The efficient market and market bubbles explained by a heterogeneous least-squares learning approach to fundamental analysis

Doug McLeod*

Version: **Staff Seminar Macquarie University 28th March 2008**

Abstract

A financial model is presented which resolves the ambiguity in the efficient markets concept. Fundamental investors do not build self-contained models of return but rather isolate particular factors out of observed return in order to understand it. Price is not the weighted average of individual views but the product of interlocking expectations, and the continual revision of expectations causes price to gravitate to the efficient point. This behaviour can be understood in terms of a hidden substrate in which price is an information storing vector, analogous to a gene in biology, and heterogeneous least squares learning is a genetic operator. The movement of price to the efficient point is emergent behaviour which cannot be determined from the initial state of the investors.

Keywords: Heterogeneous expectations, least-squares learning, efficient markets, genetic operator, coefficient model
JEL classification: G12, G14

---

1. Introduction

The concept of market efficiency was created (primarily by French mathematician Louis Bachelier and American economist Eugene Fama) as an description of the market without any particular relationship to underlying economic behaviour. A perfectly idealized version of this concept, strong market efficiency, was propounded by Fama (1970). When writers began to look for a quantitative explanation of market efficiency, tension arose between the idealized notion of efficiency and the theory of microeconomic behaviour necessary to generate it. In their seminal 1980 paper Grossman Stiglitz pointed to a paradox in the theory of financial markets. The Grossman Stiglitz study contained the following elements inter alia:

1. Two classes of individual, informed and uninformed. Subsequent literature has usually interpreted these as fundamental analysts and price watchers. The fundamental analysts do the hard work of valuing the security, and the price watchers look at price only.

---

* doug.mcleod@optusnet.com.au

2. The fundamental analysts incur cost whereas the price watchers do not.
3. The price watchers learn from price according to the standard rational expectations (RE) assumption which is that in equilibrium they have full knowledge of the return distribution conditioned on price.
4. There is 'strategy switching', which means that traders can move between fundamental analysis and price watching according to their results.
5. Trader demand decisions derive from explicit risk preference functions in a universe which contains a risk free security.

Grossman Stiglitz found that in a non stochastic environment price is perfectly revealing of the information of the fundamental analysts, and no equilibrium can exist. If anyone is paying for information then everyone gets it, but if no one is paying for information then everyone wants it. They conclude (1980, pg 405), "because information is costly, prices cannot perfectly reflect the information which is available, since if it did, those who spent resources to obtain it would receive no compensation". Grossman Stiglitz employ a further element in their model to prevent price being fully revealing:

6. A source of noise which prevents price being fully revealing. Noise can be inserted in various ways to much the same effect; the particular device employed by Grossman Stiglitz is variable security supply.

Many studies continue to work within the Grossman Stiglitz framework. In Bray(1982) the price watchers use least squares learning to determine the relationship between price and return instead of RE omniscience. Bray is able to confirm the Grossman Stiglitz result.

Brock Hommes (1997) broke new ground in their use of mathematics to characterize the dynamics of market equilibrium. They extended learning from estimation of parameter values to the choice of estimator itself . Within their model sophisticated estimators are available at a cost, and simples ones are free of charge. Use of the sophisticated estimators implies a globally stable equilibrium. Use of simple estimators implies the same unique steady state, but it is unstable. The model oscillates as the effective estimators only pay for themselves when the price is somewhat inaccurate. Brock Hommes show that if agents are fast to adapt the best estimator (what they term 'intensity of choice') a 'homoclinic orbit' exists which implies complicated dynamics.

Strategy switching within Grossman Stiglitz relies on the rational expectations assumption that investors automatically know which strategy is better (they choose the alternative with higher utility). Within the Brock Hommes study the choice of estimator is based on the accuracy and cost of the estimator. An alternative and more realistic way to model strategy switching is the genetic algorithm, whereby investors do not study the properties of the strategy directly, but rather emulate the more successful investors.

Goldbaum (2005) uses least squares learning and a random walk dividend return. For the switching mechanism he employs a method known as replicator dynamics, which is a game theory based model of biological evolution in which superior strategies attract traders from the other strategies. At equilibrium every strategy in use must have the same payoff (return). He finds convergence to the efficient point providing traders choose either fundamental or price information. If however traders can use both types of information then the price becomes unstable as the learning process has no equilibrium value.

Goldbaum is particularly concerned to show that noise can arise endogenously in a model. It is reasonable to suppose that a model not using exogenous noise devices would be a more 'elegant' resolution of the Grossman Stiglitz paradox. In a further study (2005b) he assumes a fixed security supply of zero rather than a noisy supply, and is able to show that the interaction between the learning process and the strategy switching process will generate endogenous noise. He finds that if inferior strategies are allowed to survive (via a Discrete Choice Dynamic evolutionary algorithm rather than a Replicator Dynamic algorithm) then the market will converge to an efficient point with heterogeneous traders.

Muendler (2005) evaluates the Grossman Stiglitz framework as follows: "The fundamental problem disguised behind the no-equilibrium paradox is a non-standard equilibrium definition… the equilibrium definition does not allow for a marginal analysis that yields a well-defined trade-off between marginal benefit and marginal cost of an additional signal." Muendler employs a statistical model in which traders purchase signals from an underlying one-dimensional Poisson distribution of the security return. There is a market for information as well as a market for the security. An equilibrium is derived in both markets. Muendler finds that all investors have an incentive to purchase information and this is another resolution of the Grossman Stiglitz paradox.

As a general remark, the models put sophisticated mathematics at the service of economic scenarios which are somewhat artificial. There is an unrealistic emphasis on cost in the strategy switching models. Once businesses have installed systems the ongoing costs are relatively small, and the systems are not changed often – certainly much less regularly than price fluctuates in a financial market.

1.2. Heterogeneous agents in financial simulations

The literature in financial simulation, also referred to as computational finance, uses computer simulation rather than analysis to investigate a wide variety of questions about markets. This question as to whether heterogeneity makes a qualitative difference to the way markets function has received considerable attention. Chiarella Gallegati Leombruni Palestrini (2003) uses the standard dichotomy between fundamentalist and price watching agents. The price watchers try to learn the best way to interact with the fundamentalists. A probabilistic framework using Gibbs distributions determines the excess demand of each group of agents and price movements are a stochastic process based on excess demand. The agents are heterogeneous in their ability to modify their behaviour in the light of experience and a genetic algorithm is used to update the proportion of agents at various levels of ability in the direction of the more successful ones. The study concludes that price watching investors may make greater profits than fundamentalists, and alter the dynamics of the system.

It appears that on balance heterogeneity is not helpful to the effective functioning of financial systems. It may be asked if this conclusion depends critically on the type of heterogeneity which has been studied: primarily heterogeneity in the price watchers who are free riders on the system. This study looks at heterogeneity in the fundamental analysts, those analysts looking at data other than price. Hitherto fundamental analysts have spoken with one voice – a correct but costly voice. As it turns out, what they are really doing is far more interesting than that.

1.3. The coefficient model

The point of departure of this study is that it is a ‘coefficient’ model in which security return is characterized in terms of the data used by the market rather than as an independent series. Its relationship to the Grossman Stiglitz elements is:

1. The investors combine the characteristics of both fundamental analysts and price watchers. In the standard model the fundamental analysts access data which has only one component and they have perfect knowledge of that component’s impact on return. Here the fundamental data has many components and the significance of each component can only be established by least squares learning. Every investor chooses some set of variables to analyze using multiple regression. Price is one more data series which may or may not be included in an investor’s data set. There is no dichotomy between fundamental analysts and price watching analysts. Rather the investors have different degrees of ‘informedness’ according to the number and relevance of the data series which they analyze. Even an investor who looks only at price is partly informed because they must have analyzed price. An investor who has not analyzed anything will never trade in this model.
2. Cost differentials are not used to motivate the model. If cost differentials were eliminated in analyst/price taker models then those models would favour analysts and resolve the paradox. This is not a natural thing to do because those models rely on cost for realism and would become trite in its absence. In this model heterogeneous information is the source of economic context. The author considers that cost is a second order consideration which can be enfolded into the discussion on return.
3. Any investor can learn from the price using least squares learning.
4. There is no strategy switching by traders. This is relevant to the economics of a cost based analysis but is an unnecessary elaboration here.
5. Risk and portfolio considerations are kept in the background. The demand and supply equations derive from a mean-variance framework.
6. There is a fixed security supply of zero and noise does not enter via this mechanism.

The contribution of the paper is as follows. Section 2 presents the coefficient model. Price is shown to be a linear function of market data, and this allows data to be eliminated from the model to reveal the underlying structure. Section 3 develops the price change equation, which states that price changes are not in the direction of the change of estimates *new minus old* as might be expected, but in the direction of *new* estimates. This is because both price and data coefficients are estimated, and the *old* components cancel out. Price changes push price in the direction of the unknown return parameters and the market is efficient. Section 4 undertakes the technical task of demonstrating that that the price regression coefficient is negative in a model where it is derived from regression rather than prior expectation. A negative value of the price coefficient is required for the market to make a price. Section 5 investigates the economics of the model. Contrary to the Grossman Stiglitz paradox, the model provides normal economic returns in equilibrium and investors are rewarded according to the value of their information. Section 6 demonstrates that security markets embody a genetic algorithm. An isomorphism with a natural population is derived from the properties of statistical estimation rather than from the selection of successful investors or strategies. Section 7 extends the coefficient model to a multiperiod context. The Gordon growth model can be found within the coefficient model as a special case, which gives the Gordon growth model a basis in observable data. Section 8 reviews major themes. The flip side of the requirements for market efficiency are the reasons for market bubbles and this is discussed. The genetic analogy is not incidental but critical to the paper's central conclusion – that price is not simply a function of other variables but a store of information with an independent existence, like a gene.

2. Definition of the coefficient model

2.1. Premises of the coefficient model

Premise 1: basic framework. There is a security. Payments at the rate of y per unit are made to security holders at fixed points of time spaced out at equal intervals, for instance every midday. The payment may be negative: for instance the security may be insurance policies. At the start of each period, investors indicate their interest by bidding for the stock. There is zero net supply, investors can long or short the stock and the Walrasian auctioneer sets the market clearing price.

The period between one payment and the next is referred to as the 'observation' period. In the one period model, the security is extinguished once the payment is made so the payment must include any return of capital. Another security of identical characteristics comes into being after the payment is made. In the multiperiod model, the payment is a dividend and the security continues on.

The market for the security consists of J investors each of whom derive their own model of security return by carrying out OLS regression on observed security returns using market price and variables of their own choosing. Each investor uses the regression coefficients which he or she has estimated to predict future returns.

The sequence of observation periods is divided up into longer segments, each T observation periods long, which are referred to as 'estimation' periods. The division of observation periods into estimation periods is the same for every investor. At the end of each estimation period, a *small proportion* of investors re-estimate their model using the T observations made in the period. They replace their old coefficient estimates with the fresh estimates and use them henceforth. The other investors keep using the coefficient estimates they already have. The only time coefficient estimates are updated is at the end of each estimation period; for all observation periods within a particular estimation period the coefficient estimates are kept the same.

The least squares learning assumption may seem to be unduly restrictive, but virtually any method used in practice to estimate stock market returns can be rendered at least approximately in OLS form:

- Measuring the return of discrete categories is equivalent to the use of a dummy variable in a regression.
- Applying a filter rule (a rule such that investment takes place if certain conditions are met) is equivalent to the use of discrete categories.
- Non-linear relationships can be captured by including powers of a variable in the regression.
- Mean reversion models, which cover standard charting approaches, can be cast into OLS form.
- The approach embraces multiple regression, not just simple regression.

Premise 2: data set. Investors use different data, but the data set $\mathbf{X}^{original}_{T*N\ orig}$ (covering T observation periods for a set of $N^{original}$ variables) is a compilation which lists every variable used by every investor, probably with repetition. Some of this data may have been derived by first differencing a primary series. Also available is the set of absolute returns for the security $\mathbf{y}_{T*1}$ which is the same for every investor. Each item of data is available at the start of the observation period and the corresponding return is measured at the end of the observation period.

