p. 321) .

In quantitative terms, the fractal geometry is nothing more than numeric measurements which are, at least, partly preserved in different scales, and the proportions are kept the same.

The Mandelbrot set, widely used in association with chaos theory, allows us to consider some notes about this subject. For example, a set can be seen in smaller resolutions without ever reaching the threshold, even though the forms never be the same but smaller reproductions of themselves with small changes.

In mathematical and statistical terms, the fractal geometry has numeric measurements that are preserved in different scales.

Fractal geometry is referred on this economic phenomenon for explaining the consequences of the use of nanotechnology today in energy production that can generate a great effect in a few years. Yet it is from general knowledge that the accurate prediction of this effect is impossible to be measured.

In practice the fractal geometry can be presented in a simple example in this subject, using the philosophy of the example used by Taleb (2009, p. 293), explaining the difference between the Gaussian model and the Mandelbrotian model. Mandelbrotian model allows us to illustrate how the use that today is made from nanotechnology in electrical energy production can increasingly promote its use making it a long-term trend.

Consider, for illustration, the probability of getting rich in Europe, accepting for simplification of analysis that wealth in Europe is Mandelbrotian (scalable).

Tabela 1: Scalable Distribution

Scalable distribution of wealth	
People with Liquid assets greater than 1 million Euro	1 in 62,5
Greater than 2 million Euro	1 in 250
Greater than 4 million Euro	1 in 1000
Greater than 8 million Euro	1 in 4000
Greater than 16 million Euro	1 in 16000
Greater than 32 million Euro	1 in 64000
Greater than 320 million Euro	1 in 6400000

Source: Adapted from Taleb (2009).

As can be seen in table 1, in the course of the evolution of the probability process of wealth in no moment any event diminishes the intensity of the movement or trend. As wealth increases, the chance to get rich decreases. In mathematical terms, this implies the use of power laws. Note in the example of Taleb (2009), that the inequality between the super-rich is the same inequality between the simple-rich, i.e. the speed does not diminish.

Now let this to occur with the use of nanotechnology with the extensive use of photovoltaic nano-panels. The market acceptance could be something very similar to this effect, except that instead of the falling in probability occurs an increase, i.e. there is a scalable growth of the number of organizations that use solar photovoltaic energy of nano-panels.

In general terms, the use of nanotechnology for photovoltaic solar energy production can increase significantly, reflecting it a case of complete acceptance of the technology. In other words, the tendency to spread on the market is considerable.

In summary, the photovoltaic nano-panels are economically efficient and environmentally friendly. Considering the economies of scale, the positive effects could be unimaginable and the effects very positive in the economy, businesses, or in the preservation of the environment. The development of this technology would permit to overcome one of the greatest problems related to solar energy, it would eliminate the need for large areas for the installation of photovoltaic panels. After all, nanotechnology ensures for example that the photovoltaic cells can be adjusted to the ink in the companies buildings.

6. Ethics and Nanotechnology

According to Richard Feynman (1959), the principles of physics do not speak against the possibility of manipulating things atom by atom; it would not be a violation of the law; it is something that, theoretically, can be done, but that, in practice, was never carried out because we are too big.

The new characteristics of nanotechnology raise some hypotheses about the possibilities and risks brought by this new technology. The materials are changed on a scale that is not visible to human beings, even with the aid of a conventional microscope. Just to show the type of change made on nanoscale, a big investment in scientific equipment is need (see DIEESE, 2008, p. 03).

In the development of new technologies - and nanotechnology is not different on this aspect - economic agents should have in mind the search for the creation of mechanisms to monitor and control the entire process, allowing the society to be aware of both, the positive possibilities and the possible problems arising from the adoption of nanotechnology in the production of goods and services in the economy.

The principle should be established: what will be controlled and who will make the implementation of the control to be done. After all, we will be talking about something invisible to human eyes, therefore, in this case all the details are important.

It is possible that what we are dealing with is not as simple as it seems, it is a unprecedented challenge in human history. For example, in cases of negative externalities derived from leaks of toxic waste into a

river, these problems are easily perceived. However, in terms of nanotechnology, this kind of perception will not be possible. This means that the public control of the events will be quite different from what exists today.

In order to have some balance between the control and the exploitation of resources with nanotechnology, the communication between the actors involved will be important, especially among companies which use it and in the society in general.

Finally, the monitoring of the role of nanotechnology in the relationship between business and society will be crucial, by submitting it to permanent ratings, especially in what concerns to the evident importance for economic development.

7. Conclusions

The possibilities that nanotechnology offer to humanity are very huge and cover all the possible ways people connect with each other and with the natural resources for the production of goods and services.

At first, the recognition that there are significant dangers when there is a wrong use of them is evident. These risks are materialized under the prism of problems arising from human nature itself, as corruption and bad faith, or over-exploitation of resources.

However, the benefits arising from the use of nanotechnology may break paradigms and may contribute significantly to satisfy consumption needs of humanity. These advantages can be visible in particular in terms of reducing costs and enhancing the quality of both the products and the production method, and can be particularly seen for example for the case of electric power production that is based on nano-photovoltaic panels.

The use of nanotechnology requires some precautions. There are risks involved. This always happens with any kind of a new available technology with so many possibilities of use and which contours are difficult to define. It is important to stress that the benefits are not only to the productive units, but for the society as a whole. Not using it would certainly bring to the humanity a reduction in wealth and quality of life. Anyway, the balance between costs and benefits may take into account all the relevant elements of the analysis.

Acknowledgments

This work was financially supported by FCT through the Strategic Project PEst-OE/EGE/UI0315/2011.

References

[1] Aubin, J. P. (1979), *"Applied Functional Analysis"*. John Wiley & Sons Inc., New York.

[2] Becquerel, A. E. (1968). *La Lumiére, ses causes et ses effets.* Paris: Dominio Publico.

[3] Campos, M. (2002). *História da ética.* Rio de Janeiro: Tecnoprint.

[4] Chavaglia, J. N. (2010). *Energia solar fotovoltaica como vantagem competitiva em empresas industriais da Amazônia: ganhos de eficiência da empresa e preferências do consumidor.* Lisboa: ISCTE/IUL.

[5] Chavaglia, J. N. (2009). *Sucesso Profissional: as seis habilidades.* São Paulo: Barauna.

[6] Departamento Intersindical de Estatística e Estudos Sociais - DIEESE. (2008). Nanotecnologia: conhecer para enfrentar os desafios. *Nota Técnica* .

[7] Euroresidentes. (2011). *Introdução a Nanotecnologia: O que é Nanotecnologia*. Available in www.euroresidentes.com, assessed on 26 January 2011.

[8] Filipe, J. A. (2009). A note on the drama of the commons. How to deal with an old problem in natural resources: The resources overexploitation. *International Journal of Academic Research*.

[9] Filipe, J. A. (2006). *O drama dos recursos comuns. Um caso de aplicação da teoria dos jogos aos comuns da pesca.* Lisboa: PhD thesis ISCTE.

[10] Filipe, J. A., Coelho, M. F., and Ferreira, M. A. (2007). *O drama dos recursos comuns.* Lisboa: Edições Sílabo.

[11] Filipe, J. A., Ferreira, M. A. M., Coelho, M. and Pedro, M. I. (2010), Managing Complexity: a Problem of Chaos in Fisheries Policy. *China-USA Business Review. David Publishing Company, Chicago, IL, USA. 9 (3) / pp. 15-24.*

[12] Filipe, J. A., Ferreira, M. A. M., Coelho, M. and Chavaglia, J. N., Nanotechnology: Overpassing the limits of the human being. An ethical discussion applied to medical applications, in Silva, N. and Costa, G. J. M. (Ed.), *Ethical*

Dimensions in Bio-Nanotechnology- Present and Future Applications in Telemedicine, Hershey, USA: IGI Global. Forthcoming.

[13] Hansenclerver, L., and Kupfer, D. (2002). *Economia industrial.* Rio de Janeiro: Campus.

[14] IBGE. (2008). *O Brasil em numeros.* Rio de Janeiro: IBGE.

[15] Ikki, O. (2006). *National survey report of PV power applications in Japan.* Tóquio: International Agency Energy.

[16] Mateache, A. G. (s.d.). El parque solar más grande del mundo está en lá Mancha. *Mundo Magazine.*

[17] Mcafee, a., and Brynjolfsson, E. (2008). Diferença competitiva. *Havard Business Review* .

[18] Mcdonough III, E. F., Zach, M. H., Lin, H.-E., and Berdrow, I. (2009). Integrando a inovação e o conhecimento na estratégia. *Revista Portuguesa e Brasileira de gestão* .

[19] Mckibben, B., Rogers, J., and Nijhuis, M. (2009). *National Geographic* .

[20] Mckibben, B., Rogers, J., & Nijhuis, M. (2009). *National Geographic* .

[21] Nascimento, C. A. (2004). *principio de funcionamento da célula fotovoltaica.* Lavras: UFL.

[22] Nogami, O., and Passos, C. R. (1999). *Princípios de economia.* São Paulo: Pioneira.

[23] Peters, T. (1998). *O circulo da inovação.* São Paulo: Harbra.

[24] Pindyck, R., and Rubenfield, D. (1994). *Microeconomia.* São Paulo: Makron Books.

[25] Sandroni, P. (2002). *Novíssimo dicionário de economia.* São Paulo: Best Seller.

[26] Schumpeter, J. A. (1982). *Teoria do desenvolvimento economico.* São Paulo: Abril Cultural.

[27] Silva, A. C. (2003). *Nanotecnologia: uma iniciativa recomendada pela aeronáutica*. Acesso em 26 de Janeiro de 2011, disponível em www.defesabr.com.br.

[28] Taleb, N. N. (2009). *A Lógica do Cisne Negro*. Rio de Janeiro: Best Seller.

[29] Toothman, J.; Aldous, S. (2000), How solar cells work. Retrieved from http://science.howstuffworks.com/environmental/energy/solar-cell.htm.

[30] Varian, H. R. (2006). *Microeconomia: princípios básicos.* Rio de Janeiro: Campus.

[31] Vásques, A. S. (1982). *Ética.* Rio de Janeiro: Civilização.

[32] Wald, M. L. (2008). O charme e o poder das renováveis. *Scientific american*. Brazil. Special Edition 32.

[33] Wang, Z. L. (2010). *Nova Técnica para alimentação de dispositivos em escala nanométrica sem a necessidade de fontes de energia volumosas*. Available in http://www.azonano.com/news.aspx?newsID=2149&lang=pt. Assessed on 14 February 2011.

[34] Webber, M. E. (2008). Uma questão de opção. *Scientific american*. Brasil, São Paulo: Duetto.

[35] Wissing, L., Forschunzentrum, J., and Projekttrager, J. (2006). *National Report of PV power applications in Germany*. International Agency Energy.

[36] Zweibel, K., Mason, J., and Fthenakis, V. (2008). Energia solar. *American Scientific*. Brasil, São Paulo: Duetto.

Int. J Latest Trends Fin. Eco. Sc. Vol-2 No. 3 September, 2012

Output Persistence in Portugal

António Caleiro
Departamento de Economia & CEFAGE-UE
Universidade de Évora
Largo dos Colegiais, 2
7000-803 Évora
Portugal
Email:caleiro@uevora.pt

***Abstract* - The measurement of output persistence in Portugal is the main goal of this paper. By the use of a non-parametric methodology, it is shown that the level of output in Portugal exhibits a relatively high degree of persistence. This result is essential from a (contractionary/expansionary) policy point of view as the magnitude and duration of policy effects depend upon the level of persistence in output.**

Keywords: *Output, Persistence, Portugal.*

1. Introduction

The recent world economic and financial crises have been mitigated by several kinds of economic policies, in particular by fiscal measures designed to obtain immediate growth through countercyclical stimulus or to make the economy grow after a period of austerity. Plainly, the success of those policies, in particular the contractionary ones, depend upon the reaction of economic agents, in particular, and of the whole economy, in general.

The presence of persistence, here understood as inertia, can substantially change the reaction of the economy to a policy shock or to innovations. Persistence can reduce the incidence, length, and severity of shocks and of changes in economic conditions. Furthermore, measuring the response of output to a shock is also important because it may show when it is more essential to act to overcome the harmful effect of a shock.

Traditionally, macroeconomic policies play the dominant role in smoothing the business cycle, but the effectiveness of those policies depends upon the economy's resilience. That is, the success of those policies depends upon the ability of the economic system to absorb the policy shocks and to return to the baseline. Therefore, given the presence of persistence in output, the key question is whether it is viable and effective to design countercyclical policies that act through expenditures, even if they are optimal.

The literature on the importance of persistence in macroeconomics is inexplicably insufficient. The first macroeconomic studies incorporating the issue of persistence appeared only in the early 1980s, and only recently did a factual interest, from an empirical point of view, in the phenomenon emerge. The importance and the need to (theoretically and empirically) study the phenomenon are further strengthened by the current economic and financial crisis, in which the persistence of the recession is a central issue.

The first studies that explicitly considered the importance of persistence were of a macroeconomic nature and began by highlighting the role of both staggered wage-setting and staggered price-setting as a source of persistent real effects of monetary shocks (Taylor, 1980; Rotemberg and Woodford, 1997; Huang and Liu, 2002; see also Ascari (2003) for a critique of the real role of staggered wage-setting and staggered price-setting as sources of inertia). On the other hand, given the alleged inability of standard real business cycle models to reproduce the evolution of output shown under real-world conditions (Cogley and Nason 1995), the inertial hypothesis was also used to explain the (strong) persistence of output that was observed in reality (Bouakez and Kano, 2006; Maury and Tripier, 2003). However, this development did not lead to a consensus, and the possibility of monetary policy shocks affecting aggregate output remained central to the debate. Indeed, the persistence of shocks on aggregate output has been, and still is, one of the issues most often subject to examination, and this will probably be the case for some time.

Multiple theoretical explanations have been proposed for the empirical evidence that monetary policy shocks can have a permanent effect on aggregate output (or unemployment). These explanations include imperfect information and short-run nominal price stickiness (Kiley, 2000; Wang and Wen, 2006). Furthermore, Jonsson (1997), Lockwood (1997) and Svensson (1997) have analyzed the consequences of inflation contracts on output or unemployment persistence. All these studies share the idea that whether or not price rigidity is responsible for output or unemployment persistence, this should be seen as an empirical issue rather than a theoretical one.

Another interesting consequence of output persistence is that it may invert the political business cycle, which is typically associated with depressions at the beginning of the mandate followed by pre-election inflationary expansion (Gärtner, 1996, 1999;

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.004.12/ijltfes.2047.0916

Caleiro, 2009). Quite recently, increased interest in analyzing the persistence of output and inflation has been registered, and this has included studies of their relationship with the degree of openness of the economies (Guender, 2006), the exchange-rate regime (Giugale and Korobow, 2000) or the structural changes in the preferences of consumers, firms or policy-makers. For the case of consumption in Portugal see Belbute and Caleiro (2009).

The goal of the paper is to contribute to the understanding of the effects of the policy measures recently taken in Portugal in the aftermath of the financial aid requested in May 2011. We do so by measuring the degree of persistence in output through the use of a non-parametric methodology proposed by Marques (2004) and Dias and Marques (2010). This new measure of persistence can be defined as the unconditional probability that a stationary stochastic process will not cross its mean during period *t*.

Our results show that the level of output in Portugal exhibits a fairly high degree of persistence. Plainly, this result is essential from a (contractionary/expansionary) policy point of view as the magnitude and duration of policy effects depend upon the level of persistence in output.

The remaining of the paper is organized as follows. In Section 2, some methodological notes about persistence are offered. Section 3 presents the data. Section 4 is fulfilled with the empirical results. Section 5 concludes the paper.

2. Methodological notes about the persistence

Starting with a simple definition, persistence is the speed with which a certain variable, such as output, returns to baseline (its previous level) after, say a shock, i.e. some event (for instance, a fiscal policy measure) that provoked an increase (or decrease) in output. This definition, in other words, implies that the degree of output persistence is associated with the speed with which output responds to a shock. When the value is high, output responds quickly to a shock. On the contrary, when the value is small, the speed of adjustment by output is low. To put it clearer, a variable is said to be the more persistent the slower it converges or returns to its previous level, after the occurrence of a shock.

Quantifying the response of output to a shock is indeed important not only because it may allow assessing the effectiveness of economic policy measures but also because it may, indeed, show at what time is more essential to act, through those measures, in order to overwhelm a harmful effect of a shock over output. By definition, quantifying the response of output to shocks implies evaluating the persistence of output.

As the estimates of persistence at time *t* will express how long we expect that a shock to output will take to die off (if ever), given present and *past* output, authors have proposed to obtain those estimates by the use of *autoregressive models*. As it is well known, a univariate AR(k) process is characterised by the following expression:

$$f_t = \mu + \sum_{j=1}^{k} \alpha_j f_{t-j} + \varepsilon_t \tag{1}$$

where f_t denotes output at moment *t*, which is explained by a constant μ, by past values up to lag *k*, as well as by a number of other factors, whose effect is captured by the random variable ε_t. Plainly, (1) can also be written as:

$$\Delta f_t = \mu + \sum_{j=1}^{k-1} \delta_j \Delta f_{t-j} + (\rho - 1) f_{t-1} + \varepsilon_t \tag{2}$$

where

$$\rho = \sum_{j=1}^{k} \alpha_j \tag{3}$$

and $\delta_j = -\sum_{i=j+1}^{k} \alpha_i$.

In the context of the above model (1), or (2), persistence can be defined as the speed with which output converges to its previous level after a shock in the disturbance term that raises output at moment *t* by 1%.[1]

The techniques allowing for measuring the persistence are based on the analysis of the autoregressive coefficients α_j in (1) or (2), which are subject to a statistical estimation. Plainly, the most simple case of the models (1) or (2) is the so-called AR(1) model:

$$f_j = \mu + \alpha_1 f_{t-1} + \varepsilon_t . \tag{4}$$

[1] Given that the persistence is a long-run effect of a shock to output, this concept is intimately linked to a concept usually associated to autoregressive models such as (1) or (2), i.e. the impulse response function of output, which, in fact, is not a useful measure of persistence since its infinite length.

Clearly, the variable ε_t in this kind of models has a particular importance given that it may be associated with policy measures leading to a shock in output. A positive shock, at moment *t*, will significantly last for future moments the higher is the autoregressive coefficient α_1. Following this approach, Andrews and Chen (1994) proposed the sum of the autoregressive coefficients, $\rho = \sum_{j=1}^{k} \alpha_j$, as a measure of persistence.[2] The rationale for this measure comes from realizing that for $|\rho| < 1$, the cumulative effect of a shock on output is given by $(1-\rho)^{-1}$.

Quite recently, Marques (2004) and Dias and Marques (2010) have suggested a non-parametric measure of persistence, γ, based on the relationship between persistence and mean reversion. In particular, Marques (2004) and Dias and Marques (2010) suggested using the statistic:

$$\gamma = 1 - \frac{n}{T}, \qquad (5)$$

where *n* stands for the number of times the series crosses the mean during a time interval with $T + 1$ observations, to measure the absence of mean reversion of a given series, given that it may be seen as the unconditional probability of that given series not crossing its mean in period *t*.[3]

As Dias and Marques (2010) have shown, there is a one-to-one correspondence between the sum of autoregressive coefficients, ρ, given by (3) and the non-parametric measure, γ, given by (5), when the data are generated by an AR(1) process, but such a one-to-one correspondence ceases to exist once higher order autoregressive processes are considered. In other words, only in the particular case of a first-order autoregressive model, AR(1), either one of the two measures can be used to quantify the level of persistence, as both transmit the same result, but as soon as higher order autoregressive models are considered, *i.e.*, AR(*k*) with $k \geq 2$, the monotonic relationship between ρ and γ no longer exists, therefore leading to possibly crucial differences when measuring persistence in the series.

As Dias and Marques (2010) show, using the alternative measure of persistence, γ, given by (5), has some important advantages.[4] Given its nature, such measure of persistence does not impose the need to assume a particular specification for the data generation process, therefore does not require a model for the series under investigation to be specified and estimated.[5] This is so given that γ is indeed extracting all the information about the persistence from the data itself. As it measures how often the series reverts to its means and (high/low) persistence exactly means that, after a shock, the series reverts to or crosses its means more (seldom/frequently), one does not need to specify a particular form for the data generation process.

3. The data

We use annual data for the period from 1970 to 2011 for Portuguese GDP, measured in millions of euros, at constant prices, OECD base year = 2005 (see Figure 1).

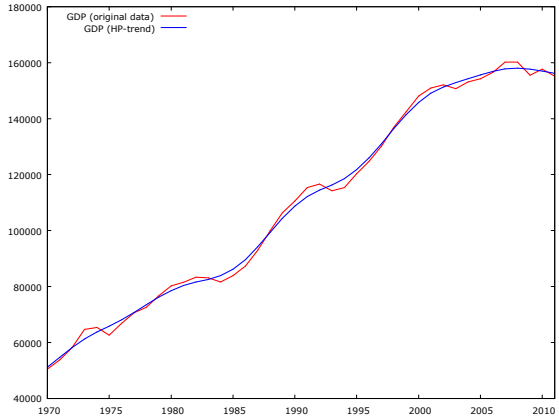


Figure 1: The evolution of GDP (source OECD) and its HP-trend

The smooth line in Figure 1 corresponds to the trend – obtained by the Hodrick-Prescott filter, to be described below – clearly identifies a relatively long-lasting period of generalized growth, followed by a decline in output after 2007. Naturally, around this trend, some cyclical component can also be (easily) identified.

4. The Level of Output Persistence in Portugal

Clearly, the time series of GDP exhibits a non-stationary behavior, which makes it necessary to use

[2] Authors have, indeed, proposed other alternative measures of persistence, such as the largest autoregressive root, the spectrum at zero frequency, or the so called half-life. For a technical appraisal of these other measures see, for instance, Marques (2004) and Dias and Marques (2004).
[3] As acknowledged in Marques (2004), values close to 0.5 indicate the absence of any significant persistence (white noise behaviour) while figures significantly above 0.5 signal significant (positive) persistence.
[4] The statistical properties of γ are extensively analysed in Marques (2004) and Dias and Marques (2010).
[5] In technical terms, this means that the measure is expected to be robust against potential model misspecifications and given its non-parametric nature also against outliers in the data.

a non-parametric measure of persistence, such as given by (5). In order to compute the estimative γ, the mean has to be computed. As suggested in Marques (2004) and Marques and Dias (2010), a time varying mean is more appropriate than the simple average for all the period under investigation. In our case we followed that suggestion by using the well-known Hodrick-Prescott (HP) filter in order to compute the mean (Hodrick and Prescott, 1981).

