


Article

A New Measure of Market Inefficiency

Christopher R. Stephens ^{1,2} , Harald A. Benink ^{3,4,*}, José Luís Gordillo ¹ and Juan Pablo Pardo-Guerra ⁵

¹ C3-Centro de Ciencias de la Complejidad, Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Ciudad de México 04510, Mexico; stephens@nucleares.unam.mx (C.R.S.); jlgr@super.unam.mx (J.L.G.)

² Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Ciudad de México 04510, Mexico

³ School of Economics and Management, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

⁴ Financial Markets Group, London School of Economics, Houghton Street, London WC2A 2AE, UK

⁵ Department of Sociology, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA; jpardoguerra@ucsd.edu

* Correspondence: H.A.Benink@tilburguniversity.edu

Abstract: Financial crises, such as the Great Financial Crisis of 2007–2009 and the COVID-19 Crisis of 2020–2021, lead to high volatility in financial markets and highlight the importance of the debate on the Efficient Markets Hypothesis, a corollary of which is that in an efficient market it should not be possible to systematically make excess returns. In this paper, we discuss a new empirical measure—Excess Trading Returns—that distinguishes between market and trading returns and that can be used to measure inefficiency. We define an Inefficiency Matrix that can provide a complete, empirical characterization of the inefficiencies inherent in a market. We illustrate its use in the context of empirical data from a pair of model markets, where information asymmetries can be clearly understood, and discuss the challenges of applying it to market data from commercial exchanges.

Keywords: efficient market hypothesis (EMH); excess trading returns; investor's behaviour; behavioural finance



Citation: Stephens, Christopher R., Harald A. Benink, José Luís Gordillo, and Juan Pablo Pardo-Guerra. 2021. A New Measure of Market Inefficiency. *Journal of Risk and Financial Management* 14: 263. <https://doi.org/10.3390/jrfm14060263>

Academic Editor: Victor Dragotă and Shigeyuki Hamori

Received: 6 April 2021

Accepted: 7 June 2021

Published: 10 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Financial crises, such as the Great Financial Crisis of 2007–2009 and the COVID-19 Crisis of 2020–2021, lead to high volatility in financial markets and highlight the importance of the debate on the Efficient Markets Hypothesis, a corollary of which is that in an efficient market it should not be possible to systematically make excess returns. Because of this, during the past few decades, it has become ever more popular to consider financial markets from other than a neoclassical rational expectations point of view. The latter considers financial markets to be in continuous equilibrium with informationally efficient prices. However, empiricists have questioned the validity of this model, pointing to evidence of inefficiencies and anomalies (for recent reviews of the literature on efficiency and anomalies see [Titan 2015](#); [Woo et al. 2020](#); [Jacobs and Müller 2020](#)).

[Pesaran \(1989\)](#) notes that the idea of a rational expectations equilibrium (REE) involves much more than the familiar concept of the equilibrium of demand and supply. A REE can be characterized by three main features: (1) all markets clear at equilibrium prices, (2) every agent knows the relationship between equilibrium prices and private information of all other agents, and (3) the information contained in equilibrium prices is fully exploited by all agents in making inferences about the private information of others. Thus, in a REE prices perform a dual role—apart from clearing the markets they also reveal to every agent the private information of all other agents. In effect, the concept of the REE requires that everybody knows (in a probabilistic sense) everything about the way the market economy functions. However, as [Von Hayek \(1937\)](#) puts it: “The statement that, if people know everything, they are in equilibrium is true simply because that is how we define

equilibrium. The assumption of a perfect market in that sense is just another way of saying that equilibrium exists, but does not get us any nearer an explanation of when and how such a state will come about. It is clear that if we want to make the assertion that under certain conditions people will approach that state we must explain by what process they will acquire the necessary knowledge.”¹

The preceding implies that, for the REE to have any operational meaning, it is necessary that the processes by means of which people learn from experience and acquire the common knowledge necessary for the achievement of the REE, are specified fully and explicitly.² In this sense, one might expect that as stock-market trading is increasingly dominated by sophisticated professionals, as opposed to individual investors, that markets are becoming more efficient. As pointed out by [Stein \(2009\)](#), however, crowding and leverage are two factors which limit this tendency to efficiency and therefore to an associated REE. Furthermore, [Rösch et al. \(2017\)](#) observe that the mainstream finance literature does not allow for market efficiency to vary through time. Market efficiency is governed by arbitrage activity and market-making capacity, both of which facilitate price convergence on efficient market benchmarks. In turn, the efficacy of arbitrage and market making is influenced by financial frictions (such as limited capital, transaction costs, short-sale constraints, and idiosyncratic volatility) whose severity varies considerably over time.

The Efficient Markets Hypothesis (EMH) has been traditionally linked to the idea that security prices fully reflect the available information. As [Shleifer \(2000\)](#) points out, this implies that when news about the value of a security hits the market, its price should react and incorporate this news both quickly and correctly. The “quickly” part means that those who receive the news late should not be able to profit from this information, while “correctly” means that the price adjustment in response to the news should be accurate on average: i.e., that prices should neither under react nor over react to particular news announcements. [Shiller \(2003\)](#) notes that we have to distance ourselves from the presumption that financial markets always work well and that price changes always reflect genuine information. According to Shiller, evidence from behavioral finance helps us to understand, for example, that the worldwide stock market boom in the late 1990s, and then crash after 2000, had its origins in human foibles and arbitrary feedback relations and must have generated a real and substantial misallocation of resources.

Even more fundamentally, the confirmation or negation of the EMH has been highly controversial, due to the existence of the joint-hypothesis problem, formulated by Fama in their seminal 1970 and 1991 [Fama \(1970, 1991\)](#) overview papers on efficient capital markets, wherein efficiency is determined only within the context of a particular asset-pricing model.

A common corollary of the EMH, that by some has been taken to be its definition, is that in an efficient market it is impossible to make excess, above average returns in a systematic fashion without accepting above average risk (for an overview see, for example [Malkiel 2003](#)), or that excess returns are unpredictable. However, [Pesaran and Timmermann \(1995\)](#) note that predictability of excess returns does not imply stock market inefficiency, and can be interpreted only in conjunction with, and in relation to, an intertemporal equilibrium model of the economy. Inevitably, all theoretical attempts at interpretation of excess return predictability will be model-dependent, and hence inconclusive. [Fama \(1991\)](#) states that it is only possible to test whether information is properly reflected in prices in the context of a pricing model that defines the meaning of “properly”. As a result, when anomalous evidence is found on the behavior of returns, the way it should be split between market inefficiency or a bad model of market equilibrium is ambiguous. Furthermore, as [Balvers et al. \(1990\)](#) have pointed out, it is possible to formulate an equilibrium model that leads to predictable returns.

As Fama [Fama \(1991\)](#) argues, it is a disappointing fact that, because of the joint-hypothesis problem, precise inferences about the degree of market efficiency are likely to remain impossible. However, a way to avoid the pitfalls of the joint-hypothesis problem is to take a completely empirical approach; defining inefficiency with respect to some measure that is not dependent on the existence of some underlying model, such as an asset-pricing

model. Thus, one evaluates the economic significance of stock market predictability by seeing if the associated information could have been exploited successfully in investment strategies, thus leading to systematic excess returns. Of course, this begs the question of how do we define systematic and excess? Excess relative to what? In the literature, it is common to measure excess relative to some ‘fixed’ benchmark, such as the risk-free interest rate, or to an index portfolio (the logical extreme of that being the market portfolio). One of the chief drawbacks of such measures is that they permit the possibility that uninformed traders acquire excess profits, even when the market is efficient. For instance, noise traders who, on average, do not change their portfolio holdings, may have excess profits relative to a fixed benchmark simply because, by chance, the price rose. Thus, as emphasized by [Bagehot \(1971\)](#), it is important to distinguish between trading gains and market gains.

The contribution of this paper is threefold. First, in order to eliminate the defects of a fixed benchmark, we propose as benchmark in Section 2 a “moving target”, where excess profit during timestep t is related to the increase in the market value of an active trading portfolio in timestep t , relative to the increase in the market value of a buy and hold portfolio in the same timestep. In this way an excess profit or loss for a given trader over the timestep t can only arise when there has been a net change in the trader’s portfolio holdings in the asset and a net change in the asset’s price. This choice of benchmark always refers the market dynamics to a zero sum game. Second, in Section 3 we develop a new measure for the relative inefficiency between two trading strategies, or trader groups, evolving from time t to time t' . The relative inefficiency will be defined as the excess returns of trading strategy i relative to trading strategy j , divided by the relevant standard error. Third, in Section 4 we propose an Inefficiency Matrix, which aims at giving a complete description of the relative inefficiencies that exist in an entire market, thus enabling us to determine the degree of market (in)efficiency. In Section 5, we illustrate the use of the Inefficiency Matrix as a diagnostic by applying it to a pair of model markets—an agent-based market and an experimental market. Section 6 contains a discussion of the application of the Inefficiency Matrix to commercial exchanges, such as stock markets, while Section 7 concludes the paper.

2. When Are Returns “Excessive” or “Abnormal”?

Canonically, “excess” returns for securities have been defined with respect to exogenous benchmarks such as a risk free interest rate, while “abnormal” returns have been defined with respect to a benchmark portfolio—the logical extreme being the market portfolio itself. Both have defects. The main defect of defining excess via the risk free interest rate is that it does not distinguish trading returns from market returns, i.e., what was earned by a general increase in the market versus a particular active trading strategy that tries to beat the market. On the other hand, judging a security’s return as being “abnormal” requires that it be benchmarked against an asset pricing model [Schwert \(2003\)](#), such as the Capital Asset Pricing Model [Sharpe \(1964\)](#), or some other empirical model, such as Fama–French three or four factor models [Fama and French \(1996\)](#) in order to determine what is “normal”. The latter type of empirical approach usually takes a firm-specific model for specifying what is considered to be a normal return.

Although in accounting for short horizon abnormal returns, where returns are close to zero anyway, one might expect the “bad model” problem, wherein the security is mispriced, to be relatively unimportant. This is certainly not the case for longer horizon returns [Fama \(1998\)](#). This is related to another problem: how often the benchmark should be updated. This is usually tied to the period that one is using for calculating excess or abnormal returns—monthly, yearly etc., especially in the case of abnormal returns defined with respect to an asset pricing model, where the standard periods tend to be short—a month or less.