The initial data set $\mathbf{X}^{original}$ ($rank = N$) can be broken into a set of independent variables $\mathbf{X}$ which are sufficient to capture every independent component of the variables in $\mathbf{X}^{original}$; the rank of $\mathbf{X}$ is also N. Of course $N \le T$. These independent variables are orthogonal so the underlying variance of $\mathbf{X}$ is given by:

$$E\left[\mathbf{X}'\mathbf{X}\right] = \mathbf{I}_{N*N} \tag{1}$$

and it is supposed with some degree of approximation that the sample variance also exhibits:

$$\mathbf{X}'\mathbf{X} = \mathbf{I}_{N*N} \tag{2}$$

Premise 3: investor strategies. The specific information which an investor uses is referred to as their strategy. Each investor j is assumed to employ jK different variables in their attempt to explain $\mathbf{y}$.

$$0 \le jK \le N \tag{3}$$

Given the data transformation specified above, the variables are not confined to the exact variables in $\mathbf{X}$ but may include linear functions of these variables. Each variable $\mathbf{X}_{jk}$ which investor *j* uses is derived from the data set using an $N*1$ strategy vector $\mathbf{a}_{jk}$ via:

$$\underset{T*1}{\mathbf{X}_{jk}} = \underset{T*N}{\mathbf{X}} \cdot \underset{N*1}{\mathbf{a}_{jk}} \tag{4}$$

The $\mathbf{a}_{jk}$ vectors can be assembled into one strategy matrix $\mathbf{a}_j = \begin{bmatrix}\mathbf{a}_{j1} & & \mathbf{a}_{jK}\end{bmatrix}$ so that the data set $\mathbf{X}_j$ used by the investor *j* is given by:

$$\underset{T*jK}{\mathbf{X}_j} = \mathbf{X}_{T*N} \cdot \underset{N*jK}{\mathbf{a}_j} \tag{5}$$

We assume that each investor uses independent variables in their regressions, so that the strategy matrix $\mathbf{a}_j$ is of full rank:

$$rank\left(\mathbf{a}_j\right) = jK \leq N \tag{6}$$

Putting together the $\mathbf{a}_j$ matrices for every investor to give one matrix $\mathbf{a}$ gives:

$$\mathbf{X}\mathbf{a}_{N*sum(jK)} = \mathbf{X}^{original} \qquad \text{where of course} \quad sum(jK) = N^{orig} \tag{7}$$

Given that $rank(\mathbf{X}^{original}) = rank(\mathbf{X}) = N$ it follows that:

$$rank(\mathbf{a}) = N \tag{8}$$

or in other words, the set of strategy vectors $\left\{\mathbf{a}_{jk}\right\}$ span the space $\mathbb{R}^N$.

Premise 4: use of particular regressors. In addition to the data in data set $\mathbf{X}$, at least one investor includes the price of the security $\mathbf{p}_{T*1}$ as a variable in their data set $\mathbf{X}_j$. Price is not included as part of another variable $\mathbf{X}_{jk}$ but only in its own right.

The regression coefficients for non-price data are denoted $\hat{\beta}$. Because of the special role of price, its regression coefficient is denoted by $\hat{\rho}$.

Premise 5: generation of returns. It is assumed that the return per period per unit (one share, not one dollar's worth) of the security is generated according to:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \mathbf{u} \quad \text{where} \;\; \mathbf{X}'\mathbf{X} = \mathbf{I}_{N*N}, \quad \mathrm{E}(\mathbf{u}) = 0, \quad \mathrm{E}\left[\mathbf{u}\mathbf{u}'\right] = \sigma^2 \mathbf{I}_{N*N} \tag{9}$$

The total variance $\mathrm{E}\left[\mathbf{u}'\mathbf{u}\right] = \sigma^2 \boldsymbol{\iota}'_{1*N} \cdot \mathbf{I}_{N*N} \boldsymbol{\iota}_{N*1} = N\sigma^2$ is denoted σ^2_{tot}.

The assumption that the mean is zero is not restrictive as OLS will automatically generate this result when an intercept is included (although it has not been assumed that the constant vector $\boldsymbol{\iota}$ occurs in $\mathbf{X}$). This relation is an empirically observed correspondence for the particular data set $\mathbf{X}$ which does not purport to be causative; if investors had a different data set available to them then a different relation would apply. Nonetheless we suppose that the return vector $\boldsymbol{\mu}$ has the character of a parameter in that it is stable over time.

Although the variance of the dividend process is finite this does not rule out an infinite variance for security return (which includes price fluctuations) in the multiperiod model.

Premise 6: Independent re-estimation. As stated, at the end of a particular estimation period a small minority of investors take the opportunity to upgrade their estimates using

X. These investors are referred to as 'new investors', although it is the estimates which are new and not the investors themselves. The majority of investors ('old investors') continue to use the estimates they already have. The proportion of this minority to the whole is denoted *w*, and is drawn randomly from the ranks of the investors. Thus some investors may re-estimate immediately, others may not re-estimate for a long time. The point of this independence assumption is that the average coefficient estimates of the old investors remains the same into the next estimation period.

The process can also be interpreted as an geometric adaptive expectations process in which the weight of new estimates is *dw*.

Premise 7: Demand proportional to expected return. A standard Markowitz portfolio model shows that given a constant level of risk, demand for a particular security is a linear function of the expected return on that security. For simplicity it is assumed here that that there is no constant, i.e.

$$q_j = B_j\left(\frac{\hat{r}_j}{p}\right) \tag{10}$$

where q_j is the amount in dollar terms of the security demanded by investor *j*, $\hat{r}_j$ is the net absolute return which the investor predicts they will receive on each unit (share) of the security, and *p* is the price of the security. $\frac{\hat{r}_j}{p}$ is the predicted net percentage return, and B_j is the amount of stock, in dollar terms, demanded by investor *j* per unit of net percentage return: the "weight of money" brought to bear by investor *j*. It is always positive.

It is helpful to define b_j , the relative weight of money for investor *j*, by $\frac{B_j}{\sum_j B_j}$, observing the denominator must be positive. The sum of relative weights is unity. Because there is no constant, a non zero quantity q_j requires that expectation $\hat{r}_j$ be non-zero, so any active investor must analyze at least one useful data series to generate a non-zero expectation.

2.2. Price theorem

The market for the security is cleared in an observation period when the net quantity is zero.

Result 1. The weighted predicted return is zero.

$$\sum_j b_j \hat{r}_j = 0 \tag{11}$$

This follows immediately as

$$0 = \sum_j q_j \tag{12}$$

$$= \sum_j B_j \frac{\hat{r}_j}{p} \tag{13}$$

$$= \sum_j b_j \hat{r}_j \quad \text{multiplying by } \frac{p}{\sum_j B_j} \text{ which is positive} \qquad \# \qquad (14)$$

Strictly speaking, this should be developed through a formal assumption that some of the information in the data set **X** is relevant (i.e. $\boldsymbol{\mu} \neq 0$), and consequently it is possible for some investor to form a non-zero expectation of return and cause price to be non-zero. This development is omitted for brevity.

Result 2: Price theorem. Price can be expressed as a function $\boldsymbol{\pi}$ of **X**, i.e.:

$$\underset{T*1}{\mathbf{p}} = \underset{T*N}{\mathbf{X}} \ \underset{N*1}{\boldsymbol{\pi}} \qquad (15)$$

where $\boldsymbol{\pi} = -\frac{\sum_j b_j \mathbf{a}_j \hat{\boldsymbol{\beta}}_{jX}}{\rho}$ is referred to as the price equation, (16)

and $\rho = \sum_j b_j \hat{\rho}_j$ is referred to as the price coefficient. (17)

Proof: For each investor,

$$\hat{\mathbf{r}}_j = \begin{bmatrix} \mathbf{X}_j & \mathbf{p} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_j \\ \hat{\rho}_j \end{bmatrix} \qquad (18)$$

$$= \mathbf{X}\mathbf{a}_j \hat{\boldsymbol{\beta}}_{jX} + \mathbf{p}\hat{\rho}_j \qquad \text{expanding and using } \mathbf{X}_j = \mathbf{X}\mathbf{a}_j \qquad (19)$$

$$\sum_j b_j \left(\mathbf{X}\mathbf{a}_j \hat{\boldsymbol{\beta}}_{jX} + \mathbf{p}\hat{\rho}_j \right) = 0 \qquad \text{substituting into (11)} \qquad (20)$$

$$\mathbf{p} = -\frac{\mathbf{X}\sum_j b_j \mathbf{a}_j \hat{\boldsymbol{\beta}}_{jX}}{\sum_j b_j \hat{\rho}_j} \qquad \text{making } \mathbf{p} \text{ the subject} \qquad (21)$$

$$= \mathbf{X}\boldsymbol{\pi} \qquad \# \qquad (22)$$

As trivial as this result appears it is of fundamental importance. If investors form expectations using a certain data set then price must be a function of that data set. Because price has a stable relationship to the data then data does not matter in a sense and the problem can be abstracted to the coefficients of data.

This functional form is constant as long as the estimations in use are unchanged, i.e. throughout the estimation period. This result may suggest that if an investor has access to all the information **X**, they could estimate the price coefficient $\boldsymbol{\pi}$ perfectly. However there is no reason for an investor to do this; explaining current price is not the same thing as predicting return which depends on dividends and future price.

Price is the ratio of estimates. If coefficient estimates were independently and normally distributed then price would have a Cauchy distribution with unstable mean and infinite variance. In fact coefficient estimates are not independently distributed but this result points the way to a theoretical basis for studies of stock return.

If the net quantity Q to be traded is positive instead of zero then (21) becomes

$$\mathbf{p} = -\frac{\mathbf{X}\sum_j b_j \mathbf{a}_j \hat{\boldsymbol{\beta}}_{jX}}{\rho} + \frac{Q}{\rho} \tag{23}$$

and this term $\frac{Q}{\rho}$ will be negative as we expect the price coefficient ρ to be negative. This is why initial public offerings must be floated at a lower price than the security subsequently trades.

Define an augmented form of the strategy matrix $\mathbf{a}_j$ which includes price, denoted $\tilde{\mathbf{a}}_{j0}$.

$$\begin{bmatrix}\mathbf{X}_j & \mathbf{p}_0\end{bmatrix} = \begin{bmatrix}\mathbf{X}\mathbf{a}_j & \mathbf{X}\boldsymbol{\pi}_0\end{bmatrix} = \mathbf{X}\begin{bmatrix}\mathbf{a}_j & \boldsymbol{\pi}_0\end{bmatrix} = \mathbf{X}\tilde{\mathbf{a}}_{j0} \tag{24}$$

If investor j does not use price then $\tilde{\mathbf{a}}_j = \mathbf{a}_j$ and if investor j does not use non-price data then $\tilde{\mathbf{a}}_j = \boldsymbol{\pi}$. The number of columns in $\tilde{\mathbf{a}}_j$ is denoted jK^{aug}. If price is not included, $jK^{aug} = jK$; if price is included $jK^{aug} = jK + 1$. We take it that if the investor includes $\boldsymbol{\pi}$ in the strategy matrix, it is independent of the other variables.

$$rank\left(\tilde{\mathbf{a}}_j\right) = jK^{aug} \tag{25}$$

3. The single period coefficient model

3.1. New coefficients

As described in Assumption 6 above, the investors are divided into *old* and *new*, and variables $b_{j\,new}, b_{j\,old}$ are introduced which represent the investor's weight of money just within the category into which they fall.

$$\sum_j b_{j\,new} = 1 \tag{26}$$

$$\sum_j b_{j\,old} = 1 \tag{27}$$

The relative weight of money of all the new investors combined is denoted *w*, and so:

$$w\sum_j b_{j\,new} + (1-w)\sum_j b_{j\,old} = 1 \tag{28}$$

The estimates of the new investors are obtained from data set $\mathbf{X}_0$ for the estimation period 0. Let $\mathbf{y}_{T*1}$ be the net absolute return (not percentage), given by the gross return less price paid. By Assumption 5,

$$\mathbf{y}_0 = \mathbf{X}_0\boldsymbol{\mu} - \mathbf{p}_0 + \mathbf{u}_0 \tag{29}$$

$$= \mathbf{X}_0\boldsymbol{\mu} - \mathbf{X}_0\boldsymbol{\pi} + \mathbf{u}_0 \qquad \text{substituting for } \mathbf{p} \text{ using (22)} \tag{30}$$

Result 3: estimation.

$$\hat{\boldsymbol{\beta}}_{j\,new} = \left(\tilde{\mathbf{a}}_{j0}'\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}'\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right) \quad \text{where} \tag{31}$$

$$\mathbf{e}_{N*1} = \left(\mathbf{X}_0'\mathbf{X}_0\right)^{-1}\mathbf{X}_0'\mathbf{u}_0, \quad \mathrm{E}(\mathbf{e}) = \mathbf{0}, \quad \mathrm{E}\left[\mathbf{e}\mathbf{e}'\right] = \sigma^2\mathbf{I}_{N*N} \tag{32}$$

Proof: Assuming that investor *j* uses both **X** and **p** in their regressors, we get:

$$\begin{bmatrix}\hat{\boldsymbol{\beta}}_{jx\ new}\\ \hat{\boldsymbol{\beta}}_{j\pi\ new}\end{bmatrix} = \left(\begin{bmatrix}\mathbf{X}_{0j} & \mathbf{p}_0\end{bmatrix}'\begin{bmatrix}\mathbf{X}_{0j} & \mathbf{p}_0\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{X}_{0j} & \mathbf{p}_0\end{bmatrix}'\mathbf{y}_0 \quad \text{standard OLS formula} \qquad (33)$$

$$= \left(\begin{bmatrix}\mathbf{a}_j & \boldsymbol{\pi}_0\end{bmatrix}'\mathbf{X}_0{}'\mathbf{X}_0\begin{bmatrix}\mathbf{a}_j & \boldsymbol{\pi}_0\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{a}_j & \boldsymbol{\pi}_0\end{bmatrix}'\mathbf{X}_0{}'\mathbf{y}_0 \qquad \text{by (5),(15)} \qquad (34)$$

$$= \left(\tilde{\mathbf{a}}_{j0}{}'\mathbf{X}_0{}'\mathbf{X}_0\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}{}'\mathbf{X}_0{}'\left(\mathbf{X}_0\boldsymbol{\mu} - \mathbf{X}_0\boldsymbol{\pi}_0 + \mathbf{u}_0\right) \qquad \text{by (24) and (30)} \qquad (35)$$

$$= \left(\tilde{\mathbf{a}}_{j0}{}'\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}{}'\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right) \qquad (36)$$

Similarly if the investor omits **X** or **p** from their regressors.