As it is well known, the HP filter defines the trend or mean, g_t, of a time series, f_t, as the solution to the minimisation problem:

$$\min_{\{g_t\}}\left\{\sum_{t=1}^{T}(f_t-g_t)^2+\lambda\sum_{t=2}^{T-1}((g_{t+1}-g_t)-(g_t-g_{t-1}))^2\right\}$$

i.e. the HP-filter seeks to minimise the cyclical component ($f_t - g_t$) subject to a smoothness condition reflected in the second term. The higher the parameter λ, the smoother will be the trend and the less deviations from trend will be penalised. In the limit, as λ goes to infinity, the filter will choose ($g_{t+1} - g_t$) = ($g_t - g_{t-1}$), for $t = 2,\ldots,T-1$, which just amounts to a linear trend. Conversely, for λ = 0, the original series is obtained.

Obviously, the HP-filter is a very flexible device since it allows us to approximate many commonly used filters by choosing appropriate values of λ. Given that the data is of yearly frequency, authors have suggested using values for λ around 6 (Ravn and Uhlig, 2002). Considering this value for λ, the measure of persistence, γ, given by (5) was computed recursively, augmenting one year to the sample, in the case starting with the time period [1970,1979], and ending with [1970,2011]. This amounts to say that the original series, f_t, was decomposed in a HP-trend, g_t, and a remaining cyclical component, from which a change in sign identify a year where output has crossed its mean; see (5).

Figure 2 shows the results of that procedure, in terms of the values registered by the non-parametric measure of persistence, γ, given by (5).

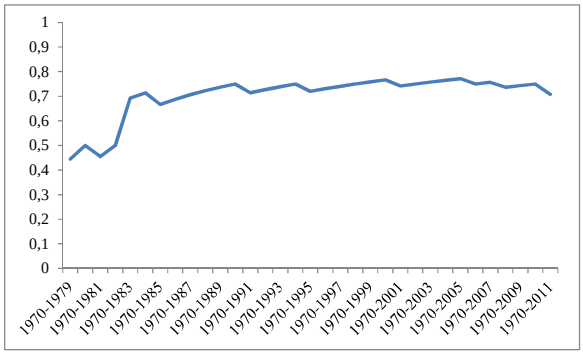


Figure 2: The level of output persistence in Portugal

As it can be seen, the Portuguese GDP exhibits a fairly high level of persistence, whose non-parametric measure has apparently stabilized at around 0.75.

5. Conclusion

This paper has explored the question of output persistence in Portugal. It is used a new methodology proposed by Marques (2004) and Dias and Marques (2010) to measure persistence using a model that is non-parametric and broader in scope than other measures used in the literature, particularly the sum of the autoregressive coefficients.

The main conclusion is that the Portuguese GDP is characterized by a quite high level of inertia, which seems to have stabilized. This result calls the attention for the long-lasting effects of economic policies, whether of contractionary or expansionary nature.

As directions for further research we would like to further explore the possible asymmetry in the effects of shocks to output (Beaudry and Koop, 1993; Elwood, 1998).

References

[1] Andrews, D. and W.K. Chen (1994), "Approximately Median-Unbiased Estimation of Autoregressive Models", Journal of Business and Economic Statistics, 12, 187-204.
[2] Ascari, G. (2003), "Price/Wage Staggering and Persistence: A Unifying Framework", Journal of Economic Surveys, 17, 511-540.
[3] Beaudry, P. and G. Koop (1993), "Do recessions permanently change output?", Journal of Monetary Economics, 31(2), 149-163.
[4] Belbute, J. and A. Caleiro (2009), "Measuring the Persistence on Consumption in Portugal", Journal of Applied Economic Sciences, IV, 2(8) (Summer), 185-202.
[5] Bouakez, H. and T. Kano (2006), "Learning-By-Doing or Habit Formation?", Review of Economic Dynamics, 9, 508-524.
[6] Caleiro, A. (2009), "How Upside Down are Political Business Cycles when there is Output Persistence", Research in Economics, 63(1), 22-26.
[7] Cogley, T. and J. Nason (1995), "Output Dynamics in Real Business Cycle Models", American Economic Review, 85, 492-511.
[8] Dias, D. and C. R. Marques (2010), "Using Mean Reversion as a Measure of Persistence", Economic Modelling, 27, 262-273.
[9] Elwood, S.K. (1998), "Is the persistence of shocks to output asymmetric?", Journal of

Monetary Economics, 41(2), 411-426.
[10] Gärtner, M. (1996), "Political Business Cycles When Real Activity is Persistent", Journal of Macroeconomics, 18, 679-692.
[11] Gärtner, M. (1999), "The Election Cycle in the Inflation Bias: Evidence from the G-7 Countries", European Journal of Political Economy, 15, 705-725.
[12] Giugale, M. and A. Korobow (2000), "Shock Persistence and the Choice of Foreign Exchange Regime: An Empirical Note from Mexico", World Bank Policy Research Working Paper No. 2371.
[13] Guender, A. (2006), "Stabilising Properties of Discretionary Monetary Policies in a Small Open Economy", The Economic Journal, 116, 309-326.
[14] Hodrick, R. and E. Prescott (1981), "Post-war U.S. Business Cycles: An Empirical Investigation", Working Paper, Carnegie-Mellon, University. Reprinted in Journal of Money, Credit and Banking, 29(1) (February 1997).
[15] Huang, K. and Z. Liu (2002), "Staggered Price-Setting, Staggered Wage-Setting and Business Cycle Persistence", Journal of Monetary Economics, 49, 405-433.
[16] Jonsson, G. (1997), "Monetary Politics and Unemployment Persistence", Journal of Monetary Economics, 39, 303-325.
[17] Kiley, M. (2000), "Endogenous Price Stickiness and Business Cycle Persistence", Journal of Money, Credit and Banking, 32, 28-53.
[18] Lockwood, B. (1997), "State-Contingent Inflation Contracts and Unemployment Persistence", Journal of Money, Credit, and Banking, 29, 286-299.
[19] Marques, C. R. (2004), "Inflation Persistence: Facts or Artefacts?", Working Paper 8, Banco de Portugal, June.
[20] Maury, T.-P. and F. Tripier (2003), "Output Persistence in Human Capital-Based Growth Models", Economics Bulletin, 5, 1-8.
[21] Ravn, M. and H. Uhlig (2002), "On adjusting the HP-filter for the frequency of observations", Review of Economics and Statistics, 84, 371-375.
[22] Rotemberg, J. and M. Woodford (1997), "An Optimization Based Econometric Framework for the Evaluation of Monetary Policy", In: Rotemberg, J., Bernanke, B., (Eds.). NBER Macroeconomics Annual 1997. Cambridge (MA): The MIT Press, 297-346.
[23] Svensson, L. (1997), "Optimal Inflation Targets, "Conservative," Central Banks, and Linear Inflation Contracts", American Economic Review, 87, 98-114.
[24] Taylor, J., (1980), "Aggregate Dynamics and Staggered Contracts," Journal of Political Economy, 88, 1-23.
[25] Wang, P.-f. and Y. Wen (2006), "Another Look at Sticky Prices and Output Persistence," Journal of Economic Dynamics and Control, 30, 2533-2552.

Comparing Agricultural Land-use Statistics from Different Sources: a Case Study in Greece*

Vasiliki Benaki [1], Luca Salvati [2], Matteo Barone[3], Carlotta Ferrara [3]

[1] *National Statistical Service of Greece, Pireos 46 & Eponiton, GR-18510 Pireas*

[2] *Italian Agricultural Research Council, Centre for the Study of Soil-Plant Interactions, Via della Navicella 2-4, I-00186 Rome*

[3] *Department of Physics "E. Amaldi", University of Roma Tre, Via della Vasca Navale, 84, 00146 Rome, Italy*

[1] vasbenaki@statistics.gr

[2] luca.salvati@entecra.it (corresponding author)

[3] ferrara.fis@uniroma3.it

** A preliminary draft of this contribution was presented in the 'First Conference of the Hellenic Association of Information and Communication Technology in Agriculture, Food and Environment' (Athens, 2002).*

Abstract **- This paper illustrates a tool integrating the Farm Structure Survey (FSS) and an improved version of CORINE Land Cover (CLC) map as a contribution to produce reliable land-use statistics at national and regional scale. To achieve compatibility between census and CLC the present tool takes into account the FSS nomenclature and definitions, and reorganizes the 44 classes of the original CLC into 16 general classes that meet the needs of the Land Use/Cover statistics in Greece. To compare the respective surface areas of the related classes and to provide the appropriate links between certain classes of the two nomenclatures four Greek regions are used in the pilot study: Kriti, and the three prefectures of Makedonia. The linkage between the two databases shows the existing differences between the administrative areas. The developed tool is able to relate data from different sources, and to display on a map, the combined spatial statistical data along with the geographical information of the area. Thus, although the new CLC seems to provide a good mapping base, the imposed minimum mapping unit of 25 ha results in an overall underestimation of the diversity of agricultural land-uses, something which is particularly important in the case of Greece for which the average size of the farm holdings is less than 5 hectares.**

Keywords *- Farm Survey Structure, Corine Land Cover, Landscape diversity, Greece*

1. Introduction

Agricultural activities are more and more combined with other activities such as environmental protection, maintaining the landscape, forestry, preserving recreational and tourist areas as well as small scale of agricultural products. Thus, there is a strong need for statistical data on rural population, and particularly, on landscape and land-use. Most of the statistical data used for policy purposes are related to populations, activities, features and other events, which are by their nature, spatial in form. The management, the process and the display of statistical data associated with spatial locations that vary geographically is mainly a spatial process. In agricultural terms, the management of agricultural resources is increasingly complex as conservation and environmental concerns play an expanding role for making conclusions. In this respect, Geographic Information Systems are needed in the production of census maps, for dealing with census logistics, for monitoring census activities, and for data dissemination (Deichmann, 1997). A wide range of spatial analysis methods has been developed for carrying out data transformations between different spatial structures. These methods help to present the data in a more meaningful and consistent manner and enable different data sets, based on different geographical units, to be brought together and overlaid. They also facilitate the spatial analysis of the statistical data required in the development of some more reliable indicators for the determination of the state and quality of the environment, able to measure the effect of the agricultural economy, across regions and countries. The use of indicators as an aid to policy decision-making in the agri-environmental context is a relatively recent phenomenon and still a developing field; however, indicators are perceived to have considerable potential as policy tools. Most policy makers

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.005.12/ijltfes.2047.0916

concerned with agri-environmental issues at the national level are confronted with fragmented information and it is accordingly difficult to use the information in a way that effectively contributes to policy decision-making. An unavoidable step in the assessment of agricultural policies and of their impact on the countryside and landscapes is the study of spatial units that constitute the underlying structure of these territories. Most statistical data in EU, by means of the Farm Structure Survey (FSS) data, is organized and presented on the basis of NUTS (Nomenclature des Unites Territoriales Statistiques) system, to provide a single, uniform breakdown of a country. Nevertheless, these units are geographical areas that may vary substantially not only in the sizes and shapes, but also over the time. In addition, this geographical level is not appropriate to carry out certain environmental studies. The need of spatial analysis and of the production of environmental indicators requires delineation of the land use data according to their natural depiction on a geographical map, beyond the administrative distribution. As a result, NUTS system cannot be applied in its present form to units that are more relevant from a geographical point of view, such as drainage areas, landscape units, biotopes, etc. This study illustrates the development of a tool interface between statistical and geographical databases by comparing Farm Structure Survey (FSS) and CORINE Land Cover (CLC) land-use figures. As a first step, the spatial disaggregation of the FSS data into an accurate geographical level requires an interface between the two nomenclatures. To reallocate the FSS data into sustainable areas a question arises of how the digital CLC map could be used to describe agro-environmental statistical structures. Note that CLC has so far been focused on land cover, rather than land use and it has been carried out once. As a result it cannot be applied to show trends. However, different countries carried it out in different years, over the period 1985- 1995. Plans already exist to upgrade CLC based on the IMAGE 2000 image data set provided by the JRC. The result is that some of the indicators based on CLC show only a snapshot rather than a trend in land-use. The developed interface is able to display on a map, accurately, the combined spatial descriptive statistical data along with the geographic information of an area of interest. Thus, the user is able to relate the FSS and the CLC data in order to find the best matching. The developed interface is able to query a database, aggregate / disaggregate the data and plot the results on a map. The comparison requires to determine the aggregation level of the classes for which the correspondence has already been set and to validate the result by comparing the respective surface areas of the related classes. After the reclassification of the above data, common classes are created and presented on a map using an embedded GIS environment. To test the interface and provide the appropriate links between certain classes of the two databases the three regions of Makedonia and the region of the island of Kriti have been chosen. The statistical data used has been provided by the Basic FSS of 1999/2000 (Census of Agricultural for Livestock breeding or simply Agricultural Census). However, to achieve compatibility between census and photo-interpretation a recently developed, improved version of the CLC geographical database has been used. The new CLC takes into account the FSS nomenclature and definitions and has provides much better acquisition period (Landsat-TM 1998 to 1999) which is the same with the census reference period (1998 to 1999). The linkage between the two databases shows the existing differences between the administrative areas of the pilot regions. The structure of the paper is as follows: chapter 2 describes the main characteristics of the FSS nomenclature, particularly addressed in the case of Greece. Chapter 3 addresses the modified CLC geographical nomenclature providing the new classification scheme. Also, in this section, the original CLC nomenclature is discussed briefly. Chapter 4 discusses the linkage between the two nomenclatures and the way it has been achieved by means of application development. Chapter 5 illustrates the results starting from the comparison of the related nomenclatures and finally, in the last section the conclusions of this work are presented.

2. The FSS database

2.1 Main issues

The effective and balanced implementation of the reformed Common Agricultural Policy requires detailed objective, quantitative data of the structure and performance of the agricultural, rural and environmental sectors. In this context the development of the structure of the agricultural holdings allows analysis of the agricultural sector and its impact to other sectors as the rural sector and the environment. The FSS is the main source to provide data on various characteristics relating to agricultural holdings, on a regular basis. These data refer to the number and size distribution of the

agricultural holdings by type of enterprise, as well as to the land improvements, crop and livestock rotations and farm practices (machinery, equipment etc.). They also refer to other structural data such as the educational level of farmer and farm labor inputs, the legal status of holder including tenure arrangements and finally other social demographic characteristics of holders. The FSS data are collected on a regular basis by the Member States and are forwarded to Eurostat, which stores them in the Eurofarm database. In order to harmonize information at the Community level, legal frameworks (Regulation and Decisions) define the methodological framework and the contents of the FSS questionnaires. Table 1 shows the FSS nomenclature, which distinguishes the detailed agricultural land use classes.

2.2 Methodological issues of the FSS in Greece

The FSS is carried out in Greece within the framework of the Community Program for the 'Statistical Surveys in the Agricultural Sector'. All the specifications and terms are defined precisely by the Regulation 571/88 as amended by the Regulation 2467/96 and the related Decisions of the Council of the E.U. The FSS is intended to collect statistical data on the structure of agricultural and livestock holdings and the employment of the population on them. The data make it possible, besides the classical tabulation of the results, to generate tables, which show the economic size and orientation of the farms (typology). In particular, the Greek FSS system aims to collect data relating to:

- The number of agricultural and livestock holdings in the country, at national, regional and local level.
- The geographical position of the holdings.
- The legal status and management of the holding.
- The agricultural training of its owner.
- The keeping of account books.
- The land uses (arable crops, permanent crops, kitchen gardens, permanent pasture-meadows and rough grazing and other areas).
- The type of ownership of the utilized agricultural area.
- The number of fields constituting the total utilized agricultural area.
- Successive crops, combined crops, irrigated crops, etc..
- Livestock raised on the farm..
- Agricultural machinery and milking equipment used.
- Employment of members of the farm owner's family.
- Employment of family members in other gainful activities besides agriculture.
- Employment of permanent, seasonal and other workers.

Sample FSS is carried out every two years, in the years ending with an odd number. The National Statistical Service of Greece (NSSG) carried out the first sample survey of the Structure of Agricultural and Livestock breeding in 1966/67, when Greece was still an associated member of the EU. The next sample survey took place in 1977/78. After the accession of the country into the EU further surveys were carried out in 1983, 1985, 1987, 1989, 1993,1995 and 1997 i.e. every two years. Every ten years an exhaustive survey (Basic FSS or Agricultural Census) is carried out. The first Agricultural Census conducted in 1950, after the Second World War. Agricultural Census of 1991 was the last census carried out at the same time with the General Censuses for population, households, agriculture etc. However, Agricultural Census in 2000 was the first census carried out before the General Population Census dated 2001 and it was based on the Farm Register. The reference period for the data collected on crops and employment is from October 1st of year t-1, to 30 September of year t, i.e. the survey year. Exceptions to this are a farm's livestock and machinery, questions relating to which have a reference date of 30 September in the year t, for the machinery and 1st November for the livestock. The statistical unit for the FSS is defined as an agricultural or livestock holding4 which during the reference period comprises at least one of the following:

- at least 0.1 ha of utilized agricultural area or at least 0.05 ha of greenhouse area, regardless of its own ship and location, or
- at least one cow, or

- at least two other large animals (oxen, buffaloes, horses, etc.), or
- at least five small animals (sheep, goats, pigs), or
- at least 50 poultry birds, or
- at least 20 beehives.

The FSS was carried out by filling in a special questionnaire after interviewing the owner of the agricultural or livestock breeding farm. The sample survey is carried out by applying the method of multi-stage stratified area sampling. In the most recent Agricultural Census in Greece the Basic FSS covered all agricultural and livestock holdings in the country (nearly 814.000 holdings).

3. Materials and methods

3.1 The CLC geographical database

CORINE (Co-ORdination on INformation of the Environment) Land Cover (CLC) is a geographic land cover/land use database encompassing most of the countries of the European Community, with aim to gather information associated with the environment on certain priority topics. It describes land cover (and partly land use) according to a nomenclature of 44 classes organized hierarchically in three levels (Dueker, 1979). CLC was elaborated based on the visual interpretation of satellite images *(Spot, Landsat TM* and *MSS).* The smallest surfaces mapped (mapping units) correspond to 25 hectares. Linear features less than 100m in width are not considered. The scale of the output product was fixed at 1:100.000. Thus, the location precision of the CLC database is 100m. Although its exploitation is just starting, it offers the potential for a wide array of uses. It can be used on its own for simple cartographic or statistical presentations and as a base for European-wide landscape analyses or more generally in combination with other data sets.

3.2 The CLC database of Greece

The CLC database has been developed in Greece in order to cover the needs of land use/cover statistics as far as the distribution of the total area of Greece in the basic categories of land use is concerned. These statistics are included in the preparatory work carried out in the context of every Agricultural Censuses. The aim is to prepare the census and to obtain data covering all the territory of Greece. Until Agricultural Census of 1991, this work was done by completing seven (7) months before the Census a 'pre-census questionnaire of total land area in the municipality or commune', using estimates by the municipal or communal working parties set up for the census and with the help of local agronomists. To facilitate completion of the pre-census questionnaire, these groups had at their disposal the land distribution data from the previous census, as well as other auxiliary data held by the municipality or commune, such as land registers, land distribution tables, etc. Land was divided up into seven basic categories of use:

- Cultivated areas and fallow land resting fallow for 1 to 5 years.
- Communal or municipal pasture land.
- Other pasture land (owned by privates, State, monasteries, etc.)
- Forests
- Areas under water (lakes, marshes, seashores, river beds)
- Build-up areas (buildings, courtyards and roads, squares etc.)
- Other areas (e.g. rocky areas, mines, etc.).

Note that the pre-census questionnaire was the only data source covering also the state-owned land, which is mostly, forest and pastures. Nevertheless, since the agricultural census is carried out by interviews of farmers it concerns only private lands that are somehow agriculturally used.

In the light of the recent developments concerning land use statistics, NSSG decided to use an up-to-date methodology using GIS techniques in order to produce more objective information on this sector. Therefore, the use of spatial analysis is required. Spatial analysis of the information to be recorded is realized by determining the area of the minimum recorded surface, which is taken according to the proposed nomenclature, the methodology of use/cover definition, the requirements of 1:100.000 scale and the user needs. The method with which the theme information drawn up, is the comparative photo-interpretation of the satellite data collected in 1998-99 in relation to those of the time period 1997-98 used for the creation of the CLC database in Greece. The digital photo-interpretation of the new satellite data is made using image processing software and other data such as those from land recordings. The recording planning and the use of the

data from the field works are also defining the reliability of the specific photo-interpretation.

The new CLC database is properly generalized as reference data and harmonized with the FSS nomenclature, by means of characteristics and definitions, linkage of the two databases to meet the needs of the NSSG. Thus, the distribution of the main land uses in Greece has been reorganized into the following sixteen (16) classes:

- Artificial surfaces
  1. Urban fabric
  2. Industrial and commercial units
  3. Transport units
  4. Mine, dump and construction sites
  5. Artificial, non-agricultural vegetated areas sport and cultural activity sites
- Agricultural areas
  6. Arable land
  7. Permanent crops
  8. Pastures
  9. Heterogeneous agricultural areas
- Forest and semi-natural areas
  10. Forests
  11. Transitional woodland /shrub
  12. Shrub and/or herbaceous vegetation associations
  13. Open spaces with little or no vegetation
- Surfaces under water
  14. Inland water
  15. Inland wetlands
  16. Coastal wetlands

The new CLC geographical database for the country's area has numerous advantages, the most important of which are the following:

- It provides a land use/cover map covering all Greece for 16 categories, compiled with the seven land use classes in the above mentioned pre-census questionnaire of the NSSG.
- The new geographical database takes into account the FSS nomenclature and definitions.
- It enables comparability between the two sources of information, namely census versus photo-interpretation. In the case of Greece the acquisition period of the data is spread over 2 years for both, the CLC (Landsat-TM 1998 to1999) and the FSS 1999/2000, (reference year the 1998-1999 crop year).
- It enables the integration of the chrono-geographical co-ordinates of the satellite images sources of CLC. This will help in the identification of districts for which CLC's image interpretation is one year apart (minus or plus) from the census year (1990 or 2000, respectively). In addition, using the intermediate FSS data that correspond closely to the date of the satellite image it will be possible to mitigate the effect of time.

