12

TOWARDS A THEORY OF HETEROGENEOUS MARKETS

At the end of this in-depth review of some of the techniques and models used with high frequency data, there is a clear evidence that foreign exchange and other financial asset price movements for short to medium-term horizons are predictable to some extent. This is substantiated by a positive forecast quality and high real-time trading model returns (e.g. Dacorogna *et al.* (1992), Pictet *et al.* (1992), Gençay *et al.* (1998)). More generally, financial asset returns, whatever asset, depart substantially from the random walk model and are being predicted with some success by market participants.

Where does this sustained predictability originate from? Are the real-time trading models, for instance, successful in capturing the inefficiencies of the FX market? Since this market is widely held to be the most efficient of the financial markets, does this success not conflict with the theory of efficient markets, which precludes the ability to forecast and denies the existence of profitable trading models? Should we conclude from this evidence that markets are inefficient? We believe that we should rather adapt our theory of financial market to the reality of the stylized facts and of very efficient markets.

The motivation of this chapter is to explain why and how markets can be at the same time highly efficient and to some extent predictable. There are a number of reasons for this that are all associated with market dynamics. We want to put in perspective the current theory of efficiency and suggest to move beyond it. This is one of the big challenges ahead in the theory of finance. Many researchers in this field are working on it, being the whole movement of "behavioral finance" around Robert Shiller[1], or the econophysics group and many others who see the need to find ways of moving from a rather static definition to a more dynamic one.

12.1 DEFINITION OF EFFICIENT MARKETS

In conventional economics, markets are assumed to be efficient if all available information is reflected in current market prices (Fama (1970), Fama (1991)). Economists have embarked on weak, semi-strong and strong-form efficiency tests. The weak-form tests investigate whether market prices actually reflect all available information. The semi-strong tests are based on so-called event studies, where the degree of market reaction to "news announcements" is analyzed. The strong-form tests, finally, analyze whether specific investors or groups have private information from which to take advantage. By and large, most studies conclude that the major financial markets are efficient and that all information is reflected in current prices. However, the conclusions of such studies have been bogged down by methodological questions; in particular, whether any observed departures from market efficiency are due to any genuine market inefficiency or whether a deficiency of the market pricing model is being used as a yardstick to compare actual with theoretical prices.

The inference, that in an efficient market no excess return can be generated with trading models, is based on the assumption that all investors act according to the rational expectation model (Shiller (1989), Fama (1970)). If this assumption is wrong, the conclusion that forecasting is impossible is also questionable. The assumption of rational expectations has been called into question in various platforms and the idea of heterogeneous expectations has become of increasing interest to specialists. Shiller (1989), for example, argues that most participants in the stock market are not "smart investors" (following the rational expectation model) but rather follow trends and fashions. The modelling of "noise trader" has become a central subject of research in market microstructure models. On the FX market, there is much investigation of "speculative bubbles" and the influence of technical analysis on the dealer's strategy (see, for example, Frankel and Froot (1990)). Some attention has also been caught by the possibility of time varying expectations, which is closer to our view of the

[1] See, for instance, Shiller (2000) where the author claims that the market agents are essentially acting irrationally.

market, (Bekaert and Hodrick (1992)). Variation over time in expected returns poses a challenge for asset pricing theory because it requires an explicit dynamic theory in contrast to the traditional static capital asset pricing model (CAPM).

In summary, the conclusion that financial asset prices are not predictable is based on three assumptions: market prices reflect all the information available, news and events that hit the market are Gaussianly distributed and the market is composed of homogeneous agents. The two first assumptions are very reasonable starting points for the definition. It is really the third one that poses a problem. It is clear that all market agents have in fact bounded rationality. They cannot be omniscient and do not all enjoy the same freedom of action or access to the markets. Recent works by Kurz (1994) or Gouree and Hommes (2000) present new theoretical models to tackle this problem. Introducing the heterogeneity of agents can give rise to very interesting nonlinear effects in the models. They show that many of the price fluctuations can be explained by endogenous effects. Similar conclusions are reached by Farmer and Lo (1999) in their discussion of market efficiency. They base their analysis on a comparison with the evolution of ecological systems. Farmer (1998) develops a market model inspired by ecological systems that contains agents with various trading strategies.