Observe $\mathbf{e}_0 = \mathbf{X}_0{}'\mathbf{u}_0$, so $E\left[\mathbf{e}_0\mathbf{e}_0{}'\right] = E\left[\mathbf{X}_0{}'\mathbf{u}_0\mathbf{u}_0{}'\mathbf{X}_0\right] = \sigma^2\mathbf{I}_{N*N}$ # (37)

We see the interesting result that regression of return against data is equivalent to regression of the underlying parameters $\boldsymbol{\mu} - \boldsymbol{\pi}$ against $\tilde{\mathbf{a}}_{j0}$, the coefficients used by investor i to derive their variables. The investor's problem in a sense does not involve data but rather is a test of how well the investor's variable coefficients capture the return coefficients $\boldsymbol{\mu} - \boldsymbol{\pi}$. We have abstracted the problem from 'data space' $\mathbb{R}^T$ to 'coefficient space' $\mathbb{R}^N$; this is a substantial gain in simplicity. For this reason the model developed here is referred to as the 'coefficient model'. ('Genetic', 'constructive' or even 'memetic' model are other possibilities.)

3.2. Updating coefficients

The average regression coefficient for investor type *j* is a weighted average of new and old regression coefficients:

$$\hat{\boldsymbol{\beta}}_{j1} = \frac{\sum_{all} b_j\hat{\boldsymbol{\beta}}_j}{\sum_{all} b_j} \qquad (38)$$

$$= \frac{(1-dw)\,b_{j\,old}\hat{\boldsymbol{\beta}}_{j0} + dw\cdot b_{j\,new}\hat{\boldsymbol{\beta}}_{j\,new}}{(1-dw)\,b_{j\,old} + dw\cdot b_{j\,new}} \qquad (39)$$

$$= (1-dw)\hat{\boldsymbol{\beta}}_{j0} + dw\cdot\hat{\boldsymbol{\beta}}_{j\,new} \qquad \text{by Assumption 6} \quad b_{j\,old} = b_{j\,new} \qquad (40)$$

The new price regression coefficient $\hat{\rho}$ uses every investor type *j* who includes price in their regression and because this is multiple regression, the price coefficients $\hat{\rho}_j$ may differ between investor types. Define:

$$b_{price} = \sum_{all} b_j \qquad (41)$$

Then $b_{price} = \sum_{new} b_{j\,new} = \sum_{old} b_{j\,old}$ by Assumption 6 (42)

Define $\hat{\rho}_{new} = \dfrac{\sum_{new} b_{j\,new}\hat{\rho}_{j\,new}}{\sum_{new} b_{j\,new}}$ (43)

and $$\hat{\rho} = \frac{\sum_{all} b_j \hat{\rho}_j}{\sum_{all} b_j} \tag{44}$$

Then $$\hat{\rho}_1 = \frac{(1-dw)\sum_{old} b_{j\,old}\hat{\rho}_{j0} + dw\sum_{new} b_{j\,new}\hat{\rho}_{j\,new}}{b_{price}} \quad \text{expanding (44) and using (41)} \tag{45}$$

$$= (1-dw)\,\hat{\rho}_0 + dw \cdot \hat{\rho}_{new} \quad \text{by (43),(44)} \tag{46}$$

The price coefficient ρ used in the price equation (16) includes the weighting:

$$\rho_1 = \sum_j b_j \hat{\rho}_{j1} \tag{47}$$

$$= b_{price}\hat{\rho}_1 \quad \text{rearranging definition (44)} \tag{48}$$

$$= b_{price}(1-dw)\,\hat{\rho}_0 + b_{price} dw \cdot \hat{\rho}_{new} \quad \text{by (46)} \tag{49}$$

ρ is smaller than $\hat{\rho}$ because in general $b_{price} < 1$.

3.3. Price change theorem

We determine how the new return estimates derived in estimation period 0 affects the price in estimation period 1. We are now dealing with data set $\mathbf{X}_1$. The change in price $\mathbf{dp}$ is not measured relative to the prices $\mathbf{p}_0$ in estimation period 0 obtained with data $\mathbf{X}_0$, but relative to the price which would obtain with current data set $\mathbf{X}_1$ if there were no new estimates. Data set $\mathbf{X}_1$ is therefore taken as constant for purposes of differentiation.

Result 4: price change theorem.

$$\mathbf{d\pi} = -\frac{dw}{\rho_0}\mathbf{H}_0\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right) \tag{50}$$

where $$\mathbf{H}_{N*N} = \sum_{j\,new} b_{j\,new}\tilde{\mathbf{a}}_{j0}\left(\tilde{\mathbf{a}}_{j0}'\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}' \tag{51}$$

is referred to as the "estimation matrix" and

$$\rho_0 = \sum_{j\,old} b_{j\,old}\hat{\rho}_{j\,old} \tag{52}$$

where subscript 0 refers to period 0, is referred to as the "price coefficient".

Proof:

$$0 = \sum_j b_j \hat{\mathbf{r}}_j \quad \text{by Result 2.1} \tag{53}$$

$$= w\sum_{j\,new} b_{j\,new}(\mathbf{X}_{1j}\hat{\boldsymbol{\beta}}_{j\,new} + \mathbf{p}_1\hat{\rho}_{j\,new}) + (1-w)\sum_{j\,old} b_{j\,old}(\mathbf{X}_{1j}\hat{\boldsymbol{\beta}}_{j\,old} + \mathbf{p}_1\hat{\rho}_{j\,old}) \tag{54}$$

dividing the investors into *new* and *old* as per (28). Differentiate with respect to variables *w*,**p**:

$$\begin{aligned}\mathbf{0} = dw\sum_{j\,new} b_{j\,new}\mathbf{X}_{1j}\hat{\boldsymbol{\beta}}_{j\,new} - dw\sum_{j\,old} b_{j\,old}\mathbf{X}_{1j}\hat{\boldsymbol{\beta}}_{j\,old} \\ + dw\sum_{j\,new} b_{j\,new}\mathbf{p}_1\hat{\rho}_{j\,new} - dw\sum_{j\,old} b_{j\,old}\mathbf{p}_1\hat{\rho}_{j\,old} \\ + \mathbf{dp}\cdot w\sum_{j\,new} b_{j\,new}\hat{\rho}_{j\,new} + \mathbf{dp}\cdot(1-w)\sum_{j\,old} \mathbf{b}_{j\,old}\hat{\rho}_{j\,old}\end{aligned} \tag{55}$$

This can be evaluated at the point $w=0$ (*new* investors are incremental); at this point

$$\mathbf{p}_1 = \mathbf{X}_1\boldsymbol{\pi}_0 \quad \text{and} \tag{56}$$

$$\sum_{j\,old} b_{j\,old}\mathbf{X}_{1j}\hat{\boldsymbol{\beta}}_{j\,old} + \sum_{j\,old} b_{j\,old}\mathbf{p}_1\hat{\rho}_{j\,old} = \mathbf{0} \text{ by (54)} \tag{57}$$

Making these substitutions into (55) yields:

$$\mathbf{0} = dw\sum_{j\,new} b_{j\,new}\mathbf{X}_{1j}\hat{\boldsymbol{\beta}}_{j\,new} + dw\sum_{j\,new} b_{j\,new}\mathbf{p}_1\hat{\rho}_{j\,new} + \mathbf{dp}\sum_{j\,old} b_{j\,old}\hat{\rho}_{j\,old} \tag{58}$$

$$= dw\sum_{j\,new} b_{j\,new}\begin{bmatrix}\mathbf{X}_1\mathbf{a}_j & \mathbf{X}_1\boldsymbol{\pi}_0\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\beta}}_{j\,new} \\ \hat{\rho}_{j\,new}\end{bmatrix} + \mathbf{dp}\sum_{j\,old} b_{j\,old}\hat{\rho}_{j\,old} \text{ by (56)} \tag{59}$$

$$= dw\sum_{j\,new} b_{j\,new}\mathbf{X}_1\tilde{\mathbf{a}}_{j0}\hat{\boldsymbol{\beta}}_{j\,new} + \mathbf{dp}\sum_{j\,old} b_{j\,old}\hat{\rho}_{j\,old} \qquad \text{where } \hat{\boldsymbol{\beta}}_{j\,new} = \begin{bmatrix}\hat{\boldsymbol{\beta}}_{j\,new} \\ \hat{\rho}_{j\,new}\end{bmatrix} \tag{60}$$

$$= dw\sum_{j\,new} b_{j\,new}\mathbf{X}_1\tilde{\mathbf{a}}_{j0}\left(\tilde{\mathbf{a}}_{j0}{}'\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}{}'\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right) + \mathbf{X}_1\mathbf{d}\boldsymbol{\pi}\sum_{j\,old} b_{j\,old}\hat{\rho}_{j\,old} \tag{61}$$

$$\text{by Result 3.1 and noting } \mathbf{dp} = \mathbf{X}_1\mathbf{d}\boldsymbol{\pi} \tag{62}$$

Now this statement is true for all observations $\mathbf{x}_{1*N}$ which form the rows of dataset **X**, and therefore is true of the coefficients alone:

$$\mathbf{0} = dw\sum_{j\,new} b_{j\,new}\tilde{\mathbf{a}}_{j0}\left(\tilde{\mathbf{a}}_{j0}{}'\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}{}'\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right) + \mathbf{d}\boldsymbol{\pi}\sum_{j\,old} b_{j\,old}\hat{\rho}_{j\,old} \tag{63}$$

$$= dw\cdot\mathbf{H}\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right) + \mathbf{d}\boldsymbol{\pi}\hat{\rho}_0 \qquad \# \tag{64}$$

This theorem may appear to be a trivial technical result, but it is contended here that it underlies the efficiency of financial markets. Note what it does not say – the price change is given by the *change* in expectation, i.e.

$$\mathbf{d}\boldsymbol{\pi} = -\frac{dw}{\rho_0}\sum b_j\mathbf{a}_j\left(\hat{\boldsymbol{\beta}}_{new} - \hat{\boldsymbol{\beta}}_{old}\right) \qquad \text{based on (60)} \tag{65}$$

In this case, prices would simply be given by $\sum_{new\ and\ old} b_j\mathbf{a}_j\hat{\boldsymbol{\beta}}_j$ - a weighted sum of everyone's expectation. Here the change in price is determined by the new estimates alone, and so the price is pushed in the direction of the underlying return parameter $\boldsymbol{\mu}$.