Further, in the case of a risk free interest rate benchmark, it is possible for traders to make excess profits simply by chance. To understand this one must remember that a market is unique, in that although it is a realization of an underlying stochastic process, one

observes only one particular realization of that process. It may be that a realization over an interval $[t, t']$ was such that the asset price $p(t) > p(t')$, whereupon one would find that many agents were apparently making “excess” profits relative to the risk free benchmark. Just how much profit depends on the “luck of the draw”, i.e., how atypical the particular observed realization was relative to its expected behavior. This problem is exacerbated if one thinks of noise traders with heterogeneous trading horizons, wherein an agent with a long horizon, who randomly makes one single trading decision to buy in a market that goes up by random chance, would be seen to have a potentially large profit relative to the risk free benchmark. Finally, in commercial markets, there is a general increase in asset prices, such as stocks, that is traditionally viewed as being compensation for the extra risk taken on. However, traditional pricing models have not been able to properly account for this increase, leading to the equity premium puzzle (see, for example, [Kocherlakota 1996](#); [Mehra and Prescott 1985](#).) The implication of this is that one can earn excess returns relative to the risk free interest rate just by doing “nothing”.

So, the question is: How does one arrive at a measure that can completely distinguish between market gains and trading gains and that avoids the pitfalls of the joint hypothesis problem? Essentially, we are led to ask: What are the “excess” returns associated with a particular trade, or sequence of trades in a market, relative to what one would have earned just by doing nothing, i.e., by not trading? We will denote this as the Excess Trading Return (ETR) [Benink et al. \(2004\)](#).³ The magnitude of the returns one would have from the market by maintaining a particular set of portfolio holdings obviously depends on the precise values of those holdings. Hence, in this case, any relevant benchmark will be both specific to a particular agent and potentially dynamic. With these thoughts in mind one may define the ETR associated with a particular trade by evaluating the return on the active trading portfolio that included that trade against the return on the passive Buy and Hold portfolio wherein the trade was not executed.

For an agent i , using a trading strategy α , we denote the agent’s portfolio value at time t by $V_{i\alpha}(t)$. Hence, the ETR associated with a single trade, n , executed at time t_n and made in the time period $[t, t']$, where $t' \leq t_n \leq t$, by an active trading portfolio using a trading strategy, α , relative to the portfolio that uses a Buy and Hold strategy over that time interval, is

$$e_{i\alpha}(t, t', t_n) = (\delta V_{i\alpha}(t, t') - \delta V_{iBH}(t, t')) \quad (1)$$

where $\delta V_{i\alpha}(t, t') = V_{i\alpha}(t) - V_{i\alpha}(t')$ is the increase in portfolio value between t' and t for agent i using trading strategy α , while $\delta V_{iBH}(t, t') = V_{iBH}(t) - V_{iBH}(t')$ is the same quantity for the Buy and Hold portfolio. To go further one needs to specify the portfolio in terms of the agent’s assets. Thus, we define

$$V_{i\alpha}(t) = \sum_{k=1}^K V_{ik\alpha}(t) = \sum_{k=1}^K n_{ik\alpha}(t) v_k(t) \quad (2)$$

where $V_{ik\alpha}$ is the portfolio value associated with asset k and we are considering a universe of K assets. $n_{ik\alpha}(t)$ is the holding that an agent i using a strategy α would have in the asset k at time t , and $v_k(t)$ is the value of one unit of the asset. For example, for stocks, ignoring dividends, $v_k = p_k$, the price of the k th asset. Note that, in principle, any asset type may be included in this definition, not only stocks. One obtains then for the ETR

$$\delta V_{i\alpha}(t, t') = \sum_{k=1}^K (n_{ik\alpha}(t) v_k(t) - n_{ik\alpha}(t') v_k(t')) \quad (3)$$

$$\delta V_{iBH}(t, t') = \sum_{k=1}^K (n_{ik}(t') v_k(t) - n_{ik}(t') v_k(t')) \quad (4)$$

and hence

$$e_{i\alpha}(t, t', t_n) = \sum_{k=1}^K \Delta n_{ik\alpha}(t, t_n) v_k(t) \quad (5)$$

where $\Delta n_{ik\alpha}(t, t_n) = (n_{ik\alpha}(t) - n_{ik\alpha}(t'))$ is the change in holdings of asset k at t_n . If one denotes the total number of agents in the market by N_a then one has the trivial constraint $\sum_{i=1}^{N_a} e_{i\alpha}(t, t', t_n) = 0$. Now, it is important to remember that (5) is also a constrained sum as one may only transfer resources between assets. Thus, one must impose the constraint $\sum_{k=1}^K \Delta n_{ik\alpha}(t, t_n) v_k(t_n) = 0$, i.e., that a trade⁴ is just a transfer of resources between assets, so that total portfolio value for a given agent must be the same immediately before and after the trade in the absence of trading costs. Taking the trade under consideration to be a transfer between assets k_1 and k_2 then $\Delta n_{ik_1\alpha}(t, t_n) v_{k_1}(t_n) = -\Delta n_{ik_2\alpha}(t, t_n) v_{k_2}(t_n)$. Thus,

$$e_{i\alpha}(t, t', t_n) = \Delta n_{ik_1\alpha}(t, t_n) \Delta v_{k_1}(t, t_n) \quad (6)$$

where $\Delta n_{ik_1\alpha}(t, t_n)$ is the change in holdings of asset k_1 due to the trade, and $\Delta v_{k_1}(t, t_n) = (v_{k_1}(t) - (v_{k_1}(t_n) v_{k_2}(t) / v_{k_2}(t_n)))$ is the change in the relative value of the asset from the trade t_n to t . Note that this relative asset value change is the difference between the value of asset k_1 at time t and its value at t_n multiplied by the relative increase in the value of k_2 . Hence, for $\Delta n_{ik\alpha}(t, t_n) > 0$, for example, if the value of asset k_2 does not increase proportionately more than that of asset k_1 , then the ETR is negative, i.e., it is only worth trading if the net increase in portfolio value due to the trade is more than that associated with doing nothing. If the holding of asset k_1 is exchanged for an asset, k_2 , whose value does not change, such as cash, then $v_{k_2}(t) = v_{k_2}(t_n)$ and, hence, $\Delta v_{k_1}(t, t_n) = (v_{k_1}(t) - v_{k_1}(t_n))$, i.e., the pure change in the value of the asset k_1 . If the cash is invested at a continuously compounded interest rate r , then $v_{k_2}(t) = (1 + r)^{(t-t')} v_{k_2}(t_n)$.

Thus, we see that in order to generate an ETR for a given trade one requires both a change in relative asset value and a change in asset holding. This is the chief distinction with any measure of excess returns defined via static, global benchmarks, where returns can be made in the absence of changes in portfolio holdings. The key difference here is that the ETR tests directly the profitability of a particular trade, i.e., what was the return on that particular trade versus maintaining one's portfolio constant. This is a crucial point, as it means that, by construction, as desired, the ETR distinguishes completely between market gains and trading gains by explicitly removing the market return that would have accrued if the trade had not taken place.

It is clear that the net ETR across the market associated with this trade is zero, as for the agent i , changing holdings by $n_{ik_1\alpha} \rightarrow n_{ik_1\alpha} + \Delta n_{ik_1\alpha}$, $n_{ik_2\alpha} \rightarrow n_{ik_2\alpha} + \Delta n_{ik_2\alpha}$ there is a counterparty, j , to the trade, using a strategy β , whose holdings change by $n_{jk_1\beta} \rightarrow n_{jk_1\beta} + \Delta n_{jk_1\beta}$ and $n_{jk_2\beta} \rightarrow n_{jk_2\beta} + \Delta n_{jk_2\beta}$, such that $\Delta n_{ik_1\alpha} + \Delta n_{jk_1\beta} = 0$ and $\Delta n_{ik_2\alpha} + \Delta n_{jk_2\beta} = 0$. Thus, the net ETR for the entire market due to the trade is zero. Trading in this sense, unlike the case of the standard benchmarks, is a zero-sum game.

To proceed further, it is necessary to decide how to determine the total ETR for an agent associated with more than a single trade. One could, of course, merely sum $e_{i\alpha}(t, t', t_n)$ over all the agent i 's trades, n , that take place at times t_n such that $t' \leq t_n \leq t$, the total number of trades being taken to be $N_{i\alpha}$, which depends on the trading frequency of the agent. Thus, with this definition, for an agent i , using a trading strategy α , the total ETR is

$$Q_{i\alpha}(t, t') = \sum_{n=1}^{N_{i\alpha}} e_{i\alpha}(t, t', t_n) \quad (7)$$

Although an adequate definition for a single trader, it has the defect that when summed across the market over multiple traders the net ETR is non-zero. This is because it is being applied to agents with potentially very different trading horizons.

Requiring a measure that corresponds to a zero-sum game when taken across multiple traders, we may consider two different measures of ETR. For the first, we consider the ETR associated with looking ahead to the next trade in the market Benink et al. (2004, 2010), i.e., the next “tick”.⁵ In this case, the ETR is

$$R_{i\alpha}(t, t') = \sum_{n=1}^N e_{i\alpha}(t, t', t_n) \quad (8)$$

where N is the total number of trades in the interval and, now

$$\begin{aligned} e_{i\alpha}(t_n, t_{n-1}) &= (\delta V_{i\alpha}(t_n, t_{n-1}) - \delta V_{iBH}(t_n, t_{n-1})) \\ &= \Delta n_{ik_1\alpha}(t_n, t_{n-1}) \Delta v_{k_1}(t_n, t_{n-1}) \end{aligned} \quad (9)$$

where we have assumed that the trade is between assets k_1 and k_2 . In distinction to (7), in (8) the sum is over all trades in the interval irrespective of whether or not trader i was a party in the trade or not. This definition is natural for markets where trading takes place at fixed, regular intervals, such as a call market, where the above refers to returns from the t_{n-1} th auction to the t_n th one. Behind this definition of ETR is the notion of “perfect foresight”, that an agent is judged over what was done at every single trade that takes place. To illustrate this, if we consider a market with one risky asset and cash, with fixed trade size V per trade, say, then the maximum ETR possible over a time interval, t' to t is given by

$$R_{i\alpha}(t, t') = V \sum_{n=1}^N |\Delta v(t_n, t_{n-1})| \quad (10)$$

An advantage of having excess profits measured at such high frequency is that statistical inference is enhanced due to the greater sampling. For example, one would expect to be able to better judge the utility of an agent’s proprietary strategy if it has been used to make 1000 trades as opposed to 10. Of course, this presumes that relevant agents trade frequently. In principle, this definition of ETR could be implemented using real, high-frequency trading data, where it could be used to evaluate the ability of traders to exploit short-term profit making opportunities. In the case where trading is typically over longer multi-period time horizons, it would not be sensible to use this criteria—for instance, to judge between different fund managers. For example, a manager might increase their portfolio holding in a given stock to 2.6% from 2.1% in January as a result of expecting an appreciation in the price of the stock over a six month interval. It would be unfair to judge this investment over the first tick after the purchase!