12.2 DYNAMIC MARKETS AND RELATIVISTIC EFFECTS

We just saw that conventional economics makes its inferences on efficient markets on the basis of a model in which economic agents are entities that act according to the rational expectation strategy. Any differences in planning horizons, frequency of trading or institutional constraints are neglected. Besides, there is substantial empirical evidence that investors have heterogeneous expectations, Müller *et al.* (1993a), Müller *et al.* (1997). Surveys on the forecasts of participants in the FX market reflect the wide dispersion of expectation at any point in time. The huge volume of FX trading[2] is another indication reinforcing this idea since it takes differences in expectation among market participants to explain why they trade. In Chapter 7, we presented the heterogeneous market hypothesis; at the end of this book the need for such a view becomes clear. It is the most elegant way to reconcile market efficiency with the stylized facts. Recently, Lux and Marchesi (1999) have developed simulation models of financial markets that include agents with different strategies (fundamentalists and chartists). They were able to show (Lux and Marchesi (2000)) that this model can reproduce most of the empirical regularities (fat tails, long mem-

[2] Over 1500 billion US$ is traded every day in the different centers like Tokyo, London and New York according to the 3-yearly survey of the Bank of International Settlements Bank for International Settlements (1999).

ory and scaling law) even though they use in their simulations Gaussianly distributed news.

There are many ways to describe these heterogeneous expectations. We believe that the most promising approach is to differentiate the expectations according to their time dimension because we consider the *different time scales* of the market participants the key characteristic of the market. Some trade short-term, others have long-term horizons with market makers at the short-term end of the scale and central banks at the long-term end. Contrary to the usual assumption, there is no privileged time scale in the market. The interaction of components with different time scales gives rise to characteristically relativistic effects[3] such as certain properties of volatility clusters, trend persistence, lag between interest rate adjustment and FX rate adjustment. The latter is a good example of what conventional theory considers an inefficiency while we see it as an effect arising from the different time scales involved in the market. To take advantage of the lag in adjustment between interest rate and exchange rate moves, an investor needs to tie up his money for months or even years. This is a very long time for a FX trader. Some investors will thus tend to ignore these profit opportunities while others invest in them, as is testified by the development of managed currency funds based on this property. The combination of all of these effects ultimately enables the construction of successful forecasting and trading models.

In long time intervals, market price changes are "flatter" and have fewer relevant movements (trend changes) than in short-term intervals. The higher the resolution and the smaller the intervals, the larger the number of relevant price movements. The long and the short-term traders thus have different trading opportunities: the shorter the trading horizon, the greater the opportunity set. A market participant's response to outside events should always be viewed as relative to their intrinsic opportunity set. A short-term trader does not react in the same way as a long-term trader. Economic decision makers, such as traders, treasurers and central bankers, interpret the same information differently. The variation in perspective has the effect that specific price movements cannot lead to a uniform reaction; rather, they result in individual reactions of different components. In turn, these reactions give rise to secondary reactions, with the different components reacting to their respective initial response. Watching the intraday price movements, one clearly sees the sequences of secondary reactions triggered by the initial events. (See, for example, Goodhart and Curcio (1991), Almeida *et al.* (1998) on news effect on the FX market or Franke and Hess (1997) on the Deutsche Termin Börse). The existence of

[3] We use here the term "relativistic" to express the dynamic interaction between different market components relative to each other rather than relative to the news that has impacted the market. These effects are sometimes called endogenous effects in the literature, Kurz (1994).

different trading strategies in the market was also put forward in Chapter 7 to explain the HARCH effect of asymmetry in the information flow at different frequencies. LeBaron (2000) shows that introducing agents with different time horizons in his market model gives rise to heteroskedasticity effects in the resulting price volatility.

The delay with which the secondary reactions unfold is called the *relaxation time*. If diverse components with different time scales interact in the market, there is typically a mixture of long and short relaxation times following the impact of outside events. If different relaxation times are combined, the resulting autocorrelation decays hyperbolically. This is a natural explanation of the long memory effects detected in financial markets. Dacorogna *et al.* (1993)'s study of the autocorrelation function for short-term absolute price changes confirmed the hyperbolic decay and revealed that volatility clusters tend to have a longer memory than assumed by other studies of the subject. We saw in Chapter 7 that there are many studies confirming this effect.

There is yet another phenomenon which originates from the fact that financial markets are spread worldwide. Economic and political news and trading activity are not stationary. They have a clear-cut pattern of moving around the world in a 24-hour cycle. The price data of foreign exchange rates reflects this in terms of a 24-hour seasonality in market volatility, Müller *et al.* (1990). This seasonality can be accounted for by introducing a *business time scale* such as in Dacorogna *et al.* (1993). The 24-hour cycle implies that market reactions to an event cannot be simultaneous and that there are distinct relaxation times following the event. Geographical components related to the business hours of the different trading centers must be added to the time components. This is generally called the "heat wave" effect as proposed in Engle *et al.* (1990).

12.3 IMPACT OF THE NEW TECHNOLOGY

The realization that there is value in the data to define an investment strategy has brought to life many new firms that specialize in modeling financial markets and in providing trading advices on the basis of technical models. The question is, of course, will the impact of the new technology be a passing phenomenon or will it have a long-term effect? As the relativistic phenomenon arises from the interaction of components with different time scales, it will remain appropriate as long as heterogeneous expectations continue to exist in the market. The interaction process may become more complex, but it cannot disappear.