It is worth validating the price change formula by comparing it with a straight differentiation of price with respect to price coefficient, on the assumption that only the price coefficient changes. (22) can be rewritten as:

$$\boldsymbol{\pi} = -\frac{\sum_j b_j \mathbf{a}_j \hat{\boldsymbol{\beta}}_{jX}}{\rho_0} \tag{66}$$

so differentiation of price gives:

$$\mathbf{d\pi} = \frac{\sum_j b_j \mathbf{a}_j \hat{\boldsymbol{\beta}}_{jX}}{\rho_0^2} \cdot d\rho \tag{67}$$

$$= -\frac{\boldsymbol{\pi}_0}{\rho_0} \cdot (b_{price} dw \cdot \hat{\rho}_{new}) \qquad \text{using (66) and } d\rho = b_{price} dw \cdot \hat{\rho}_{new} \text{ from (49)} \tag{68}$$

and if investors only reestimate price then price change theorem yields:

$$\mathbf{d\pi} = -\frac{dw}{\rho_0}\left(\sum_{j\,new} b_{j\,new} \tilde{\mathbf{a}}_{j0} \left(\tilde{\mathbf{a}}_{j0}{}' \tilde{\mathbf{a}}_{j0}\right)^{-1} \tilde{\mathbf{a}}_{j0}{}' \left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right)\right) \qquad \text{by (50),(51)} \tag{69}$$

$$= -\frac{dw}{\rho_0} b_{price} \boldsymbol{\pi}_0 \hat{\rho}_{new} \quad \text{using (42), } \tilde{\mathbf{a}}_{j0} = \boldsymbol{\pi}_0, \ \hat{\rho}_{new} = \left(\tilde{\mathbf{a}}_{j0}{}' \tilde{\mathbf{a}}_{j0}\right)^{-1} \tilde{\mathbf{a}}_{j0}{}' \left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right) \tag{70}$$

which is indeed identical.

3.4. Properties of the estimation matrix

It is useful to introduce an orthogonalized and normalized version, $\boldsymbol{\alpha}_j$, of the augmented strategy matrix $\tilde{\mathbf{a}}_j$.

Result 5: normalized strategy matrix. The matrix $\tilde{\mathbf{a}}_{j0}\left(\tilde{\mathbf{a}}_{j0}{}'\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}{}'$ can be expressed as $\boldsymbol{\alpha}_{j0}\boldsymbol{\alpha}'_{j0}$ where $\boldsymbol{\alpha}_{j0}$ is an $N * jK^{aug}$ matrix of orthogonalized and normalized strategy vectors.

Proof: The matrix $\tilde{\mathbf{a}}_{j0}\left(\tilde{\mathbf{a}}_{j0}{}'\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}{}'$ is real symmetric so can be represented as $\boldsymbol{\Lambda\lambda\Lambda}'$ where $\boldsymbol{\lambda}$ is a diagonal matrix and $\boldsymbol{\Lambda}$ consists of real orthogonal eigenvectors. It is idempotent so all the eigenvalues must be 0 or 1. Multiplying by $\tilde{\mathbf{a}}_{j0}$ on either side of the matrix gives $\tilde{\mathbf{a}}_{j0}{}'\tilde{\mathbf{a}}_{j0}$ which has rank jK^{aug} so the matrix must have rank jK^{aug} not less. It therefore has jK^{aug} unity eigenvalues. Without loss of generality, place the unity eigenvalues and corresponding eigenvectors first in the matrix. So

$$\tilde{\mathbf{a}}_{j0}\left(\tilde{\mathbf{a}}_{j0}{}'\tilde{\mathbf{a}}_{j0}\right)^{-1}\tilde{\mathbf{a}}_{j0}{}' = \begin{bmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_2\end{bmatrix}\begin{bmatrix}\mathbf{I}_{jKaug*jKaug} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_1{}'\\ \boldsymbol{\Lambda}_2{}'\end{bmatrix} \tag{71}$$

$$= \boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1{}' \qquad \text{Take } \boldsymbol{\alpha}_{j0} = \boldsymbol{\Lambda}_1. \qquad \# \tag{72}$$

Observe that:

$$\boldsymbol{\alpha}_{j0}{}'\boldsymbol{\alpha}_{j0} = \boldsymbol{\Lambda}_1{}'\boldsymbol{\Lambda}_1 = \mathbf{I}_{jKaug*jKaug} \tag{73}$$

as required. The estimation matrix $\mathbf{H}$ can be expressed as:

$$\mathbf{H}_0 = \sum_{j\,new} b_{j\,new} \boldsymbol{\alpha}_{j0} \boldsymbol{\alpha}_{j0}' \tag{74}$$

It is also useful to partition $\boldsymbol{\alpha}_j$ into columns:

$$\boldsymbol{\alpha}_j = \begin{bmatrix} \mathbf{o}_{j1} & \mathbf{o}_{j2} & \end{bmatrix} \tag{75}$$

Result 6. The estimation matrix **H** is (i) can be expressed as $\boldsymbol{\Lambda}_H \boldsymbol{\lambda}_H \boldsymbol{\Lambda}_H'$ where $\boldsymbol{\Lambda}_H$ is an N*N matrix of orthogonal eigenvectors and $\boldsymbol{\lambda}_H$ is an $N*N$ matrix of positive eigenvalues (ii) is positive definite.

Proof: By (8), $\{\mathbf{a}_{jk}\}$ spans the space $\mathbb{R}^N$. So for all non-zero N*1 vectors **x**, there exists an N*1 vector $\mathbf{a}_{Jk}$ such that:

$$\mathbf{x}'\mathbf{a}_{Jk} \neq 0 \tag{76}$$

and corresponding normalized strategy matrix $\boldsymbol{\alpha}_{J0}$ which uses $\mathbf{a}_{Jk}$ such that:

$$\mathbf{x}_{N*1} = \boldsymbol{\alpha}_{J0} \boldsymbol{\beta}_{JK*1} + \boldsymbol{\varepsilon}_{N*1}, \tag{77}$$

where $\boldsymbol{\beta} \neq \mathbf{0}_{JK*1}$ (78)

and $\boldsymbol{\varepsilon}_{1*N}{}' \boldsymbol{\alpha}_{N*jKaug} = \mathbf{0}_{1*jKaug}$ (79)

Now $$\mathbf{x}'\mathbf{H}_0\mathbf{x} = \mathbf{x}'\left(\sum_{j\,new} b_{j\,new} \boldsymbol{\alpha}_{j0} \boldsymbol{\alpha}_{j0}'\right)\mathbf{x} \tag{80}$$

$$\geq b_{J\,new} \mathbf{x}'\boldsymbol{\alpha}_{J0} \boldsymbol{\alpha}_{J0}'\mathbf{x} \qquad \boldsymbol{\alpha}_{j0}\boldsymbol{\alpha}_{j0}' \text{ is non-negative definite} \tag{81}$$

$$= b_{J\,new}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}_{J0}' + \boldsymbol{\varepsilon}'\right)\boldsymbol{\alpha}_{J0}\boldsymbol{\alpha}_{J0}'\left(\boldsymbol{\alpha}_{J0}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \qquad \text{by (77)} \tag{82}$$

$$= b_{J\,new}\left(\boldsymbol{\alpha}_{J0}'\boldsymbol{\alpha}_{J0}\boldsymbol{\beta}\right)_{1*JK}'\left(\boldsymbol{\alpha}_{J0}'\boldsymbol{\alpha}_{J0}\boldsymbol{\beta}\right)_{JK*1} \qquad \text{by (79)} \tag{83}$$

$$> \mathbf{0} \text{ as } \boldsymbol{\alpha}_{J0}'\boldsymbol{\alpha}_{J0} \text{ is full rank so multiplying by } \boldsymbol{\beta} \neq \mathbf{0} \text{ gives non-zero product.} \tag{84}$$

Further **H** is real symmetric so can be represented as $\boldsymbol{\Lambda\lambda\Lambda}'$ where $\boldsymbol{\lambda}$ is a diagonal matrix and $\boldsymbol{\Lambda}$ consists of real orthogonal eigenvectors. Since **H** is positive definite, $\boldsymbol{\lambda}$ must be positive diagonal. #

Result 7. The eigenvalues $\boldsymbol{\lambda}$ of the estimation matrix **H** are such that

$$0 < \boldsymbol{\lambda} \leq 1 \tag{85}$$

Proof: Let **v** be the eigenvector of $\mathbf{H}_0$ corresponding to eigenvalue λ.

$$\lambda^2 \mathbf{v}'\mathbf{v} = \mathbf{v}'\mathbf{H}_0{}'\mathbf{H}_0\mathbf{v} \qquad \text{from definition of eigenvector} \tag{86}$$

$$= \sum_i \sum_j b_i b_j \hat{\mathbf{v}}_i{}'\hat{\mathbf{v}}_j \qquad \text{given } \mathbf{H}_0\mathbf{v} = \sum_{j\,new} b_{j\,new}\boldsymbol{\alpha}_{j0}\boldsymbol{\alpha}_{j0}'\mathbf{v} \tag{87}$$

$$= \sum_{j\,new} b_{j\,new}\hat{\mathbf{v}}_j \tag{88}$$

$$\leq \sum_i \sum_j b_i b_j \mathbf{v}'\mathbf{v} \qquad \text{given } \hat{\mathbf{v}}_i{}'\hat{\mathbf{v}}_j = \left|\hat{\mathbf{v}}_i\right| \cdot \left|\hat{\mathbf{v}}_j\right| \cdot \cos\theta \tag{89}$$

$$\leq |\mathbf{v}| \cdot |\mathbf{v}| \quad \text{since } \left|\hat{\mathbf{v}}_j\right| \leq |\mathbf{v}| \tag{90}$$

$$= \mathbf{v}'\mathbf{v} \qquad \text{noting } \sum_i \sum_j b_i b_j = 1 \tag{91}$$

Hence $\lambda^2 \le 1$ dividing both sides by $\mathbf{v}'\mathbf{v} = 1$ # (92)

3.5. The Efficient Market Theorem

We introduce two working assumptions which are replaced by an endogenous stability condition in section 4 below:
Assumption 1: negative price coefficient. The price coefficient ρ is negative and less than the expression $-dw \cdot \lambda_H^{\max}$.

$$\rho < -dw \cdot \lambda_H^{max} \tag{93}$$

Assumption 2: independence price coefficient. The price coefficient ρ is independent of the current price $\boldsymbol{\pi}$.

The estimation matrix **H** is a function of the augmented strategy matrix $\tilde{\mathbf{a}}_j$ which is a function of the price vector $\boldsymbol{\pi}$. Since price varies, it is useful to develop an approximation for $\mathbf{H}(\boldsymbol{\mu}-\boldsymbol{\pi})$.

Lemma 8. A first order approximation of $\mathbf{H}(\boldsymbol{\mu}-\boldsymbol{\pi})$ is $\mathbf{H}_\mu(\boldsymbol{\mu}-\boldsymbol{\pi})$, where $\mathbf{H}_\mu$ denotes **H** evaluated at $\boldsymbol{\pi} = \boldsymbol{\mu}$.