4. Linking the two databases

As it is well known, data collection methods are optimized for a particular need and therefore the resultant data structures are not usually readily comparable in a cross-sectional study. Thus, although a particular census may be analyzed in detail comparing censuses with each other have been proved problematic since they may use different administrative units, or they may use the same unit system, which includes many boundary changes that make the comparison difficult. To the best of our knowledge, three types of data incompatibilities have been distinguished so far (Frank, 1999; Gregory, 2000) and will be described, briefly, below.

4.1 Differences in Data Models

Raster and vector data models are the GIS approaches for the spatial presentation of natural vegetation, the forest area and generally the development of land use. In a raster data model, a uniform grid, each cell of which is assigned a unique descriptor depending on the coordinate system used, represents space. Raster models can be directly imported into the software and immediately become available for use (Burrough, 1986). They are well suited for the representation of remotely sensed digital data and are commonly employed in the environmental sciences. In contrast, in a vector data model, the spatial data is based on geometric shapes of points, lines, and polygons. This model is object-oriented and is based on the coordinate system used.

Vector GIS knows where the spatial feature (line, point, polygon) exists, as well as the relationship with the other features. Vector data models are particularly suited for the representation of linear data features like roads, or clearly delineated areas, such as, property lines and city limits After the representation of the spatial features, their associated properties must be specified in a separate database. For simultaneous use of data from both, raster and vector models a conversion of one data set to the respective model of the other data set needs to be performed. Data conversions, however, are often ambiguous and typically result in a loss of information (Maffini et al., 1987). It is difficult, for example, to derive the best fitting vector representation from a given raster grid. Furthermore, it is likely that a set of transformations from vector to raster and back to vector will result in a target feature whose shape differs from the original source feature. These transformation functions may not be accessible to the end user (Ehlers et al., 1991, Maffini et al., 1987). The data transformation from analog paper maps and tabulations to digital data falls into the same category. Loss of data, spatial inaccuracy and error is introduced by conversion techniques like scanning, digitizing, rasterization, vectorization and manual data input (Goodchild, et al., 1989). Digital data creation is also extremely time consuming.

4.1.1 Inconsistencies of areal units

Comparing census data with other data sources of some specific area of interest may not be the same, either because of boundaries changes over the time or because of the different definition of administrative/areal units used for the data collection (Xie et al., 1995). Two key issues need to be addressed in terms of areal unit comparability. One is related to data integration and map overlay (non-matching areal units). The other is related to data analysis and statistical comparability of areal units of different sizes and shapes (modifiable areal unit problem).

4.1.2 Non-matching Areal Units

Integrated analysis of spatial and attribute data is based on map overlay operations. Non-matching areal units require a transformation of data from one system of areal units to another in which data values are apportioned to the newly created spatial units. Then, the newly created zones allow data overlay and analysis. These transformations are known as "areal interpolation".

4.1.3 Modifiable Areal Unit Problem (MAUP)

Generally, the statistical data, whose distribution and characteristics are not well known, are presented by an appropriate aggregated variable of some higher class. In addition, censuses base their statistics on well-defined areal units that tend to vary in size and shape leading to inconsistent and misleading statistical results. This is known as a Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984). A possible approach to face this problem would be the reaggreation of the available data into homogeneous subunits and the increase the spatial detail using ancillary land cover data in order to display the census data on a map (Yuan et al., 1997).

4.1.4 Temporal incompatibilities

Data collection of land cover data and monitoring of physical changes relies on remote sensing via aircraft or satellites. Coverage cycles, for example, for the different LandSat orbiters (Lillesand, et al., 1994) range from 8-14 days. This data density, however, may be deceptive since data for certain regions is usually available for much fewer dates due to the fact that frequent cloud cover prohibits data collection. In the case of agricultural cultivations, the above problem has to be considered in more detail since the cultivations are usually visible on specific periods over the year.

4.2 Technical procedure

To describe the methodology adopted in the investigated issue, one has to take into account the non-matching areal units and the MAUP problems mentioned in section 4.1. The temporal incompatibilities problem and the procedure of matching the data points by non-matching due to collection cycles will not be considered because in our case, the data used has been interpreted by an independent intuitional organization. The Non-matching Areal Units problem arises due to the following reasons.

The different boundaries definition of the administrative units that have been used in the Hellenic censuses of the year 1991 and 2000. To solve the problem there is the need to transform the data between different spatial structures. As transformation may be described the process of aggregation and disaggregation within nested and non-nested neighbor polygons. To overlay the data together the conceptual model has been designed containing and maintaining all polygons and the

related geometric data (lines, nodes etc), representing the areal units. To link the descriptive and the spatial information, the data of the geographic area of interest is broken down into smaller parts in order to determine the field that identifies the specific entity (`AreaCode') in order to be used as a reference key to the GIS. A set of spatial queries also has been developed for carrying out the transformations.

Moreover, the different geodetic datums used for the presentation of statistical and for the ancillary geographical data. To use the ancillary data along with the other geographical data a target datum has to be selected as the reference datum for all data and an automated procedure has to be developed to convert the data between the source and the selected target geodetic datum.

To automate both the transformation between different definitions of administrative units and the conversion between different geodetic datums, an object has been designed, called "Geo-Object". This Object can be used as the basic map layer on any similar application. The MAUP problem is faced using ancillary geographic data (Flowerdew et al., 1991) such as contour lines, lines representing rivers or polygons representing lakes. This methodology permits the synthesis of geographical data along with the studied statistical data and allows the combination of different scenarios in order to simulate the plotting of the descriptive data containing quantitative and areal information on a map. For validation and / or prediction purposes, the results are compared visually with other spatial quantitative information or sampling data presented on thematic maps.

A database entity object provides the connection between the conceptual model and the input quantitative data. It has only one method that is used to insert a new 'AreaCode' into the conceptual model and finally it connects the new inserted 'AreaCode' with the appropriate areal information. A tab-delimited text file contains some quantitative information in a country level. To link this text with the appropriate geographic feature in the GIS environment, the entity "Details" of the conceptual model must be updated with the 'AreaCode' of countries that this text file contains. This can be done easily by using the properties and methods of the above described class of objects.

The advantage of the described methodology is the capability to combine quantitative data from different sources, and to compare them with the available spatial features concerning the distribution of similar quantitative data (thematic map) of the same area of interest, into an integrated geographical environment. This environment can contain more geographical features than the ordinary thematic maps such as contour lines, roads, rivers, airport, etc.

This can also be helpful for the methods that have been developed to solve the problem of geographic missing values. The precision of those methods is depended on the availability and quantity of historical data. Using this methodology, it is obvious that the integration of all types of GIS data with quantitative data available from other sources is crucial for someone to decide about the data correctness.

4.3 Study site application

As it has been pointed out, the linkage of the two nomenclatures, by means of the FSS the CLC databases require computer-based application software able to display maps and descriptive data in a tabular form. This has been achieved using geographical information from CLC database linked with tabular information of the multi-dimensional tables of FSS (Table 2). The user becomes part of the GIS without the necessity of specific skills and intimate knowledge of the data used. The application consists of the following parts:

- A relational database
- The class of objects for data manipulation
- The class of objects for GIS manipulation
- The main body of the application software containing the above items along with the functions required by the end user.

To begin with, a step-by-step analysis of the software design is required. However, for the purpose of this research it is assumed that the pilot area is already known. Then, the appropriate design steps are as follows:

1. On the CLC's geographic layer of the area of interest we add the remaining geographic characteristics (contour lines, roads, cities, lakes, rivers etc.). This will help in the understanding the exact location of the CLC data.
2. From the FSS database we select only the themes, which associated with agriculture products. The data selected is at prefecture

level, in thousands of hectares of agriculture products, reported in 2000 census.

3. The data provided by the FSS and the CLC databases is studied in order to develop the entity relationship model, and then the database system of the application.

4. CLC data is stored in some database tables of the application, using especially developed programs, while NC's data were stored manually. NC provides also the appropriate DLLs in order to develop programs for automated data transfer.

5. We pointed out the appropriate functions and queries, and we developed object classes to satisfy the requirements for uniformity at both, user and developer levels.

6. We developed an application in which are used the RDBMS, the GIS and the pre-mentioned object classes. The basic capabilities offered by this application are the following:

   - Ability to compose (aggregate) a new FSS theme by selecting one or more CLC classes, and vice versa.
   - Ability to decompose (disaggregate) an existing FSS theme to one or more CLC classes, and vice versa.
   - Ability to correspond (relate) the new FSS themes to CLC classes.
   - Ability to classify (sort) the results either by date, or by county (region), or by CLC class.

Ability to observe the results plotted on the map and to classify these by geographic characteristics, such as allocation of the selected growth by elevation.

5. Results

Table 2 presents the linkage between the 2000 FSS and the new CLC nomenclatures. Although the new CLC nomenclature has been harmonized with the FSS nomenclature there are still some problems related to the two different methodologies. The analysis of the above problems has been carried out throughout of a comparison between the respective areas of the related classes, and has been allowed to make proposals for a future work. The available data from the 2000 FSS has been based at Municipality/Commune level (NUTS V or LAU-1 level), whereas the data has been drawn from the new CLC at the district level (NUTS III). The data of two databases has been compared in a pilot study of four regions of Greece at a district level (NUTS III). The comparison shows large deviations in the agricultural areas. Generally, the examined agricultural areas in new CLC are greater than the corresponding agricultural areas in the 2000 FSS. The problem of large deviations is caused mainly because of the difficulties in correlating the pastures areas between the two databases, whereas the differences of the arable areas and the areas under permanent crops are related to the different methodologies.

The results found so far are presented in Tables 3 to 5. Table 3 presents the differences (%) in arable areas, areas under permanent crops, and cultivated areas (aggregation of D+E), as they recorded in the districts (NUTS III) of the examined regions, between the two nomenclatures. Positive sign is in favor of the new CLC nomenclature, whereas negative sign is in favor of the FSS nomenclature. Note that the actual differences in the above classes are not as high as they are in the remaining classes, namely pastures and meadows (Table 4), heterogeneous areas and agricultural areas (Table 5). To facilitate the comparison for the last cases the actual values are presented.

As it may be observed (Table 3) the above differences (%) in the regions (NUTS II) are generally smaller from the corresponding inter-regional ones (district level; NUTS III). This is due to the fact that the mapping unit of 25 ha in the new CLC is not able to identify parcels of smaller size. This is the case of Greece, in which the average holding size is around 4.5 ha and the average parcel size is around 0.7 ha. An additional reason is that in FSS all the holdings are recorded at the place of residence of the holder (natural person) or headquarter (legal person) of the holding. In the following some preliminary comparison of these results are summarized:

- *Arable areas*

Region comparison shows that the difference for the region of Kriti is about 66% in favor of the FSS nomenclature. However, the differences in the regions of Makedonia are not as high (at most 33%) and are in favor of the new CLC nomenclature (Table 3). Generally, the differences in the arable crop areas are moderate and are in favor either of the FSS nomenclature or of the new CLC nomenclature (NUTS III level). Interesting to note that in some

districts of the regions of Kentriki and Dytiki Makedonia the results are almost the same.

- *Areas under permanent crops*

In general terms the situation is opposite of the one described in the arable crop areas. As it may be observed from Table 3 in the region of Kriti the differences (%) between the two nomenclatures are very small (about 6%). In the regions of Makedonia these differences (%) are moderate (at most 61%) and are in favor of the FSS nomenclature (NUTS II). Furthermore, in the districts of some regions these differences are substantial and/or in the opposite direction (e.g. Evros, Rodopi).

- *Pastures and meadows*

The total areas of pastures and meadows are generally larger in the new CLC than the corresponding areas recorted by the FSS. In all regions (NUTS II) the differences are very high. In the region of Kriti, the two districts of Rethimno and Chania the recorded areas in the new CLC are smaller than the corresponding areas of FSS. This is because only the private areas are recorded in the FSS, whereas all pastures (such as state-owned pastures, private pastures, etc.) are recorded in the new CLC.

- *Heterogeneous areas*

FSS and new CLC present very high actual differences in the class of the heterogeneous areas. Even the two nomenclatures are harmonized there is still a methodological problem of how to relate the two nomenclatures. In particular, in the FSS the survey unit is the agricultural holding, which comprises of at least 0,1 ha. Therefore the heterogeneity (combined crops) of these areas is referred to this small area. In new CLC the heterogeneity is examined within the mapping unit of 25 ha. Under these circumstances a polygon in the new CLC that includes different parcels of a single crop is recorded as heterogeneous area, whereas in the FSS the corresponding parcels are recorded as single crops.

- *Agricultural areas*

All the Agricultural Areas (AA=D+G+F) resulting from the new CLC nomenclature show larger values than the corresponding areas in the FSS nomenclature, particularly in the districts. The differences are generally high with exception of two districts of the region of Kriti. As it has been pointed out previously, the large deviations observed between the agricultural areas as they recorded in the new CLC and the FSS are due to the large deviations in the pastures.

- *Cultivated areas*

Given the problems of the large deviations in the total agricultural areas that are caused mainly from the pastures, the aggregation of the arable areas and the areas under permanent crops into the new class of "Cultivated areas" shows that the differences presented in this class are not significant.

6. Discussion

This study has been based on the provisional data of the 2000 FSS and the new CLC databases and it may be considered as a first step in the direction of present georeference statistical data. The difficulties appeared in the linkage of the two databases can be generally explained from the following points.

The different methodology used as far as the data collection methods and the coverage are concerned. In particular, the FSS is a census using as a reference unit the farm, whereas the new CLC is based on photo—intepretation of the whole area of the country using as a reference unit the mapping unit of at least 25 ha. In addition, CLC has so far been focused on land cover, rather than land use. The minimum size of 25 ha of CLC mapping units presents the difficulty of identifying parcels of smaller size. Thus, a number of non-agricultural areas are classified as agricultural whereas they are only partially agriculture. This is a common problem in areas with forest and olive-trees. Besides, areas classified as non-agricultural areas in CLC may include part of an agricultural area. This explains a number of differences within the agricultural classes. For example, part of meadows or permanent crops can be included in areas with arable crops and conversely.

Despite the harmonization between the new CLC and FSS nomenclatures there are still problems as far as pastures and heterogeneous areas are concerned. In the new CLC, the non-agricultural classes defined by the codes 11, 12, and 13 ("Transitional woodland/shrub", "areas with mixed shrub/grass vegetation" and "areas with little or no vegetation" respectively) may include surfaces classified as "permanent meadows and pastures" in the FSS. Furthermore the FSS does not record the state-owned meadows, which in the new CLC are recorded under the code 8 ("areas under meadow or

pastures"). The special features of Greek agriculture that is marked by the diversity of the holdings in terms of area of production (mixed holdings), the small size of the holdings (average size 4,5 ha), the fragmentation of their area (6 parcels approximately per holding and average parcel size of 0,7 ha). In quite a number of cases the parcels of the same holding are normally located far away from the farmhouse or from the headquarter, but they are recorded at the place of the farmhouse or the headquarter (by definition).

7. Conclusion

The work presented so far is a pilot study merging, by means of a software tool, the statistical data, available at the administrative level, with the geo-referenced land cover in order to identify and explain the most significant differences encountered between the aggregates of agricultural land cover classes. This has been achieved with the use of the 2000 FSS and the new CLC databases already under development in Greece. The new CLC seems to provide a good mapping base for Greece, which could be improved further by using suitable satellite images that are able to produce scaled maps of at least 1:50000. Note that the imposed minimum mapping unit of 25 ha results in an overall underestimation of the diversity of landscapes something, which is particularly important in the case of Greece for which the average size of the holdings is 4,5 ha. Apart of CLC, additional sources may be used providing detailed complementary information, such as aerial ortho-photographs, the cadastral map of Greece, IACS (Integrated Administrative Control System), MARS (Monitor Agriculture with Remote Sensing), NATURA2000 database, or other ongoing analysis of the European landscape.

When the final data from the remaining regions of Greece will be available a quality analysis of the two databases will be carried out and a finer level of nomenclature will be examined. This will allow final conclusions to be drawn and further actions to be taken in the future. Future research is to continue improving the idea of interoperable geo-object by adding methods and properties for uncertainty manipulation and to investigate requirements of GIS in a fuzzy object data model. Our final objective is to embody in the Geo-Object, the ability to generate and visualize transitions from one state to another, using the rules of an expert spatiotemporal system.

References

[1] A. Frank (19-21 June 1999), Tracing socioeconomic pattern of urban development, Proceedings of Geoinformatics'99 conference, Ann Arbor, Michigan.

[2] Burrough P.A (1986), Principles of Geographic Information Systems, 1st ed, pp.336, Oxford University Press, New York.

[3] Deichmann U. (6-10 December 1997), Geographical information systems in the census process —Technology options, costs and benefits, Workshop on Strategies for the 2000 Round of Population and Housing Censuses in the ESCWA Region Cairo.

[4] Dragicevic V. and Marceau D. (2000), An application of fuzzy logic reasoning for GIS temporal modeling of dynamic processes, Fuzzy sets and Systems 113, Elsevier, pp.69-80.

[5] Dueker, K.J. (1979), Land Resource Information Systems: A Review of Fifteen Years Experience, Geo-Processing, 1, pp.105-128.

[6] Ehlers, M., Greenle, D., Smith, T. and Star, J. (1991), Integration of Remote Sensing and GIS: Data and Data Access, Photogrammetric Engineering & Remote Sensing 57(6), pp.669-675.

[7] Flowerdew, Robin and Mick Green (1994), Areal Interpolation and Types of Data. In Spatial Analysis and GIS edited by Stewart Fortheringham and Peter Rogerson. London, Bristol: Taylor & Francis Ltd.

[8] Goodchild, M. and Gopal S. (eds), (1989), Accuracy of Spatial Databases, London: Taylor and Francis.

[9] Ian N Gregory (23 - 25 August 2000), An evaluation of the accuracy of the areal interpolation of data for the analysis of long-term change in England and Wales, Proceedings of the 5th International Conference on GeoComputation University of Greenwich, United Kingdom.

[10] Lillesand, T.M. and Kiefer, R.W. (1994), Remote Sensing and Image Interpretation (3rd ed.), New York: Wiley.

[11] Maffini and Guilio (1987), Raster versus Vector Data Encoding and Handling: A Commentary. Photogrammetric Engineering & Remote Sensing 53 (10), pp.1397- 1398.

[12] Openshaw, S. (1984), The Modifiable Areal Unit Problem, Norwich: Geo Books.

[13] Ozemoy, V.M., Smith, D.R., and Sicherman, A. (1981), Evaluating Computerized Geographic Information Systems Using Decision Analysis, Interfaces, 11, pp.92-98.

[14] Parent C., Spaccapietra S., Zimanyi E., ' Conseptual Modeling for Federated GIS over the Web'. Ecole Polytechnique Federale de Lausanne — Laboratoire de Bases de Donees, 1.

[15] Sambrakos M., Fillis I., Tsiligiridis T. (2001), Integration of Spatial Descriptive Statistical Data and Geographic Information, Proceedings of 8th Panhellenic Conference on Informatics, paper 91C, Nicosia, Cyprus.

[16] Xie and Yichun (1995), The Overlaid Network Algorithms for Areal Interpolation Problem, Computers, Environment and Urban Systems, 19 (4), pp.287-306.

[17] Yuan, Y., Smith, R.M. and Limp, W.F. (1997), Remodeling Census Population with Spatial Information from Landsat TM Imagery, Computers, Environment and Urban Systems, 21(3/4), pp.245-258.

Table 1. Classification of land use in the 2000 FSS nomenclature.