News technologies enable users to identify additional trading opportunities to increase their profits. This quickens their pace of trading and contributes to higher market volume and liquidity. The improved liquidity lowers the spreads between bid and ask prices. Lower spreads reduce transaction costs, which in turn increase the opportunity horizon for profitable trading. The new technology introduces a shift in perspective, with components starting to focus on increasingly more numerous short-term time intervals.

As components become increasingly short-term in their focus, the spectrum of short-term components increases. This has the effect that relative differences among components become more significant and the relativistic effects more pronounced. Contrary to accepted notions which assume that sufficient buying power can "trade away" any phenomenon, the increased buying power will have the overall effect of enhancing the relativistic effects. Thus the very basis of our ability to forecast and build profitable trading models will be enhanced. This statement must be qualified in the sense that the reaction patterns will become increasingly diversified, and therefore, more complex and the speed of adjustment will increase requiring more and more sophisticated models.

12.4 ZERO-SUM GAME OR PERPETUUM MOBILE?

Conventional thought has it that financial markets must be a zero-sum game. This is true if we take a static view. In reality, the financial markets are dynamic and they are highly complex.

Markets are a platform for components to take advantage of the diversity of interests. They are able to match their opposing objectives when one component buys and another component sells. The lower the friction, the easier a counterpart for a particular transaction is found and the larger, therefore, is the particular component's opportunity set. By being able to go ahead with a particular transaction, the flexibility of the respective components is increased and their profit potential improved.

The new technology fosters the ability of the market to provide an environment for the generation of wealth. As explained, interaction within the market gives rise to relativistic effects and relaxation times. To the extent that these relativistic effects are understood and incorporated into forecasting and trading model technology, market participants have the opportunity to generate additional profit or limit their losses. In our terminology, the profit which is generated is energy extracted from the market. Improved efficacy of component interaction generates additional energy and reduces the friction associated with buying and selling within the market. The process may be compared to the search for more efficient engines in the automobile industry where everybody gains from it in the long term.

Have we achieved a *perpetuum mobile*? The answer is clearly no. Like any other technological innovation, the new technology does not generate energy from nothing, but it does take advantage of the energy potential existing in the financial markets. By offering a service to the economic agents, financial markets are not closed systems but do get a permanent input of money. This makes them highly open systems in terms of energy. Besides, a lot of resources have been put into the new technology in the form of extensive research, development work and hardware to treat the information. There are numerous studies that have shown that simple trading rules do not work in efficient markets. Only elaborated treatment of the data allows the identification of profitable trading rules. This treatment is not free, it has a price. Moreover, as the relativistic reaction patterns become increasingly diversified, research and development efforts will have to increase in the future to keep up with the ever changing nonlinear patterns.

12.5 DISCUSSION OF THE CONVENTIONAL DEFINITION

As the markets consist of a diversity of components, there are different relaxation times, occurring because of the underlying relativistic effects between different components. It follows that the weak form of efficiency coupled with the rational expectation model cannot be attained. Because of the presence of different time components with heterogeneous expectations, current market prices cannot reflect all available information. The price discovery mechanism follows rather a dynamic "error correction model" where the successive reactions to an event unfold in the price. Why, then, did this not show up more clearly in previous scientific investigations? Some of the several reasons include:

- High-frequency data is a prerequisite for the empirical investigation of relativistic phenomena.
- Extensive computing power is needed to show the predictability in financial markets. Access to reasonably priced computing power has become available only recently.
- It is in the last few decades that an increasing awareness for dynamic and nonlinear processes has been gained. Such an awareness is crucial for the study of relativistic effects.

The presumption of conventional economics that forecasting is impossible per definition has had a powerful impact on the research on market efficiency. Economists have focused on structural studies which were hamstrung by a lack of high-frequency data and theoretical shortcomings. Little academic research has been invested in actually trying to predict shorter-term price movements and build successful trading models.

12.6 AN IMPROVED DEFINITION OF "EFFICIENT MARKETS"

Although the current definition of efficient markets has shortcomings, we do not think that this concept should be abandoned; rather, it should be adapted to the new findings. It is important to find a good measure of how well a market operates.

From a dynamic perspective, the notion of reduced friction should be central to the notion of efficiency. We consider an efficient market to be a market where all market information must be available to the decision makers and there must be participants with different time scales and heterogeneous expectations trading with each other to ensure a minimum of friction in the transaction costs.

A quantitative measure of efficiency might be derived from the bid-ask spreads (those between real bid and ask prices being more appropriate for such a measure than the nominal spreads quoted in information systems). Spreads are not only a measure of "friction", they also contain a risk component. The volatility or, more precisely, the probability of extreme price changes within short time intervals should be considered together with the spread in the quantitative measure of market efficiency to be proposed. We are sure that in the years to come this definition will prevail and we shall find precise measures of efficiency as it is the case in thermodynamics and engineering.