Proof: $\mathbf{H} = \sum_{j\,new} b_j \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j{}'$ by (74) (94)

$$= \sum_j b_j \begin{bmatrix} \mathbf{o}_{j1} & \mathbf{o}_{j2} & \end{bmatrix} \begin{bmatrix} \mathbf{o}_{j1}{}' \\ \mathbf{o}_{j2}{}' \\ \\ \end{bmatrix} \qquad \text{by (75)} \tag{95}$$

$$= \sum_j b_j \sum_k \mathbf{o}_{jk} \mathbf{o}_{jk}{}' \tag{96}$$

Differentiate $\mathbf{H}(\boldsymbol{\mu}-\boldsymbol{\pi})$ with respect to $\boldsymbol{\pi}$ (in the following, $\mathbf{c}_1$ and $\mathbf{c}_2$ are constants):

$$\frac{\partial \mathbf{H}(\boldsymbol{\mu}-\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} = \sum_j b_j \sum_k \frac{\partial}{\partial \boldsymbol{\pi}} \left(\mathbf{o}_{jk} \mathbf{o}_{jk}{}' (\boldsymbol{\mu}-\boldsymbol{\pi}) \right) \tag{97}$$

$$\frac{\partial}{\partial \boldsymbol{\pi}} \left(\mathbf{o}_{jk} \mathbf{o}_{jk}{}' (\boldsymbol{\mu}-\boldsymbol{\pi}) \right) = \frac{\partial \mathbf{o}_{jk}}{\partial \boldsymbol{\pi}} \left(\mathbf{o}_{jk}{}' (\boldsymbol{\mu}-\boldsymbol{\pi}) \right) + \frac{\partial}{\partial \boldsymbol{\pi}} \left(\mathbf{c}_1 \cdot \mathbf{o}_{jk}{}' \cdot \mathbf{c}_2 \right) \Big|_{\substack{\mathbf{c}_1 = \mathbf{o}_{jk} \\ \mathbf{c}_2 = \boldsymbol{\mu}-\boldsymbol{\pi}}} + \mathbf{o}_{jk} \mathbf{o}_{jk}{}' \frac{\partial}{\partial \boldsymbol{\pi}} (\boldsymbol{\mu}-\boldsymbol{\pi}) \tag{98}$$

$$= \frac{\partial \mathbf{o}_{jk}}{\partial \boldsymbol{\pi}} \left(\mathbf{o}_{jk}{}' (\boldsymbol{\mu}-\boldsymbol{\pi}) \right) + \frac{\partial}{\partial \boldsymbol{\pi}} \left(\mathbf{c}_1 \cdot \mathbf{c}_2{}' \cdot \mathbf{o}_{jk} \right) \Big|_{\substack{\mathbf{c}_1 = \mathbf{o}_{jk} \\ \mathbf{c}_2 = \boldsymbol{\mu}-\boldsymbol{\pi}}} + \mathbf{o}_{jk} \mathbf{o}_{jk}{}' (\mathbf{-I}) \quad \text{interchanging } \mathbf{c}_2, \mathbf{o}_{jk} \tag{99}$$

$$= \frac{\partial \mathbf{o}_{jk}}{\partial \boldsymbol{\pi}} \left(\mathbf{o}_{jk}{}' (\boldsymbol{\mu}-\boldsymbol{\pi}) \right) + \mathbf{o}_{jk} (\boldsymbol{\mu}-\boldsymbol{\pi})' \frac{\partial \mathbf{o}_{jk}}{\partial \boldsymbol{\pi}} - \mathbf{o}_{jk} \mathbf{o}_{jk}{}' \tag{100}$$

so
$$\frac{\partial \mathbf{H}(\boldsymbol{\mu}-\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} = \sum_j b_j \sum_k \left(\frac{\partial \mathbf{o}_{jk}}{\partial \boldsymbol{\pi}} \left(\mathbf{o}_{jk}{}' (\boldsymbol{\mu}-\boldsymbol{\pi}) \right) + \mathbf{o}_{jk} (\boldsymbol{\mu}-\boldsymbol{\pi})' \frac{\partial \mathbf{o}_{jk}}{\partial \boldsymbol{\pi}} - \mathbf{o}_{jk} \mathbf{o}_{jk}{}' \right) \tag{101}$$

and the required first order approximation is given by

$$\mathbf{H(\mu - \pi)} \approx \mathbf{H(\mu - \pi)}\Big|_{\boldsymbol{\pi}=\boldsymbol{\mu}} + \frac{\partial \mathbf{H(\mu - \pi)}}{\partial \boldsymbol{\pi}}\Bigg|_{\boldsymbol{\pi}=\boldsymbol{\mu}} (\boldsymbol{\pi} - \boldsymbol{\mu}) \tag{102}$$

noting the final factor is $\boldsymbol{\pi} - \boldsymbol{\mu}$, as per a Taylor series, not $\boldsymbol{\mu} - \boldsymbol{\pi}$

$$= \mathbf{0} + \sum_j b_j \sum_k \left(\mathbf{0} + \mathbf{0} - \mathbf{o}_{jk}\mathbf{o}_{jk}'\right)\Big|_{\boldsymbol{\pi}=\boldsymbol{\mu}} (\boldsymbol{\pi} - \boldsymbol{\mu}) \qquad \text{using (101)} \tag{103}$$

$$= -\mathbf{H}_\mu (\boldsymbol{\pi} - \boldsymbol{\mu}) \tag{104}$$

$$= \mathbf{H}_\mu (\boldsymbol{\mu} - \boldsymbol{\pi}) \qquad \# \tag{105}$$

This technical result is straightforward. A first order approximation is quite accurate in the region around the mean $\boldsymbol{\mu}$. Consideration of the underlying process suggests that points in the other, outlying regions will be rapidly pushed into the region where the approximation is good.

Let $\bar{\bar{\rho}}$ denote the harmonic mean of price coefficient ρ.

$$\bar{\bar{\rho}} = \left[E\left[\frac{1}{\rho}\right]\right]^{-1} \quad \text{ie} \quad \frac{1}{\bar{\bar{\rho}}} = E\left[\frac{1}{\rho}\right] \tag{106}$$

Lemma 9. The eigenvalues $\dot{\boldsymbol{\lambda}}$ and eigenvectors $\dot{\mathbf{v}}$ of the matrix $\mathbf{I} + \frac{dw}{\bar{\bar{\rho}}}\mathbf{H}_\mu$ are given by

$$\dot{\boldsymbol{\lambda}} = 1 + \frac{dw}{\bar{\bar{\rho}}}\boldsymbol{\lambda}_H \tag{107}$$

$$\dot{\mathbf{v}} = \mathbf{v}_H \tag{108}$$

where λ_H, $\mathbf{v}_H$ denote the eigenvalue and eigenvector of the estimation matrix $\mathbf{H}_\mu$.

Proof: $$\left[\mathbf{I} + \frac{dw}{\bar{\bar{\rho}}}\mathbf{H}_\mu\right] \cdot \boldsymbol{\Lambda}_H = \boldsymbol{\Lambda}_H \left[\mathbf{I} + \frac{dw}{\bar{\bar{\rho}}}\boldsymbol{\lambda}_H\right] \qquad \# \tag{109}$$

Lemma 10. The eigenvalues $\dot{\lambda}$ of the matrix $\mathbf{I} + \frac{dw}{\bar{\bar{\rho}}} \cdot \mathbf{H}_\mu$ fall in the range

$$0 < \dot{\lambda} < 1 \tag{110}$$

Proof: $$1 > -\frac{dw \cdot \lambda_H}{\bar{\bar{\rho}}} \qquad \text{rearranging (93), true for all } \lambda_H \text{, and noting } \bar{\bar{\rho}} < 0 \tag{111}$$

so $$0 < -\frac{dw \cdot \lambda_H}{\bar{\bar{\rho}}} < 1 \qquad \text{expression is positive} \tag{112}$$

$$1 > 1 + \frac{dw \cdot \lambda_H}{\bar{\bar{\rho}}} > 0 \qquad \text{multiply by -1, add 1} \tag{113}$$

$$0 < \dot{\lambda}^{\max} < 1 \qquad \text{by (107)} \qquad \# \tag{114}$$

It is now possible to prove the central result of this paper.

Theorem 11: the Efficient Market Theorem.

$$\lim_{n\to\infty} E(\boldsymbol{\pi}_n) = \boldsymbol{\mu} \tag{115}$$

Proof: $\mathbf{d\pi} = -\frac{dw}{\rho_0}\mathbf{H}\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right)$ Equation (50) (116)

$\boldsymbol{\mu} - \boldsymbol{\pi}_1 = \boldsymbol{\mu} - \boldsymbol{\pi}_0 + \frac{dw}{\rho_0}\mathbf{H}\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right)$ rearranging (117)

$\boldsymbol{\mu} - \boldsymbol{\pi}_1 \approx \boldsymbol{\mu} - \boldsymbol{\pi}_0 + \frac{dw}{\rho_0}\mathbf{H}_\mu\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right)$ applying Lemma 8 (118)

Taking expectations, ρ_0 is independent of $\boldsymbol{\pi}$ under Assumption 2 and $\mathbf{H}_\mu$ is constant.

$E\left[\boldsymbol{\mu} - \boldsymbol{\pi}_1\right] = E\left[\boldsymbol{\mu} - \boldsymbol{\pi}_0\right] + dw \cdot E\left[\frac{1}{\rho_0}\right]\mathbf{H}_\mu E\left[\boldsymbol{\mu} - \boldsymbol{\pi}_0\right]$ noting $E\left[\mathbf{e}_0\right] = \mathbf{0}$ (119)

$= \left(\mathbf{I} + \frac{dw}{\overline{\overline{\rho}}}\mathbf{H}_\mu\right)E\left[\boldsymbol{\mu} - \boldsymbol{\pi}_0\right]$ (107), factorizing (120)

$= \boldsymbol{\Lambda}\dot{\boldsymbol{\lambda}}\boldsymbol{\Lambda}' E\left[\boldsymbol{\mu} - \boldsymbol{\pi}_0\right]$ substituting $\boldsymbol{\Lambda}\dot{\boldsymbol{\lambda}}\boldsymbol{\Lambda}'$ for $\mathbf{I} + \frac{dw}{\overline{\overline{\rho}}}\mathbf{H}_\mu$ by Lemma 9 (121)

so $E\left[\boldsymbol{\mu} - \boldsymbol{\pi}_2\right] = \boldsymbol{\Lambda}\dot{\boldsymbol{\lambda}}^2\boldsymbol{\Lambda}' E\left[\boldsymbol{\mu} - \boldsymbol{\pi}_0\right]$ (122)

and result follows by Lemma 10. #

For a stock which has traded for a few months after its initial listing, this result can be restated more simply as

$E\left[\boldsymbol{\pi}\right] = \boldsymbol{\mu}$ (123)

The process of convergence is illustrated in the following diagram:

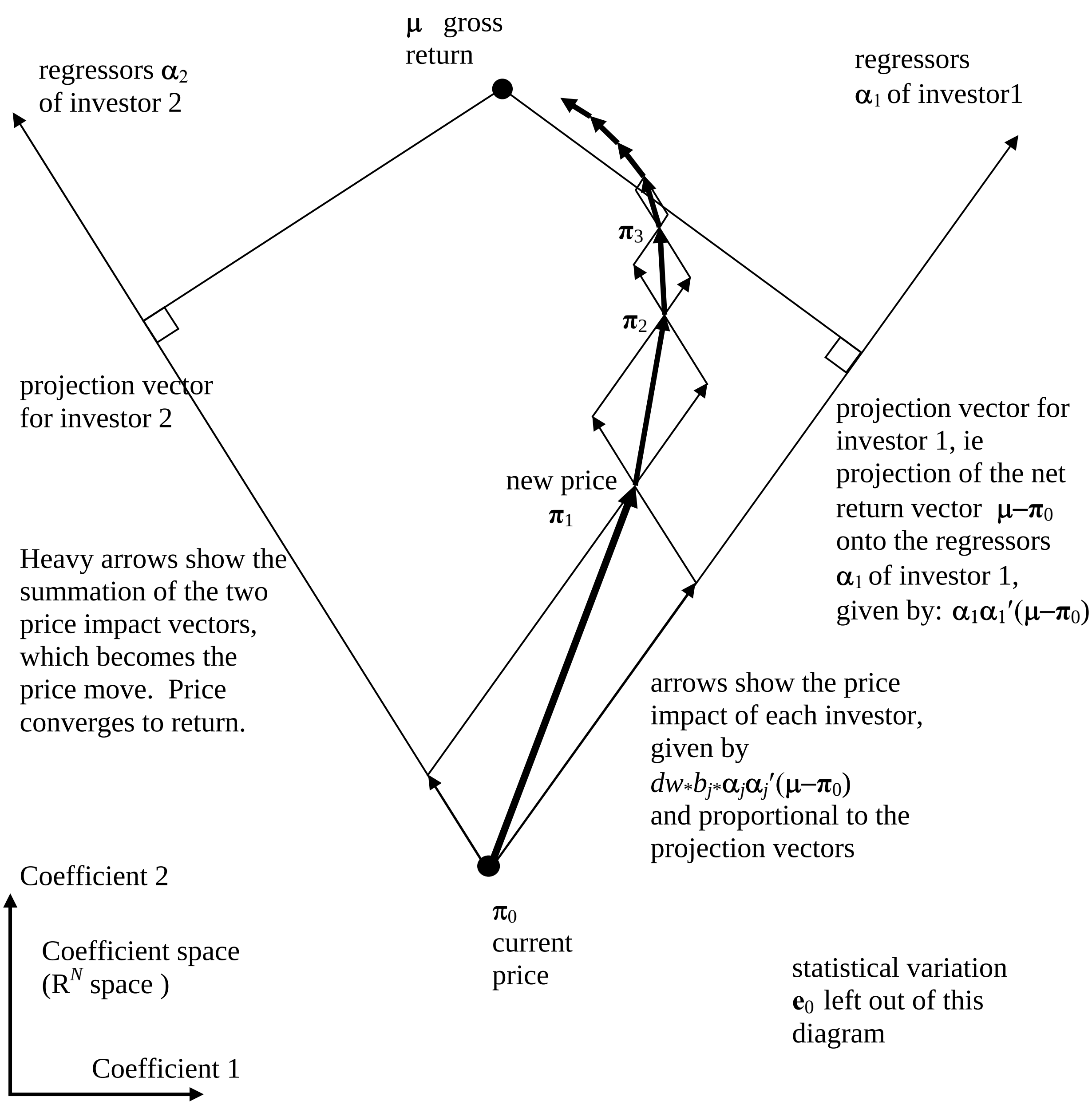

Fig. 1. Convergence of price to return in the coefficient model. Investor i regress the net return $\boldsymbol{\mu} - \boldsymbol{\pi}$ against a set of regressors $\boldsymbol{\alpha}_i$. This corresponds to the projection of the net return vector onto the regressor vector. The investor then buys according to this information, which moves the price along the line of the regressor – this is shown by the unbolded arrows. The total impact on price is the vector summation of the individual impacts of each investor and this is shown by the bold arrows. Successive iterations of this regression process move price along the path of the bold arrows to converge with return.