		DO1	Common wheat and spelt
D:	D01-D08: CEREALS	D02	Durum wheat
ARABLE LAND		D03	Rye
		D04	Barley
		D05	Oats
		DO6	Grain maize
		D07	Rice
		D08	Other cereals
		DO9C	Pulses-fodder peas
	D09: DRIED PULSES	DO9D	Pulses-fodder field beans
		DO9E	Pulses-other than fodder peas and field beans
		D10	Potatoes
	D10-D12: ROOT CROPS	D11	Sugar beets
		D12	Fodder roots and brassicas
		D13A	Tobacco
	D13: INDUSTRIAL PLANTS	D13B	Hops
		D13C	Cotton
		D13D	Other industrial plants
		D13D1	Other oil seeds or fibre plants
		D13D1A	Rape and turniprape
		D13D1B	Sunflower
		D13D1C	Soya
		D13D1D	Other oil seeds or fibre plants-others
		D13D2	Aromatic-medicinal and culinary plants
		D13D3	Industrial plants-others
		D14A	Fresh vegetables, mellons, strawberries-outdoor-openfield
	D14-D15:		
	FRESH VEGETABLES, MELLONS, STRAWBERRIES	D14B	Fresh vegetables, mellons, strawberries-outdoor- market gardening
		D15	Fresh vegetables, mellons, strawberries under glass
	D16-D17:	D16	Flowers and ornamental plants outdoor
	FLOWER AND ORNAMENTAL PLANTS	D17	Flowers and ornamental plants under glass
		D18A	Forage plants-temporary grass
	D18: FORAGE PLANTS	D18B	Forage-plants-other green fodder-total
		D18B1	Forage-plants-other green fodder-green maize
		D18B2	Forage-plants-other green fodder-leguminous plants
		D18B3	Forage-plants-other green fodder-others
	D19-D20:	D19	Seeds and seedlings
	OTHER ARABLE CROPS	D20	Other crops
	D21: FALLOW LAND	D21	Fallow land
E: KITCHEN	E: KITCHEN GARDENS	E	Kitchen gardens
GARDENS			
F:	F:	F01	Permanent grassland and meadow-pasture and meadow
PERMANENT	PERMANENT PASTURES AND		
PASTURES AND	MEADOWS	F02	Permanent grassland and meadow-rough grazing
MEADOWS			
G:	G1: FRUIT AND BERRY PLANTATIONS	GO1A	Fruit and berry plantations-temperate climate
PERMA		GO1B	Fruit and berry plantations-subtropical climate
NENT		GO1C	Fruit and berry plantations-nuts
CROPS	G2: CITRUS PLANTATIONS	G02	Citrus plantations
	G3: OLIVE PLANTATIONS	GO3A	Olive plantations-table olives
		GO3B	Olive plantations-oil production
	G4: VINEYARDS	GO4A	Vineyards-quality wine
		GO4B	Vineyards-other wines
		GO4C	Vineyards-table grapes
		GO4D	Vineyards-raisins
	G5: NURSERIES	G05	Nurseries
	G6: OTHER PERMANENT CROPS	GO6	Other permanent crops
	G7: PERMANENT CROPS UNDER GLASS	G07	Permanent crops under glass

H: OTHER LAND	H0103: NON-UTILIZED AGRICULTURAL LAND	H01	Unutilized agricultural land which is no longer farmed, for economic, social or other reasons
		H03	Other land occupied by buildings, pleasure gardens, etc.
	H02: WOODED AREA	H02	Woodland
1: COMBINED AND SUCCESSIVE SECONDARY CROPPING, MUSHROOMS,	101: SUCCESSIVE SECONDARY CROPS	101A	Successive secondary crops-non fodder cereals
		101B	Successive secondary crops-non fodder pulses
		101C	Successive secondary crops-non fodder oil-seed plants
		101D	Successive secondary crops-others total
	102: MUSHROOMS	102	Mushrooms
	103: IRRIGATED AREA	103A	Total irrigable area
		103B	Irrigated once a year-total
	104: AREA COVERED BY GREENHOUSES IN USE	104	Area covered by greenhouses in use
	105: COMBINED CROPS	105A	Combined crops-agricultural-forestry
		105B	Combined crops-permanent-annual
		105C	Combined crops-permanent-permanent
		105D	Combined crops-others

Table 2. Linkage between the 2000 FSS and the new CLC nomenclatures in Greece

New CLC		FSS	
LEVEL 1	**LEVEL 2**	**LEVEL 1**	**LEVEL 2**
1. Artificial surfaces (Man-made areas)	1.1 Urban fabric (Build-up areas, urban agglomerations)		-
	1.2 Industrial and commercial units (Industrial or commercial zones)		-
	1.3 Transport units		-
	1.4 Mine, dump and construction sites (Mines, waste disposal sites)		-
	1.5 Artificial, non-agricultural vegetated areas sport and cultural activity sites (Artificial or non-agricultural green areas)		-
2. Agricultural areas	2.1 Arable land (Areas under arable crops)	Utilized agricultural areas D+G+E	D=D01+D02+D03+ D04+D05+ D06+D07+D08+D09+D10+ Dll+D12+D13+D14+D15+ D16+D17+D18+D19+
	2.2 Permanent crops (Areas under permanent crops)		G=G01+G02+G03+G04+G05 +G06+G07
	2.3 Pastures (Areas under meadow or pasture)		F=F01+ F02
	2.4 Heterogeneous agricultural areas (Areas with mixed uses -mixed farmland)		105A+I05B
3. Forests and semi-natural areas	3.1 Forests (Forested areas)		H02: only the private forests
	3.2 Transitional woodland /shrub		H01: only the private uncultivated areas for economic, social or other reasons
	3.3 Shrub and/or herbaceous vegetation associations (Areas with mixed shrub/grassy vegetation)		
	3.4 Open spaces with little or no vegetation (Areas with little or no vegetation)		
4. Surfaces under water	4.1 Inland water		
	4.2 Inland wetlands		
	4.3 Coastal wetlands		

Table 3. Results showing the differences (%) in arable areas, areas under permanent crops and cultivated areas (D+E) as they recorded by the 2000 FSS and the new CLC nomenclatures.

Regions (NUTS II)	Districts (NUTS III)	Arable Areas (% difference)	Areas under Permanent Crops (% difference)	Cultivated Areas (% difference)
ANATOLIKI MAKEDONIA & THRAM	DRAMA	45	-93	42
	KAVALA	64	-45	31
	EVROS	24	44	25
	XANTHI	33	-67	32
	RODOPI	31	89	32
TOTAL		33	-27	30
KENTRIKI MAKEDONIA	IMATHIA	42	-91	-12
	SALONIKI	4	-49	3
	KILKIS	-7	-39	-7
	PELLA	-31	-77	-47
	PIERIA	-7	-79	-14
	SERRES	42	-81	37
	CHALKIDIKI	54	-9	34
TOTAL		15	-61	4
DYTIKI MAKEDONIA	GREVENA	20	-68	18
	KASTORIA	-21	-35	-22
	KOZANI	4	27	5
	FLORINA	-3	-44	-4
TOTAL		3	-14	2
TOTAL MAKEDONIA		18	-52	12
KRITI	IRAKLIO	-71	4	-4
	LASITHI	54	47	48
	RETHIMNO	-91	-7	-24
	CHANIA	-72	4	-4
TOTAL		-66	6	-3

Table 4. Results showing the actual values and the corresponding differences in the class of pastures and meadows as they recorded by the 2000 FSS and the new CLC nomenclatures.

	Districts	Pastures and meadows (ha)		
(NUTS II)	(NUTS III)	2000 FSS	new CLC	Difference
ANATOLIKI	DRAMA	1,294	31,380	30,086
MAKEDONIA	KAVALA	760	19,810	19,050
	EVROS	4,353	13,870	9,517
	XANTHI	81	11,910	11,829
	RODOPI	1,733	13,520	11,787
TOTAL		8,221	90,490	82,269
KENTRIKI	IMATHIA	860	9,840	8,980
MAKEDONIA	SALONIKI	473	25,020	24,547
	KILKIS	5,310	40,680	35,370
	PELLA	2,458	25,910	23,452
	PIERIA	3	6,570	6,567
	SERRES	6,246	28,520	22,274
	CHALKIDIKI	2,780	5,330	2,550
TOTAL		18,130	141,870	**123,740**
DYTIKI	GREBENA	315	25,890	25,575
MAKEDONIA	KASTORIA	822	29,840	29,018
	KOZANI	794	70,610	69,816
	FLORINA	5,477	27,200	21,723
TOTAL		7,408	153,540	146,132
TOTAL MAKEDONIA		33,759	385,900	**352,141**
KRITI	IRAKLIO	36,412	69,070	32,658
	LASITHI	16,817	61,631	44,814
	RETHIMNO	62,470	53,241	-9,229
	CHANIA	63,410	40,167	-23,243
TOTAL		179,109	224,109	**45,000**

Table 5. Results showing the actual values and the difference (%) in the class of agricultural areas as they recorded by the FSS and the new CLC nomenclatures, It also shows the average parcel area.

Regions (NUTS II)	Districts (NUTS III)	Agricultural Areas (ha)		Agricultural Areas (% difference)	Average parcel area (ha)
		2000 FSS	New CLC		
ANATOLIKI MAKEDONIA	DRAMA	47,193	104,720	122	0.78
	KAVALA	44,860	92,390	106	0.55
	EVROS	150,252	231,060	54	0.64
	XANTHI	37,214	69,940	88	0.69
	RODOPI	74,941	121,230	62	0.62
TOTAL		354,460	619,340	75	0.64
KENTRIKI MAKEDONIA	IMATHIA	53,894	95,690	78	0.82
	THESSALONIKI	129,483	222,840	72	0.79
	KILKIS	106,027	172,420	63	0.90
	PELLA	77,660	151,640	95	0.61
	PIERIA	45,543	74,950	65	0.75
	SERRES	144,947	234,670	62	0.58
	CHALKIDIKI	77,274	147,270	91	0.77
TOTAL		634,828	1,099,480	73	0.72
DYTIKI MAKEDONIA	GREVENA	41,432	93,810	126	0.80
	KASTORIA	24,887	74,260	198	0.58
	KOZANI	88,170	166,260	89	0.58
	FLORINA	52,952	90,960	72	0.55
TOTAL		**207,441**	**425,290**	**105**	**0.60**
TOTAL MAKEDONIA		**1,196,729**	**2,144,110**	**79**	**0.67**
KRITI	IRAKLIO	139,733	221,982	59	0.40
	LASITHI	37,864	127,252	236	0.44
	RETHIMNO	101,182	115,842	14	0.87
	CHANIA	109,191	116,472	7	0.83
TOTAL		**387,970**	**581,548**	**50**	**0.57**

Efficiency Analysis by using Data Envelop Analysis Model: Evidence from Indian Banks

Majid Karimzadeh
Islamic Azad University of Saravan
Aligarh, Uttar Pradesh, India,202002
Karimzadeh111@yahoo.com

Abstract **- The structure of Indian banking has substantially changed over the past decades, partially as a result of adoption of new technologies and process of reforms and accompanying deregulation has embodied an incentive for bank management to focus on improving efficiency, especially given the more competitive banking environment. This study aims to examine the efficiency of Indian commercial banks during 2000 – 2010 by utilizing Data Envelopment Analysis (DEA). Based on the sample of 8 commercial banks, our findings reveal that the mean of cost (economic) efficiency, technical efficiency, and allocative efficiency are 0.991, 0.995, and 0.991 in VRS model and 0.936. 0.969, and 0.958 in CRR model, respectively using DEA approach. Inputs and outputs of this study were analyzed based on intermediation approach. In addition, the results suggest that Bank of India and ICICI bank are more efficient as compare to other banks in India and result confirmed that selected Public Sector Banks are more efficient than Private sectors during the study period in India.**

Keyword: Data envelop analysis, economic efficiency, allocative and technical efficiency, Indian banks.

1. Introduction

Investigating the efficiency of the financial system and in particular banks has gained a lot of popularity in recent times for several reasons. First, the efficiency of banks is directly linked to the productivity of the economy. Banking system assets constitute a substantial proportion of total output (Bauer Paul et al, 1992). Banks provide liquidity, payments and safekeeping for depositors` and channel these funds into investment and working capital requirements. In addition, banks are supposed to play a special role in funding small businesses that often have very limited access to other sources of external finance. Banks also play a major role in ensuring a smoothly functioning payment system, which allows financial and real resources to flow freely to their highest-returns uses. A basic benefit of enhanced efficiency is a reduction in spreads between lending and deposit rates. This is likely to stimulate both greater loan demands for industrial investment (and thus contribute to higher economic growth) and greater mobilization of savings through the banking system. Banks in most developing countries operate with relatively wide spreads. Although government policies and regulations are considered major causes of such wide spreads, studies on banking efficiency has pointed at operating inefficiencies as one other possible source that needs to be investigated. Wide spreads affect intermediation and distort prices thus impairing the role of the financial system in contributing to rapid economic growth (Ikhide. S, 2000).

Indian financial services industry is dominated by the banking sector and the banking structure in India is broadly classified into public sector banks, private sector banks and foreign banks. The public sector banks continue to dominate the banking industry, in terms of lending and borrowing, and it has widely spread out branches which help greatly in pooling up of resources as well as in revenue generation for credit creation.

The Indian financial sector reform of 1991 has greatly changed the face of Indian Banking system. In addition to the nationalized banks, several private Banks were newly founded or created by previously extant financial institutions. India has also seen the entry of over two dozen foreign banks since the beginning of financial reforms. In the face of increased competition, the banks have to operate more efficiently in order to sustain and perform better. In the context of increased competition and the importance of banks in financial markets, it becomes very much essential to evaluate whether these banks operate efficiently. Primarily, there are two chief reasons to measure the efficiency of banking institutions. Firstly, this assists to identify the most efficient banks and benchmarks the relative efficiency of individual banks against the most efficient banks. Secondly, it helps to evaluate the

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.006.12/ijltfes.2047.0916

impact of various policy measures on the performance of banks.

The objective of this paper is to estimate technical and total economic efficiency of commercial banks in India for the period 2000-2010. The paper is structured as follows: the first section will discuss review of literature in banking followed by methodology, data and specification of bank inputs and outputs. Empirical findings are discussed in the next section followed by the conclusion.

2. Literature review

During the late 1980s and particularly in the 1990s, the DEA method has been used extensively to evaluate banking institutions. Sathye (2003) used DEA to study the relative efficiency of Indian banks in the late 1990's with that of banks operating in other countries. He found that the public sector banks have a higher mean efficiency score as compared to the private sector banks in India, but found mixed results when comparing public sector banks and foreign commercial banks in India. Kumbhakar and Sarkar (2004) estimated the cost efficiency of public and private sector banks in India by using the stochastic cost frontier model with specification of translog cost function. The study used data of 50 banks for the analysis and necessary information have been collected from the various issues of the annual reports published by the Indian Banks' Association for the period 1986-2000. The empirical results revealed that deregulation not only increased the cost inefficiency but also affected the rate of fall in inefficiency of banks. During this period private banks were more efficient than the public sector banks according to study.

Rammohan and Ray (2004) compared the revenue maximizing efficiency of banks in India in 1990's. Deposits and operating costs were taken as inputs while loans, investments and other income were taken as outputs. Their research found that public sector banks were significantly better than private sector banks on revenue maximization efficiency. However it was found that the difference in efficiency between public sector banks and foreign banks was not significant.

Das et al, (2004) examined the efficiency of Indian banks by using DEA model. Four input measures: deposits and other borrowings, number of employees, fixed assets and equity, and three output measures: investments, performing loan assets and other non-interest fee based incomes were used in the analysis. He found that Indian banks did not exhibit much of a difference in terms of input or output oriented technical and cost efficiency. However, in terms of revenue and profit efficiencies prominent differences were seen. He also found that size of the bank, ownership of the bank, and listing on the stock exchange had a positive impact on the average profit and revenue efficiency scores.

Soori et al, (2005) analyzed efficiency of Iranian banking system and the main Purpose of the study was to investigate the comparative efficiency of commercial banks in Iran using stochastic frontier function as a parametric and data envelopment analysis as a non-parametric approaches. The data used cover the period 1996-2004. The findings of this paper show that there is a significant difference between non–parametric and parametric methods in measuring the efficiency in the commercial banks of Iran. Debasish (2006) also attempted to measure the relative performance of Indian banks, using the output-oriented CRR DEA model. The analysis used nine variables and seven output variables in order to examine the relative efficiency of commercial banks over the period 1997 – 2004.

Mostafa, M. (2007) investigated the efficiency of top 85 Arab banks using DEA and Neural networks for the year 2005. He found that, eight banks as per the CCR Score and four banks as per BCC Score were positioned on the efficient frontier. He suggested that future studies should test the existence of positive rank-order correlations between efficiency scores obtained from DEA analysis and traditional efficiency measures such as financial ratios. His results further demonstrate that, Al-Rajhi Bank and National Commercial Bank were placed among the top ten Arab banks with a relative ranking of eight and ten respectively.

Moh'd Al-Jarrah (2007) is used data Envelopment Analysis (DEA) approach to investigate cost efficiency levels of banks operating in Jordan, Egypt, Saudi Arabia and Bahrain over 1992-2000. The estimated cost efficiency is further decomposed into technical and allocative efficiency at both variable and constant return to scale. Later on, the technical efficiency is further decomposed into pure technical and scale efficiency. Cost efficiency scores ranged from 50 to 70% with some variations in scores depending on bank's size and its geographical locations. The results suggested that the same level of output could be produced with approximately 50-

70% of their current inputs if banks under study were operating on the most efficient frontier.

Chansarn (2008) conduct a study aimed to examine the relative efficiency of Thai commercial banks during 2003 – 2006 by utilizing Data Envelopment Analysis (DEA). Based on the sample of 13 commercial banks, findings revealed that the efficiency of Thai commercial banks via operation approach is very high and stable while the efficiency via intermediation approach is moderately high and somewhat volatile. In term of size, large, medium and small banks, in average, were efficient via operation approach with the average efficiencies of 100%. However, small banks were the most efficient banks via intermediation approach

AlKhathlan and Abdul Malik (2008) used basic DEA models i.e. CCR and BCR to evaluate the relative efficiency of Saudi Banks using annual data from 2003 through 2008. The results showed that, on a relative scale, Saudi banks were efficient in the management of their financial resources. In addition, the results would provide crucial information about Saudi banks' financial conditions and management performance for the benefit of bank regulators, managers and bank stock investors.

Kumar and Gulati (2008) conducted a study aimed to measure the extent of technical, pure technical, and scale efficiencies in 27 public sector banks (PSBs) operating in India in the year 2004/05. The empirical findings of study revealed that PSBs (Public sector banks) operate at 88.5 percent level of overall technical efficiency i.e., inputs could be reduced by 11.5 percent without sacrificing output if all banks were efficient as 7 benchmark banks identified by DEA. Further, the contribution of scale inefficiency in overall technical inefficiency has been observed to be smaller than what been observed due to managerial inefficiency (i.e., pure technical inefficiency). The findings pertaining to returns-to-scale in Indian public sector banking industry highlight that the predominant form of scale inefficiency is decreasing returns-to-scale. The results of logistic regression analysis also provide that the exposure of the banks to off-balance sheet activities (i.e., non-traditional activities) has a strong and positive impact on the overall technical efficiency of banks in India.

San O et al, (2011) in their study utilizes non parametric Data Envelopment Analysis (DEA) to analyze and compare the efficiency of foreign and domestic banks in Malaysia. The analysis was based on a panel data set of 9 domestic banks and 12 foreign banks in Malaysia over the period of 2002-2009. Intermediation approach is used to define the inputs and outputs in computerizing the efficiency scores. Surprisingly, the findings are inconsistent with most of the findings of previous studies where the foreign banks were outperforming their domestic peers in term of efficiency. Conversely, the finding of this study shows that domestic banks have a higher efficiency level than foreign banks, this imply that domestic banks are relatively more managerially efficient in controlling their costs. The second stage of the empirical results was based on the Tobit model, which suggests that the pure technical efficiency (PTE) of banks in Malaysia is mainly affected by capital strength, loan quality, expenses and asset size.

3. Methodology

The literature distinguishes two main approaches in measuring banking efficiency; a parametric and a non-parametric approach in which the specification of a production cost function is required in both approaches.

The parametric approach engages in the specification and econometric estimation of a statistical or parametric function, while the non-parametric method offers a linear boundary by enveloping the experimental data points, known as "Data Envelopment Analysis" (DEA). This study uses non-parametric approach-Data Envelopment Analysis (DEA) to estimate technical and economic efficiency of Indian commercial banks. The main objective of DEA is to determine which firms are operating on their efficient frontier and which firms are not. If the firm's input-output combination lies on the DEA frontier, the firm is considered efficient; and the firm is considered inefficient if the firm's input-output combination lies inside the frontier. The present study uses the latest available published data for the year 2000 compiled by 2010.

3.1 Data envelop analysis

Data Envelopment Analysis (DEA) developed by Charnes et al. (1978) is a linear programming based technique. DEA occasionally called frontier analysis is a performance measurement technique which can be used for analyzing the relative efficiency of productive units, having the same multiple inputs and multiple outputs. It is a non-parametric analytic technique which allows us to

compare the relative efficiency of units as benchmark and by measuring the inefficiencies in input combinations in other units relative to the benchmark. One of the earliest studies on DEA is the study of Farrell (1957) who attempted to measure the technical efficiency of production in single input and single output case. DEA was originally developed by Charnes, Cooper and Rhodes (1978) with the assumption of constant return to scale (CRS) in attempt to propose a model that generalizes the single-input, single output measure of a DMU to a multiple inputs, multiple outputs setting. Thus DMU is an entity that uses input to produce output. DEA was extended by Banker, Charnes and Cooper (1984) to include variable return to scale (VRS). Up to now the DEA measure has been used to evaluate and compare educational departments, health care, agricultural production, banking, armed forces, sports, market research, transportation and many other applications.

DEA is a deterministic methodology for examining the relative efficiency, based on the data of selected inputs and outputs of a number of entities called decision–making units (DMUs). From the set of available data, DEA identifies relative efficient DMUs (which are used as reference points) which define the efficiency frontier and evaluate the inefficient of other DMUs which lie below that frontier.

DEA is an alternative analytic technique to regression analysis. Regression analysis approach is characterized as a central tendency approach and it evaluates DMUs relative to an average. In contrast, DEA is an extreme point method and compares each DMU with the only best DMU. The main advantage of DEA is that, unlike regression analysis, it does not require an assumption of a functional form relating inputs to outputs. Instead, it constructs the best production function solely on the basis of observed data; hence statistical tests for significance of the parameters are not necessary (Chansarn, 2008).

Return to scale

Return to scale refers to increasing or decreasing efficiency based on size. For example, a manufacturer can achieve certain economies of scale by producing thousand Integrated Circuits at a time rather than one at a time. It might be only 100 times as hard as producing one at a time. This is an example of increasing returns to scale (IRS). On the other hand, the manufacturer might find it more than trillion times difficult to produce a trillion Integrated Circuits at a time because of storage problems and limitations on the worldwide Silicon supply. This range of production illustrates Decreasing Returns to Scale (DRS). Combining the extreme two ranges would necessitate Variable Returns to Scale (VRS). Constant Return to Scale (CRS) means that the producers are able to linearly scale the inputs and outputs without increasing or decreasing efficiency. This is a significant assumption. The assumption of CRS may be valid over limited ranges but its use must be justified. But, CRS efficiency scores will never be higher than that of VRS efficiency scores. In a CRS model, the input-oriented efficiency score is exactly equal to the inverse of the output-oriented efficiency score. This is not necessarily true for inefficient DMUs in the case of other return to scale assumptions. The CRS version is more restrictive than the VRS and yields usually a fewer number of efficient units and also lower efficient score among all DMUs. In DEA literature the CRS model is typically referred to as the CCR model after the originators of the seminal publication, by Charnes, Cooper and Rhodes (1978).

CCR's model: The model has developed the Farrell's efficiency measurement concept from several inputs and one output to several inputs and several outputs. In this model (Charnes et al (1978)) using a linear combination, different inputs and outputs are changed into one virtual input and output which the ratio of these virtual combinations of outputs to inputs will be the estimation of efficiency boundary for the measurement of relative efficiency given that the yield is constant.