Of course, one can readily generalize (8) to the case where t_n and t_{n-1} refer not to trades, i.e., ticks but, rather that $(t_n - t_{n-1})$ refers to a fixed time interval, τ . In the case that trading is carried out only at regular intervals this reduces to (8) above. When trading occurs continuously, then a fixed time interval would be appropriate for evaluating the performance of traders with trading horizons not too dissimilar to τ . So, just as it is inappropriate to consider a fund manager’s performance over a short time interval, so it would be inappropriate to consider the performance of a day trader using a time horizon of a year, i.e., to consider the net change in the portfolio weighting over the year weighted by the net change in price over the year.

In the above, changes in portfolio holdings associated with a trade were always referred to a dynamic reference point—the actual holdings of the agent before the trade. Keeping to the idea that ETR should measure the returns for a particular agent of executing a sequence of trades relative to not having executed those trades, we may also define the ETR associated with a trade at t_n between assets k_1 and k_2 as

$$e_{i\alpha}(t, t', t_n) = \Delta n_{ik_1\alpha}(t, t') \Delta v_{k_1}(t, t_n) \quad (11)$$

Superficially, this expression looks the same as (6). However, the crucial difference here is that $\Delta n_{ik_1\alpha}(t, t') = (n_{ik_1\alpha}(t) - \dots - n_{ik_1\alpha}(t'))$ refers to the change in portfolio holdings relative to the initial time t' . The ETR associated with the sequence of trades is then, in analogy with (8),

$$E_{i\alpha}(t, t') = \sum_{n=1}^N e_{i\alpha}(t, t', t_n) \quad (12)$$

To illustrate the difference between the two, consider an agent trading a specific risky asset and cash. The specific sequence of four trades is taken to be the purchase of one share at each of the first two ticks, a held position on the third tick and the sale of one share at the fourth tick. In this case, for $R_{i\alpha}(t, t')$, one has

$$\begin{aligned} R_{i\alpha}(t, t') &= \sum_{n=1}^4 e_{i\alpha}(t_n, t_{n-1}) \\ &= (-1)(p(t) - p(t=4)) + (0)(p(t=4) - p(t=3)) \\ &+ (+1)(p(t=3) - p(t=2)) + (+1)(p(t=2) - p(t=1)) \end{aligned} \quad (13)$$

whereas $E_{i\alpha}(t, t')$ is given by

$$\begin{aligned} E_{i\alpha}(t, t') &= \sum_{n=1}^4 e_{i\alpha}(t, t', t_n) \\ &= (+1)(p(t) - p(t=4)) + (+2)(p(t=4) - p(t=3)) \\ &+ (+2)(p(t=3) - p(t=2)) + (+1)(p(t=2) - p(t=1)) \end{aligned} \quad (14)$$

Thus, for $R_{i\alpha}(t, t')$ one considers the change in portfolio holdings from one tick to the next, whereas for $E_{i\alpha}(t, t')$ one considers the cumulative change relative to some initial holding. A consequence of this is that $E_{i\alpha}(t, t')$ tends to be a more volatile measure as a large net position leads to potential large changes in ETR of opposite sign every time there is a price change.

It should be noted that all three measures of ETR have been tested in the context of both agent-based markets Benink et al. (2004, 2010) and experimental markets Stephens et al. (2006). Although it is possible to imagine situations where the different measures can give quite different results, these tend to be rather pathological. Up to now, in applications, all three measures have led to remarkably coherent but complementary results, giving different perspectives as to how market inefficiencies evolve.

Having defined the ETR associated both with a single trade, a fixed time interval, or a sequence of trades, one may proceed to define an average ETR per trade, $\bar{E}_{i\alpha}(t, t')$, and an average ETR per unit time, $\hat{E}_{i\alpha}(t, t')$ defined by

$$\bar{E}_{i\alpha}(t, t') = \frac{1}{N_{i\alpha}} E_{i\alpha}(t, t') \quad (15)$$

$$\hat{E}_{i\alpha}(t, t') = \frac{1}{(t - t')} E_{i\alpha}(t, t') \quad (16)$$

Furthermore, one may define a total, or average, ETR associated with any relevant subset of traders, \mathcal{G} , consisting of $N_{\mathcal{G}}$ agents, simply by summing, $E_{\mathcal{G}}(t, t') = \sum_{i \in \mathcal{G}} E_{i\alpha}(t, t')$, or averaging, $\bar{E}_{\mathcal{G}}(t, t') = (1/N_{\mathcal{G}} N_{i\alpha}) E_{\mathcal{G}}(t, t')$, the ETR over those agents $i \in \mathcal{G}$ to obtain the average ETR per trader, or the average ETR per agent per trade, in the group \mathcal{G} . In the former case we will refer to it as the ETR of a “representative” agent. As an illustration, one could consider trying to determine if financial institutions are making excess profits at the expense of the rest of the market, or whether the particular investment decisions of a fund of funds style hedge fund have resulted in excess returns, where in this case the sum will be over the funds in the group.

One could also choose a trading strategy, α , itself as \mathcal{G} , whereupon one would sum over all agents using that strategy. At this juncture it is worth discussing what we mean by “strategy”. This can best be done by appealing to a biological analogy—that between phenotype and genotype, the former being the physical manifestation of the organism whose genetic makeup is encoded in the genotype. In this context the phenotype of a trading strategy used during an interval of time $[t, t']$, is the actual pattern of trades $\{\Delta n_{i\alpha}(t_1), \Delta n_{i\alpha}(t_2), \dots, \Delta n_{i\alpha}(t_n)\}$ that lead to changes in portfolio holdings between t and t' . Both the agent’s portfolio holdings and wealth depend uniquely on this sequence. However, there may be many underlying trading strategies that lead to this sequence. This underlying strategy, which will depend on the agent’s information set, utility function etc. can be thought of as the genotype. In this sense, by α , one could imply either the phenotypic or genotypic strategy.

ETR as a Stochastic Variable

As excess profit is a stochastic variable, there is always a non-zero probability that, over a given time interval, an agent makes a profit just by chance. Hence, it is natural to refer the magnitude of any ETR (“signal”) to the degree of variance (“noise”) in the ETR, measured, for example, in units of the standard deviation. One may consider different statistical ensembles associated with different ways of grouping trades together—across time for a particular trader, or across a set of traders at a particular time, or a combination of the two. Which one is the most natural will depend on the question under investigation. For a given agent, i , one may consider the volatility associated with a particular sequence of trades, or the volatility per trade, or per unit time. For instance, the variance per trade is

$$\sigma_{i\alpha}^2(t, t') = \frac{1}{(N-1)} \sum_{n=1}^N (e_{i\alpha}(t, t', t_n) - \bar{E}_{i\alpha}(t, t', t_n))^2 \quad (17)$$

while the variance per trader is

$$\sigma_{i\alpha}^2(t, t') = \frac{1}{(N_G-1)} \sum_{n=1}^{N_G} (e_{i\alpha}(t, t', t_n) - \bar{E}_G(t, t'))^2 \quad (18)$$

Now, one may construct analogs of the Sharpe ratio [Sharpe \(1964\)](#) that measure the ETR associated with a particular trader, set of traders or trading strategy. These may be constructed by taking any measure of ETR and dividing by the corresponding volatility measure. For an agent i , using a strategy α , we may introduce a reward to variability ratio $S_{i\alpha}(t, t')$ given by

$$S_{i\alpha}(t, t') = \frac{\bar{E}_{i\alpha}(t, t')}{\sigma_{i\alpha}(t, t')} \quad (19)$$

3. ETR and Measures of Relative Inefficiency between Agents

In the last section, we introduced ETR as the most appropriate measure of excess returns due to trading. We now adopt the canonical stance that in an efficient market it should not be possible to make excess or abnormal returns. However, due to the defects discussed in Section 2 of the usual benchmarks we will define efficiency in terms of the ETR. This will have the advantage of, in principle, being a completely empirical measure, independent of any underlying asset pricing model, while at the same time focusing on the performance of a specific set of trades within a zero-sum game.

To proceed we first introduce the idea of the relative ETR⁶, $\mathcal{E}_{ij}(t, t')$, between one trader or set of traders versus another, defining the relative ETR between traders or trader groups i and j to be

$$\mathcal{E}_{ij}(t, t') = (E_{i\alpha}(t, t') - E_{j\beta}(t, t')) \quad (20)$$

Thus, we take $\mathcal{E}_{ij}(t, t')$ as a measure of the extent to which an agent or agent group, i , may make excess returns relative to another agent or agent group j . However, as $\mathcal{E}_{ij}(t, t')$ is

a stochastic variable, once again it is natural to measure it in terms of a volatility measure. As (20) represents a “signal” of inefficiency, that is, the difference between two means, it is most natural to consider the standard error associated with the stochastic variables $E_{i\alpha}(t, t')$ and $E_{j\beta}(t, t')$ defined by

$$\sigma_{err} = \left(\frac{\sigma_i^2(t, t')}{N_i} + \frac{\sigma_j^2(t, t')}{N_j} \right)^{\frac{1}{2}} \quad (21)$$

where N_i and N_j are the numbers of observations in the statistical samples used to calculate σ_i and σ_j . For instance, in the case where i and j represent agent groups, then σ_i and σ_j are the standard deviations associated with the representative agents of the groups and hence N_i and N_j are the number of agents in the two groups, respectively. On the other hand, for a particular agent they would represent the number of trades in the interval $[t, t']$ associated with the agents i and j , respectively. Thus, the more trades an agent participates in the higher the degree of statistical confidence one may have about any ETR.