3.6. Simulation

As results depend on first order approximations the coefficient model was simulation tested. Return is assumed to be a function of two parameters, and there are two classes of investor. The first type look at data only and the second type look at price only.

		Investor 1: data only	Investor 2: price only
Relative proportion	b_j	0.5	0.5
Augmented strategy matrix	$\tilde{\mathbf{a}}_j$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\boldsymbol{\pi}_{2*1}$
Regression coefficient using (31)	$\hat{\boldsymbol{\beta}}_j$	$\hat{\boldsymbol{\beta}} = \boldsymbol{\mu}_{2*1} - \boldsymbol{\pi}_{2*1} + \mathbf{e}_{2*1}$	$\hat{\rho} = (\boldsymbol{\pi}'\boldsymbol{\pi})^{-1}\boldsymbol{\pi}'(\boldsymbol{\mu}_{2*1} - \boldsymbol{\pi}_{2*1} + \mathbf{e}_{2*1})$

Table 1: Characterization of investor types used in the simulation.

Estimates $\hat{\boldsymbol{\beta}}_1, \rho_1$ are updated as per (40),(49) and price is calculated by $\boldsymbol{\pi}_1 = -\dfrac{\hat{\boldsymbol{\beta}}_1}{\rho_1}$. There are no approximations in this stratagem. Typical results are shown in Figure 2. The price coefficient converges to a negative value providing the update proportion dw is kept low. This is necessary because of the low degrees of freedom $N = 2$ and is explained in Section 6 (high r_{error}). Parameters of the simulation shown below are:

$$\boldsymbol{\mu} = \begin{bmatrix} 0.2 \\ 0.3 \end{bmatrix};\ \boldsymbol{\sigma}_e = \begin{bmatrix} 0.10 \\ 0.05 \end{bmatrix};\ \hat{\boldsymbol{\beta}}_{x0} = \begin{bmatrix} 0.000012 \\ 0.000001 \end{bmatrix};\ \hat{\rho}_0 = -0.00002;\ \boldsymbol{\pi}_0 = \begin{bmatrix} 0.6 \\ 0.1 \end{bmatrix};\ dw = 0.000001$$

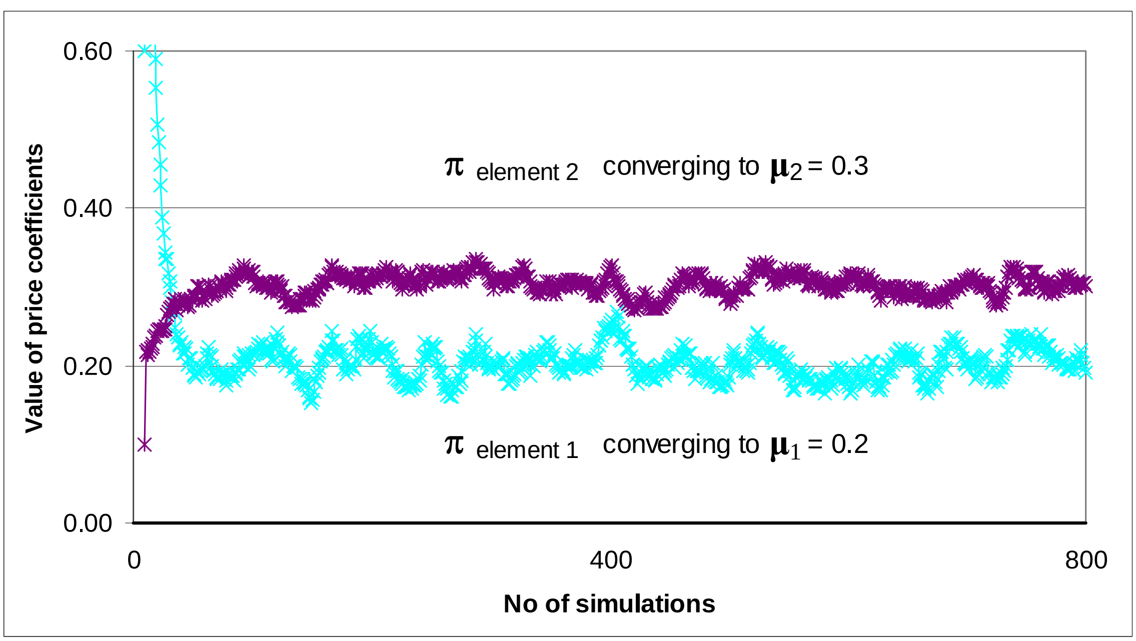

Fig. 2. Simulation showing convergence of price vector $\boldsymbol{\pi}$ to return vector $\boldsymbol{\mu}$.

4. The price coefficient (omitted)

5. The economics of the coefficient model (omitted)

6. Heterogeneous least squares learning as a genetic algorithm

Any process which uses a blind multipronged hill-climbing algorithm can be described as a genetic algorithm because this is the presumed mechanism of natural selection. We show that heterogeneous least squares learning is a genetic algorithm by constructing a genetic algorithm for a natural system and showing that it operates identically to heterogeneous least squares learning within the coefficient model.

The similarity of the processes does not depend on the profitability and survival of the economic actors, which is the traditional avenue for constructing an analogy of the 'economic Darwinism' type. Nor does it depend on the selection of particular strategies, which is the version of this concept found in modern rational expectations literature. For instance Marimon McGratton (1995) demonstrate an isomorphism between adaptive learning and evolutionary learning. The process described by the coefficient model will operate without any elimination of less profitable investors or strategies. It is the information processing itself – least squares learning – which is analogous to natural selection.

Let a creature *j*, i.e. a single organism, be represented by an *N**1 vector $\mathbf{g}_j$ in 'gene space'. Each gene represents a separate dimension of variation, variations are ranked in some physical order, and there are *N* separate genes so gene space is $\mathbb{R}^N$. It is a question whether genetic variation has a continuous character but we will ignore this. Let the population mean of the **g** vectors be denoted by $\boldsymbol{\pi}$, and the optimal genotype - the genotype which is superior to all others in reproductive capability - by $\boldsymbol{\mu}+\mathbf{e}$. Optimum fitness has a stochastic component, denoted **e**, which is set by prevailing environment conditions, the situation with predator/prey species etc. We suppose that the mean is sufficiently close to the optimum that reproductive capability is a smooth concave function of $\boldsymbol{\pi}$, with maximum at $\boldsymbol{\mu}+\mathbf{e}$. For the species this presents as a hill climbing problem.

Let the relative weight of creature *j* (the number of individuals with genotype $\mathbf{g}_j$) be denoted by b_j. At period 0

$$1=\sum_j b_j \tag{124}$$

$$\boldsymbol{\pi}_0=\sum_j b_j\mathbf{g}_j \tag{125}$$

Variation around the mean, referred to as the 'variation' vector, can be denoted by $\boldsymbol{\alpha}_j$:

$$\boldsymbol{\alpha}_j=\mathbf{g}_j-\boldsymbol{\pi}_0 \tag{126}$$

Observe $\sum_j b_j\boldsymbol{\alpha}_j=\mathbf{0}$ (127)

Natural selection premise 1. The fitness of each creature is directly proportional to its coordinate in the direction of the optimum, $(\boldsymbol{\mu}+\mathbf{e})-\boldsymbol{\pi}$. This vector, referred to as the fitness vector, is denoted **r**.

Natural selection premise 2. The Net Reproduction Rate (*NRR*) of a creature (i.e. production of offspring after replacing itself) is given by the fitness of the creature relative to other members of the species.

Applying these two assumptions yields:

$$NRR_j = w\left(creature\ \ fitness - average\ \ fitness\right) = w(\mathbf{g}_j{}'\mathbf{r} - \boldsymbol{\pi}_0{}'\mathbf{r}) = w\boldsymbol{\alpha}_j{}'\mathbf{r} \tag{128}$$

where w is some positive constant. *NRR* will be positive for some creatures and negative for others; on average it is zero as shown in (131). The Gross Reproduction Rate (*GRR*), which includes the replacement creature, is given by

$$GRR_j = NRR_j + 1 = w\boldsymbol{\alpha}_j{}'\mathbf{r} + 1 \tag{129}$$

so the weighting of descendents produced by creatures of type j will be given by

$$b_{j\ descendents} = b_j \cdot GRR_j = b_j\left(1 + w\boldsymbol{\alpha}_j{}'\mathbf{r}\right) \tag{130}$$

Observe that the total weighting attributed to descendents is unity:

$$\sum_j b_{j\ descendents} = \sum_j b_j\left(1 + w\boldsymbol{\alpha}_j{}'\mathbf{r}\right) = \sum_j b_j + w\left(\sum_j b_j\boldsymbol{\alpha}_j{}'\right)\mathbf{r} = 1 \quad \text{by (124),(127)} \tag{131}$$

so these weights $b_{j\ descendents}$ can be used for the purpose of finding the new population mean in period 1.

Natural selection premise 3. The average genotype $\overline{\mathbf{g}_{ij}}$ of the descendents i of creature j equals the genotype of the parent, i.e.

$$\mathbf{g}_{ij} = \mathbf{g}_j + \mathbf{u}_i \qquad \text{where } \mathbf{u}_i \text{ is an error term} \tag{132}$$

$$\overline{\mathbf{g}_{ij}} = \frac{\sum_i \mathbf{g}_{ij}}{GRR_j} = \mathbf{g}_j \tag{133}$$

Result 33: genetic algorithm theorem. Heterogeneous least squares learning is a genetic algorithm.

Proof: The above three assumptions define a genetic algorithm. We proceed by showing that this algorithm embodies the same process as heterogeneous least squares learning. Evaluate the population mean in period 1, $\boldsymbol{\pi}_1$.

$$\boldsymbol{\pi}_1 = \sum_j \overline{\mathbf{g}_{ij}} \cdot b_{j\ descendents} \tag{134}$$

$$= \sum_j \left(\boldsymbol{\pi}_0 + \boldsymbol{\alpha}_j\right) \cdot b_j\left(1 + w\boldsymbol{\alpha}_j{}'\mathbf{r}\right) \qquad \text{by (133),(126),(130)} \tag{135}$$

$$= \sum_j b_j\boldsymbol{\pi}_0 + \sum_j \boldsymbol{\pi}_0 b_j w\boldsymbol{\alpha}_j{}'\mathbf{r} + \sum_j \boldsymbol{\alpha}_j b_j + \sum_j \boldsymbol{\alpha}_j b_j w\boldsymbol{\alpha}_j{}'\mathbf{r} \tag{136}$$

$$= \sum_j b_j\boldsymbol{\pi}_0 + \mathbf{0} + \mathbf{0} + w\sum_j b_j\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j{}'\mathbf{r} \qquad \text{by (127)} \tag{137}$$

$$= \boldsymbol{\pi}_0 + w\mathbf{H}\left(\boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0\right) \tag{138}$$

where $\mathbf{H} = \sum_j b_j\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j{}'$ (139)

and $\mathbf{r} = \boldsymbol{\mu} - \boldsymbol{\pi}_0 + \mathbf{e}_0$ (140)

Effectively each creature is forming a regression coefficient $\boldsymbol{\alpha}_j{}'(\boldsymbol{\mu}-\boldsymbol{\pi}_0)$ which is applied to explanatory variable $\boldsymbol{\alpha}_j$ to yield prediction vector $\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j{}'(\boldsymbol{\mu}-\boldsymbol{\pi}_0)$. Matrix **H**, the weighted sum of the projection matrices $\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j{}'$, is cognate with the estimation matrix **H** defined at (74).