BCC's model: In contrast to constant yield in the above mentioned model, the BCC's model (Banker et al (1984)) assumes a variable output with respect to the scale. In the model, the technical efficiency is decomposed to pure technical efficiency and scaled efficiency in order to measure the output to scale as well as efficiency itself.

Mathematically, relative efficiency of a DMU is defined as the ratio of weighted sum of outputs to weighted sum of inputs. This can be written as:

$$\boldsymbol{ho} = \frac{\int_{r=1}^{s} Ur\ Yro}{\int_{i=1}^{m} Vi\ Xio} \quad (1)$$

Where:
S= number of outputs:
Ur= weight of output r:
Yro= amount of r produced by the DMU:
M=number of inputs:

Vi= weight of input I :and,
Xio= amount of input I used by the DMU:

Equation 1 assumes CRS and controllable inputs. While outputs and inputs can be measured and entered in this equation without standardization, determining a common set of weights can be difficult (Avkiran, 1999). DMUs might assess their outputs and inputs in a different way. This issue is answered in the Charnes, Cooper and Rhodes (known as CCR) model. Charnes et al. (1978) developed the CCR model that had an input orientation and assumed

CRS. The result of CCR model indicates a score for overall technical efficiency (OTE) of each DMU. In other words, this model calculates the technical efficiency and scale efficiency combined for each DMU. The CCR model addressed the above problem by allowing a DMU to take up a set of weights that maximize its relative efficiency ratio without the same ratio for other DMUs exceeding one. Thus equation 1 is rewritten in the form of a fractional programming problem:

$$\max ho = \frac{\int_{r=1}^{s} Ur\ Yro}{\int_{i=1}^{m} Vi\ Xio} \quad (2)$$

Subject to:

$$\frac{\int_{r=1}^{s} Ur\ Yro}{\int_{i=1}^{m} Vi\ Xio} \leq \text{For each DMU in the sample}$$

Where j=1,……, n (number of DMUs)

To measure efficiency, equation 2 is converted into a linear programming problem. In equation 3, the denominator is a set of constant and the numerator is maximized:

$$\max ho = \int_{r=1}^{s} Ur\ Yro \quad (3)$$

$$\int_{i=1}^{m} Vi\ Xio = 1$$

$$\int_{r=1}^{s} Ur\ Yrj - \int_{i=1}^{m} Vi\ Xij \leq 0,$$

$$Ur, Vi \geq \in,$$

Therefore, in order to avoid the exclusion of an output or an input in the calculation of efficiency, weights *u* and *v* are not permitted to fall below non-Archimedean small positive numbers (∈). Equation 3 utilizes controllable inputs and CRS. It is a linear programming problem that models input minimization.

Then, Banker et al. (1984) introduced the usage of VRS that splits OTE into two components, namely pure technical efficiency (PTE) and scale efficiency (SE). This is popularly referred as Banker, Charnes and Cooper (known as BCC) model. The BCC linear programming problem that calculates pure technical efficiency is depicted in equation 4:

$$\max ho = \int_{r=1}^{s} Ur\ Yro + Co \quad (4)$$

$$\int_{i=1}^{m} Vi\ Xio = 1$$

$$\int_{r=1}^{s} Ur\ Yrj - \int_{i=1}^{m} Vi\ Xrj - Co < 0,$$

$$Ur, Vi \geq \in,$$

On the whole, the former concerns about the capability of managers to use the firms' given resources, while the latter refers to utilizing scale economies by working at a point where the production frontier shows CRS.

To discuss DEA in more detail it is necessary to look at the different concepts of efficiency. The most common efficiency concept is technical efficiency: the conversion of physical inputs (such as the services of employees and machines) into outputs relative to best practice. In other words, given current technology, there is no wastage of inputs whatsoever in producing the given quantity of output. An organization operating at best practice is said to be 100% technically efficient. If operating below best practice levels, then the organization's technical efficiency is expressed as a percentage of best practice. Managerial practices and the scale or size of operations affect tech Allocative efficiency refers to whether inputs, for a given level of output and set of input prices, are chosen to minimize the cost of production, assuming that the organization being examined is already fully technically efficient. Allocative efficiency is also expressed as a percentage score, with a score of 100% indicating that the organization is using its inputs in the proportions that would minimize costs. An organization that is operating at best practice in engineering terms could still be allocatively inefficient because it is not using inputs in the proportions which minimize its costs, given relative input prices. Finally, cost efficiency (total economic efficiency) refers to the combination of technical and allocative efficiency. An organization will only be cost efficient if it is both technically and allocatively

efficient. Cost efficiency is calculated as the product of the technical and allocative efficiency scores (expressed as a percentage), so an organization can only achieve a 100% score in cost efficiency if it has achieved 100% in both technical and allocative efficiency.

3.2 The Data and model specification

This study includes 8 major commercial banks of India, State Bank of India (SBI), Bank of India (BOI), and Central Bank of India (CBI), Panjab National Bank (PNB), and Union Bank of India (UBI) as public bank and, ICICI Bank, HDFC Bank, and Axis Bank as private bank. The annual balance sheet and income statement used were taken from different reports of Reserve Bank of India.

In the literature in the field, there is no consensus regarding the inputs and outputs that have to be used in the analysis of the efficiency of the activity of commercial banks (Berger and Humphrey, 1997). In the studies in the field, five approaches for defining inputs and outputs in the analysis of the efficiency of a bank were developed, namely: the intermediation approach; the production approach; the asset approach; the user cost; the value added approach. The first three approaches are developed according to the functions banks fulfill (Favero and Papi, 1995). The production and the intermediation approaches are the best known ones and the most used in the quantification of bank efficiency (Sealy and Lindley, 1997).

In the production-type approach, banks are considered as deposit and loan producers and it is assumed that banks use inputs such as capital and labor to produce a number of deposits and loans. According to the intermediation approach, banks are considered the intermediaries that transfer the financial resources from surplus agents to the fund deficit ones. In this approach it is considered that the bank uses as inputs: deposits, other funds, equity and work, which they transform into outputs such as: loans and financial investments. The opportunity for using each method varies depending on circumstances (Tortosa- Ausina, 2002). The intermediation approach is considered relevant for the banking sector, where the largest share of activity consists of transforming the attracted funds into loans or financial investments (Andrie and Cocris, 2010).

In the analysis we will use the following set of inputs and outputs to quantify the efficiency of banks in India:

- Outputs: loans and investments
- Inputs: fixed assets, deposits, and number of employees.

Before explaining the empirical DEA models for estimating cost and profit efficiency, we discuss the data and selection of inputs and outputs in the subsequent section.

This study uses the intermediation approach to define bank inputs and outputs. Under the intermediation approach, banks are treated as financial intermediaries that combine deposits, labour and capital to produce loans and investments. The values of loans and investments are treated as output measures; labour, deposits and number of employees is inputs. Price information is necessary for analyzing cost efficiency therefore in this section we will explain prices of inputs and calculation of them:

For the price of employees we used Employee expense per capita (P1) which means employees expense divided by number of employees. For the price of deposits we used Average of interest paid by the banks (P2) that could be calculated as an interest expense over total value of deposits and for price of fixed assets (P3) we used depreciation costs on fixed assets.

Table 1. Total economic efficiency (CRS model)

Bank	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	Average
SBI	0.819	0.882	0.980	1	0.999	0.896	0.851	0.933	0.968	1	1	0.938
BOI	0.853	0.933	0.972	1	0.973	0.952	0.872	0.955	0.989	0.989	1	0.953
CBI	0.815	0.896	0.919	1	0.947	0.910	0.834	0.915	0.920	0.924	1	0.916
UBI	0.822	0.851	0.952	1	0.936	0.956	0.957	0.969	0.954	1	1	0.945
PNB	0.793	0.819	0.885	0.928	1	0.817	0.851	0.996	0.918	0.965	1	0.906
ICICI	0.875	0.927	1	0.993	0.979	0.947	0.945	1	0.984	0.966	1	0.965
HDFC	0.935	1	0.920	1	1	0.920	0.921	0.865	0.890	0.947	1	0.945
Axis	0.883	0.981	0.858	0.730	0.899	1	0.990	0.983	1	0.895	0.937	0.923
Average	0.849	0.911	0.935	0.956	0.966	0.924	0.902	0.952	0.952	0.960	0.992	0.936

4. Empirical result

The summary result for the analysis via intermediation approach is presented in Tables

Table 2. Total Economic Efficiency (VRS Model)

Bank	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	Average
SBI	1	0.982	1	1	1	1	1	1	0.970	1	1	0.995
BOI	1	1	1	1	1	1	0.995	1	1	1	1	0.999
CBI	1	0.978	1	1	0.999	1	0.972	0.949	0.935	0.940	1	0.979
UBI	1	0.990	1	1	0.981	1	1	0.987	0.968	1	1	0.993
PNB	1	0.986	1	0.990	1	0.953	1	0.996	1	0.990	1	0.992
ICICI	1	1	1	1	1	1	0.902	1	1	0.988	1	0.998
HDFC	1	1	0.953	1	1	0.989	0.945	1	0.997	0.950	1	0.985
AXIS	1	1	0.897	0.796	0.921	1	0.992	0.994	1	0.929	1	0.957
Average	1	0.992	0.985	0.993	0.987	0.995	0.989	0.995	0.988	0.982	1	0.991

Table 3. Technical Efficiency (CRSS)

Bank	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	average
SBI	0.897	0.921	1	1	1	0.987	0.976	0.982	1	1	1	0.978
BOI	0.934	0.974	1	1	1	1	1	1	1	1	1	0.992
CBI	0.921	0.900	0.974	1	1	1	1	0.877	0.925	0.953	1	0.959
UBI	0.883	0.896	0.977	1	0.964	1	1	0.993	0.971	1	1	0.971
PNB	0.912	0.906	0.983	0.976	1	0.920	1	0.967	0.965	0.984	1	0.965
ICICI	0.932	0.957	1	1	1	1	1	1	1	0.983	1	0.988
HDFC	0.953	1	0.937	1	1	0.979	0.980	1	0.980	0.986	1	0.942
AXIS	0.908	1	0.887	0.750	1	1	0.995	0.999	1	0.963	1	0.955
Average	0.918	0.944	0.969	0.965	0.995	0.986	0.994	0.977	0.980	0.983	1	0.969

Table 4. Technical efficiency (VRS model)

Bank	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	Average
SBI	1	1	1	1	1	1	1	1	1	1	1	1
BOI	1	1	1	1	1	1	1	1	1	1	1	1
CBI	1	1	1	1	1	1	1	0.962	0.989	0.985	1	0.994
UBI	1	0.998	1	1	0.999	1	1	1	0.972	1	1	0.997
PNB	1	1	1	0.994	1	0.978	1	1	1	0.993	1	0.997
ICICI	1	1	1	1	1	1	1	1	1	1	1	1
HDFC	1	1	0.965	1	1	1	0.981	1	1	0.989	1	0.994
AXIS	1	1	1	0.839	1	1	1	1	1	0.963	1	0.982
Average	1	0.999	0.979	0.979	0.999	0.997	0.997	0.995	0.995	0.991	1	0.995

All computation was performed using DEA Frontier program. The efficiency of commercial banks in Indian was first examined by applying the DEA approach for each year by using a common frontier. We then examine the analysis by examining the efficiency of private banks only, public banks only and a pooled common frontier for all banks for all years.

Tables 1 and 2, give results of efficiency scores estimated according to the DEA method respectively under the assumption of CRS and VRS. Scores efficiency is obtained by calculating the average score for each bank. The average efficiency score over all the period is 0.936 with CRS and 0.985 with VRS.

Average total economic efficiency that shows ability of firm in efficient allocation of inputs according to their prices is equal to 0.991. According to tables efficiency trend is increasing as it decreased from 2000 but again increased to efficient level in 2011. Among the banks during this period Bank of India (BOI) and ICICI Bank respectively with 0.999 and 0.998 had the highest average efficiency and Axis bank had the lowest average efficiency (0.957).

The evolution of technical efficiency scores by banks (VRS assumptions) over the considered period reveals that Bank of India (BOI), State bank of India (SBI), and ICICI bank have an stable tendency, while Central bank of India (CBI) and Axis bank have unstable ones. The special case of CBI, decreasing tendency from 2006 to 2009 and increasing after is due to the raise of the investment and loans at the end of the period, while the inputs levels remained steady. Therefore, CBI banks were more efficient in producing that specific asset with almost the same level of inputs than the other years.

A more detailed analysis, of efficiency degrees per banks groups (state owned and private) shows that on average public banks are more efficient ones except ICICI bank which is pioneer private bank in the case of technology adoption in India.

For each year in the testing period, there are more technically efficient Indian banks than allocative and cost efficient banks (see tables). The mean technical efficiency score peaked at 1 in the years 2000. It then decreased slowly for the rest of the examining period till 2011. This could be partially explained by the inability of most Indian banks to capture the full benefits of upgrading their equipment and systems, particularly in respect to staffing level and branch locations.

Among the public banks, Bank of India (BOI) and State Bank of India (SBI) show better performance and are the most efficient banks and Central bank of India (CBI) has the lowest efficiency as compare to other public banks and among private banks ICICI bank is the most efficient bank and Axis bank has the lowest rank.

The results show a fluctuating trend in efficiency scores of banking sector operating in India. As per CCR models, banks' efficiency increased from 2000-01 to 2004-05 then from 2005-06 trend faced a slight decline to 2007-08 and after that increased slightly to efficient level as the score is 1 which means 100% efficiency in 2010-11. According to BCC models, bank's efficiency decreased from 2000-01 to 2002-03. In the next year it starts to increase to 2005-06 and from 2006-07 to 2009-10 trends was unstable till to 2010-11 it increased up to efficient level 1.

The efficiency scores from the analysis clearly indicate from the selected Banks, Public banks more efficient with the highest efficient level as close to 1 in all the years by both the models. It is clearly shown that Indian financial market is still dominated by public banks.

5. Conclusion

Using non-parametric approach Data Envelopment Analysis (DEA) methodology enables us to estimate economic, technical, and allocative efficiency. We have run tests for each year, Public banks, private banks, and for all banks for all years.

The results suggest that the mean overall or economic efficiency was 100 percent in 2000, decreasing to 98 percent in 2002, and remained unstable from 2003 to 2009 with fluctuating in percentage till 2010-11 which reached to 100 percent again.

The cost efficiency estimated for the banks under study averaged 93% when the estimates are derived under constant return to scale while the estimates averaged around 99% under variable return to scale over 2000-2010. The efficiency scores vary across banks based on their relative size and across their geographical locations. Based on the size, the largest banks are found to be relatively the most cost efficient. These cost estimates suggest that the same level of output could be produced with approximately 93- 99% of their current inputs if banks under study were operating on the most efficient frontier.

When we decomposed the cost efficiency into technical and allocative efficiency, the allocative

efficiency scores in particular, vary considerably based on bank's size and bank's geographical location. The technical efficiency averaged around 99% for the banks under study with insignificant differences among the banks under study. This suggests that the banks under study might increase one or more of their current outputs by around 1% without reduction in their other outputs or without a need for more inputs. Bank of India averaged the highest technical efficiency in both model while the Central bank of India along with Axis bank averaged the least under both constant and variable returns to scale.

The allocative efficiency scores averaged around 0.991 for the banks under study and the bank of India, ICICI bank, and State bank of India are found to be the most allocative efficient and realized an efficient score the highest while the Axis banks are found to be the least.

Finally, while the India have implemented many economic and financial reforms over the last decades or so, these do appear to have positive impact on the efficiency of the respective banking systems under study and it shows an increasing trend in performance of Indian banks caused by IT innovation, competition, better supervision, and enlarged investment in new information technology during the recent time period (2000-01 to 2010-11). The banks were left with no option but to improve their functional attitude, strategies and policies. In this paper, while the author proposes ways to achieve compromise solutions, recommend further research in the area to incorporate the dynamic nature of such decisions.

In comparison with international standards, Indian banks would need to improve their technological orientation, to continue their efforts to reduce the percentage of non-performing assets and expand the possibilities for augmenting their financial activities in order to improve their profit efficiency in the near future.

References

[1] Andrie, A, M and Cocris, V, (2010). Comparative Analysis of the Efficiency of Romanian Banks. *Romanian Journal of Economic Forecasting*,13(4): 54-75.

[2] AlKhathlan,KH and Malik,S, A. (2008). Are saudi Banks Efficient? Evidence using Data Envelopment Analysis (DEA). *International journal of economics and finance.* 2(2):53-58.

[3] Bauer Paul, Allen Berger, and David Humphrey. (1992). *Efficiency and Productivity Growth in US Banking. in The Measurement of Productivity Efficiency; Techniques and Applications*, edited by HO Fried, C.A.K Lovell, and P.Schmidt, (Oxford University Press) 386-413

[4] Berger, A. and Humphrey, D. (1997). Efficiency of Financial Institutions: International Survey and Directions for Future Research. *European Journal of Operational Research.* 98(5): 175-212.

[5] Charnes A., Cooper W.W. and Lewin A. Y. and Seiford L. M. (1994), *Data Envelopment Analysis: Theory, Methodology and Applications*, Kluwer Academic Press, Boston, 3-12.

[6] Debasiah S S. (2006). Efficiency Performance in Indian Banking –Use of Data Envelopment Analysis. *Global Business Review*, 7(2).325 – 333.

[7] Das.A and Ghosh. S,(2006). Financial deregulation and efficiency: An empirical analysis of Indian banks during the post reform period. *Review of Financial Economics.* 15(3). 197-199.

[8] Moh'd Al-Jarrah, I. 2007. The use of DEA in measuring efficiency in Arabian banking. *Banks and Bank Systems Journal*. 2(4) :21-30.

[9] Ikhide S.(2000). Efficiency of Commercial Banks in Namibia. BON Occasional Paper No. 4. Available at: www.bon.com.na/docs/pub/Efficiency%2520of%2520Commercial%2520Banks%2520in%2520Namibia.pdf.

[10] Favero, C. and Papi, L. (1995). Technical Efficiency and Scale Efficiency in the Italian Banking Sector: A Non-parametric Approach. *Applied Economics.* 27: 385-395.

[11] KUMAR, S and GULATI, R. (2008). An Examination of Technical, Pure Technical, and Scale Efficiencies in Indian Public Sector Banks using Data Envelopment Analysis. *Eurasian Journal of Business and Economics*.1 (2): 33-69.

[12] Kumbhakar. S.C. and Sarkar, S. (2004). Deregulation, Ownership and Efficiency in Indian Banking: An application of Stochastic Frontier Analysis. IGIDD working paper. Available at: www.igidr.ac.in/conf/finwrk/workshop.pdf.

[13] Mostafa, M. M. (2007). Modeling the efficiency of top Arab banks: A DEA–neural network approach. *Expert Systems with Applications* 36(1): 309-320.

[14] Satye. M, (2003). Efficiency of Banks in developing Economy. The case of India. European *Journal of Operational Research*. 148(3). 662-671.

[15] Soori. A. R, Garshasbi. A. R, Oryani. B, (2005). Comparative Analysis of the Efficiency in the Commercial Banks of Iran: Using DEA and SFA Approaches. *New Economy and Commerce Journal*. 8(2):33-60.

[16] San.O.T, Lim Y. T, and Teh B, H .(2011). A Comparison on Efficiency of Domestic and Foreign Banks in Malaysia: A DEA Approach. *Business Management Dynamics.*1 (4):33-49.

[17] Sealey, C. and Lindley, J. (1977). Inputs, Outputs and a Theory of Production and Cost at Depository Financial Institutions. *Journal of Finance*. 32(4): 1251-1266.

[18] Tortosa-Ausina, E.(2002).Banks Cost Efficiency and Output Specification. *Journal of Productivity Analysis*. 18(3): 199-222.

The Influence of Celebrity Endorsements on Stock Prices

Ryan Ladner

School of Adult & Graduate Studies, Bryan College
721 Bryan Drive, United States
Ryan.Ladner@bryan.edu

Abstract **- Celebrity endorsements have been long been used to promote companies' products and strengthen brands; however, celebrity endorsements can also be linked to an increase in company stock prices. The purpose of this paper is to examine the influence of celebrity endorsements on stock prices. An overview of the theoretical framework of the celebrity will be applied to stock analysts and a review of the abnormal returns of this influence will be presented.**

Keywords *- Marketing, stock-price, investor, celebrity endorsement, efficient market, inefficient market, behavioural finance*.

1. Introduction

In traditional financial theory, capital markets have been regarded as adherent to the *efficient-market hypothesis*. The efficient-market hypothesis assumes a market where all information is available to all investors. When prices reflect this availability, the markets are considered efficient (Fama, 1970). In the efficient-market hypothesis it is stated that investors with a well-diversified portfolio cannot consistently earn more or less relative to the market average (Mayo, 2010). According to the efficient-market hypothesis, investors react so quickly to changes in the news that markets remain efficient; however, based on research, this has not always been the case. The emerging field of behavioral finance has placed this theory and the rationality of investors in question. The argument in behavioral finance is that individual investors make decisions based on heuristics and biases due to limited time and information, and thereby develop shortcuts to making decisions (Ackert & Deaves, 2009, p. 83). Investors who make decisions based on heuristics and biases do not make decisions with all existing information, which leads to anomalies in the market. These anomalies are counterintuitive to the concept of market efficiency (Ackert & Deaves, 2009). Upon study, there are many reasons why these anomalies exist; however, one that has acquired much attention deals with endorsements by Celebrity stock endorsers (Barber, Lehavy, McNichols, & Trueman, 2001; Barber & Loeffler, 1993; Karniouchina, Moore, & Cooney, 2009; Metcalf & Malkiel, 1994; Womack, 1996). With the influence of the Internet, online television, podcasts, and YouTube, individual investors have unprecedented access to information and celebrities and experts can reach a broader audience than ever before. This reach has produced several event-study analyses that show market inefficiency and raise questions regarding the implications of endorsements and stock price returns for companies.

2. Celebrity Endorsements

Celebrity endorsements provide more value than merely attracting customers to products. For example, just the announcement of a celebrity endorsement can result in a rise of a company's stock price. For example, studies have shown when a company announces that a celebrity will speak for their product, an increase in the price of company stock results (Agrawal & Kamakura, 1995). Not only does the increase in stock price occur during the initial announcement, the increase remains in effect for the life of the advertisements or the celebrity's popularity. Celebrity endorsements have offered positive stock returns simply by virtue of the favourable mention of their financial performance. The presumed trustworthiness of endorsers entices investors to purchase the stock without performing due diligence. Marketing executives can use this phenomenon to make brand and advertising decisions and to increase the return on investment of marketing operations for their companies.