Now we may make the following definition:

Definition 1. The Relative Inefficiency, $I_{ij}(t, t')$, between two strategies, or agent groups, i and j , over the time interval $[t, t']$ is

$$I_{ij}(t, t') = \frac{(E_i(t, t') - E_j(t, t'))}{\left(\frac{\sigma_i^2(t, t')}{N_i} + \frac{\sigma_j^2(t, t')}{N_j} \right)^{\frac{1}{2}}} \quad (22)$$

Dividing by the standard error, σ_{err} , gives a result that measures the statistical confidence one can have in the ETR of the two agents or agent groups. A reasonable criterion for concluding that two agents or agent groups are relatively inefficient over a given time interval is that $I_{ij}(t, t') > n$ over that time interval, where n is a measure of the confidence interval required. A reasonable value of n is $n = 1.96$ corresponding to the 95% confidence interval in the case that the distribution can be approximated by a normal distribution, though, of course, one may require a stricter criterion.

4. A New Measure of Market Inefficiency

With a definition of the relative inefficiency between agents or agent groups in hand, one may consequently attack the question of inefficiency in the market as a whole. To this end, to consider the entire market, the division into agent groups should be a partition, i.e., it should cover all agents, and any agent should appear in one and only one group—the finest partition being one where every agent is considered separately. Thus, we characterize the inefficiency of a market by making the following definition:

Definition 2. The Inefficiency Matrix, \mathbf{I}^m , for a market m evolving from time t to time t' is the matrix with matrix elements $I_{ij}(t, t')$ from Equation (22).

We take the Inefficiency Matrix to give a complete description of the relative inefficiencies that exist in a market. Note that this definition of inefficiency, with the measures of ETR we have proposed, is totally endogenous, making no reference whatsoever to any external benchmark. Naturally, the Inefficiency Matrix may also be defined by defining its matrix elements with respect to some exogenous benchmark, such as the risk free interest rate. Such a fixed, universal benchmark would clearly cancel from $\mathcal{E}_{ij}(t, t')$, being the same for both i and j .

Note that \mathbf{I}^m is antisymmetric, i.e., $I_{ij}(t, t') = -I_{ji}(t, t')$ as, if agent group i is making profits relative to trader group j , then agent group j is making losses of exactly the same magnitude relative to trader group i . Furthermore, of course, a trader cannot make profits from themselves. A market will be defined as inefficient over the interval $[t, t']$ if any

$I_{ij}(t, t') > n$ over the interval, where n is chosen according to the degree of confidence one requires.

At the most fine grained level, for a given definition of ETR, \mathbf{I}^m is unique when calculated in terms of individual agents. However, in this case the dimensionality of the matrix is that of the number of agents. Any reduction in the dimension of the matrix via an aggregation or coarse graining to agent groups will not be unique. In this case a market may be observed to be inefficient with respect to one market partition but not with respect to another. This is not a defect of the definition but rather a caution about how to aggregate. For instance, imagine a market where there are two similar informed traders making profits relative to a group of two noise traders. The market may be divided up into two groups, 1 and 2 in several different ways. If group 1 corresponds to the two informed traders, and group 2 to noise traders, then $I_{12}(t, t') > 0$ and could be made $> n$ depending on the relative informational advantage of the informed relative to the uninformed. On the other hand, if both groups consist of one informed and one noise trader then $I_{12}(t, t') \sim 0$.

An associated single inefficiency measure for the whole market is

$$\mathcal{I}^m(t, t') = \frac{1}{N'} \left(\frac{1}{2} \text{Tr}(-(\mathbf{I}^m)^2) \right)^{\frac{1}{2}} \quad (23)$$

where the trace is over all strategies or agent groups and the normalization factor $N' = N(N-1)/2$, where N is the number of strategies or agent groups in the market. With this single market measure we could in principle also consider the relative inefficiency of one market versus another.

It is important to emphasize that in a real-world commercial market the question of whether a market is efficient or not is really an empirical one, as we do not have a valid underlying theory that can demonstrably prove a market to be efficient or not. Moreover, it is one that can only be answered statistically, given that the evolution of a market is stochastic. In that sense the empirical question boils down to one of: Can one infer that a financial market is efficient from a set of data?

5. Applying the Inefficiency Matrix to Empirical Data

As an illustration of how the Inefficiency Matrix can be applied to empirical data to test for and understand inefficiencies we will consider data from two different studies, one simulational and one experimental. In the former, as we have the advantage of being able to create the market, we can choose to introduce informational asymmetries directly that can and should result in inefficiencies. For the experimental market the detection of efficiencies is purely inferential. However, due to the fact that we know the explicit identity of every market participant and the counterparties to every trade the subsequent analysis is greatly facilitated compared to standard market data.

5.1. Artificial Financial Market

Agent-based markets have been gaining more acceptance in recent years [LeBaron \(2000\)](#). In [Benink et al. \(2010\)](#) such a market was used to consider market inefficiency and learning. The market consisted of one risky asset, without dividends, and one riskless asset ("cash") without interest. The market-clearing mechanism was essentially a call-auction, a simulation being carried out for a prescribed number of auctions/ticks with shares bought in fixed size lots of one share. At each auction/tick an agent takes a position (buy/sell/neutral). Resources are finite and hence traders have portfolio limits associated with either zero cash or zero stock. Short selling and borrowing were not permitted. After each tick, price is updated exogenously via a supply/demand type law which is common to several artificial markets, including the original Santa Fe model [Palmer et al. \(1994\)](#). The wealth of agent i at time t is $W_i(t) = (E_i(t) + H_i(t)p(t))$, where $E_i(t)$ and $H_i(t)$ are the cash position and holding in the risky asset at time t . For each auction: At every tick each agent takes a position with an associated volume and at a given price, each agent being able to value the asset independently but with prices that are not too different⁷.

Trading strategies are taken from a simple one-parameter family described by a “bias”, d , ($d \in [0, 1]$), where the associated position probabilities are:

$$P(\text{buy}) = 2d/3, \quad P(\text{neutral}) = 1/3, \quad P(\text{sell}) = 2(1 - d)/3 \quad (24)$$

We denote a trading strategy by the pair $(100d, 100(1 - d))$. For example, for $d = 1/2$, $(50, 50)$ corresponds to a liquidity or noise trader. An alternative, or complementary, interpretation is that it corresponds to a trader or investor that believes (correctly or incorrectly) that the market is efficient, having no statistical bias in favor of one position versus another. In contrast, a trader with $d = 1/0$ corresponds to a trader with a strong belief that the market will rise/fall. One could also interpret a biased trader to have a superior understanding of the underlying market dynamics—knowing that, all else being equal, a preponderance to buy/sell will lead to excess demand/supply, which in its turn will lead to a price increase/decrease, which will lead to a higher valued portfolio. In this sense, these traders are informed relative to their noise trading counterparts. In this case, the logic follows that of [De Long et al. \(1990\)](#): That asset markets can fruitfully be analysed as though they are populated by two types of agents—those who are rational and understand the market, and those who trade on market noise. The point is that there are important precedents for this schematic choice of trading strategies. Furthermore, intuitively, one would expect the informational asymmetry between the strategies to be associated with inefficiency when they were present at the same time.

5.1.1. Characterization of Efficient Markets

We first consider the characterization of an efficient market. As one of the usual ways of viewing market efficiency is in terms of rational expectations, as meaning homogeneous expectations, we begin by considering markets where agents have homogeneous beliefs. In the left graph of [Figure 1](#) we see a histogram of the ETR after 3001 auctions for an homogeneous group of informed $(90, 10)$ traders, the distribution at $t = 0$ being a single peak at the origin. The evolving distribution is always unimodal (modulo noise effects). Its mean is always zero, as we are playing a zero-sum game, where the losses of the traders to the left of the origin are equal to the profits of the traders to the right. The variance increases in time as $A(n)t$, where $A(n)$ is a constant that measures market liquidity—the less liquid the market the more volatile, and therefore the greater the dispersion. Thus, the unimodal distribution spreads. These characteristics hold for any homogeneous group of traders, irrespective of how informed they are, the only variation being in the variance, as controlled by the market liquidity. Thus, we may conclude that any homogeneous market is efficient, in that there are no systematic ETR. Although this covers the standard efficient markets equilibrium of homogeneous rational expectations, meaning that the agents agree on a fundamental equilibrium price, it is obviously much more general, covering cases where agents are boundedly rational or irrational—as long as they are all in the same boat!

Although ETR distributions for different homogeneous markets are similar, price behaviour can be very different. For example, in a market consisting only of noise traders, price exhibits a random walk around the origin, while for a market of informed traders with a large bias, price increases very rapidly due to the large excess demand. This also shows that efficiency, in the sense of no systematic, excess profits, is generically quite different from the notion of allocative efficiency [Code and Sunder \(1993\)](#), a homogeneous market of informed traders always being in disequilibrium. Thus, the hallmark of an efficient market is that no subgroup of traders makes systematic excess profits at the expense of any other. This is manifest in the histograms of ETR by the fact that, on average, the distribution is unimodal and by the fact that if we followed the trajectory of a given agent it would perform a random walk around the origin, spending, on average, as much time winning as losing. Thus, homogeneity of the market should be a sufficient condition for efficiency. It is not clear whether it is also necessary.