Now this expression can be rearranged and expectations taken to give

$$E[\boldsymbol{\mu}-\boldsymbol{\pi}_1]=E[\boldsymbol{\mu}-\boldsymbol{\pi}_0]-w\mathbf{H}\cdot E[\boldsymbol{\mu}-\boldsymbol{\pi}_0] \tag{141}$$

$$=(\mathbf{I}-w\mathbf{H})\cdot E[\boldsymbol{\mu}-\boldsymbol{\pi}_0] \quad \text{which conforms with (120)} \qquad \# \tag{142}$$

One apparent difference is that the vectors $\boldsymbol{\alpha}_j$ are normalized in the financial model, i.e. $\boldsymbol{\alpha}_j{}'\boldsymbol{\alpha}_j=\mathbf{I}$ whereas this is not the case for the genetic algorithm; but the scaling of genetic variation is in any case arbitrary. The estimation matrix **H** is a genetic operator – it takes the system state (price, mean genotype) and an observation of the environment (return, mean fitness) and produces the next state.

The following figures and table examine the correspondence between financial and natural systems for its own interest.

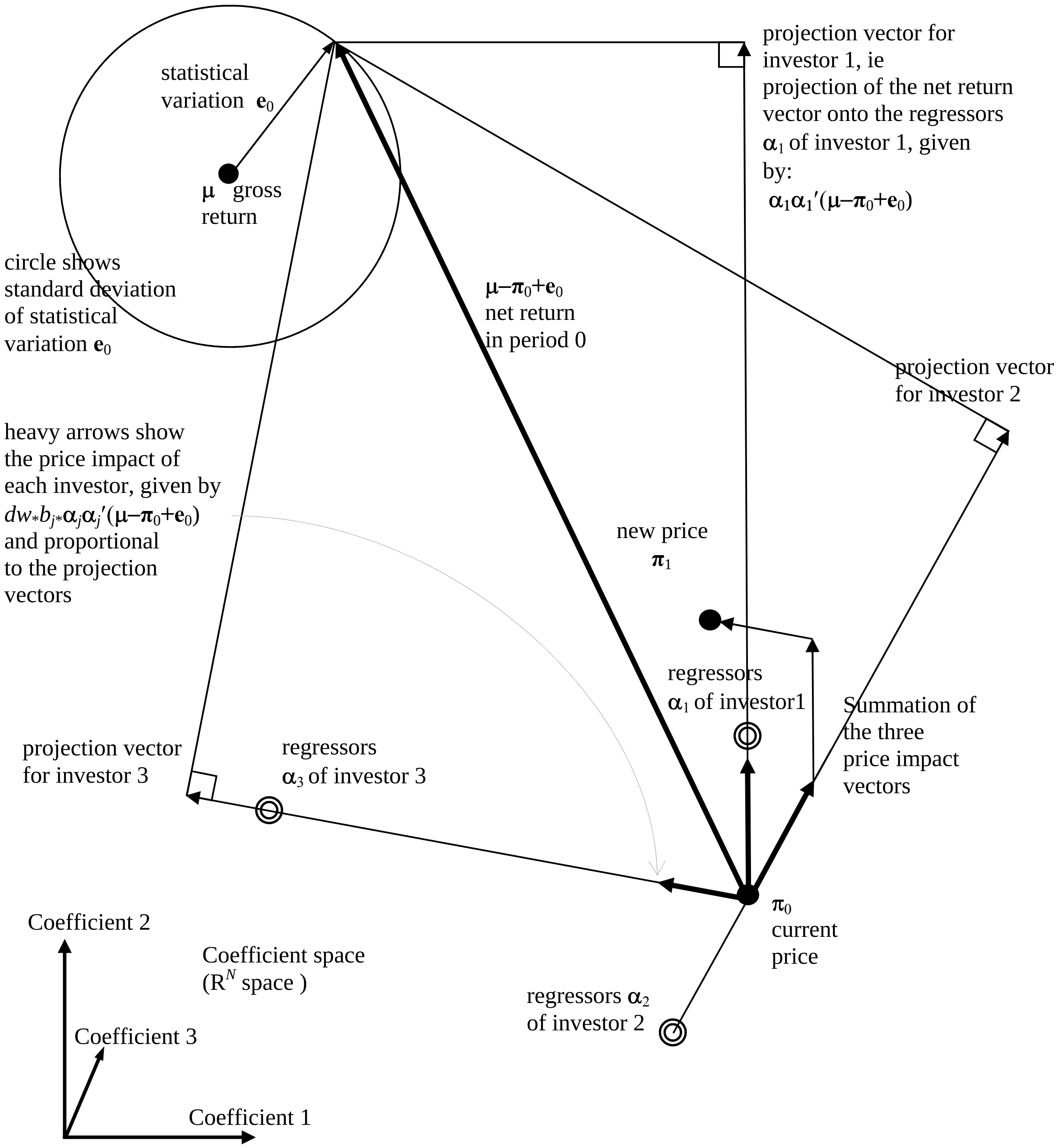

Fig. 6. **Coefficient model applied to finance**. The market is initially at price $\boldsymbol{\pi}_0$, the lowest ● symbol. Each investor attempts to find a profitable strategy by regressing the net return $\boldsymbol{\mu}+\mathbf{e}_0-\boldsymbol{\pi}_0$ onto particular regressor vectors $\boldsymbol{\alpha}$ which they have chosen, shown by the target symbols ◎. This yields a predicted return given by the geometric projection of the net return vector onto the regressors. The consequent impact of the investor's position on the price is in the direction of the regressors $\boldsymbol{\alpha}$ and proportional to this geometric projection: the impact is shown by the heavy arrows. Effect is to move the price to $\boldsymbol{\pi}_1$ in the next generation and ultimately to the gross return vector $\boldsymbol{\mu}$.

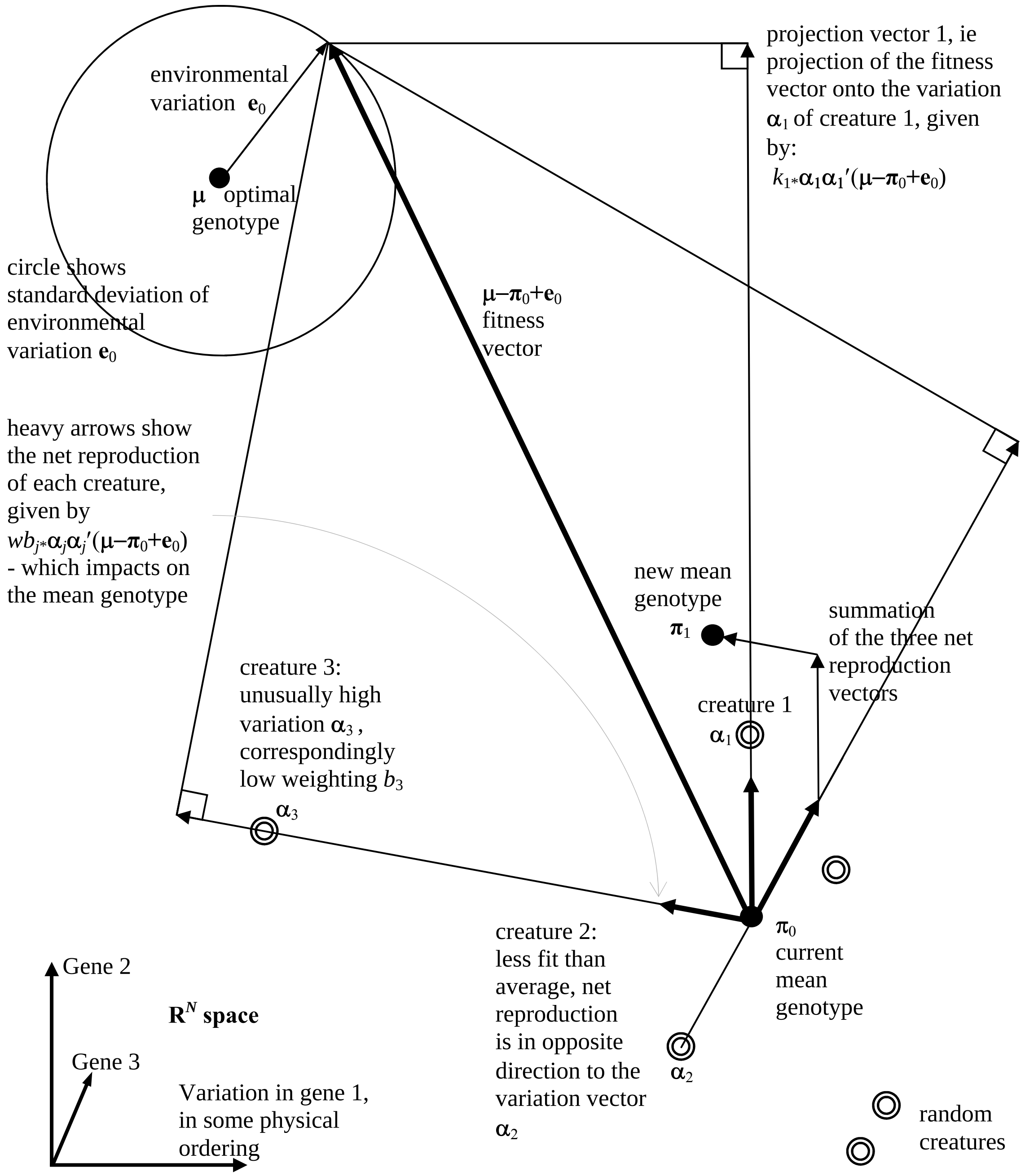

Fig. 7. **Coefficient model applied to natural selection**. The species initially has mean genotype π_0, the lowest ● symbol, and individual creatures are shown by the target symbols ◎. Each creature's fitness is determined by the covariance of its genetic variation α and the fitness vector $\mu + e_0 - \pi_0$, suggested by geometric projection of the fitness vector onto the variation line α. Consequent net reproduction of the creature perpetuates its variation α, but is proportional to fitness: net reproduction is shown by the heavy arrows. Effect is to move the mean genotype to π_1 in the next generation and ultimately to optimal fitness μ.

Table 6: Correspondence of the elements in financial and natural systems.

	FINANCE	NATURAL SELECTION
STARTING POINT	*Price* Price is a vector $\boldsymbol{\pi}_0$ in 'coefficient space', not a single piece of information; it stores all the information which generates an accurate price.	*Mean genotype* The mean genotype is a vector $\boldsymbol{\pi}_0$ in 'gene space'**.** A creature is a function, not a single piece of information, formed by all the information in its gene vector (set of chromosomes).
CREATION OF VARIATION	Different variables are selected by actors for investigation. Each actor *j* relies on the price to explain the majority of return, and chooses only a small number of variables $\boldsymbol{\alpha}_j$ for further investigation in a process of econometric *specification*.	Different genotypes are created by genetic crossover, i.e. for sexually reproducing species this comprises meiosis & fertilization. Different creatures *j* show variation along every genetic axis to produce genotype $\mathbf{g}_j$ with variation $\boldsymbol{\alpha}_j$ from the mean.
TEST OF VARIATION	Set of statistical *trials i* forming an *estimation period*, in which security returns and explanatory variable values are generated.	The creature's efforts to reproduce represent separate statistical *trials*.
EVALUATION OF TEST	*Estimation* i.e. regression to determine new estimates.	*Natural selection*, i.e. differential rates of survival. As shown above, this is equivalent to an estimation by regression.
TEST ERROR	The accuracy of estimates is affected by error (statistical variation), denoted **e** in coefficient space.	Creature lives are affected by random variations – different environmental conditions, different numbers of predators and prey at various times etc. (The role of chance has received attention in ecological literature in recent years.)
RECOMBINATION	Intermixing of the new estimates and the old through a process of price formation to generate a new *price* $\boldsymbol{\pi}_1$.	Intermixing of members of the next generation of the population to produce a new *mean genotype* $\boldsymbol{\pi}_1$.

7. The multiperiod coefficient model (omitted)

8. Conclusion

8.1. The nature of financial market processes: heterogeneous least squares learning

Typically studies of financial market equilibrium view security valuation as a black box process, following Grossman Stiglitz. Valuation is carried out by a subset of the traders, fundamental analysts, and it is neither possible nor necessary to investigate the process further. An alternative paradigm is offered here whereby instead of one source of information which is costly to obtain, there are many sources of information which yield various returns. From this premise it has been shown that there is an economic process at work which drives financial markets to a fully informed equilibrium even though each actor possesses only part of the information. Agents do not "add together" information to arrive at a value for a security. Rather they look at security returns and "subtract", or back out, factors which they believe are relevant.