Company leaders have long used celebrity endorsements to increase awareness and sales of their products. To understand the effects of celebrity endorsements and advertising, it is important to provide a theoretical framework of the celebrity and the foundational marketing definition. A celebrity is traditionally defined as "a person who is well known by the public" (Friedman & Friedman, 1979, p. 63). This person is usually attractive and/or likeable and may possess some type of expertise or achievement (Kamins, Brand, Hoeke, & Moe, 1989).

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.007.12/ijltfes.2047.0916

Celebrities are used in advertisements because they enhance brand awareness and message recall and produce a higher probability that consumers will buy the product (Agrawal & Kamakura, 1995). A celebrity's achievements and likeability alone will not entice consumers to purchase products. To be effective as a celebrity, and therefore influence stock prices, a celebrity must possess traits that consumer's desire and one of those traits is credibility. There are two models in the literature that focus on the credibility of a celebrity – the source credibility model and the source attractiveness model (Erdogan, 1999).

2.2 Source Credibility and Attractiveness Models

The source credibility model represents someone who has expertise, trustworthiness, and attractiveness (Dholakia & Sternthal, 1977; Doss, 2011; Ohanian, 1991). Expertise "refers to the amount of knowledge that a source is perceived to have about a subject" (Erdogan, 1999, p. 298). Expertise is a key factor in determining the credibility of the celebrity and has been shown to be the most influential aspect of selling to the consumer (Erdogan, 1999). Trustworthiness represents the confidence that consumers place in the celebrity advertisers and whether or not they believe he or she is making valid statements (Amos, Holmes, & Strutton, 2008; Ohanian, 1990).

Amos et al. (2008) concluded that the trustworthiness of the celebrity is the most effective part of the source credibility model and with advertisements featuring celebrities as a whole. The source attractiveness model integrates neatly into the source credibility model. The overlapping factor in the two models is attractiveness (Ohanian, 1990). It is important to note that attractiveness does not mean just the physical aspect of the celebrity advertiser. Attractiveness also represents the celebrity's familiarity and likeability (Erdogan, 1999). When analyzing the effect of celebrity endorsements on the price of stocks, it is important that the endorser represent all three traits in the source credibility model to be effective. For many individual investors, the endorser of stocks may be the only source he or she uses for stock purchase information for retirement funds or other important monetary aspects of their life and credibility is an important factor in making those decisions.

The celebrity endorser is represented in three main personas: (a) the spokesperson, (b) the endorser, and (c) the testifier (Erdogan, 1999). An overview of the personas will be provided with a focus on a professional expert. A spokesperson or testimony by a celebrity is generally not going to appeal to someone who is purchasing stocks. It is critical that the endorser of the purchase be an expert in his or her field with experience of stock purchases. In the case of these stock purchases, the most influential celebrity will be the professional expert endorser. The professional expert endorser is "an individual or group possessing superior knowledge regarding the product class endorsed and (who) has obtained this knowledge as a result of experience, study, and training" (Kamins et al., 1989, p. 63). Because the purpose of this study is to look at financials, it is important to note that researchers have found that experts perform better for products with high financial performance and therefore are the most relevant for this review (Friedman & Friedman, 1979).

2.3 Celebrity Stock Endorsers

One of the most popular celebrity endorsers of stocks is Jim Cramer. Cramer is the host of *Mad Money* on CNBC. More than 250,000 viewers watch *Mad Money* each day (Karniouchina et al., 2009, p. 245). Cramer's show is described as a mix of "professional wrestling, infomercial, pitching, and hyperkinetic game shows, all the while dispensing stock tips to the couch potato investors" (Becker, 2005, p. 10). Cramer's dynamic show attracts many rookie investors who look to him for recommendations. Neumann and Kenny (2007) described Cramer as "a man who blurs the line between creating business news and covering it" (p. 603). Cramer has critics and may not be appropriate for all investors; however, for the many people who watch his show, there is no doubting the influence that he holds (Lawler, 2009).

Cramer has source credibility and can easily be categorized as a professional expert. Cramer is a graduate of Harvard College and a former hedge fund manager at Cramer Berkowitz. During his tenure as senior partner, he amassed a 24% rate of return for 15 years (Kadlec, 2002) As far as trustworthiness goes, Cramer does not invest his own money and therefore does not directly profit from the recommendations he makes. Cramer holds a charitable trust and does not have to make any disclosures when recommending stocks to individual investors except that it is held by his charitable trust. Knowing that Cramer is not making money for himself may entice individual investors to trust him without inhibition.

In addition to television show hosts like Jim Cramer, other popular stock analysts must be considered as celebrity endorsers. One of the most popular and widely used stock analysts are The Motley Fools, whose website has been described as "the most popular internet stock chat website" (Giacomino & Akers, 2011, p. 37). The Motley Fools website often provides information that is "contradictory to academics, often makes mistakes, and should be critically evaluated" (Giacomino & Akers, 2011, p. 44). However, even though this information is well known, The Motley Fools and their recommendations affect the stock market just as much as Jim Cramer and other analyst's recommendations (Giacomino & Akers, 2011, p. 44).

It is interesting to note that the mission statements of Jim Cramer and of The Motley Fools are very similar. Cramer states that he wants to "educate people, entertain people, and help them make money" (Becker, 2005, p. 10). The Motley Fools mission statement says that they are "here to educate, amuse, and enrich" (Giacomino & Akers, 2011, p. 37). These mission statements have entertainment and influence as core values. If the efficient-market hypothesis was completely valid and investors only bought and sold stocks based on effective financial evaluations of financial performance, then the financially unrelated investment advice and entertainment value of Cramer and The Motley Fools would have no weight in the capital markets; however, event studies show otherwise.

3. Event Study Analysis

3.2 Mad Money

An event study analysis is critical in determining the celebrity endorsement effect of Jim Cramer on the efficient-market hypothesis. An event study focuses on variables that occur during a specified period of time. In the past, event study analyses have been used to analyze what type of effect events have on profitability (Agrawal & Kamakura, 1995). Results from event analysis indicate that celebrity endorsements have positive effects on the return of stock prices around the day of endorsement. (Agrawal & Kamakura, 1995).

The three event studies analyzed for Jim Cramer all show similar results: an increase in volume of shares traded after Cramer's recommendations and abnormal stock returns greater than 1.00% that decrease back to normal levels over time (Karniouchina et al., 2009; Neumann & Kenny, 2007). Neumann and Kenny (2007) analyzed 216 recommendations by Jim Cramer from July 26, 2005 to September 9, 2005. During the first trading day after a buy recommendation aired, "abnormal returns of 1.06%, 1.09%, and 1.00% relative to the market model, CRSP index, and historical mean (were) realized" (Neumann & Kenny, 2007, p. 605). In addition, the estimated volume for the particular stocks increased by 27.78% the day after the show date, suggesting abnormal buying and selling after mentioning them on his show (Neumann & Kenny, 2007). Engleberg, Sasserville, and Williams (2006) found similar results in their event study of 246 initial recommendations given by Cramer between July 28, 2005 and October 10, 2005 (p. 2). The cumulative abnormal return for the study was 6.71% for its value three days before the recommendation and 1.96% overnight (Engleberg et al., 2006, p 2). A turnover ratio was also computed in the study showing that the volume of shares traded of "smaller firms is 317% of its typical size the day of recommendation, 890% the day following recommendation, and 451% on the second day following the recommendation" (Engleberg et al., 2006, p. 7). Karniouchina et al. (2009) also showed abnormal returns of 1.07% and 1.23% on the opening and closing sections of the show (p. 251). The event studies all portray the increased volume and abnormal returns of the stock picks of Jim Cramer.

When looking at Jim Cramer as a celebrity and professional expert, it is apparent that even though his show is not an advertisement, it is definitely persuasive and is lacking ambiguous statements normally involving stock picks (Karniouchina et al., 2009). Contrary to an advertisement where individuals watch paid promotions, people watching Jim Cramer's show are actively looking for recommendations (Karniouchina et al., 2009). Cramer's expertise as a celebrity is a reflection of his track record. Even though Cramer is not trying to sell anything, his track record is important to his career as a stock analyst (Karniouchina et al., 2009). It is because of this endorsement and professional value offered to naïve investors that abnormal returns occur after his recommendations.

3.3 Analyst Recommendations

According to Barber and Odean (2008), investors are more likely to buy rather than sell stocks that catch their attention (Barber & Odean, 2008). With the explosion of information available on the Internet, it has become increasingly easy for stocks to catch the attention of investors. When analysts such as The Motley Fools change recommendations, they advertise them publicly and investors are able to buy or sell, reacting to them immediately (Barber et al., 2001). Analyst recommendations are key determinants of trading volume from day to day. Investors react to changes in recommendation by security analysts at the end of the trading day (Barber et al., 2001). Abnormal gross returns from analyst recommendations are on average 4.13% for buy recommendations and 4.91% for the sell recommendations (Barber et al., 2001, p. 561)

One of the most famous studies of analyst recommendations occurred with the very popular "Dartboard" column of *The Wall Street Journal*. In the study by Barber and Loeffler (1993), stocks picked in the "Dartboard" column by professional experts of *The Wall Street Journal* earned abnormal returns of 4% and abnormal volume for six days after the picks appeared in the column (Barber & Loeffler, 1993, p. 277). In addition, one of the most famous analysts, The Motley Fools, also shows abnormal volume and returns. When The Motley Fools announce a buy recommendation there is an average price increase of $3.36 to $3.72 per share and during the three day period following there is an average increase of $6.08 to $6.87 per share. In addition, there is a 126.53% increase in the volume of trades during an announcement, followed by a 114.43% increase the day after (Hirschey, Richardson, &

Scholz, 2000, p. 68). This reinforces the belief that analyst's recommendations influence the buying behavior of investors. The research suggests that investors fail to do their due diligence when they purchase stocks and instead follow the patterns of celebrity endorsements by Jim Cramer and other popular analysts.

3.4 Marketing Strategy

Investors seem to be looking to take shortcuts when it comes to investing in stocks. The heuristics involved show that many investors fail to do their own research when it comes to stock picks. These shortcuts using the recommendations of Jim Cramer, The Motley Fools, and other analysts provide an astonishing opportunity for marketing executives. First, attention grabbing analysts and news stories seem to produce abnormal stock returns. It is the individual investors and not institutional investors who are the most susceptible to purchasing stocks that grab attention and produce news (Barber & Odean, 2008). Marketing executives could monitor these shows and analyst recommendations for opportunities to sell products to investors.

One factor being studied is the effect that ownership of a company stock leads to the purchase of company products (Aspara & Tikkanen, 2008; Frieder & Subrahmanyam, 2005). There is evidence that the investors may buy the products of companies because they are a stock owner (Aspara & Tikkanen, 2008). In traditional consumer behavior, the positive attitude a person has about a company influences their purchases (Aspara & Tikkanen, 2008). Consumers are influenced on stock ownership and product purchases because the stock has personal relevance to them and the company's brand (Aspara & Tikkanen, 2010).

Aspara and Tikkanen (2010) found that consumers manifest the personal relevance of a product purchase in two ways. First, if a consumer is evaluating two different stocks and they both have the same financial returns and risks, the consumer will be more willing to invest in the one that is personally relevant. Second, if a stock purchase is not producing the returns expected, a consumer may be more willing to deal with the low financial returns of a stock in which the consumer has no personal relevance (Aspara & Tikkanen, 2010). According to the survey sent out in the study, only 14.3% of the respondents stated that they did not care which stock to invest in if both had similar financial returns and risk; accordingly, 85.7% of the respondents were willing to invest in a stock for a reason beyond its expected financial returns (Aspara & Tikkanen, 2010, p. 21).

From the research of stock purchases and celebrity endorsements, it is apparent that investors make decisions for reasons other than financial returns. There are some heuristics and biases that arise from investing and research shows that anomalies exist and challenges to the efficient-market hypothesis are evident. Marketing executives can use this information to promote products and even their stock in new ways. First, marketing executives can identify domains the company's product represent and then target customers' personal relevance of the company (Aspara & Tikkanen, 2010). Second, companies could place reminders on packaging or messaging related to investing in the company. Significant partnerships with an online broker could increase the reach for the company and the broker. Third, companies could market products to buy and stock to invest in. They could accomplish this through shareholder advertising and promotion to customers who already own company stock (Aspara & Tikkanen, 2010). Celebrity endorsers could provide the avenue through which to communicate these strategies to customers.

Following the marketing framework segmentation, targeting, and positioning, a plan could be developed using celebrity endorsers and investors. The segmentation of the market represents a group of customers who own the stock of a corporation. Studies could be done to analyze whom this market represents and products could be marketed directly to shareholders through the annual report or other avenues. Using the demographics of the shareholders, companies could target specific populations with specific products that a company could develop. The *hybrid marketing* concept of selling not only stock but also products to the consumer could produce a target audience that companies do not normally serve and may produce a form of brand loyalty to the products and the stock. Finally, a company could position itself to offer products just to the investor. The opportunity to market to these investors could open up a new product or product line when dealing with these implications. A celebrity endorser could be used to introduce this information to customers. Sponsorship on shows such as Mad Money and analyst recommendation sites offering discounts to shareholders on products advertised by the endorsers open the door for company leaders to develop strategies for stock price and brand loyalty.

4. Conclusion

The efficient-market hypothesis remains a long-standing theory on how capital markets operate. Though regarded by many as fact, the literature examining the decisions based on analyst recommendations and famous celebrity endorsers brings to the surface doubts regarding an efficient market. There is a psychological process involved in choosing investments and companies can analyze the decision making process to market products and stock purchases to shareholders and future shareholders. Further research needs to be completed regarding this literature. It is important to understand how widespread the effects of recommendation

changes are and the possibilities of using them for *hybrid marketing* strategies of stock purchases and product purchases. In addition, could the benefit of running ads with positive analyst recommendations, or placing a celebrity stock analyst in a commercial for the companies, increase not only the brand purchase but also the ownership in the company's stock? To further research the phenomenon that occurs with celebrity endorsements and stock prices, a correlation could be conducted between the two variables using larger than a three-month sample size. This would help to establish that this was a consistent anomaly in the capital market. In addition, more analyst and television stock advice could be analyzed to determine exactly what type of celebrity endorsement credibility and trustworthiness is required to produce these anomalies in stock prices.

References

[1] Ackert, L., & Deaves, R. (2009). *Behavioral finance: Psychology, decision-making, and markets* (1st ed.). South-Western College Publisher: Oklahoma City, OK.

[2] Agrawal, J., & Kamakura, W. A. (1995). The economic worth of celebrity endorsers: An event study analysis. *Journal of Marketing*, *59*(3), 56.

[3] Amos, C., Holmes, G., & Strutton, D. (2008). Exploring the relationship between celebrity endorser effects and advertising effectiveness. *International Journal of Advertising*, *27*(2), 209-234.

[4] Aspara, J., & Tikkanen, H. (2008). Interactions of individuals' company-related attitudes and their buying of companies' stocks and products. *Journal of Behavioral Finance*, *9*(2), 85-94.

[5] Aspara, J., & Tikkanen, H. (2010). Consumers' stock preferences beyond expected financial returns: The influence of product and brand evaluations. *International Journal of Bank Marketing*, *28*(3), 193-221.

[6] Babin, B. J., & Harris, E. (2010). *CB2 (with review cards and CB4ME.COM Printed Access Card) (Student Edition)* (2nd ed.). South-Western College Pub.

[7] Barber, B. M., & Loeffler, D. (1993). The "Dartboard" column: Second-hand information and price pressure. *Journal of Financial & Quantitative Analysis*, *28*(2), 273-284.

[8] Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, *21*(2), 785-818.

[9] Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *Journal of Finance*, *56*(2), 531-563.

[10] Becker, A. (2005). Jim Cramer is going "Mad." *Broadcasting & Cable*, *135*(31), 10.

[11] Dholakia, R. R., & Sternthal, B. (1977). Highly credible sources: Persuasive facilitators or persuasive liabilities? *Journal of Consumer Research*, *3*(4), 223-232.

[12] Doss, S. (2011). The transference of brand attitude: The effect on the celebrity endorser. *Journal of Management & Marketing Research*, *7*, 1-11.

[13] Engleberg,J., Sasserville, C., & Williams J. (2006). *Is the market Mad? Evidence from Mad Money*. Unpublished manuscript, Kellogg School of Management, Northwestern University, Evanston, Illinois, United States.

[14] Erdogan, B. Z. (1999). Celebrity endorsement: A literature review. *Journal of Marketing Management*, *15*(4), 291-314.

[15] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*(2), 383-417.

[16] Frieder, L., & Subrahmanyam, A. (2005). Brand perceptions and the market for common stock. *Journal of Financial and Quantitative Analysis*, *40*(01), 57-85.

[17] Friedman, H. H., & Friedman, L. (1979). Endorser effectiveness by product type. *Journal of Advertising Research*, *19*(5), 63-71.

[18] Giacomino, D. E., & Akers, M. D. (2011). Examining an online investment research service: The Motley Fools. *Journal of Business & Economics Research*, *9*(1), 37-48.

[19] Hirschey, M., Richardson, V. J., & Scholz, S. (2000). How "foolish" are internet investors? *Financial Analysts Journal*, *56*(1), 62.

[20] Kamins, M. A., Brand, M. J., Hoeke, S. A., & Moe, J. C. (1989). Two-sided versus one-sided celebrity endorsements: The impact on advertising effectiveness and credibility. *Journal of Advertising*, *18*(2), 4-10.

[21] Karniouchina, E. V., Moore, W. L., & Cooney, K. J. (2009). Impact of Mad Money stock recommendations: Merging financial and marketing perspectives. *Journal of Marketing*, *73*(6), 244-266.

[22] Kadlec, D. (2002). Cramer vs. Cramer. *Time*, *159*(13), 54.

[23] Lawler, J. (2009). Who's got your back? *Entrepreneur*, *37*(9), 60.

[24] Mayo, H. B. (2010). *Investments: An introduction* (10th ed.). South-Western College Publishers: Oklahoma City, OK.

[25] Metcalf, G. E., & Malkiel, B. G. (1994). The Wall Street Journal contests: The experts, the darts, and the efficient-market hypothesis. *Applied Financial Economics*, *4*(5), 371-374.

[26] Neumann, J. J., & Kenny, P. M. (2007). Does Mad Money make the market go mad? *Quarterly Review of Economics & Finance*, *47*(5), 602-615.

[27] Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39-52.

[28] Ohanian, R. (1991). The impact of celebrity spokespersons' perceived image on consumers' intention to purchase. *Journal of Advertising Research*, *31*(1), 46-54.

Efficiency-wage Hypothesis and the Operational Production Pattern

Brian W. Bresnahan[#1], Nadeem Naqvi*[2], Carolin Schürg[+3]
[#]*Department of Radiology, University of Washington, Seattle, WA U.S.A.*
[*]*Department of Economics, Justus Liebig University, Giessen, Germany*
[+]*Department of Finance, Frankfurt School of Finance & Management, Frankfurt, Germany*
[1]bres@uw.edu
[2]nadeem.naqvi@wirtschaft.uni-giessen.de
[3]c.schuerg@fs.de

***Abstract* - An economy's production set is the collection of all net output vectors that the economy is capable of producing with a given technology and fixed quantities of primary factors of production. The boundary of this set is called the production possibility frontier or PPF. We show that, if the efficiency-wage hypothesis holds, a country's PPF, though conceptually valid, is an operationally irrelevant concept, because the economy never operates on the PPF, which is a view that ought to be appreciated in light of persistent unemployment in the new structure of economies of the post-21st-Century-crisis world.**

***Keywords*:** general equilibrium, production set, production possibility frontier, efficiency wage, economic policy, labor market, factors of production

1. Introduction

In mainstream economics, an economy's production set is taken to be the collection of all net output vectors that the economy is capable of producing, by transforming inputs into outputs, with a given technology and with fixed quantities of primary factors of production.[1] The boundary of this set is called the production possibility frontier (PPF). An essential feature of the production set of an economy is that it is invariant to changes in market prices.[2] In such a case, there is a *unidirectional* relationship insofar as exogenous changes in technology or fixed factor supplies can induce changes in general equilibrium market prices, but changes in market prices do *not* affect the structure of the production set, nor shift its boundary in any manner whatsoever. At a theoretical level, this paper forms part of a program, the purpose of which is to establish that *operationally* the economy's market-invariant PPF is an irrelevant concept insofar as there is a *bidirectional* relationship between an economy's *actual* production possibilities on the one hand, and the general equilibrium market prices, on the other. We accomplish the goal of establishing this claim by utilizing the efficiency-wage hypothesis in the manner of Solow (1979).

Egbert and Naqvi (2011) also achieve this objective. However, they do so by embedding a small open economy in an integrated world capital market, so that under endogenous international capital mobility, a government policy change can alter the quantity of capital that locates in the country, and thereby change the structure of its economy's production set, and *inter alia* shift the boundary of such a set. In this paper, we show that if the efficiency-wage hypothesis holds, the *operational* production pattern in the economy has nothing to do with the concept of a PPF in economics as we know it. Moreover, in contrast with Egbert and Naqvi, we demonstrate that this is so both (a) without international capital mobility in a small open economy, and (b) in a closed economy.

Related work by Albert and Meckl (2001) provides a canonical formulation of the Heckscher-Ohlin (HO) model with the efficiency-wage phenomenon exhibited in both sectors of the economy. This is based on the formulation of Summers (1988). By making the effort function in each sector depend on fixed, though intersectorally differential, mark-ups on the economy-wide, weighted-average wage rate (their reference wage), their model simultaneously exhibits, both involuntary unemployment and stable inter-industry wage differentials. In addition, they demonstrate that (1) all properties, including the Stolper-Samuelson theorem, the Ryczynski theorem, and so on, hold, if factor intensities are interpreted in the cost-share sense, and (2) that the result of immigration can be a lowering of the unemployment rate, whereas Foreign Direct Investment (FDI) can raise unemployment. These latter results are both quite startling – though entirely correct and intuitive – once the dependence of unemployment changes on sectoral distribution of employment is noted.