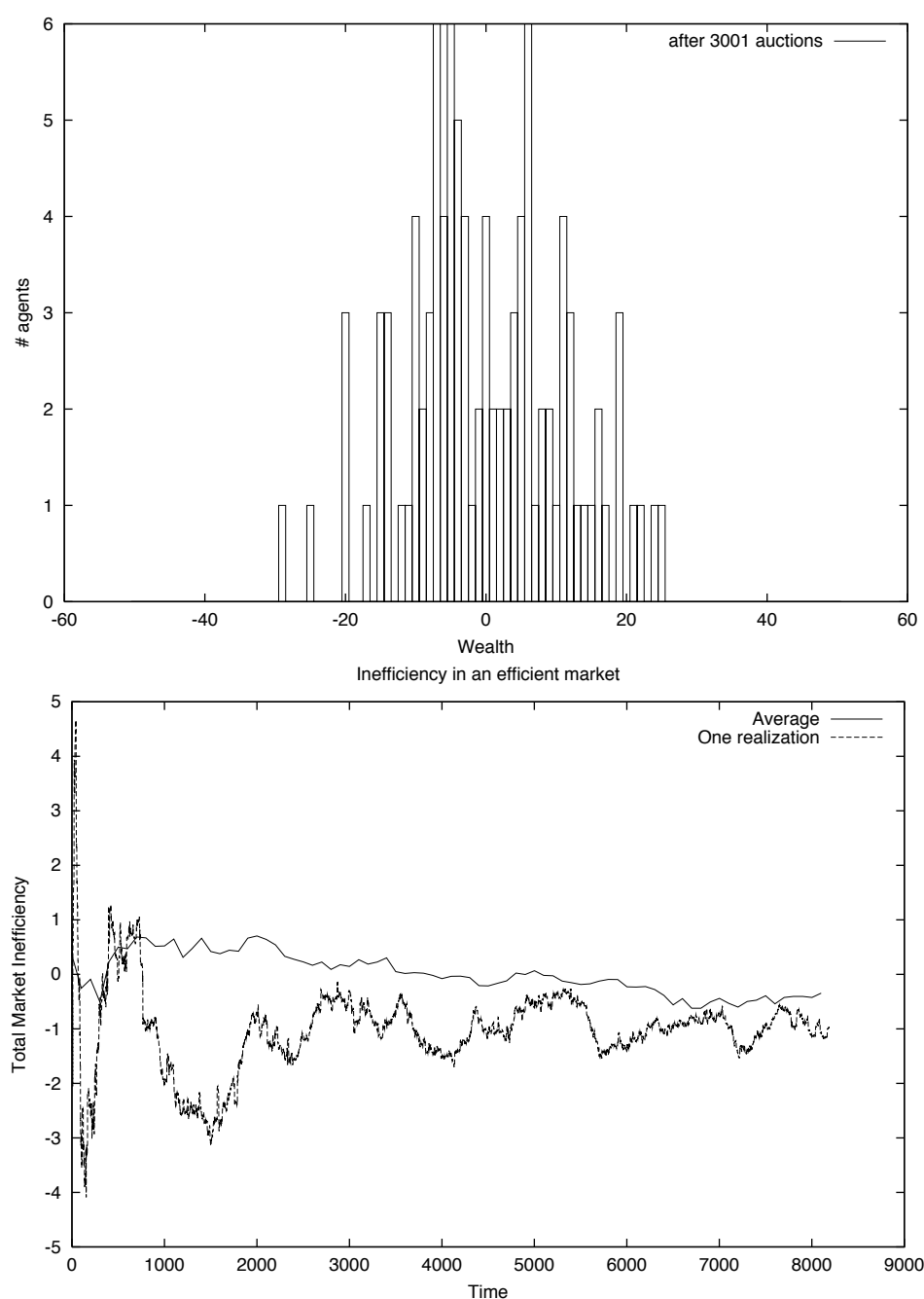


Figure 1. Histogram of ETR for 100 (90, 10) traders (**top**) and Graph of $I_{ij}(t, 0)$ for a fixed random subgroup of 50 noise traders in a homogeneous market (**bottom**).

In terms of the Inefficiency Matrix, we can check that any chosen subgroup of traders has no ETR by considering I_{ij} , where i refers to the group of interest and j refers to the rest. Even though we know this market to be homogeneous by construction we can ask whether it is possible to deduce it to be by observation? This would be achieved by examining the Inefficiency matrix for different subgroups of agents. In the right graph of Figure 1 we plot $I_{ij}(t, 0)$ for a group of 50 noise traders chosen randomly from a total of 100. Apart from an initial transient, $|I_{ij}(t, 0)| < 2$. In fact, taking an average over 10 different experiments leads to a resultant curve that leads to consistency of the null hypothesis that the market is efficient. We can also see that the volatility of the curve diminishes as a function of time. This is a consequence of the definition of the Inefficiency Matrix where, for an efficient market, the numerator is zero on average, while the denominator increases linearly with time. Thus, as t increases, the gradual reduction in I_{ij} allows us to more reliably infer that

the market is efficient. This also allows us to see that, even if we decided that excess profits (or in this case losses) were being made, say over the first 1500 ticks, we may conclude that they are not systematic, in that after this time there are no excess profits (or losses).

It is worth emphasizing again here why we are using a dynamic Buy and Hold benchmark: In a market of informed (90,10) traders, all agents will make large profits relative to a global benchmark, such as a risk free interest rate. However, no trader in the market is making systematic excess profits relative to any other. In distinction to the case of a market with noise traders, though, the (90, 10) market does not correspond to a rational expectations equilibrium, in that there is a strong, continuous excess demand. In both cases, however, agents may be acting perfectly rationally with respect to optimizing their own utilities. It should be clear that, when we talk about a homogeneous market here, we mean that the agents use strategies that are statistically the same. It does not necessarily follow that what might be behind those strategies—utilities, risk preferences, information sets etc.—is the same. Agents may be rational, boundedly-rational, irrational, risk-neutral, risk averse etc. Recall, as well, the discussion on the genotypic and phenotypic points of view on strategies—it may well be that two statistically different strategies give rise to the same pattern of buys and sells. In other words, observationally, trying to determine whether two trading strategies are statistically the same or not is a difficult inference problem.

5.1.2. Characterisation of Inefficient Markets

Having determined homogeneous markets to be efficient, we now turn to a characterization of inefficient markets. Although it is of great interest to ask how inefficiencies appear and disappear, for example through learning, here we want to consider an idealized situation, where a time period exists such that the inefficiency persists across the entire time period. Additionally, we will restrict for the moment to the case where there are only two different trader groups in the market, i.e., two different representative agents. As a concrete illustration, we consider a market consisting of 50 noise traders and 50 informed traders with a (90, 10) bias. In the left graph of Figure 2 we see a histogram of the associated ETR after 3001 ticks. The most marked difference with Figure 1 is the bimodal nature of the distribution, with the informed traders to the right and the uninformed to the left, thus indicating that the informed traders are making excess profits at the expense of the noise traders. Such a multimodal structure is symptomatic of a market inefficiency. For instance, with three statistically different trader groups one would expect to see a trimodal distribution. The mean of the distribution remains at the origin and, once again, the variance increases. However, in this case the variance of the overall distribution increases very rapidly due to the bimodal nature of the distribution. If we focused on the trajectory in the distribution of a particular agent, then an informed agent would perform a random walk within the right hand part of the bimodal distribution, while an uninformed agent would perform a random walk in the left part. The velocity at which the two peaks separate depends on the difference in bias between the two agent groups. This may be thought of as a measure of the “informational advantage” one group has over the other.

Turning to our other metric—the Inefficiency Matrix—we see in the right graph of Figure 2 the behavior of $I_{(100-x,x)(50,50)}$ for $x = 5, 10, 15, 20, 25, 30, 35, 40$ and 45 , for a market consisting of 50 traders of type $(100 - x, x)$ and 50 noise traders. Clearly the market is unambiguously inefficient, i.e., $I_{(100-x,x)(50,50)} > 2$ for all markets that include informed and uninformed traders together, where the inefficiency increases monotonically with time due to the monotonic increase in excess profits of the informed. Note that the degree of inefficiency strongly depends on the degree of bias of the informed agents. The higher the bias the higher the excess demand and therefore the higher the average price increase between auctions. This, in turn, leads to higher excess profits for those traders that have a bias to buy. For larger biases the inefficiency increases approximately as $t^{1/2}$. Eventually the curves of Figure 2 begin to flatten out as time goes on. This is a simple consequence of the existence of portfolio limits. As cash is used up exponentially by the informed as a function of time (due to the exponential increase in the price of the

stock when both informed and uninformed agents are present), whereas the stock of the informed is used up only linearly (only one unit of stock can change hands every auction), the informed are the first ones to decouple from the market and for that reason the rate of increase of the inefficiency goes to zero and the inefficiency itself becomes constant. The stronger the bias of the informed the quicker they decouple. So, analogous to the situation discussed in [Gode and Sunder \(1993\)](#), in this sense, efficiency can come to the market just by the existence of rules (no short selling or borrowing) as opposed to the more normally considered route of learning. Clearly then, we see that heterogeneous trading strategies can lead to market inefficiencies, the size of the inefficiency depending on the relative informational advantage of one trading strategy versus another.