An engineer building a computer uses an additive process to create something from first principles. Perhaps this analogy has led writers to assume that "financial engineers" carrying out comprehensive analysis with all the available information must form the foundation of a financial market. It is demonstrated here that this need not be the case.

8.2. Economics of the efficient market and the Grossman-Stiglitz paradox

The properties exhibited by the market in the coefficient model suggest that efficiency is a journey rather than a destination. It is the expected value of price which equals return: $E[\boldsymbol{\pi}] = \boldsymbol{\mu}$ rather than price itself which equals return: $\boldsymbol{\pi} = \boldsymbol{\mu}$. The process by which price gravitates to return is a stochastic one in which investors are rewarded according to their contributions. Specifically, the return received by each investor is determined by:

- The distance of the current estimate of value $\boldsymbol{\pi}$ from the true value $\boldsymbol{\mu}$
- The relevance R^2 of the investor's information
- The weight of money b_i backing that information
- The sampling error in the estimate derived from the information
- Whether the estimate is up to date or not (new or old)

and the returns may be positive or negative depending on these factors. All of these factors could be predicted on a priori grounds, and all are related to profit as one would expect intuitively – with the one exception that sometimes *new* estimates lose money because of statistical error.

A cyclical process has been identified whereby the return to new estimates will be negative on average when price, the market's current estimate of value, $\boldsymbol{\pi}$ approaches too closely to the return $\boldsymbol{\mu}$. If this estimate $\boldsymbol{\pi}$ drifts out again as a result of discouraged analysts leaving the market, return will increase. The process is in line with the standard microeconomic principle that markets in equilibrium yields normal returns. All investors contribute to processing information and are rewarded according to their contribution.

8.3 The coefficient superstrate and the reinterpretation of price

When the original data model is reduced to coefficient space every variable and process takes a new form. The representation of the least squares learning process in coefficient space is simpler and the behaviour of the market is more intuitive. Coefficient space can be understood as a hidden superstrate which constructs the market. This expression 'constructs the market' can be given a formal meaning:

In any well-formed explanation of why the market price tends towards the realized return in a heterogeneous least squares learning market, the data **X** *will drop out and we will be left with the coefficient representation.*

Although the word 'substrate' is used in this paper's Abstract, the less familiar term 'superstrate' is more accurate in describing the concept of the coefficient layer. 'Constructs the market' is used rather than 'determines market behaviour' because from a causal point of view it is the investors operating at the data level who determine market behaviour. Notwithstanding, this behaviour can only be understood at the superstrate level.

Viewed within the superstrate, price is not a single piece of information but a collection of information which behaves like a computer memory. It stores a valuation formula, supplies it to investors, and updates. The process by which price gravitates to the return can be regarded as emergent behaviour – it is not operating on the data level which is seen by the participants nor are they aware of it.

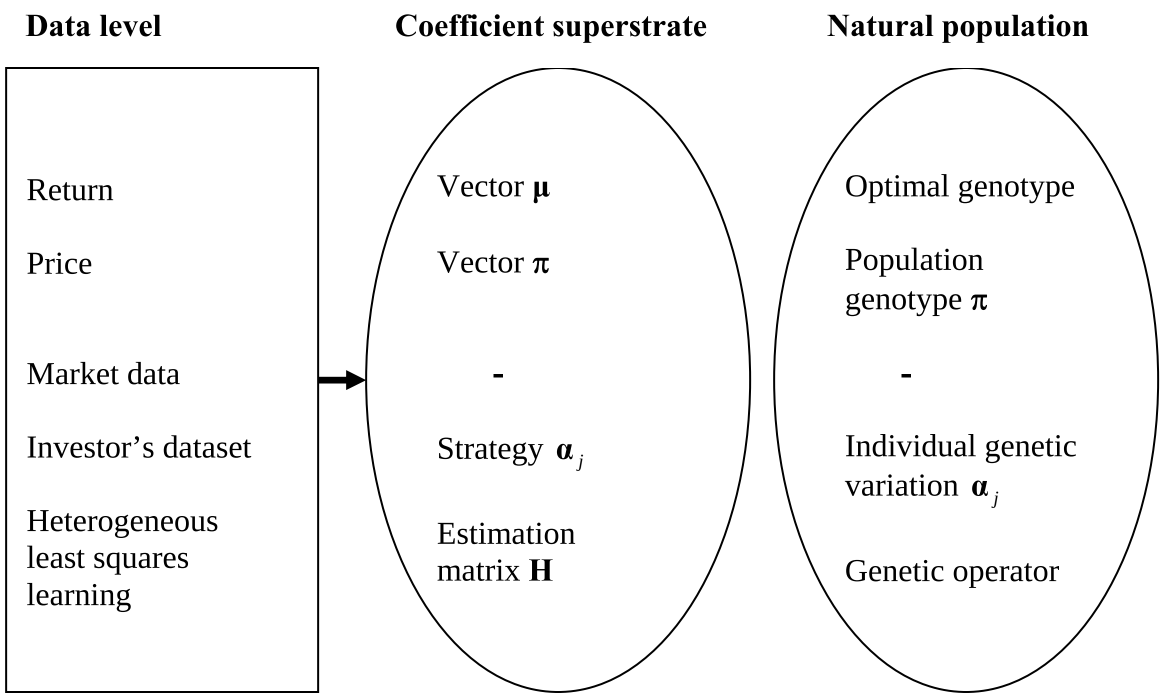

Fig. 8: Shows the correspondences between the 'data level' which the economic actors see, the coefficient superstrate to which market processes can be reduced, and the corresponding entities in a natural population. There is a 1-1 mapping between market and genetic elements.

In the superstrate the way in which the market works corresponds to natural selection in biology. Price corresponds to the average genotype. It stores a complete valuation formula, just as a set of genes provides a full blueprint for constructing an organism. Heterogeneous least squares learning is a genetic operator which performs selection and recombination. The correspondence may be seen as an analogy or isomorphism, but the stronger interpretation is that the same underlying process is expressed in two different self-organizing systems.

Because price is not a single value which is immediately dependent on demand and supply conditions but a vector with a persistent value and genetic properties, we can think of it having an independent existence as an economic entity. This reification of price puts analytical clothing on the bones of Dawkin's (1976) meme concept – purportedly an equivalent of the gene in social systems.

It follows from the independent existence of price that there are endogenous reasons why the market will move around, quite apart from exogenous movement due to news. Two separate cycles have been identified:

- Short cycles as price $\boldsymbol{\pi}$ rotates around return $\boldsymbol{\mu}$.
- Longer cycles as fundamental investors move in and out of the market, as per the discussion on profitability

These gyrations in the absence of a change in underlying value suggest the market will exhibit mean reversion and may provide a theoretical basis for that army of practitioners who claim to discern such patterns through 'technical analysis'. The implications of technical analysis are beyond the scope of this paper which as the title states is concerned only with fundamental analysis.

8.4. The fragile nature of financial market processes

The process of forming a price in financial markets is inherently fragile because it relies on the generation of a negative price coefficient to act as denominator, and it turns out that the price coefficient is close to zero. It is only a fortuitous interplay of parameters which allows markets to operate at all, and it is not surprising that markets are susceptible to occasional malfunction for a variety of different reasons. Financial markets and natural selection may both use a genetic algorithm, but one difference between the two is that (I assume) natural selection is more stable because it does not require a denominator.

In foreign exchange (fx) trading there are two frames of reference – the two different currencies. In this case the argument for the negativity of the price coefficient of the security may break down because investors are estimating the price coefficient from two different sides. This may explain why fx markets often do not seem to reflect fundamentals but trade perversely: their mechanism is fundamentally flawed.

Four mechanisms for market bubbles have been identified. The first three arise as natural consequences of coefficient theory rather than a conscious attempt to model bubbles. Market efficiency and market bubbles can be regarded as two sides of the same coin: if the market is not heading to the efficient point it is heading to infinity.

- the stability condition of the price convergence process is not satisfied, because the frequency with which investors update their estimates, measured by dw, is too great.

$$dw > \sigma_X^2 \left(\frac{b_{price}}{\lambda_P} \right) \left(\frac{\lambda_X}{\lambda_P} \right)^2 \tag{143}$$

- There is insufficient variance in the component of return which is orthogonal to the true value of return to generate a negative value for the price coefficient. A rule of thumb derived from simulation testing is that instability sets in where:

$$r_{error} = \frac{\sigma_{error\ y}}{\sigma_{error\ x}} > 20\% \tag{144}$$

- the rate of growth in the data is greater than the discount rate. The result

$$E[\boldsymbol{\pi}] = \frac{\boldsymbol{\mu}}{r - g} \tag{145}$$

  picks up the actual rate of growth in the data *g* rather than the expected rate of growth in the dividends, so if growth *g* exceeds the interest rate *r* then the price will increase without limit.
- price increases may lead to even greater subsequent increases if the accelerator factor Q is greater than unity.

$$Q_0 = \frac{dw \cdot \lambda_H}{\rho_0 T} > 1 \tag{146}$$

This factor depends on the update proportion, the maximal eigenvalue of the estimation matrix, the price coefficient and the sample length. All of these will change so as to increase accelerator Q in a bubble environment.

In practice these situations are particularly likely for startup issues which are not yet making a profit in a time of buoyant economic growth. There is too little non-price information available and investors are concentrating on price alone. Three famous historical bubbles fall into this category: the Dutch tulip mania of 1636-37, the South Sea bubble of 1720, and Dot-com mania in 2000.

8.5. Summary- the investor's life in the bush of ghosts

A model is constructed for heterogeneous least squares learning where market return is expressed in terms of data coefficients. It has been shown how a market can converge to an efficient equilibrium, and why the price coefficient will be negative. These results have been extended to a multiperiod model and an explanation of market bubbles from first principles has been presented.

Within the model, investor strategies support each other in a self-sustaining equilibrium. Each investor relies on the pattern of investment of the other investors for the accuracy of their own estimations, and no part can be removed without disrupting the other parts. Notwithstanding the system is "reducibly complex" – it can bootstrap itself up over time from a low information starting point to a situation characterized by specialization and dependency. The mechanism may offer clues to the origin of other complex systems.

One of the oldest and most important themes in economics, going back to Adam Smith's "invisible hand", is that the whole is greater than the sum of the parts. The coefficient model locates the parts which have previously been obscured. Behind the outward appearance of a market – the trades and the current market price – is a hidden superstrate, in which price is not a single piece of information but a vector with its own independent existence as an economic entity. The vector stores information and makes it available to traders in a same way as a gene in biology (a 'meme'). Heterogeneous least squares learning is a genetic operator which moves price to the point of optimum explanatory power. The gravitation of price to the efficient point is emergent behaviour in that it cannot be determined from the information initially possessed by the investors. The process is as invisible to the investors as natural selection is to creatures.

References

Brock W., Hommes C. 1997. A rational route to randomness. Econometrica 65, pp. 1235-1274.

Brock W., Hommes C. 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control 22, pp. 1235-1274.

Chiarella C.,Gallegati M., Leombruni R., Palestrini A. 2003. Asset price dynamics among heterogeneous agents. Computational Economics 22, pp. 213-223.

Dawkins R. 1976. The Selfish Gene. Oxford University Press, Oxford, Ch 11.

Fama E. 1970. Efficient capital markets: a review of theory and empirical work. Journal of Finance 25, pp. 383–417.

Goldbaum D. 2005, Market efficiency and learning in an endogenously unstable environment. Journal of Economic Dynamics and Control 29(5), pp. 953-978.

Grossman S. 1976. On the efficiency of competitive stock markets where traders have diverse information. Journal of Finance 31(2), pp. 573-85.

Grossman S., Stiglitz J. 1980, On the impossibility of informationally efficient markets. American Economic Review 70(3), pp. 393-408.

Levy M., Levy H. 1996. The danger of assuming homogeneous expectations. Financial Analysts Journal 52, pp. 65-70.

Marimon R., McGrattan E., 1995. On adaptive learning in strategic games. In: Kirman A., Salmon M. (eds), Learning and Rationality in Economics, Basil Blackwell, Oxford, pp 63-101.

Muendler M. 2005. Rational information choice in financial market equilibrium. SESIFO Working Paper No. 1436.