Our purpose here is different from that of Albert and Meckl (2001). It is to demonstrate the endogenous character of the operational PPF in the face of persistent unemployment, in a Specific Factors model.[3] To this end, we adopt the efficiency wage function used by Solow (1979).[4]

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.008.12/ijltfes.2047.0916

Shiller (2010) and Stiglitz (2011), among others, call for endogenizing some variables that were in the pre-21st-Century-crisis world taken to be exogenously specified in economic models. Blinder (2010) goes further and asks for a complete overhaul of the macroeconomics curriculum, and talks about planning to drop the assumption of a single-interest-rate economy from macroeconomic models in the next, 12th Edition, of his joint text with Baumol.[5] These observations are driven by the failure of macroeconomics to adequately predict or fully comprehend the economic crisis impacting the global economy starting in 2007 to 2008. Our primary purpose in writing the current paper is to heed this call for change in the spirit of Blinder, Shiller and Stiglitz.

Section 2 lays out the Specific Factors model with the efficiency-wage hypothesis, and shows how the pattern of production and the unemployment rate are market dependent for a small open economy without international capital mobility. Section 3 considers the effects of government-policy induced market price changes on the production pattern in the economy. Section 4 extends our result to the case of a closed economy, so as to rule out any presumption that such market-dependence of production pattern arises from the openness of the economy. Section 5 contains concluding remarks.

2. Specific Factors Model with the Efficiency-wage Hypothesis

Consider the standard Specific Factors model with the efficiency-wage hypothesis built into it.[6] Let all economic activity in an economy be divided into two parts: Manufactured goods, *M*, and Services, *S*, produced by the technology embodied in the following production functions.

$$M = F(e(w_L)L^d, K_m) \quad (1)$$

and

$$S = G(\bar{H}, K_s), \quad (2)$$

where L^d is the number of unskilled workers demanded in the manufacturing sector, and K_m the endogenously determined quantity of capital employed in the manufacturing sector of the economy, whereas $\bar{H}$ andK_s are, respectively, the number of skilled workers and the amount of capital employed in service-sector production. In this economy, capital is intersectorally mobile, but unskilled labor is specific to manufacturing whereas skilled labor is specific to services. Moreover, in Eq. (1), $e(w_L)$ is the number of efficiency units of labor delivered by each worker, with the property that it rises as the unskilled wage rate rises, but at a diminishing rate, so that $e'(w_L) > 0$ and $e''(w_L) < 0$, and w_L is the unskilled wage rate.[7] If unskilled labor employed in manufacturing is seen in terms of efficiency units, instead of in terms of the number of workers employed, it would be $N = e(w_L)L^d$ effective units.

Each firm in the manufacturing sector is perfectly competitive in the commodity market and the capital market, where it behaves as a price taker. It chooses how much capital to rent and how many unskilled workers to hire. However, it also chooses what wage rate to offer the unskilled workers so as to maximize profit. In the labor market, therefore, the manufacturing firms have a limited ability to determine the wage offer. The unskilled wage rate is thus a decision variable for each firm because a higher wage rate induces each hired worker to deliver greater efficiency. All firms are identical, as are all workers, and capital is homogenous as well.

Here $F(e(w_L)L^d, K_m)$ and $G(\bar{H}, K_s)$ are *concave* production functions that are characterized by (i) the Inada conditions, including indispensable inputs, (ii) constant returns to scale, and (iii) the law of diminishing returns, which together imply that (iv) inputs are co-operative.[8]

First consider the case in which this is a small open economy. Further, let services be the *numéraire* commodity, so that the price of the service sector output equals one. Then, p_m is the relative price of the manufactures in terms of services. Additional relationships that hold are

$$p_m F_K(e(w_L)L^d, K_m) = G_K(\bar{H}, K_s) = r \quad (3)$$

and

$$K_m + K_S = \bar{K}. \quad (4)$$

Equation (3) asserts that, given that capital is mobile across sectors, the values of marginal product of capital are equal in both sectors, and their common value equals the endogenously determined domestic rental rate of capital, *r*, measured in terms of services, and Eq. (4) asserts that the demand and hence employment of capital in the two sectors equals its fixed supply $\bar{K}$, on the assumption that *r* is perfectly flexible.

Additionally, the unskilled wage rate in the economy is endogenously determined by

$$p_m F_L(e(w_L)L^d, K_m) = w_L/e(w_L) \quad (5)$$

and

$$\frac{e'(\hat{w}_L)\hat{w}_L}{e(\hat{w}_L)} = 1, \quad (6)$$

where Eq. (6) is the well-known Solow elasticity condition.[9] The unskilled wage rate, while in principle perfectly flexible, is sticky at $\hat{w}_L$, which is determined solely by Eq. (6), due to the incentive of the manufacturing firms to offer the wage that minimizes cost with respect to a unit of effective, effort-adjusted unskilled labor, rather than with respect to raw, unskilled labor that is not adjusted for efficiency.[10]

Finally, we have

$$G_H(\bar{H}, K_S) = w_H, \quad (7)$$

which determines the skilled wage rate. Notice that in the model, the two wage rates are also expressed in terms of the same *numéraire* commodity, *viz.*, services.

Without loss of generality, assume that manufactures are imported (implying that some services are outsourced to this country by some foreign countries). Then p_m is the *domestic* relative price of the manufactures in terms of services, and $p_m = p_m^* + t$, where p_m^* is the *world* price of manufactures that this country takes as exogenously given, since p_m^* is determined on the world market for manufactured goods. Here, $t \geq 0$ is the import tariff that is exogenously imposed by the country's government. This tariff has a value that is less than t_p, which is the prohibitive tariff that snuffs out all imports.

As already noted, from Eq. (6) alone, the value of $w_L = \hat{w}_L$ is uniquely determined. With this, for $t = 0$, which means that in free trade equilibrium, from Eq. (3), Eq. (4) and Eq. (5), $K_m = \hat{K}_m$, $K_S = \hat{K}_S$ and $L^d = \hat{L}^d$ are uniquely determined, as is $r = \hat{r}$, all of which are functions of the exogenous variables $\bar{K}$, $\bar{H}$, and p_m. Finally, substituting for $\hat{K}_S$ in Eq. (7) uniquely determines $\hat{w}_H$ as a function of the three exogenous variables. Notice also that unemployment of unskilled workers in the economy is equal to $\hat{u} = \bar{L} - \hat{L}^d > 0$, which is a function also of the fourth exogenous variable $\bar{L}$, and this unemployment arises as an equilibrium phenomenon, because the manufacturing firms do not reduce the unskilled wage rate for fear of facing reduced efficiency of incumbent workers.

Once the equilibrium values of these endogenous variables are plugged into the production functions Eq. (1) and Eq. (2), we also obtain the pattern of production in the economy, $\hat{M}$ and $\hat{S}$, which lies strictly inside the PPF because of the unemployment of unskilled labor, despite full employment of both capital and skilled labor in the economy. It can be verified that the restrictions placed on the production functions and on the efficiency function ensure that the general equilibrium supply curves for both final commodities are upward sloping. This is true in spite of the fact that the economy operates strictly inside the PPF.

3. Policy Intervention

To see the effects of policy intervention, in particular the effect of an import tariff, a diagram is helpful.

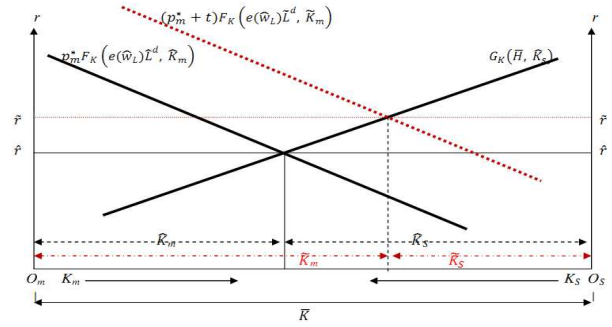


Figure 1
Graph of Equations (3) and (4), with $t > 0$

In Figure 1, the fact that $\hat{w}_L$ is determined from Eq. (6), which also determines $e(\hat{w}_L)$, from the efficiency wage function, is utilized. Substituting for these two values in Eq. (3), Eq. (4) and Eq. (5), we obtain $\hat{K}_m$, $\hat{K}_S$ and $\hat{L}^d$ under free trade, for $t = 0$, as displayed in Figure 1. Suppose next that the government imposes a positive tariff on the imports of manufactures. That means $t > 0$. Then, the left-hand side (LHS) of Eq. (3) becomes higher, displayed by a rightward shift of the curve representing the value of marginal product of capital in manufacturing, as the (red) dashed curve. To restore equilibrium, given the law of diminishing returns, the employment of capital must rise in manufacturing. In Figure 1, this is given by $\tilde{K}_m > \hat{K}_m$. With a given quantity $\bar{K}$ of capital in the economy, this means that less capital is now employed in the service sector as a consequence, given by $\tilde{K}_S < \hat{K}_S$ in Figure 1.

Notice also that with more capital employed in manufacturing, given that inputs are cooperative, the value of marginal product of unskilled labor rises, so that the LHS of Eq. (5) becomes higher. Since the RHS of Eq. (5) remains fixed due to Eq. (6), given the law of diminishing returns, the employment of unskilled labor must rise in manufacturing for equilibrium to be maintained, so that $\tilde{L}^d > \hat{L}^d$, thereby reducing unskilled unemployment in the economy, which is now $\tilde{u} = \bar{L} - \tilde{L}^d < \hat{u}$.

Due to the positive import tariff on manufactures, a change in income distribution in favor of unskilled workers occurs, simply because the unskilled wage rate remains unchanged but

unemployment is lower. Further, since the supply of skilled labor, $\bar{H}$, is exogenously given, the decline in capital employed in the services sector lowers the productivity of skilled workers, and *inter alia* reduces the skilled wage rate, w_H. This means that $\tilde{w}_H < \hat{w}_H$ and total earnings of skilled workers decline. Also, with a fixed quantity of capital in the country, a higher rental rate of capital implies that the aggregate earnings of capital owners rise. Notice that the rightward shift of the curve representing the value of marginal product of unskilled labor, the (red) dashed curve, occurs due to two forces: an increase in the domestic relative price of manufactures *and* an increase in the employment of unskilled labor in manufacturing.

Once again, under tariff protection, the new equilibrium values of these endogenous variables can be substituted in Eq. (1) and Eq. (2) to obtain the new pattern of production in the economy, $\tilde{M} > \hat{M}$ and $\tilde{S} < \hat{S}$, which also lies strictly inside the PPF because of the unemployment of unskilled labor despite full employment of both capital and skilled labor in the economy. The manufacturing sector expands, because more unskilled labor and more capital are employed under tariff protection, but the service sector contracts since some capital gets extracted from this sector.

4. Closed Economy

The economy considered so far is a small open economy for which the domestic relative price of manufactures p_m is exogenously fixed, either because of equality with the parametric world price in free trade, or due to that reason and a government specified fixed import tariff. However, if the economy is closed, p_m has to be endogenously determined, which happens due to the fact that domestic demand for manufactures must be met solely by domestic output supply, and by Walras law, the service market will also clear. So, for a closed economy, we have,

$$C^m(p_m,\ S + p_m M - \tau) + \gamma^m = M(p_m,\ \bar{K}, \bar{H})\ . \quad (8)$$

Where, in Eq. (8), C^m is domestic demand for manufactured goods, τ is the *exogenous* lump-sum tax revenue collected by the government, rendering $S + p_m M - \tau$ as the disposable personal income of the country, and γ^m is the *exogenous* government demand for manufactures. With this specification, Eq. (1) – Eq. (8) constitutes the model of a closed economy, and Eq. (8) serves to endogenously determine the Walrasian general equilibrium domestic relative price of manufactures, at a value of $p_m = \tilde{\tilde{p}}_m$.[11] In a closed economy, the relative price is an endogenous variable, but it is replaced here by two exogenous variables, γ^m and τ.

Straightforward reasoning will show that an increase in government demand for manufactures will result in a rightward shift of the curve representing the value of marginal product of capital in manufacturing, not unlike that represented in Figure 1 as the (red) dashed curve. This is analogous to an increase in the import tariff on manufactures, because it will also raise the domestic relative price of manufactures, although this will be an *induced* change rather than one that is exogenous in the tariff case.

To see this, consider an increase in γ^m. As is evident from Eq. (8), this increases the domestic aggregate demand for manufactures, which, by itself, leads to an increase in p_m to, say, $\hat{\hat{p}}_m$. As a consequence, given upward sloping general equilibrium supply curves, the output of manufactures rises to $\widehat{\widehat{M}}$, and this is accomplished, one, by increased employment of capital in manufacturing to $\widehat{\widehat{K}}_m$. Moreover, due to the cooperative nature of inputs, greater employment of capital induces increased productivity of unskilled workers, so that the LHS of Eq. (5) rises. Since the RHS of Eq. (5) remains unchanged due to Eq. (6), to restore the equality in Eq. (5), given the law of diminishing returns, manufacturing firms must hire more unskilled workers, say $\hat{\hat{L}}^d$, which constitutes the second cause of an expansion of the manufacturing sector output. Naturally, unemployment falls to $\hat{\hat{u}}$. Clearly, consequent upon an increase in γ^m, there is a rightward shift of the curve representing the value of marginal product of capital in manufacturing in Figure 1, due to both an increase in K_m and an increase in L^d.

Since capital supply is fixed, more capital employed in manufacturing implies less of it will be employed in the service sector, causing the service sector output to contract to $\hat{\hat{S}}$, and the skilled wage rate also to fall to $\widehat{\widehat{w}}_S$ because of the reduced productivity of skilled workers, since they have less capital to work with, as seen from Eq. (7).

As with an import tariff on manufactures, there is a similar change in factoral income distribution, with skilled workers earning less, $\widehat{\widehat{w}}_S \bar{H}$, while unskilled workers as a group becoming better off due to lower unemployment at the same wage rate, $\hat{w}_L \hat{\hat{L}}^d$. Also, from (3) it is evident that with lower employment of capital in the service sector, the RHS becomes higher due to the law of diminishing returns, so the rental rate of capital rises to $\hat{\hat{r}}$, leading to an increase in the income of capital owners to $\hat{\hat{r}}\bar{K}$.

It can be verified that the consequences are exactly the opposite if the government increases its demand for services, instead of increasing its demand for manufactures. Are the consequences of a reduction of government taxation on *all* the endogenous variables exactly the same as the effects of an increase in government demand for manufactures? The answer is "no." This is due to the fact that a reduction in lump-sum taxes leads to an increase in disposable personal income, and given positive marginal propensities to consume *both* manufactures and services (adding up to 1), the consumer demand for services also rises. This leads to an increase in the aggregate demand for services also, which, by creating an excess demand for services at the pre-tax-reduction relative price of manufactures, generates a force to reduce this relative price. The final outcome, therefore, depends on the *relative magnitude* of the marginal propensity to consume manufactures versus that to consume services.

This much is clear: the increase in the relative price of manufactures, in magnitude, will necessarily be less under a tax reduction scheme than under an increase in government demand for manufactures. It could well be the case that the relative price of manufactures falls, if the marginal propensity to consume services is sufficiently greater than that to consume manufactures. The output supply response of manufactures and services will correspondingly be dependent on the direction and magnitude of the change in the relative price of manufactures. Thus the consequences for the economy are much more complicated due to a tax reduction than due to an increase in government demand for manufactures.

To obtain more precise inferences from the effects of a tax-reduction policy, more restrictions would have to be placed on consumers' preferences, for example, that personal preferences are both identical and homothetic. Such restrictions would be helpful for positive-theoretic purposes of description and prediction, though not for the normative purpose of policy prescription.[12] If all consumers do have preferences that are identical and homothetic, the behavior of all consumers can be portrayed by a single consumption function for descriptive and predictive purposes. Still, more information is needed regarding both marginal propensities to consume the two commodities and regarding the price elasticities of demand for the two commodities, to reach more refined conclusions. This, incidentally, also exposes the weakness of macroeconomic thinking relative to general equilibrium analysis, since in macroeconomic analysis such issues are assumed away.

It is noteworthy that all of the phenomena described thus far, whether for a small open economy or for a closed economy, arise well inside the PPF as we know it. Hence the redundancy of the concept of the PPF, as immune from the influences of market forces, is evident in our examination of the economic issues involved.

5. Concluding Remarks

If stylized facts dictate, the efficiency-wage hypothesis can be built into the service sector and would apply to skilled workers. The exercise would progress along similar lines, except that in this case the skilled wage rate would become sticky, leading to the unemployment of skilled workers, instead of the unemployment of unskilled workers, as in the model of the current paper. The analysis, however, would trivially be along the same lines, with the only qualification that "unskilled labor" would be replaced by "skilled labor" in terms of the conclusions that we have reached. The actual economy under examination would determine which version of the model is more applicable. Of course, there is nothing wrong in assuming that the efficiency-wage phenomenon arises in both sectors, and with regard to both unskilled workers and skilled workers: for unskilled due to the fact that a higher wage rate provides more nutrition and hence more efficiency, and for the skilled workers because a higher wage rate serves as a disincentive to shirk by raising the opportunity cost of getting fired from the job. Both wage rates would, in such a case, become sticky.

Another extension would be to include an internationally non-traded commodity, in the manner of Batra and Naqvi (1989). This has the advantage of permitting an analysis of real exchange rate variations, since this rate is the ratio of the price of the non-traded commodity to the index of prices of the internationally traded commodities.

Returning to the fundamental reason for our writing this paper, we wish to emphasize, as Egbert and Naqvi (2011) have done in the context of international capital mobility, that a re-evaluation of the concepts and analyses we as economists are prone to undertaking is warranted. In this instance, we suggest a reconsideration with respect to the concept of the PPF, on the so-called supply side of the economy. We are, in fact, attempting to answer the Blinder-Shiller-Stiglitz call for thinking anew the conceptual basis of economic analysis in the 21st Century, post-economic-crisis world.

The fundamental point we wish to emphasize is that extremely valuable conceptual structures have been built by economists in the past 200 years or so, and we are in the enviable position to utilize them, so that only relatively minor, though considerably judicious, decisions need to be made to direct our attention to the economic reality that now faces us, as opposed to the one that faced us in the past. We argue that it is not a case of agency failures. Rather, it is a case of a change in the structure of economies, especially with respect to persistent unemployment,

so that a slight change in the *approach to observation* will actually give us a way to comprehend the structure of this new economic reality that has emerged in the post-economic-crisis world.

Let us heed the call for a change in perspective, while retaining and preserving the enormous intellectual legacy we have been left by our economist forefathers and foremothers. Amendments to theory that Blinder, Shiller and Stiglitz have called for, i.e., to endogenize some variables and relationships that have hitherto been treated as exogenous in the post-21st-Century-crisis economies, are quite sufficient for explaining phenomena of the structurally-different economies of the post-crisis world.

References

[1] Albert, M., and Meckl, J. (1997), 'Sectoral Structure and Unemployment in Open Economies," in Koch, K-J.and Jäger, K. (eds.) *Trade, Growth, and Economic Policy in Open Economies – Essays in Honor of Hans-Jürgen Vosgerau*, Springer, Germany.

[2], "Wage Differentials, Factor Intensities, and Unemployment: A Heckscher-Ohlin Trade Model," *German Economic Review* 2(3), 2001, 287-301.

[3] Batra, R.N., and Naqvi, N. (1989), 'Non-traded and Intermediate Goods and the Theory of Protection', *European Economic Review*, 33(4), 721-735.

[4] Baumol, W.J., and Blinder, A.S. (2009), *Macroeconomics:Economics: Principles and Policy*, 11th edition, Mason, Cengage.

[5] Blinder, A.S. (2010), 'Teaching Macro Principles *after* the Financial Crisis', *Journal of Economic Education* 41(4), 385-390.

[6] Debreu, G. (1959), *Theory of Value: An Axiomatic Analysis of General Equilibrium*, New Haven: YaleUniversity Press.

[7] Egbert, H., and Naqvi, N. (2011), 'Market-dependent Production Set', MPRA Paper No. 33829, 02.October 2011, available at http://mpra.ub.uni-muenchen.de/33829/.

[8] Jones, R.W. (1971), 'A Three-Factor Model in Theory, Trade and History', in J.N. Bhagwati, R.W. Jones, R.A. Mundell, and J. Vanek (eds.), *Trade, Balance of Payments and Growth: Papers in International Economics in Honor of Charles P. Kindleberger*, North-Holland, Amsterdam.

[9] Koopmans, T.C. (1957), *Three Essays on the State of Economic Science*, McGraw-Hill, New York.

[10] Neary, J.P. (1985), 'International Factor Mobility, Minimum Wage Rates, and Factor-Price Equalization: A Synthesis', *Quarterly Journal of Economics* 100(3), 551-570.

[11] Sen, Amartya. "The Welfare Basis of Real Income Comparisons: A Survey," *Journal of Economic Literature*, 17(1), 1- 45.

[12] Shiller, R.J. (2010), "How Should the Financial Crisis Change How We Teach Economics?", *Journal of Economic Education*, 41(4), 403-409.

[13] Solow, R.M. (1979), "Another Possible Source of Wage Stickiness", *Journal of Macroeconomics*, 1(1), 79-82.

[14] Stiglitz, J.E. (2011). "Re-thinking Macroeconomics: What Failed, and How to Repair It", *Journal of the European Economic Association*, 9(4), 591-645.

[15] Summers, L.H. (1988), "Relative Wages, Efficiency Wages, and Keynesian Unemployment", *American Economic Review*, 78(2), 383-388.

Appendix

This appendix contains notes that are material to the issue, but could detract from the coherence of the narrative in the body of the paper.

1. Typically, among other restrictions imposed on this set, it is assumed that the set is (a) non-empty, (b) compact (closed and bounded) and (c) convex.

2. For a classic treatment, see Koopmans (1957) and Debreu (1959).

3. Unlike our objective, the Albert-Meckl (2001, p. 287) stated goal is to demonstrate that "all HO results have close … analogues" in the HO model with the efficiency wage function in each sector dependent on a fixed sector-specific mark-up on their reference wage. The more general formulation in Albert and Meckl (1997) has the additional advantage that it unifies and synthesizes a large body of diverse efficiency wage models, primarily because they show that their model entails a very useful property, which also obtains in their 2001 work, that the general equilibrium outcome maximizes the gross domestic product (GDP) of the country, although under an *additional* linear resource constraint, so that the envelope theorem properties of the constrained-GDP function are preserved.