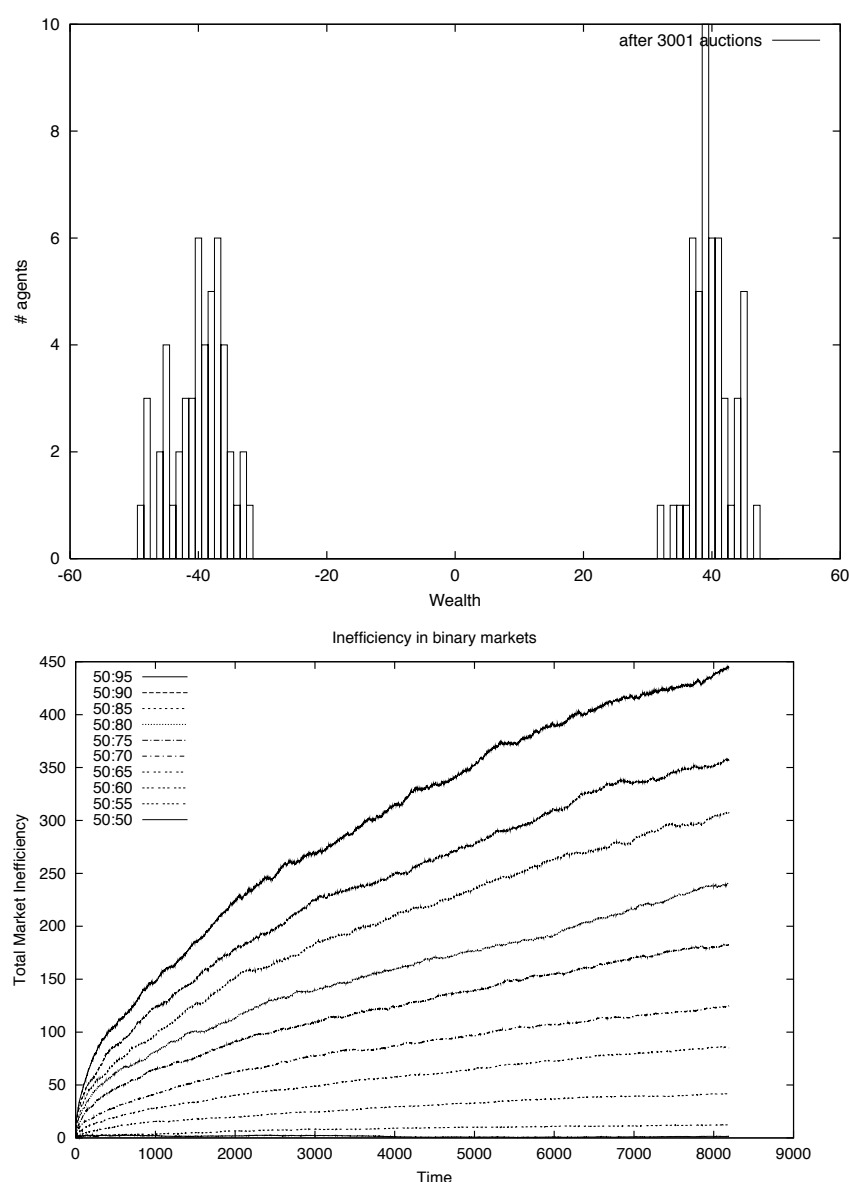


Figure 2. Histogram of excess profits for 50 (90, 10) traders and 50 (50, 50) traders after 3001 ticks (**top**) and graph of $I_{ij}(t, 0)$ for markets with 50 $j = (50, 50)$ traders and 50 $i = (100 - x, x)$ traders. In the legend, 50:65, for example, corresponds to a market with 50 noise traders and 50 (65:35) traders.

5.2. Experimental Markets

Experimental markets are an interesting area for applying our inefficiency measures, as they are intermediate between an artificial market, where the trading strategies are transparent but algorithmic, and real-world commercial markets, such as the LSE or Euronext,

where obtaining data that allows us to evaluate the success of every market participant is very difficult to obtain. An experimental market also is sufficiently simple that the relationship between trading success and information may be analysed or that characteristics of the agent's trading strategy may be inferred as we have access to every trade of each agent. The application of our excess profit measures was applied to experimental markets in [Stephens et al. \(2006\)](#).

The markets we will examine are political stock markets (PSM), which were the first experimental prediction markets [Forsythe et al. \(1992\)](#), where agents freely trade contracts (called "stocks") written on the vote-shares of candidates or parties (vote-share markets) in elections. Prices in PSMs are derived by participants setting limit or market orders for the different stocks. For instance, if a trader believes a party is overvalued, they may sell the share; if they consider a party undervalued they may buy it. Equilibrium prices should then be a good forecast for the election outcome, reflecting the aggregate opinions of all traders.

The fact that the intrinsic values of contracts traded on capital markets are never revealed is a major constraint concerning the possibility to test those markets for information efficiency. In this sense, PSMs have a clear advantage, in that the intrinsic value of its contracts is known after the election. Taking into account that many of the numerous PSMs conducted in the past produced forecasts better than those of opinion polls, we can imagine those markets to be quite informationally efficient in general. Despite that, there is evidence in the literature that traders can make systematic profits under certain circumstances (see [Stephens et al. 2006](#) and references therein).

We will illustrate the use of ETR and the Inefficiency Matrix using a PSM that was conducted from mid-August to mid-September 2004 for the regional parliament elections of the German state of Brandenburg. The market was of continuous double-auction type, with a partially open order book, traders having access to the three best bids and offers for every stock. Participation was basically open to anyone with internet access. Each participant started with an initial endowment of 100,000 units—Wahldollars—of currency for free.

As emphasized, a market is efficient or not with respect to a given partition of the market into agent groups—the most fine grained partition being that of individual agents. However, in that case, fluctuations are generally large and it is more difficult to reliably infer inefficiencies. Additionally, with a large number of matrix elements being tested, it is quite probable that some will exhibit an apparent signal just by chance. If we are to group agents, on what basis should this be done? In order to have in hand an efficient benchmark with a good degree of statistical reliability, we considered first a partition consisting of two random subgroups of agents. As the groups were chosen randomly, one would not expect to see any statistically significant inefficiency. So, where might one most likely find inefficiency?

As there are very many possible aggregations, one is naturally led to think what are the most likely partitions that may exhibit inefficiency. Although we cannot group by strategy—genotype—per se, we may group by performance. To this end it is natural to surmise that the most successful traders, i.e., those with the highest ETR, are the most likely to be associated with an inefficient partition. However, if we look at the ETR over the period of the entire market we are potentially doing nothing more than verifying that we had chosen a very biased sample in the first place. In order to be able to conclude that there is an inefficiency, the chosen group must be selected using a particular sample—a training set—and subsequently must be shown to be making ETR over a test set of data as well. If this is the case, and it is statistically significant, then we may conclude that there is an inefficiency.

In the in-sample set, the agents were ranked in terms of performance by ETR and divided into deciles, decile 10 corresponding to the 10% of traders with the highest ETR in the first half of the market. The matrix element $I_{ij}(t, 0)$ was then determined, where i corresponded to decile 10 traders and j corresponded to the rest of the market, thereby

determining if there is a relative inefficiency between the best performing traders and the rest.

The results can be seen in Figure 3. The left hand graph corresponds to an arbitrary, random choice of trader groups. Several other divisions were performed with similar results. Of course, as the Inefficiency Matrix elements are essentially *t*-tests, as stated, one must be aware of the possibility that many comparisons can lead to the possibility of a significant result just by chance. Note how, as expected, the inefficiency is initially quite volatile. Relatively quickly however, it becomes smoother as the increasing size of the statistical sample (total number of trades) allows for better inference. The overall magnitude of the inefficiency varies between about 0 and 1, hence we may conclude that there is no statistically significant inefficiency here, and therefore that this market is, as expected, for this partition, efficient.

On the right hand side of Figure 3 we see the graph of inefficiency as a function of time in the in-sample period for the most successful trading strategy versus the rest of the market. As can be seen, there is an apparent strong inefficiency. However, there is, by definition, also a strong sampling bias that needs to be accounted for. If, indeed, these traders are “superior” traders, systematically over time, then in the out-of-sample results we should also see evidence of inefficiency.

The out-of sample results can be seen in Figure 4. For the randomly chosen trader groups we see that, as in the in-sample data, apart from a brief, initial transient there is no significant inefficiency. On the contrary, in the case of the best performing trader group, we see that, although the overall corresponding inefficiency is much reduced relative to the in-sample set, there is still a steady increasing trend that eventually reaches statistically significant levels. If the ETR associated with the first period were purely due to sampling bias then one would not, on average, expect to see any signal at all in the out-of-sample period. On the other hand, if these superior traders were uniformly superior over time, then one would expect to see just as strong a signal in the out-of-sample period as in the in-sample period.

Another interesting characteristic of the inefficiency of the best performing traders, especially in the out-of-sample data, is that the inefficiency curves show gradually diminishing slopes, thus indicating that the rate at which ETR were made by the winners was decreasing. This is not so manifest in the in-sample result shown in the right hand side of Figure 3 due to the presence of the strong selection bias, but is, however, more noticeable in the out-of-sample result shown in the right hand side of Figure 4, where from the 7th to the 14th the change in inefficiency is twice what it is from the 14th to the 20th.

A possible explanation of the fact that the overall level of inefficiency is reduced and decreasing in the out-of-sample data, is that over the course of this market learning takes place. This explanation can be corroborated [Stephens et al. \(2009\)](#) by a more detailed look at the individual trading profiles of the agents. In the first half of the market, the high-performance decile 10 traders made profits principally at the expense of the low-performance decile 1 traders, who had corresponding excess trading losses. However, in the second half of the market, the participation of the decile 1 traders decreased significantly and, hence, the profit opportunities of the decile 10 traders diminished. Essentially, the decile 1 losers learned that trading led to losses and so became risk averse and declined to trade. This meant that the decile 10 winners had to look for profits by trading against better traders than the decile 1 losers and therefore their excess profits were smaller.