4. In our formulation, as in Solow (1979), the efficiency of unskilled workers (only in one sector), depends on the *real* wage rate. While it is measured in terms of services in our formulation, nothing would be lost if the real wage rate were measured in units of manufactures.

5. See Baumol and Blinder (2009).

6. See Jones (1971) for the original formulation.

7. These properties have to hold locally for the existence of equilibrium, though not for uniqueness. Uniqueness requires additional properties including that the these conditions hold $\forall w_L \in (o, \infty)$, with $e'(0^+) = \infty$ and $e'(\infty^-) = 0$.

8. The cross-partial derivatives of the two production functions are both positive. Intuitively this means that more capital increases the marginal productivity of unskilled labor in manufacturing, and conversely. Also, more capital employed in the service sector raises the marginal productivity of skilled labor, and conversely.

9. By substituting the right-hand side (RHS) of Eq. (5) for the value of marginal product of labor, this follows from the first order condition of profit maximization with respect to the unskilled wage rate.

10. Solow (1979) writes, "The upshot is: if the wage enters the short-run production function, a cost minimizing firm will leave its wage offer unchanged no matter how its output varies if and only if the wage enters the production function in a labor augmenting way." (p. 81). This stickiness of the wage rate arising from the behavior of firms, rather than its rigidity, is precisely the cause of unemployment emerging as an equilibrium phenomenon despite the willingness of unemployed workers to offer to work for less. Moreover, this *entailed* constancy, instead of *imposed* constancy of the wage rate, can permit the use of the fixed-factor-price constrained GDP function that Neary (1985) proposes for a fruitful unification and synthesis of the literature on international capital mobility and minimum wage rates in general equilibrium.

11. Other endogenous variable values, $\tilde{\tilde{z}}$, generically speaking, are to be similarly denoted in general equilibrium for the closed economy.

12. For the soundness (acceptability) of normative policy prescription, it would also have to be the case that all persons in the country actually have exactly the same income, to which equal weights are attached in social *evaluation*; unless, of course, the value judgment is also advanced that equal weights *should* be attached to unequal personal incomes in social evaluation. Indeed, Sen (1979) writes, "given homothetic preferences identical for all, … the market behavior of a *group* of consumers can be treated as if it were that of one consumer … While this renders distribution of income irrelevant for explaining or predicting market *behavior*, it does not, of course, make distribution irrelevant for *social welfare!* … While for the purpose of studying market behavior that assumption [of relevance of distribution] can be dropped still retaining the aggregation over the consumers, the same clearly does not hold for making social welfare judgments." (pp. 27-28)

Thus, if it *is* the case that homothetic preferences are identical for all persons in society, then this, by itself, fails to constitute a justification as to why equal weights *ought* to be assigned to persons with different incomes, which is the implicit value judgment inherent in the use of a "representative consumer's" *personal* utility function as a *social* welfare evaluation function, simply because "ought" cannot be deduced from "is." On the other hand, in addition to homothetic preferences identical for all persons, if it is also assumed that all persons in society have exactly the same income, in which case equal weighting may be justifiable as a value judgment for social evaluation, then the country becomes indistinguishable from a person, as, for instance, Robinson Crusoe without Man Friday constituting a one-person society. We make no such claim.

Domestic Consumer Market for Indian Tea: A Survey Report

Kakali Hazarika
Department of Humanities and Social Sciences
National Institute of Technology; Silchar
Silchar- 788010; Assam; India
Email: ziasarmah@gmail.com Contact no. +91-9854835154

***Abstract* - A survey conducted by the Indian Tea Board reveals that 89 per cent people take tea as their habit, 8 per cent for refreshing their mind and 3 per cent for appeasing hunger. Tea industry in India is directly responsible for economic and social development in far-flung areas where almost no other economic activity exists. The industry gives a prestigious identity to India as well as Assam in the international platform. Due to weak international and national (auction) prices over last few years, a worldwide tea sector crisis had been observed, which is felt at its worst in India. It is not easy to cope up with the changing situations in the tea market. However, India has an alternative as far as India has a large domestic market, with an increasing trend. In this paper, an attempt to observe the present consumer market in India and behaviour of the urban non poor Indian tea consumers is made.**

Key words: *Auction, consumer, domestic market, price, tea.*

1. Introduction

Tea, the world's most popular and cheapest beverage, also known as 'chai' in many languages has a long history of about 5000 years. Tea has occupied an important place in Indian economy for the last several decades [3]. Tea industry is a source of revenue for the country and it provides employment opportunities to millions of people. However, the industry is passing through a crisis phase since late 1990's. The market for Indian tea is changing day by day. In the international market, Indian tea is losing its position due to high price and poor quality [10]. Increasing competition and new competitors are eating up India's share. However, India has a huge domestic market for tea. Here also tea has tough competition with soft drinks and other beverages like horlicks, bournvita etc. The traditional tea industry was production oriented. Now to meet the competition, a new market orientation is emerging. Now both, market creation and market matching are needed to face the demand of the market [4]. New techniques of product promotion are in demand to meet the competition.

Earlier, India had a higher share in export and a small portion of total tea production was retained for home consumption. At the time of independence, only 79 million kg or about 31 per cent of the production of 255 million kg was retained for home consumption. During that time, India had the highest position in export market. However, in 1997-98, 660 million kg or almost 80 per cent of the production of 838 million kg was domestically consumed, showing an opposite situation [5]. Indian tea is loosing its position in international field and hence it has to manage its sale in the domestic market only. The main reasons of this degradation are emergence of new tea producers, increasing global competition, higher cost of production, global standard of pesticide control, poor quality of tea, defective auction system, etc [8].

2. Indian Tea Consumers

It is very difficult to monitor tea consumption in India. Firstly, it is a huge market and secondly a large number of local players have entered into the market. The absence of accurate estimates had affected the formulation of long-term industry wide action plans. Tea companies have realized the importance of the growing domestic market and changing consumer taste. Hence, packet tea has been gaining importance and the number of tea brands in the markets has been increasing. Another significant change in the tea market is that now manufacturers try to attract consumers with the latest forms in packaging technology.

Tea has a very large universe of consumers. In this study 215 (two hundred fifteen) urban non poor tea consumers belonging to the age group 20 to 60 are randomly selected from Assam; the largest tea producer of India for interview and information are collected about their tea drinking habits, preferred

International Journal of Latest Trends in Finance & Economic Sciences
IJLTFES, E-ISSN: 2047-0916
DOI: 10.009.12/ijltfes.2047.0916

quality of tea, budget on tea, etc. The purchasing behaviour of tea consumers has effected by many factors. These are:

Table 1: The factors effecting consumer behaviour

Sl No.	*Factors*	*Percentage*
1	Price	5 %
2	Brand	10 %
3	Appearance	2 %
4	Taste	45 %
5	Colour	30 %
6	No idea	8 %

Source: Field Survey

Table (Table 1) shows that only 5 per cent consumers consider price as an important factor for purchasing a tea. To find out the impact of price on the preference of consumers, χ^2 (chi square) test has been used.

Table 2: Price of tea and Consumer's preference

Price	*Loose (N=45)*	*Packet; Local (N=34)*	*Packet; National (N=25)*	*Packet without brand (N=15)*
100 - 200	48.8 %	50 %	12 %	40 %
201 - 300	48.8 %	41 %	48 %	46 %
301- above	2.4 %	9 %	40 %	14 %

To find out the relationship between price of tea and consumer's preference to purchage; the null hypothesis is that these two attributes are independent that means price can not influence on purchasing behaviour. (Table 4.62)

The table value of χ^2 (Chi square) for 6 degree of freedom and 5 % level of significance is 12.592 and the calculated value is 20.24.

Therefore it shows that the calculated value of χ^2 (Chi square) is much higher than table value which means that the calculated value cannot be said to have arisen just because of chance. Hence the hypothesis does not hold good or is rejected. Hence price of tea can not change the preference of consumers for a particular tea.

To see the consumer behaviour and price of tea, the present study has collected primary data from the retailer.

Table 3: Tea Price and consumer preference

Type of tea	*Price range of tea (in Rs)*	*Consumer's preference (in %)*
Packet (Without branded)	120 – 200	15
Loose tea	120 – 190	25
Packet (Local branded)	120 – 250	35
Packet (National branded)	270 – 350	25

Source: Field Survey

Table 3 shows that 25 per cent consumer prefers loose tea and another 15 percent prefer packet tea without brand. The price of both the type of tea (non-brand and loose) is lower compare to branded tea. The observation of the retailer was that those consumers who prefer loose tea and non branded tea they belongs to lower income group. On the other hand consumers belonging to higher income group do not give much importance on price but they prefer there brand even at a higher price.

Table 4: level of income and type of tea purchase

Income level	Packet tea (N= 129)	Loose tea (N= 86)
5000 – 10,000	27.9 %	52.3 %
10,001 -20,000	58.1 %	30.2 %
20,001 – 30,000	14 %	17.5 %

To know if there is any co relation between two attributes i.e. monthly income level and type of tea purchase of consumers χ^2 (chi square) test was undertaken. It is assumed that there is no relation between these two parameters.

Calculated value of χ^2 (chi square) is 17 while tabulated value for 2 degree of freedom and 5 % level of significance is 5.991. Hence, it seems that the calculated value is greater than the tabulated value and hypothesis is rejected. So it can conclude that there is some type of relation between monthly income level and type of tea purchase by the consumer.

The table (Table 4.64) shows that consumers of comparatively higher income group (Rs 10,000 to 30,000), prefer packet tea more. The price of packet tea is more than that of loose tea. But the consumers who prefer packet tea are indifferent about the price of the tea. They follow only their choice (taste, colour and brand) to select a particular brand of tea. Consumer survey reveals that consumer has a believe that packet tea is more hygienic than loose tea. Hence, they are indifferent about price of tea.

The survey of consumer behaviour for tea reveals that 60 per cent of the consumers are indifference to price, which signifies that they prefer quality (taste and colour) even at a higher price.

Literature on consumer behaviour also reported that price of a product has very negligible impact on consumers, human beings are creature of habit; hence, habit has a strong role on purchasing a product, rather than its price.

Table 5: Preference for tea drinking

Reason of tea drinking	*Percentage*
Habit	50
Refreshment	30
Without any reason	20

Source: Field survey

In the consumer survey, it was observed that 50 per cent of the consumers take tea as a habit. Habit of consumers refers to preference for a particular type of brand and quality.

Another observation of consumer behaviour over quality is that they prefer the best quality by using various sources of information like opinion of friends (30 per cent), opinion of neighbour (20 per cent) and opinion of work companion (25 per cent) on a particular tea brand.

The primary survey revealed that 70 per cent of the consumers do not know about the type of tea they are consuming (whether it is CTC, orthodox, green tea, Assam tea or Darjeeling tea).consumers understand about quality of tea like taste and colour. Study also observed that the expenditure on tea per household comes less than 1 per cent of their monthly income. It signifies that they give less importance in selecting brand and quality. Consumers are very flexible in deciding a particular brand; they switch over from one brand to another without any reason.

Many factors other than quality influence the mind of the consumer at the time of buying a particular brand of tea. These are

1. *Shop display:*

Attractive shop display can influence on the buying decision of the consumers. Shop display about health benefits of a tea or gift offer with a tea can motivate the consumers.

2. *Advertisement:*

In the primary study it was found that advertisement have very negligible impact (1 per cent) on buying behaviour. However the advertisement, that touches the emotion or given by some renowned person (sports star, film actor etc.) can influence the consumer to change his preference.

3. *Future benefit:*

Consumers are more aware about immediate future. Hence if the tea will gives such promises (reduce heart disease, increase immunity etc.) to protect the future, consumer will purchase it.

4. *Self statement:*

Most of the times people purchase products to make a statement about them.

5. *Influence of other:*

It is a common nature of consumers that they follow other people. Consumers always try to purchase observing other's behaviour and lifestyle.

To observe the importance of brands in consumer's buying behaviour of tea, researcher has collected primary data.

Branding is one of the most important factors influencing an item's success or failure in today's marketplace.

The primary sources of the study have reported that 58 per cent of the consumers prefer branded tea as branding gives them some assurance of quality.

The study observed that all the sample tea consumers are aware about various tea brands available in the market. However, while purchasing a particular tea, consumers prefer performance of that tea. The 35 per cent of the consumers stated that local brands could satisfy their need, while other 25 per cent of the consumer prefers national brands.

Table 5: Awareness of consumers about tea brands

Type of tea	*Percentage of awareness*
National brands	65 %
Local brands	13 %
Both	22 %

From the study, it was found that all the sample consumers are aware about national brands of tea. They can promptly tell the names of national tea brands like Tata, Redlebel, Taz etc. Around 65 per cent of the consumers do not know about local tea brands. The consumers who aware about local tea brands are not the consumers of local brands. In this study only one garden has its own brand name and market share of local consumers.

The study also revealed that the entire branding process of tea is done by some big corporate auction buyers like Tata, Tetley, etc. The tea producer produce tea, but they are not bound to sell this to retail market, so producers prefer auction considering it an effortless process.

Around 60 per cent of the observed gardens do not want to go for branding, but in today's market economy, branding is the only weapon for survival.

Brand awareness is increasing among consumers and consumers are ready to pay for a quality product. However, information was found that all these consumers are not loyal to a particular brand and always curious to search the best product. Consumers frequently change their preferences for a particular tea, influenced by

- Friends

  Friends have great influence on consumer's purchasing behaviour. Consumers are strongly motivated by friends to change his preference.

- Experience
  Experience gathered from daily life influence the consumer to change his preference for a particular tea. For example: if a consumer is offered a cup of tea in a party and he finds the taste to be more good than that of he consumes regularly, definitely he will enquire for it. If his budget allows, he will revise his own preference.
- Advertisement
  Consumers do not blindly follow advertisement; but some influential advertisement, which can touch the emotion of the consumers, will alter their preference.
- Health consciousness
  Increasing health consciousness can influence a consumer to modify his choice for a particular tea.

All the above factors influence the tea consumers to be disloyal for their tea. The observed consumers of the study also stated that they are not getting any particular brand of tea, which will fulfill all the components of their preference (taste, colour, flavour and price).

From retailer's point of view also, it was found that 88 per cent of the retailers prefer packet tea (branded or without brand) to sold. They also found that packet tea is better in quality preservation than loose tea.

All the respondents know about local tea brands. The study found only 35 per cent of the consumer who purchase local brands. Others are not regular consumer of local brands because:

- Local brands are not available in every stall
- According to them, national brands are better than local brands
- They wish that the tea they are consuming would be an international one. This is not possible for local brands.
- Packaging of national brands is more attractive.

On the other hand, the sample consumers (25 per cent) who prefer national brands, have reported that the prices of national brands like Tata, Redlebel, Taz, are comparatively higher than loose tea and the local brands.

Hence, all the observations reported that there is not any single brand of tea available in Assam, bearing promises of quality assurance and value. The available local as well as national brands have very limited influence on consumer's buying behaviour.

3. Findings

From the present research, following observations are drawn about the marketing system followed by tea producers and about the role of ultimate tea consumers in retail tea market:

1) The consumers of tea are indifferent about their quality of tea. They do not know about the type of tea whether it is CTC, orthodox, organic, Assam tea or Darjeeling tea. The tea marketer takes this as a big advantage for market their brand as tea in general without mentioning the type of tea they are selling.
2) Around half (50 per cent) of the consumers take tea as habit and 45 per cent of them prefer taste in tea purchasing. Therefore, they will pay more for the tea having better taste and suit to their habit.
3) The buying behaviour of a tea consumer is a peculiar behaviour, because knowledge of tea consumers on tea is very less.
4) Tea consumers have very flexible buying behaviour. They frequently change their preference for a particular tea.
5) Expenditure of consumer's on tea is very minimum, around 1 per cent of total monthly income of the households. This is one cause of ignorance about tea.
6) Among the various ingredient of tea, consumers prefer taste, colour and flavour.
7) Consumers are always ready to pay more for the best tea. They always prefer the best performance.
8) Though it is said that advertising is the powerful technique of product promotion, the study revealed negligible impact of advertising on tea consumers.
9) Consumers are highly influenced by friends, neighbour and work companion while purchasing a tea.
10) Price is a very important constituent of other product, but in case of tea, price plays minimum role in consumer's buying behaviour while purchasing a tea. In case of tea, consumers give more emphasis on taste, colour and brand.
11) The result shows that brand awareness is increasing among consumers. They are always curious about the innovations and do not hesitate to try a new product.
12) Consumers have knowledge about national tea brands but possess limited knowledge about local brands.
13) It was observed that consumers always discuss about the new things they are experiencing with their family, friends, companions and neighbours. Opinion of friends, neighbour, work companion and

family discussion has very strong impact on consumer's purchasing behavior.

14) When consumers are not satisfied with a particular tea, 90 per cent of the consumers discourage others from buying it.
15) It was readily apparent from the data that the competition is growing tougher and tougher among the various types of tea i.e. loose, packet without brand, packet national brands and packet local brands of tea. All the consumers of the study are willing to buy a local brand of Assam tea. They will pay more for better product. Only colorful packaging and gift offer cannot attract consumers. People always prefer best quality.
16) Chi square test revealed that the type of tea consumption does not depend on the monthly income of the consumers. Consumers are always ready to pay more for the best product.
17) Retailers find it difficult to predict consumer's demand for a particular tea.
18) It was observed that the price of national brands is comparatively higher than the loose tea and local brands. Some varieties of loose tea and local brands have equal price with national brands. But consumers always prefer national brands while choosing from the three having equal price.
19) Not any strong local brand is found which have a large universe of consumers.

4. Summary

The information gathered from the consumers of tea shows that consumers are very much conscious about their taste. They are ready to pay more for better product. Colourful packaging and gift offer can attract very limited number of tea consumers. People always prefer best quality. Their view is that teas of National brands are serving better. They also find local brands better but according to them, Assam tea have not any powerful local brand, which can compete with popular national brands like Tata, Hindustan Lever Limited, Tetley etc. However, consumers are waiting for such a local brand.

5. Market Promotion

Ever since the commercial production of tea had started in India, the tea growers did not give much attention on marketing as they always enjoy a readymade market for their products (auction). In the recent past, due to oversupply of tea against demand the market strategy has shifted from the seller market to the buyer market [10]. In the present market economy consumers are the kings and market decide the products to be produced and sold [4]. Therefore, the grower must recognize the new challenges of the market. Marketing is the key to withstand the present deadlock. . Marketing helps in the development of the standard of products and services and increases the standard of all fields. Though export market of Indian tea is losing its position, the next alternative will be domestic market that has a trend of increase in demand [1].

6. Conclusion

Drinking habits and lifestyles have changed in the last 15-20 years and people are now willing to pay more for quality tea. However increasing consumer attention to the quality of products, growing brand loyalty and active promotions by manufacturers reflect a shift from unbranded to the branded products. Advertising is the best option for tea companies to explain their specialty. To capture the domestic market tea companies or packers can target fairs, festivals, mandis (a local village market), expos etc. for advertisement. Development from grass root level will be beneficial in the long run. So the focus of the tea companies or producers should be the small 'chaiwalas' and the small hoteliers. Trial packs offered at low price and promotions through village fairs and targeting the women folk have helped drive sales in the rural areas. Though consumers are shifting towards packet tea, consumption of loose tea is not very low. India has a large domestic market, which is expanding day by day. The reason for enormous domestic consumption is not due to a high per capita consumption but because of a huge population. Although competition is strong in the domestic market but this market should be the basic platform for the Indian tea growers.

REFERENCES

[1] Baruah, Dr. Prodip. (2008). *The Tea Industry of Assam, Origin and Development*. Guwahat. India: EBH Publisher.

[2] Dahiya, Suraj Bhan. (2002). *Tough time for tea industry ahead*. Chandigarh:The Tribune.

[3] Dutta, A. K. & Das S.C. (1991). *Field management in tea*. Tea culture. India: Tea Research Association, Tocklai.

[4] Pillai & Bhagavati. (1998). *Modern marketing-principles and practices*. New Delhi: S. Chand & Company Ltd.

[5] Sahu, Bishnuprasad. (2004). *Human resource development for industrial workers*. New Delhi; India: Mittal Publication.

[6] Kadavil, S. M. (2008). *Partners in change*. Indian Tea Research Journal.

[7] Sarkar, G.K. (1972). *A study of tea prices in: the world tea economy***.** Calcutta; India: Oxford University Press.

[8] Sen, Benoy B. (2008). *Origin and development of tea*. Guwahati, India: EBH Publisher.

[9] Sengupta, U. (1993). *Indian market*. Tea Culture, Processing and Marketing. New Delhi, India: Oxford & IBH Publications. Co. Pvt. Ltd.

[10] Singh, Narin & Kumar. (2007). *Socio economic and political problems of tea garden workers: a study of Assam*. New Delhi, India: Mittal Publication.

[11] Krishnakumar, P.K. (2008). ET Bureau. The Economic Times. http://www.economictimes.indiatimes.com/articleshow.

[12] Lee, Hope. (2004). *All the tea in India*. http://www.all_the_tea_in_india/Market_Trends_Tea_/0504.html.

[13] Choudhury, Dr. Rabindra Kr. (2006). *India's tea industry and Assam.* http://www.nenanews.com.

International Journal of Latest Trends in Finance & Economic Sciences

Call for Papers

The journal welcomes high quality original research papers, survey papers, review papers, tutorial, technical notes as well as discussion papers.

- International Economics
- Monetary and Fiscal Policy Decision-Making
- Financial Economics
- Financial Analysis
- Financial Risk Analysis
- Corporate Finance
- Forecasting Financial Markets
- International Finance
- Pricing and Risk of Financial Instruments
- Public Finance Decision-Making
- Economic Development
- International Banking and Portfolio Management
- International Political Economy
- Financial Market Regulation
- Transition Economies
- Exchange Rate Modelling
- Economic Policy
- Portfolio and Asset Management
- Advances in Financial Econometrics and Statistics