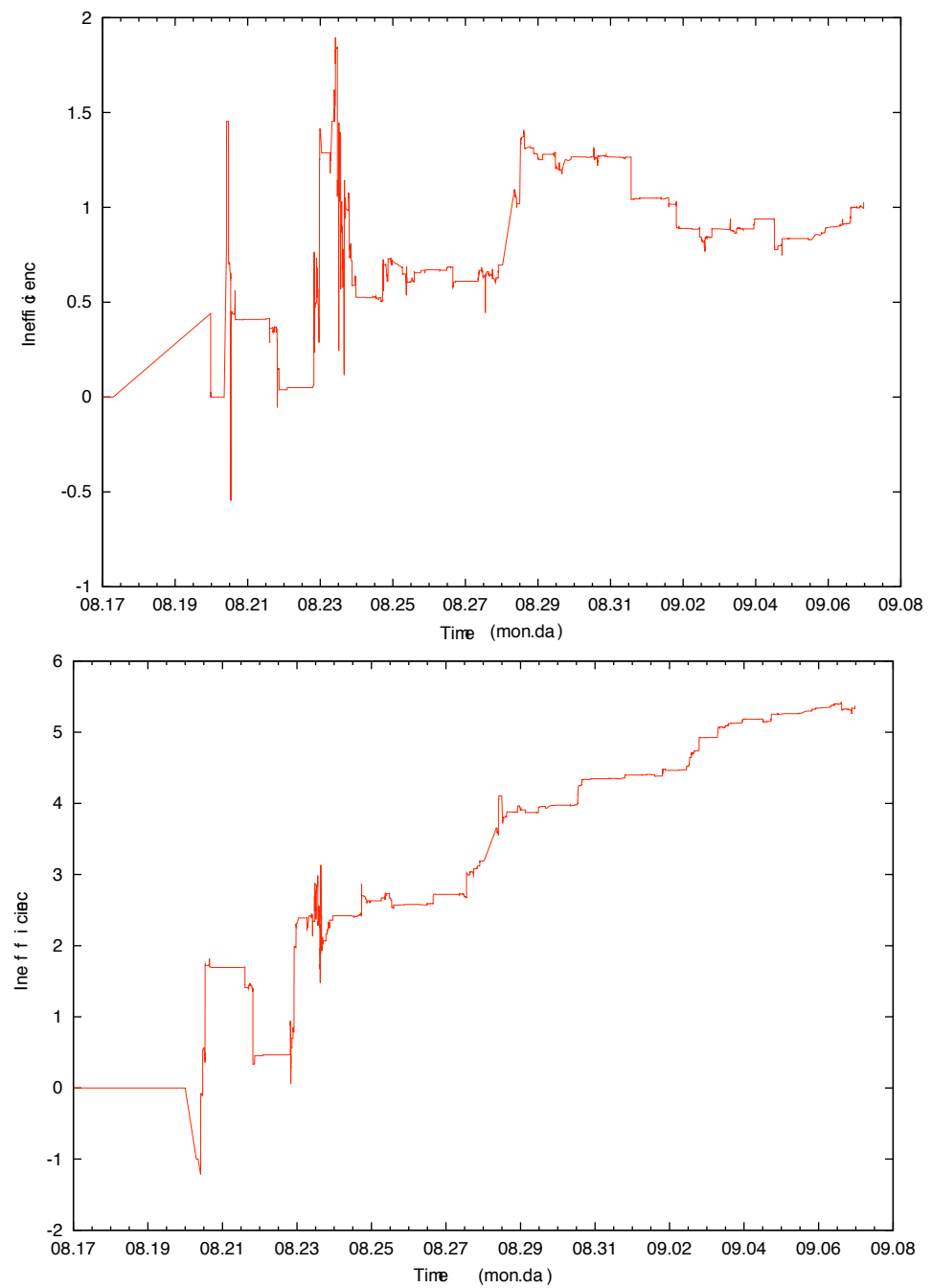


Figure 3. Inefficiency $I_{ij}(t,0)$ for two agent groups in the in-sample (first half) set. On the **(top)**, two randomly chosen groups and on the **(bottom)** the decile of agents with the highest ETR against the rest of the market.

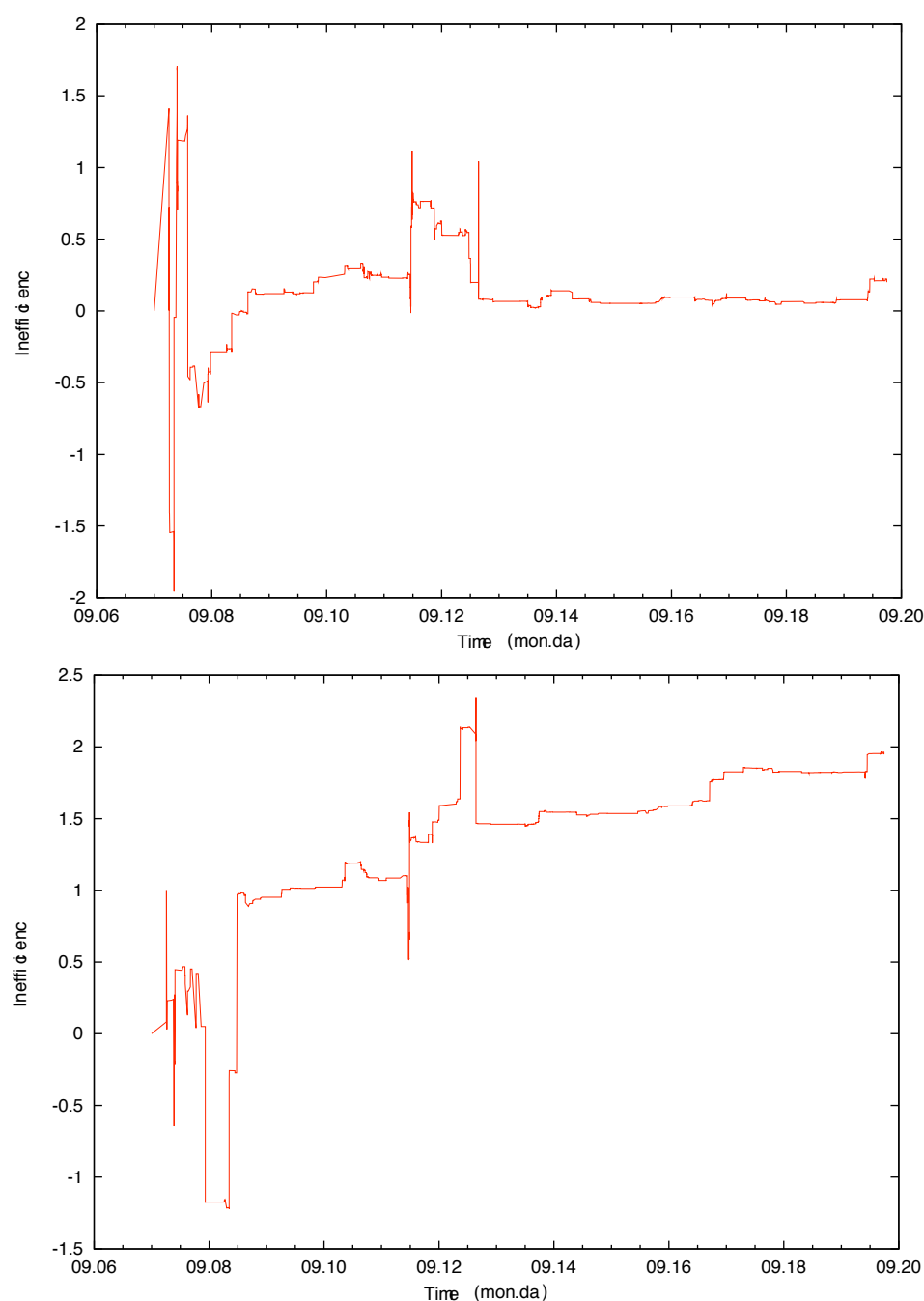


Figure 4. Inefficiency $I_{ij}(t, 0)$ for two agent groups in the out-of-sample (second half) set. On the (top), two randomly chosen groups and on the (bottom) the decile of agents with the highest ETR against the rest of the market.

6. Applications to Real-World Commercial Stock Markets

Using two types of financial market—artificial and experimental—we have shown how our inefficiency measures can be used to not only measure the degree of inefficiency in a market, and how it evolves over time, but also to understand how and why it arises. An obvious question then arises: why not just apply it to data from commercial stock markets such as the LSE or Euronext? Our inefficiency measures can clearly be applied to such stock markets. The question is more: what data is required in order to apply them?

The data challenges are the following: firstly, we require trade data that is labelled, i.e., that we can unambiguously label the counterparties in any trade. However, generic market data is anonymous. Some datasets have broker identifiers, but very few have the

associated buy-side counterparties identified. The labelling is important in that, at the most fine-grained level, we wish to determine if a particular market participant, or group, is making excess profits or, indeed, excess losses. We thus need an identifier. Moreover, we require that all market participants have such an identifier.

If we have a labelled set of trades then the inefficiency matrix can be constructed on an agent by agent basis. However, the statistical significance of each element depends on how many trades were done by the corresponding agent. As a statistical significance measure, although an agent might have made a large excess profit, if this was a result of only a few trades then this might not be distinguishable from luck [Stephens et al. \(2009\)](#).

By grouping agents we may obtain better statistics. However, as emphasized, the degree of inefficiency is dependent on how agents are grouped. As we have shown, a natural grouping is that of choosing those agents that are the most/least successful. However, in order to eliminate pure selection bias as the cause of the inefficiency, we must consider analysing an out-of-sample set of data.

There are many interesting labels by which trades could be grouped—trades by mutual funds, funds of a given type, institutional trades in general etc. In all cases however, we require that every trade is labeled by a buyer and a seller. For instance, one might imagine using publicly available mutual fund data to try to identify inefficiencies. However, although we may note the change in holdings of a fund from one period of time to another, we do not know how the portfolios of the counterparties changed. Additionally, mutual funds do not form a partition, i.e., there are trades done where the counterparty is not another mutual fund.

In summary: our empirical characterization of inefficiency does not suffer from any theoretical barrier, as is manifest with information based approaches, which are limited by the joint-hypothesis problem. As our artificial and experimental market examples show, any difficulty is purely practical, in that the application of our formalism requires a data set with particular characteristics. Moreover, it is not that such data does not currently exist but, rather, that such data is highly confidential and therefore very difficult to obtain. Finally, as our examples show, our inefficiency measures are not just “black boxes” that simply determine if there is a market inefficiency or not. By studying the characteristics of the groups that exhibit inefficiencies we may begin to infer why such inefficiencies exist in the first place and therefore potentially allow for the design of more efficient markets.

7. Conclusions

The Efficient Markets Hypothesis has been traditionally linked to the idea that security prices fully reflect available information. Unfortunately, its confirmation or negation has been highly controversial, due to the existence of the joint-hypothesis problem. To avoid the joint-hypothesis problem, we defined in this paper a purely empirical, quantitative measure of efficiency, defining an inefficient market in terms of whether or not there exist traders making systematic excess trading returns. To distinguish between trading returns and market returns over a given time interval, excess was defined for any agent by subtracting off from the agent’s change in portfolio value those gains that would have accrued if the agent had not traded over the interval. It was argued that a suitable measure should also correspond to a zero-sum game, as then for every net winner there is one or more net losers. Several candidates were introduced. All were of the form such that the ETR associated with a given trade is proportional to “ $\Delta n \Delta v$ ”. Thus, a non-zero ETR for a trade requires both a change in portfolio holdings and a change in asset values.

As ETR is a stochastic variable it is most naturally measured in units of the standard deviation of the excess profits. In this way one can not only distinguish between traders profiting from an active trading strategy, as opposed to those profiting purely from market gains, but one can also distinguish between those traders who have a “lucky” trading strategy versus an “intelligent” one. We also discussed the statistical inference problem associated with trying to infer the trading strategy and associated information set (“geno-

type”) of an agent from their observed trades (“phenotype”) in order to better understand the potential origin of any observed inefficiency.

Next, we developed a new measure for the relative inefficiency between trading strategies and an Inefficiency Matrix which aims at giving a complete description of the relative inefficiencies that exist in an entire market, thus enabling us to determine the degree of market (in)efficiency.

Finally, we illustrated the use of ETR and the Inefficiency Matrix in a pair of model markets. We first considered an agent-based market as in this case the trading strategies of the individual agents are explicitly known and it was therefore possible to know everything about the map between trading strategy and trades (the genotype-phenotype map). We saw that, generically, informationally homogeneous markets were efficient, in that no trader had an ETR that was significantly greater than any other. In contrast, in an informationally heterogeneous market we saw how informed traders generated statistically significant ETR by trading against uninformed traders. In the experimental market we considered, the underlying trading strategies of the agents are unknown. However, by grouping together by ETR in an in-sample set of data, we could determine that there existed a group of agents with a statistically significant inefficiency relative to the rest of the market. We argued that the fact that the rate of increase of the inefficiency decreased as a function of time was due to learning.

Our inefficiency measures can clearly be applied to financial market data from commercial exchanges and, in particular, in periods of high volatility, as in recent crises such as the Global Financial Crisis of 2007–2009 and the COVID-19 Crisis of 2020–2021. [Brown \(2020\)](#) notes that, in times of economic crisis, adaptations of behavior can imply that, although the wisdom of crowds implied by the efficient markets hypothesis governs over extended periods, it does not govern at every point in time. Because of this, asset markets should be more volatile in times of economic crisis. Establishing the existence of inefficiencies related to crisis events is also important from a regulatory and societal perspective. [Haslam et al. \(2018\)](#) argue that the phenomenon of stock market crashes, which most see as at least ex post evidence of manifestations of poor quality pricing, is for many linked to very negative outcomes. Even given some doubt about pricing in this context, a regulator concerned about social well-being should seek to do something to intervene.

Author Contributions: Conceptualization, H.A.B.; Formal analysis, C.R.S. and J.P.P.-G.; Software, J.L.G. All authors have read and agreed to the published version of the manuscript. .

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Peter Bossaerts, Doyne Farmer, Martin Gruber, Florian Hauser, Michael Kirchler, Marco Pagano, Christian Rieck, Jeremy Stein and Henri Waelbroeck for helpful comments on an earlier version of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ [Benink and Bossaerts \(2001\)](#) have applied the work of Hayek and related Neo-Austrian theory specifically to financial markets.
- ² For a recent discussion of these issues, see [Pastor and Veronesi \(2009\)](#).
- ³ A related notion—the return gap—was considered in [Kacperczyk et al. \(2008\)](#) in the context of measuring mutual fund performance.
- ⁴ Of course, we are here neglecting considerations of any costs, such as trading costs, information costs etc.
- ⁵ In the case of a market with fixed trading periods then one would be updating ETR every trading period.

- ⁶ In the context of entire markets it has been suggested in [Farmer and Lo \(1999\)](#) that “relative” efficiency of markets is a more useful notion than that of absolute efficiency.
- ⁷ Details can be found in [Benink et al. \(2010\)](#).

References

- Bagehot, Walter. 1971. The Only Game in Town. *Financial Analysts Journal* 27: 12–14. [\[CrossRef\]](#)
- Balvers, Ronald J., Thomas F. Cosimano, and Bill McDonald. 1990. Predicting Stock Returns in an Efficient Market. *Journal of Finance* 45: 1109–28. [\[CrossRef\]](#)
- Benink, Harald, A. and Peter Bossaerts. 2001. An Exploration of Neo-Austrian Theory Applied to Financial Markets. *Journal of Finance* 56: 1011–27. [\[CrossRef\]](#)
- Benink, Harald, A., Jose Luis Gordillo, Juan Pablo Pardo-Guerra, and Christopher R. Stephens. 2004. A Study of Neo-Austrian Economics Using an Artificial Stock Market. *EFA 2004 Maastricht Meetings*. Paper No. 3218, Available online: <http://ssrn.com/abstract=567125> (accessed on 9 June 2021).
- Benink, Harald, A., Jose Luis Gordillo, Juan Pablo Pardo-Guerra, and Christopher R. Stephens. 2010. Market Efficiency and learning in an Artificial Stock Market: A perspective from Neo-Austrian Economics. *J. Empirical Finance* 17: 668–88. [\[CrossRef\]](#)
- Brown, S. J. 2020. The Efficient Market Hypothesis, the Financial Analysts Journal, and the Professional Status of Investment Management. *Financial Analysts Journal* 76: 5–14. [\[CrossRef\]](#)
- De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy* 98: 703–38. [\[CrossRef\]](#)
- Fama, Eugene F. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance* 25: 383–417. [\[CrossRef\]](#)
- Fama, Eugene F. 1991. Efficient Capital Markets: II. *Journal of Finance* 46: 1575–617. [\[CrossRef\]](#)
- Fama, Eugene F., and Kenneth R. French. 1996. Multifactor Explanations of Asset Pricing Anomalies. *Journal of Financial Economics* 51: 55–84. [\[CrossRef\]](#)
- Fama, Eugene F. 1998. Market Efficiency, Long Term Returns and Behavioral Finance. *Journal of Financial Economics* 49: 283–306. [\[CrossRef\]](#)
- Farmer, J. Doyné, and Andrew W. Lo. 1999. Frontiers of Science: Evolution and Efficient Markets. *Proceedings of the National Academy of Sciences* 96: 9991–92. [\[CrossRef\]](#)
- Forsythe, Robert, Forrest Nelson, George R. Neumann and Jack Wright. 1992. Anatomy of an Experimental Political Stock Market. *American Economic Review* 82: 1142–61.
- Gode, Dhananjay K., and Shyam Sunder. 1993. Allocative Efficiency of Markets with Zero Intelligence Traders: Market as a Partial Substitute for Individual Rationality. *The Journal of Political Economy* 101: 119–37. [\[CrossRef\]](#)
- Haslam, Jim, Jiao Ji, and Hanwen Sun. 2018. Towards a Well-Functioning Stock Market in Context, Critically Appreciating Issues in Interpreting Efficient Markets Research and Its Regulatory Implications. *Journal of Capital Markets Studies* 2: 21–35. [\[CrossRef\]](#)
- Von Hayek, Friedrich A. 1937. Economics and Knowledge. *Economica* 4: 33–54. [\[CrossRef\]](#)
- Jacobs, Heiko, and Sebastian Müller. 2020. Anomalies across the Globe: Once public, no longer consistent. *Journal of Financial Economics* 135: 213–30. [\[CrossRef\]](#)
- Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng. 2008. Unobserved Actions of Mutual Funds. *The Review of Financial Studies* 21: 2739–416. [\[CrossRef\]](#)
- Kocherlakota, Narayana R. 1996. The Equity Premium: It’s Still a Puzzle. *Journal of Economic Literature* 34: 42–71.
- LeBaron, Blake. 2000. Agent-based Computational Finance: Suggested Readings and Early Research. *Journal of Economic Dynamics and Control* 24: 679–702. [\[CrossRef\]](#)
- Malkiel, Burton G. 2003. The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives* 17: 59–82. [\[CrossRef\]](#)
- Mehra, Rajnish, and Edward C. Prescott. 1985. The Equity Premium: A Puzzle. *Journal of Monetary Economics* 15: 145–61. [\[CrossRef\]](#)
- Palmer, Richard G., W. Brian Arthur, John H. Holland, Blake LeBaron, and Paul Tayler. 1994. Artificial Economic Life: A Simple Model of a Stock Market. *Physica* 75: 264–74.
- Pastor, Lubos, and Pietro Veronesi. 2009. Learning in Financial Markets. *Annual Review of Financial Economics* 1: 361–81. [\[CrossRef\]](#)
- Pesaran, M. Hashem. 1989. *The Limits to Rational Expectations*. Oxford: Basil Blackwell.
- Pesaran, M. Hashem, and Allan Timmermann. 1995. Predictability of Stock Returns: Robustness and Economic Significance. *Journal of Finance* 50: 1201–28. [\[CrossRef\]](#)
- Rösch, Dominik M., Avanidhar Subrahmanyam, and Mathijs A. Van Dijk. 2017. The Dynamics of Market Efficiency. *The Review of Financial Studies* 30: 1151–87. [\[CrossRef\]](#)
- Shleifer, Andrei. 2000. *Inefficient Markets—An Introduction to Behavioral Finance*. Oxford: Oxford University Press.
- Schwert, G. William. 2003. Anomalies and Market Efficiency. In *Handbook of the Economics of Finance*. Edited by G. M. Constantinides, M. Harris and R. Stulz. Amsterdam: Elsevier Science, pp. 937–71.
- Sharpe, William F. 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance* 19: 425–42.
- Shiller, Robert J. 2003. From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives* 17: 83–104. [\[CrossRef\]](#)
- Stein, Jeremy C. 2009. Presidential Address: Sophisticated Investors and Market Efficiency. *Journal of Finance* 64: 1517–48. [\[CrossRef\]](#)

- Stephens, Christopher Rhodes, Jose Luis Gordillo, and F. Hauser. 2006. Testing Efficiency in an Experimental Market using Excess Trading Returns. Paper Presented at WEHIA 2006, Bologna Italy. Available online: http://www.dse.unibo.it/wehia/paper/parallel/Stephens_5.4.pdf (accessed on 9 June 2021).
- Stephens, Christopher Rhodes, Jose Luis Gordillo, and Enrique Martínéz Miranda. 2009. Who's Smart and Who's Lucky? Inferring Trading Strategy, Learning and Adaptation in Financial Markets through Data Mining. In *Natural Computing in Computational Finance, Studies in Computational Intelligence*. Berlin and Heidelberg: Springer, vol. 2, pp. 95–114.
- Titan, Alexandra Gabriela. 2015. The Efficient Market Hypothesis: Review of Specialized Literature and Empirical Research. *Procedia Economics and Finance* 32: 442–49. [CrossRef]
- Woo, Kai-Yin, Chulin Mai, Michael McAleer, and Wing-Keung Wong. 2020. Review on Efficiency and Anomalies in Stock Markets. *Economies* 8: 1–51. [CrossRef]